2018

Safety Investigation of Traffic Crashes Incorporating Spatial Correlation Effects

Khalid Alkahtani
University of Central Florida

SAFETY INVESTIGATION OF TRAFFIC CRASHES INCORPORATING SPATIAL CORRELATION EFFECTS

by

KHALID FALEH ALKAHTANI

B.Sc., King Saud University, Saudi Arabia, 2007

M.Eng.Sc., University of New South Wales, Australia, 2011

M.Sc., University of Central Florida, United States, 2016

A dissertation submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

in the Department of Civil, Environmental and Construction Engineering

in the College of Engineering and Computer Science

at University of Central Florida

Orlando, Florida

Summer Term

2018

Major Professor: Mohamed A. Abdel-Aty

ABSTRACT

One main interest in crash frequency modeling is to predict crash counts over a spatial domain of interest (e.g., traffic analysis zones (TAZs)). The macro-level crash prediction models can assist transportation planners with a comprehensive perspective to consider safety in the long-range transportation planning process. Most of the previous studies that have examined traffic crashes at the macro-level are related to high-income countries, whereas there is a lack of similar studies among lower- and middle-income countries where most road traffic deaths (90%) occur. This includes Middle Eastern countries, necessitating a thorough investigation and diagnosis of the issues and factors instigating traffic crashes in the region in order to reduce these serious traffic crashes. Since pedestrians are more vulnerable to traffic crashes compared to other road users, especially in this region, a safety investigation of pedestrian crashes is crucial to improving traffic safety. Riyadh, Saudi Arabia, which is one of the largest Middle East metropolises, is used as an example to reflect the representation of these countries' characteristics, where Saudi Arabia has a rather distinct situation in that it is considered a high-income country, and yet it has the highest rate of traffic fatalities compared to their high-income counterparts. Therefore, in this research, several statistical methods are used to investigate the association between traffic crash frequency and contributing factors of crash data, which are characterized by 1) geographical referencing (i.e., observed at specific locations) or spatially varying over geographic units when modeled; 2) correlation between different response variables (e.g., crash counts by severity or type levels); and 3) temporally correlated. A Bayesian multivariate spatial model is developed for predicting crash counts by severity and type. Therefore, based on the findings of this study, policy makers would be able to suggest appropriate safety countermeasures for each type of crash in each zone.

ACKNOWLEDGMENT

All the praises and thanks be to the Almighty Allah for His blessing in completing this dissertation.

I would like to express my sincere thanks and appreciation to my advisor, Dr. Mohamed Abdel-Aty, for his continuous guidance and help. His encouragement and constant motivation and inspiration pushed me to do better everytime.

I would also like to thank the dissertation committee members, Dr. Essam Radwan, Dr. Naveen Eluru, Dr. Qipeng Phil Zheng, and Dr. Jaeyoung Lee for their invaluable discussion and feedback.

I wish to thank The High Commission for the Development of Riyadh (HCDR), The General Directorate of Traffic, The General Department of Studies and Designs in the Riyadh municipality and NAJM company for providing the data. Also, thank you to those who have helped me in these departments.

I appreciate King Saud University for providing the financial support to pursue my postgraduate studies. Also, thank you for the college of engineering and department of civil engineering for helping me throughout my journey to obtain the data.

I owe a debt of gratitude to my family and best friends and colleagues for their countless contribution to my success.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
BCI	Bayesian Credible Interval
BG	Block Group
BIC	Bayesian Information Criterion
CAR	Conditional Autoregressive
CB	Census Block
CCD	Census County Division
CT	Census Tract
CV	Cross Validation
DIC	Deviance Information Criterion
FB	Full Bayesian
GLM	Generalized Linear Model
GMCAR	Generalized Multivariate Conditional Autoregressive Model (CAR)
GWNBR	Geographically Weighted Negative Binomial Regression
GWPR	Geographically Weighted Poisson Regression
GWR	Geographically Weighted Regression

HAY	Zonal system
HCDR	High Commission for the Development of Riyadh
HICs	High-Income Countries
ICAR	Intrinsic Conditional Autoregressive
KSA	Kingdom of Saudi Arabia
LMICs	Lower- and Middle- Income Countries
MAD	Mean Absolute Deviation
MCMC	Markov chain Monte Carlo
MTR	Mass Transit Railway
MVPLN-CAR	Multivariate Poisson-lognormal Conditional Autoregressive
NB	Negative Binomial
PDO	Property Damage Only
PG	Poisson-gamma
PLN	Poisson-lognormal
PLN-CAR	Univariate Poisson-lognormal Conditional Autoregressive
PSI	Potential for Safety Improvements
RE-NB	Random Effects Negative Binomial

RMSE	Root Mean Squared Errors
RP	Random Parameter
RP-NB	Random Parameter Negative Binomial
RP-P	Random Parameter Poisson
SAS	Statistical Analysis System
S-GWPR	semi-parametric geographically weighted Poisson regression
STAZ	Statewide Traffic Analysis Zone
SVC	Spatially Varying Coefficients
TAD	Traffic Analysis District
TAZ	Traffic Analysis Zone
TSAZ	Traffic Safety Analysis Zone
VIF	Variance Inflation Factor
VKT	Vehicle-Kilometers-Traveled
WHO	World Health Organization
ZCTA	ZIP-Code Tabulation Area

CHAPTER 1: INTRODUCTION

1.1 Research Motivation and Problem Statement

Traffic crashes are considered as one of the top ten contributors to the total death toll in the world. By 2030, it is expected to be one of the top five causes of death (WHO, 2015a). A report published by the World Health Organization (WHO) envisages that the road traffic crash death toll will rise from the 1.3 million reported in 2004 to 2.4 million by 2030. Such a huge increase is believed to be as a result of the consistently increasing numbers of vehicle ownership and the vehicle use associated with economic growth in developing countries.

Furthermore, more than one-third of road traffic fatalities in developing countries were mostly pedestrian and cyclist related. Due to the ever-increasing demand for both travel and transportation related developmental infrastructures, ensuring the efficient movement and safety of road users has proven to be challenging for governments. In addition, the aftermath of increasing vehicular operations, such as traffic congestion, air pollution, and oil dependency, have worsened the situation.

These alarming transportation-related issues have resulted in more governmental sanctions with a view to implementing more sustainable solutions. This has led governments in many countries to initiate policies that encourage walking and cycling because they are sustainable, inexpensive, environmentally friendly and, most importantly, they curb the alarming transportation-related issues. In communities that aim to be sustainable, walking has been recognized as one of the best active modes of transportation (Babalik-Sutcliffe, 2013; Loo and Chow, 2008). However, high crash rates of pedestrians have been a deterrent to choosing walking as a major mode of transportation. Thus, one of the ways to promote walking is to ensure the safety of pedestrians.

According to the WHO (WHO, 2015a), nearly 50% of registered motorized vehicles are in high-income countries (HICs), and only 10% of road traffic deaths take place there. The Kingdom of Saudi Arabia (KSA) has a rather distinct situation where it is considered a high-income country, and yet it has the highest rate of traffic fatalities compared to their high-income counterparts. The road traffic death burden (per 100,000 population) in the KSA was 65% higher than the average of the world's road traffic deaths and 3.5 times the average of HICs in 2015 (WHO, 2015b). In the context of neighboring countries with which the KSA shares many characteristics, the KSA's rate of traffic fatalities is 75% higher than the average of other Gulf countries. Based on these estimates, the KSA ranked 157 out of 180 countries in terms of road safety, with a rate of 27 fatalities per hundred thousand of the population.

The rapid increase in population and continued economic growth throughout the KSA has led to an enormous increase in the number of vehicles, which has consequently led to a high level of traffic crashes. According to the National Strategy for Traffic Safety (2014), the number of traffic crashes increased by 92% between 2006 and 2011 to reach 544,179 crashes in 2011. In that same period, the percentage of registered vehicles increased by 42% to reach 9.7 million vehicles; this is in contrast to the total population, which recorded only a 14% increase. Moreover, more than 20 fatalities per day can be directly linked to traffic crashes, and to every fatal crash, there are at least two permanent disabilities. In fact, statistics have shown that 73% of the total fatalities due to traffic crashes in the KSA were abled individuals below the age of 40, which makes up a large percentage of the nation's active and productive adult population. This makes this ever-increasing fatality rate a serious issue that requires national attention and interventions to improve road safety. Correspondingly, road safety has become increasingly important over the last few years in the KSA. An instance of drastic measures taken by the government is The National Transformation Program 2020, whose vision related to traffic crashes is to reduce the traffic fatality rate by 25% by 2020 (Saudi Arabia's Vision 2030, 2016).

Saudi Arabia is divided into 13 provinces, and its second-largest province overall is the Riyadh Province, as shown in Figure 1-1. Riyadh is the capital city of the Riyadh Province and the largest area in the region (5,961 km^2); it is also one of the largest Middle East metropolises.

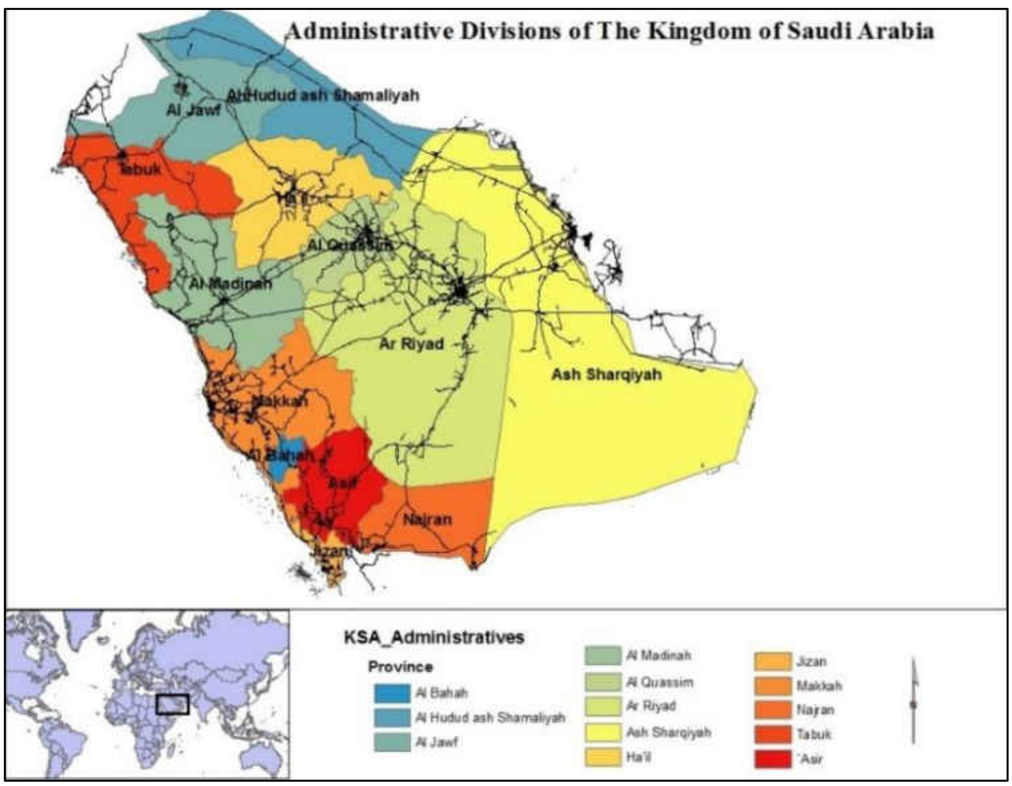


Figure 1-1: Provinces of the Kingdom of Saudi Arabia

The rapid population and economic growth has led to an increase in travel demands, which, as a result, led to a high rate of traffic crashes. According to the Central Department of Statistics and Information (2010), Riyadh's population in 2010 was rounded up to approximately 5,339,400 and is expected to rise to seven million by the year 2021. In 2013, there were 5.7 million people in Riyadh, with 29% being non-Saudi. In addition, Riyadh has an estimated road network length of 13,850 km, which has 7.4 million daily trips (Central Department of Statistics and Information, 2010). Figure 1-2 displays the distributions of the total number of traffic crashes in the three most heavily populated regions in the KSA (i.e., Makkah, Riyadh, and Eastern Regions) between 1997 and 2012. During this period, more than 5.5 million traffic crashes occurred in the KSA, with 26.4% of them occurring in Riyadh (Figure 1-2) (Directorate General of Traffic, 2013). This implies that over one-fourth of road traffic crashes in the KSA occurred in Riyadh. Furthermore, the total number of crashes is relatively high within the Riyadh Region, particularly in the last few years.

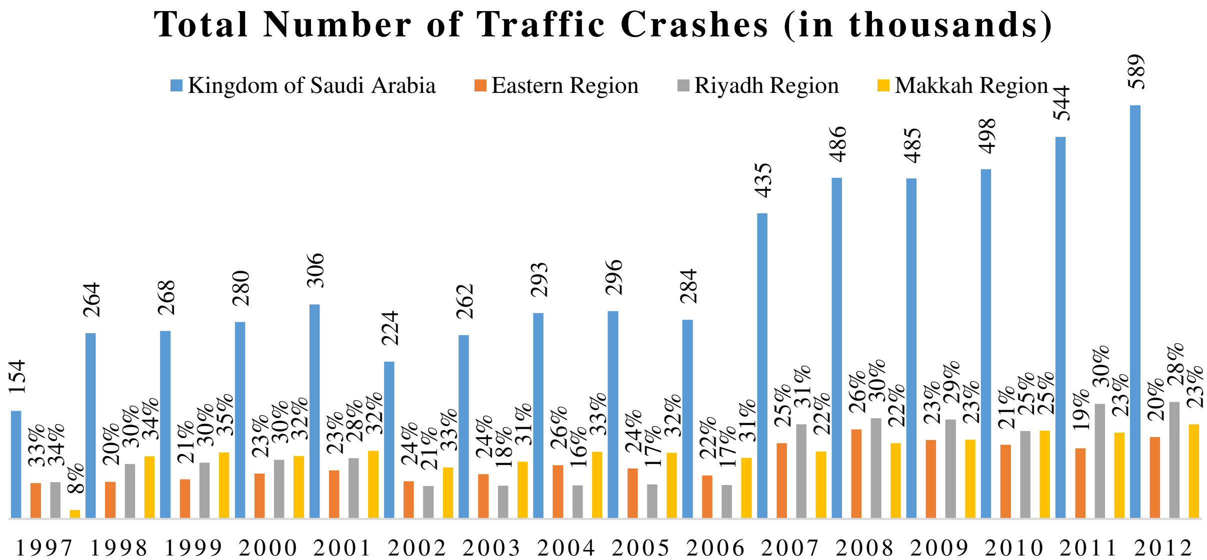


Figure 1-2: Distributions of the total number of traffic crashes in Saudi Arabia and its proportions of Eastern, Riyadh, and Makkah Regions between 1997 and 2012

Also, between 1997 and 2012, around 60,000 people were either killed or injured due to road traffic crashes in Riyadh (Figure 1-3) (Directorate General of Traffic, 2013). In 2015, road traffic crashes caused 667 fatalities; 175 of them were pedestrians. This implies that over one-fourth of road traffic fatalities were pedestrians, necessitating a thorough investigation and diagnosis of the issues and factors instigating traffic crashes in the region in order to reduce these serious traffic crashes (Al-Ghamdi, 2003, 1996; Alarifi et al., 2017; Alkahtani et al., 2018; Altwaijri et al., 2011; Hassan and Al-Faleh, 2013; Koushki and Al-Ghadeer, 1992).

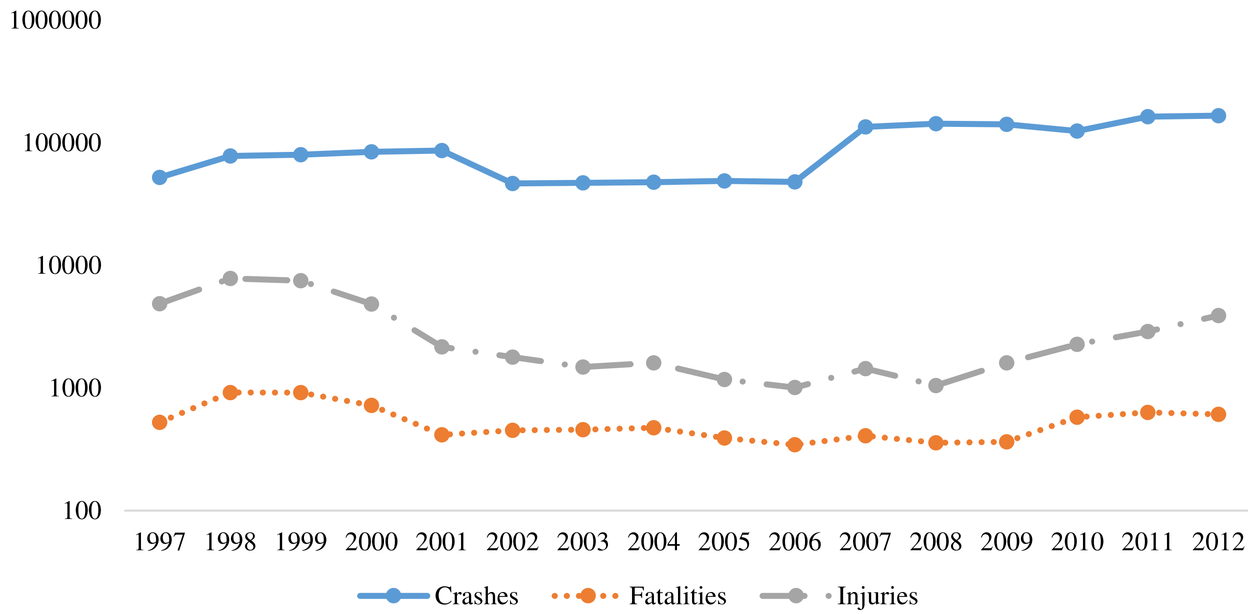


Figure 1-3: Total Road Traffic Crashes of Riyadh, KSA from 1997 to 2012 (Directorate General of Traffic, 2013)

1.2 Research Objectives

In order to achieve the objectives of this research, statistical models of crash frequency can be used to investigate the associations between traffic crash frequency and contributing factors, including traffic volume, land-use, socio-demographics and roadway characteristics. One of the motivation to use crash frequency models is to predict some dependent variables over geographical units (e.g., census tracts/blocks, zip/post codes, counties, traffic analysis zones (TAZs)). The macro-level crash prediction models provide transportation planners with a broad-spectrum perspective to consider safety in the long-range transportation planning process (Washington, 2006).

The purpose of this research is to investigate the contributing factors that cause traffic crash frequency by incorporating overdispersion and spatial effects in the City of Riyadh and to determine potential countermeasures in order to improve traffic safety. The following procedures could achieve the specific objective:

1- Examining the role of local variations of parameters in spatial data;
2- Safety investigation of pedestrian crashes at the macroscopic level; and
3- Predicting traffic crash counts by severity and type at the macro-level as well.

These research objectives are broken down into a number of research tasks:

The first objective could be accomplished by conducting a comparison between different approaches that incorporate spatial effects for crash counts modeling using the following tasks:

a) Exploring the effect of random parameter on severe crashes.
b) Exploring the effect of spatial random parameter on severe crashes.
c) Exploring the effect of geographically weighted Poisson regression on severe crashes.
d) Exploring new approaches that incorporate overdispersion in geographically weighted regression, i.e., geographically weighted negative binomial regression (i.e., GWNBR).
e) Comparing the SVC, GWNBR, and random parameter models.
f) Comparing the outcomes using different measures, e.g. DIC, MAD, and RMSE.

Although many previous studies have been done that analyzed pedestrian crashes in developing countries, no study has explored pedestrian crashes using data from Middle East countries, such as Saudi Arabia. Thus, the second objective is to identify the causes and characteristics of pedestrian crashes in Riyadh. Specifically, it will estimate the effects of traffic volume, roadway characteristics, socio-economics, and land-use as credible factors associated with the involvement of pedestrians in road traffic crashes at the zonal level. It would also be meaningful to compare the contributing factors to pedestrian crashes with those in other developing countries.

In addition, most of the previous studies that have examined traffic crashes at the macro-level are related to HICs, whereas there is a lack of similar studies among lower- and middle-income countries (LMICs) where most of road traffic deaths (90%) occur, including Middle Eastern countries. The lack of laws that are related to some key risk factors for road safety, including speed, seatbelts, etc., is separating the KSA from their high-income counterparts. Moreover, there is an absence of safe infrastructure for the more vulnerable road users, including pedestrians and cyclists, as the typical mode of transportation in the KSA is the private car.

Therefore, the third objective would be achieved by developing univariate and multivariate spatial models and comparing them with those of aspatial counterparts using the full Bayesian (FB) technique for analyzing crash counts by crash severity levels (i.e., fatal, injury, and property damage only (PDO)) and by crash type levels (i.e., pedestrian, bicycle, single- vehicle and multi-vehicle) at the zonal level. Specifically, investigating the effects of contributing factors on crash counts by severity and contributing factors to crashes by type. That could be accomplished by using extensive crash data, including road networks, traffic volume, socio-economics and demographics, and land-use data, which are commonly used for long-range transportation plans.

1.3 Organization of the Dissertation

The organization of this dissertation is as follows: First, Chapter 2, following this chapter, summarizes the literature review on previous macroscopic traffic safety researches. Chapter 3 presents a detailed description of the ways and efforts of the data collection and preparation used in this research. Chapter 4 explains the statistical methodologies that were followed to achieve the research objectives. Chapter 5 explores the spatial dependency (autocorrelation) and heterogeneity in spatial crash data at the macroscopic level using most broadly used approaches for modeling spatially correlated data, i.e., geographically weighted regression (GWR) and spatially varying coefficients (SVC). Chapter 6 examines the research analyses and results of macroscopic crash modeling for pedestrian crashes in Riyadh, Saudi Arabia. Chapter 7 develops a multivariate spatial model for crash counts by severity and type levels. Several implications for traffic safety policies in Riyadh are suggested based on the results in Chapter 8. Finally, Chapter 9 summarizes the overall dissertation.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

There have been enormous studies in the literature focused on spatial predictions of crash data from a Bayesian perspective using a wide array of statistical methodologies based on data from different spatial units at the macro-level. A summary of these approaches, the spatial units, and the crash levels that were used in previous studies are listed in Table 2-1.

2.2 Random Parameter model and Geographically Weighted Regression

Crash data is observed at a specific location and then aggregated to a zonal level for analysis and modeling. Two common approaches that are employed to model such data are geographically weighted regression with Poisson (GWPR) and Negative Binomial (GWNBR) and random parameter or spatially varying coefficients (SVC) models. The first one explains the correlations across observations at different geographic units, whereas the latter one is to account for spatial heterogeneity when modeling the association between crash counts and contributing factors at the zonal level.

A study by Hadayeghi et al. (2003) estimated total and severe crashes at the TAZ level in Toronto, Canada, using negative binomial regression at the macro-level. The authors revealed that the total crashes per zone increased as the vehicle-kilometers-traveled (VKT), the major roadway length, the total number of employment and the minor roadway length increased. However, higher posted speed limit and higher volume to capacity had a negative impact on total crashes. Moreover, the VKT, the major road length and the number of household members had a significant positive association with severe crashes in the morning peak hour. In contrast, the congestion and the posted speed limit had a significantly negative association with severe crashes during the same period. In addition, the authors developed Geographically Weighted Regression (GWR) models to inspect the spatial variation between zonal crashes and covariates. The results indicated inconsistent improvements. Li et al. (2013) used the GWPR to study the spatial heterogeneity in the relationship between zonal fatal crash counts and contributing factors at the zonal level in California. The results uncovered that the GWPR successfully captured the spatial variation and outperformed the traditional model, i.e., generalized linear model (GLM), in predicting fatal crashes. Recently, Xu and Huang (2015) conducted a study to investigate the spatial heterogeneity in regional safety modeling for total and severe crashes using two advanced models that are commonly employed to account for spatial autocorrelation, i.e., a semi-parametric geographically weighted Poisson regression and a random parameter negative binomial model. The results revealed that both models captured the spatial variation between the zonal crash frequency and covariates, yet sets of outcomes were different. The semi-parametric geographically weighted Poisson regression (S-GWPR) revealed a better goodness-of-fit, e.g., highest value of R-Squared, lowest mean absolute deviance (MAD) and Akaike information criterion (AIC). However, the S-GWPR may be affected by overdispersion in crash data. Therefore, it would be important in crash modeling to account for the overdispersion in GWR. In a recent study conducted by Gomes et al. (2017), the GWNBR was proposed for crash count modeling. The authors concluded the model is a promising tool in providing a better fit for the regional data because it accounts for overdispersion and reduces spatial dependence. However, one of the limitations of their study is the absence of the traffic exposure variable.

2.3 Contributing Factors to Pedestrian Crashes

Several prior studies have been devoted to identifying the contributing factors associated with pedestrian crashes in the context of developing countries. Some of these studies analyzed the factors that affect the frequency of pedestrian crashes (Rifaat et al., 2017; Tulu et al., 2015; Wang et al., 2016). Tulu et al. (2015) conducted a study to estimate pedestrian crash frequency for two-lane rural roads in Ethiopia using the random parameter negative binomial model. The authors revealed that younger pedestrians have a lower risk of crashes compared to older pedestrians. Moreover, it was found that crossing the road in pairs or in groups is safer than crossing alone. In a recent study carried out by Wang et al. (2016), the association between pedestrian crashes and several factors related to roadway characteristics, socio-economic features, and land-use data was examined in Shanghai, China. The authors found that the population had a positive impact on pedestrian crashes. In addition, length of major and minor arterials, higher roadway density, area of TAZ, and land-use intensity were statistically significant and positively associated with pedestrian crashes. However, average intersection spacing in kilometer and percentage of 3-way intersections were statistically significant and were revealed to be negatively associated with pedestrian crashes. Rifaat et al. (2017) explored the factors contributing to the number of vehicle-pedestrian crashes at intersections in Dhaka, Bangladesh. The authors stated that most of the contributing factors are in line with those found in prior studies carried out in the context of developed countries.

In addition, most of these studies analyzed pedestrian-involved crash severity (Amoh-Gyimah et al., 2017a; Mabunda et al., 2008; Sarkar et al., 2011; Zhang et al., 2014). Mabunda et al. (2008) investigated the association between temporal factors, e.g., age, gender, and day/time of death, and pedestrian fatalities in South Africa. The authors uncovered that during weekday afternoons and evenings, pedestrian-related crashes (child and adolescent) occurred more often. However, a study by Amoh-Gyimah et al. (2017), in Ghana, found that during weekends, pedestrian crashes are more likely to occur, in addition to at night-time on roads with no street lights, on untarred roadways, and in the middle of road intersections. Sarkar et al. (2011) employed logistic regression models to investigate significant factors affecting the probability of pedestrian fatalities along Bangladesh's roadways. The authors found that the involvement of elderly pedestrians (population aged 55 years old and more) and young pedestrians (population aged less than 15 years old) increased the likelihood of a fatality. Furthermore, pedestrians who crossed the road had a higher risk of fatality compared to those who walked along the road. Moreover, there were higher odds of pedestrian fatalities on national highways with no traffic control or stop signs compared to those with traffic lights or police. Zhang et al. (2014) examined significant factors associated with fault and severity in pedestrian-involved crashes. The authors found that crashes involving elderly pedestrians (i.e., 45 years or older), drunk driving, other illegal driving, and speeding are more likely to cause serious injuries or death.

In Saudi Arabia, their representation is rare in similar studies that address pedestrian safety in the region. Al-Ghamdi (2002) illustrated the factors affecting pedestrian-related crashes in the city of Riyadh using 638 pedestrian-related crashes over three years, from 1997 to 1999. The results showed that the pedestrian fatality rates were relatively high within the age group of 1–19 years old and 60–80 years old, which indicates that the youngest and oldest age groups are more likely to be involved in fatal pedestrian crashes as compared to other age groups. In addition, 77.1% of the pedestrians involved in these crashes were struck while crossing a road due to either no crosswalk existing or crosswalk was not utilized. Additionally, Al-Shammari et al. (2009) analyzed 460 pedestrian crashes out of 1500 crashes over a 3-year period. They found that two-thirds of drivers and close to half of the abled individuals below the age of 30. In addition, between 4:00 pm and midnight was the time span in which two-thirds of pedestrian crashes happened. Regarding the crash conditions, they found that pedestrians being struck while crossing the road were about two-thirds.

These studies, however, lack a thorough investigation of the effects of contributing factors on pedestrian crashes at the macro-level. Investigating the impact of traffic, socio-economic factors, and land-use on pedestrian safety could fill the gap of studying the safety of pedestrians in this region.

There have been several studies that investigate the contributing factors for pedestrian crashes, which are mostly Western-related studies. They found several socioeconomic factors for pedestrian crashes at the macro-level. LaScala et al. (2000) pointed out that traffic flow, population density, and some socio-economic factors such as age composition of the local population, unemployment, gender, and level of education have an association with pedestrian injury rates. Ng et al. (2002) examined the risk of traffic crashes across 274 traffic analysis zones (TAZs) in Hong Kong. In terms of pedestrian-related crashes, the authors found that the number of cinema seats, commercial areas, flatted factory areas, market stalls, and mass transit railway (MTR) catchment areas had a positive impact on pedestrian crashes, whereas the greenbelt areas, specialized factory areas, and territory school places (primary, secondary, and tertiary) had a negative impact on pedestrian crashes. Noland and Quddus (2004) revealed that severe pedestrian crashes increase when total population, a lower percentage of the population aged between 45 and 64, and a higher percentage of the population aged between 64 or over increased. Loukaitou-Sideris et al. (2007) showed that pedestrian crashes are more likely to take place in zones with high population and employment density and high traffic volumes. Regarding land-use type, a large concentration of commercial/retail and multifamily residential land uses contributes to pedestrian collisions. Further, a higher concentration of the Latino population was associated with more pedestrian crashes than other ethnicities.

Wier et al. (2009) stated that there is a positive association between pedestrian crashes and traffic volume, arterial roads without transit, the proportion of commercial and residential land uses, employee and resident populations, and the proportion of people living below the poverty level. However, the land areas and the proportion of the population aged 65 years old or over had negative effects. Siddiqui et al. (2012) found out that there was a positive association between the number of pedestrian crashes and roadway length with a 35 mph posted speed limit, number of intersections, total number of dwelling units, population density, percentage of households with 0 or 1 vehicle, and number of employment. However, association with family income was negative. Lee et al. (2015b) found that the vehicle-miles-traveled, population density, proportion of African Americans and Hispanics, proportion of families without vehicles, rooms of hotel/motel, time-share employment, and the number of intersections and traffic signals were found to significantly and positively affect pedestrian crashes. In contrast, the proportion of roadways with speed limits more than or equal to 55 mph had a negative effect on pedestrian crashes.

2.4 Multivariate Crash Frequency Prediction Models

There is a substantial body of previous studies devoted to exploring the correlations among crash counts by severity (number of crashes resulting in fatalities, injuries, and PDO) simultaneously. For example, Ladron de Guevara et al. (2004) analyzed crash frequency simultaneously by severity level (i.e., no injury, injury, and fatal) using traffic analysis zone (TAZ) crash data. Recently, Aguero-Valverde (2013) used the multivariate conditional autoregressive (CAR) model by applying a full Bayesian hierarchical approach to simultaneously estimate crash counts by crash severity (i.e., fatal, injury, and PDO) and compare it with its univariate counterpart at the macro-level. Five years of 81 cantons crash data in Costa Rica was utilized. The author revealed that the multivariate spatial model performed better than the univariate one. Yasmin and Eluru (2017) proposed a joint negative binomial-ordered logit fractional split model framework to jointly analyze motor vehicle crash severities (no injury, minor injury, incapacitating injury and fatal injury) at the statewide traffic analysis zone (STAZ) level in Florida. The authors confirmed the superiority of the multivariate model in terms of data fit compared to the univariate model.

In addition, there are many studies which have also examined correlations among crash counts by type or mode (Narayanamoorthy et al., 2013; Nashad et al., 2016; Osama and Sayed, 2017; Song et al., 2006; Wang and Kockelman, 2013). Song et al. (2006) developed Bayesian multivariate conditional autoregressive (MCAR) models that can account for the spatial effect using 254 counties in Texas to simultaneously analyze crash frequency by roadway locations (i.e., intersection, intersection-related, driveway-access, and non-intersection). The authors revealed that the model with multivariate CAR and the univariate model with correlated CAR performed much better than the model with univariate CAR. Narayanamoorthy et al. (2013) and Wang and Kockelman (2013) employed a bivariate spatial modeling approach for counts of non-motorized traffic crashes (i.e., pedestrian and bicycle) by injury severity crash analysis at the census tracts (CTs) level. Nashad et al. (2016) developed a bivariate model by adopting a copula-based negative binomial model for pedestrian and bicycle crash frequency considering crashes at the STAZ level in Florida. The authors illustrated the importance of incorporating dependence between pedestrian and bicycle crashes in the macro-level analysis. Osama and Sayed (2017) developed a bivariate CAR to analyze counts of crash types of pedestrians and bicyclists in each of 134 TAZs of the city of Vancouver between 2009 and 2013. The authors concluded that the crash modes and spatial correlations are found to significantly affect the models' performance in the macro level safety analysis.

Also, several previous studies have shown the presence of correlations between non-motorized crashes (i.e., pedestrian and bicycle) and counts of motorized-related traffic crashes. Recently, Lee et al. (2015) (macro-scale) and Huang et al. (2017) (micro-scale) developed a multivariate spatial model to simultaneously analyze the frequency of motor vehicle, bicycle, and pedestrian crashes. These two studies concluded that the multivariate spatial model significantly outperformed the corresponding univariate spatial model, and that the spatial error component played an important role in significantly improving the model performance. In addition, the highly correlated heterogeneous residuals in modeling crash risk among these three transportation modes were confirmed. However, the results were different regarding spatial correlation. The correlations for spatial residuals between different crash modes at neighboring locations were not statistically significant in the micro-level safety analysis.

Regarding motorized-related crashes, Geedipally and Lord (2010) asserted that motorized crashes (i.e., single- and multi-vehicle) should be analyzed separately, and a joint NB model should be utilized. Ma et al. (2016) developed a random effect bivariate Poisson-lognormal model to explore the effect of geometric features, weather, and traffic conditions on crash frequency at the micro-level. The authors confirmed the existence of correlations between single- and multi-vehicle crashes. Therefore, it would be meaningful to explore the correlation between counts of non-motorized (i.e., pedestrian and bicycle) and motorized (i.e., single- and multi-vehicle) crashes and to examine the contributing factors at the macro-level.

Table 2-1: Summary of Previous Studies Related to Macroscopic Crash Analysis

Study	Methodology	Spatial Unit	Crash level	Study	Methodology	Spatial Unit	Crash level
(Aguero-Valverde, 2013)	Multivariate and Univariate, FB	Canton	Fatal, injury and PDO	(Cai et al., 2017a)	Bayesian joint modeling	TAD	Total crash, Non-motorist crash, Proportion of non-motorist crashes
(Cai et al., 2018b)	Grouped random parameters multivariate spatial	TAD	Total crashes	(Ladron de Guevara et al., 2004)	Simultaneous NB	TAZ	Fatal crashes, Injury crashes, PDO crashes
(Lee and Abdel-Aty, 2017)	Multivariate Bayesian Poisson lognormal CAR	ZIP	Total Pedestrian crashes, Total bicycle crashes	(Lee et al., 2015b)	Multivariate Poisson lognormal conditional autoregressive model	TAZ	Total crashes
(Narayanamoorthy et al., 2013)	Spatial multivariate, generalized ordered-response (GOR) framework	CT	Pedestrian and bicycle Possible injury, Non-incapacitating injury, Incapacitating injury, Fatal injury crashes	(Nashad et al., 2016)	Copula based bivariate NB	STAZ	Total Pedestrian crashes, Total bicycle crashes

Study	Methodology	Spatial Unit	Crash level	Study	Methodology	Spatial Unit	Crash level
(Osama and Sayed, 2017)	Multivariate conditional autoregressive (CAR)	TAZ	Total Pedestrian crashes, Total bicycle crashes	(Song et al., 2006)	Multivariate conditional autoregressive (CAR)	County	Severe crashes
(Wang and Kockelman, 2013)	Multivariate Poisson-lognormal conditional-autoregressive, GMCAR	TAZ	Severe crash counts, non-Severe crash counts	(Yasmin and Eluru, 2018)	Joint NB-Ordered Logit Fractional Split	STAZ	Total crashes, no injury crashes, minor injury crashes, incapacitating injury crashes, fatal injury crashes
(Abdel-Aty et al., 2011)	NB	TAZ	Total crashes, Severe crashes, Peak hour crashes, Pedestrian and Bicycle crashes	(Abdel-Aty et al., 2013)	Poisson-lognormal	BG, CT, TAZ	Total crashes, Severe crashes, Pedestrian crashes
(Aguero-Valverde and Jovanis, 2006)	Full Bayes hierarchical model	County	Fatal crashes, Injury crashes	(Amoros et al., 2003)	NB	County	Total crashes, Fatal crashes
(Cai et al., 2016)	NB, zero-inflated NB and hurdle NB	TAZ	Total pedestrian crashes, Total bicycle crashes	(Cai et al., 2017b)	aspatial Poisson-lognormal model (PLN), Poisson-lognormal Conditional	TAZs, TADs	Total crashes severe crashes, total bicycle and pedestrian crashes

Study	Methodology	Spatial Unit	Crash level	Study	Methodology	Spatial Unit	Crash level
					Autoregressive mode		
(Cai et al., 2018a)	Bayesian Poisson-lognormal	TAD	Total crashes	(Cho et al., 2009)	Path analysis	Community analysis zones	Total crashes
(Cottrill and Thakuriah, 2010)	Poisson Regression with heterogeneity	CT	Total crashes	(Guo et al., 2017)	Poisson-lognormal, spatial Poisson-lognormal,	TAZ	Total, severe and slight
(Hadayeghi et al., 2007)	NB	TAZ	Total crashes, Severe crashes	(Hanna et al., 2012)	Conditional logistic regression	County	Fatal crash
(Huang et al., 2010)	Bayesian spatial model	County	Total crashes, Severe crashes	(Huang et al., 2016)	Bayesian spatial model with CAR prior, Bayesian spatial joint model	TAZ	Total crashes
(Karim et al., 2013)	NB, Spatial Poisson-Gamma, Full Bayes Estimation Technique	TAZ	Total number of crashes, Number of injury crashes, Property damage only crashes PDO	(Karlaftis and Tarko, 1998)	Cluster analysis, NB	County	Total crashes, Urban crashes, Rural crashes
(Kim et al., 2006)	NB	Grid structure	Total crashes, Bicycle crashes, Pedestrian crashes	(LaScala et al., 2000)	Spatial autocorrelation corrected regression	CT	Injury crashes

Study	Methodology	Spatial Unit	Crash level	Study	Methodology	Spatial Unit	Crash level
(Lee et al., 2014b)	Bayesian Poisson Lognormal	TAZ, TSAZ	Total crashes, Severe crashes	(Lee et al., 2014a)	Bayesian Poisson-lognormal	ZIP	Total crashes
(Lee et al., 2015a)	Bayesian Poisson lognormal simultaneous equations spatial error model	ZIP	Total Pedestrian crashes	(Lee et al., 2017)	Mixed-effects NB	BG, TAZ, CT, ZCTA, TAD, CCD, and County	Total and severe crashes, total pedestrian crashes and total bicycle crashes
(Loukaitou-Sideris et al., 2007)	OLS	CT	Total pedestrian crashes	(Lovegrove and Sayed, 2006)	GLM assuming NB	Neighborhoods	Total/severe
(MacNab, 2004)	Bayesian spatial and ecological regression model	Local health areas	Injury crashes	(Moeinaddini et al., 2014)	NB	City	Fatalities per million inhabitants
(Naderan and Shahi, 2010)	NB	TAZ	Total crashes, PDO crashes, Injury crashes, Fatal crashes	(Ng et al., 2002)	Cluster analysis, NB	TAZ	Total crashes, Fatal crashes, Pedestrian crashes
(Noland, 2003)	NB	State	Fatal injury crashes	(Noland, 2003)	RE-NB	State	Fatal crashes, Injury crashes
(Noland and Oh, 2004)	NB	County	Total and fatal	(R. Noland and Quddus, 2004)	NB	Standard statistical regions	Fatal/Serious injury crashes, Slight injury crashes
(R. B. Noland and Quddus, 2004)	NB	County	Total crashes, Fatal crashes	(R. B. Noland and Quddus, 2004)	NB	Census wards	Fatal crashes, Serious injury crashes, Slight injury crashes

Study	Methodology	Spatial Unit	Crash level	Study	Methodology	Spatial Unit	Crash level
(Noland et al., 2013)	NB	BG	Total severe crash and total severe pedestrian crashes	(Osama and Sayed, 2016)	GLM and FB	TAZ	Total Cyclist-motorist
(Pulugurtha et al., 2013)	NB with log-link, Wald Chi-square,	TAZ	Total number of crashes, Number of injury crashes, Property damage only crashes PDO	(Quddus, 2008)	NB, Spatial autoregressive model, Bayesian hierarchical model	Census ward	Fatal crashes, Serious injury crashes, Slight Injury crashes
(Siddiqui and Abdel-Aty, 2012)	Bayesian Poisson-lognormal	TAZ	Total Pedestrian crashes	(Siddiqui and Abdel-Aty, 2016)	NB, Bayesian Poisson-lognormal	TAZ	Total and severe crashes
(Siddiqui et al., 2012)	NB, Bayesian log-normal model	TAZ	Total pedestrian crashes, Total bicycle crashes	(Siddiqui et al., 2014)	Bayesian Poisson-lognormal	TAZ	Total Pedestrian crashes
(Stamatiadis and Puccini, 2000)	Quasi induced exposure method	State	Fatal crashes	(Ukkusuri et al., 2012)	NB, NB with heterogeneity in dispersion parameter, Zero-inflated NB	CT	Total crashes
(Wang et al., 2009)	NB	Census ward	Fatalities Serious injuries Slight injuries	(Wang et al., 2016)	Bayesian CAR	TAZ	Total pedestrian crashes
(Wei and Lovegrove, 2013)	NB	TAZ	Total crashes	(Wier et al., 2009)	Ordinary least square regression	CT	Injury crashes

Study	Methodology	Spatial Unit	Crash level	Study	Methodology	Spatial Unit	Crash level
(Xu et al., 2014)	regionalization with dynamic constrained agglomerative clustering and partitioning, Bayesian Poisson lognormal model, Bayesian spatial	TAZ	Total crashes, severe crashes	(Yasmin and Eluru, 2016)	Latent segmentation	TAZ	Total pedestrian crashes
(Amoh-Gyimah et al., 2017b)	GLMs, RP-NB, S-GWPR	Statistical area, Thiessen polygon-based TAZ, state electoral divisions, postal areas, grid	Total crash Serious crash Minor crash	(Gomes et al., 2017)	GWNBR	TAZ	Severe crashes
(Hadayeghi et al., 2003)	Negative binomial Geographically weighted regression	TAZ	Total crashes, Severe crashes	(Hadayeghi et al., 2010a)	GWPR, Full Bayesian Semiparametric Additive	TAZ	Total crashes, Severe crashes
(Hadayeghi et al., 2010b)	GWPR	TAZ	Total crashes, Severe crashes	(Li et al., 2013)	GWPR	County	Fatal crashes
(Rhee et al., 2016)	GWR, OLS, spatial error model, spatial lag model	TAZ	Total crash, fatal and severe injury and minor injury	(Xu and Huang, 2015)	GWPR, RP-NB	TAZ	Total and severe crashes

Study	Methodology	Spatial Unit	Crash level	Study	Methodology	Spatial Unit	Crash level
(Zhang et al., 2015)	GWPR	CT	Total crashes involving pedestrians and bicyclists	(Amoh-Gyimah et al., 2016)	NB, RP-NB, and Poisson-Gamma-CAR.	Statistical Area census classification	Total crash Serious crash Minor crash
(Coruh et al., 2015)	RP-NB	City	Total crash	(Ukkusuri et al., 2011)	RP-NB	CT	Total crashes
(Xu et al., 2017)	Bayesian spatially varying coefficients	TAZ	Severe crashes	(Tasic et al., 2016)	Generalized Additive Models, Bayesian Hierarchical Models	CT	Total crashes, total pedestrian crashes, total bicycle crashes
(Dissanayake et al., 2009)	Generalized Linear Model, Generalized Poisson model, NB GLMs	Census ward	Total, severe, school time slight, school time severe, non-school time slight, non-school time severe	(Dong et al., 2016)	Poisson Lognormal model, the Bayesian spatial and temporal model, the Bayesian spatio-temporal interaction model	TAZ	Total crash
(Hadayeghi et al., 2006)	NB	TAZ	Total and severe crashes	(Truong et al., 2016)	RE-NB, RP-NB, ST-CAR	Provinces	Fatal
(Levine et al., 1995)	Spatial lag	CB	Total crashes	(Wang et al., 2018)	Poisson lognormal CAR	TAZ	Total crash

2.5 Literature Review Summary "Current Issues"

Considering previous studies, the research gaps can be summarized as follows:

1- A lack of studies that used complete and accurate traffic crash data to represent the situation in LMICs, including Middle Eastern countries;

2- Few studies addressed the contributing factors that affect frequency of traffic crashes in these countries, especially the most vulnerable road users, i.e., pedestrians;

3- A need to conduct a comparison to investigate the effect of spatial dependency and heterogeneity in crash count modeling incorporating overdispersion at the zonal level using advanced spatial models; and

4- A lack of studies that addressed the contributing factors that affect frequency of traffic crashes in Riyadh at the macro-level.

5- Therefore, Riyadh, Saudi Arabia, which is one of the most densely populated and heavily motorized cities in the Middle East, is used as an example to reflect the representation of these countries' characteristics, where Saudi Arabia has a unique situation in that it is considered a high-income country, and yet it has the highest rate of traffic fatalities compared to its high-income counterparts.

CHAPTER 3: DATA DESCRIPTION

3.1 Introduction

In order to achieve useful results for this study, substantial data was obtained. Several meetings and communications were done with different authorities and departments in Riyadh to collect relevant data. At the beginning, it was hard to figure out which department had the related data and even harder to obtain it due to the requirements, many of which take a long time to process and respond to.

There are two zoning systems that are used by the Higher Commission for the Development of Riyadh (HCDR) and The General Department of Studies and Designs in the Riyadh municipality: TAZ and HAY. However, the most complete data is based on the HAY-level from the HCDR, which was used in this study. The HCDR is using a HAY zone system for transportation planning and modeling. HAY is the term traditionally referring to a neighborhood, and it is relatively comparable to the traffic analysis districts (TADs), (a group of TAZs), in the USA. The city of Riyadh is composed of 208 HAYs, as can be seen in (Figure 3-1).

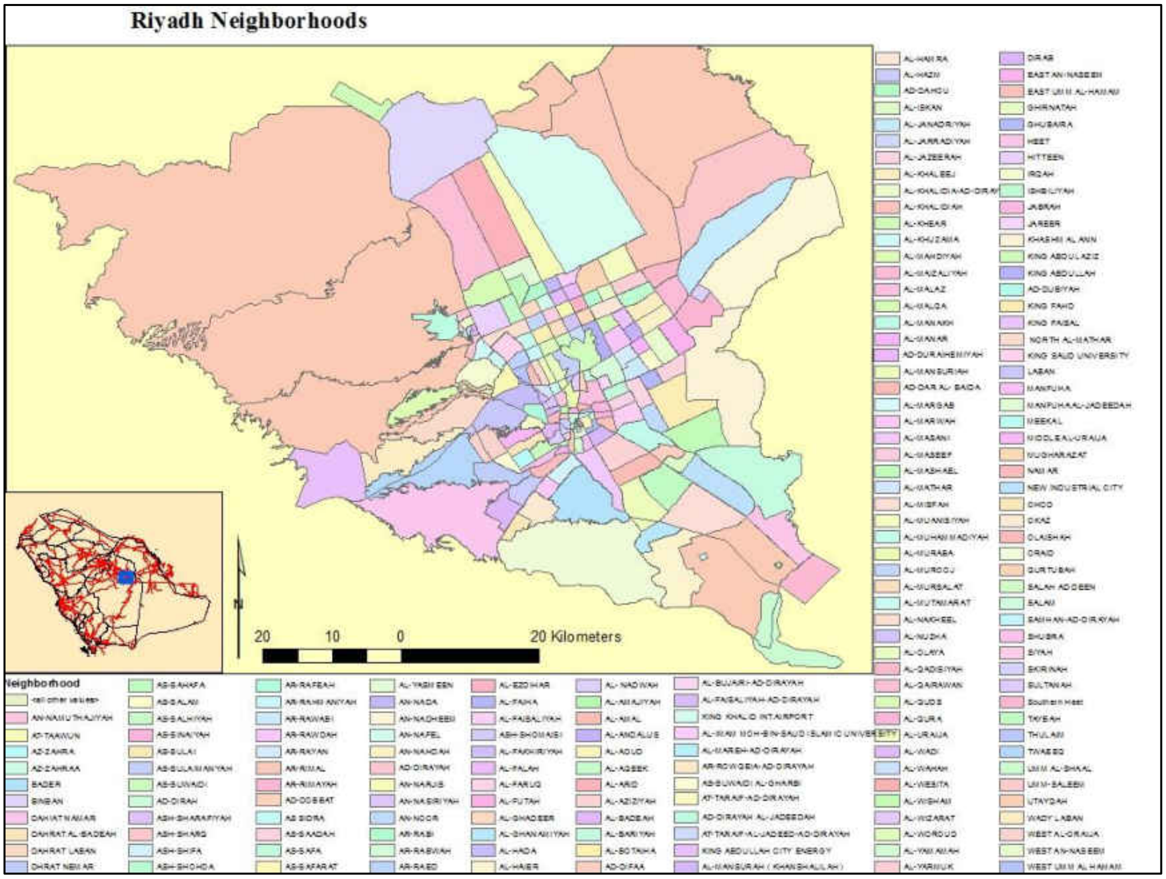


Figure 3-1: HAYs of Riyadh City

These 208 HAYs are considered to be the development protection boundaries that contain all types of land-use as shown in Figure 3-2 (dotted lines). However, around 80% of the area in these 208 HAYs is either uninhabited or unused, including valleys and empty or undeveloped areas. Figure 3-2 and Figure 3-3 display the spatial distribution of total crashes and total pedestrian crashes per HAY of the study areas (179 HAYs) utilized in this study (discussed later in following sections).

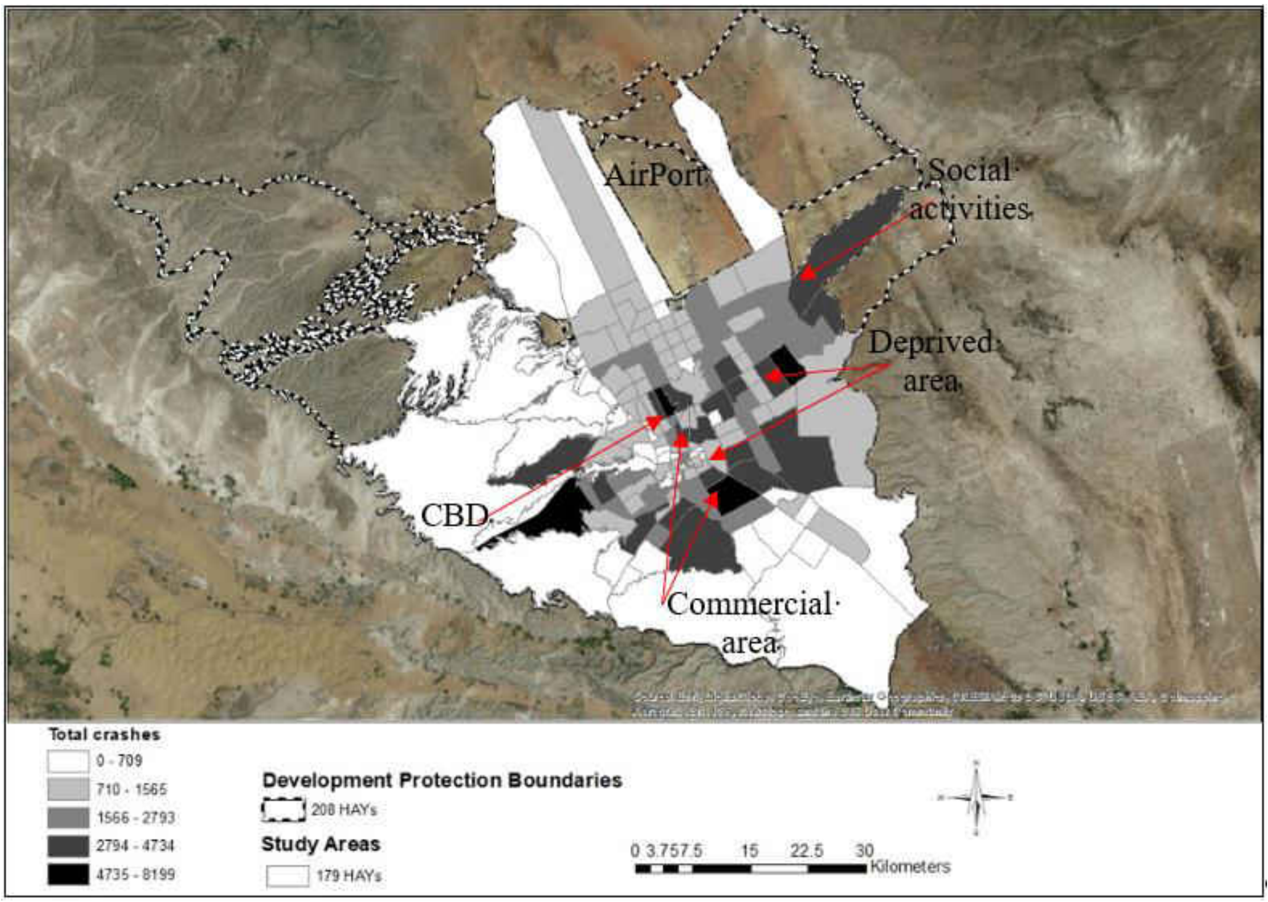


Figure 3-2: Map of development protection boundaries, study areas, and spatial distribution of total crashes per HAY

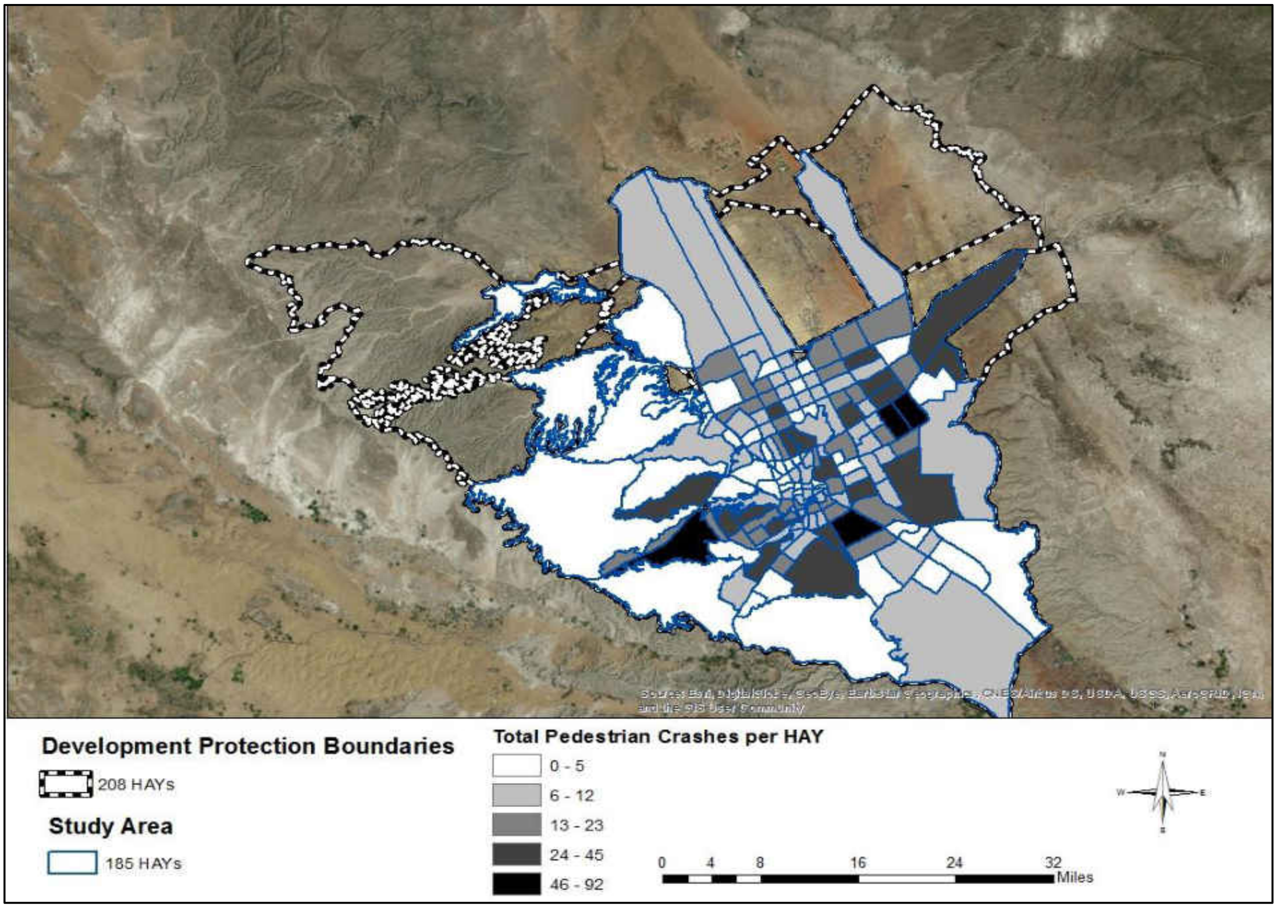


Figure 3-3: Map of Pedestrian Crash Spatial Distributions per HAY

3.1.1 Crash Data

The main source of road crash data is the traffic police. The General Directorate of Traffic publishes the road crash statistics each year, which contains all reported crashes with general information on the crash, vehicle(s) and individual(s) involved. The available reports cover the period from 1997 to 2012. Since 2004, the Higher Commission for the Development of Riyadh "HCDR" has collected road crash details from the Riyadh Traffic Department using a new system to store this data for analysis and study. Table 9-1 shows the traffic crash report template that the traffic department is using to report the details of the crash. Thus, the details of crash data were essentially received from the HCDR in Excel file format that has three worksheets containing details of the crash, vehicle(s) involved and the person(s) that were involved in the crash for the period from 2004 to 2016. This crash data is provided for all types of severity, including PDO crashes. Moreover, part of the property damage only crashes (PDO) are collected by a private company, namely, NAJM. Two conditions are required in order for the NAJM company to handle and report the PDO crash: 1) at least one of the vehicles or individuals that were involved in the crash has insurance and 2) there is no injury caused by the crash. The NAJM company was established in 2007 and covers 27 cities in the KSA. That is, enough to have the most recent and complete crash data.

The crash dataset contains the details of the crash date and time, crash type, crash location (i.e., latitudes, longitudes), street, HAY, zone, district, crash point, direction, type of private and public damage, lighting, land surface and weather conditions, crash severity, crash reasons, crash type, e.g., head-on and angle, number of major and minor injuries, and number of fatalities. In order to investigate crash data, the contributing factors for the crashes as marked by the police are provided relevant to the driver, passenger, pedestrian, road and vehicle. For example, Figure 3-4 and Figure 3-5 show the contributing factors for pedestrian crashes relevant to driver and pedestrian respectively for this research. Figure 3-4 shows the contributing driver factors for pedestrian crashes. The most relevant factor is distraction at 42.93%, followed by sudden deviation at 32.15% and speeding at 17.62%. In addition, Figure 3-5 illustrates the contributing pedestrian factors for pedestrian crashes. The most prominent factor is the pedestrian crossing the road at non-allowed places at 87.47%.

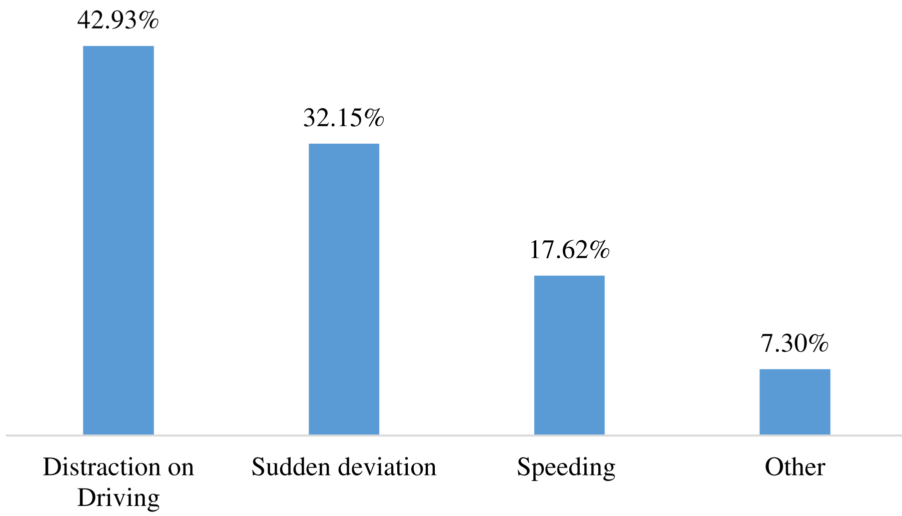


Figure 3-4: Pedestrian Crash Causes Relevant to Driver

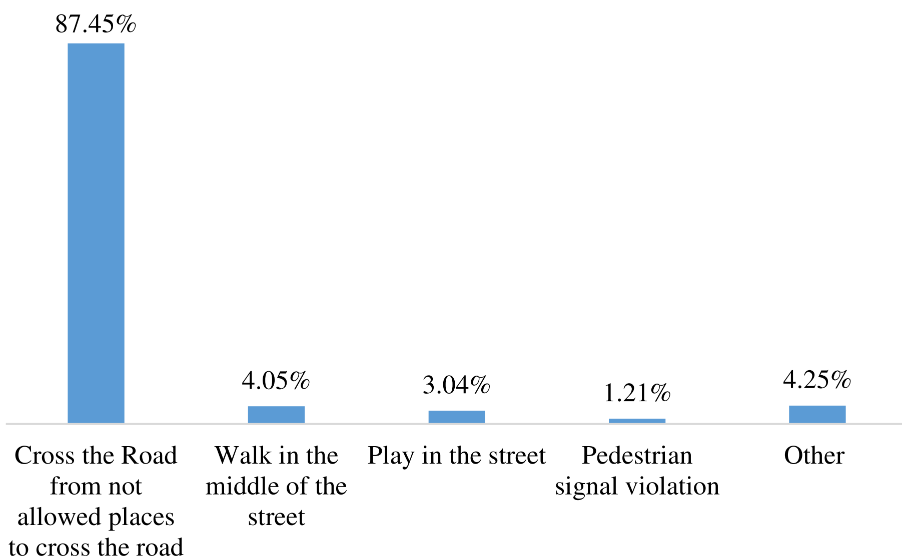


Figure 3-5: Pedestrian Crash Causes Relevant to Pedestrian

The second file contains the vehicle(s) dataset and contains details of the percentage of blame, vehicle make, model and color, driver direction, vehicle registration type, registration country, vehicle status, and point of collision.

The third file contains the person(s) dataset and contains details of the age, birth year, person type, e.g., driver, pedestrian etc., health status, gender, license type, and nationality. In regard to this research, Figure 3-6 and Figure 3-7 illustrate the distribution of age-cohort fatality rates per hundred thousand people related to pedestrian and driver respectively in 2015. It can be seen that the people aged 60 years old and over had the highest pedestrian fatality rate. Moreover, the young population (aged 14-25 years old) was the highest group involved in pedestrian crashes as a driver.

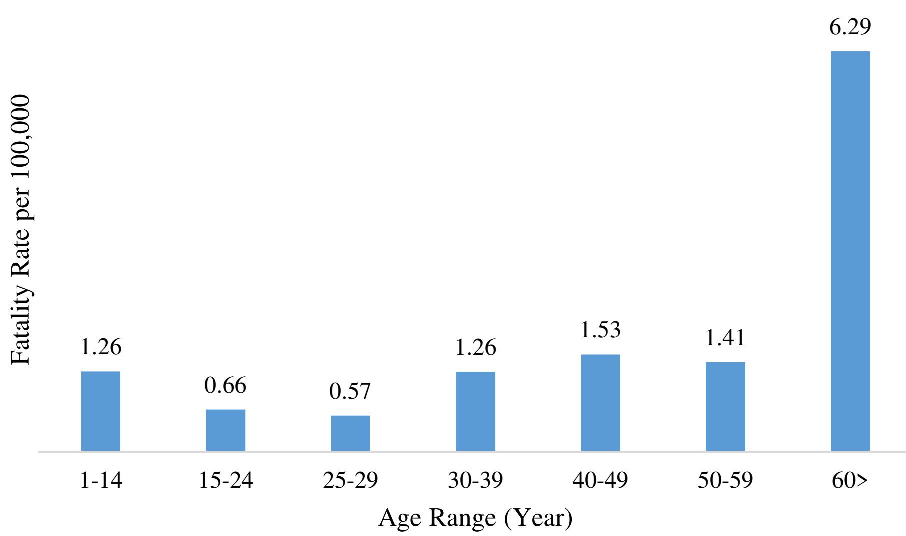


Figure 3-6: Distribution of Fatality Rates by Age Cohort (Death per 10^5 Population) in 2015, Riyadh Related to Pedestrian

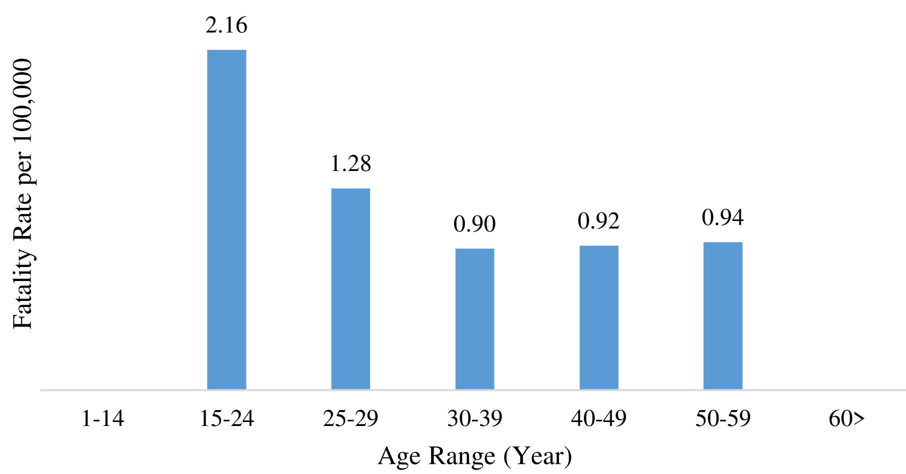


Figure 3-7: Distribution of Fatality Rates by Age Cohort (Death per 10^5 Population) in 2015, Riyadh Related to Vehicle Driver

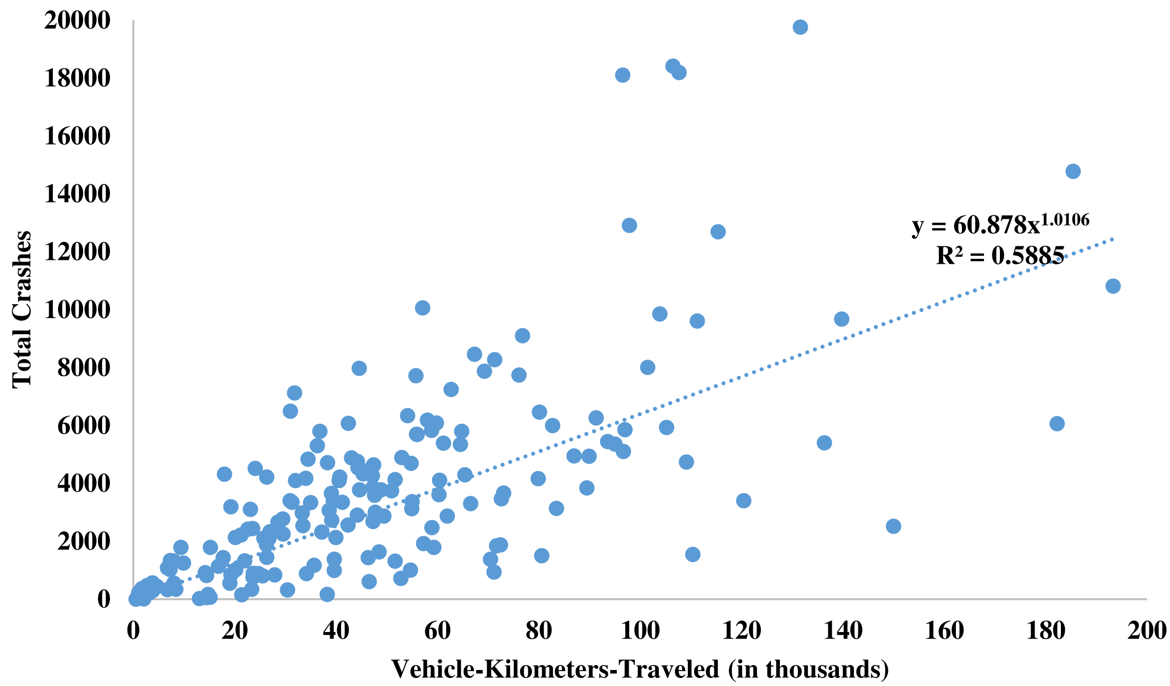
There are two key identifications to link these three datasets; first is to link the crash record to each vehicle(s)- and person(s)-related record, and the second is to link each involved vehicle(s) to each involved person(s). These two key IDs were used to identify single- and multi-vehicle crashes and pedestrian and bicycle crashes as well. There were 1,487,140 crash reports in the period from 2004-2016. In addition, there were 104 different nationalities reported in crash reports. Pedestrian crashes were extracted using the two key IDs where the person type is pedestrian. There were 10,075 pedestrian crashes. Then, using ArcGIS to locate these records, only 4,655 pedestrian crashes were located in the study area and considered in this research. However, some of the records were not geocoded in the data, especially older ones. In addition, using several years of data could be misguided because of the changes in the road network (e.g., adding a new road), rapid change in land-use and population growth, or some other changes which have developed during the study period that may skew the analysis (Herbel et al., 2010). Therefore, 2,131 pedestrian-vehicle crashes during the period from 2013 to 2015 were used in this study for investigation of pedestrian crashes. In addition, a total of 253,217 crashes within the period from 2012 to 2015 were used in this study for predicting crash counts by severity and type.

3.1.2 Road Network and Traffic Volume Data

There were several maps of the road network. The General Department of Studies and Designs in the Riyadh municipality has the road network data for 2012, which contains the traffic volume data, speed limits of roads, road functions and sub-functions, number of lanes and other information. This data was obtained in the beginning of this research for the initial analysis. The issue with this data was that the data did not align with the ArcGIS Basemap for which the projection needs to be corrected (recently, they have been working to update the map). Therefore, with the help of ESRI, the network was aligned to the Basemap. After several trials, the following steps were able to fix the projection issue:

- Transforming the geographic coordination system from "Ain_el_Abd_1970" to the "GCS_WGS_1984".
- Using the project tool (Data Management) in ArcMap to project the data from "GCS_WGS_1984" (Geographic Coordinate System) to "WGS_1984_Web_Mercator_Auxiliary_Sphere" (Projected Coordinate system).
- Editing in ArcMap and zooming to a major street intersection for the reference.
- Dragging and moving the features to match the street intersections shown in ESRI Basemaps.

However, during the research, it was found that the HCDR has an accurate map that covers most of the road network. Therefore, the road network data was also obtained from HCDR in a shapefile format, which includes road classification (limited-access highway, arterial, and collector), posted speed, number of lanes, and traffic volume. In addition, the number of intersections and number of traffic signals were extracted from an additional layer using ESRI ArcMap 10.3 to be used in this study. One of the significant factors in safety-related studies is the use of traffic exposure. Thus, traffic volume data was provided by the HCDR as average weekly daily traffic, which is calibrated and modeled based on traffic counts, demographic survey, and land-use data. Based on this, the vehicle-kilometers-traveled was calculated as the length of the road multiplied by the traffic volume on each road, and then aggregated to HAY-based data also using ESRI ArcMap 10.3 software. Figure 3-8 and Figure 3-9 illustrate the positive association between the vehicle exposure (vehicle-kilometers-traveled (VKT)) and population and total crashes and pedestrian crashes per HAY.



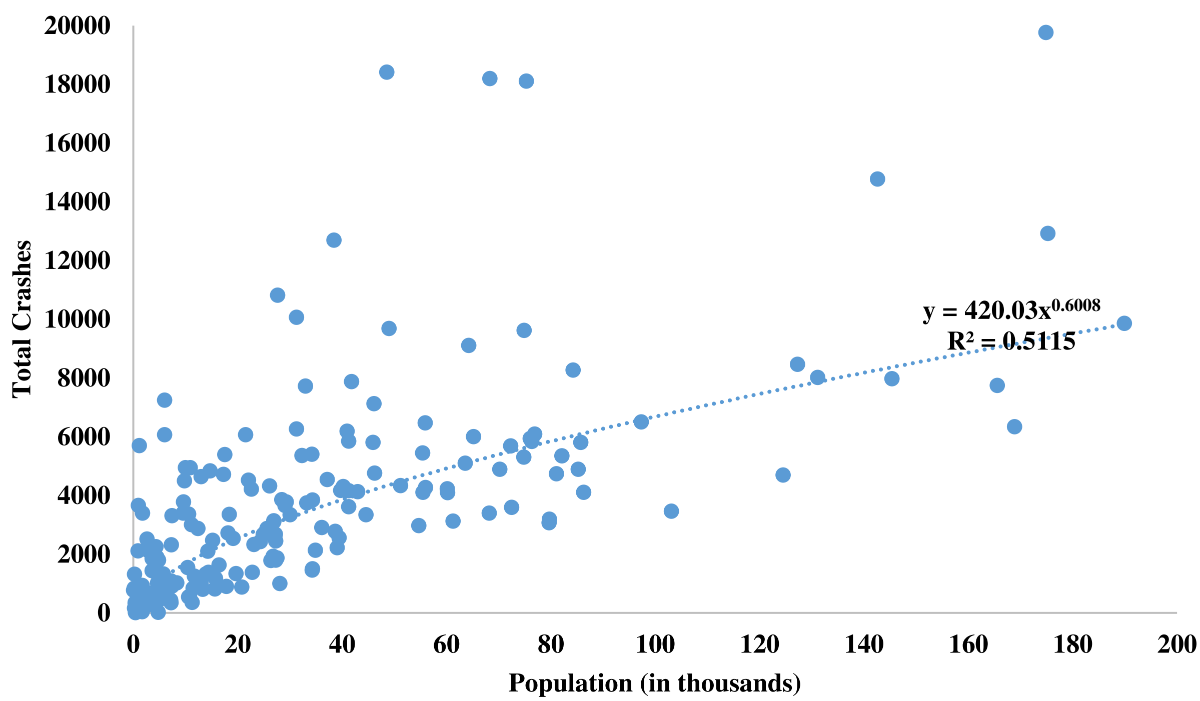


Figure 3-8: Scatter plot of the VKT and population vs. total crashes

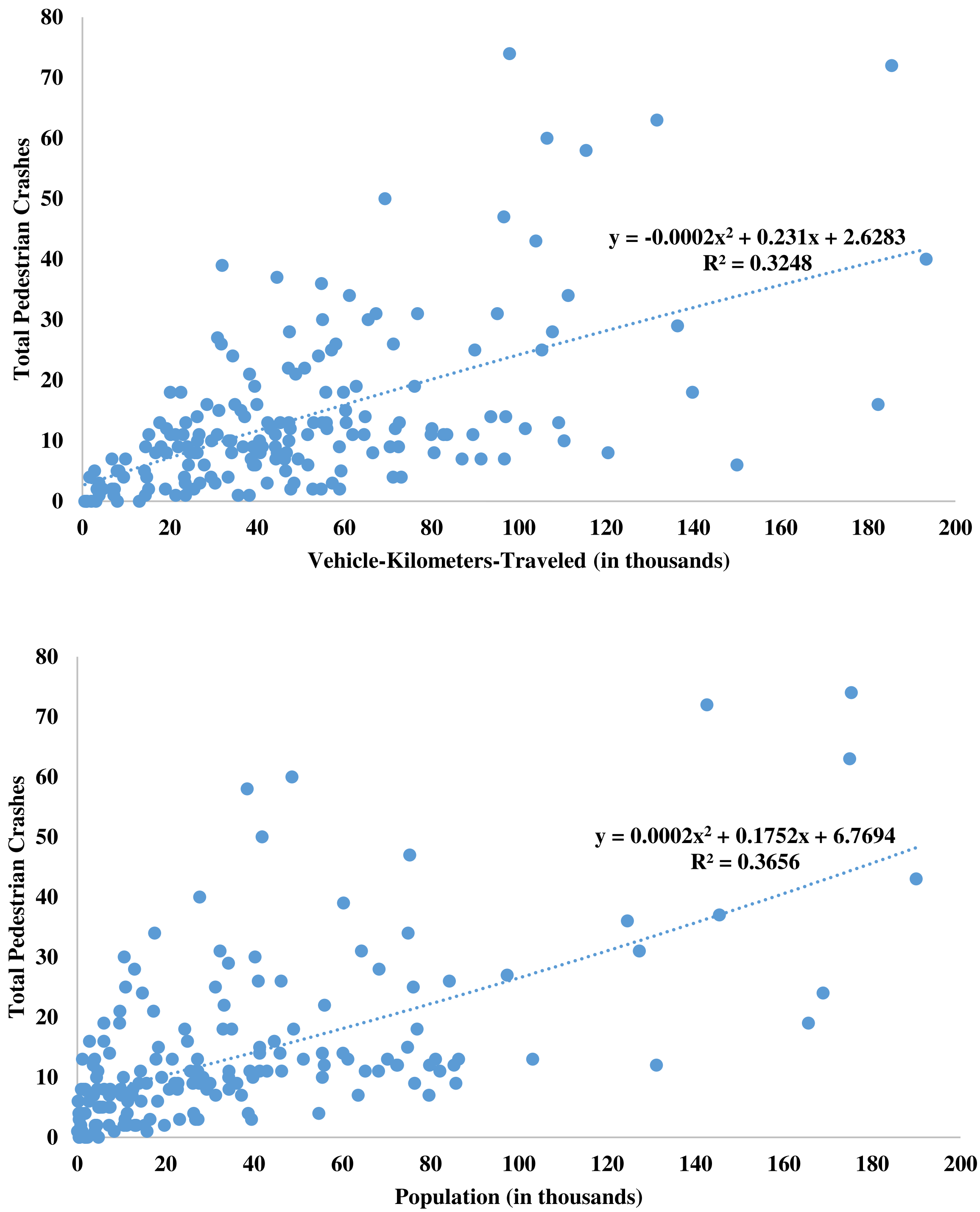


Figure 3-9: Scatter plot of the VKT and population vs. total pedestrian crashes

3.1.3 Socio-Economic and Demographic Data

The socio-economic and demographic data was collected from the field household survey conducted by HCDR in 2012. It includes 16 age groups of population, gender, Saudi and non-Saudi populations, home and vehicle ownership, employment status, education level, and marital status. The socio-economic survey was based on 185 HAYs that are fully developed and inhabited. Therefore, these 185 HAYs were considered in this research.

3.1.4 Land-Use Data

The land-use data was provided in a shapefile format. It was categorized mainly by 17 different types: Residential, Industrial, Warehouse, Transportation Services, Communication and Public Facilities, Commercial, Professional and Business Services, Governmental, Cemeteries, Health, Educational, Mosques, Cultural, Recreational, Agriculture, Empty (Unused Area), and Unknown. For example, the governmental area includes police, traffic, civil defense and ministry buildings and any government authority buildings. In addition, the transport services area contains the railway network and building services, passenger and cargo utilities, transport vehicle parking and airports. Recreational Areas include the areas of sports facilities and stadiums, sports arenas, swimming pools, amusement parks and exhibition halls, and all the campgrounds and parks and children's playgrounds. It also includes the public garden recreational parks. Lastly, agricultural areas include agriculture lands and farms and their related activities, mining extraction activities, crude oil and natural gas mineral areas and any other resources. As shown in Figure 3-10, the largest area is the residential area with 31.4%, followed by the agricultural and governmental areas with 18.32% and 10.5% respectively.

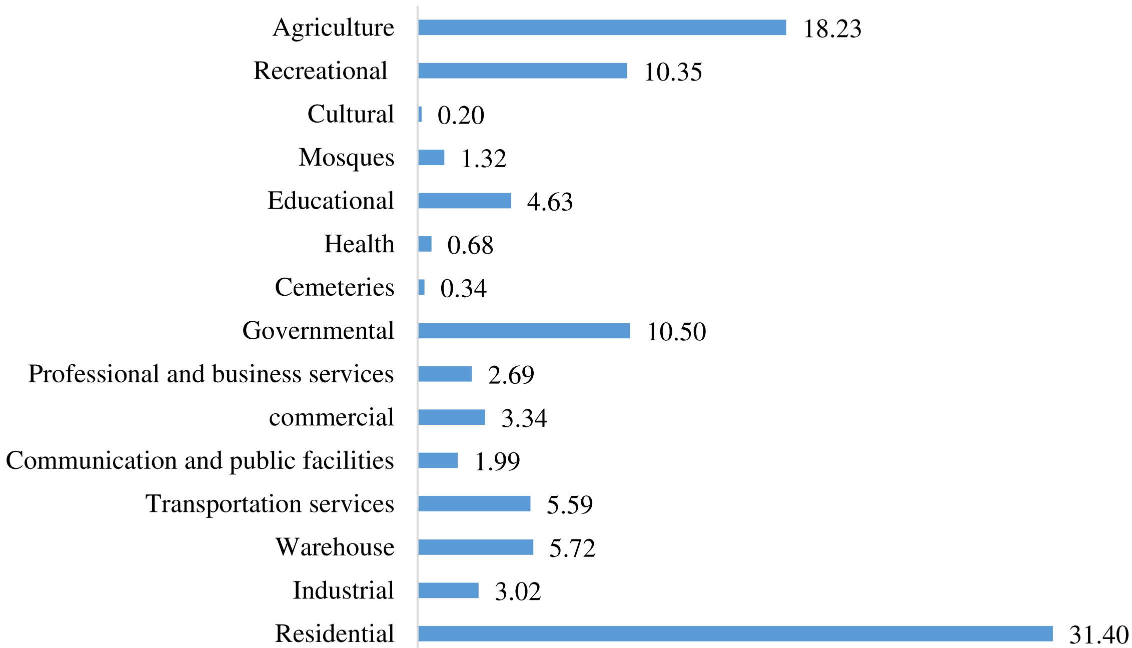


Figure 3-10: Percentages of Land-Use Categories in Riyadh

3.2 Data preparation

The ESRI ArcMap 10.3 GIS and SAS softwares were used to integrate, aggregate, and manage all these datasets. The layer of road network does not cover six HAYs out of the 185 HAYs; these six HAYs were omitted from the analysis, thus, 179 out of 208 units were utilized in the analysis.

Prior to the model estimation, transformation of some variables is important to reduce the variance and minimize the heteroscedasticity and correlation among the variables (Lee et al., 2015b; Quddus, 2008). For example, VKT and population density are transformed into the natural logarithmic, and numbers of young and elderly people are transformed into a proportion. Descriptive statistics of the data are listed in Table 3-1.

Table 3-1: Descriptive statistics of the data (N=179)

Type	Description	Mean	Std Dev	Min	Max
Dependent Variables	Total fatal crashes	7.430	10.027	0	78
	Total injury crashes	32.486	35.094	0	247
	Total PDO crashes	1378.940	1245.290	0	7997
	Total pedestrian crashes	14.251	13.666	0	75
	Total bicycle crashes	3.670	4.062	0	22
	Total single-vehicle crashes	258.145	333.735	0	2212
	Total multi-vehicle crashes	1138.550	1009.260	0	5874
	Severe Pedestrian Crashes	8.201	10.597	0	70
	Severe Crashes	33.905	38.104	0	267
Traffic Demand	Vehicle-Kilometers-Traveled	49878.49	36206.59	566.444	193305.31
	Log of Vehicle- Kilometers -Traveled(LVKT)	10.472	1.007	6.339	12.172
Socioeconomic and Demographic	Population density (population/m2)	0.008	0.008	0	0.043
	Log of Population Density(LPoD)	-5.857	2.046	-12.736	-3.14
	Proportion of Non-Saudi people(NonSa)	0.407	0.300	0	1.00
	Proportion of young people (15–24 years old) (YNGP)	0.201	0.088	0	0.8
	Proportion of elderly people (65 years old or older) (EldGP)	0.020	0.018	0	0.143
	Proportion of Retired people(RetP)	0.023	0.019	0	0.08
	Proportion of Households without vehicles(VEH0P)	0.023	0.039	0	0.257
	Proportion of Illiteracy (LitP)	0.038	0.073	0	0.667
	Proportion of unemployed people(UEmP)	0.024	0.046	0	0.571
	Proportion of people whose educational attainment higher than high school(UniUpP)	0.19	0.115	0	0.747
Land-Use	Residential Area(Res)	0.240	0.163	0	0.530
	Commercial Area(Com)	0.024	0.035	0	0.200
	Educational Area(Edu)	0.025	0.043	0	0.336
	Governmental Area(Gov)	0.025	0.052	0	0.266
	Agricultural Area(Agr)	0.032	0.110	0	0.904
Road Network	Number of traffic signals(SIs)	1.905	2.981	0	19.00
	Proportion of divided Collector Roads(CLDP)	0.223	0.196	0	0.752
	Proportion of Undivided Collector Roads(CLUDP)	0.205	0.237	0	1.00
	Proportion of Collector Roads(CLP)	0.425	0.234	0	1.00
	Proportion of Freeway Roads(FWP)	0.138	0.166	0	1.00

In addition, multicollinearity between independent variables was examined using the most commonly used Pearson correlation coefficient, which measures the strength of a linear association between two continuous variables (Table 3-2). A high correlation between independent variables can lead to high uncertainty and a large standard error. In this study, the correlation between two variables was deemed high if the absolute value of the Pearson correlation coefficient was 0.6 or higher according to Evans (1996). Moreover, multicollinearity has been detected by using the variance inflation factors (VIFs) method for all independent variables. As a common rule of thumb, a variable is considered highly collinear if the VIF exceeds 10 (Chatterjee and Hadi, 2015; Gujarati, 2009). In this study, all VIF values were less than 2, indicating no evidence of multicollinearity. Therefore, considering all above variables as candidate variables to be included in the model, only statistically credible variables were left in the final models.

Table 3-2: Pearson and Spearman correlation coefficients for variables

Var	LVKT	LPoD	NonSa	UniUpP	RetP	VEH0P	LitP	UEmP	YNGP	EldGP	Res	Com	Edu	Gov	Agr	SIs	CLUDP	CLDP	CLP
LVKT	1	-0.126	0.312	0.099	0.327	-0.010	0.026	0.103	0.131	0.227	-0.022	0.256	-0.013	0.077	-0.088	0.368	-0.208	0.272	-0.147
LPoD	-0.059	1	0.716	-0.036	0.085	0.394	0.220	0.218	-0.083	0.169	0.740	0.569	0.583	0.136	-0.333	0.300	0.322	0.209	0.436
NonSa	0.262	0.478	1	0.051	0.165	0.381	0.230	0.235	-0.119	0.254	0.521	0.545	0.390	0.064	-0.206	0.498	0.186	0.236	0.291
UniUpP	0.119	-0.007	-0.055	1	0.111	-0.303	-0.552	-0.260	-0.205	0.108	0.210	-0.047	0.130	0.169	-0.198	-0.035	-0.044	0.148	-0.028
RetP	0.235	0.209	-0.074	0.046	1	-0.198	0.157	0.343	0.396	0.617	0.265	0.174	0.199	-0.105	-0.015	0.153	-0.192	0.307	-0.004
VEH0P	-0.156	0.065	0.073	-0.204	-0.372	1	0.267	0.087	-0.282	-0.030	0.131	0.299	0.216	0.106	-0.074	0.219	0.235	-0.016	0.178
LitP	0.000	-0.244	-0.035	-0.361	-0.118	-0.023	1	0.318	0.117	0.154	0.132	0.205	0.151	-0.019	0.105	0.070	0.119	0.058	0.161
UEmP	-0.030	-0.109	-0.020	-0.251	0.003	-0.098	-0.021	1	0.353	0.450	0.175	0.219	0.228	0.055	0.090	0.214	0.146	0.114	0.245
YNGP	0.101	0.050	-0.084	-0.257	0.319	-0.367	0.207	0.192	1	0.211	-0.024	0.040	0.018	-0.139	-0.039	-0.020	-0.261	0.239	-0.020
EldGP	0.052	0.122	0.031	-0.020	0.456	-0.251	-0.109	0.555	0.194	1	0.275	0.263	0.200	0.012	0.035	0.182	0.066	0.096	0.087
Res	-0.028	0.723	0.363	0.209	0.224	-0.069	-0.136	-0.080	0.023	0.184	1	0.494	0.603	0.212	-0.337	0.264	0.232	0.282	0.380
Com	0.084	0.326	0.207	-0.124	-0.025	0.137	0.019	0.012	-0.010	0.064	0.239	1	0.479	0.163	-0.285	0.398	0.286	0.141	0.260
Edu	-0.002	0.174	-0.002	0.104	0.054	-0.059	-0.061	-0.014	0.104	0.011	0.169	0.053	1	0.378	-0.257	0.204	0.272	0.188	0.345
Gov	0.055	-0.023	-0.049	0.224	-0.228	0.002	-0.092	-0.091	-0.052	-0.111	-0.016	-0.054	0.340	1	-0.265	0.135	0.251	0.025	0.209
Agr	-0.241	-0.250	-0.160	-0.099	-0.172	-0.090	0.012	0.152	-0.137	-0.080	-0.288	-0.066	-0.120	-0.110	1	-0.135	0.102	-0.255	-0.062
SIs	0.341	0.205	0.489	-0.016	0.095	-0.021	-0.065	-0.011	-0.006	0.126	0.162	0.159	-0.008	-0.059	-0.127	1	0.139	0.153	0.182
CLUDP	-0.326	0.214	0.172	-0.098	-0.227	0.091	0.050	0.024	-0.198	0.019	0.133	0.252	0.032	0.071	0.374	-0.003	1	-0.355	0.573
CLDP	0.211	0.294	0.091	0.084	0.287	-0.087	-0.133	-0.063	0.199	-0.021	0.259	-0.077	0.126	-0.047	-0.250	0.085	-0.435	1	0.429
CLP	-0.147	0.472	0.253	-0.032	0.019	0.022	-0.059	-0.026	-0.018	0.008	0.358	0.193	0.140	0.035	0.167	0.072	0.650	0.407	1

Note: The numbers in the lower left triangle of the matrix are the Pearson correlations, and the ones in the upper right are the Spearman correlations.

CHAPTER 4: STATISTICAL METHODOLOGIES

4.1 Introduction

Initially, a traffic crash is, in theory, the output of a Bernoulli trial. Each time a vehicle enters any type of entity (a trial), e.g., an intersection, a highway segment, etc., on a given transportation network, it will result in either a crash or no crash. For the sake of consistency, a crash is termed a "success" when occur and a "failure" otherwise; more details can be found in the study of Lord et al. (2005). Because crash count data is discrete, random events and non-negative integers, the Poisson distribution is employed to estimate the crash occurrence model. Based on that, several statistical frequency models have been adapted.

For data collected for areal units, many statistical modeling approaches have been developed. One of the advantages of using crash frequency models is the ability to predict crash count levels over spatial units (e.g., census units, wards, counties, traffic analysis zones (TAZs)). The macro-level crash prediction models provide transportation planners with a wide-spectrum perspective to consider safety in the long-range transportation planning process (Washington, 2006). Statistical methodology is a vital tool to investigate the association between traffic crash frequency and contributing factors for crash data that are characterized by 1) geographically referenced based on the crash locations or varying over geographic units when modeled; 2) correlation among crash count levels; and 3) temporally correlated. Unfortunately, the latter one cannot be incorporated due to lack of data for several years. Therefore, in this research, a wide array of statistical models at the macro-level will be developed for crash frequency modeling, including traditional regression models, Poisson and Poisson-gamma (PG) models, Poisson-lognormal (PLN), random-effects and spatial models for univariate and multivariate crash counts. In addition, the spatial variations can be investigated using two common approaches; geographically weighted regression and random parameter or spatially varying coefficients models.

4.2 Local variations of parameters

Because regional crash data is observed and collected at specific geographic units (e.g., number of crashes per zones, socio-economic and land-use changes across census units, etc.), spatial dependency (autocorrelation) may occur between the observations. In addition, when modeling the relationship between crash counts and contributing factors, spatial heterogeneity may develop due to spatial correlation among neighboring sites. Therefore, the impact of local variations of parameters in spatial data can be investigated using more commonly employed approaches: geographically weighted regression (GWR) and random parameter (RP) or spatially varying coefficients (SVC) models. That is to account for the spatial dependency (autocorrelation) and heterogeneity in crash count modeling at the zonal level.

4.2.1 Geographically Weighted Regression

4.2.1.1 Overview and Model Development

The most commonly used global negative binomial regression is NB-2, and it can be obtained by introducing heterogeneity into the conditional mean of the Poisson as follows (Greene, 2007; Hilbe, 2011):

$$y_j \sim NB\left[t_j exp(\textstyle\sum_k \beta_k \, x_k), \alpha\right] \tag{1}$$

where y_j is the response variable, i.e., number of observed crashes per HAY j, α represents the overdispersion parameter, and β_k is the model fixed estimate related to the corresponding covariate x_k, for k = 1, …, k.

The local negative binomial regression, i.e., geographically weighted negative binomial regression (GWNBR) is the extension of the global negative binomial in Eq. (1) as the following formula. Readers are referred to da Silva and Rodrigues (2016) for more details:

$$y_j \sim NB\left[t_j exp\left(\sum_k \beta_k \left(u_j, v_j\right) x_{jk}\right), \alpha\left(u_j, v_j\right)\right] \tag{2}$$

where (u_j, v_j) are the locations (coordinates) of the HAY's centroids j, for j = 1, …, n. A special case of this methodology is the Geographically Weighted Negative Binomial Regression with global overdispersion (GWNBRg), in which only the β_k are allowed to spatially vary.

4.2.1.2 Model Comparison

The Corrected Akaike Information Criterion (AICc) can be employed to select the best model where smaller values indicate the preferred model (McMillen, 2004; Nakaya et al., 2005).

4.2.2 Spatially varying coefficients model

A wide array of spatial statistical approaches have been used to account for spatial dependence in crash count models. Mainly, the Bayesian hierarchical models are employed in such analyses by including a set of random effects at the second level of hierarchy to incorporate spatial correlation (Xu et al., 2017). The Gaussian Conditional Autoregressive (CAR) prior with a normal $(\overline{\emptyset}_i, \tau_i^2)$, which was originally proposed by Besag (1991), is commonly used to model this effect.

One of the alternative formulations for the CAR model is the one proposed by Leroux et al. (2000). As opposed to CAR, which mainly formulates the random effect into two components (i.e., heterogeneity effects and spatial correlation effects, which will be discussed in the following section (4.3)), the authors used only one random effect. Lee (2011) compared four of the most commonly used conditional autoregressive prior distributions in Bayesian analysis, including these two models. The author found that the model proposed by Leroux et al. (2000) was the best overall, as it consistently presents better in the presence of independence and strong spatial correlation.

The formulations of the model are as follows (Congdon, 2014):

$$\log(\mu_i) = \beta_0 + \sum_{J=1}^{J} \beta_j X_i + \varepsilon_i \quad (3)$$

$$\varepsilon_i | \varepsilon_{[i]} \sim N\left(\frac{\rho}{1-\rho+\rho d_i} \sum_{j\epsilon\delta_i} \varepsilon_i, \frac{\sigma_\varepsilon^2}{1-\rho+\rho d_i}\right) \quad (4)$$

where $\rho(0 \leq \rho \leq 1)$ presents the spatial correlation parameter and where $\rho = 0$ indicates independence, while $\rho = 1$ the Leroux priors for the spatial effects correspond to the CAR model (Banerjee et al., 2014), d_i is the number of neighbors of HAY i.

$$\rho \sim Uniform(0,1) \quad (5)$$

$$\sigma_\varepsilon^2 \sim Gamma(1,0.001) \quad (6)$$

For the SVC model, the coefficients were spatially varying as follows:

$$\log(\mu_i) = \beta_0 + \sum_{J=1}^{J} (\beta_j * X_{ji} + \beta^*{}_{ji} * X_{ji}) \quad (7)$$

where β^*_{ji} represents the corresponding estimate of spatially varying coefficient over HAY i for J (j=1, 2, …, J) covariate X_{ji} with mean μ_{ji} and precision matrix $\Sigma_{\tau j*j}$ with a hyper-prior $Rj * j$ estimated by a Wishart distribution as follows:

$$\mu_{ji}|\mu_{k\in\partial_{ji}} = \frac{\rho \sum_{k\in\partial_i} \mu_{jk}}{1-\rho+\rho d_i} \quad (8)$$

$$\Sigma_{\tau j*j} = [1 - \rho + \rho d_i]Rj * j \quad (9)$$

4.2.2.1 <u>Model Comparison</u>

The Deviance Information Criterion (DIC) is mainly used for model comparison. Small values of the (DIC) suggest preferred models (Guo and Carlin, 2004).

Other goodness-of-fit measures were employed for the comparison between all developed models: Mean Absolute Deviation (MAD) and Root Mean Squared Errors (RMSE). The (MAD) can be obtained by dividing the sum of absolute errors by the size of the sample as follows:

$$MAD = \frac{1}{n}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_i\right| \quad (10)$$

where N is the size of the sample, Y_i and $\hat{Y}_i$ are the observed and predicted values, respectively.

The RMSE can be obtained by the square root of the total of the squared errors divided by the size of the sample, as shown in the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \quad (11)$$

Models with smaller (MAD and RMSE) values are preferred.

4.3 Full Bayesian analysis

Recent developments in Markov chain Monte Carlo (MCMC) allow a Full Bayes analysis of sophisticated models for complex spatial data (Banerjee et al., 2014). A Full Bayes hierarchical technique (FB) was used to estimate univariate and multivariate crash models with and without spatial effects.

Poisson regression is appropriate to model crash count data since it is random, taking integer values and non-negative events. The probability of areal unit$_i$ (i.e., HAY$_i$ in this study) having y_i crashes per time period is given by the following formula (Lord and Mannering, 2010):

$$\mathrm{P}(y_i) = \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!} \tag{12}$$

where $\mathrm{P}(y_i)$ is the probability of HAY$_i$ having y_i crashes per time period, and μ_i is the Poisson parameter, which shows the expected number of crashes per period.

Poisson regression models are estimated by specifying the Poisson parameter μ_i as a function of explanatory variables, with the most broadly used functional form being:

$$\mu_i = \exp(\beta X_i) \tag{13}$$

where X_i is a row vector of explanatory variables indicating characteristics of HAY$_i$, and β is a coefficient estimate of model covariates βX_i.

One problem when using the Poisson regression model is not accounting for overdispersion where the variance of the crashes is limited to being equal to the mean. Most crash frequency data is overdispersed where the variance is greater than the mean. To overcome this issue, the Poisson-gamma (PG) modeling methodology is appropriate because it includes a disturbance term, which accounts for the overdispersion (Agresti and Kateri, 2011; Washington et al., 2010). Hence, several studies employed it (Abbas, 2004; Abdel-Aty and Radwan, 2000; Amoros et al., 2003; Cafiso et al., 2010; Carson and Mannering, 2001; El-Basyouny and Sayed, 2006; Hauer and Hakkert, 1988; Karlaftis and Tarko, 1998; Kim and Washington, 2006; Kulmala, 1995; Lee et al., 2013; Lord et al., 2010, 2005; Lord and Bonneson, 2007; Malyshkina and Mannering, 2010; Maycock and Hall, 1984; Miaou, 1994; Miaou and Lord, 2003; Milton and Mannering, 1998; R. B. Noland and Quddus, 2004; Persaud and Nguyen, 1998; Poch and Mannering, 1996; Pulugurtha et al., 2013; Shankar et al., 1998, 1995; Ukkusuri et al., 2011):

$$\mu_i = \exp(\beta X_i + \theta_i) \tag{14}$$

where θ_i is a gamma distributed error term with mean 1 and variance σ^2.

However, Poisson-lognormal (PLN) regression has been widely used by many researchers, as it is more flexible than the Poisson-gamma to handle the overdispersion issue (Abdel-Aty et al., 2013; Lord and Mannering, 2010; Lord and Miranda-Moreno, 2008). The Poisson-lognormal model has a structure similar to the Poisson-gamma model except that the $\exp(\theta_i)$ term used to compute the Poisson parameter is assumed to follow lognormal distribution instead of gamma distribution. Specifically, θ_i is suggested to follow a normal distribution with zero mean and variance σ^2.

4.3.1 Univariate Bayesian Spatial Poisson-lognormal Modeling

A Bayesian Spatial Poisson-lognormal model is identified as follows:

$$y_i \sim \text{Poisson}(\mu_i) \tag{15}$$

$$\log(\mu_i) = \beta_0 + \sum_{J=1}^{J} \beta_j X_i + \theta_i + \emptyset_i \qquad (16)$$

$$\theta_i \sim \mathrm{Normal}(0, \tau_\theta) \qquad (17)$$

Where y_i is the number of observed crashes, β_0 is the intercept term, β's are the fixed effect parameters, X_i denotes a vector of covariates in the i^{th} HAY, θ_i is the error term of the model that accounts for overdispersion or unobserved heterogeneity (because of excluded variables in the model) across HAYs, $\emptyset_i$ is the spatial autocorrelation error term in which both variables represent the random effects (Aguero-Valverde and Jovanis, 2006), and τ_θ is the precision parameter which is the reciprocal of the variance, and it follows a prior gamma (0.5, 0.0005). This variance ($1/\tau_\theta$) provides the amount of variation not explained by the Poisson assumption (Lawson et al., 2003). Moreover, a non-informative normal (0, 100,000) prior is assumed for β_0 and β's.

For the spatial distribution, the Gaussian Conditional Autoregressive (CAR) prior with a normal ($\overline{\emptyset}_i, \tau_i^2$), which was originally proposed by Besag (1974), was used to model spatial correlation (Banerjee et al., 2014). Mean of $\overline{\emptyset}_i$ is defined by:

$$\overline{\emptyset}_i = \frac{\sum_{i \neq j} \emptyset_j \times W_{ij}}{\sum_{i \neq j} W_{ij}} \qquad (18)$$

$$\text{Where,} \qquad W_{ij}(\text{the matrix of spatial weights}) \begin{cases} 1, & \text{i and j are adjacent} \\ 0, & \text{Otherwise} \end{cases}$$

The impact of spatial correlation proportion is calculated as follows (Huang et al., 2010):

$$\alpha = \frac{\sigma_{\emptyset_i}}{\sigma_{\theta_i} + \sigma_{\emptyset_i}} \qquad (19)$$

Where $\sigma_{(.)}$ is the empirical marginal standard deviation function.

Univariate crash frequency models could be developed for specific crash levels; however, if there are correlations between crash levels, a potential statistical issue might develop. Thus, multivariate models are necessary in crash-frequency modeling because they explicitly consider the potential correlation among crash counts by severity or type level (Lord and Mannering, 2010).

4.3.2 Multivariate Bayesian Spatial Poisson-lognormal Modeling

Crash count data can be classified into k categories (k=1, 2, …, k) (e.g., crash severity counts, i.e., fatality, injury, and PDO, and crash type counts, i.e., pedestrian, bicycle, single-vehicle, and multi-vehicle). Because crash outcomes are rare, the observed counts of crashes Y_{ik} of the crash category k in HAY_i (i=1, 2, ..., 179) are assumed to follow independent Poisson distributions, conditional on expected Poisson rate μ_{ik}.

$$Y_{ik} \sim Poisson\ (\mu_{ik}) \tag{20}$$

The Poisson rate is modeled as a function of the covariates following a log-normal distribution, as follows (Aguero-Valverde, 2013; Thomas et al., 2004):

$$\log(\mu_{ik}) = \beta_{0k} + \sum_{j=1}^{J} \beta_{kj}\, X_{ij} + \theta_{ik} + \emptyset_{ik} \tag{21}$$

where β_{0k} is the intercept for crash category k, $\beta_{kj}{}'$s are the corresponding regression estimates for J (j=1, 2, …, J) covariate and crash category k, X_{ij} is the value of the J covariate for HAY_i of the crash category k, and θ_{ik} and $\emptyset_{ik}$ denote the unstructured (heterogeneity) and the structured (spatially correlated) random effect term in HAY_i for each crash category k, respectively. The first one is to account for overdispersion, while the latter is to account for spatial correlation. These random effects represent unobserved covariates by capturing the extra-Poisson heterogeneity among HAYs. It is common to incorporate both heterogeneity and the spatial random effects to determine the importance of the spatial dependency. If the spatial random effect dominates the unstructured random effect, then estimated risks will show spatial structure and, if it is the opposite, then the effect will be to shrink the estimated risks towards the overall mean (Besag, 1991). This spatial model borrows strength from neighboring locations when estimating variables, thus improving model estimation (Ancelet et al., 2012).

Obtaining the fully Bayesian posterior estimates requires the specification of prior distributions. The intercept β_{0k} and the regression coefficient β_{kj} were assigned highly non-informative Normal priors (Aguero-Valverde and Jovanis, 2009):

$$\beta_{0k} \sim N(0,10{,}000) \tag{22}$$

$$\beta_{kj} \sim N(0,10{,}000) \tag{23}$$

Following, specifications of prior distributions for random effects used in the univariate model were specified and contrasted to their corresponding multivariate model.

4.3.2.1 <u>Priors specification for the heterogeneity effects in the univariate model (PLN)</u>

The univariate Bayesian spatial model assumes that the random effects for different crash categories are independent, which does not account for the correlations between heterogeneities of crashes by categories *k*. The prior distributions for the heterogeneity θ_{ik} was proposed as an independent normal distribution with an expected mean of zero and a variance of $\tau_{k\theta}$:

$$\theta_{ik} \sim N\left(0, \tau_{\theta k}^{-1}\right) \tag{24}$$

where $\tau_{\theta k}$ is the precision parameter, which is the reciprocal of the variance for each crash category k, and follows a commonly used gamma distribution as suggested by Wakefield et al. (2000) and used by Aguero-Valverde and Jovanis (2006):

$$\tau_{\theta k} \sim Gamma(0.5, 0.0005) \qquad (25)$$

This variance ($1/\tau_{\theta k}$) provides the amount of variation not explained by the Poisson assumption.

4.3.2.2 Priors specification for the spatial effects in the univariate spatial model (PLN-CAR)

The previous model assumes sites are independent of one another and ignore the spatial interactions among sites. In contrast, spatial dependency or autocorrelation exists among the regional crash observations because, by nature, the crash data is observed with reference to location (Lai et al., 2008; LeSage, 1997; Quddus, 2008). Neglecting spatial autocorrelations may invalidate the assumption of the random distribution (LeSage and Pace, 2004) and hence lead to a biased estimation of the model parameters. Several traffic safety studies have found that spatial autocorrelations exist in the crash count data (Aguero-Valverde, 2013; Hadayeghi et al., 2010b; LaScala et al., 2000; Levine et al., 1995).

Spatial data from near locations is more likely to be similar than data from distant locations, which is known as spatial autocorrelation (O'Sullivan and Unwin, 2003; Schabenberger and Gotway, 2004). A spatial autocorrelation or dependence may exist if a particular variable of a geographic location is affected by the same variable of the adjacent location. The presence of spatial autocorrelation between neighboring units may introduce bias into the statistical analyses because of the violation of the assumption of unit independence (LeSage and Pace, 2004).

Although Poisson log-normal models can handle the effect of unobserved heterogeneity among HAYs, such models cannot account for the spatial autocorrelation. In order to account for the spatial autocorrelation, a spatial error term ($\emptyset_i$) was involved in the model specification.

Two common models that incorporate spatial dependence are the simultaneously autoregressive (SAR), which was originally developed by Whittle (1954), and the conditionally autoregressive (CAR), which was proposed by Besag (1974). The SAR model considers adding an explanatory variable in the form of a spatially lagged dependent variable or adding spatially lagged error structure into a linear regression model using likelihood methods. In contrast, the CAR model formulates the random effect into two components to account for both spatial dependence and uncorrelated heterogeneity using Gibbs sampling (Banerjee et al., 2014). The intrinsic Gaussian Conditional Autoregressive (ICAR) prior with a normal ($\overline{\emptyset}_{ik}, \tau^2_{\emptyset k}$), which was originally proposed by Besag (1974), was used to model spatial correlation (Banerjee et al., 2014). Mean of $\overline{\emptyset}_{ik}$ is defined by:

$$\overline{\emptyset}_{ik} = \frac{\sum_{i \neq j} \emptyset_{jk} \times W_{ij}}{\sum_{i \neq j} W_{ij}} \tag{26}$$

where, W_{ij} (the matrix of spatial weights) $\begin{cases} 1, & \text{i and j are adjacent} \\ 0, & \text{Otherwise} \end{cases}$

This specification is essentially equivalent to the specification of the univariate Poisson log-normal for each crash category. Nevertheless, it offers the possibility of being directly comparable with the multivariate specifications.

4.3.2.3 Priors specification for the heterogeneity effect in the multivariate model (MVPLN)

Unlike the univariate model specification for the heterogeneity effect, the multivariate model takes into account the correlations across random effects. Therefore, the main differences between the multivariate model and the univariate model are the prior specifications of the random effects.

For the multivariate model, correlated priors in the heterogeneity random effects vector are estimated using multivariate normal priors (Aguero-Valverde, 2013; Ma and Kockelman, 2006; Park and Lord, 2007):

$$\theta_i = MN(\mu_i, \Sigma_\theta) \tag{27}$$

where μ_i is a vector with all elements zeroes and Σ_θ is the variance–covariance matrix with a hyper-prior estimated by a Wishart distribution. The Wishart distribution is commonly used as a conjugate prior for the inverse of the variance–covariance parameters of multivariate normal distributions (Gelman et al., 2003). The diagonal elements of the variance–covariance matrix Σ_θ i.e., $\Sigma_{11}, \Sigma_{22}, \dots, \Sigma_{kk}$ represent the heterogeneous variances of $\theta_{i1}, \theta_{i2}, \dots, \theta_{ik}$ respectively, whereas the off-diagonal elements e.g., Σ_{12} and Σ_{21}, represent the heterogeneous covariance between θ_{i1} and θ_{i2}:

$$\Sigma_\theta^{-1} \sim Wishart(R, d) \tag{28}$$

where Σ_θ^{-1} is a symmetric positive definite matrix, also called the precision matrix, while R and d are the scale matrix and the degrees of freedom, respectively. In order to produce a non-informative prior for the precision matrix Σ_θ^{-1}, the following values of $R\ and\ d$ were used (Aguero-Valverde, 2013; Gelman et al., 2003):

$$R = d \times d = \begin{bmatrix} 0.100 & 0.005 & 0.005 & \cdots \\ 0.005 & 0.100 & 0.005 & \cdots \\ 0.005 & 0.005 & 0.100 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, d = 3\ or\ 4 \quad (29)$$

4.3.2.4 Priors specification for the spatial effects in the multivariate spatial model (MVPLN-CAR)

The specification in the multivariate model is a zero-centered multivariate conditional autoregressive (MCAR), which is an extension of the univariate conditional autoregressive model (CAR) in the univariate model (Aguero-Valverde, 2013; Banerjee et al., 2014):

$$\emptyset_{ik} | (\emptyset_{-i1}, \emptyset_{-i2}, \dots, \emptyset_{-ik}) \sim MN\left(\overline{\emptyset_{ik}}, \frac{\Sigma_\emptyset}{n_i}\right) \quad (30)$$

where $\emptyset_i$ is the mean vector $\overline{\emptyset_i} = (\sum_{j \neq i} \omega_{i,j} \emptyset_{j1}/n_i, \sum_{j \neq i} \omega_{i,j} \emptyset_{j2}/n_i)^T$, and $\Sigma_\emptyset$ is the $k \times k$ covariance matrix.

Similar to the univariate model, the intrinsic MCAR model depends only on the adjacency matrix. Analogous to the specification of $\sum_\theta$ in Eq. (27), the diagonal elements of the variance–covariance matrix $\Sigma_\emptyset$ represent the spatial variances, while the off-diagonal elements represent the spatial covariances of different category k. Furthermore, a highly non-informative Wishart distribution is employed for the precision matrix $\Sigma_\emptyset^{-1}$ as defined by Eq. (28).

The posterior correlation between the total random effects is also calculated. In addition, the posterior proportion of variation explained by the spatial effects in univariate and multivariate models is determined as follows (Aguero-Valverde, 2013; Banerjee et al., 2014):

$$\alpha_k = \frac{\mathrm{sd}\ \phi_k}{\mathrm{sd}\ \theta_k + \mathrm{sd}\ \phi_k} \tag{31}$$

where (sd) means the marginal standard deviation.

4.4 Detecting the Existence of Spatial Autocorrelation

In order to identify the existence of spatial autocorrelation among the residuals, Moran's I statistic was used. Moran's I is one of the standard statistics that has been developed to measure strength of spatial association among areal units (Banerjee et al., 2014).

Moran's I is given by the following formula:

$$\mathrm{I} = \frac{\mathrm{n} \times \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{w_{ij}} (\mathrm{y_i} - \bar{\mathrm{y}})(\mathrm{y_j} - \bar{\mathrm{y}})}{(\sum_{\mathrm{i} \neq \mathrm{j}}^{\mathrm{n}} \mathrm{w_{ij}}) \times \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \bar{\mathrm{y}})^2} \tag{32}$$

where n is the number of HAYs units indexed by i and j, and $\bar{\mathrm{y}}$ is the mean of crashes y.

4.5 Model Comparison

The Deviance Information Criterion (DIC) and posterior mean deviance $\bar{\mathrm{D}}$ are used for model comparison. The Deviance Information Criterion is defined in analogy with the Akaike information criterion (AIC) (Spiegelhalter et al., 2002). It is used to assess model complexity and compare different models, and it can be obtained once the model has converged after running a number of iterations.

DIC is defined as:

$$\mathrm{DIC} = \overline{\mathrm{D}} + \mathrm{pD} = 2 * \overline{\mathrm{D}} - \widehat{\mathrm{D}} \qquad (33)$$

where the pD denotes the effective number of parameters, the $\overline{\mathrm{D}}$ is the posterior mean of the deviance $(-2 * \log(\text{likelihood}))$ and the $\widehat{\mathrm{D}}$ is a point estimate of the deviance obtained by substituting in the posterior means $\overline{\theta}$ of θ; therefore, $\widehat{\mathrm{D}} = -2 * \log(\mathrm{p}(\mathrm{y}|\overline{\theta}))$.

Since small values of the deviance indicate good fit, while small values of the effective number of parameters indicate a parsimonious model, small values of the sum (DIC) indicate preferred models (Guo and Carlin, 2004). According to Spiegelhalter et al. (2005), differences in the DIC values of more than 10 might definitely rule out the model with a higher DIC and a difference between 5 and 10 can be considered substantial.

CHAPTER 5: SPATIAL DEPENDENCY AND HETEROGENEITY IN CRASH COUNT MODELING

5.1 Introduction

This chapter compares most commonly used methods for modeling spatially correlated data, i.e., geographically weighted regression (GWR) and spatially varying coefficients (SVC) for severe (fatal and serious injuries) crash data based on the past 2 years (2014 and 2015). Figure 5-1 illustrates the distribution of severe crashes per HAY. Several models were attempted, including global negative binomial (NB) and random parameter (RP) models and locally varying coefficients models, i.e., geographically weighted regression with Poisson distribution (GWPR) and negative binomial regression with global overdispersion (GWNBRg) and local overdispersion (GWNBR). In addition, Bayesian hierarchical models such as SVC and the one proposed by Leroux et al. (2000) as an alternative formulation for the Gaussian Conditional Autoregressive (CAR) prior, which was originally proposed by Besag (1991), were used to model spatial correlation.

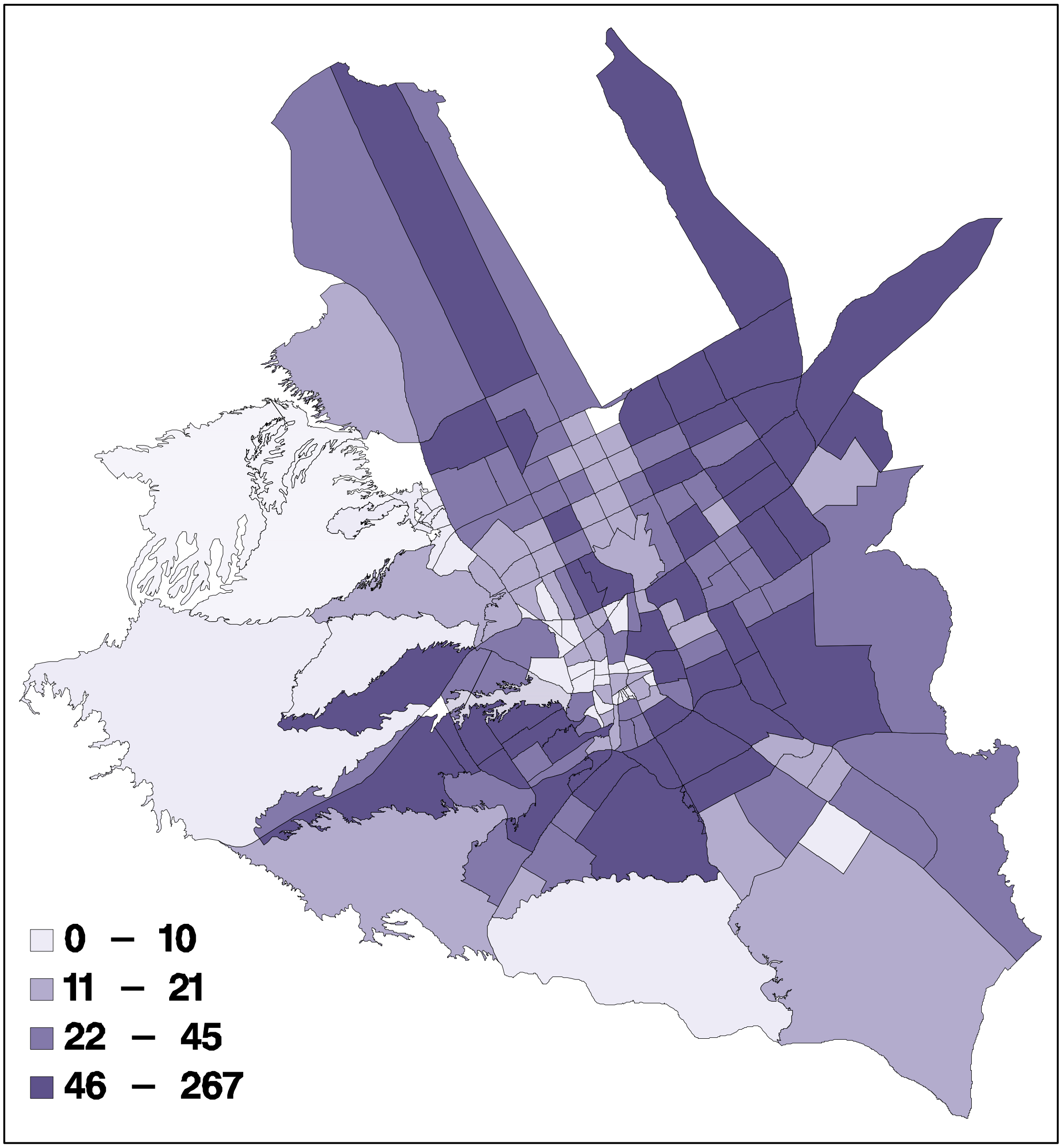


Figure 5-1: Distribution of severe crashes per HAY

5.2 Model development and discussion

Before proceeding to the model estimation, some variables were transformed to reduce the correlation between them as discussed in section 3.2). Table 3-1 presents the descriptive statistics of the data. In addition, the Pearson and Spearman correlation coefficients were calculated to detect if there is any correlation between two variables to avoid multicollinearity due to high correlation between the local variable estimates (Wheeler and Tiefelsdorf, 2005). The results are presented in Table 3-2. Based on the suggested methodologies in part (4.2) and available data, seven models were developed to examine the parameters' variations in crash frequency models. Table 5-1 presents the results of the models and the models' performance measures. Following are the discussions about each model's outputs.

5.2.1 Global negative binomial and random parameter models

Initially, the random effect model is considered as a special case of the random parameter model where the intercept is random, and all other parameters are fixed. However, the random parameter negative binomial (RP-NB) model does not appear to be appropriate to fit the data. That may be due to the small sample size. Moreover, the overdispersion parameter estimated with the NB model is about 0.18. Fitting the RP model with only a random constant revealed a convergence, but the overdispersion increased dramatically with a very high standard error. This extreme result means that the NB-RP model is inappropriate for the data and the Poisson distribution will be used alternatively. That means that the RP-NB model is a Poisson model with two forms of unobserved disturbance, as follows (Greene, 2007);

$$E[y|x,u,e] = \exp(beta'x + u + e) \qquad (34)$$

where u has a log-gamma distribution and e has a normal distribution. Since this is a cross section, there are two random terms (implied) in the Poisson model. Introducing the random effect into a negative binomial model essentially adds a heterogeneity term to a model that is obtained by adding a heterogeneity term to a lower level (the Poisson) model. As such, it will be common that attempts to fit the negative binomial model with random effects will be unsuccessful (Greene, 2016).

The negative binomial (NB) models (global) and Poisson random parameter (RP-P) models (local) were employed using the NLOGIT 6.0 software (Greene, 2007). The RP-P models were estimated by specifying a functional form of the parameter density function and using simulation-based maximum likelihood with 200 Halton draws and some parameters found to be random with best fit of the normal distribution. Based on the results, it is found that the random parameter does not improve the fit of the data compared to the global model (NB).

5.2.2 Geographically weighted regression (GWR)

The SAS macro developed by da Silva and Rodrigues (2014) and (2016) was used to estimate the GWPR, GWNBRg and GWNBR models. In addition, the software GWR 4.0 developed by Nakaya et al. (2012) was used to confirm the results of the GWPR models. The main objective of using the GWNBR model was to account for overdispersion and to reduce the spatial dependency. The adaptive bi-square kernel was used to determine the geographical weights using the nearest-neighbor bandwidth. Both models were fitted using AIC and Cross-Validation (CV) approaches, but the comparison between the models was conducted based on log-likelihood.

The results revealed that the GWNBR model was superior compared to all other developed models in terms of the goodness-of-fit measures for severe crash models. Regarding the MAD and RMSE measures, it can be shown that the GWPR had lower values compared to the GWNBRg and GWNBR, which may be due to two reasons: 1) the GWPR is more vulnerable to extreme values, as can be observed from its smaller bandwidth size, and 2) as can be shown in Figure 5-2 to Figure 5-5, the GWNBRg and GWNBR had a more homogeneous spatial variation of the coefficients than the GWPR model (Gomes et al., 2017). However, they present a better fit in terms of a lower value of the goodness-of-measure (AICc). The findings of these models would be compared with the results of the SVC model, as discussed in the following section. Table 5-2 summarizes the models' results and their performances based on goodness-of-fit measures.

5.2.3 Spatially varying coefficients (SVC)

The software WinBUGS 1.4 was used to estimate the Leroux and SVC models using two chains taken to 10,000 iterations. The first 4,000 iterations were set as a burn-in sample. The results show that the SVC model performed better than the Leroux model in terms of the DIC.

To sum up, the GWNBR model provides the best fit of the data in terms of the goodness-of-fit measure, i.e., the AICc with its GWR counterparts. In addition, in comparing the GWNBR with the two Bayesian spatial models (i.e., Leroux and SVC), the GWNBR had significantly lower values of the MAD and RMSE.

Table 5-1: Summary of the coefficient estimates of the models and the models' performances for severe crashes (1)

Variable\Model	NB		RP-P		GWPR					GWNBRg					GWNBR				
	Mean	**P-value**	**Mean**	**P-value**	**Mean**	**Min**	**Max**	**Lq**	**Uq**	**Mean**	**Min**	**Max**	**Lq**	**Uq**	**Mean**	**Min**	**Max**	**Lq**	**Uq**
Intercept	-6.958	<.0001	-6.95	0.0000	-7.816	-11.967	-3.655	-8.826	-6.811	-7.527	-8.242	-6.068	-7.943	-7.224	-7.728	-9.014	-5.978	-8.023	-7.382
S.D. of Ran. Parameter Dist.			0.043	0.0000															
Log of vehicle-kilometers-traveled	0.687	<.0001	0.693	0.0000	0.732	0.104	1.314	0.568	0.918	0.734	0.534	0.882	0.670	0.803	0.745	0.486	0.922	0.672	0.825
S.D. of Ran. Parameter Dist.			0.041	0.0000															
Log of population	0.285	<.0001	0.283	0.0000	0.291	-0.051	0.814	0.168	0.390	0.285	0.175	0.443	0.222	0.334	0.286	0.137	0.499	0.207	0.344
S.D. of Ran. Parameter Dist.			0.014	0.0000															
Proportion of Freeway Roads lengths	0.607	0.0175	0.616	0.0000	-0.093	-3.068	3.107	-1.078	0.975	0.014	-1.072	1.781	-0.593	0.475	-0.003	-1.380	2.394	-0.722	0.476
Proportion of population aged (15-24 years old)	0.939	0.039	0.546	0.0000	1.235	-2.820	6.852	0.360	1.921	1.004	-0.003	2.565	0.554	1.308	1.058	-0.513	2.666	0.658	1.492
S.D. of Ran. Parameter Dist.			0.080	0.0560															
Proportion of population aged (>=60 years old)	2.772	0.0256	1.79	0.0000	1.825	-9.569	11.321	-0.412	4.560	2.066	-1.257	5.159	0.790	3.494	2.033	-3.506	6.282	0.243	4.010
S.D. of Ran. Parameter Dist.			2.5	0.0000															
Residential Area	-1.525	<.0001	-1.66	0.0000	-0.577	-2.243	1.335	-1.207	0.056	-1.126	-1.938	-0.366	-1.531	-0.702	-0.980	-2.043	0.079	-1.500	-0.441
Number of traffic signals	0.035	0.0104	0.032	0.0000	0.029	-0.050	0.186	0.004	0.040	0.022	0.004	0.072	0.012	0.028	0.021	-0.003	0.103	0.009	0.027
Overdispersion	0.188		-		0					0.188					-				
AIC	1364.191		1367.473		1303.668					1333.582					1303.765				
AICc	1365.256		1369.679		1386.619					1340.896					1316.205				
Bandwidth	-		-		54					153					130				
MAD	11.592		11.861		5.243					9.360					8.622				
RMSE	17.994		19.7047		7.601					14.933					13.733				

Table 5-2: Summary of the coefficient estimates of the models and the models' performances for severe crashes (2)

Variable\Model	Leroux				SVC			
	Mean	Std Dev	BCI		Mean	Std Dev	BCI	
			2.5%	97.5%			2.5%	97.5%
Intercept	-6.301	0.263	-6.768	-5.669	-6.051	0.428	-6.954	-5.286
D. of Ran. Parameter Dist.	-	-	-	-	2.648	1.064	0.568	0.703
Log of vehicle-kilometers-traveled	0.645	0.028	0.590	0.710	0.633	0.042	0.568	0.703
S.D. of Ran. Parameter Dist.	-	-	-	-	0.635	0.040	0.570	0.281
Log of population	0.239	0.024	0.190	0.278	0.229	0.026	0.181	0.280
S.D. of Ran. Parameter Dist.	-	-	-	-	0.229	0.026	0.181	0.948
Proportion of Freeway Roads lengths	0.316##	0.222	-0.132	0.733	0.375##	0.269	-0.124	0.947
Proportion of population aged (15-24 years old)	0.402	0.351	-0.277	1.082	0.535##	0.401	-0.279	1.313
S.D. of Ran. Parameter Dist.	-	-	-	-	0.375##	0.269	-0.125	1.313
Proportion of population aged (>=60 years old)	2.875	0.899	1.101	4.600	2.647	1.064	0.571	4.827
S.D. of Ran. Parameter Dist.	-	-	-	-	0.535##	0.401	-0.280	4.826
Residential Area	-0.660	0.260	-1.152	-0.141	-0.685	0.292	-1.272	-0.119
Number of traffic signals	0.040	0.009	0.021	0.057	0.042	0.011	0.021	0.063
DIC	1171.940				1165.940			
MAD	12.752				12.597			
RMSE	23.951				23.485			

##statistically significant at 80% Bayesian credible interval

Figure 5-2 to Figure 5-5 illustrate the distributions of the local estimates over the 179 HAYs of GWPR, GWNBRg, GWNBR, and SVC models for severe crashes. The parameters demonstrate patterns of spatial variations. However, the maps of the coefficients of the three GWR models display more smoothness than the SVC model. In addition, all local estimates had a positive impact on severe crashes. However, the minimum and maximum of the values vary across all three GWR models. The GWPR produces the highest number of HAYs with negative signs. The number of HAYs that had negative values in GWPR for log of population, proportion of population aged 15-24 years old, and proportion of population aged 65 years and more were 2, 16, and 50, respectively. This indicates that because the GWPR does not account for the overdispersion in crash count data, the model estimates may have unexpected signs (Gomes et al., 2017).

Lastly, Figure 5-6 depicts the distribution of posterior mean of the overdispersion in GWNBR. It can be found that the highest values were distributed in the middle and south, while the lowest values were distributed in the north and north east. The first regions may have unobserved factors that affect severe crashes. This may be explained by the main role of incorporating overdispersion in modeling crash counts, as it is to account for unobserved/omitted variables. For example, the models' results indicated that increases of severe crashes were associated with increases of "logarithm of VKT", some of the areas may have missing information about traffic volumes, e.g., on minor roads. This is also illustrated by the models' performances as discussed earlier, as the GWNBR was superior compared to GWPR and GWNBRg. Therefore, it would be better to allow overdispersion to spatially vary over regions in spatial data, as the model GWNBR showed better fit compared to its counterpart (GWNBRg).

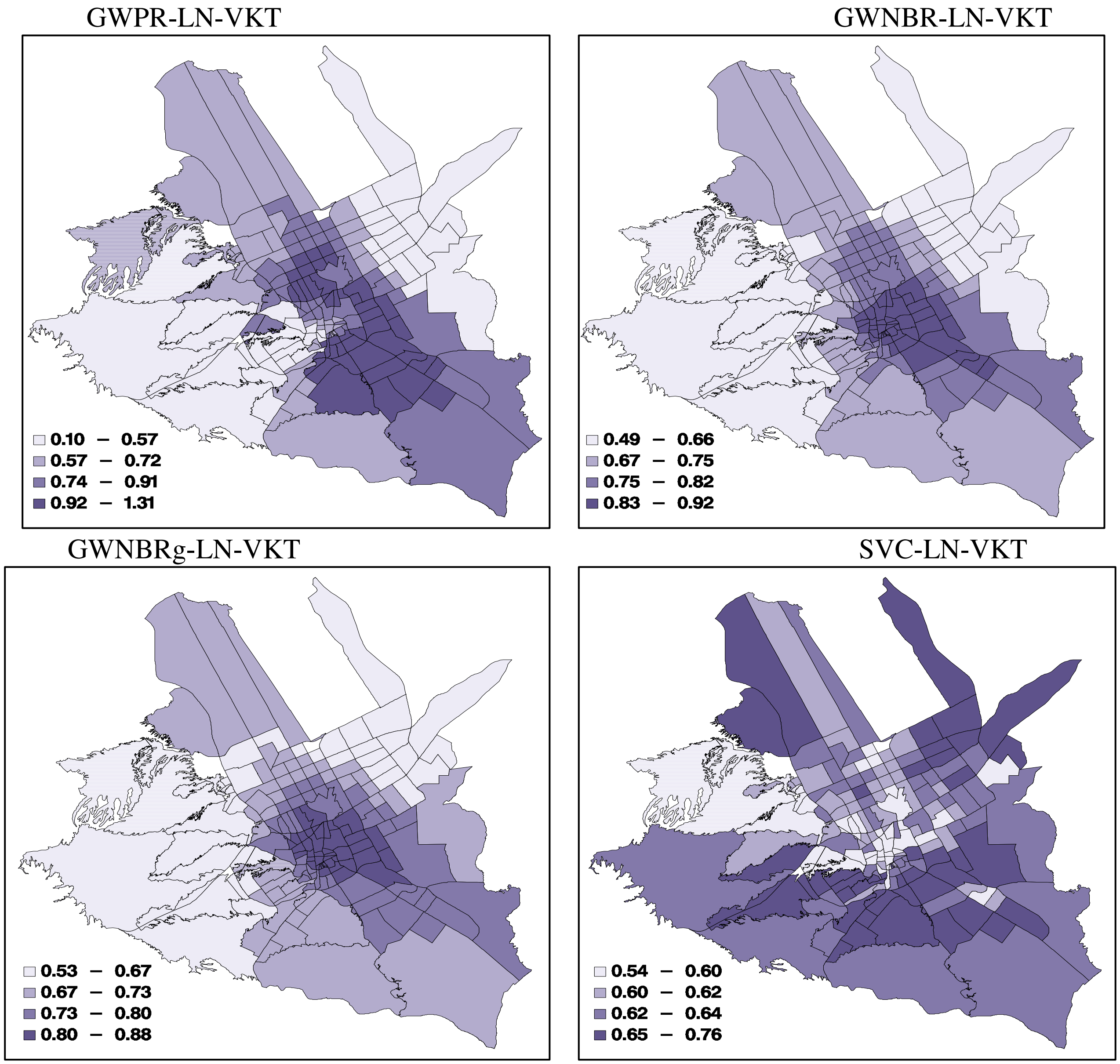


Figure 5-2: Distributions of the Posterior mean of the GWPR, GWNBRg, GWNBR, and SVC models, LN-VKT

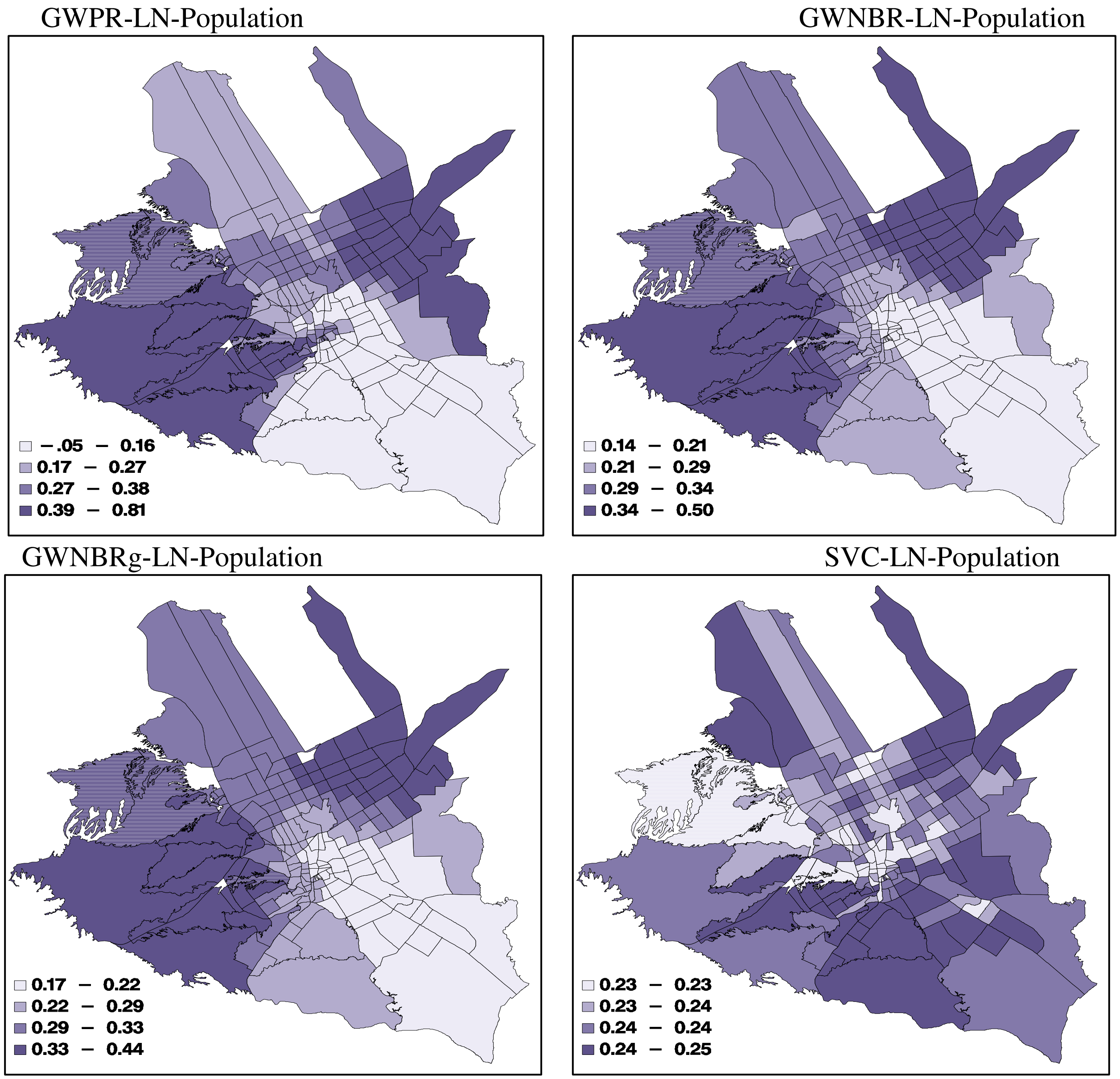


Figure 5-3: Distributions of the Posterior mean of the GWPR, GWNBRg, GWNBR, and SVC models, LN-Population

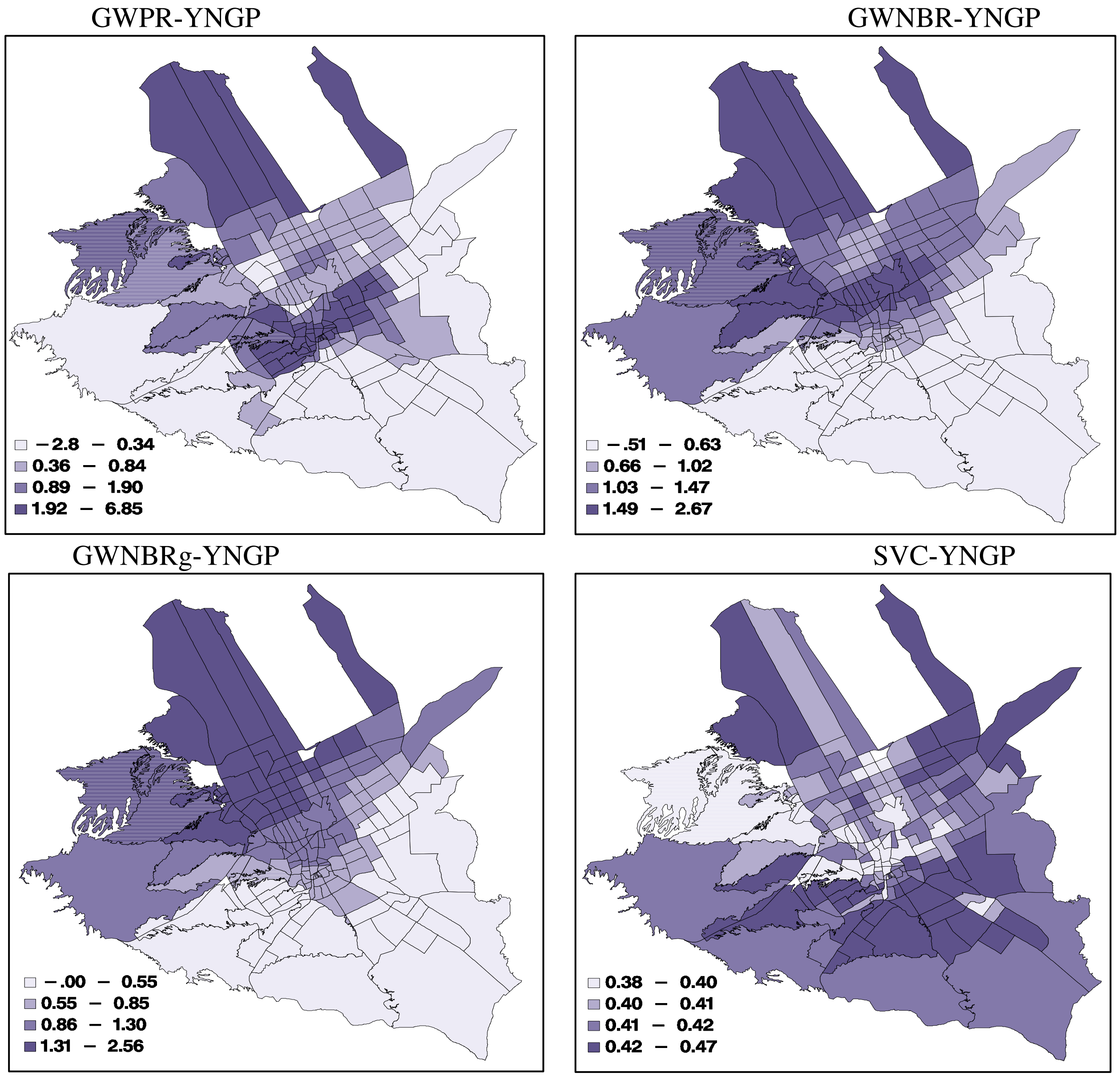


Figure 5-4: Distributions of the Posterior mean of the GWPR, GWNBRg, GWNBR, and SVC models, YNGP

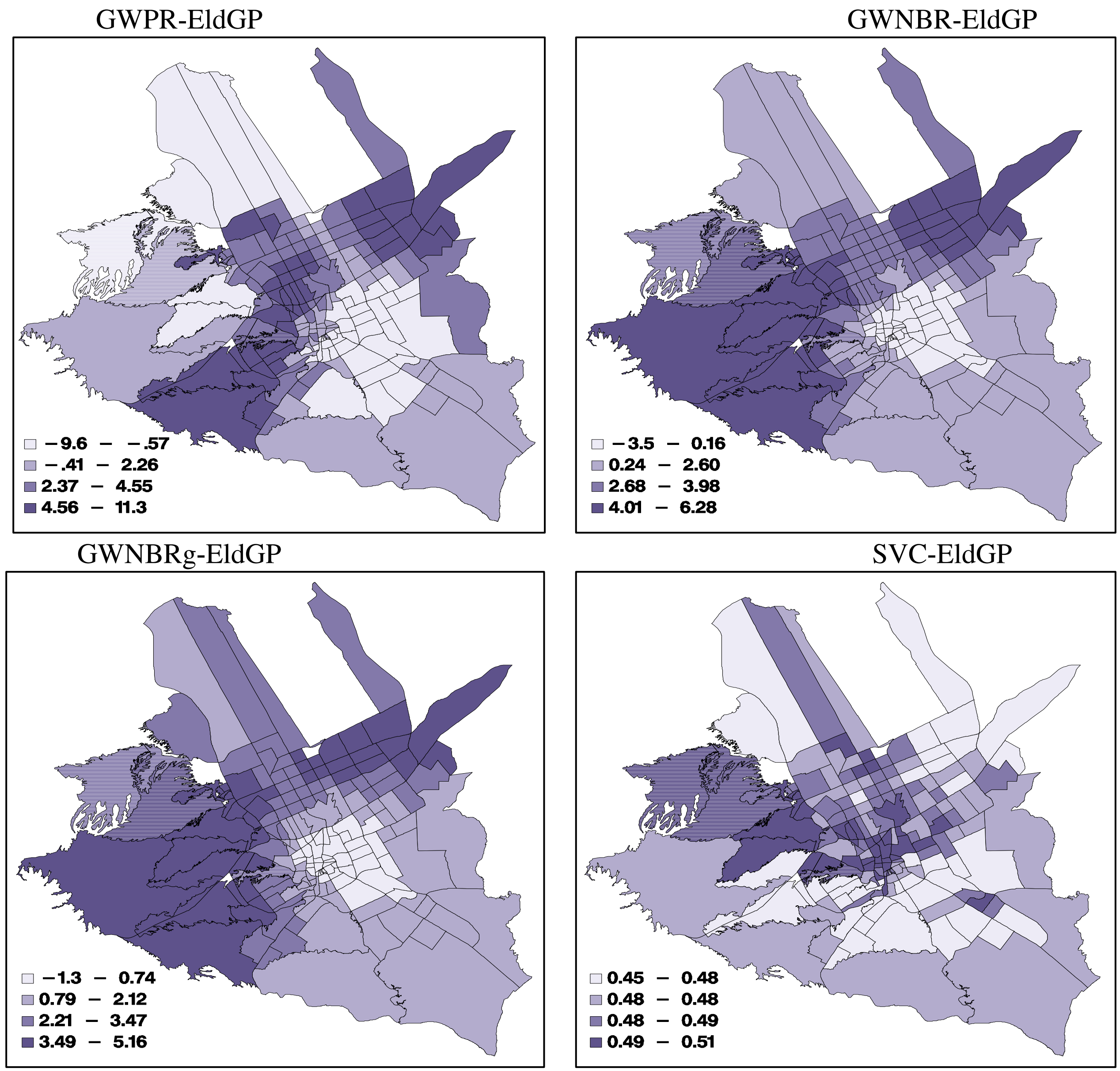


Figure 5-5: Distributions of the Posterior mean of the GWPR, GWNBRg, GWNBR, and SVC models, EldGP

GWNBR-Local Overdispersion

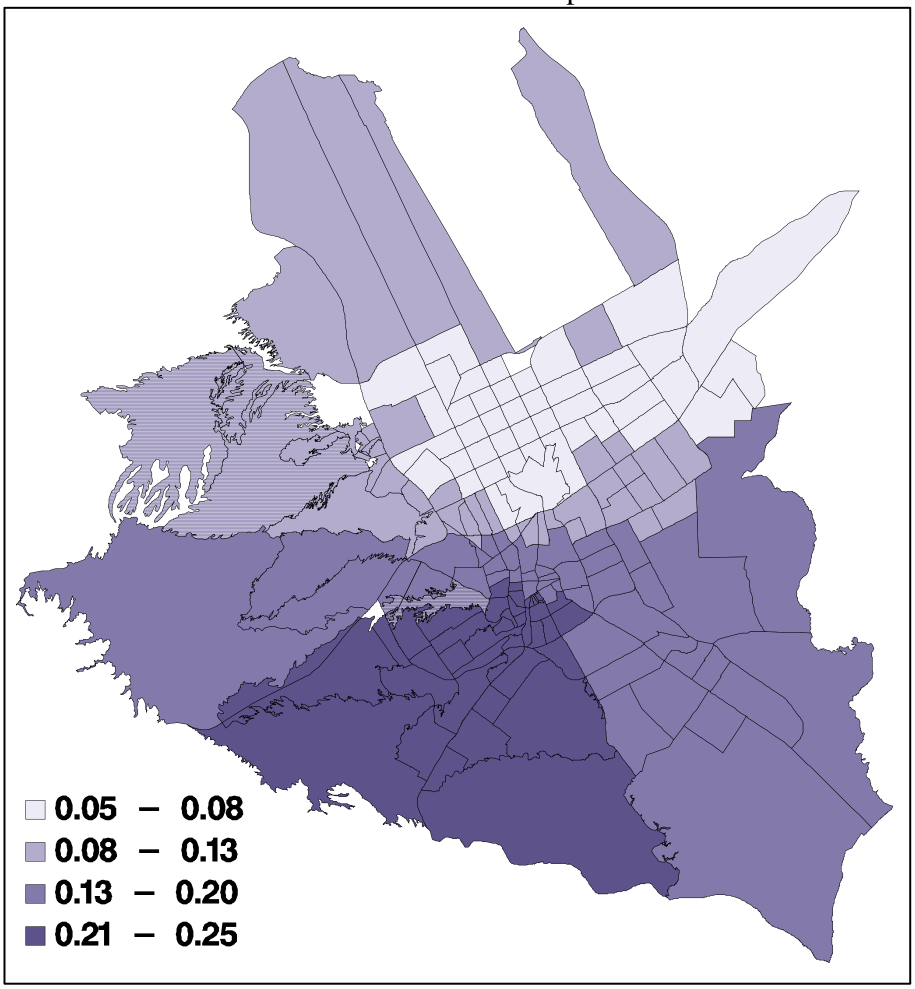


Figure 5-6: Distributions of the Posterior mean of the GWNBR, overdispersion

5.3 Conclusion

This chapter examined the issue of the spatial dependency and heterogeneity in crash count modeling at the zonal level. Two advanced methods that are commonly employed for modeling spatially correlated data, i.e., geographically weighted regression (GWR) and spatially varying coefficients (SVC), were used for severe crashes. The results show that the GWNBR model outperformed all other developed models in terms of fit of the data.

CHAPTER 6: SAFETY INVESTIGATION OF PEDESTRIAN CRASHES AT THE ZONAL LEVEL

6.1 Select Appropriate Pedestrian-Involved Crash Exposure

The most common vehicular exposure in traffic safety is the vehicle-kilometer-traveled. Pedestrian-involved crash frequency is highly affected by traffic volume and population due to the collisions between motor vehicles and pedestrians (Lee et al., 2015a). The population density can be used as a measure of exposure (Cottrill and Thakuriah, 2010). There were two related factors: population density as a socioeconomic surrogate exposure and proportion of residential area as a land-use surrogate exposure, and both of these factors can be used in the model. Both variables after transformation were attempted at the same time in the model beside the vehicular exposure, "i.e., Log of vehicle-kilometers-traveled"; however, due to the high correlation between them, (r = 0.723) as shown in Table 3-2, the most statistically credible variable was used in the model. Table 6-1 indicates that the model using the population density (LPoD) variable had a better goodness-of-fit measure. Both VKT and population density were used in this study as surrogate exposures to pedestrian-involved crashes.

Table 6-1: Selection of Exposure Variables for Total Pedestrian Crashes

	Mean	S.D.	BCI			Mean	S.D.	BCI	
			2.5%	97.5%				2.5%	97.5%
Intercept	-4.1361	0.6078	-5.3274	-2.9448	**Intercept**	-5.8332	0.6468	-7.1010	-4.5654
LNVKT	0.7355	0.0559	0.6259	0.8451	**LNVKT**	0.7402	0.0595	0.6236	0.8568
LPoD	0.2100	0.0379	0.1602	0.2598	**Res**	1.8638	0.3117	1.2529	2.4747
DIC			1144.106		**DIC**			1169.612	

6.2 Model Development

Based on the suggested methodology in part 4.3.1 and available data, the aspatial Bayesian models (PG and PLN) and the Bayesian Poisson-lognormal Conditional Autoregressive (PLN-CAR) model were employed to develop the macro-level safety models. The variables were aggregated at the HAY level, such as VKT, proportion of collector roadways, and number of signalized intersections, in addition to the socioeconomic and land-use variables. Any variables that were consistently not credible have been omitted from the final model. The final model was determined based on the smallest DIC among candidate models.

The spatial distribution of total and severe (i.e., fatal and serious injury) pedestrian-involved crashes varied among HAYs, which indicates different models need to be attempted to identify the contributing factors. Therefore, in addition to the model estimation outcome for pedestrian crashes of all severities combined per HAY, the severe pedestrian crash (fatal and serious injury) model was estimated to examine the factors affecting the severity level of pedestrian-involved crashes, which may differ based on previous studies (Aguero-Valverde and Jovanis, 2006; Quddus et al., 2012). Several models were developed to explore the best model that can estimate the contributing factors that affect pedestrian crashes and all outcomes in accordance with intuition.

The software SAS 9.4 was used to estimate the PG models using procedure GENMOD Bayes analysis. Using one chain, the convergence in these two models was obtained after 10,000 iterations and 2,000 burn-ins size.

The software WinBUGS 1.4 was used to estimate the Bayesian PLN and the Bayesian PLN-CAR models using two chains taken to 30,000 and 20,000 iterations for total and severe pedestrian crash models respectively. The convergence was obtained after 4,000 iterations. The model convergence was assured based on Brooks-Gelman-Rubin statistics, with overlaps among the Markov chains, and density plots. To determine statistical significance of the model parameters, the assessment was based on 95% and 90% Bayesian credible intervals (BCIs). A BCI provides an estimated range of values (interval) that likely contains the true value of the parameter.

6.3 Results and Discussion

The number of credible variables in the severe pedestrian-involved crash model was higher than that of the total pedestrian-involved crashes. The Bayesian Poisson-lognormal Conditional Autoregressive models show substantial improvement compared to the aspatial Bayesian models (Poisson-gamma and Poisson-Lognormal models) for both total and severe pedestrian-involved crash frequency, and it is in line with previous studies (e.g., Amoh-Gyimah et al., 2016; Lord and Miranda-Moreno, 2008; Siddiqui et al., 2012).

The variations explained by the spatial correlations relative to variations in total and severe pedestrian-involved crashes at HAYs were 94.4% and 93.1% respectively.

Table 6-2 presents the Global Moran's *I* statistics calculated from the pedestrian crash models' residuals and the corresponding Z-score and P-value. It showed that the spatial correlation could be controlled by spatial models. The results demonstrated statistically significant spatial autocorrelations at a 99.9% confidence level in aspatial models' residuals, both in total and in severe pedestrian crashes. Thus, the spatial autocorrelation should be accounted for in the estimation of the models. With respect to the total pedestrian crash model, the residuals of the Poisson-lognormal spatial model were revealed to be spatially correlated, which shows the existence of a spatial autocorrelation. However, for the severe pedestrian crash model, there seems to be no significant spatial autocorrelation among spatial Poisson-lognormal model residuals.

Table 6-2: Moran's I Statistics of Pedestrian Crashes' Residuals

Model	**Total Pedestrian Crash**			**Severe Pedestrian Crash**		
	Global Moran's I	**Z-Score**	***P*-value**	**Global Moran's I**	**Z-Score**	***P*-value**
PG	0.247154	5.42797	0.0000	0.148146	3.31719	0.0009
PLN	0.247785	5.44147	0.0000	0.147344	3.28695	0.0010
PLN-CAR	-0.108031	-2.20010	0.0278	-0.063218	-1.23978	0.2151

The model results in Table 6-3 and Table 6-4 showed that the spatial correlation affects the credible level of some variables. Some variables became not credible, e.g., proportion of undivided collector roads in total pedestrian crash models, proportion of population aged 15-24 years old, and proportion of divided collector roads in the severe pedestrian model. The parameter interpretation will be based on the models that showed the best goodness-of-fit, which were the spatial Poisson lognormal models.

Table 6-3: Total Pedestrian-Involved Crash Models for HAYs in Riyadh

Variable Description	Total Pedestrian Crashes											
	PG				PLN				PLN-CAR			
	Mean	S.D.	BCI 2.5%	BCI 97.5%	Mean	S.D.	BCI 2.5%	BCI 97.5%	Mean	S.D.	BCI 2.5%	BCI 97.5%
Intercept	-3.880	0.653	-5.161	-2.600	-3.724	0.609	-4.989	-2.543	-4.351	0.605	-5.506	-3.194
Log of vehicle-kilometers-traveled	0.674	0.062	0.553	0.795	0.658	0.059	0.546	0.777	0.732	0.058	0.620	0.836
Log of population density	0.145	0.029	0.088	0.201	0.154	0.029	0.099	0.214	0.168	0.030	0.108	0.226
Proportion of people whose educational attainment higher than high school	-0.633#	0.371	-1.360	0.094	-0.676#	0.401	-1.473	0.099	-1.147	0.476	-2.102	-0.200
Proportion of Retired people	-4.784#	2.520	-9.723	0.155	-5.447	2.601	-10.63	-0.418	-3.334##	2.562	-8.300	1.744
Commercial Area	2.545	1.248	0.100	4.991	2.679	1.305	0.122	5.258	2.343	1.082	0.215	4.454
Agricultural Area	-1.208#	0.643	-2.468	0.053	-1.176#	0.652	-2.486	0.074	-1.231#	0.670	-2.597	0.039
Number of traffic signals	0.054	0.014	0.026	0.081	0.058	0.014	0.030	0.086	0.060	0.012	0.037	0.084
Proportion of Undivided Collector Roads	0.417#	0.238	-0.050	0.884	0.355##	0.240	-0.117	0.830	0.191	0.253	-0.303	0.691
Dispersion	0.188	0.031	0.137	0.259	-				-			
θ	-				0.451	0.039	0.379	0.531	0.051	0.038	0.014	0.160
$\emptyset$	-				-				0.845	0.082	0.690	1.014
α	-				-				0.944	0.040	0.830	0.984
DIC	1117.092				1019.28				1001.75			

#statistically significant at 90% Bayesian credible interval ##statistically significant at 80% Bayesian credible interval

Table 6-4: Severe Pedestrian-Involved Crash Models for HAYs in Riyadh

Variable Description	**Severe Pedestrian Crashes**											
	PG				**PLN**				**PLN-CAR**			
	Mean	**S.D.**	**BCI**		**Mean**	**S.D.**	**BCI**		**Mean**	**S.D.**	**BCI**	
			2.5%	**97.5%**			**2.5%**	**97.5%**			**2.5%**	**97.5%**
Intercept	-6.738	0.757	-8.221	-5.254	-6.831	0.621	-7.928	-5.486	-5.658	0.803	-7.248	-3.951
Log of vehicle-kilometers-traveled	0.7998	0.071	0.660	0.939	0.801	0.059	0.662	0.896	0.754	0.077	0.592	0.904
Log of population density	0.131	0.032	0.069	0.194	0.135	0.033	0.068	0.196	0.192	0.037	0.124	0.267
Proportion of Households without vehicles	3.232	1.393	0.502	5.963	3.002	1.453	0.106	5.840	2.319##	1.652	-0.967	5.555
Proportion of Illiteracy	1.313#	0.715	-0.089	2.714	1.223##	0.753	-0.291	2.668	1.033##	0.797	-0.569	2.562
Proportion of unemployed people	2.936	1.061	0.857	5.014	2.989	1.083	0.796	5.052	2.406	1.071	0.259	4.464
Proportion of population aged (15-24 years old)	1.557	0.718	0.150	2.964	1.576	0.696	0.249	2.930	0.607	0.730	-0.884	1.935
Commercial Area	3.774	1.360	1.109	6.440	3.752	1.420	0.998	6.536	2.402#	1.323	-0.225	4.946
Educational Area	-4.313	1.228	-6.719	-1.907	-4.268	1.310	-6.878	-1.743	-3.612	1.235	-6.085	-1.241
Number of traffic signals	0.077	0.015	0.049	0.106	0.081	0.015	0.052	0.111	0.077	0.015	0.047	0.105
Proportion of divided Collector Roads	0.638	0.254	0.141	1.136	0.634	0.271	0.107	1.166	0.136	0.300	-0.459	0.714
Dispersion	0.181	0.037	0.121	0.271	-				-			
θ	-				0.453	0.049	0.362	0.553	0.070	0.065	0.014	0.255
$\emptyset$	-				-				0.936	0.116	0.713	1.162
α	-				-				0.931	0.064	0.757	0.986
DIC	938.124				871.69				854.491			

#statistically significant at 90% Bayesian credible interval ##statistically significant at 80% Bayesian credible interval

6.3.1 Parameters Interpretation

6.3.1.1 Effect of Traffic Demand and Socio-economic and Demographic Factors

As expected, the model indicates that total and severe pedestrian crashes increase when the vehicle-kilometers-traveled (vehicular exposure) and population density increase. These results are in line with the studies conducted in developed countries (LaScala et al., 2000; Lee et al., 2014; Loukaitou-Sideris et al., 2007; Cai et al., 2017a, 2016; Noland and Quddus, 2004; Siddiqui et al., 2012; Wier et al., 2009). It is expected that a large part of traffic trips are made by foreign drivers, since they make up a significant part of the population of Riyadh (around one-third of the population). However, it has a high correlation with some variables, e.g.; veh0p (r= 0.53528, p-value= <0.0001) and proportion of retired population (r= -0.52394, p-value= <0.0001). Furthermore, the model that includes the proportion of non-Saudi has a worse goodness-of-fit and has fewer credible variables as compared to the developed model in this study, which may be due to overfitting when having too many variables in a model. A model that has been overfitted may perform poorly when applied to a new sample drawn from the same population (Hadayeghi et al., 2003). In addition, the DIC as a goodness-of-fit measure is penalizing larger variable models (Spiegelhalter et al., 2005).

Road users involved in road collisions are affected by their own characteristics and not only the environment and the behavior of other individuals (Loo and Anderson, 2015). These characteristics are related to their social status, e.g., income and education level, and/or to the characteristics of the places in which they live. In general, people with low socio-economic status or who live in more deprived areas (low educational attainment, unemployed status, etc.) and low-income families are more likely to be pedestrians. The data about the direct measure of income was not available. However, there is information about education level and employment status. There have been many studies that have shown that an education level has a strong positive relationship with an income level (Bailey and Dynarski, 2011; Belley and Lochner, 2007). That is because low-income families might not be able to support their young ones academically. Although income and education are often correlated, education can be used as a proxy for academic abilities, intelligence, knowledge, awareness, etc. Moreover, the proportion of unemployed people can theoretically be a better proxy for income than education. The results showed that people with higher education had a lower possibility to be involved in pedestrian crashes, which implies that households with a high proportion of people that achieved a higher education were less exposed to pedestrian activities. This finding is consistent with the previous study of LaScala et al. (2000).

Furthermore, the proportion of unemployed people was credible at a 95% level and had a positive correlation with severe pedestrian crashes. Hence, in HAYs where unemployed people make up a significant part of the population, a high number of pedestrian-involved crashes might develop.

These findings are consistent with previous studies (e.g., Cottrill and Thakuriah, 2010; LaScala et al., 2004, 2000) that have shown that populations with a low socio-economic status, in a “deprived area,” are more prone to severe pedestrian crash causalities. Several implications of these findings for improved pedestrian safety measures for areas forming a significant part of these population groups. Education is the first priority and the most effective way to build knowledge of pedestrians’ rights and safety. Therefore, continuing safety education/campaigns should be implemented, especially in deprived areas, to improve people’s pedestrian safety awareness.

In the literature, there have been several studies which have examined different age cohorts to explore the variation in zonal pedestrian crash frequency. Therefore, different disaggregated age cohorts were included to investigate age cohorts' impact on pedestrian-involved crashes. The aspatial model uncovered that the likelihood of severe pedestrian crash frequency increased when the proportion of the young population (aged 15-24 years old) increased. However, it was not statistically credible in the spatial Poisson-lognormal model. That might be because the effect of this variable was captured by the spatial effects. In addition, the proportions of the population aged under 14 (young population) and the elderly population (aged 60 and more) were consistently found to be statistically not credible.

The proportion of the young population and older people may have positive, or, conversely, negative effects on pedestrian crashes. For example, older people are fragile, with slower reaction times and poorer vision, which may increase the likelihood of being involved and severely injured in crashes. However, their activities are much lower compared to younger people and/or they are mostly served by their children, so they are less likely to be exposed to pedestrian activities, e.g., shopping in nearby stores. Moreover, children under 14 are fragile, more impulsive, and unpredictable, and have poorer judgment of vehicle speed. In contrast, they may still be under the protection of their parents, e.g., they do not walk alone. Therefore, maybe these two opposite effects canceled out and the variable turned out to be not credible. In addition, that might be due to the fact that the impacts of these variables were captured by another variable, e.g., retired people or the surrogate exposure population density. It is recommended to initiate well-structured safety campaigns for young people or apply strict enforcement in order to reduce this type of crash.

6.3.1.2 Effect of Land-Use

Land-use is considered as another significant category for modeling pedestrian-involved crash occurrence. The results show that commercial areas increase the likelihood of total and severe pedestrian crash frequency. This is also consistent with studies conducted by Loukaitou-Sideris et al. (2007), Wier et al. (2009), and Kim and Yamashita (2002). It is interpreted that more commercial areas in HAYs trigger the rise of pedestrian-involved crash frequency. Therefore, around dense pedestrian activities (e.g., commercial areas), crosswalks should be marked clearly and designed safely.

In addition, there are two more credible land-use variables, as agricultural and educational areas had a negative association with total and severe pedestrian-involved crashes respectively. The effects may be due to low vehicular and pedestrian traffic. The HCDR defines an agricultural area as one that includes all agricultural lands, and related activities, including places of industry and mineral resources. Furthermore, educational areas include not only schools but also areas of universities and colleges for all different educational levels. In a recent study conducted by Altwaijri et al. (2012), they stated that educational areas are negatively correlated with fatal crashes in Riyadh. Unlike general educational areas, Ng et al. (2002) found out that the primary, secondary, and tertiary school places had a credible and negative correlation with pedestrian-related crash frequency. On the other hand, Loukaitou-Sideris et al. (2007; 2011) pointed out that a greater number of schools in a census tract is associated with increased pedestrian crash frequency. In a recent study by Lee et al. (2015a), they found that the school areas of kindergarten (4-6-year-olds) through twelfth grade (7-19-year-olds) is more vulnerable to pedestrian crashes. The classification of land-use is useful for planning and development purposes and may not necessarily be the best category for crash analysis (Kim and Yamashita, 2002). Therefore, it would be useful if further studies are conducted to investigate the effects of schools in pedestrian-involved crash frequency.

6.3.1.3 Effect of Road Network

The aspatial models also indicate that HAYs with more collector road kilometers have more pedestrian crashes. However, it was not statistically credible in the spatial Poisson-lognormal model. That may be because of the impact of the spatial effects. It is worth mentioning that the maximum posted speed on collector roads is 70 kph. In addition, divided collector roads increase the likelihood of developing severe pedestrian crashes. However, it was not statistically credible in the spatial Poisson-lognormal model, as well. Moreover, it confirms the results that are shown in Figure 3-5. Police have to be present more at these locations and there has to be enforcement to prevent pedestrians from crossing the roads at non-allowed places. However, some locations may suffer from lack of crosswalks. Several countermeasures can be implemented to improve pedestrian safety, e.g., pedestrian bridges with escalators seem to be an effective solution to promote pedestrian bridge use (Räsänen et al., 2007). Figure 6-1 and Figure 6-2 show two different situations related to crosswalks. The first photo (Figure 6-1) shows an inappropriate pedestrian access design and no markings. The second photo (Figure 6-2) illustrates that there are no crosswalk pavement markings at an intersection. It is well known that intersections are among the most hazardous locations for pedestrians because of the traffic conflicts with pedestrians crossing at intersections. The total and severe pedestrian crash models indicated that pedestrian crashes increased when the number of traffic signals increased. Therefore, a better engineering design of the signalized intersections that accommodate pedestrian timing and appropriate signage to restrict pedestrian movements can be implemented to reduce these conflicts and enhance pedestrian safety at intersections, e.g., adequate time for pedestrians to cross the street safely using countdown signals and clear crosswalk pavement markings. Regarding enforcement, installing crosswalk cameras to catch drivers who enter an intersection or marked crosswalk on red or fail to stop or yield the right-of-way to pedestrians could be helpful. A clear sign should be installed to order drivers to stop before the pedestrian crosswalk/intersections, or they would be fined.

Figure 6-1: Unmarked Crosswalk on Collector road

Figure 6-2: Unmarked Crosswalk at an Intersection

6.4 Summary and Conclusions

This chapter aims to examine the variations in pedestrian crashes in HAYs in Riyadh using 2,131 pedestrian crashes between 2013 and 2015. The results revealed the factors that contribute to pedestrian crashes. The data shows that driver distraction was the highest reason for pedestrian crashes relevant to drivers, followed by sudden deviation and speeding. Such behavior should be prevented by providing appropriate awareness, safety campaigns, and education to the drivers. In addition, the main reason for pedestrian crashes relevant to the pedestrian was that the pedestrians cross the road at undesignated places. Decision makers should take into consideration implementing pedestrian safety countermeasures to achieve better effectiveness in improving pedestrian safety. Lastly, the proportion of the young population (aged 14-25 years old) was the highest group involved in pedestrian crashes.

Several models were developed using aspatial models (i.e., Poisson-gamma and Poisson-lognormal) and spatial Poisson-lognormal models to examine the association between contributing factors and pedestrian crashes, including total and severe (i.e., fatal and serious injury) pedestrian crash models. It was illustrated that the spatial Poisson-lognormal models were superior compared to aspatial models (Poisson-gamma and Poisson-lognormal). In addition, the traffic volume, socio-economics, land-use, and roadway characteristic factors have an influence on pedestrian crashes.

Although both total and severe pedestrian crash models have common key variables, including VKT, population density, and the number of traffic signals, there were some factors affecting total pedestrian crash occurrences that are different from those affecting severe pedestrian crashes. In the total pedestrian crash model, the proportion of people whose education level is higher than high school was negatively associated with pedestrian crashes.

Meanwhile, in the severe pedestrian crash model, it was found that areas with low socio-economic status (low educational attainment, unemployed status, etc.) are prone to higher severe pedestrian-involved crash frequencies. It illustrates some of the risks related to the areas most affected by these crashes, which need to be targeted with better-designed educational and awareness programs to improve pedestrian safety.

In terms of land-use influence, the HAYs with a greater fraction of commercial areas have a greater likelihood of total and severe pedestrian-involved crashes, whereas a significantly higher fraction of agricultural and educational areas reduces it.

Overall, this study provides a better understanding of pedestrian safety effects at the macro-level and its contributing factors in Saudi Arabia. Decision makers can implement appropriate education and effective campaigns, benefit from the development of engineering countermeasures, traffic control, and management policy, as well as apply law enforcement to enhance pedestrian safety. There should be complementary improvements between all related agencies and departments in Riyadh city to integrate walking in daily life to foster the environment of a friendly transportation mode that satisfies people's safety and convenience. We believe that this study is the first study to explore pedestrian crashes at the macro-level using data from Saudi Arabia. Obtaining the most complete and recent crash data from Riyadh, Saudi Arabia, is considered a major contribution of this paper, where such data is rare in developing countries.

CHAPTER 7: A BAYESIAN MULTIVARIATE SPATIAL MODEL FOR PREDICTING CRASH COUNTS BY SEVERITY AND TYPE AT THE MACRO-LEVEL

7.1 Results and discussion

Because each HAY has its own parameters, the random parameters models were employed to allow each estimated parameter to vary spatially (Lord and Mannering, 2010). However, they did not improve the fit based on the goodness-of-fit measures. In addition, one advantage of the Bayesian perspective is that previous knowledge about the parameters can be incorporated into the analysis (Yu and Abdel-Aty, 2013). Therefore, different approaches, including two-stage updating, previous studies, moment, and maximum likelihood estimation to develop informative priors for the parameters have been investigated. Yet, the models' results showed no improvement compared to the models with non-informative priors.

Two chains with 50,000 iterations, 5,000 of which are discarded as a burn-in sample, were set up for each model using WinBUGS 1.4.3 software. The developed model convergence and performance are thoroughly assured based on the Gelman–Rubin convergence statistic set, and the ratios of the Monte Carlo errors relative to the standard deviations of the estimates are less than 0.05. A total of eight univariate and multivariate Poisson-lognormal models with and without the spatial random effects in a Bayesian framework were estimated for crash counts by crash severity (i.e., fatal, injury, and PDO) and by crash type (i.e., pedestrian, bicycle, single- and multi-vehicle) using the methodology in section (4.3.2). Table 7-1 summarizes the models' performances. The results show that the multivariate models (MVPLN & MVPLN-CAR) performed much better than the corresponding univariate models (PLN & PLN-CAR) in both crash type and crash severity models. Furthermore, the univariate and multivariate spatial models outperformed the corresponding aspatial models in both models. This implies that incorporating spatial effects improves model performance, and it is consistent with the findings of Osama and Sayed (2017). In terms of crash type model, in comparing the two multivariate models, the MVPLN-CAR model provides a significantly lower value of the posterior mean of the deviance and a slightly lower value of the DIC. This indicates that the multivariate heterogeneity random effects' residuals are partially spatially correlated, which justifies the use of a multivariate spatial model (Huang et al., 2017). Moran's I test was conducted to examine the existence of spatial autocorrelation among aspatial models' residuals (heterogeneity random effects), and the results are discussed in the following section. These findings are consistent with previous studies that demonstrated that accounting for correlation among different crash levels significantly improves the crash model performance (Aguero-Valverde, 2013; Aguero-Valverde and Jovanis, 2009; Huang et al., 2017; Lee et al., 2015b; Osama and Sayed, 2017).

Table 7-1: Summary of the models' performances

	Crash severity model			**Crash type model**		
Model	$\bar{D}$	pD	DIC	$\bar{D}$	pD	DIC
PLN	3511.76	399.432	3911.19	4572.18	560.938	5133.12
PLN-CAR	3502.83	384.798	3887.62	4549.42	525.918	5075.34
MVPLN	3502.60	343.760	3846.36	4568.49	500.335	5068.82
MVPLN-CAR	3487.17	345.240	3832.41	4563.27	504.098	5067.37

In order to control for spatial autocorrelation in the crash models' residuals (heterogeneity random effects), the Global Moran's I statistic was calculated. Table 7-2 presents the Global

Moran's I statistic calculated from the crash models' residuals and the corresponding Z-score and P-value. It revealed that the spatial correlation could be controlled by spatial models for both crash type and crash severity models. The results showed statistically significant spatial autocorrelations at a 99.9% confidence level in univariate and multivariate aspatial models' residuals both for crash type and crash severity models. Therefore, the spatial autocorrelation should be accounted for in the estimation of the models. With respect to the fatal crash model, there seems to be no significant spatial autocorrelation among the PLN model's residuals, and this is in line with the previous study by Aguero-Valverde and Jovanis (2006). However, the residuals of the counterpart model (MVPLN) were revealed to be spatially correlated, which demonstrates the existence of a spatial autocorrelation among the fatal crash model's residuals. Finally, there appears to be no significant spatial autocorrelation among MVPLN-CAR models' residuals for both models, which indicates that applying multivariate spatial models at the macro-level is appropriate, because it can account for spatial autocorrelations. Therefore, the discussion of the results will be based on the outputs of the MVPLN-CAR models.

Table 7-2: Moran's I statistic of crash models' residuals

Model	Crash Severity	Index	Z-Score	P-value	Crash Type	Index	Z-Score	P-value
PLN	**Fatal**	-0.019	-0.306	0.759	**Pedestrian**	0.261	5.735	0.000
	Injury	0.234	5.1640	0.000	**Bicycle**	0.237	5.237	0.000
	PDO	0.196	4.363	0.000	**Single-vehicle**	0.432	9.417	0.000
					Multi-vehicle	0.251	5.553	0.000
PLN-CAR	**Fatal**	-0.060	-1.163	0.244	**Pedestrian**	-0.113	-2.312	0.021
	Injury	-0.086	-1.730	0.083	**Bicycle**	-0.088	-1.792	0.073
	PDO	-0.132	-2.742	0.006	**Single-vehicle**	-0.024	-0.403	0.687
					Multi-vehicle	-0.125	-2.595	0.0095
MVPLN	**Fatal**	0.1287	2.899	0.003	**Pedestrian**	0.263	5.787	0.000
	Injury	0.211	4.666	0.000	**Bicycle**	0.296	6.504	0.000
	PDO	0.197	4.385	0.000	**Single-vehicle**	0.411	8.973	0.000
					Multi-vehicles	0.250	5.537	0.000
MVPLN-CAR	**Fatal**	-0.075	-1.497	0.134	**Pedestrian**	-0.033	-0.595	0.5514
	Injury	-0.044	-0.830	0.406	**Bicycle**	-0.007	-0.039	0.9686
	PDO	-0.064	-1.253	0.210	**Single-vehicle**	-0.048	0.9194	0.358
					Multi-vehicle	-0.061	-1.208	0.2270

Table 7-3 and Table 7-4 present the posterior variance–covariance and correlation matrices of error terms for crash severity and crash type models respectively. The whole covariance matrix of heterogeneity effects was insignificant for the crash severity model. The correlations among crash severities was insignificant as well. The same trend also appeared for the four types of crashes, except for single-vehicle and multi-vehicle crashes and for pedestrian and single-vehicle or multi-vehicle crashes, which were significant at 80% BCI. This might be because a substantial portion of the heterogeneity random errors was captured by the spatial random errors. In addition, the model contains two random effects and the majority of the variation of these effects is explained by the spatial component (discussed in the following section).

The spatial covariance was significant at 95% BCI for all response variables. In addition, the posterior correlations between the spatial errors for all response variables were significant at 95% BCI. Although the posterior correlation among all response variables for heterogeneity's residuals were insignificant, that may indicate that any residual in the data is spatially associated; the underlying covariates that explain the lack of fit in the model are spatially correlated. Therefore, the search for unobserved/missing covariates can be restricted to covariates that vary spatially. To investigate more, two more models were estimated to uncover the impact of the heterogeneity and spatial components. The heterogeneity term was removed in the first model. In the second model, the correlations among all response variables for heterogeneity component was set to zero. The results revealed no improvement of the fit in terms of the goodness-of-fit (DIC). Since the spatial effects of risk dominate, the correlations between the total random effect (heterogeneity and spatial effects) for all response variables were significant at 95% and very high (>0.8), except for the multi-vehicle and single-vehicle, which was high (0.5043) and was less (0.315 and 0.321 respectively) for the single-vehicle and pedestrian or bicycle, which indicates a strong shared geographical pattern of risk between the severities/types of crashes. In general, the highest correlation is between pedestrian and bicycle, and it is consistent with previous studies (Huang et al., 2017; Lee et al., 2015b).

These findings are inconsistent with the findings of Aguero-Valverde (2013), who found that the correlations among the three levels of crash severity (i.e., fatal, injury, and PDO) were significant for the heterogeneity effects, while insignificant for the spatial effects. In addition, the finding which revealed that the correlations between neighboring locations are more significant than among crash modes is consistent with the studies at the macro-level (Lee et al., 2015b; Osama and Sayed, 2017) and is in contrast with Huang et al.'s (2017) micro-level study.

Table 7-3: Posterior of variance–covariance/correlation matrices for crash severity model

	Fatal crash	Injury crash	PDO crash
	Unstructured (heterogeneity) errors		
Fatal crash	**0.0552(0.0327)**	0.0181(0.0189)	0.0075(0.0181)
Injury crash	0.4077(0.3028)	**0.0254(0.0138)**	0.0102(0.0119)
PDO crash	0.1204(0.3660)	0.3017(0.2991)	**0.0331(0.0163)**
	Structured (spatial) errors		
Fatal crash	**0.6740(0.2134)**	**0.6817(0.1599)**	**0.7163(0.1649)**
Injury crash	**0.9201(0.0451)**	**0.8270(0.1524)**	**0.7894(0.1350)**
PDO crash	**0.8711(0.0768)**	**0.8601(0.0473)**	**1.0210(0.1569)**
	Total random effects		
Injury crash	**0.8911(0.0446)**	-	**-**
PDO crash	**0.8284(0.0698)**	**0.8439(0.0432)**	**-**

Note: Shading cells indicate correlation matrix. Bold indicates significant coefficients at 95% BCI. Standard deviation in parentheses.

Table 7-4: Posterior of variance–covariance/correlation matrices for crash type model

	Pedestrian	Bicycle	Single-vehicle	Multi-vehicle
	Unstructured (heterogeneity) errors			
Pedestrian crash	**0.0504(0.0256)**	-0.0125(0.0191)	$0.0214^{\#\#}$(0.0198)	$0.0227^{\#\#}$(0.0193)
Bicycle crash	-0.2684(0.3377)	**0.0596(0.0373)**	-0.0065(0.0194)	-0.0038(0.0173)
Single-vehicle crash	$0.4375^{\#\#}$(0.3013)	-0.1559(0.3686)	**0.0380(0.0222)**	$0.0184^{\#\#}$(0.0168)
Multi-vehicle crash	$0.4945^{\#\#}$(0.2753)	-0.1184(0.3632)	$0.4613^{\#\#}$(0.2765)	**0.0326(0.0180)**
	Structured (spatial) errors			
Pedestrian crash	**0.7923(0.1874)**	**1.0340(0.2131)**	**0.3628(0.1641)**	**0.7833(0.1639)**
Bicycle crash	**0.9592(0.0319)**	**1.4990(0.3978)**	**0.5360(0.2215)**	**1.0770(0.2243)**
Single-vehicle crash	**0.3082(0.1153)**	**0.3355(0.1129)**	**1.6770(0.2471)**	**0.6966(0.1670)**
Multi-vehicle crash	**0.8303(0.0569)**	**0.8345(0.0691)**	**0.5043(0.0778)**	**1.1230(0.1708)**
	Total random effects			
Bicycle crash	**0.8972(0.0386)**	-	-	**-**
Single-vehicle crash	**0.3150(0.1039)**	**0.3210(0.1073)**	-	**-**
Multi-vehicle crash	**0.8163(0.0529)**	**0.8023(0.0655)**	**0.5053(0.0730)**	**-**

Note: Shading cells indicate correlation matrix. Bold indicates significant coefficients at 95% BCI. ##statistically significant at 80% BCI. Standard deviation in parentheses.

The model results of the multivariate CAR models for crash severity and crash type are presented in Table 7-5 and Table 7-6 respectively.

The posterior means for the standard deviations of heterogeneity effects for all response variables ranged from 0.154 to 0.234. However, the posterior means for the standard deviations of spatial effects were higher, ranging from 0.81 to 1.292. All of them were statistically significant at 95% BCI. Although the correlations of the heterogeneity random effect were insignificant in the MVPLN-CAR models, the proportions of variability in the random effects due to spatial effects were significant. The variations explained by the spatial random effects for fatality, injury, and PDO were 0.780, 0.854, and 0.850 respectively. The spatial correlation is high in the injury and PDO models compared to the fatality model. This finding is in line with previous studies (Aguero-Valverde and Jovanis, 2006; Quddus, 2008). Similarly, the variation explained by the spatial random effects for all models of crash types was high (>0.8). This indicates that the spatial random effect is dominating the heterogeneity random effect.

Regarding the parameter interpretation of the models, the results showed that the number of significant variables is exactly the same in both the crash severity model and the crash type model. However, signs and magnitudes of coefficients vary. There are several common significant variables for the response variables. For example, 'log of vehicle-kilometers-traveled', which was used as exposure variables, and number of traffic signals were commonly significant and positively associated with all response variables. These findings are consistent with previous studies, e.g., Aguero-Valverde (2013) and Lee et al. (2015b). In roadway networks, intersections are among the most dangerous locations for vehicles and road users due to the traffic conflicts between vehicles and with pedestrians crossing at intersections. An implication of this result is, to reduce these crashes, to enhance traffic signal timing and pedestrian safety at signalized intersections; for example, encouraging pedestrians to cross from crosswalk pavement markings and setting adequate time for pedestrians to cross the street safely using countdown signals.

In terms of land-use variables, a residential area was significant for fatal, PDO, pedestrian, and multi-vehicle crashes. Although it had a positive relationship with PDO, pedestrian, and multi-vehicle crashes, it is negatively associated with fatal crashes. This is consistent with the findings of previous studies, e.g., Ng et al. (2002) and Noland and Quddus (2005, 2004), in which they asserted that higher dense urban areas are associated with reductions in casualties. This effect may be because of the relatively low speed compared to other land-use types. In addition, the commercial area is also commonly significant for PDO, pedestrian, and multi-vehicle crashes and positively related to them. These findings are consistent with previous studies that found that the land area zoned for residential and commercial uses were found to significantly and positively affect pedestrian crashes, including Kim et al. (2006), Loukaitou-Sideris et al. (2007), and Wier et al. (2009). In contrast, the governmental area was significant and negatively associated with fatal, injury, bicycle, and single-vehicle crashes. Furthermore, the agricultural area was significantly and negatively associated with fatal, injury, PDO, pedestrian, bicycle, and multi-vehicle crashes. The effects may be due to low vehicular and pedestrian traffic. Moreover, there are many police officers patrolling around government buildings, which makes drivers more cautious and likely to observe traffic regulations. An implication of these findings is to target the affected areas (i.e., residential and commercial) with appropriate educational and awareness campaigns and programs, e.g. about pedestrian safety and security. Furthermore, applying law enforcement in commercial areas can reduce fatal crashes and other types of crashes. Lastly, revisiting the engineering design of road networks or urban planning in these areas could help, because it can play a vital role in preventing these types of crashes related to each HAY and to improve traffic safety in these areas. For example, installing traffic calming, which is the combination of physical controls and community support, can alleviate traffic hazards in these areas, which can contribute to reducing these types of crashes (Huang and Cynecki, 2000).

Different age groups were aggregated to investigate the effect of age on crash frequency for all severities and types. However, only the age groups 15-24 years old and 65 and older were significant at an 80% BCI and positively related to injury and single-vehicle crashes.

It was revealed that the proportion of collector roads was significant and positively associated with PDO, pedestrian, bicycle, and multi-vehicle crashes. Urban collector roads are characterized be their complexity in terms of driving environments, often encompassing both mobility and roadside activity accessibility. Consequently, motorized traffic and the most vulnerable road users, e.g., pedestrians, are competing continually (Domenichini et al., 2018). This finding is in line with the findings of Lovegrove and Sayed (2006), who uncovered that the association between zones containing increasing amounts of collector roads and PDO crashes is positive. This highlights the necessity of assuring road users' safety and security, especially for pedestrians and cyclists, as they are more vulnerable to traffic crashes not only at traffic signals, but also along collector roads, by implementing appropriate countermeasures. Instances of such countermeasures include dedicating exclusive lanes for pedestrians and cyclists and designing pedestrian bridges and pedestrian crossing markings, etc. In addition, enforcement can be installed along this type of road in order to reduce crashes occurrence on collector roads.

The proportion of non-Saudi population was significant and positively associated with PDO and multi-vehicle crashes. This is expected, since they make up a significant part of the population of Riyadh (around thirty percent of the population) and because several of these individuals work as private drivers or in companies, which consequently directly involves them in a large part of traffic trips every day. An implication of this result is that establishing appropriate education for foreigners, e.g., using different languages, could contribute to limiting this type of crash. Some of them are workers with no driver's license, so it is imperative to emphasize the need to apply for one. Also, since most of these workers have access to company vehicles, the company may be compelled to conduct competency assessments with a view to substantiating workers' compliance with traffic safety regulations. New arrivals need to be trained before being involved in the road environment. In contrast, the proportion of people whose educational attainment is higher than high school was significant and negatively associated with fatal and pedestrian crashes.

Table 7-5: Summary of the coefficient estimates for the crash severities models

Variable	*Fatal crash*				*Injury crash*				*PDO crash*			
	Mean	**S.D.**	**BCI**		**Mean**	**S.D.**	**BCI**		**Mean**	**S.D.**	**BCI**	
			2.5%	**97.5%**			**2.5%**	**97.5%**			**2.5%**	**97.5%**
Intercept	**-6.931**	**0.847**	**-8.586**	**-5.211**	**-5.988**	**0.590**	**-7.133**	**-4.890**	**-3.281**	**0.315**	**-3.930**	**-2.662**
Log of Vehicle- Kilometers -Traveled	**0.847**	**0.072**	**0.706**	**0.989**	**0.818**	**0.053**	**0.719**	**0.916**	**0.880**	**0.029**	**0.830**	**0.937**
Proportion of Collector Roads	0.089	0.287	-0.465	0.650	0.336##	0.234	-0.105	0.805	**0.445**	**0.228**	**0.036**	**0.924**
Proportion of Non-Saudi	-0.163	0.231	-0.611	0.290	0.129	0.184	-0.235	0.492	**0.392**	**0.162**	**0.069**	**0.703**
Proportion of young people (15–24 years old)	0.495	0.632	-0.732	1.731	0.6198##	0.476	-0.316	1.566	0.480	0.411	-0.322	1.298
Proportion of elderly people (65 years old or older)	3.525	3.012	-2.485	9.350	3.305##	2.319	-1.192	7.911	0.989	2.211	-3.265	5.382
Proportion of people whose educational attainment higher than high school	**-2.346**	**0.608**	**-3.530**	**-1.132**	-0.272	0.434	-1.109	0.595	0.014	0.389	-0.696	0.789
Residential area	**-0.863**	**0.430**	**-1.689**	**-0.002**	-0.086	0.338	-0.751	0.582	**0.906**	**0.329**	**0.267**	**1.569**
Commercial area	0.576	1.542	-2.474	3.563	0.716	1.148	-1.575	2.955	**2.413**	**1.124**	**0.088**	**4.596**
Governmental area	**-2.251**	**1.018**	**-4.267**	**-0.265**	**-1.669**	**0.788**	**-3.213**	**-0.113**	0.626	0.737	-0.826	2.059
Agricultural area	**-2.478**	**0.938**	**-4.387**	**-0.696**	**-1.350**	**0.568**	**-2.475**	**-0.258**	**-1.527**	**0.438**	**-2.385**	**-0.678**
Number of traffic signals	**0.038**	**0.014**	**0.010**	**0.066**	**0.067**	**0.012**	**0.043**	**0.090**	**0.045**	**0.012**	**0.023**	**0.068**
Heterogeneity random effect	**0.226**	**0.066**	**0.118**	**0.370**	**0.154**	**0.039**	**0.093**	**0.246**	**0.177**	**0.043**	**0.106**	**0.272**
Spatial random effect	**0.810**	**0.132**	**0.541**	**1.065**	**0.906**	**0.084**	**0.744**	**1.073**	**1.008**	**0.078**	**0.854**	**1.161**
Proportion of variation explained by the spatial effect	**0.780**	**0.070**	**0.617**	**0.888**	**0.854**	**0.038**	**0.766**	**0.912**	**0.850**	**0.037**	**0.769**	**0.910**

#statistically significant at 90% BCI ##statistically significant at 80% BCI
Bold indicates significant coefficients at 95% BCI.

Table 7-6: Summary of the coefficient estimates for the crash types models

Variable	*Pedestrian crash*				*Bicycle crash*				*Single-vehicle crash*				*Multi-vehicle crash*			
	Mean	**S.D.**	**BCI 2.5%**	**BCI 97.5%**	**Mean**	**S.D.**	**BCI 2.5%**	**BCI 97.5%**	**Mean**	**S.D.**	**BCI 2.5%**	**BCI 97.5%**	**Mean**	**S.D.**	**BCI 2.5%**	**BCI 97.5%**
Intercept	**-6.069**	**0.838**	**-7.780**	**-4.540**	**-8.579**	**1.305**	**-11.190**	**-6.067**	**-7.272**	**0.565**	**-8.544**	**-6.123**	**-3.958**	**0.621**	**-5.140**	**-2.841**
Log of Vehicle-Kilometers -Traveled	**0.741**	**0.074**	**0.608**	**0.891**	**0.859**	**0.112**	**0.644**	**1.090**	**1.076**	**0.058**	**0.957**	**1.210**	**0.923**	**0.053**	**0.805**	**1.028**
Proportion of Collector Roads	**0.594**	**0.248**	**0.094**	**1.072**	**0.942**	**0.393**	**0.174**	**1.713**	0.319	0.265	-0.182	0.866	**0.510**	**0.214**	**0.059**	**0.916**
Proportion of Non-Saudi	0.3896$^{\#}$	0.211	-0.017	0.817	0.4435$^{\#\#}$	0.325	-0.177	1.091	-0.104	0.231	-0.526	0.383	**0.416**	**0.187**	**0.091**	**0.837**
Proportion of young people (15–24 years old)	-0.080	0.622	-1.312	1.121	-0.068	1.046	-2.163	1.919	0.8147$^{\#\#}$	0.579	-0.327	1.912	0.617	0.466	-0.219	1.629
Proportion of elderly people (65 years old or older)	-0.427	2.736	-5.765	4.938	-0.303	4.604	-9.389	8.676	4.121$^{\#\#}$	2.699	-1.379	9.157	-0.388	2.204	-4.682	4.060
Proportion of people whose educational attainment higher than high school	**-1.305**	**0.510**	**-2.296**	**-0.279**	-0.193	0.756	-1.679	1.293	0.562	0.551	-0.513	1.612	0.002	0.467	-0.839	0.979
Residential area	**0.930**	**0.386**	**0.150**	**1.671**	-0.418	0.581	-1.539	0.728	0.266	0.430	-0.626	1.063	**0.969**	**0.367**	**0.207**	**1.640**
Commercial area	**3.382**	**1.227**	**0.963**	**5.768**	1.651	1.940	-2.230	5.360	1.388	1.424	-1.419	4.167	**2.499**	**1.151**	**0.178**	**4.748**
Governmental area	0.265	0.872	-1.425	1.988	**-3.453**	**1.492**	**-6.414**	**-0.543**	**-1.386**	**0.950**	**-3.232**	**0.476**	0.701	0.745	-0.701	2.204
Agricultural area	**-1.568**	**0.688**	**-2.944**	**-0.253**	**-4.099**	**1.485**	**-7.161**	**-1.371**	-0.729	0.647	-2.016	0.527	**-1.534**	**0.454**	**-2.428**	**-0.631**
Number of traffic signals	**0.074**	**0.014**	**0.046**	**0.103**	**0.048**	**0.021**	**0.006**	**0.089**	**0.036**	**0.016**	**0.004**	**0.068**	**0.046**	**0.014**	**0.019**	**0.075**
Heterogeneity random effect	**0.218**	**0.055**	**0.122**	**0.336**	**0.234**	**0.071**	**0.120**	**0.392**	**0.187**	**0.053**	**0.103**	**0.308**	**0.174**	**0.047**	**0.101**	**0.279**
Spatial random effect	**0.884**	**0.105**	**0.679**	**1.092**	**1.214**	**0.161**	**0.916**	**1.547**	**1.292**	**0.095**	**1.111**	**1.485**	**1.057**	**0.081**	**0.896**	**1.214**
Proportion of variation explained by the spatial effect	**0.802**	**0.052**	**0.688**	**0.888**	**0.838**	**0.048**	**0.731**	**0.915**	**0.873**	**0.035**	**0.794**	**0.929**	**0.858**	**0.038**	**0.771**	**0.917**

#statistically significant at 90% BCI ##statistically significant at 80% BCI
Bold indicates significant coefficients at 95% BCI.

7.2 Conclusion

This chapter used a multivariate Poisson-lognormal CAR (MVPLN-CAR) model in a Bayesian framework to simultaneously model correlated crash counts, which accounts for spatial autocorrelation and correlations among different crash counts by severity and type levels at the HAY level in Riyadh, Saudi Arabia. The MVPLN-CAR model was compared to the corresponding aspatial model.

The results reveal that the multivariate models (MVPLN & MVPLN-CAR) outperformed the corresponding univariate models (PLN & PLN-CAR) in both crash type count and crash severity count models in terms of goodness-of-fit measures. Furthermore, the univariate and multivariate models incorporating both unstructured (heterogeneity) and structured (spatially correlated) effects performed better than those not incorporating the spatial effect counterparts in both models. This indicates that adopting the multivariate models by incorporating both the unstructured (heterogeneity) random error term and spatially structured conditional autoregressive term at the zonal level significantly improves the crash model performance.

There are significant correlations between the total random effect (heterogeneity and spatial effects) for the three severity levels and the four type levels of crash counts, indicating a strong shared geographical pattern of risk between each of these levels. In terms of crash type model, the correlation is very strong between pedestrian and bicycle crashes, but it is relatively less strong between single-vehicle and pedestrian or bicycle crashes. In addition, the traffic volume, road characteristics, socio-economics and demographics, and land-use factors have a significant influence on different crash severities and types. However, the variable estimates and statistical significance vary across different models.

The findings of this study can help and guide decision makers and practitioners in selecting more appropriate safety countermeasures for enhancing traffic safety policies in Riyadh, which can contribute to achieving the goal of the KSA National Transformation Program 2020 relative to reducing traffic fatalities.

CHAPTER 8: RESEARCH IMPLICATIONS

8.1 Introduction

The findings in Chapter 7 may provide a useful implication for decision makers and practitioners using the crash hot zone screening results. The crash hot zone screening results show the overall crash distributions within the study area. Identification of sites with promise is a vital task to improving highway safety by finding spatial units where the crash risk is high (Hauer, 1996). Therefore, the outputs of the best multiple models in Chapter 7 were used for hotspot identification. Based on that, the potential for safety improvements (PSIs) was calculated for hot zone identification using the models.

8.2 Screening of the hot zones

In order to rank hot zones, a Potential for Safety Improvement (PSI) was employed in this study as a measure of how many crashes can be effectively reduced. It can be calculated by the difference between the expected and predicted crash counts (Aguero-Valverde and Jovanis, 2010; Hauer et al., 2002; Lee et al., 2015a). PSI is defined as:

$$PSI = N_{\text{expected}} - N_{\text{predicted}} \qquad (35)$$

$$PSI = \exp(\beta_{0k} + \sum_{j=1}^{J} \beta_{kj} X_{ij} + \theta_{ik} + \emptyset_{ik}) - \exp(\beta_{0k} + \sum_{j=1}^{J} \beta_{kj} X_{ij}) \qquad (36)$$

$$PSI = \exp(\beta_{0k} + \sum_{j=1}^{J} \beta_{kj} X_{ij}) * (\exp(\theta_{ik} + \emptyset_{ik}) - 1) \qquad (37)$$

All zones in the study area were classified into three categories based on their PSI values: hot 'H', warm 'W', and cold 'C' zones (Lee et al., 2015a).

$$\text{where, ranking of hot zones categories} \begin{cases} \text{H}, & \text{Top 10\% PSI} \\ \text{W}, & \text{Between PSI} = 0 \text{ and Top 10\% PSI} \\ \text{C}, & \text{PSI} < 0 \end{cases}$$

Therefore, 'H' has much more crashes compared to other zones with similar characteristics. Zones that are categorized as warm zones also have some room for crash reduction; however, the issue of safety is not as risky as in 'H' zones. With respect to cold zones, they have fewer crashes compared to other similar zones.

8.3 Results and Discussion

In order to identify hot zones, the PSI performance measure was calculated and mapped. Figure 8-1 to Figure 8-4 display the macroscopic screening results using PSI calculated based on MVPLN-CAR models. The first two figures illustrate maps of the ranking HAYs for crash severity models, while the other two illustrate maps of the ranking HAYs for crash type models. The map distribution is based on the three categories of the PSI values (i.e., hot, warm, and cold), where the hot is the top 10% PSI, the warm is between PSI=0 and less than the top 10% PSI, and the cold is the PSI less than zero (Lee et al., 2015a).

It was found that the distributions of hot zones for both models (i.e., crash severity model and crash type model) can be compared. This could help to explain which severity/type(s) of crash/es affected each HAY the most. For example, the spatial distribution of PDO crashes is mostly similar to the multi-vehicle crash distributions. In other word, the majority of PDO crashes are multi-vehicle-related crashes. The results also show common similar categories of screening results for each HAY of all crash severity/type level models. For example, in the south (the HAY indicated by the arrow in the bottom of the figure as "commercial area" as shown in Figure 3-2 is categorized as hot for all response variables. This may be as a result of its very dense population and commercial activities. It can be seen also that the middle (the HAYs located in the CBD area) shows the highest PSI for all response variables except for single vehicle crashes. These HAYs are being considered as the HAYs where most of the population and commercial activities are fiercely concentrated. In contrast, some HAYs are indicated by common similar categories of screening results for some response variables, i.e., non-motorized and/or motorized related crashes. For example, in the north (the HAY near the airport where most of the attractions (recreational) are concentrated), is being classified as hot for fatal crashes and it seems to be affected by non-motorized (pedestrian and/or bicycle) related crashes. Furthermore, in the northeast, the HAY where it is considered as one of the most-visited for social activities is being categorized as hot for fatal and injury crashes. This HAY is also indicated mostly by pedestrian and/or single-vehicle related crashes, which may be due to its high concentration of social activities. However, some HAYs seem to be affected by only one type of crash

This illustrates some of the risks related to the HAYs most affected by these crashes, which means there is a need to target the affected HAYs with better-designed educational and awareness programs, law enforcement, and engineering designs related to each category to improve traffic safety, as discussed in the earlier section.

Fatal crashes **Injury crashes**

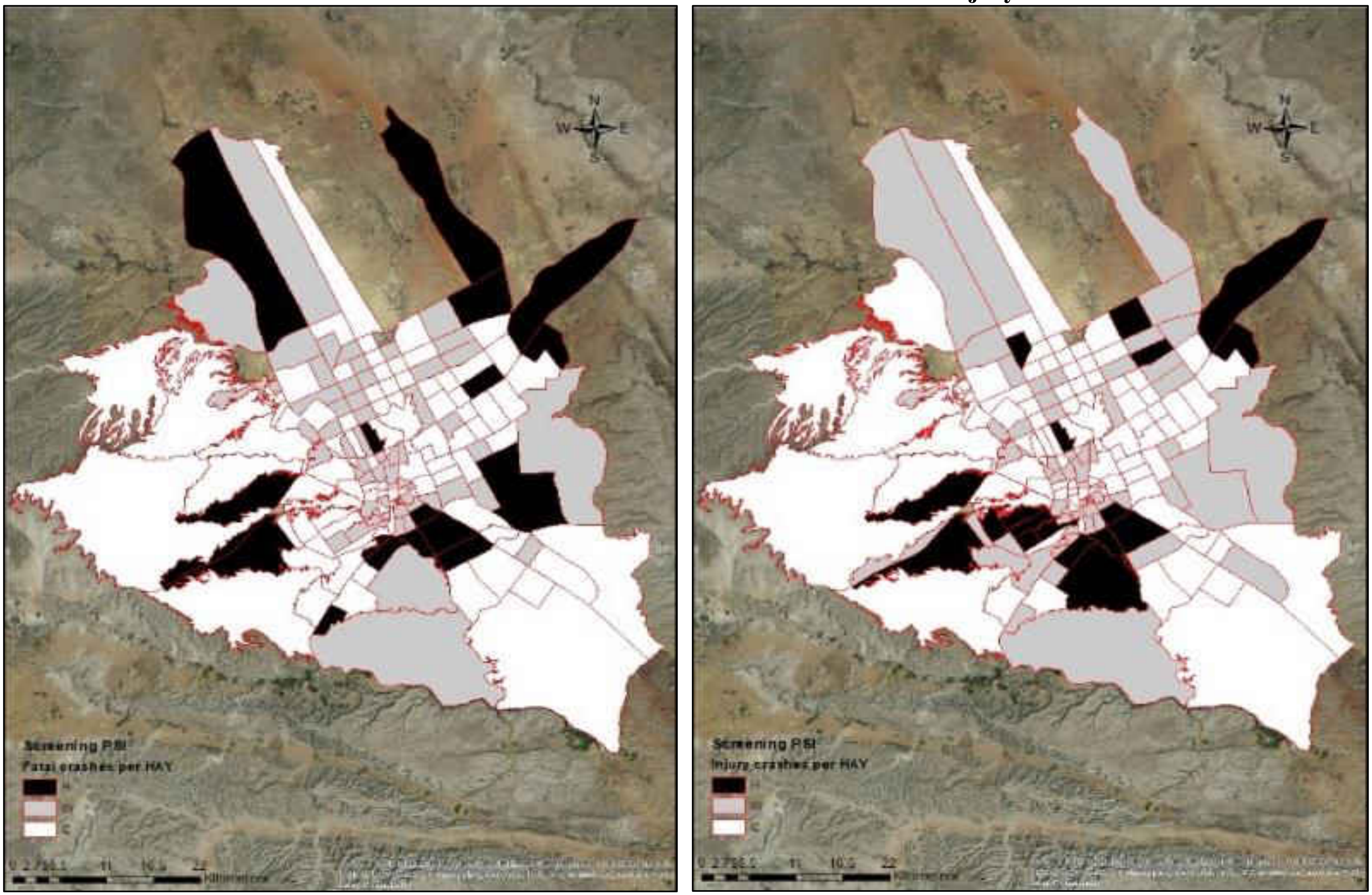


Figure 8-1: Results of screening based on PSI of fatal and injury crashes per HAY

PDO crashes

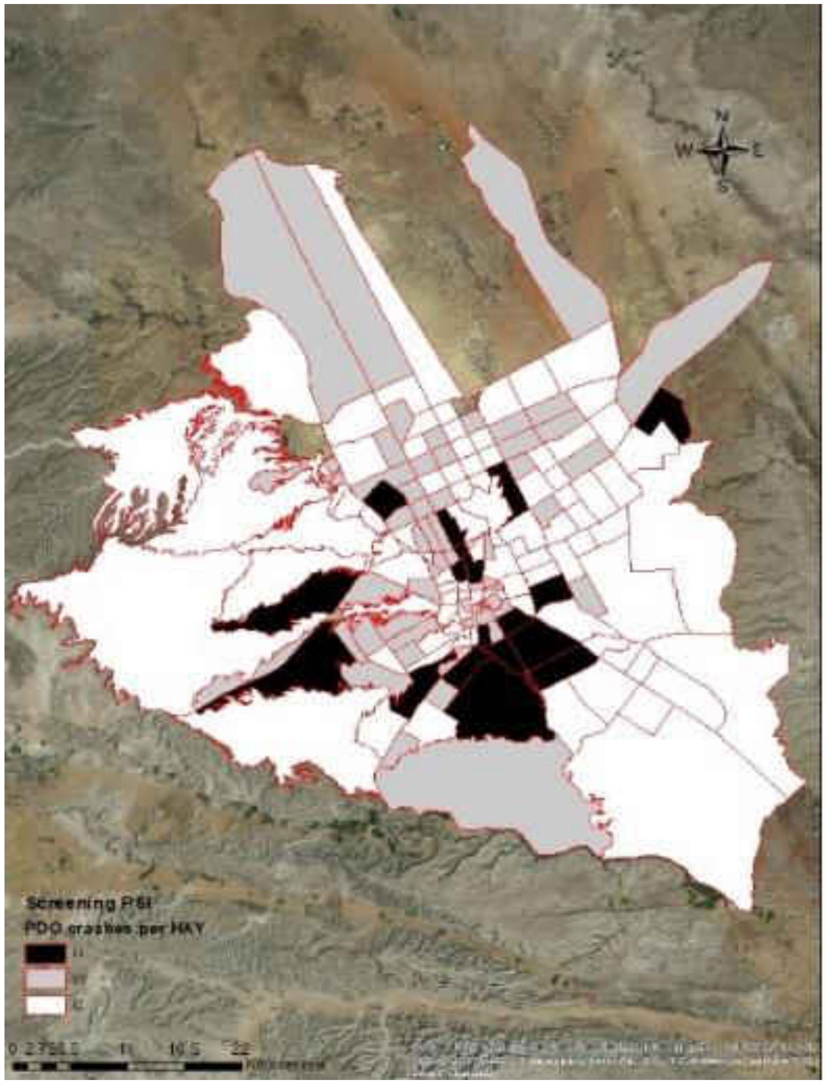


Figure 8-2: Results of screening based on PSI of PDO crashes per HAY

Pedestrian crashes **Bicycle crashes**

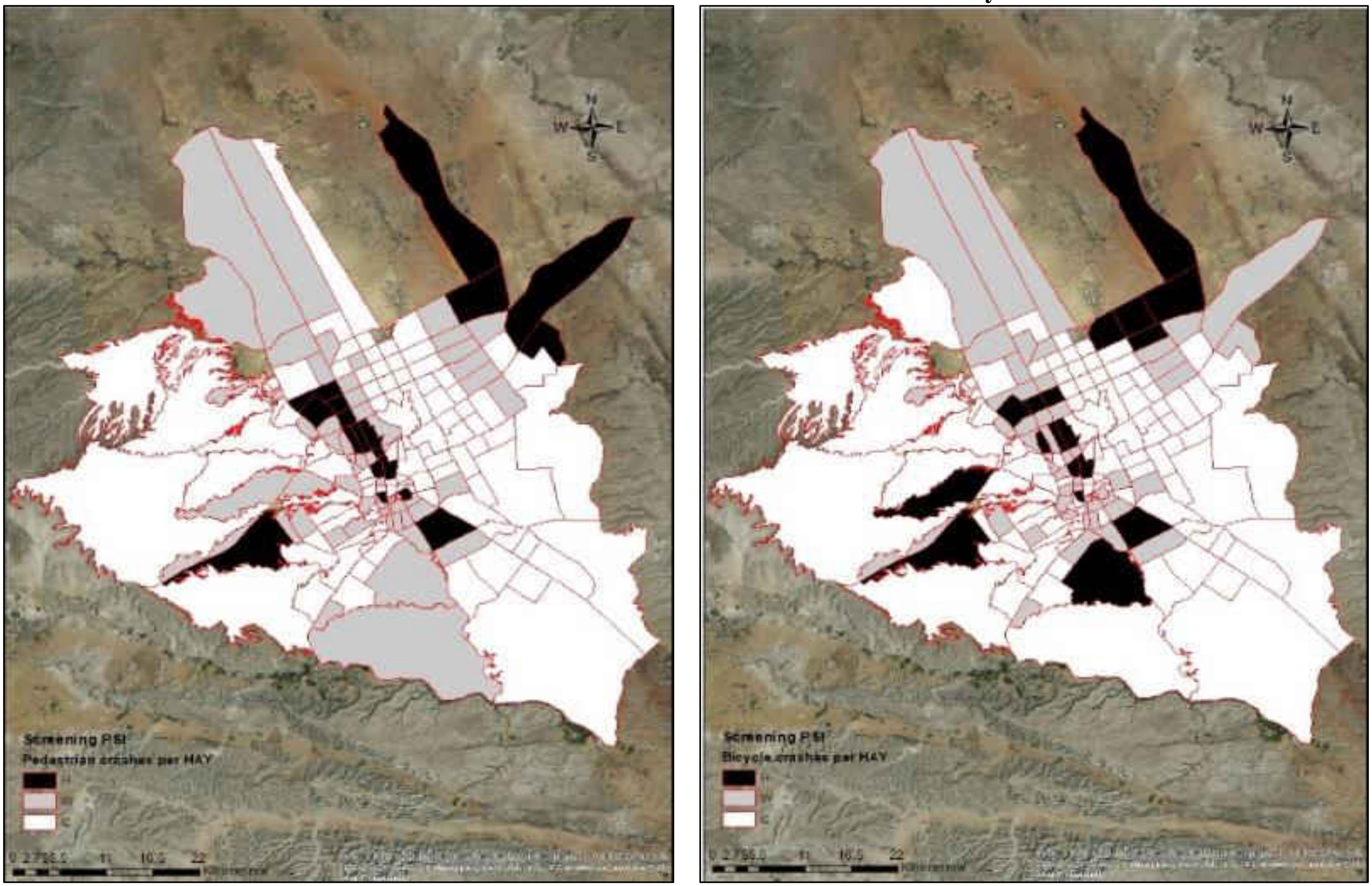


Figure 8-3: Results of screening based on PSI of pedestrian and bicycle crashes per HAY

Single vehicle crashes **Multiple vehicles crashes**

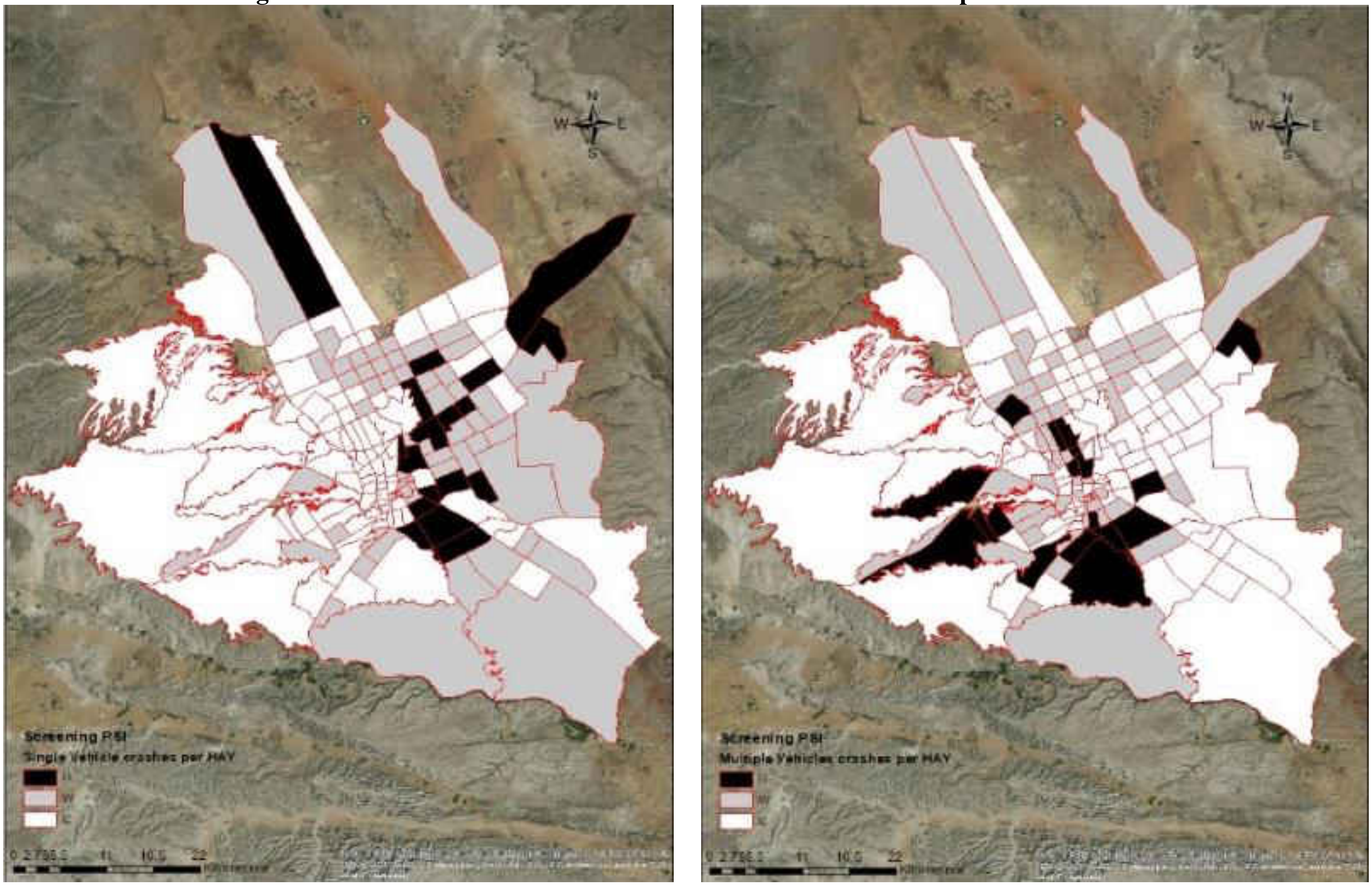


Figure 8-4: Results of screening based on PSI of single- and multi-vehicle crashes per HAY

Further analysis of the results in order to investigate the impact of percentage of residential area is based on the results in Table 7-5 and Table 7-6. Figure 8-5 shows that the maximum percentage of residential area is 90% compared to all other 17 land-use areas. Furthermore, 33.52% (60 out of 179) of HAYs in Riyadh have 15% or less of residential areas. Also, 42% of them have between 45% and 90% of residential areas. The overall average of percentage of residential areas in these 179 HAYs is 34.59%.

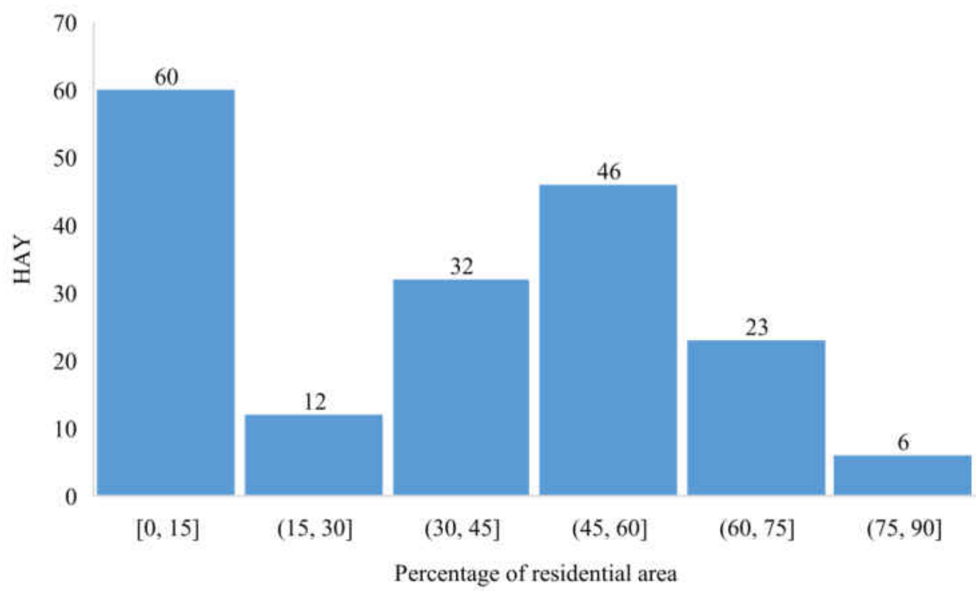


Figure 8-5: Percentage of residential area

The overall percentage of residential areas (total residential area divided by total area) is 8.3, and it was used for calculating the relative risk of fatal, PDO, pedestrian, and multi-vehicle crashes. There are 48 out of 179 HYAs that have less than 8% of the residential land-use, as is shown in Figure 8-6.

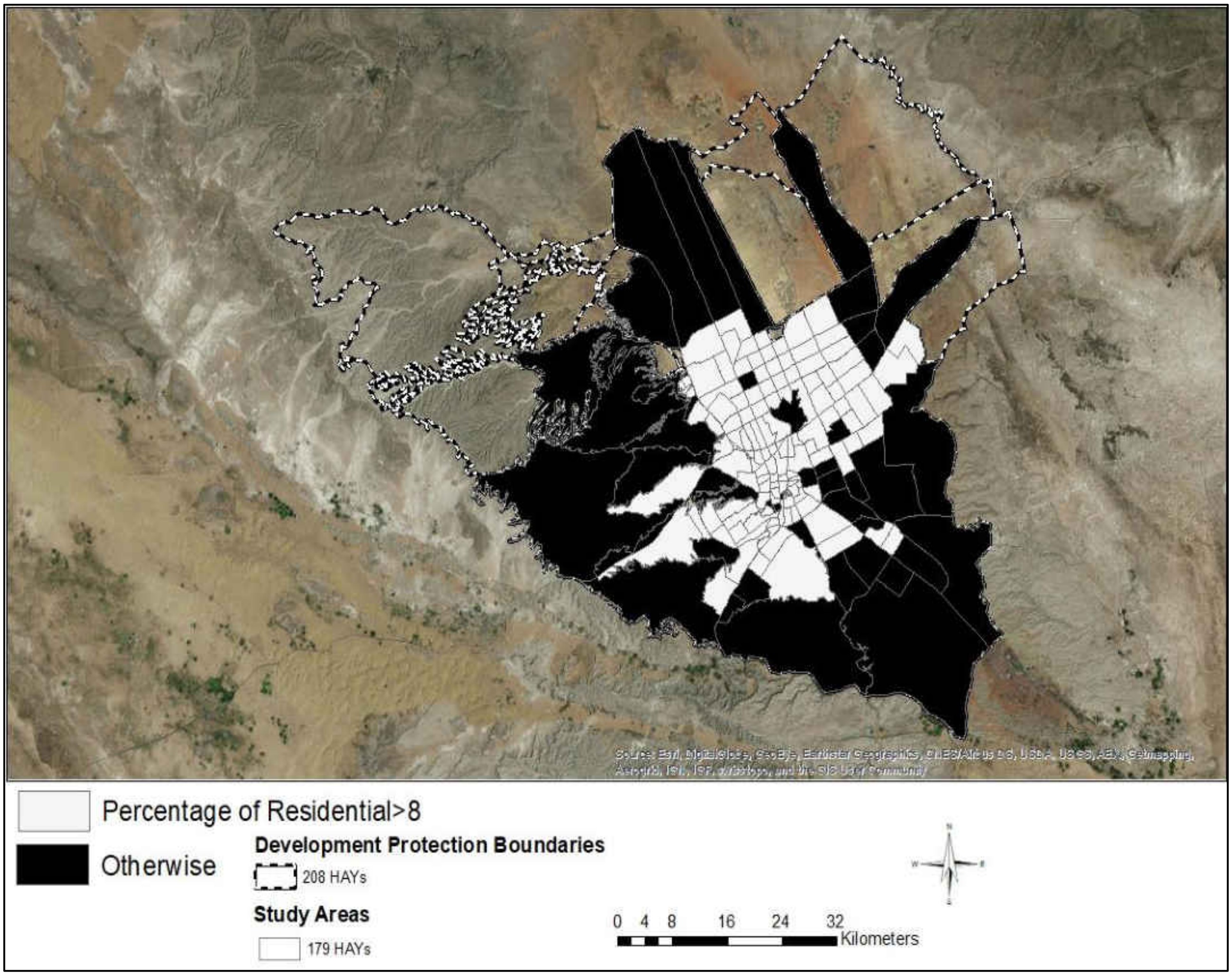


Figure 8-6: Distributions of the residential land-use based on its percentage of the total land use

The following equation was employed to compute the relative risk of these crashes for the residential area factor considering all other covariates are zero:

$$EXP(X_i * \beta_j) \tag{38}$$

Figure 8-7 shows that the relative risk of fatal crashes at 2.3% of residential areas is almost 1.87 times the overall percentage of residential area.

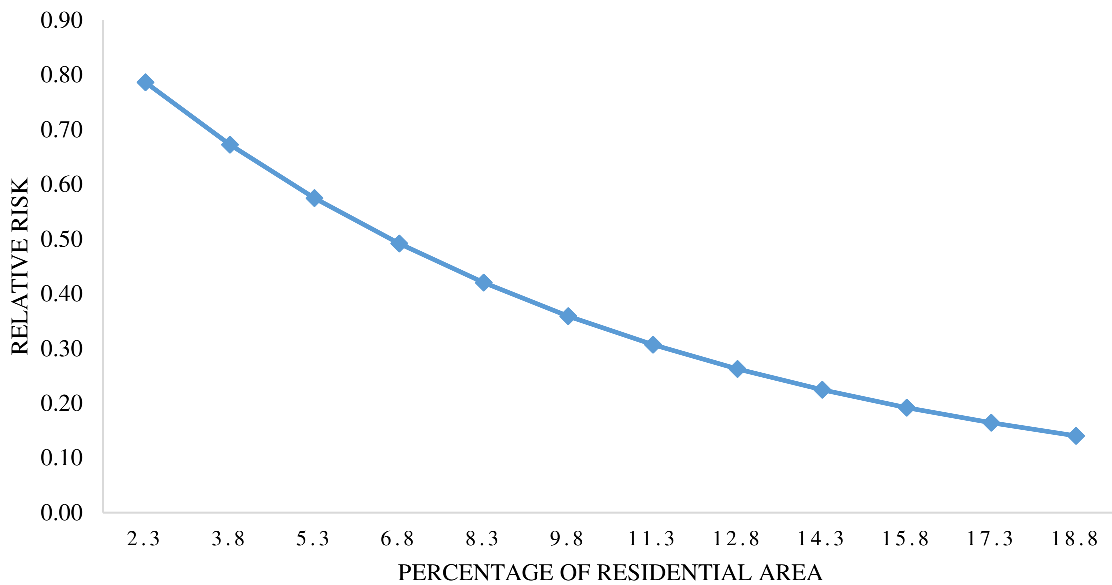


Figure 8-7: Relative risks of fatal crashes

Regarding PDO, pedestrian, and multi-vehicle crashes, the relative risk of fatal crashes at the percentage of residential areas of 2.3 were 48.23%, 49.097%, and 50.53% respectively, lower than at the overall percentage of residential areas of 8.3, as is shown in Figure 8-8. Moreover, the increased trend of the relative risk of pedestrian crashes is the lowest compared to the other type of crash (i.e., multi-vehicle).

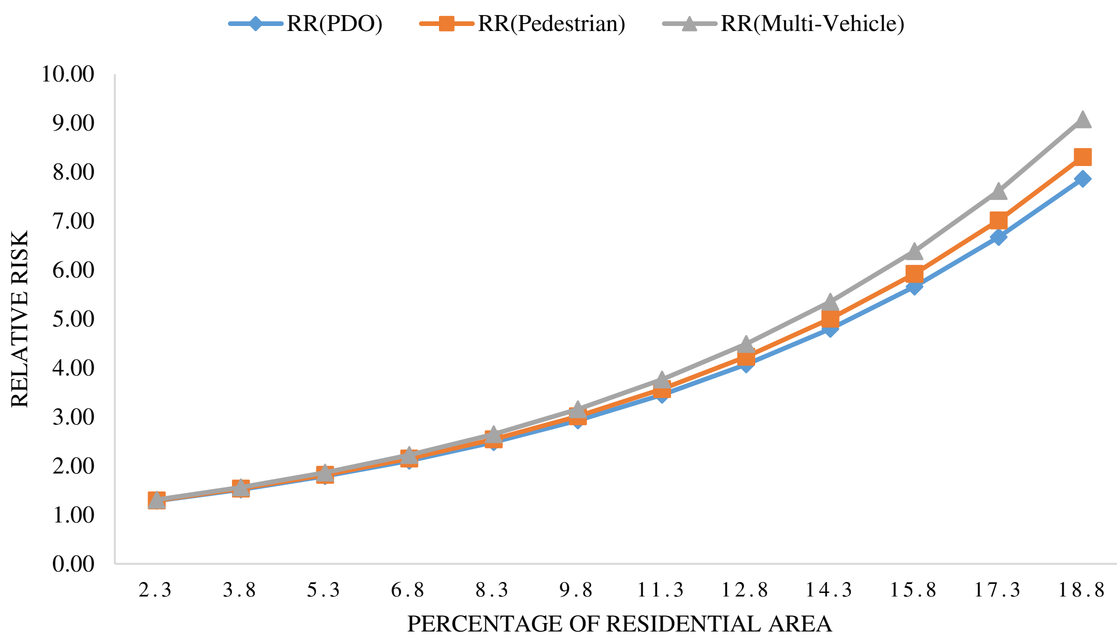


Figure 8-8: Relative risk of PDO, pedestrian, and multi-vehicle crashes

The goal of the KSA's National Transformation Program 2020 relative to reducing traffic fatalities is ambitious, but achievable. The National Strategic Plan for Traffic Safety is one of the significant tools that can be used to achieve it. Based on the plan's vision, ten key 'action areas' were identified to improve road safety throughout the KSA, including data and research. An implication of this study's results to reduce these crashes is to share mapping of crash data and the analysis results in an easy and professional way with law enforcement, traffic engineers in related departments, transportation planning agencies, and interested researchers, and even with the public to improve people's traffic safety awareness. For example, Figure 8-9 illustrates an interactive web mapping application created for Riyadh's crashes per HAY. In this app, the distributions of all of crash data by severity and type were mapped with pop-up windows for each HAY to show the number of crashes.

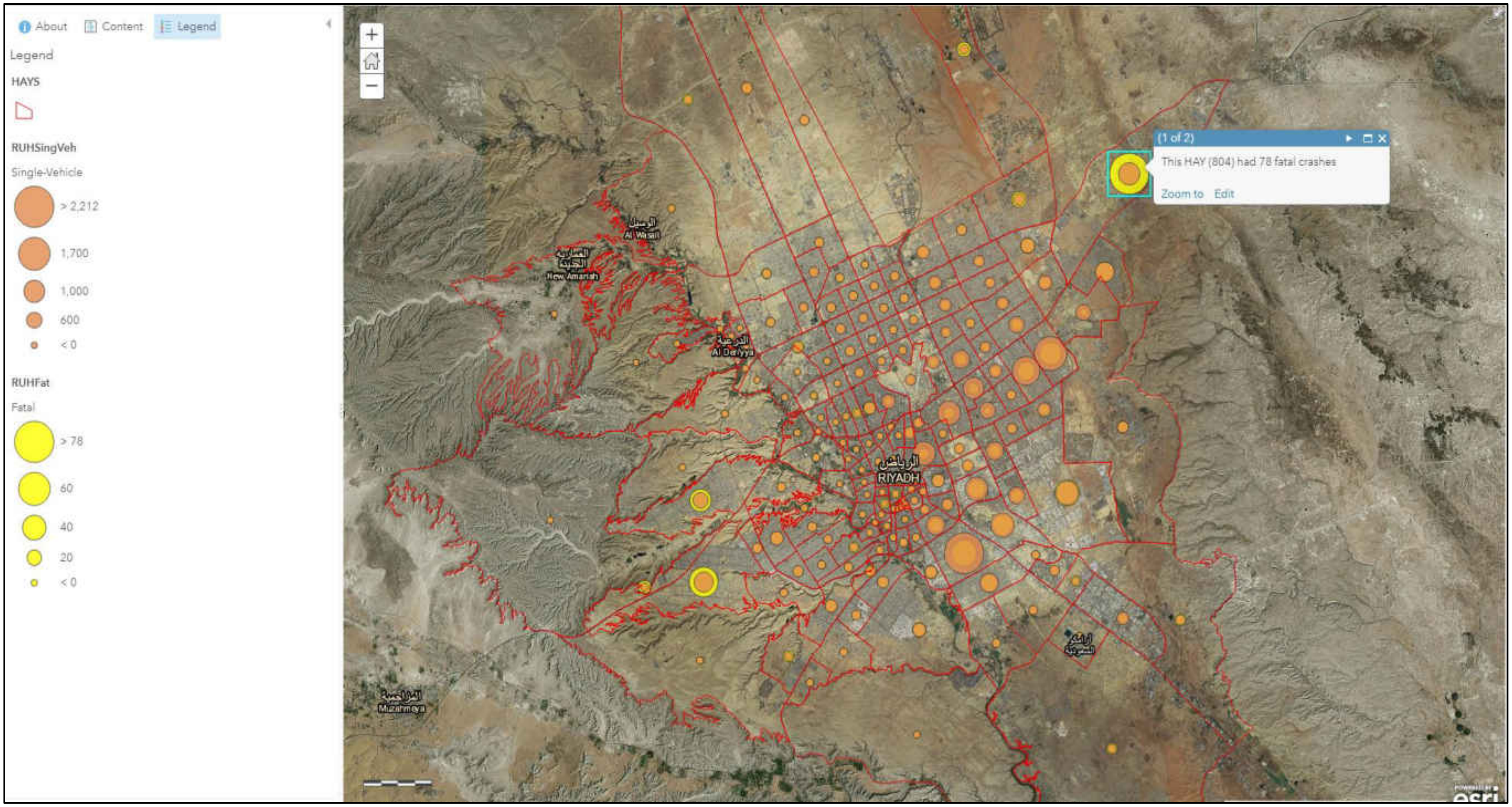


Figure 8-9: Web mapping application of Riyadh crashes per HAY

CHAPTER 9: CONCLUSIONS

This dissertation examined the contributing factors that could cause traffic crash occurrences, incorporating overdispersion and spatial effects at the macro-level, in the City of Riyadh, Saudi Arabia. It also suggests potential countermeasures based on the results in order to improve traffic safety. The primary objectives of this study are: 1) to explore the impact of local variations of parameters in spatial count data; 2) to examine the causes and characteristics of pedestrian crashes at the macroscopic level; and 3) to conduct a safety investigation of traffic crashes by severity and type at the macro-level as well.

Chapter 5 conducts a comparison between the Geographically Weighted Negative Binomial Regression (GWNBR) and spatially varying coefficients (SVC) model, which are both used for modeling spatially correlated data to account for spatial heterogeneity and to minimize spatial dependency (autocorrelation). It was found that incorporating overdispersion and spatial effects in crash frequency modeling would improve the model estimation, where the GWNBR performs better than all other developed models, including global negative binomial, random parameter, geographically weighted regression, either with Poisson distribution or negative binomial with global overdispersion, and spatially varying coefficients models.

In Chapter 6, the variation in pedestrian crashes among HAYs in Riyadh was inspected using several aspatial and spatial (i.e., Poisson-gamma and Poisson-lognormal) models. The results show that the spatial Poisson-lognormal model outperformed its counterparts in both total and severe pedestrian models.

There were common key variables that affect both total and severe pedestrian crashes (i.e., VKT, population density, and the number of traffic signals). However, there were some factors affecting total pedestrian crash frequency that are different from those affecting severe pedestrian crashes. In the total pedestrian crash model, the proportion of people whose education level is higher than high school had a negative association with pedestrian crashes. Meanwhile, in the severe pedestrian crash model, it was revealed that areas with a low socio-economic status (low educational attainment, unemployed status, etc.) are prone to higher severe pedestrian-involved crash frequency. It explains some of the risks related to the areas most affected by these crashes. Based on that, well-structured educational and awareness programs can be applied in these areas to enhance pedestrian safety. With respect to land-use effects, the HAYs with a greater proportion of commercial area have a greater likelihood of total and severe pedestrian-involved crashes, while a significantly higher proportion of agricultural and educational areas decreases it.

In addition, in Chapter 7, the multivariate Poisson-lognormal CAR model in a Bayesian framework was used to simultaneously model correlated crash counts, which accounts for spatial autocorrelation and correlations among different crash counts by severity and type levels at the zonal level. Then, it was compared to the corresponding aspatial model. It was found that the multivariate models performed better than the corresponding univariate models in both crash type and severity models in terms of goodness-of-fit measures. In addition, the spatial univariate and multivariate models outperformed the aspatial counterparts. This indicates the importance of incorporating both the heterogeneity effects and spatial correlation effects in the zonal level analysis.

The results revealed that there are significant correlations between the total random effect (heterogeneity and spatial effects) for all response variables, indicating a strong shared geographical pattern of risk between each of these variables. In terms of the crash type model, the correlation is very strong between pedestrian and bicycle crashes, yet it is relatively less strong between single-vehicle and pedestrian or bicycle crashes. Furthermore, several contributing factors, including the traffic volume, road characteristics, socio-economics and demographics, and land-use factors have a significant influence on different crash severities and types. In contrast, the variable estimates and statistical significance across different models are varying.

As practical implications of the findings of this study, Chapter 8 illustrates the overall crash distributions within the study area using the crash hot zone screening results. The results revealed some of HAYs are exposed to all of crash types, while others show only motorized and/or non-motorized-related crashes, and a few were solely highly exposed for only one type of crash. This would assist decision makers and practitioners with a useful implication to understand risks related to the HAYs most affected by these crashes.

Overall, this dissertation provides a comprehensive understanding of traffic safety impacts at the macroscopic level and their contributing factors in Riyadh, Saudi Arabia. The findings of this study could help and guide decision makers and practitioners in selecting more appropriate safety countermeasures and in implementing appropriate education and effective campaigns, as well as assist in the development of engineering countermeasures, traffic control, and management policy and applying law enforcement to enhance traffic safety policies in Riyadh. This would improve traffic safety with the aim toward achieving the goal of the KSA's National Transformation Program 2020 relative to reducing traffic fatalities. It would be expected that the results would also be beneficial to other countries in the region with the same socioeconomic structure and characteristics.

APPENDIX: TRAFFIC CRASH REPORT

Table 9-1: Traffic crash report

Time and Date	Time of	Min	Hr	Date	AM	PM	Crash plan											
	Crash																	
	Call Order																	
	First police arrival																	
	Report closed																	
location	City/district/neibourhood																	
	Latitude																	
	Longitude																	
	Street name and number																	
	Closest street or landmark																	
	Distance from landmark		Direction		KM road board													
Vehicle(s)	No.	direction	Plate no.	Reg. Type	Manufac. Country	Issue dept.	color	Model	Make	Type	Acc point				status	Name	No.	sign
											A	B	C	D				
	1																	
	2																	
Person(s)	No.	Name	Nationa.	Id	Type	Health Sta.	%	Ins. Com.	No.	Exp. Date	Lic. type							
	1																	
	2																	
Crash Summary																		
Witness	Name	ID	No															
	1																	
	2																	
	Surface cond.		Light		reason	Acc place	W.	Acc type	Prv. Dmg	Pub dmg								
	Dry	Wet	Clear	Dark														

LIST OF REFERENCES

Abbas, K.A., 2004. Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt. Accid. Anal. Prev. 36, 149–163.

Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical unit based analysis in the context of transportation safety planning. Transp. Res. Part A Policy Pract. 49, 62–75.

Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating trip and roadway characteristics to manage safety in traffic analysis zones. Transp. Res. Rec. J. Transp. Res. Board 20–28.

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32, 633–642.

Agresti, A., Kateri, M., 2011. Categorical data analysis. Springer.

Aguero-Valverde, J., 2013. Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. Accid. Anal. Prev. 59, 365–373.

Aguero-Valverde, J., Jovanis, P., 2010. Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. Transp. Res. Rec. J. Transp. Res. Board 21–32.

Aguero-Valverde, J., Jovanis, P., 2009. Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. Transp. Res. Rec. J. Transp. Res. Board 82–91.

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accid. Anal. Prev. 38, 618–625.

Al-Ghamdi, A.S., 2003. Analysis of traffic accidents at urban intersections in Riyadh. Accid. Anal. Prev. 35, 717–724.

Al-Ghamdi, A.S., 2002. Pedestrian–vehicle crashes and analytical techniques for stratified contingency tables. Accid. Anal. Prev. 34, 205–214.

Al-Ghamdi, A.S., 1996. Road Accidents in Saudi Arabia: A Comparative and Analytical Study. Proc. 2nd Int. Conf. Urban Transp. Environ. 21st Century. pp.231–253.

Al-Shammari, N., Bendak, S., Al-Gadhi, S., 2009. In-depth analysis of pedestrian crashes in Riyadh. Traffic Inj. Prev. 10, 552–559.

Alarifi, S.A., Alkahtani, K.F., Abdel-Aty, M.A., Kher, S.O., Mohammed, A.M., AlMojil, A.H., 2017. Corridor safety evaluation in a developing country using a survey based approach. J. Transp. Saf. Secur. 1–18.

Alkahtani, K.F., Abdel-Aty, M., Lee, J., 2018. A zonal level safety investigation of pedestrian crashes in Riyadh, Saudi Arabia. Int. J. Sustain. Transp. 1–13. doi:10.1080/15568318.2018.1463417

Altwaijri, S., Quddus, M., Bristow, A., 2012. Analysing the severity and frequency of traffic crashes in riyadh city using statistical models. Int. J. Transp. Sci. Technol. 1, 351–364.

Altwaijri, S., Quddus, M.A., Bristow, A., 2011. Factors affecting the severity of traffic crashes in Riyadh City, in: Proceedings of the 90th Annual Meeting of the Transportation Research Board. Washington, D.C.

Amoh-Gyimah, R., Aidoo, E.N., Akaateba, M.A., Appiah, S.K., 2017a. The effect of natural and built environmental characteristics on pedestrian-vehicle crash severity in Ghana. Int. J. Inj. Contr. Saf. Promot. 24, 459–468.

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2017b. The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models. Anal. methods Accid. Res. 13, 28–51.

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: a cross-comparison of estimation methods. Accid. Anal. Prev. 93, 147–159.

Amoros, E., Martin, J.L., Laumon, B., 2003. Comparison of road crashes incidence and severity between some French counties. Accid. Anal. Prev. 35, 537–547.

Ancelet, S., Abellan, J.J., Del Rio Vilas, V.J., Birch, C., Richardson, S., 2012. Bayesian shared spatial-component models to combine and borrow strength across sparse disease surveillance sources. Biometrical J. 54, 385–404.

Babalik-Sutcliffe, E., 2013. Urban form and sustainable transport: lessons from the Ankara case. Int. J. Sustain. Transp. 7, 416–430.

Bailey, M.J., Dynarski, S.M., 2011. Gains and gaps: Changing inequality in US college entry and completion. National Bureau of Economic Research.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2014. Hierarchical modeling and analysis for spatial data. Crc Press.

Belley, P., Lochner, L., 2007. The changing role of family income and ability in determining educational achievement. J. Hum. Cap. 1, 37–89.

Besag, J., 1991. Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. 43, 45–59.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B 192–236.

Cafiso, S., Di Silvestro, G., Persaud, B., Begum, M., 2010. Revisiting variability of dispersion parameter of safety performance for two-lane rural roads. Transp. Res. Rec. J. Transp. Res. Board 38–46.

Cai, Q., Abdel-Aty, M., Lee, J., 2017a. Macro-level vulnerable road users crash analysis: a Bayesian joint modeling approach of frequency and proportion. Accid. Anal. Prev. 107, 11–19.

Cai, Q., Abdel-Aty, M., Lee, J., Eluru, N., 2017b. Comparative analysis of zonal systems for macro-level crash modeling. J. Safety Res.

Cai, Q., Abdel-Aty, M., Lee, J., Huang, H., 2018a. Integrating macro-and micro-level safety analyses: a Bayesian approach incorporating spatial interaction. Transp. A Transp. Sci. 1–22.

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., Wang, X., 2018b. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. Anal. Methods Accid. Res. 19, 1–15.

Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: incorporating spatial spillover effects in dual state count models. Accid. Anal. Prev. 93, 14–22.

Carson, J., Mannering, F., 2001. The effect of ice warning signs on ice-accident frequencies and severities. Accid. Anal. Prev. 33, 99–109.

Central Department of Statistics and Information, 2010. Results of Population and Housing Census of 1430H. CDSI, Ministry of Economy and Planning, Riyadh.

Chatterjee, S., Hadi, A.S., 2015. Regression analysis by example. John Wiley & Sons.

Cho, G., Rodriguez, D.A., Khattak, A.J., 2009. The role of the built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. Accid. Anal. Prev. 41, 692–702.

Congdon, P., 2014. Applied bayesian modelling. John Wiley & Sons.

Coruh, E., Bilgic, A., Tortum, A., 2015. Accident analysis with aggregated data: The random parameters negative binomial panel count data model. Anal. methods Accid. Res. 7, 37–49.

Cottrill, C.D., Thakuriah, P.V., 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. Accid. Anal. Prev. 42, 1718–1728.

da Silva, A.R., Rodrigues, T.C.V., 2016. A SAS® Macro for Geographically Weighted Negative Binomial Regression. Available at: <http://support.sas.com/resources/papers/proceedings16/8000-2016.pdf>.

da Silva, A.R., Rodrigues, T.C.V., 2014. Geographically weighted negative binomial regression—incorporating overdispersion. Stat. Comput. 24, 769–783.

Directorate General of Traffic, 2013. Statistical Report for traffic accidents. Riyadh.

Dissanayake, D., Aryaija, J., Wedagama, D.M.P., 2009. Modelling the effects of land use and temporal factors on child pedestrian casualties. Accid. Anal. Prev. 41, 1016–1024.

Domenichini, L., Branzi, V., Meocci, M., 2018. Virtual testing of speed reduction schemes on urban collector roads. Accid. Anal. Prev. 110, 38–51.

Dong, N., Huang, H., Lee, J., Gao, M., Abdel-Aty, M., 2016. Macroscopic hotspots identification: a Bayesian spatio-temporal interaction approach. Accid. Anal. Prev. 92, 256–264.

El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transp. Res. Rec. J. Transp. Res. Board 9–16.

Evans, J.D., 1996. Straightforward statistics for the behavioral sciences. Brooks/Cole.

Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson–gamma models. Accid. Anal. Prev. 42, 1273–1282.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2003. Bayesian data analysis, (2nd ed.). ed. Chapman & Hall/CRC Boca Raton, FL.

Gomes, M.J.T.L., Cunto, F., da Silva, A.R., 2017. Geographically weighted negative binomial regression applied to zonal level safety performance models. Accid. Anal. Prev. 106, 254–261.

Greene, W.H., 2016. LIMDEP: Version 11.0: Reference Guide. Econometric Software.

Greene, W.H., 2007. NLOGIT 6.0 reference guide. Econom. software, inc., New York.

Gujarati, D.N., 2009. Basic econometrics. Tata McGraw-Hill Education.

Guo, Q., Xu, P., Pei, X., Wong, S.C., Yao, D., 2017. The effect of road network patterns on pedestrian safety: a zone-based Bayesian spatial modeling approach. Accid. Anal. Prev. 99, 114–124.

Guo, X., Carlin, B.P., 2004. Separate and joint modeling of longitudinal and event time data using standard computer packages. Am. Stat. 58, 16–24.

Hadayeghi, A., Shalaby, A., Persaud, B., 2010a. Development of planning-level transportation safety models using full Bayesian semiparametric additive techniques. J. Transp. Saf. Secur. 2, 45–68.

Hadayeghi, A., Shalaby, A., Persaud, B., 2007. Safety prediction models: proactive tool for safety evaluation in urban transportation planning applications. Transp. Res. Rec. J. Transp. Res. Board 225–236.

Hadayeghi, A., Shalaby, A., Persaud, B., 2003. Macrolevel accident prediction models for evaluating safety of urban transportation systems. Transp. Res. Rec. J. Transp. Res. Board 87–95.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010b. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accid. Anal. Prev. 42, 676–688.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., Cheung, C., 2006. Temporal transferability and updating of zonal level accident prediction models. Accid. Anal. Prev. 38, 579–589.

Hanna, C.L., Laflamme, L., Bingham, C.R., 2012. Fatal crash involvement of unlicensed young drivers: county level differences according to material deprivation and urbanicity in the United States. Accid. Anal. Prev. 45, 291–295.

Hassan, H.M., Al-Faleh, H., 2013. Exploring the risk factors associated with the size and severity of roadway crashes in Riyadh. J. Safety Res. 47, 67–74.

Hauer, E., 1996. Detection of safety deterioration in a series of accident counts. Transp. Res. Rec. J. Transp. Res. Board 38–43.

Hauer, E., Hakkert, A.S., 1988. Extent and some implications of incomplete accident reporting. Transp. Res. Rec. 1185, 1–10.

Hauer, E., Kononov, J., Allery, B., Griffith, M., 2002. Screening the road network for sites with promise. Transp. Res. Rec. J. Transp. Res. Board 27–32.

Herbel, S., Laing, L., McGovern, C., 2010. Highway Safety Improvement Program Manual. US Department of Transportation, Federal Highway Administration, Office of Safety.

Hilbe, J.M., 2011. Negative binomial regression. Cambridge University Press.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida: Bayesian spatial modeling. Transp. Res. Rec. J. Transp. Res. Board 27–37.

Huang, H., Cynecki, M., 2000. Effects of traffic calming measures on pedestrian and motorist behavior. Transp. Res. Rec. J. Transp. Res. Board 26–31.

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal crash prediction with application in hot zones identification. J. Transp. Geogr. 54, 248–

256.

Huang, H., Zhou, H., Wang, J., Chang, F., Ma, M., 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. Anal. methods Accid. Res. 14, 10–21.

Karim, M., Wahba, M., Sayed, T., 2013. Spatial effects on zone-level collision prediction models. Transp. Res. Rec. J. Transp. Res. Board 50–59.

Karlaftis, M.G., Tarko, A.P., 1998. Heterogeneity considerations in accident modeling. Accid. Anal. Prev. 30, 425–433.

Kim, D.-G., Washington, S., 2006. The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. Accid. Anal. Prev. 38, 1094–1100.

Kim, K., Brunner, I., Yamashita, E., 2006. Influence of land use, population, employment, and economic activity on accidents. Transp. Res. Rec. J. Transp. Res. Board 56–64.

Kim, K., Yamashita, E., 2002. Motor vehicle crashes and land use: empirical analysis from Hawaii. Transp. Res. Rec. J. Transp. Res. Board 73–79.

Koushki, P.A., Al-Ghadeer, A.M., 1992. Driver noncompliance with Traffic regulations in rapidly developing urban areas of Saudi Arabia. Transp. Res. Rec. 1.

Kulmala, R., 1995. Safety at rural three-and four-arm junctions. Development and application of accident prediction models.

Ladron de Guevara, F., Washington, S., Oh, J., 2004. Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. Transp. Res. Rec.

J. Transp. Res. Board 191–199.

Lai, P.-C., So, F.-M., Chan, K.-W., 2008. Spatial epidemiological approaches in disease mapping and analysis. CRC Press.

LaScala, E.A., Gerber, D., Gruenewald, P.J., 2000. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accid. Anal. Prev. 32, 651–658.

LaScala, E.A., Gruenewald, P.J., Johnson, F.W., 2004. An ecological study of the locations of schools and child pedestrian injury collisions. Accid. Anal. Prev. 36, 569–576.

Lawson, A.B., Browne, W.J., Rodeiro, C.L.V., 2003. Disease mapping with WinBUGS and MLwiN. John Wiley & Sons.

Lee, D., 2011. A comparison of conditional autoregressive models used in Bayesian disease mapping. Spat. Spatiotemporal. Epidemiol. 2, 79–89.

Lee, J., Abdel-Aty, M., 2017. Macro-level analysis of bicycle safety: Focusing on the characteristics of both crash location and residence. Int. J. Sustain. Transp. 1–8.

Lee, J., Abdel-Aty, M., Cai, Q., 2017. Intersection crash prediction modeling with macro-level data from various geographic units. Accid. Anal. Prev. 102, 213–226.

Lee, J., Abdel-Aty, M., Choi, K., 2014a. Analysis of residence characteristics of at-fault drivers in traffic crashes. Saf. Sci. 68, 6–13.

Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015a. Multi-level hot zone identification for pedestrian safety. Accid. Anal. Prev. 76, 64–73.

Lee, J., Abdel-Aty, M., Choi, K., Siddiqui, C., 2013. Analysis of residence characteristics of drivers, pedestrians, and bicyclists involved in traffic crashes, in: Transportation Research Board 92nd Annual Meeting.

Lee, J., Abdel-Aty, M., Jiang, X., 2015b. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accid. Anal. Prev. 78, 146–154.

Lee, J., Abdel-Aty, M., Jiang, X., 2014b. Development of zone system for macro-level traffic safety analysis. J. Transp. Geogr. 38, 13–21.

Lee, J., Abdel-Aty, M.A., Jiang, X., 2014c. Macroscopic Multivariate Crash Modeling for Motor Vehicle, Bicycle, and Pedestrian Crashes, in: Transportation Research Board 93rd Annual Meeting.

Leroux, B.G., Lei, X., Breslow, N., 2000. Estimation of disease rates in small areas: a new mixed model for spatial dependence, in: Statistical Models in Epidemiology, the Environment, and Clinical Trials. Springer, pp. 179–191.

LeSage, J.P., 1997. Bayesian estimation of spatial autoregressive models. Int. Reg. Sci. Rev. 20, 113–129.

LeSage, J.P., Pace, R.K., 2004. Models for spatially dependent missing data. J. Real Estate Financ. Econ. 29, 233–254.

Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. Accid. Anal. Prev. 27, 675–685.

Li, Z., Wang, W., Liu, P., Bigham, J.M., Ragland, D.R., 2013. Using geographically weighted

Poisson regression for county-level crash modeling in California. Saf. Sci. 58, 89–97.

Loo, B.P.Y., Anderson, T.K., 2015. Spatial Analysis Methods of Road Traffic Collisions. CRC Press.

Loo, B.P.Y., Chow, A.S.Y., 2008. Changing urban form in Hong Kong: what are the challenges on sustainable transportation? Int. J. Sustain. Transp. 2, 177–193.

Lord, D., Bonneson, J., 2007. Development of accident modification factors for rural frontage road segments in Texas. Transp. Res. Rec. J. Transp. Res. Board 20–27.

Lord, D., Geedipally, S.R., Guikema, S.D., 2010. Extension of the Application of Conway Maxwell Poisson Models: Analyzing Traffic Crash Data Exhibiting Underdispersion. Risk Anal. 30, 1268–1276.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44, 291–305.

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

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid. Anal. Prev. 37, 35–46.

Loukaitou-Sideris, A., Liggett, R., Sung, H.-G., 2007. Death on the crosswalk A study of pedestrian-automobile collisions in los angeles. J. Plan. Educ. Res. 26, 338–351.

Lovegrove, G.R., Sayed, T., 2006. Macro-level collision prediction models for evaluating neighbourhood traffic safety. Can. J. Civ. Eng. 33, 609–621.

Ma, J., Kockelman, K., 2006. Bayesian multivariate Poisson regression for models of injury count, by severity. Transp. Res. Rec. J. Transp. Res. Board 24–34.

Ma, X., Chen, S., Chen, F., 2016. Correlated random-effects bivariate poisson lognormal model to study single-vehicle and multivehicle crashes. J. Transp. Eng. 142, 4016049.

Mabunda, M.M., Swart, L.-A., Seedat, M., 2008. Magnitude and categories of pedestrian fatalities in South Africa. Accid. Anal. Prev. 40, 586–593.

MacNab, Y.C., 2004. Bayesian spatial and ecological models for small-area accident and injury analysis. Accid. Anal. Prev. 36, 1019–1028.

Malyshkina, N. V, Mannering, F.L., 2010. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accid. Anal. Prev. 42, 131–139.

Maycock, G., Hall, R.D., 1984. Accidents at 4-arm roundabouts.

McMillen, D.P., 2004. Geographically weighted regression: the analysis of spatially varying relationships.

Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26, 471–482.

Miaou, S.-P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. Transp.

Res. Rec. J. Transp. Res. Board 31–40.

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

Moeinaddini, M., Asadi-Shekari, Z., Shah, M.Z., 2014. The relationship between urban street networks and the number of transport fatalities at the city level. Saf. Sci. 62, 114–120.

Naderan, A., Shahi, J., 2010. Aggregate crash prediction models: Introducing crash generation concept. Accid. Anal. Prev. 42, 339–346.

Nakaya, T., Charlton, M., Lewis, P., Fotheringham, A.S., Brunsdon, C., 2012. Windows Application for Geographically Weighted Regression Modeling.

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically weighted Poisson regression for disease association mapping. Stat. Med. 24, 2695–2717.

Narayanamoorthy, S., Paleti, R., Bhat, C.R., 2013. On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transp. Res. part B Methodol. 55, 245–264.

Nashad, T., Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M.A., 2016. Joint modeling of pedestrian and bicycle crashes: copula-based approach. Transp. Res. Rec. J. Transp. Res. Board 119–127.

National Strategy for Traffic Safety, 2014. National strategic plan for traffic safety, National Strategy for Traffic Safety. King Abdul Aziz City for Science and Technology.

Ng, K., Hung, W., Wong, W., 2002. An algorithm for assessing the risk of traffic accident. J.

Safety Res. 33, 387–410.

Noland, R., Quddus, M., 2004. Analysis of pedestrian and bicycle casualties with regional panel data. Transp. Res. Rec. J. Transp. Res. Board 28–33.

Noland, R.B., 2003. Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. Accid. Anal. Prev. 35, 599–611.

Noland, R.B., Klein, N.J., Tulach, N.K., 2013. Do lower income areas have more pedestrian casualties? Accid. Anal. Prev. 59, 337–345.

Noland, R.B., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. Accid. Anal. Prev. 36, 525–532.

Noland, R.B., Quddus, M.A., 2005. Congestion and safety: A spatial analysis of London. Transp. Res. Part A Policy Pract. 39, 737–754.

Noland, R.B., Quddus, M.A., 2004. A spatially disaggregate analysis of road casualties in England. Accid. Anal. Prev. 36, 973–984.

O'Sullivan, D., Unwin, D., 2003. Geographic Information Analysis, Science/Geography. Wiley.

Osama, A., Sayed, T., 2017. Investigating the effect of spatial and mode correlations on active transportation safety modeling. Anal. methods Accid. Res. 16, 60–74.

Osama, A., Sayed, T., 2016. Evaluating the impact of bike network indicators on cyclist safety using macro-level collision prediction models. Accid. Anal. Prev. 97, 28–37.

Park, E., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transp. Res. Rec. J. Transp. Res. Board 1–6.

Persaud, B., Nguyen, T., 1998. Disaggregate safety performance models for signalized intersections on Ontario provincial roads. Transp. Res. Rec. J. Transp. Res. Board 113–120.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. J. Transp. Eng. 122, 105–113.

Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013. Traffic analysis zone level crash estimation models based on land use characteristics. Accid. Anal. Prev. 50, 678–687.

Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accid. Anal. Prev. 40, 1486–1497.

Quddus, M.A., Bristow, A., Altwaijri, S., 2012. Spatial Analysis of Traffic Crashes in Riyadh City, Saudi Arabia, Using GIS and Statistical Models, in: Proceedings of the 91th Annual Meeting of the Transportation Research Board. Washington, D.C.

Räsänen, M., Lajunen, T., Alticafarbay, F., Aydin, C., 2007. Pedestrian self-reports of factors influencing the use of pedestrian bridges. Accid. Anal. Prev. 39, 969–973.

Rhee, K.-A., Kim, J.-K., Lee, Y., Ulfarsson, G.F., 2016. Spatial regression analysis of traffic crashes in Seoul. Accid. Anal. Prev. 91, 190–199.

Rifaat, S.M., Tay, R., Raihan, S.M., Fahim, A., Touhidduzzaman, S.M., 2017. Vehicle-Pedestrian crashes at Intersections in Dhaka city. Open Transp. J. 11.

Sarkar, S., Tay, R., Hunt, J., 2011. Logistic regression model of risk of fatality in vehicle-pedestrian crashes on national highways in Bangladesh. Transp. Res. Rec. J. Transp. Res. Board 128–137.

Saudi Arabia's Vision 2030, 2030, 2016. National Transformation Program 2020, http://vision2030.gov.sa/sites/default/files/NTP_En.pdf.

Schabenberger, O., Gotway, C.A., 2004. Statistical Methods for Spatial Data Analysis, Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Shankar, V., Albin, R., Milton, J., Mannering, F., 1998. Evaluating median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model. Transp. Res. Rec. J. Transp. Res. Board 44–48.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27, 371–389.

Siddiqui, C., Abdel-Aty, M., 2016. Geographical Boundary Dependency Versus Roadway Hierarchy in Macroscopic Safety Modeling: Analysis with Motor Vehicle Crash Data. Transp. Res. Rec. J. Transp. Res. Board 59–71.

Siddiqui, C., Abdel-Aty, M., 2012. Nature of modeling boundary pedestrian crashes at zones. Transp. Res. Rec. J. Transp. Res. Board 31–40.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2014. Implications of pedestrian safety planning factors in areas with minority and low-income populations. Int. J. Sustain. Transp. 8, 360–381.

Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid. Anal. Prev. 45, 382–391.

Song, J.J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. J. Multivar. Anal. 97, 246–273.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2005. WinBUGS User Manual. MRC Biostatistics Unit, Cambridge.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B (Statistical Methodol. 64, 583–639.

Stamatiadis, N., Puccini, G., 2000. Socioeconomic descriptors of fatal crash rates in the Southeast USA. Inj. Control Saf. Promot. 7, 165–173.

Tasic, I., Porter, R.J., Brewer, S., 2016. Applications of Generalized Additive Models and Bayesian Hierarchical Models for Areal Safety Analysis of Urban Multimodal Transportation Systems: Case Study in the City of Chicago, in: Transportation Research Board 95th Annual Meeting.

Thomas, A., Best, N., Lunn, D., Arnold, R., Spiegelhalter, D., 2004. GeoBugs user manual. Cambridge Med. Res. Counc. Biostat. Unit.

Truong, L.T., Kieu, L.-M., Vu, T.A., 2016. Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam. Accid. Anal. Prev. 94, 153–161.

Tulu, G.S., Washington, S., Haque, M.M., King, M.J., 2015. Investigation of pedestrian crashes on two-way two-lane rural roads in Ethiopia. Accid. Anal. Prev. 78, 118–126.

Ukkusuri, S., Hasan, S., Aziz, H., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transp. Res. Rec. J. Transp. Res.

Board 98–106.

Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Saf. Sci. 50, 1141–1151.

Wakefield, J.C., Best, N.G., Waller, L., 2000. Bayesian approaches to disease mapping. Spat. Epidemiol. methods Appl. 104–127.

Wang, C., Quddus, M., Ison, S., 2009. The effects of area-wide road speed and curvature on traffic casualties in England. J. Transp. Geogr. 17, 385–395.

Wang, X., Yang, J., Lee, C., Ji, Z., You, S., 2016. Macro-level safety analysis of pedestrian crashes in Shanghai, China. Accid. Anal. Prev. 96, 12–21.

Wang, X., Yuan, J., Schultz, G.G., Fang, S., 2018. Investigating the safety impact of roadway network features of suburban arterials in Shanghai. Accid. Anal. Prev. 113, 137–148.

Wang, Y., Kockelman, K.M., 2013. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accid. Anal. Prev. 60, 71–84.

Washington, S., 2006. Incorporating safety into long-range transportation planning. Transportation Research Board.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. Statistical and econometric methods for transportation data analysis. CRC press.

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. Accid.

Anal. Prev. 61, 129–137.

Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geogr. Syst. 7, 161–187.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 434–449.

WHO, 2015a. Global Status Report on Road Safety 2015. World Health Organization, Geneva.

WHO, 2015b. Causes of death (2004) http://www. who. int/healthinfo/global_burden_disease. GBD_report_2004update_part2. pdf.

Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accid. Anal. Prev. 41, 137–145.

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16–25.

Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. Accid. Anal. Prev. 70, 110–120.

Xu, P., Huang, H., Dong, N., Wong, S.C., 2017. Revisiting crash spatial heterogeneity: A Bayesian spatially varying coefficients approach. Accid. Anal. Prev. 98, 330–337.

Yasmin, S., Eluru, N., 2018. A joint econometric framework for modeling crash counts by severity. Transp. A Transp. Sci. 14, 230–255.

Yasmin, S., Eluru, N., 2017. A joint econometric framework for modeling crash counts by severity. Transp. A Transp. Sci. 1–26.

Yasmin, S., Eluru, N., 2016. Latent segmentation based count models: analysis of bicycle safety in Montreal and Toronto. Accid. Anal. Prev. 95, 157–171.

Yu, R., Abdel-Aty, M., 2013. Investigating different approaches to develop informative priors in hierarchical Bayesian safety performance functions. Accid. Anal. Prev. 56, 51–58.

Zhang, G., Yau, K.K.W., Zhang, X., 2014. Analyzing fault and severity in pedestrian–motor vehicle accidents in China. Accid. Anal. Prev. 73, 141–150.

Zhang, Y., Bigham, J., Ragland, D., Chen, X., 2015. Investigating the associations between road network structure and non-motorist accidents. J. Transp. Geogr. 42, 34–47.