Investigating the link between the built environment and the incidence of pedestrian crashes in Cape Town, South Africa

by
Pascal Nteziyaremye

Dissertation presented for the degree of Doctor of Philosophy in the Faculty of Engineering at Stellenbosch University

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

100
1918 · 2018

Supervisor: Prof. Marion Sinclair

December 2018

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature:

Date: December 2018

Abstract

Pedestrians are the most vulnerable road users in the road environment, particularly in the developing world. To gain a better understanding of pedestrian crash causation, the built environment has been given much attention in the international traffic safety research. However, research of this nature is still scarce in the developing world, including South Africa. This study investigates the link between the built environment and the incidence of pedestrian crashes. The study used pedestrian crash data collected in Cape Town over a 3-year period between 2012 and 2014. The research method involved screening, geocoding and supplementing poor quality secondary data on pedestrian crashes. Moreover, the study applies a variety of analytical methods including univariate, bivariate, geospatial and multivariate analyses. Four GIS-based spatial analysis methods were used to identify clusters of pedestrian crashes within the study area. These methods include the planar kernel density estimation (KDE), the Anselin local Moran's I, the Getis-Ord Gi* and the Optimized Hot Spot Analysis (OHA). Two modelling techniques, the Generalised Linear Modelling (GLM) and Geographically Weighted Regression (GWR) modelling were used to relate the built environment and population variables to total; intersection; and killed and seriously injured (KSI) pedestrian crashes. For this analysis, the data was aggregated and analysed at the census suburb level. Among other results, it was found that population; land use mix; traffic signals; roundabouts/mini-circles; industrial use; four- and multi-legged intersections; and high mobility roads are associated with greater numbers of pedestrian crashes. The study also revealed that pedestrian crashes are positively related to socio-economic deprivation. In addition, spatial variations of the associations in the models were investigated and discussed. Hotspots of pedestrian crashes were identified mostly in the South Eastern regions of Cape Town which are also areas where economically-disadvantaged residents are concentrated. The presented models can be used to predict future pedestrian crashes using information that is easily available at the city level. The models are also crucial for the planning of safe walking environments which are particularly needed in South Africa and other developing countries.

Opsomming

Voetgangers is die kwesbaarste padgebruikers in die padomgewing, veral in die ontwikkelende wêreld. Om ŉ beter begrip van voetgangerbotsing oorsake te verkry, is baie aandag aan die geboude omgewing gegee in internasionale verkeerveiligheid navorsing. Navorsing van hierdie aard is egter steeds skaars in die ontwikkelende wêreld, insluitend Suid-Afrika. Hierdie studie ondersoek die skakel tussen die geboude omgewing en die voorkoms van voetgangerbotsings. Die studie het voetgangerbotsing data, ingesamel in Kaapstad oor ŉ drie-jaar periode tussen 2012 en 2014, gebruik. Die navorsingsmetode het sifting, geo-kodering en die aanvulling van lae-kwaliteit sekondêre data oor voetgangerbotsings ingesluit. Verder pas hierdie studie ŉ verskeidenheid van analitiese metodes toe, insluitend univariante, bivariante, geo-ruimtelike en multivariante analises. Vier GIS-gebaseerde ruimte-analise metodes is gebruik om voetgangerbotsing-bundels in die studiearea te identifiseer. Hierdie metodes sluit die planêre kerndigtheidsberaming (KDB), die Anselin plaaslike Moran's I, die Getis-Ord Gi * en die Geoptimaliseerde Warmkol Analise (GWA) in. Twee modelleringstegnieke, die Veralgemeende Lineêre Modellering (VLM) en Geografies Geweegde Regressie (GGR) modellering is gebruik om die geboude omgewing en populasie veranderlikes te verbind aan totale, interseksie, en noodlottige en ernstig beseerde (NEB) voetgangerbotsings. Vir hierdie analise is die data gesommeer en geanaliseer op die sensus voorstad vlak. Dit is, onder andere, bevind dat populasie, grondgebruik samestelling, verkeerseine, verkeersirkels/mini-sirkels, industriële gebruik, vier- en multi-been interseksies, en hoë mobiliteitspaaie geassosieer word met groter hoeveelhede voetgangerbotsings. Die studie het ook onthul dat voetgangerbotsings positief verbind is aan sosio-ekonomiese ontneming. Daarbenewens is ruimtelike variasies van die assosiasies in die modelle ondersoek en bespreek. Warmkolle van voetgangerbotsings is meestal in die Suid-Oostelike streke van Kaapstad, wat ook areas is waar ekonomies benadeelde inwoners gekonsentreerd is, geïdentifiseer. Die voorgestelde modelle kan gebruik word om toekomstige voetgangerbotsings te voorspel deur inligting te gebruik wat maklik op die stadsvlak beskikbaar is. Die modelle is ook noodsaaklik vir die beplanning van veilige loop omgewings wat veral in Suid-Afrika en ander ontwikkelende lande benodig word.

Acknowledgement

I wish to thank the following organisations and individuals who were instrumental in the completion of my PhD research:

I would like to express my deepest appreciation to my God and saviour, Jesus Christ for being with me and pouring out His kindness and love on me throughout this journey. His grace and strength has enabled me to embark on and complete this task and all glory, praise and honour belong to Him alone.

Thank you to my study leader, Prof Marion Sinclair, for her guidance, advice and mentorship. Your enthusiasm and continuous support in this project as well as over the years is deeply appreciated.

Thank you to Ilse Combrinck at the Transport and Urban Development Authority (TDA) Cape Town who provided the pedestrian crash data used in this work. I appreciate your friendliness and willingness to assist me.

Thank you to the Strategic Development Information and GIS (SDI & GIS) for providing me with spatial data on land use and transportation systems.

Thank you to Leon Croukamp from Stellenbosch University for his assistance in obtaining GIS software needed for this project.

My deepest appreciation goes to my wife Saajidah Nteziyaremye. Thank you for your words of wisdom, constant prayers, joyful demeanour and unstinting support throughout this venture that has greatly encouraged and emboldened me during times of difficulty. Thank you also for your professional input and intellectually stimulating conversations. You are a blessing to me in more ways than I can count.

Thank you to my family and friends for their emotional support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1 Background

The safety of pedestrians in the road environment has been a global concern since the introduction of motorised travel modes in the late nineteenth century. According to the World Health Organization (WHO), pedestrian deaths account for 22 percent of about 1.2 million road traffic deaths occurring every year worldwide (World Health Organization, 2015). On the African continent, pedestrian deaths represent 39 percent of all traffic-related deaths occurring on the African region (World Health Organization, 2015).

South Africa is ranked among the countries with the highest rate of traffic-related deaths worldwide and with higher proportions of pedestrian deaths. Disproportionate rates of traffic deaths among different road user categories are well documented in South Africa, with pedestrians representing over a third of all traffic deaths according to data published by the Road Traffic Management Corporation (RTMC, 2016, 2017). The majority of vehicle-pedestrian crashes occur when pedestrians are crossing the roadway, and behavioural factors significantly influence the occurrence of vehicle-pedestrian crashes (Albers *et al*., 2010).

A variety of variables affect the incidence and the severity of a pedestrian crash. These factors relate to the behaviour of road users, road geometric characteristics as well as vehicle and environmental factors. Engaging in risky behaviour on the part of both pedestrians and motorists is the major factor contributing to pedestrian crashes in South Africa. Research has shown that about 75 percent of crashes in South Africa are attributed to human factors (Vogel & Bester, 2005) and similar trends may be expected for pedestrian-vehicle crashes. Contributing factors which are consistently reported to be associated with pedestrian crashes are speeding; driver's disregard of pedestrians; jaywalking; illegal crossings on freeways; alcohol use; and poor visibility of pedestrians at night times (Luke & Heyns, 2014; RTMC, 2011; Sinclair & Zuidgeest, 2016). Roadway and environmental factors such lack of pedestrian facilities, inadequate road design and lighting conditions are other factors that may increase the likelihood of pedestrian crashes (Ojungu-Omara & Vanderschuren, 2006; Vogel & Bester, 2005).

Urban environments are often locations of higher crash risk for pedestrians due to the presence of high traffic volumes, higher density of streets and higher pedestrian activity. Recently, the influence of the urban built environment on pedestrian safety has been given attention in road safety research. The aspects of the built environment which are often included in safety analyses are land use patterns, urban design features and elements of transportation systems. Research on the relationships between the attributes of the built environment and the incidence of pedestrian crashes is often coupled with black spot analysis to detect high-risk locations for pedestrians on the transportation system. However, research in this area is still scarce in the developing world.

Active transport modes such as walking and cycling are currently being promoted in many countries for their health and environmental benefits. Promoting these modes is one of the strategies adopted in practice to alleviate negative consequences resulting from motorised traffic, such as traffic crashes, congestion, delays, traffic delays, land consumption, air and noise pollution among others (Ewing *et a*l., 2011; Gomez-Ibanez *et al*., 2009). Research in South Africa demonstrated that more than half of the low- and middle-income households rely on walking to reach school, work, shops and other services (Behrens, 2002). As walking is already an important transport mode in South Africa and as people are currently encouraged to walk more, the effort to reduce pedestrian crashes is crucial to make walking environments safer. One way of achieving this goal, is to rethink how the built environment is planned and designed.

In terms of prevention, road safety investigations play an important role in uncovering a range of factors which may contribute to the high incidence of pedestrian crashes. Road safety analysis is a prerequisite for road safety programs to ensure a safe and efficient transportation system. There is a variety of approaches that are aimed at quantitatively estimating the safety of transportation systems. The most widely used approaches are statistical methods and geospatial analysis methods. The latter approach has become popular owing to recent advances in Geographical Information Systems (GIS). There has been a growing interest among traffic researchers to use GIS as a tool to enhancing visual presentation, helping to discover and visualise spatial patterns and communicating information in more expressive way. Both methods are applied in this study to investigate the associations between the attributes of the built environment and pedestrian crashes in the context of South African urban spaces.

1.2 Problem statement

Road planning practice in South Africa has been concerned with improving the mobility of motorised travel modes with little or no consideration given to the needs and safety of non-motorised travel modes. In addition, the built environment in South African cities differs in many respects from other cities in the world. The form of the built environment in South Africa has been shaped by past policies of racially segregated human settlements and this has inevitably affected travel behaviour as well as the extent to which pedestrian safety was prioritised. Another key challenge in urban planning practice in South Africa has been a lack of coordination between land use and transportation systems. All these factors have led to higher vulnerability of pedestrians marked by a disproportionate share of crash risk particularly in urban environments.

Pedestrian crashes occur as a result of one or more factors, namely human, road environment and vehicle factors. Despite the general recognition that human factors play a significant role as a contributory factor to pedestrian crashes, research has shown that the design of the built environment plays a significant role in influencing road user behaviour, and as such has the potential to impact the incidence and the severity of pedestrian crashes. This topic has attracted the attention of many researchers in the field of road safety in the effort to provide safe walking environments which are given higher priority in the contemporary planning practice. However, research of this nature is dominated by the developed world where travel patterns, motorisation level, infrastructure, availability of funds, policies and other safety-related aspects differ from those in the developing world. Contextual differences often pose limitations when applying safety-related research findings from the developed world and stress the need to conduct research of this nature in the developing world.

In South Africa, a few attempts have been made to investigate the relationships between pedestrian crashes and site-specific elements (e.g. intersections, schools, etc.) or street-scale elements (e.g. crosswalks, sidewalks, intersection design elements etc.). From literature searches, it would seems that research into the associations between pedestrian crashes and the built environment at a macro-scale level is non-existent in South Africa. Therefore, little is known about the influence of the built environment features on the frequency and severity of pedestrian crashes in the context of South Africa.

Traditionally, road safety analyses are carried out by using a variety of analytical methods with statistical methods and geospatial analyses being the most widely applied. Estimating road safety using these methods requires good quality, reliable and accurate information on crash occurrence. It also requires other zonal information on road facilities and traffic operational characteristics, travel behaviour (e.g. vehicle kilometres travelled), exposure variables (e.g. traffic volumes, pedestrian volumes and speed), land use, and population characteristics. In the context of South Africa and other many countries in the developing world, this information is not always available and data which is available is often subjected to a number of data deficiencies. Data on road traffic crashes collected by transportation agencies in South Africa is not geocoded and this deficiency poses serious limitations when applying conventional methods of safety analysis such as statistical methods and GIS-based spatial analyses. This study attempted to overcome these challenges by developing a methodology to improve historical crash data and by gathering additional information necessary for the application of statistical methods and geospatial methods. This allowed the researcher to investigate a number of research questions which are formulated as follows:

1. Is there a measurable link between the built environment and pedestrian crashes?
2. If the link exists, what is its extent?
3. If the link exists, does it vary spatially?
4. What are the characteristics of pedestrian crashes in the study area?
5. Where are pedestrian crashes more likely to occur?
6. Where are hot spots for pedestrian crashes located in the study area?
7. How suitable are the crash analysis methods used in this study to the context of South Africa?
8. Are the findings from this study comparable to those in the international literature?

1.3 Aims and objectives

This study investigates the link between the built environment features and the incidence of pedestrian crashes in the context of the South African urban environment. The specific objectives of this study are:

1. To investigate the extent to which the frequency of pedestrian crashes are associated with the built environment and population characteristics through the use of statistical methods. The specific intentions within this objective are:

   a. To identify the aspects of the built environment that are related to the incidence of pedestrian crashes;
   b. To identify population characteristics that are associated with pedestrian crashes;
   c. To quantify the associations between pedestrian crashes, the built environment and population characteristics;
   d. To investigate spatial variability of associations across the study area
2. To evaluate the performance of statistical methods used in this study to predict the incidence of pedestrian crashes.
3. To identify high-risk locations for pedestrian crashes through the use of geospatial analysis techniques.
4. To evaluate the performance of geospatial analysis methods applied in this study for the identification of hot spots for pedestrian crashes.
5. To describe pedestrian crash profiles and contributing factors. Within this objective, the specific intents are:
   a. To describe pedestrian casualties by demographic characteristics
   b. To analyse pedestrian casualties by location of crash occurrence
   c. To describe pedestrian casualties by injury severity
   d. To assess design deficiencies at intersections where pedestrian crashes took place
   e. To identify pedestrian behavioural aspects and actions contributing to pedestrian crash occurrence.

1.4 Definition of terms

Connectivity of the street network: The directness and availability of alternative routes from one point to another within a street network.

Crash rate: The number of crashes in accordance with a measure of exposure (e.g. population, pedestrian/vehicle volumes, time and distance).

Cul-de-sac: It is a dead-end street or an endpoint of a link that has no other connections.

Exposure: The proximity to potentially harmful situations in traffic environment or a precondition that must be present in order to have a traffic crash.

Fatality: A death resulting from a road traffic crash (usually within a 30 day period after the crash occurrence).

Injury: An injury is defined as physical damage to a human body caused either by sudden transfers of energy exceeding the threshold of physiological tolerance or by the lack of one or more vital elements, such as oxygen.

Intersection: The endpoint of a link that connects to other links.

Intersection-related pedestrian casualty: the term is used in this study to signify any pedestrian casualty that occurred at an at-grade junction of roads.

Land use mix: The relative proximity of different types of land use within a given area.

Land use: The distribution of spatially located activities across a geographic area, including the location and the density of different activities, where activities are grouped into relatively broad categories, such as residential use, commercial use, industrial use, offices, parks, transport facilities, schools, brownfield sites, open spaces, etc.

Link: A road segment between two nodes.

Midblock-related pedestrian casualty: A pedestrian casualty that occurred on a link or a section of the road between two consecutive road junctions.

Pedestrian casualty: In this study, a pedestrian casualty is defined as any pedestrian who was killed or injured, and any person who was involved in a crash with a vehicle but for which the injury severity was not known.

Police reported crash: A crash that is reported to the police and is recorded in the crash database.

Street density: The measure of the length of a road network per unit of area.

The built environment: The built environment in this study consists of land use, urban design, and transportation systems, and encompasses patterns of human activity within the physical environment.

Traffic crash: A term that is sometimes preferred to traffic accident due to the fact that it reflects an element of causality rather than an unavoidable event caused by chance or occurring without human responsibility.

Transportation system: The physical infrastructure of roads, sidewalks, bicycle paths, railroad tracks, bridges as well as the level of transport services provided.

Urban form: The design of the city and the physical elements within it, including both their arrangement and their appearance. The term urban form and urban design are used interchangeably in this study.

Vulnerable road users (VRUs): A group of road users most at risk in traffic in view of their susceptibility to injury in the event of a crash, and generally these include pedestrians, cyclists and motorcyclists.

1.5 Assumptions

The study is based on pedestrian casualty data including people who were killed, injured, those without injuries or those whose injuries were not known but there was evidence that they were involved in car crashes. However, it is recognised in this study that a single pedestrian crash event could result in more than one pedestrian casualties. This implies that that the number of pedestrian casualties does not necessarily reflect the same number of pedestrian crash events. In spite of that, the study considers a pedestrian casualty as a proxy for a pedestrian crash and assumes that the findings obtained by using pedestrian casualty data could be generalised to pedestrian crashes.

The types of land use included in this study are the zonings and subzonings allocated to each property located within the boundaries of the city as approved by the zoning scheme regulations of the City of Cape Town. However, literature searches could not find any study on the extent to which the zoning map of Cape Town is representative of the actual land use development. A quick test on a small sample of properties in the zoning map was carried out in this study to verify whether the permitted land use development matches the actual use. The results of this test confirmed that the zoning map is sufficiently accurate and that it can be used as an acceptable proxy for land use development. Therefore, it is assumed in this study that the permitted land use designated in the zoning maps is the same as the actual land use development.

1.6 Scope of the study

The conceptual framework adopted in this study was inspired and supported by previous researchers who recognise the existence of both the direct and indirect link between the built environment and pedestrian crash occurrence (Cho, Rodríguez & Khattak, 2009; Ewing & Dumbaugh, 2009; Miranda-Moreno, Morency & El-Geneidy, 2011; Ukkusuri, Miranda-Moreno, Ramadurai & Isa-Tavarez, 2012; Wier, Weintraub, Humphreys, Seto & Bhatia, 2009). The concept is illustrated in Figure 1-1. In this conceptualisation, exposure variables which are vehicular speed, traffic volumes and pedestrian activities play a mediating role in the link. Salient elements of the framework that are concerned in this study are the built environment, socio-demographic variables (or population characteristics) and pedestrian crashes. In the context of this study, data on exposure variables (i.e. pedestrian volumes, traffic volumes and vehicular speed) was not available at the census suburb level and could not be included in the study. It follows that the current study is concerned with only the link between the built environment and pedestrian crashes and does not report on the mediating effect of exposure variables.

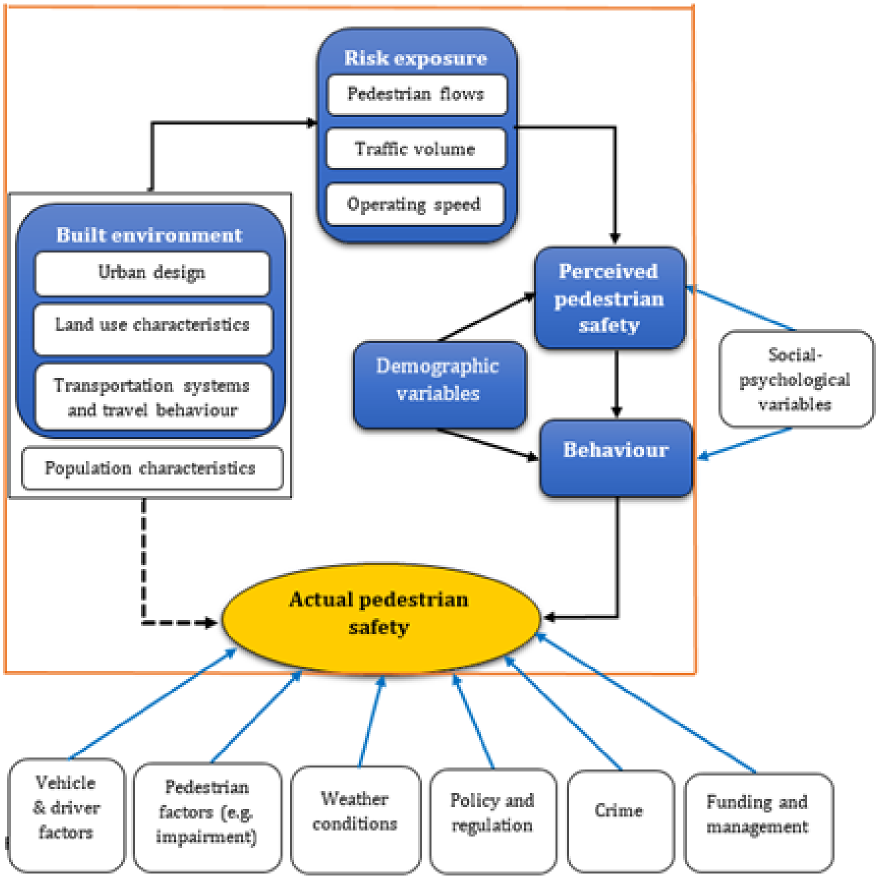


Figure 1-1: Conceptual framework adopted in this study

1.7 Delineations and limitations

Pedestrian safety is affected by a number of factors as portrayed in the conceptual framework illustrated in Figure 1-1. Among factors that influence pedestrian safety, the aspects of the built environment (i.e. land use patterns, urban design features and transportation systems) and population characteristics (i.e. demographic and socio-economic characteristics) are the only aspects of the conceptual framework included in the modelling processes. The analysis of behavioural aspects reported in historical crash records are restricted to descriptive analyses in this study. Pedestrian crash data used in this study was retrieved from police-reported crash data collected in the City of Cape Town. Due to the limited information detail of the crash data used in this study, the study does not examine the relationships between pedestrian crashes and other aspects presented in the conceptual framework such as weather conditions, vehicle factors, driver behaviour, alcohol impairment, crime and so forth. Furthermore, certain geospatial analyses such as Kernel Density Estimation (KDE) were restricted to only intersection-related pedestrian crashes since the description of crash locations in the crash data does not allow proper location identification for non-intersection pedestrian crashes (i.e. crashes occurring on links or road sections between two consecutive intersections).

1.8 Thesis statement

This study is based on the idea that spatially aggregated data on the built environment and population characteristics can be used to predict the incidence of pedestrian crashes. The thesis statement in this study is formulated as follows: “There is a measurable link between the attributes of the built environment and the incidence of pedestrian crashes”.

1.9 Significance of the study

Understanding the influence of the attributes of the built environment on pedestrian safety is of vital importance for both research and practice in the effort to address pedestrian safety problems and develop safer walking environments in urban spaces. From a theoretical perspective, this study adds to the limited existing body of knowledge on the influence of the built environment and population characteristics on pedestrian crashes. To the South African context, this might be the first research attempt to examine the associations between the built environment and pedestrian crash incidence at a metropolitan scale.

From a practical perspective, understanding the influence of the aspects of the built environment on pedestrian crash risk through research is a crucial step towards the development of evidence-based safety interventions and strategic plans appropriate to the South African context. For instance, the objectives of this study are in line with the vision of the National Road Safety Strategy (NRSS) for the 2016-2030 period currently being implemented in South Africa, which is to ensure "safe and secure roads" (Department of Transport, 2014). The expected new insights from this study would support a number of priority areas for interventions identified in the NRSS, including identifying and addressing high risk locations; developing and redefining infrastructure design aimed at protecting vulnerable road users; identifying and addressing shortcomings in road safety data management system; developing comprehensive programmes to improve road user behaviour; and increasing road research relevant to South Africa. A better understanding of the influence of the built environment on pedestrian crash risk can assist with the achievement of the NRSS goals and may guide interventions aiming at addressing pedestrian safety problems and supporting the planning of safe walking environments.

1.10 Thesis approach

This research is an empirical study that uses a mix of analytical methods, including descriptive analysis, inferential analysis, geospatial analyses and modelling techniques. The research questions are of three types: descriptive, causal and predictive. Figure 1-2 illustrates a summary of the research process adopted to achieve the research objectives. The research process illustrated in Figure 1-2 indicates how the research was conducted and outlines the expected outcomes from the study.

The literature review provides a summary of existing works relevant to the research topic and was structured according to emergent themes. A preliminary scan of existing literature influenced the choice of the research topic and the formulation of the research problem. After the development of the proposal, a secondary literature survey was performed to provide a theory base from the previous works that underpin the key subjects of the research topic. The secondary literature review also provided information on how other scholars have approached studies of the similar area of interest to achieve their research objectives.

The study used two types of data - primary data and secondary data - both collected from different sources including local governmental agencies and online open data. The choice of the method used in this study was guided by the existing theory base from previous research. In addition, the choice of the method also depended on the nature of data that was available for the analysis. The research approach illustrated in Figure 1-2 presents research strategies for responding to the research questions and proving the thesis statement.

The study attempted to address data deficiencies by improving data on pedestrian crashes and the built environment. The improved data was used to investigate associations between the built environment and pedestrian crashes and this was performed in four main steps:

- Geocoding and mapping a dataset of pedestrian casualties in ArcMap;
- Aggregating and measuring the attributes of the built environments at the census suburb level;
- Identifying clusters (hot spots and cold spots) of pedestrian casualties using two geospatial methods: local statistics of spatial autocorrelation and the planar kernel density estimation (KDE);
- Using statistical methods to investigate associations between the built environment and pedestrian casualties in the City of Cape Town.

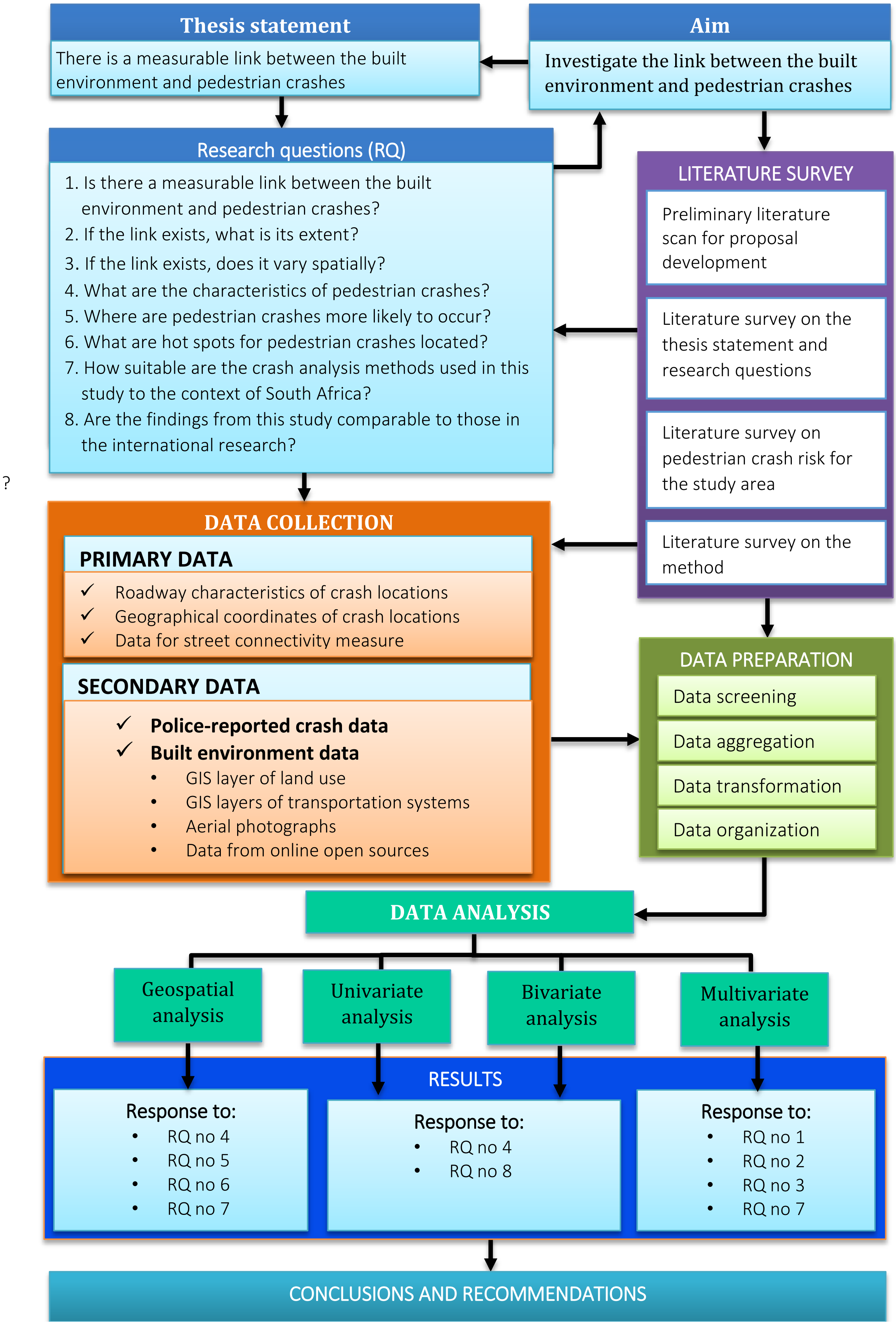


Figure 1-2: Research process

1.11 Chapter overview

Chapter 1 provides the introduction of the study. This chapter outlines the background, the research problem, the aims and objectives, the definitions of terms, the assumptions, the scope of the study, the delineations and limitations of the study, the significance of the study the thesis and the research approach. Chapter 2 provides a theoretical basis of the research problem through a review of existing literature relevant to the research questions investigated in this study. Chapter 3 describes the methodological approach applied to investigate the research questions. Chapter 4 presents and discusses the results of empirical analyses and Chapter 5 presents the conclusions drawn from the results, the original contributions of the study, practical implications of the study, limitations, and considerations for future research.

Chapter 2: Literature review

2.1 Introduction

The literature review chapter provides a review of previous works relevant to the scope of this study. The chapter provides the theoretical background to the research questions stated in the first chapter. The chapter also feeds into the identification of appropriate methods applied in this study to investigate the research questions. The body of the literature is broken down into the following sub-chapters:

1. Pedestrian casualty profile
2. Risk factors of pedestrian crashes
3. The built environment as a risk factor of pedestrian crashes
4. Measures of the attributes of the built environment
5. Associations between the built environment and the incidence of pedestrian crashes in South Africa.
6. Crash modelling techniques and
7. Concluding notes on the literature review.

2.2 Pedestrian casualty profile

2.2.1 Pedestrian casualty profile worldwide

Road traffic injuries are currently regarded as a global public health problem by the World Health Organization (WHO). More than 1.2 million people die each year as a result of road traffic crashes and about 50 million more people sustain non-fatal injuries globally (World Health Organization, 2015). The burden of road traffic injuries is disproportionately distributed across different geographical locations on the globe with low- and middle-income countries being the most affected regions. While the global traffic fatality rate is 17.4 per 100 000 population, the traffic fatality rates in low-and middle-income regions stand at 24.1 and 18.4 deaths per 100 000 population, respectively (World Health Organization, 2015).

In addition to geographical differences, there is also an imbalance in the burden according to the type of road user. The risk of dying as a result of a road traffic crash is higher for vulnerable road users (pedestrians, cyclist and motorcyclists) than for car occupants in all countries. According to WHO (2015), half of road traffic deaths recorded globally are vulnerable road users. The African region has the highest pedestrian fatality rate when compared with other parts of the globe. While pedestrian deaths account for 22 percent globally, the proportion of pedestrians killed in the African region stands at 39 percent (World Health Organization, 2015). The lowest proportion of pedestrian deaths (13 percent) is recorded in the South-East Asia region (see Figure 2-1).

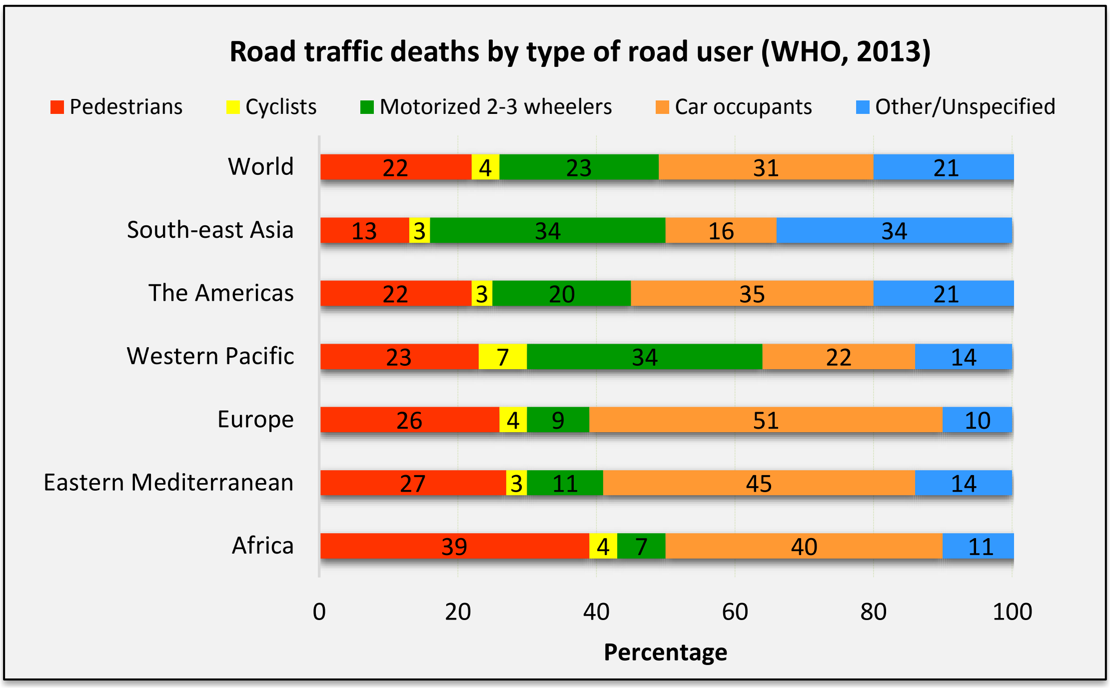


Figure 2-1: Road traffic deaths by type of road user (WH0, 2013)

2.2.2 Pedestrian casualty profile in South Africa

According to many sources, South Africa is ranked among the countries with the highest rate of road traffic deaths in the world. In the study conducted by the World Health Organization (WHO), South Africa was ranked 177th of the 182 countries that participated in the study with a traffic mortality rate of 31.9 per 100 000 population (World Health Organization, 2013). Approximately 14 000 people die every year as a result of traffic crashes (RTMC, 2016). Disproportionate rates of traffic deaths among different road user categories are well documented, with pedestrians representing nearly 40 percent of all traffic deaths in South Africa according to data published by the Road Traffic Management Corporation (RTMC, 2016, 2017). Of these pedestrian deaths, approximately a quarter of them are children and young people below the age of 20 years (RTMC, 2016, 2017). It has been reported that transport-related injuries are among the leading causes of injury and death among children in South Africa. According to data collected by ChildSafe South Africa, road traffic injuries (RTI) are ranked to be the second leading cause of injury among children aged 0 to 12 years, representing 15.7 percent of all child injuries and 25 percent of all hospitalised cases (Herbert et *al*., 2012). In Cape Town, pedestrian crash figures are higher than national figures and pedestrian crashes represent more than 60% of all traffic-related deaths (Liebenberg & Garrod, 2005).

2.3 Risk factors of pedestrian crashes

Research has identified a number of factors influencing pedestrian safety. In general, four main risk factors influence the incidence and severity of pedestrian crashes (Peden *et al*., 2004):

a) The exposure to risk, which is described by the amount of movement within a transportation system, commonly defined as traffic volume, pedestrian volume and speed (Lassarre *et al*., 2007; Ukkusuri *et al*., 2012);
b) The underlying probability of a crash given a particular exposure;
c) The probability of injury; and
d) The outcome or severity of injury.

The main risk factors of pedestrian crashes can be summarised in four categories: (1) factors influencing exposure to risk; (2) factors influencing crash occurrence; (3) factors influencing injury severity; and (4) factors influencing post-crash care (Peden *et al*., 2004). Following this classification, risk factors reviewed in this study are summarised in Figure 2-2. Elements of the built environment identified in Figure 2-2 as risk factors are colour-coded in blue.

Factors influencing exposure to risk
- Motorization level
- Traffic volumes
- Demographic factors
- Socio-ecomomic factors
- Activity factors: e.g. playing, travelling to school
- Rural versus urban environments
- Transport, land use and roadway planning
- Legislation on minimum age for drinking alcohol, driving and employment

Factors influencing crash ocurrence
- Human factors
  - Developmental factors
  - Age and gender
  - Risk-taking behaviour
  - Peer influence
- Behavioural aspects
- Roadway and environmental factors
- Vehicle-related factors

Factors influencing post-crash injury outcome
- Pre-hospital care factors
- Acute care factors
- Rehabilitation

Factors influencing injury severity
- Road class (e.g. Freeways, arterials, access roads)
- Inappropriate and excessive speeds
- Human tolerance to impact

Figure 2-2: Risk factors of pedestrian crashes [adopted from Peden et al. (2004)].

2.3.1 The built environment as a risk factor of pedestrian crashes

This section provides the definition of the built environment and a detailed description of the attributes of the built environment. In addition, the section provides a review of studies that documented the relationships between the attributes of the built environment and the incidence of pedestrian crashes.

2.3.1.1 Definition of the built environment

The built environment is a broad concept that has been described by different authors according to their research purposes. The attempt to describe the built environment has originated from a number of studies that, from around 1990, introduced the framework of dimensions of the built environment while considering the impact of the built environment on travel behaviour and physical activity (Ouyang & Bejleri, 2014). Three dimensions or 3Ds were originally defined by Cervero and Kockelman (1997) to describe the built environment characteristics. These dimensions include density, diversity and design.

A few years later, the concept of 3Ds was extended to 5Ds by including other two dimensions, destination accessibility and distance to transit (Ewing & Cervero, 2010). According to these authors, these dimensions are defined as follows. Density is defined as a measure of variables of interest per unit of area. Diversity is characterised as the number of land use types in a given area and the degree to which they are represented in a spatial unit of analysis (land area, floor are etc.). Design relates to characteristics of street networks in a given area. Destination accessibility is referred to as the measure of ease of accessing destinations. Distance to transit is defined as the shortest distance from residences or places of employment to the nearest transit station or stop. A number of other studies have integrated other dimensions into the concept of the built environment such as demand management (encompassing parking supply and cost) and demographic variables (Ewing & Cervero, 2010).

In this study, the built environment is defined as the physical environment that is human-made or human-altered with the intention to facilitate and enhance opportunities for human activities (Smith & Brooks, 2013). The four main elements of the built environment concerned with this study include density; land use patterns; urban design features; and the transportation system. These are the characteristics of the built environment which have been often included in research which is concerned with travel behaviour and traffic safety (e.g. Handy *et al.*, 2002; TRB, 2005).

1. Density

Density is usually expressed as the number of population, jobs or households or other aspects of interest per unit of area (Ewing & Cervero, 2010; Handy *et al.*, 2002). Population density is the density measure that is often included in safety analyses and is usually expressed in population numbers per hectare of land. Job density is often included in studies that are concerned with travel behaviour and this measure has been adopted in traffic safety research as well (Miranda-Moreno *et al.*, 2011; Quistberg *et al.*, 2015).

2. Land use patterns

Land use patterns relates to the distribution of spatially located activities across a geographic area, including the location and the density of different activities, where activities are grouped into relatively broad categories, such as residential, commercial, office, industrial, parks, transport facilities, schools, brownfield sites, open spaces, etc. (TRB, 2005).

3. Urban form

Urban form is a term that has a broad meaning. Generally, urban form refers to the aesthetic, physical, and functional qualities of the built environment, such as the design, arrangement and appearance of buildings and streetscapes, and relates to both land use patterns and the transportation system (Handy *et al.*, 2002; TRB, 2005). Many research have identified a number of urban form features that are related to non-motorised transport modes, such as street connectivity, accessibility, density and land use mix (Badoe & Miller, 2000; Hess *et al*, 2001; Wedagama *et al.*, 2008).

i. <u>Street connectivity</u>

Connectivity refers to the number of transportation connections (road segments, walking and cycling paths) linking people to their destinations (Marshall, 2005). Connectivity and permeability are often used interchangeably, although a distinction between the two terms is underlined by several researchers. Permeability is defined by the extent to which urban form permits or restricts movement of people or vehicles in different directions(Forsyth & Southworth, 2008; Pafka & Dovey, 2017). Based on this definition, the extent to which an area is permeable is dependent not only on the number or density of connections but also on the capacity of those connections to carry people and vehicles. Consequently, widening roads within a street network of an urban area would make that area more permeable but leaves its connectivity unaltered (Marshall, 2005). Nevertheless, this distinction is often overlooked in research due to difficulties in objectively measuring the permeability of an urban area.

Connectivity is an important factor in urban design practice as it impacts the efficiency of public transport, travel choices, emergency access and the liveability of a community (Chandra & Quadrifoglio, 2013). A highly connected street or path network has many short links, a high density of intersections and a small number of dead-ends (or cul-de-sacs). It is also recognised that a well-connected road system offers more route options and decreases travel times by allowing more direct trips between destinations in a neighbourhood (Victoria Transport Policy Institute, 2012).

ii. Accessibility

Research has adopted the directness of a trip between an origin and a destination as a central aspect of defining and measuring another dimension of urban form called accessibility. Accessibility is defined as the directness of links and availability of alternative routes between origins and destinations within a road network (Handy *et al.*, 2002). From a pedestrian/cyclist point of view, short travel times and more route options are the main influential factors for walking and cycling. This claim makes accessibility an important feature of urban design which is very sensitive to the attractiveness of non-motorised modes.

iii. Land use mix

Increasing transportation-related problems (traffic crashes, congestions, delays, air and noise pollution) have pushed planners and policy makers to consider the design of cities which are accessible, sustainable and conducive to human-scaled transportation modes such as walking and cycling. Land use mix is the forefront of the *New Urbanism* [1] which is concerned with creating human-scale, walkable, functional and sustainable neighbourhoods (Ohm & Sitkowski, 2004). The benefits of land use mix are evaluated through its impact on travel behaviour, environmental, social and economic contexts (Musakwa & van Niekerk, 2012).

Land use mix is defined as the relative proximity of different land use types in a given area or the degree to which different land use types are contained in a given geographic area (Handy *et al.*, 2002). It can be taken as a measure of diversity previously mentioned in the works by

[1] New Urbanism is an urban planning and design approach based on principles of reducing car dependence and creating liveable and walkable neighbourhoods by locating high-density housing, jobs and commercial sites closer to each other (Ohm & Sitkowski, 2004).

Cervero & Kockelman (1997) and Ewing & Cervero (2010). From a transportation perspective, numerous scholars have reported that land use mix has a powerful influence on travel choice behaviour (Bashirul Haque *et al*., 2013; Hannan, 2013; Leck, 2006; Rajamani *et a*l., 2003; Sarkar & Chunchu, 2016; Sarkar & Mallikarjuna, 2013; Zhang *et al.*, 2012). Land use mix shortens trip lengths by locating origins and destinations closer to each other (Ewing *et al.*, 2011). Short distances between origins and destinations have the potential to induce changes in travel behaviour since the use of non-motorised transport modes is most favourable for short-distance trips (Rietveld, 2000). Factors such as the distance from home to a commercial centre play a central role in choosing a transport mode to use to reach the destination in question. In addition, land use mix is thought to improve accessibility and to promote transit use(Ewing & Cervero, 2010).

From economical point of view, land use mix has been proven to reduce private vehicle use, to raise property values and to promote better employment mix and to boost street activity (Matthews & Turnbull, 2007; Yang, Song & Choi, 2016). From an environmental perspective, land use mix helps to reduce land consumption (Gehrke & Clifton, 2015). Furthermore, a reduction in vehicle miles travelled is associated with a decrease in emissions and energy consumption (Liu & Shen, 2011; Steemers, 2003; Zhang *et al.*, 2012). From a social context, neighbourhoods with higher levels of land use mix are deemed to be livelier and have the potential to promote a "sense of place[2]" for local community (Song, Merlin & Rodriguez, 2013) and to support spatial as well as community integration (Musakwa & van Niekerk, 2012).

A review of the literature has shown inconsistency in the definition of the concept of land use mix. Some definitions diverge on certain points such as the number and types of land use required to have a mixed-use development and dimensions of land use spread. According to some literature, a mixed-use development can be achieved by combining at least two main land use types (Aygoren, 2004). In other literature, a mixed-use development is characterised by a combination of three or more functionally and physically integrated revenue-producing uses such as retail, entertainment, office, residential, hotel, cultural or recreational (Rabianski *et al*., 2009; Witherspoon *et al.*, 1976). According to the Institute of Transportation Engineers (ITE), a mixed-use development consists of a combination of at least two land use types for which land use interaction can be achieved by the use of local streets, without the need for using major

[2] A sense of place refers to "lens through which people experience and make meaning of their experiences in and with place" (Adams, 2013) and reflects both place attachment (i.e. a bond between people and a place) and place meaning (i.e. symbolic meanings people attribute to a place) (Kudryavtsev, Stedman & Krasny, 2012).

streets. The land use types may include residential, retail, restaurant, hotel, office and/or entertainment (Institute of Transportation Engineers, 2012, 2017).

In addition to the number and the types of land use, spatial dimensions (i.e. horizontal and vertical) also shape the typology of the concept of land use mix. Vertical land use mix can be achieved by mixing different types of land use in a single vertical building or development when, for instance, at least one floor of a building accommodates different activities allocated for revenue-producing uses (retail, office, recreational etc.) while other floors accommodate residential use (Sarkar *et al*., 2014). The vertical clustering of land use types is often referred to as multiple land use (Rabianski *et al.*, 2009). On the other hand, horizontal land use mix consists of different single-use buildings on adjacent or near-adjacent parcels of land (Sarkar *et al.*, 2014).

This dimension-based typology was extended by Hoppenbrouwer and Louw (2005) who included punctual and temporal dimensions in their definition of land use mix. The authors argued that land use mix can be achieved by integrating two or more distinct land use types within a single point in space which is referred to as shared premises dimension. A simple example of this concept is a combination of housing and employment opportunities in a single building. The shared premises dimension is praised in some literature as the ultimate form of land use mix as two or more activities are accommodated within the walls of a single building (Louw & Vries, 2002). This notion is currently facilitated by technological advances in telecommunication whereby homeworking[3] and remote working[4] are receiving greater recognition as alternative way to commuting to the traditional office. Furthermore, a single space can accommodate various land use types at different times. For example, a school hall can be used as a worship place during evenings or on Sundays, and a conference hall can be transformed into a theatre or cinema during evenings or weekends. This temporal dimension of land use mix is also denoted as sequential use of space (Hoppenbrouwer & Louw, 2005). Sequential land use in one structure is supported by adaptation of a space to host different activities and the nature of human activities concerned.

In addition to the four dimension-based typologies, geographical scales are other aspects of land use mix that gained researchers' attention in the effort to provide the best methodological analyses of land use mix (Gehrke & Clifton, 2015). It should be noted that a larger geographical

[3] Homeworking refers to working from a dwelling located in a residential area (Louw and Vries, 2002).
[4] Remote working refers to working from home, hotel, coffee shop, co-working space etc. (Gerdenitsch, 2017).

area is more likely to encompass a wide variety of land use types than a smaller one. A mixture of land use types can be analysed and measured at different geographical scales, such as metropolitan regions, census aggregations, neighbourhoods, urban blocks, buffers around individual urban form aspects (e.g. roads, households), and so forth. From this perspective, defining and analysing land use mix requires two conceptual considerations: quantity and distance (Gehrke & Clifton, 2015). From this approach stems another notion of land use mix which reflects how land use types or activities within close proximity potentially have an influence over each other across a limited spatial range (Song *et al.*, 2013). It is from this perspective that the concept of land use mix is employed in this study. In summary, land use mix is defined in this study as a balanced blend of land use types (including residential, commercial, industrial, recreational, cultural and institutional uses) co-located in an integral way that allows interaction between activities, supports sustainable development and enhances neighbourhood amenity across a limited spatial range.

4. Transportation systems

The transportation system refers to the physical infrastructure of roads, sidewalks, bicycle paths, railways and so on, and services that provide the spatial links or connectivity among activities (Handy *et al.*, 2002).

2.3.1.2 Influence of the attributes of the built environment on pedestrian crashes

This section provides a review of previous studies that reported on the associations between the attributes of the built environment and pedestrian crashes and the injury severity. The review of existing literature is broken down according to the attributes of the built environment.

1. Influence of land use patterns on pedestrian crashes

While examining the relationships between the built environment and pedestrian crashes, a larger number of studies found that land use influences the frequency of pedestrian crashes and the injury severity resulting from the occurrence of a pedestrian crash. Literature searches have found that the majority of studies that reported on the relationship between land use and pedestrian crashes were undertaken in North America.

A study by Kim *et al.* (2010) examined crash data collected over a 3-year period collected in the city and county of Honolulu, Hawaii. The study applied a binary logistic regression technique to model eight dichotomous dependent variables using data on demographic variables, land use and accessibility, aggregated within uniform grids of 0.1 square miles (or 0.259 square kilometres). Six land use categories included in the analysis were agricultural, business, commercial, high-density residential, low-density residential and military use. Results from the model developed for pedestrian crashes indicated that business and commercial uses were strongly associated with increased numbers of pedestrian crashes. In addition, the model showed a fairly positive relationship between low-density residential and pedestrian crashes.

The influence of land use on pedestrian crashes was reported in a study undertaken by Wier *et al.* (2009) in California, US. The study used the ordinary least squares regression (OLS) method to model pedestrian crashes collected in the City of San Francisco over a 5-year period, using data on population characteristics, exposure variables and the built environment. Data was aggregated at the census tract level and the analysis covered a study area consisting of 176 census tracts. The study included street characteristics and land use characteristics as proxy variables for the built environment. Seven variables describing land use were extracted from zoning district data and these include land area and proportions of commercial use; industrial use; neighbourhood commercial use; residential use; high-density residential use; and a mix of residential and commercial uses (referred to as "residential-neighbourhood commercial use"). The study found that the proportion of land area zoned for commercial use and residential-neighbourhood commercial use were positively associated with pedestrian crashes. Land area had a negative relationship with pedestrian crashes, simply because an increase in land area potentially results in a decrease of population density (Wier *et al.*, 2009).

Ukkusuri *et al.* (2012) carried out a study with one of the purposes being to investigate the extent to which pedestrian safety is related to land use and road design characteristics. The study used data on land use, socio-demographic characteristics, transit supply, road network and travel characteristics as well as pedestrian crash data collected in the City of New York over a 5-year period. The data was aggregated at two different levels – zip code and census tract level. The study included nine variables describing land use which were the proportion of residential, commercial, industrial, office space, retail space, and open space land uses as well as the total number of schools, parks and acres of parks. Separate models were developed for counts of pedestrian crashes and counts of fatal pedestrian crashes using the generalised linear modelling (GLM) technique. Model results for both the total number of pedestrian crashes and the number of fatal pedestrian crashes at the census tract level indicated that tracts with greater intensity of industrial, commercial and open space uses were more likely to experience larger numbers of pedestrian crashes. Moreover, the number of schools in a census tract was found to be positively related to the frequency of both pedestrian crashes and fatal pedestrian crashes. The analysis of elasticities showed that the number of schools had a greater impact on the frequency of pedestrian crashes than other land use variables. However, residential use was found to be associated with fewer pedestrian crashes and fewer pedestrian fatalities. The model results regarding the effect of land use on pedestrian crashes at the zip code level were quite consistent with those at the census tract level, except the fact that the relationships were not statistically significant in most of the cases.

Using crash data collected between 2003 and 2007 in the San Antonio-Bexar County metropolitan region in the United States, Dumbaugh and Li (2010) investigated the influence of the built environment on urban crash incidence. The authors first developed a GIS-based dataset consisting of crash and urban form data. Separate models for crashes involving motorists, pedestrians and cyclists were developed using the negative binomial regression modelling technique. Model results for vehicle-pedestrian crashes indicated that strip commercial use and big-box stores[5] (i.e. megastores) were related to increased numbers of vehicle-pedestrian crashes. According to the model for vehicle-pedestrian crashes, each additional commercial use would result in 3 percent increase in vehicle-pedestrian crashes and each additional big-box store would result in 8.7 percent increase in vehicle-pedestrian crashes. However, negative associations were found between pedestrian-scaled retail uses – commercial or retail use of 20,000 square feet (1858 square metres) or less and having a flour-area ratio of 1.0 or greater – and the number of vehicle-pedestrian crashes (Dumbaugh & Li, 2010).

In another the study undertaken in the US, Zhang *et al.* (2015) investigated associations between the road network structure and non-motorist crashes. Their study applied the geographically weighted regression (GWR) technique to model non-motorist crashes collected in 321 census tracts in Alemada County, California. The models were based on the built environment data including variables describing the road network, land use and transportation system as well as other zonal variables, such as traffic behaviour and demographic variables. Land use characteristics were described using three proxy variables, including the number of commercial properties, the number of housing units and the rate of housing units built before

[5] In the study by Dumbaugh and Li (2010), a big-box store is defined as a retail use having a building area of at least 50,000 square feet (4645 square metres) and with a floor-area ratio (FAR) of 0.4 or less.

1950[6]. The authors found that the number of commercial properties in a census tract was positively related to the frequency of pedestrian and cyclist crashes. Their findings suggest that areas with more commercial properties are more likely to experience greater numbers of pedestrian and cyclist crashes. According to the authors, this finding can be explained by the presence of higher volumes of pedestrians and cyclists in areas with more commercial activities such as shopping, dining and entertainment.

In England, Wedagama *et al.* (2006) analysed pedestrian and cyclist crash data collected between 1998 and 2001 in Newcastle upon Tyne. The authors applied the generalised linear modelling (GLM) to data on non-motorised crashes (as outcome variables), and data on land use, population and intersection facilities (as explanatory variables). The data was aggregated at the Enumeration District as the spatial unit of analysis. The study area was split up into two analysis zones, Newcastle City Centre and Gosforth. Separates models were developed for each analysis zone, road user category (pedestrians and cyclists) and each of four subsamples of casualty data – child/adult by working hours/non-working hours. For the entire study area, model results for the pedestrian casualty subsamples demonstrated that retail[7] and community use were associated with increased number of pedestrian casualties during working hours. For the city centre zone (i.e. Newcastle City Centre), more pronounced positive relationships between retail use and pedestrian casualties were found during non-working hours. The authors argued that the predominance of certain retail activities after working hours, such as bars, public houses and restaurants are the possible explanation of these relationship trends. In addition, the model results showed that industrial use was positively related to pedestrian casualties in the city centre zone while the opposite (i.e. negative relationship) was found in the suburban zone. In this study, residential use was excluded from the analysis as it was found highly correlated with population density (Wedagama *et al.*, 2006).

Again, the same team of researchers carried out a study in the UK to investigate associations between urban land use and injury severity by focusing on three pedestrian categories: children (under 16 years); adults (between 16 and 64 years); and elderly (older than 64 years) (Wedagama et al., 2008). As with their previous study, the authors analysed pedestrian casualty

---

[6] The inclusion of this variable was based on the research evidence that the safety performance of neighbourhoods built before 1950 differs from those built more recently in the US (Marshall & Garrick, 2010).

[7] According to the Office of the Deputy Prime Minister (ODPM), retail includes shops (e.g. shops, boutiques, post offices, travel agencies, filling stations , car dealerships, internet cafes, etc.), financial and professional services (e.g. banks, insurance brokers, betting offices etc.), public houses and bars (e.g. pubs, wine bars, etc.) as well as restaurants and cafes (Office of the Deputy Prime Minister, 2006).

data collected in the city of Newcastle upon Tyne during the period between 1998 and 2001, and applied the generalised lineal modelling technique to fit the data. The developed models include five variables to describe land use patterns: industrial use; offices; retail; community[8] building; and leisure[9] building. The model results indicated that only retail use was significantly associated with increased number of killed and seriously injured (KSI) and slightly injured casualties among adult pedestrians. However, the authors reported the absence of associations between land use and injury severity among child and elderly pedestrians. The models for adult pedestrians indicated that an increase in retail land use by just 1 percent would result in an increase in the number of KSI adult pedestrians by 30 percent over weekdays, and by 50 percent during weekend non-working hours. Moreover, the authors found that an increase in retail intensity by 1 percent would result in slight injuries among adult pedestrians increasing by 40 percent and 30 percent during weekdays and weekend non-working hours, respectively.

2. The influence of land use mix on the incidence of pedestrian crashes

The influence of land use mix on the frequency of pedestrian crashes and injury severity was investigated in a study carried out by Amoh-Gyimah *et al.* (2016). The authors used three modelling techniques, the random parameter negative binomial (RPNB), the non-spatial negative binomial (NB) and the conditional autoregressive model (CAR), to model total, serious injury and minor injury pedestrian crashes. In this study, land use mix was measured by the Balance Index (BAL). For the entire sample of pedestrian crashes, the Balance Index had estimates of 1.17; 0.89; and 1.49 for NB, CAR and RPNB, respectively. For minor injury pedestrian crashes, the estimates of BAL were found to be 1.30; 1.03; and 1.44, for NB, CAR and RPNB, respectively. For serious injury pedestrian crashes, the magnitude of BAL estimates reduced to 0.99; 0.87; and 1.36 for NB, CAR and RPNB, respectively. The interpretation given to partial effect results is that an increase of 1 percent in land use mix would result in an average increase of 11.68; 5.91; and 4.51, for total, minor injury and serious injury pedestrian crashes, respectively (Amoh-Gyimah *et al.*, 2016).

3. Influence of urban design features on pedestrian crashes

The influence of urban design features on the incidence of pedestrian crashes has been confirmed in numerous studies. Associations between pedestrian crashes and the number of

---

[8] Community buildings comprise health, educational, community and religious buildings, police stations, and fire stations (Land use Change in England to 1997: LUCS-14)
[9] Leisure building consists of, for instance, museums, cinemas, theatres, bowling alleys, sport halls, holiday camps etc. (Land use Change in England to 1997: LUCS-14)

intersections aggregated at a uniform gird of 0.1 square miles (i.e. 0.259 square kilometres) was found in study conducted by Kim et al. (2010) carried out in the US. In another US study, the number of intersections per road length was shown to be significantly associated with a reduced risk of sustaining a fatal injury for pedestrians (Mohan *et al*., 2017). The results from this study also demonstrated that the length of the road network was associated with higher pedestrian fatality rates. With regard to street density, Zhang e*t al.*(2015) reported a negative correlation between street density and the number of pedestrian and cyclist crashes in their study conducted in the US. The authors argued that presence of safety countermeasures in densely populated areas are the possible explanation for this finding.

4. Influence of transportation system features on pedestrian crashes

The impact of the urban road network on road safety has been documented in previous research (Moeinaddini *et al.*, 2014; Mohan *et* al., 2017). Some studies have tested the relationship between the urban road network and traffic safety in general. A small number of studies have examined this relationship specifically focusing on associations between elements of the road network structure and pedestrian crashes. Studies that are reviewed in this study fall in the second category.

The structure of road network has been documented in a number of studies for its potential impact on traffic safety. A study by Mohan et al. (2017) used a stratified random sample method on a five-year dataset (between 2005 and 2010) to investigate the influence of the road network structure and intersection density on road traffic fatality rates in 16 American cities. Three major road types were included in the analysis and these are: Primary roads (generally divided, limited-access highways coded as “S1100”); secondary road (main arterial roads coded as “S1200”); and paved non-arterial street, road, or byway with a single lane of traffic in each direction (coded as “S1400”). The number of intersections on different road types was expressed in terms of intersection density (intersections per km of roads, intersections per squared km and the ratio of km of various road types per square km of area). Eight models were developed separately depending on the outcome variable considered in regression models, with one model applied to a dataset of pedestrian fatalities. The pedestrian model includes two explanatory variables describing intersection density: (1) Length (in km) of various road types (“S1100”, “S1200” and “S1400”) per square kilometre of area; and (2) the number of intersections per linear kilometre. The results from the pedestrian fatality model demonstrated that an increase in non-arterial roads (“S1100” and “S1400”) was slightly but nevertheless significantly associated with increased pedestrian fatality rates. Interestingly, the authors found that an increase in both the number of intersections per road and in the length of main arterial roads ("S1200") were associated with reduced pedestrian fatality rates. The authors argued that the latter finding may be justified by the fact that few pedestrians walk on main arterial roads in the United States.

Hanson *et al*. (2013) evaluated injury severity among pedestrians using a case-control methodology on pedestrian crash data collected between 2007 and 2009 in New Jersey, United States. Results from this study demonstrated that pedestrians are more likely to sustain more severe injuries when: (1) crashes occurred on roads with six or more lanes with a median; (2) it was dark (i.e. no street lighting is provided); and (3) sidewalks and buffers between the road and the sidewalk (e.g. the presence of planted areas, bicycle lanes, on-street parking on one or both sides of the street) were not present. In addition, pedestrian crashes on high speed roads were more likely to result in severe injuries, confirming that speed is an important factor determining the severity of crash outcome. The authors also found that elderly pedestrians (aged 65 years and older) are more likely to sustain severe injuries than other age-groups.

The study by Dumbaugh and Li (2010) included two variables describing the road network structure – freeways miles and arterial miles – in the model developed for pedestrian crashes. The model results demonstrated that arterial roads were associated with higher numbers of vehicle-pedestrian crashes. The model estimates indicated that an increase of one mile (1.61km) linear length of arterial roads would contribute to 9.3 percent increase in vehicle-pedestrian crashes. However, freeway facilities were found to be associated with fewer numbers of vehicle-pedestrian crashes in this study.

The influence of street characteristics such as functional road class on the incidence of pedestrian crashes was investigated in the US by Wier *et al*. (2009). The authors used the ordinary least squares regression model on pedestrian crash data collected over a 5-year period and aggregated at the census tract level. Four road classes – residential streets, arterial streets without public transit, arterial streets with public transit, and a combination of freeways and highways – are among variables included in the model to describe the street network. The model results demonstrated that arterial streets without public transit were significantly associated a greater number of pedestrian crashes. However, the authors did not discuss why the absence of transit on arterial streets would increase the likelihood of pedestrian crash incidence.

Ukkusuri *et al*. (2012) reported on the relationship between the road width and pedestrian crash risk. The authors found that the proportion of wide roads (width larger than 50 feet) was positively related to pedestrian crash frequency. The authors also found that pedestrian crash risk was greater on roads with more than four lanes. With regards to road class, the authors found that local roads were associated with a reduced pedestrian crash risk while primary roads without access restriction increased the likelihood of pedestrian crash occurrence. Similar findings on the influence of functional road class emerged in the study by Mohan et al. (2017). The authors reported that the likelihood for pedestrians to sustain a fatal injury increases with higher proportions of higher road classes (main arterials and highways).

In Greece, Papadimitriou (2016) used a field survey to develop an integrated methodology for the analysis of pedestrian behaviour and exposure in urban areas. Among the objectives of the study, two were (1) to identify and to quantify the combined effect of road, traffic and human factors on pedestrian behaviour and exposure; and (2) to establish the link between pedestrian behaviour and exposure in the light of the effects of road, traffic and human factors. The results of this study demonstrated that pedestrian behaviour and exposure were significantly affected by road type, traffic volume and pedestrian risk-taking. Increased risk exposure was found on principal urban arterials – where risk-taking behaviour is low but the associated exposure is very high – and minor arterials – where risk-taking behaviour is more frequent and the associated exposure is still high (Papadimitriou, 2016).

Zhang *et al*. (2015) applied the geographically weighted regression (GWR) technique to examine the associations between the road network structure and traffic crashes affecting non-motorized modes (walking and cycling). The road network structure was described using three measures: “Average geodesic distance”[10]; “network betweenness centrality”[11]; and “overall clustering coefficient”[12]. Further information on these three topological measures of the road network can be found in Zhang *et al*. (2011); Crucitti *et al*. (2006); Hanneman & Riddle (2005); and Zhang *et al*. (2012). With respect to the three structural measures of the road network, the results showed that road networks with higher values of average geodesic distance, higher

[10] Average geodesic distance shows how far each road is from other roads and is defined as the number of links in the shortest possible route from one node to another (Zhang *et al.*, 2015).
[11] Network betweenness centrality reflects how much a network is centred on some individual streets and is defined as the frequency with which a point falls between pairs of other points on the shortest paths connecting them (Zhang *et al.*, 2015).
[12] The overall clustering coefficient indicates the tendency for a road network to be centred toward local subnetworks or how a single node is close to its neighbouring nodes in a subnetwork of streets (Zhang *et al.*, 2015).

betweenness centrality, and greater clustering coefficients were associated with lower non-motorist crash frequency. According to the authors, these findings suggest that road networks that are more indirectly connected, more centred and with a greater number of sub-clusters (i.e. with some roads highly clustered in several subnetworks as part of the network) may contribute to a safer environment for pedestrians and cyclists. Based on their findings, the authors inferred that the cul-de sac road network (i.e. characterised by higher geodesic distance, higher betweenness centrality and a higher clustering coefficient) may be associated with lower pedestrian and cyclist crash frequencies.

Several other studies have documented the impact of street network on pedestrian safety. Gårder (2004) used state-wide pedestrian crashes collected in United States and reported that the vast majority (71 percent) of pedestrian crashes occurred on straight sections of the road network with adequate sight distances. Curves were locations of only 4 percent of pedestrian crashes, whereas straight roads with a grade accounted for 19 percent of pedestrian crashes. Pedestrians were reported to face a greater crash risk on wider streets – streets with more than 2 lanes. In addition, the author found that 21 percent of fatal pedestrian crashes were reported on local streets, 23 percent on collectors and 56 percent on arterial roads. Arterial roads and major collectors alone were locations for 75 percent of all fatal pedestrian crashes. Furthermore, the author reported higher pedestrian crash risk at locations with higher vehicular speeds (average speed greater than 32 km/h) and strong associations between crash severity and speed.

Other aspects of the transportation system that have been documented in many studies include intersection geometry and type of traffic control. Ukkusuri *et al*. (2012) observed that all-way-stop and three way intersections were associated with a reduced pedestrian crash risk. In addition, the authors reported that the likelihood of a pedestrian crash occurrence increased with the presence of intersections with four and five approaches. The authors argued that the latter finding can be attributed to high levels of pedestrian activity as well as higher traffic volumes and vehicular speeds at four- and multi-legged intersections compared with those at intersections controlled by the all-way-stop sign and those at three-legged intersections.

Zhang *et al*. (2015) used census tract data to investigate associations between road crashes affecting non-motorised modes and zonal factors including intersection configuration types, among others. The model results demonstrated positive associations between the number of four-legged intersections and the crash rate. This finding implies that areas with more four-legged intersections are likely to experience a greater number of pedestrian and cyclist crashes (Zhang *et al.*, 2015). The authors provided two reasons that may account for this finding. The first possible reason was a greater number of traffic conflicts between motorists and non-motorized modes compared with the three-legged intersection type. The second reason was the fact that a greater number of four-legged intersections are found in downtowns which are characterised by higher levels of pedestrian and cyclist activity (i.e. higher pedestrian and cyclist volumes).

In the study by Dumbaugh and Li (2010), four-legged intersections were shown to be associated with increased numbers of pedestrian crashes. The authors reported that an increase of one intersection of this type would result in 0.9 percent increase in the number of pedestrian crashes. Contrary to this finding, the authors revealed that the three-leg intersection type was associated with fewer pedestrian crashes.

Gårder (2004) used pedestrian crash data collected in the US and analysed crash frequency according to intersection configuration type. The author found that pedestrian crashes were most frequent (19 percent) at three-legged intersections compared to other type of intersection geometry. The four-legged intersection type accounted for 17 percent of pedestrian crashes and driveways were locations for only 5 percent of pedestrian crashes (Gårder, 2004). The author also reported that marked crosswalks controlled by traffic signals were associated with a greater pedestrian crash risk than uncontrolled crosswalks, suggesting that pedestrian crash risk is greater at signalised crossing locations.

A number of accessibility variables such as the number of bus stops have been reported in several studies as being associated with a greater pedestrian crash risk. As an example, a study carried out by Ukkusuri *et al*. (2012) reported that the increased number of subway stations and the proportion of commuters who use active modes increase the likelihood of pedestrian crash occurrence. Similar findings were reported by Kim *et al*. (2010) who observed that the number of bus stops was associated with increased pedestrian crash risk. One variable describing accessibility, the number of bus lines in the census tract, was included in the study by Zhang *et al*. (2015). The authors found this variable was significantly associated with a greater number of pedestrian and cyclist crashes. Two possible reasons were provided to explain this finding: The presence of higher volumes of pedestrians and cyclists attracted by transit systems and the presence of facilities for non-motorised modes in the vicinity of bus stops (Zhang *et al.*, 2015).

2.3.2 Factors influencing pedestrian crash occurrence

2.3.2.1 Behavioural aspects

Researchers have long known that the most significant factor exacerbating pedestrian-vehicle crash incidence is risky behaviour on the part of pedestrians, of which can include failure to adhere to the right of way of vehicles, contravening traffic signals, failure to use designated crosswalks and pedestrian facilities, running into the roadway, entering the roadway parked vehicles, alcohol intoxication and travelling in the direction of traffic rather than against it (Baltes, 1998; Hunter *et al*., 1996; Stutts *et al*., 1996).

In South Africa, behavioural aspects were reported to have significantly influence the occurrence of pedestrian crashes (Albers *et al.*, 2010). Human factors are the main contributory factors of road crashes in South Africa, leading to more than 75% of all traffic crashes (Ojungu-Omara & Vanderschuren, 2006; Vogel & Bester, 2005). With respect to pedestrian safety, jaywalking is the most risky behaviour, contributing nearly to half of deaths among pedestrians (Ojungu-Omara & Vanderschuren, 2006).

Pedestrians are naturally not the only party who may be at fault or cause traffic crashes. With respect to driver contributing factors, failure to yield to pedestrians, distractions, reckless driving and intoxication are common factors on the part of drivers that contribute to pedestrian-vehicle crashes (Hunter *et al.*, 1996). As an example, drivers' failure to yield the right-of-way to pedestrians was reported to be the predominant contributing factor of pedestrian crashes in the US, and this behavioural pattern was often linked with speeding (Hunter *et al.*, 1996).

1. **Age-related factors**

The age of a pedestrian has a significant effect on crash risk and injury. In the US, an examination of crash data collected in 2010 revealed that the highest fatality rate was among pedestrians aged 25-44 years old (NHTSA, 2012). A third of all pedestrian fatalities affected pedestrians in this age group. In this study, children younger than 15 years old and elderly pedestrians (65 years and older) were found to be at a greater risk of sustaining life-threatening injuries in the event of a crash. While a high percentage of crashes (29.9 percent) involved children, a smaller number of these crashes (10.4 percent) represented the number of fatalities among pedestrians in this group.

A study into crash occurrence and injury severity among children and elderly pedestrians demonstrated that child pedestrians are also at a greater risk of experiencing severe physical trauma in crashes (Kröyer, 2015). Unlike elderly pedestrians, children demonstrated greater resilience in recovering from injuries and therefore they have a greater chance of surviving crashes than elderly pedestrians (Kröyer, 2015).

In South Africa, Mabunda *et al*. (2008) analysed a database of 7 433 pedestrian deaths that occurred between 2001 and 2004 in four South African cities. The authors found that the average age of pedestrians killed in traffic crashes was 33 years old and almost 50 percent of all pedestrian fatalities affected young adults aged 20-39 years old. In the study by Matzopoulos (2004), the age group from 0 to 14 years old was the most involved in pedestrian fatalities. The number of traffic fatalities observed in the 0-14 age group represented 65 percent of all traffic-related fatalities analysed in this study. Another South African study conducted in 2001 by the National Injury Mortality Surveillance System (NIMSS) reported that pedestrian fatalities among children and adolescent (i.e. ages between 0 and 19 years old) represented 60 percent of all road traffic crashes recorded in Cape Town (Prinsloo, 2001). Consistent results were also found in another South African study by du Toit and van As (2001). The authors reported that that traffic fatalities among child pedestrians under 8 years of age represented nearly half (49 percent) of all traffic fatalities.

2. Gender-related factors

In general, trends for fatal pedestrian crashes by gender show that males are always overrepresented in pedestrian crashes. In the US, Hunter et *al*. (1996) reported that the ratio of male to female fatalities varied from 3.6 to 1 in the 21 to 24 age group, and from 1.3 to 1 in the oldest age group. However, the trends for non-fatal pedestrian crashes by gender were somewhat different as the dominance of males was not seen in every age category. Females were overrepresented in the 21-24 and 65-74 age groups, with male-to-female ratios ranging from 0.6 to 1 and 0.8 to 1 in respective age groups. The authors argued that the discrepancy in the male-to-female ratio could be attributed to fundamental differences between the behaviour of males and females in different age groups. The study concluded that males were overrepresented in pedestrian crashes, and the crash risk among males was greater in fatal crashes than in non-fatal accidents.

Gårder (2004) analysed fatal pedestrian crashes for the 1994-1998 period in the State of Maine, US. The author reported that females represented 37 percent of the fatal injured pedestrians while males represented 63 percent.

In South Africa, male pedestrians were found overrepresented in fatal crashes recorded between 2001 and 2004, accounting for 76 percent of total fatal crashes (Mabunda *et al.*, 2008). A male-to-female ratio of 3.3 was found in this study, suggesting that the likelihood of sustaining a fatal crash is more than 3 times higher for male pedestrians than female pedestrians. The highest level of gender imbalance in fatal pedestrian crashes was found in the 20-39 age group, with male-to-female ratio being 4.6 (Mabunda *et al*., 2008). Few other South African researchers also confirmed gender differences in pedestrian crashes. The analysis of pedestrian crashes in these studies indicated that young males were at the highest crash risk compared to other pedestrian categories (MacKenzie *et al*., 2008).

2.3.2.2 Alcohol related factors

Alcohol is a psychoactive drug, usually ingested in a drink in the form of ethanol or ethyl alcohol (Shinar, 2007). The concentration of alcohol in the blood is expressed by means of a standard measure, the Blood Alcohol Concentration (BAC). As an example, a BAC of 100% is equivalent to a concentration of 1 gram (1000 milligram) of alcohol per 1 millilitre of blood. Thus 5 milligram of alcohol per 1 millilitre of blood would yield a BAC equivalent to 0.5 %. While measuring the BAC, the road user is asked to blow into a portable breather tester to analyse their lung air. According to Vanlaar (2005), the breath alcohol concentration is proportional to the BAC by a factor of 2.27. Thus, as an example, a breath alcohol concentration of 0.44 mg alcohol per litre of exhaled air is equivalent to 1mg/ml in the blood, or 0.10% BAC.

The level of alcohol impairment is directly related to the amount of alcohol consumed. However, with regards to sensitivity to alcohol, gender differences may take place. A higher BAC will be produced in a female than a male of equal weight after the consumption of the same amount of alcohol (Shinar, 2007). This is because the alcohol impairment is a function of its dilution in the blood, and water constitutes 58 percent of an average man's weight, whereas it is only 49 percent of women's weight (Shinar, 2007).

Intoxicated pedestrians are more likely to be involved in alcohol-related crashes as their impairment affects their ability to judge distances and vehicular speeds especially in darkness, resulting in longer perception-reaction times and poor decision-making (Dultz & Frangos, 2013). For instance, Oxley *et al*.(2006) carried out an experimental study and reported that alcohol intoxication affects pedestrian's crossing behaviour by impairing their ability to select safe gaps in traffic. In United States, an analysis of pedestrian crash data revealed that alcohol impairment was reported in 17 percent of fatal pedestrian crashes (Gårder, 2004).

In South Africa, 6 billion litres of alcohol beverages are consumed every year (Meel, 2007) and the estimation of adult per capita consumption of absolute alcohol is between 9 and 10 litres per year, placing South Africa amongst the higher alcohol consumption nations (Parry & Bennetts, 1998). The social costs of alcohol-related trauma and traffic crashes in South Africa far exceed the revenue collected (Meel, 2007). Thus, alcohol misuse and abuse is a major burden on South African society and has a great impact on the incidence of road traffic crashes.

Alcohol impairment has been reported to be a contributing factor in 76 percent of all deaths after interpersonal violence in South Africa (Van der Spuy, 2000). In Cape Town, alcohol was reported as a leading cause of traffic fatalities among pedestrians and was found to be a contributing factor in 61 percent of all pedestrian fatalities. Of these alcohol-related fatalities, more than half (59.5 percent) of the examined victims had BACs at or above 0.08% (Van der Spuy, 1991). In another South African study, an examination of blood samples collected in 2003 by the National Injury Mortality Surveillance System (NIMSS) indicated that 53 percent of traffic fatalities tested for alcohol had positive BACs (Matzopoulos, 2004). Of these alcohol-related fatalities, pedestrians were the most impaired road users representing 61 percent. A high prevalence of alcohol use among pedestrians is a concern in South Africa. Moreover, Mabunda *et al*. (2008) analysed pedestrian fatalities for the 2001-2004 period and found that more than half (58 percent) of the 4 004 individuals tested were positive for alcohol.

Significant gender differences in the distribution of alcohol-related crashes have been reported in a number of studies. Male pedestrians are generally overrepresented in alcohol-related crashes. For example, the proportion of male pedestrians involved in alcohol-related crashes was found to be 76.7 percent while that of female pedestrians stood at 23.3 percent in South Africa (Peden et *al*., 1996). In the study by Mabunda *et al*. (2008), higher levels of alcohol concentrations were found among male pedestrian fatalities (mean BAC of 0.22 g/100 ml) than female pedestrian fatalities (mean BAC of 0.21 g/100ml). There were also more male fatalities (62.3 percent) tested positive for alcohol than female fatalities (42 percent).

Age is also one of the contributory factors of alcohol-related crashes involving pedestrians. Several studies have identified pedestrian age groups that are prone to alcohol impairment. In South Australia, Hutchinson *et al*. (2009) reported that 71 percent of pedestrians involved in alcohol-related crashes were in the 20-49 age group. In the United States, alcohol-related pedestrian crashes were observed predominantly in the 21-45 age group, whereas crashes involving sober pedestrians peaked among pedestrians younger than 18 years old and those older than 55 years old (Jehle & Cottington, 1988; Wilson *et al*., 2003). In South Africa, the study by Mabunda *et al*. (2008) who used crash data from the NIMSS for the 2002-2004 period revealed that the majority of pedestrian fatalities tested positive for alcohol were in the 20-44 age group and the 45-and-older age group – 60.9 percent of tested cases in the 20-44 age group were found positive for alcohol and 53.6 percent of tested cases in the 45-and-older age group were found positive for alcohol (Mabunda *et al*., 2008). Contrary to these findings, Matzopoulos (2004) who used crash data collected in Cape Town did not found a significant difference in age distribution by alcohol involvement.

2.3.2.3 Vehicle-related factors

In South Africa, vehicles factors contribute in about 10 percent of total crashes and fatalities or in December 2002 in South Africa (NDoT, 2003 cited in Ojungu-Omara & Vanderschuren, 2006; Vogel & Bester, 2005). The major contributing vehicle factors are tyre bursts, brakes and light (NDoT, 2003 cited in Ojungu-Omara & Vanderschuren, 2006). According to statistics published by the Road Traffic Management Corporation (RTMC) in South Africa, vehicle factors contributed to 14.1 percent of all fatal crashes recorded between 2004 and 2014 (Department of Transport, 2014). The literature search could not find a South African study that reported on the contribution of vehicle factors to pedestrian crashes. In the United States, vehicle factors contributed to 12 percent of pedestrian crashes, where extended mirrors, defective brakes, foggy/dirty windshield, defective tires, defective lights and oversized vehicle/load were the main predominant factors contributing to pedestrian crashes (Hunter *et al*., 1996). In another US study, trucks and vans were reported to be the types of vehicles that killed and injured a greater number of pedestrians (Gårder, 2004). While these vehicle types consisted of less than a third of all vehicles during the analysis period, they accounted for 46 percent and 36 percent of the vehicles involved in fatal pedestrian crashes and in non-fatal pedestrian crashes, respectively (Gårder, 2004).

2.3.2.4 Roadway and environment factors

The design of roadway and the pedestrian environment has an effect on pedestrian safety and pedestrian behaviour. Good roadway design holds considerable promise in enhancing pedestrian safety. The presence, width, spacing and quality of sidewalks and pedestrian crosswalk, signal timings, and distance between crosswalks all affect the actions, behaviour and safety of pedestrians. A good pedestrian environment reduces the conflict between motorists and pedestrians, promotes and reinforces pedestrian behaviour and reduces the opportunity for pedestrians to resort to unsafe behaviour.

Urban road structure plays a significant role in pedestrian walking and crossing decisions, pedestrian compliance with traffic rules and the associated safety implications (Papadimitriou, 2016). Several studies have sought to understand pedestrian walking and crossing choices in the urban environment by using three levels of behavioural analysis; strategic level, tactical level and operational level (Hoogendoorn & Bovy, 2004; Ishaque & Noland, 2008; Papadimitriou *et al*., 2009). According to these studies, pedestrian choices at the strategic level include for instance elaborating a list of activities to perform and deciding on departure time. At the tactical level, pedestrians make a number of off-road decisions such as scheduling activities, choosing activity areas and selecting routes to take to reach the selected activity places. Along the trip however, pedestrians can make on-road decisions to respond to unpredicted conditions met during the trip. Such conditions may be adverse weather conditions, external factors (e.g. presence of obstacles, stimulation of the walking environment) and internal or personal factors such as time-pressure, attitudes of pedestrians, etc. (Hoogendoorn & Bovy, 2004). At the operational level, pedestrians make instantaneous decisions regarding walking tasks such as adjusting walking speeds, avoiding obstacles, crossing the road, interacting with other pedestrians or other road users and so forth (Papadimitriou et al., 2009).

The three-level approach of pedestrian behavioural analysis is illustrated in Figure 2-3. More information on the interdependences among the three pedestrian behavioural levels depicted in Figure 2-3 is provided in the studies by Ishaque & Noland (2008) and by Papadimitriou *et al*. (2009). Some features of urban road structure have the potential to influence pedestrian choices made at the three behavioural levels.

Much research has investigated the influence of design features of the road networks on pedestrian choices (e.g. mode choice, route choice) and behaviour. Route directness has been reported in many studies as an important design feature influencing pedestrian route choice

(Stangl, 2012). Route directness pertains both the length of the route and the ease to reach their destinations with the smallest amount of effort possible – mostly expressed in terms of the number of direction changes encountered along the way (Hoogendoorn & Bovy, 2004; Stangl, 2012; Venter *et al*., 2014). Pedestrian are more inclined to choose the shortest route and this has been confirmed in a big number of studies (Behrens, 2010; Cantillo *et al*., 2015; Jamil *et al*., 2015; Sinclair & Zuidgeest, 2016).

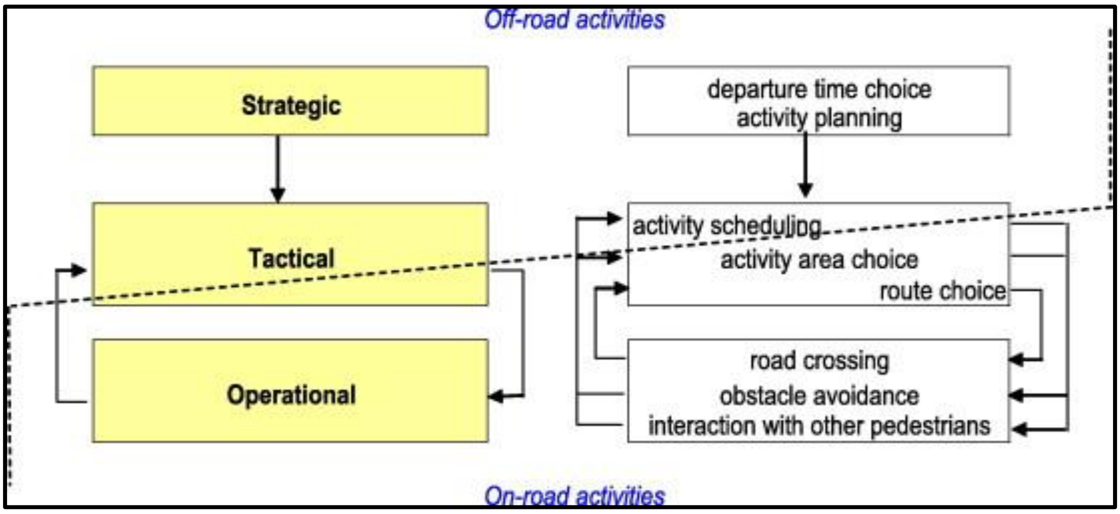


Figure 2-3: The three-level approach of pedestrian behavioural analysis (Papadimitriou *et al.*, 2009)

Other design features which play an important role in influencing pedestrian route choice behaviour include attractiveness of the walking environment, expected amount of interactions with other road users, crossing distances, pedestrian facilities, pedestrian movement network, transport services available along the route, to name a few (Hodgson *et al.*, 2004). In South Africa, few studies have highlighted conditions on different functional road classes that influence pedestrian route choice and crossing behaviour (Behrens, 2010; Sinclair & Zuidgeest, 2016).

With regard to environmental factors affecting pedestrian safety, Hunter *et al*. (1996) showed that poor visibility was the major contributing factor to pedestrian crashes in the United States. Other researchers also showed that poor luminous intensity is the major contributory factor of many pedestrian crashes at night time (Elvik, 1995; Plainis *et al*., 2006).

Road safety research in South Africa has shown that the roadway and environmental factors contribute to approximately 15 percent of road crashes in South Africa (Ojungu-Omara & Vanderschuren, 2006; Vogel & Bester, 2005). In the category of roadway and environmental factors, poor visibility, slippery road and poor road pavement are deemed to cause a significant number of road traffic crashes (Ojungu-Omara & Vanderschuren, 2006). With regard to pedestrian safety, poor visibility has been reported in many studies to be the main contributing factor to a significant number of pedestrian crashes, and this explains why higher pedestrian crash rates are observed during night times. For example, Mabunda et *al.* (2008) reported that over 45 percent of pedestrian fatalities occurred during hours of darkness (between 18:00 and 24:00) with the highest incidence reported between 18:00 and 21:00. The same study also found a link between the time of fatal pedestrian crash occurrence and certain age groups of pedestrians. Pedestrian fatalities among children and adolescents peaked in the late afternoon between 16:00 and 19:00 whereas fatalities among young adult pedestrians (20-39 years age group) peaked between 18:00 and 21:00. Most of the female and elderly pedestrian fatalities occurred between 18:00 and 23:00, but another significant number of fatalities among these pedestrian categories (female and elderly pedestrians) occurred in the morning between 06:00 and midday (Mabunda *et al*., 2008).

Further evidence of the implications of poor visibility on pedestrian safety was reported by Liebenberg & Garrod (2005). These authors reported that approximately 60 percent of pedestrian crashes recorded in 2002 in Cape Town occurred under conditions of poor visibility, such as night times, or dawn/dusk conditions. Pedestrian fatalities that occurred in dawn and dusk conditions accounted for 21 percent and those that took place during night-time conditions accounted for 43 percent of all pedestrian fatalities recorded in Cape Town (Liebenberg & Garrod, 2005).

The influence of road infrastructure features on pedestrian crash occurrence were documented in several South African studies. Ribbens (1996) reported that more than half of pedestrian crashes occurred when pedestrians were crossing at a non-designated crossing point. An early study conducted in the 1980s also reported an increasing incidence of pedestrian crashes at signalised intersections (located in the central business district, on main arterials, and in suburban shopping areas) compared to that at uncontrolled intersections (Ribbens, 1985). The author identified further causal factors of pedestrian crashes, including non-compliance with traffic signals by both motorists and pedestrians, pedestrians walking/running into vehicles, turning vehicle-pedestrian conflicts, and visibility problems.

In a more recent study, Nteziyaremye and Sinclair (2013) investigated pedestrian behaviours at various pedestrian crossing facilities in the Western Cape. This study observed pedestrian crossing behaviour including walking speeds, gaze behaviour, delays, pedestrian-vehicle conflicts, and temporal and spatial compliant behaviour. Difficulty in negotiating crossing facilities was measured by walking pedestrian speeds, the number and the type of evasive actions, waiting times at the kerbsides and in the middle of the roadway, the number of head movements performed in visual search and unsafe crossing behaviour. Signalised intersections and mid-block crosswalks on four-lane undivided roads emerged to be the facilities with the highest levels of discomfort and safety risk for pedestrians. In this study, crossing against the red man signal was observed in the range from 82 percent to 87 percent of the observed crossing events at signalised pedestrian crossings (Nteziyaremye and Sinclair, 2013).

2.3.3 Factors influencing pedestrian exposure to risk

2.3.3.1 Definitions and measures of exposure

Pedestrian exposure is defined as a rate of pedestrian contact with motorised traffic that can create opportunity for road traffic crashes (Greene-Roesel *et al*., 2007; Lassarre *et al.*, 2007). Researchers in the field of traffic safety have applied various metrics to estimate pedestrian exposure to crash risk. These metrics can be grouped into four main categories: area-based measures; trip-based measures; point-based measures; and activity-based measures (Lam *et al*., 2014; Lam *at al*., 2013; Yao *et al*., 2015). Data on these measures can be collected in two ways. The first approach is to gather data while trips are in progress by the use of traffic counters or video observations while the second approach involves collecting data after trips are completed by means of surveys and interviews (Wolfe, 1982).

Area-based methods involve the use of variables such as population size or population density, the number of registered vehicles or the number of licenced drivers within a unit of area (e.g. census tracts, TAZs or buffer zones) as proxy measures of pedestrian exposure (Van den Bossche *et al*., 2005; Chakravarthy *et al*., 2010; Cottrill & Thakuriah, 2010; Wier *et al*., 2009). An advantage of these methods is that they make use of data that is readily available from census or travel surveys (Lam *et al*., 2014). However, these measures are criticised for obscuring the intensity of pedestrian activities within a geographic unit and for being subjected to potential bias commonly referred to as the 'issue of the Modifiable Areal Unit Problem'(MAUP) (Lam *et al*., 2014). The MAUP refers to the situation in which statistical inferences and interpretations in spatial analyses are influenced by both the shape and scale of spatial units used to aggregate data (Wong, 2009; Xu, Huang & Dong, 2018).

Alternatively, pedestrian exposure is quantified by using trip-based measures which take account of the distance travelled, time spent travelling or the number of trips made (Van den Bossche *et al*., 2005; Jonah & Engel, 1983). The most widely used measure of this category is vehicle kilometres travelled (VKT) or vehicle miles travelled (VMT) to quantify distance travelled by motor vehicles on the road network during a given period of time (Abbas, 2004; Van den Bossche *et al*., 2005; Hakim *et al*., 1991; Pei *et al*., 2011).

In some studies, the total distance travelled by motor vehicles is estimated based on fuel and energy efficiency data (Blum & Gaudry, 2000; Fournier & Simard, 2000; Fridstrøm *et al*., 1995; Jaeger & Lassarre, 2000; Tegnér *et al*., 2000). In South Africa, data on monthly fuel sales collected at the province level was used as a proxy of travel-related exposure in a study by Sukhai *et al*. (2011). In a similar way, other studies have attempted to quantify pedestrian exposure by the use of pedestrian kilometres travelled (PKT) (e.g. Lam *et al*., 2013). However, a couple of shortcomings associated with the use of trip-based exposure measures have been pointed out in literature. These measures require data which is often not easy to obtain, and when the data is available, it is often kept in a format that cannot be used in whatever type of analysis, or data can be less relevant depending on the scope of analysis (Van den Bossche *et al*., 2005). Furthermore, trip-based exposure measures examine a single trip type at a time and fail to consider trip chaining effects and collecting data on large study areas involves high costs (Lam *et al*., 2014).

Point-based exposure measures use volumes usually collected by counting the number of vehicles or pedestrians passing through a designated measurement point during a given observation period (Davis & Braaksma, 1988). Similar to the trip-based exposure measures, problems arise when applying this method to a large study area. Traffic volumes can be easily collected on a section of a road but counts for regional and local roads are rarely available (Van den Bossche *et al*., 2005). Specifically for pedestrians, the method can be suitable while collecting pedestrian volumes at specific areas of interest such as intersections, but the method is not easily adopted to collect pedestrian volumes on the whole road network or links (Lam *et al*., 2014).

Recently, several researchers proposed activity-based approaches in the effort to address the shortcomings associated with the previous exposure metrics (Lam *et al.*, 2014, 2013). The activity-based measures use travel diary data to collect individuals' travel behaviour in the context of time geography (Lam *et al.*, 2014, 2013). Two activity-based methods have been used in research and these are the space-time path (STP) and potential path tree (PPT), both developed by Lam *et al*. (2014). More detail on the activity-based approaches can be found in Lam *et al*. (2013) and in Lam *et al*. (2014).

2.3.3.2 Influence of traffic volume on pedestrian safety

Traffic volume is an important exposure risk factor for pedestrian crashes. Research in traffic safety has demonstrated a correlation between pedestrian crashes and traffic volume. In the study by Wier *et al*. (2009) carried out in the US, traffic volume was shown to be the variable with the strongest correlation with pedestrian crashes. Again in the US, Zhang *et al*. (2015) reported a negative association between vehicle miles travelled (VMT) and pedestrian and cyclist crash rates in 65 to 68 percent of census tracts. This finding suggests that areas with higher traffic volumes on the road network are likely to have fewer pedestrian and cyclist crashes, simply because fewer people are using these non-motorised modes (Zhang *et al.*, 2015).

In Canada, Miranda-Moreno *et al*. (2011) analysed crash data collected in the City of Montreal and found a significant influence of traffic volume on pedestrian crash frequency at signalised intersections. The results from this study indicated that a reduction of 30 percent of traffic volume would lead to a reduction of 35 percent in the total number of injured pedestrians and a reduction of 50 percent in pedestrian crash risk at signalised intersections.

In contrast, a few studies have reported inconsistent results with regard to the impact of traffic volume on the incidence of pedestrian crashes. In Hong Kong, Yao *et al*. (2015) applied the Negative Binomial regression technique to motorist and pedestrian crash data. The authors found that traffic volume, measured in annual average daily traffic (AADT), was negatively related to the frequency of vehicle-pedestrian crashes. This finding suggests that pedestrian crash risk is greater on roads with lower traffic volumes. Reduced pedestrians' attention in situations of low vehicular traffic volumes was given as a possible reason that may increase the likelihood of pedestrian crash occurrence (Yao *et al*., 2015).

2.3.3.3 Socio-economic variables

Research has shown a relationship between pedestrian crash risk and low socio-economic status of populations (Cubbin & Smith, 2002; Marcin *et al.*, 2003; Zoni, Domínguez-Berjón *et al*., 2016). A study that analysed crash data in Hawaii reported significant associations between pedestrian crashes and socio-economic variables, such as the number of jobs and the number of people living below the poverty level (Kim *et al*., 2010). In the study by Wier *et al.* (2009) carried out in San Francisco, an analysis of pedestrian crash data aggregated at the census tract level indicated that pedestrian crash risk is higher among resident populations living below the poverty level and among employed populations.

2.4 Impact of the built environment and pedestrian safety in Cape Town

2.4.1 General information on the City of Cape Town

The City of Cape Town is a large metropolitan area covering a land area of 2,461 square kilometres (Statistics South Africa, n.d.). Cape Town is the oldest city in South Africa, and has the second biggest population of all other South African Cities (City of Cape Town, 2012a). The City of Cape Town is the second largest metropolitan municipality after Johannesburg (WESGRO, 2016). The City of Cape Town includes eight planning districts and these are the Northern District, Blaauwberg, Cape Flats, Table Bay, Southern District, Khayelitsha/Mitchell's Plain, Helderberg District and Tygerberg District (City of Cape Town, 2015a) (see Figure 2-4).

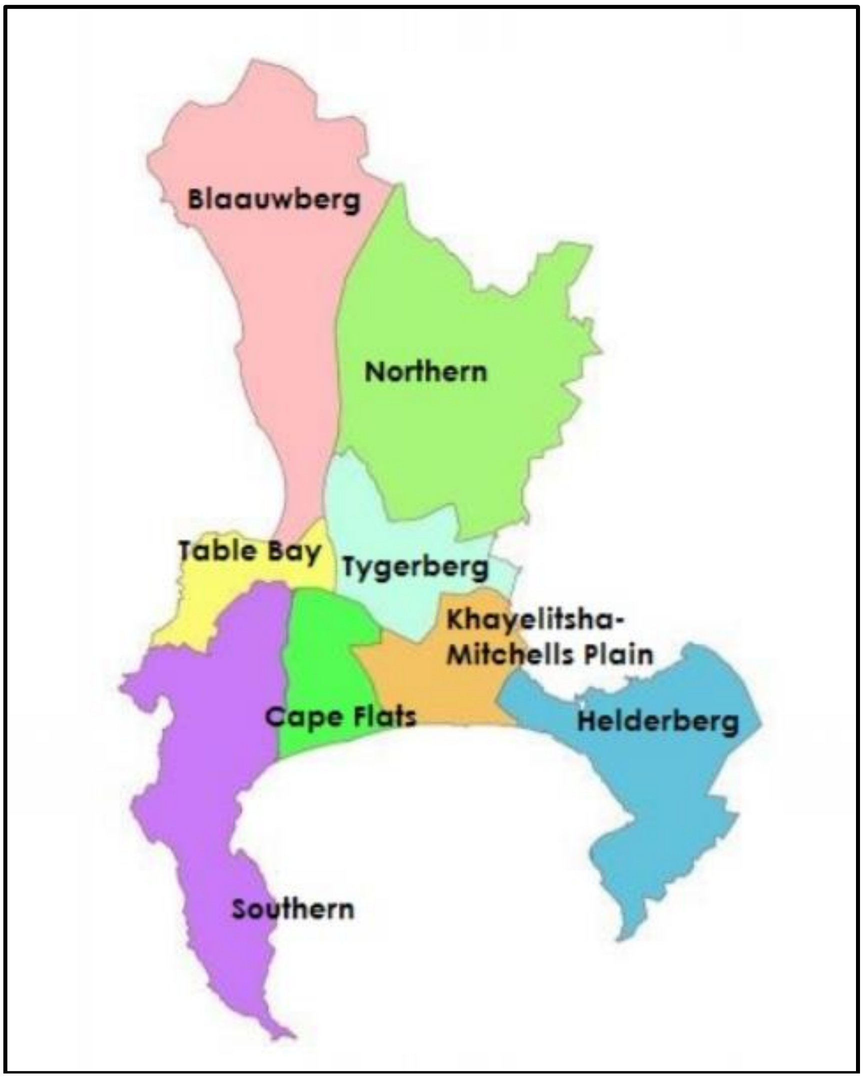


Figure 2-4: Eight planning districts of the City of Cape Town (City of Cape Town, 2015a)

Based on the 2015 population data, the distribution of the population by district is illustrated in Figure 2-5. In 2011, the population of Cape Town was 3 740 025 and this figure corresponds to an increase of 29.3 percent since 2001. The number of households in 2011 was 1 068 572, corresponding to an increase of 37.5 percent since 2001. Of the city's households, 14 percent (144 000 households) are reported to live in informal dwellings (City of Cape Town, 2012b).

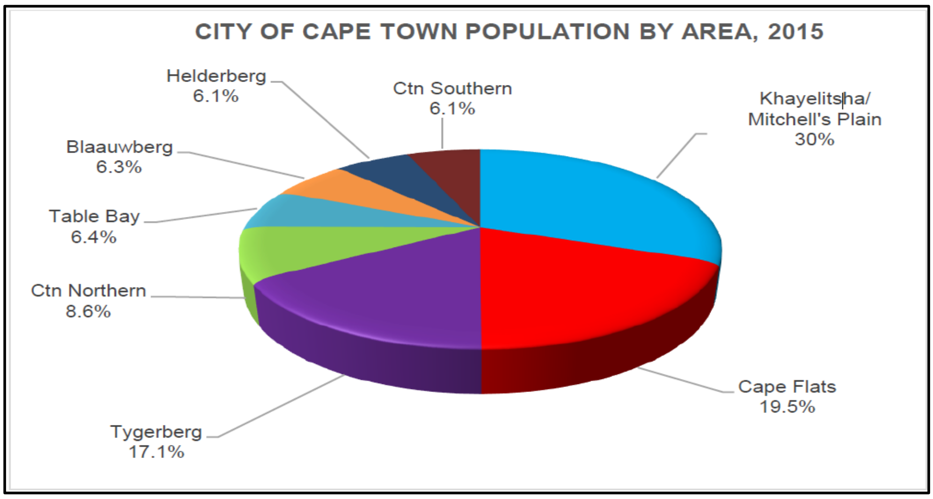


Figure 2-5: City of Cape Town population by district, 2015

As the population histogram for Cape Town shows, the majority of the population are around the ages of 25 to 29 years for both males and females (City of Cape Town, 2012b). There is an apparent graph tapering in the age groups between 5 and 19 years (see Figure 2-6) and the dip might be the result of the infant mortality pattern that prevailed between 1996 and 2011(Statistics South Africa, 2015).

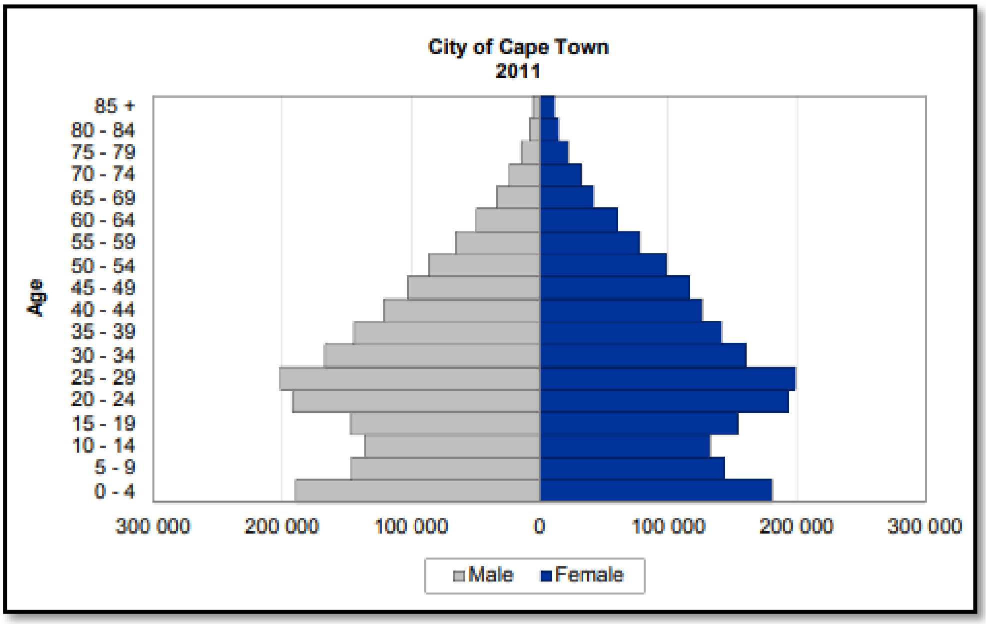


Figure 2-6: Histogram of Cape Town population, 2011(City of Cape Town, 2012b)

2.4.2 The influence of the built environment on pedestrian safety in Cape Town

The demand for transport and travel mode patterns are influenced by the nature of built-up areas and open spaces and their spatial distribution in the area. The past land use planning in South African cities created fragmented and dispersed urban activity patterns by promoting rigid mono-functional zoning of land. The 2004 State of the Cities Report states that "the Apartheid city was a political economy of space that had two central features: racially-based spatial planning and a political economy that meant development for some at the expense of the majority" (SAC Network, 2004). Apartheid urban planning was shaped by policies of strict social segregation which reserved well-located land for the specific races and classes and forced the poor people, especially poor Black residents to live in sprawling, overcrowded and dysfunctional settlements which are devoid of work and economic opportunities, social services and recreational facilities (SAC Network, 2004; Turok, 1994).

A deliberate separation of population groups through vacant land, railway lines, highways and major arterial roads, also perpetuated the victimization of poorer communities (SAC Network, 2011). This resulted in longer travel distances and higher expenditures on public transport for the poor and unattractiveness of non-motorized travel modes. In addition, the access to important urban amenities across railway lines, freeways and major arterial roads affects the safety of residents of those poorer communities as they are forced to either cross these physical barriers in unsafe manners or endure longer travel distances to safely cross at designated crossing facilities. There is compelling evidence that the most hazardous locations of pedestrian crashes in the City of Cape Town are located in or adjacent to informal settlements and along wider and high-speed roads (City of Cape Town's Transport Authority, 2005).

A few studies have reported on the impact of the road network structure on pedestrian safety in Cape Town. As illustrated in Figure 2-7, the highest rates of pedestrian crashes over a 6-year period (1997-2002) were found on a segment of Lansdowne Road situated between Strandfontein Road and Baden Powell Drive, followed by a section of the N1 (City of Cape Town's Transport Authority, 2005). The majority of roads ranked to be the most hazardous locations for pedestrian crashes are urban freeways facilities (see Figure 2-7). These are also locations where pedestrian injuries resulting from road traffic crashes tend to be more severe (i.e. serious injuries and fatal injuries), highlighting the impact of vehicular speed on pedestrian injury severity. However, it is not clear whether an exposure measure (e.g. number of pedestrian crashes by km linear length) was applied to the number of pedestrian crashes analysed in this study.

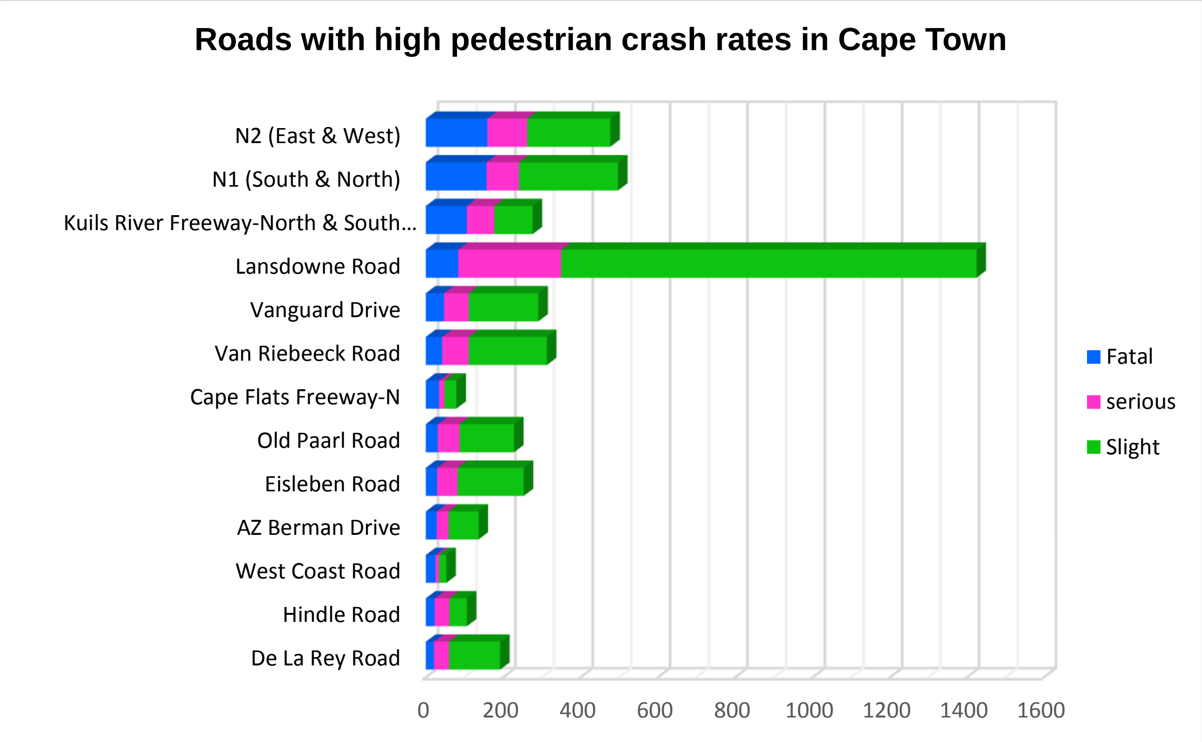


Figure 2-7: Roads with high pedestrian crash rates in Cape Town (City of Cape Town's Transport Authority, 2005)

Freeway facilities are locations of a significant number of pedestrian crashes in South Africa. About 2 000 pedestrian crashes occur annually on freeway facilities (Ribbens, 1996). A number of unsafe pedestrian behaviours, such as pedestrians walking through the interchange area to destinations on the other side of the freeway; pedestrians using the interchange as a modal transfer point; and pedestrians engaging in retail activities significantly contribute to pedestrian crash occurrence on freeway facilities (Ribbens, 1996).

In Cape Town, Behrens (2010) carried out an investigation into pedestrian crossing behaviour on selected arterials and freeways in Cape Town. Different patterns of pedestrian crossing behaviour, including distance between footpath crossings, the nearest formal crossing facility and pedestrian movement desire lines were observed. Additionally, an exploratory roadside intercept survey was conducted to explore pedestrian crossing attitudes and reasons for illegal crossing preferences. The results showed that 62 percent of crossing events on freeways and 93 percent of crossing events on arterials were illegal. Even though the distribution of footpath crossing distances from the nearest crossing facility showed a greater use of freeway crossing facilities compared to arterial crossing facilities, patterns of crossing were found to be strongly associated with the location of crossing facilities in relation to dominant pedestrian movement desire lines. The results from the intercept survey on grade-separated facilities revealed that the most common reason to choose a particular route was the desire to walk the shortest route, followed by concerns for personal security (Behrens, 2010).

The next stage of the study by Behrens (2010) investigated pedestrian crossing points on two arterials (Klipfontein Road and Buitengracht Street), a major collector (Cavendish Street) and two freeways (the N2 and the R300) all located in Cape Town. Results showed that only 15 percent of crossings occurred at the crossing facility on Klipfontein Road. The remaining (85 percent) were distributed away from the designated crossing points. On Buitengracht Street, a rate ranging from 1 to 5 percent of crossing events was observed at the designated crossings and the remaining (95-99 percent) were observed away from the designated crossing points, with the highest concentration being located at 61-70 metres from the crossing facility. A high rate of spatial compliance (80 percent) was observed on Cavendish Street. According to the author, the higher level of spatial compliance was attributed to the presence of a crossing facility located close to the pedestrian movement desire lines for that particular road. On freeways, the study indicated that the concentration of crossing points was between 100 and 300 metres from the nearest crossing facility, whereas only 5 to 15 percent was located at crossing facilities (Behrens, 2010).

2.5 Measure of the attributes of the built environment

The built environment has been subjected to close scrutiny over a wide array of studies for its potential correlations with travel behaviour, health problems, traffic crashes, and so forth. Numerous researchers have studied the built environment at various scale levels, from building or site level to the neighbourhood and regional levels. These researchers have used three approaches to measure the attributes of the built environments: Objective measure, subjective measure and a combination of both measures (Lin & Moudon, 2010; Saelens & Handy, 2008).

Two studies carried out by Lin and Moudon (2010) and Saelens and Handy (2008) provide a review of existing studies that applied the three approaches of measuring the built environment. Objective measures of the built environment involve the use of field-collected data and data stored in different non-spatial and spatial databases to describe and quantify the attributes of the built environment (Lin & Moudon, 2010; Orstad *et al*., 2017). Subjective measures are self-reported perceptions on the aspects of the built environment obtained from survey questionnaires (Humpel *et al*., 2004; Lin & Moudon, 2010; Nyunt e*t al*., 2015; Orstad *et al.*, 2017). Objective measures of the built environment have benefited from the recent emergence of Geographical Information Systems (GIS) technology (Koohsari *et al*., 2015). GIS is an automated system designed to capture, store, manipulate, analyse, manage and present spatial data (Dangermond, 1992). GIS provides an easier and cost-effective alternative for analysing and measuring the built environment especially for large study areas. This subchapter provides a review of objectives measures used in literature to quantify aspects of the built environment.

2.5.1 Measures of land use patterns

Land use patterns are usually measured in terms of proportion of a particular land use types within a geographic unit of analysis, number of properties of a particular use within a unit of analysis or intensity of use. Land use intensity is often described in terms of counts (e.g. number of land-use types), percentage (e.g. percentage of parcels with residential use) and proportions (e.g. jobs-to-housing ratio) of specified land-use types within a geographic unit of analysis (Gehrke & Clifton, 2015; Song & Rodriguez, 2005). Land use intensity can also be measured in terms of density such as employment density (i.e. number of employees per area of analysis) (Brownson *et al*., 2009; Song *et al*., 2013).

2.5.2 Measures of land use mix

As defined in the sections above, the concept of land use mix reflects the quantity and interaction between land use types within a particular limited space. Existing methodological approaches of mixed land-use measures are often grouped according to two conceptual elements: land-use interaction and geographic scale (Gehrke & Clifton, 2015). Put simply, some measures of land use mix are used to quantify the interaction between co-located land use types and another set of measures seek to quantify this interaction with a great emphasis on controlling the effect of size on spatial units of analysis.

Measures of land-use interaction are often classified into three categories: accessibility, intensity and pattern measures (Brownson *et al.*, 2009; Song & Rodriguez, 2005). From a broader perspective, accessibility measures denote the ability of individuals to access human activities important to their quality of life and wellbeing or simply, the ease with which individuals can reach a particular land-use (Morris *et al*., 1979; Song & Rodriguez, 2005) Intensity measures describe the extent or magnitude to which land use types are represented in an area and pattern measures describe the manner in which land use types are spatially distributed and arranged within an area (Gehrke & Clifton, 2015; Song & Rodriguez, 2005).

2.5.2.1 Accessibility-based measures

Reaching spatially separated land use activities involves human effort, time, cost, availability of services, individual willingness to partake in a particular activity and attractiveness of an activity. Some of these constraints encapsulate a behavioural dimension (e.g. perceived activity attractiveness and service supply, travel choices) and others reflect a spatial separation dimension such as distance, travel time and cost (Morris et al., 1979). As a result, measures of land use mix which attempt to conceptualise accessibility in terms of these constraints vary widely depending on intended application. In this study, a review of these measures is restricted to those which encapsulate the notion of spatial proximity of land use types. Two main categories of measures commonly used include distance-based measures and gravity-based measures (Brownson *et al.*, 2009; Kwan, 2013; Du Plessis, 2015; Song & Rodriguez, 2005).

Distance-related measures quantify spatial separation in terms of linear or street network distance between an origin location and a closest specified land-use type. Gravity-based measures capture the level of attractiveness of each potential destination land-use and weigh that attraction by the willingness to travel (distance decay function) or simply, travel costs as impedance to travel to a destination from an origin location. The travel impedance function is also referred to as travel disutility, and is expressed in terms of distance, travel time and money. Travel time is the most common form of travel disutility adopted in gravity-based measures and accessibility in this sense represents the cumulative number of land-use types accessible within a given amount of travel time from a point of origin (Chen *et al*., 2011). A gravity-based measure can quantify, for instance, the number of jobs that are reachable within 30 minutes by car or transit.

Although not widely explored by research, a further concept referred to as "temporal availability" is explored in several studies (Gehrke & Clifton, 2015; Kwan, 2013). This novel idea uses the concept of time-geography (Miller, 2005) as a conceptual framework to shed light on the failure of current accessibility-based measures to account for temporal availability of land use. Within this conceptualisation, in addition to spatial constraints, time constraints (e.g. facility or service opening hours, transit schedules) are also important factors that influence the accessibility of activities in urban spaces. In many circumstances, spatial proximity doesn't always account for better accessibility. Staying in close proximity to a governmental service does not necessarily means that the service is more accessible to a person as space-time constraints (e.g. work schedules and service opening hours) can hinder the service accessibility (Kwan, 2013). Integrating the notion of temporal availability into accessibility measures is believed to improve the accuracy levels of accessibility measures (Gehrke & Clifton, 2015; Kwan, 2013) and therefore, consideration of spatial-temporal accessibility measures is recommended for future research.

2.5.2.2 Intensity based measures

Intensity-based measures describe land use mix in terms of counts (e.g. number of land-use types), percentage (e.g. percentage of parcels with residential use) and proportions (e.g. jobs-to-housing ratio) of specified land-use types within a geographic unit of analysis (Gehrke & Clifton, 2015; Song & Rodriguez, 2005). Measures in terms of density such as employment density (number of employees per area of analysis) also fall within the category of intensity-based measures (Brownson *et al.*, 2009; Song *et al.*, 2013).

2.5.2.3 Pattern-based measures

These measures quantify the spatial distribution and configuration (spatial arrangement) of different land use types within a spatial unit of analysis. Song and Rodriguez (2005) categorised these measures into three groups: evenness and diversity measures; exposure measures; and clustering measures. Common measures of evenness and diversity include (1) the Balance Index; (2) Entropy measures; (3) the Herfindahl-Hirschman Index; (3) the Gini Coefficient and (4) the Atkinson Index (Song & Rodriguez, 2005). The exposure Index and Clustering Index represent the remaining categories, respectively. A further classification of these measures into integral and divisional measures is proposed in the study carried out by Song et *al*. (2013) based on sensitivity to area-wide land-use distribution. Integral measures reflect land use balance and are determined based on overall distribution of land use types within an area. In contrast, divisional measures provide an assessment of the evenness of land use types and make use of significantly smaller subdivisions (Song et *al*., 2013).

2.5.3 Urban form measures

The features of urban form that are commonly considered in many studies include land use mix, connectivity and accessibility.

2.5.3.1 Street connectivity measures

There is a growing interest in the literature of urban planning and transportation studies to identify an appropriate approach to measuring street connectivity. A review of this literature provides various quantitative measures of connectivity most commonly used by researchers and practitioners. This section describes several measures of street connectivity commonly used by previous studies.

1. **Connectivity measures based on block density and size**

Numerous studies have used block characteristics as a tool to measure connectivity. A block is defined as the smallest fully enclosed polygon bounded on all sides by streets, roads, railway tracks or geopolitical boundaries lines (Ewing *et al*., 2003; Frank *et al*., 2000). The quality and the quantity of connections within a street network depend heavily on the length and size of blocks.

Block length was applied in a small group of studies as a standard tool to measure connectivity (Berrigan *et al*., 2010; Cervero & Kockelman, 1997a; Cervero & Radisch, 1996; Handy *et al*., 2003). The Block length is usually the distance measured from the kerb of one intersection to the kerb of the next intersection (kerb-to-kerb distance). When working on large study areas, some researchers simplified the measurement of the average block length by dividing the total roadway length by number of the blocks adjacent to the road (Parks & Schofer, 2006). Shorter block lengths mean more street intersections and more intersections provide flexibility and a greater number of route alternatives between locations. Neighbourhoods that are considered most pedestrian-friendly have shorter blocks and higher intersection density. Higher intersection densities means more links (road segments between intersections) with slower speed as traffic must slow down and possibly stop, thus providing more opportunity for pedestrians to safely cross the roadway and access their destinations. Many cities have regulations on block length to allow adequate intersection spacing. For instance, according to several sources, the maximum block lengths in the US usually ranges from 300 to 600 feet or 91.4 to 182.9 meters (Association, 2006; Handy *et al.*, 2003). In South Africa, a fairly short block with a length of approximately 100 meters is considered as the most appropriate (CSIR Building and Construction Technology, 2005).

Another set of studies has used block size as a standard tool of connectivity measure. Contrary to block length that considers only one dimension of the block (length), block size captures a two-dimensional structure of the block (length and width). Studies that have adopted this approach assessed connectivity within a given study area in terms of average block area (Ewing *et al*., 2014; Hess *et al*., 1999; Krizek, 2000; Marshall & Garrick, 2010). In addition to these studies that based connectivity measure on average block area, other (though fewer) studies determined block size in terms of block perimeter and an average block perimeter was adopted as a proxy of connectivity level within a study area (e.g. Song & Knaap, 2004). Other researchers have used block density as another block-related proxy measure of connectivity (Berrigan *et al.*, 2010; Cervero & Radisch, 1996; Frank *et al.*, 2000; Parks & Schofer, 2006; Song & Knaap, 2004). In these studies, block density was defined as the number of blocks per unit area of urban land.

2. Intersection density

A large number of studies used intersection density as a measure of connectivity (Berrigan *et al.*, 2010; Dill, 2004; Ewing & Cervero, 2010; Marshall & Garrick, 2010; Parks & Schofer, 2006). In these studies, intersection density refers to the number of street intersections per unit area. As a supplement to other connectivity measures, proportions or density of a specific type of intersection (e.g. T-intersection, four-way intersections or intersection with more than three legs) are another way of measuring connectivity employed by a few studies (Cervero & Kockelman, 1997a; Greenwald & Boarnet, 2001; Parks & Schofer, 2006; Wood, Frank &

Giles-Corti, 2010). Intersection density has also been determined at street scale level in many other studies, as the number of street intersections per kilometre or mile of street network length (Kang & Oh, 2016; Larsen & El-Geneidy, 2011; Troped *et al*., 2010).

3. Density of cul-de-sac

A cul-de-sac is a dead-end street mostly ending in a circular turnaround. Research has shown that a large number of cul-de-sacs within a given area is a hindrance for walking and cycling because of longer walking distances to destinations as a result of limited permeability of the area (Marshall & Garrick, 2010; Rajamani *et al*., 2003). Density of the cul-de-sacs as a measure of connectivity has been adopted in a number of studies. For some studies, the density was calculated as the number of cul-de-sacs per locality (Cervero & Radisch, 1996) or percentage of cul-de-sac (dead-end) intersections in a street network (Rajamani *et al*., 2003). For other studies, the density was calculated as the number of cul-de-sacs per unit area (Marshall & Garrick, 2010). Aside from the density of cul-de-sacs, literature searches have found an uncommon measure of connectivity in the study by Song & Knaap (2004), expressed in terms of median length of cul-de-sac, measured in linear units.

2.6 Literature on traffic crash modelling

Over the past few decades, modelling traffic crashes has been a very important topic in road safety research. Researchers in traffic safety employ various techniques for analysing crash data. The most widely used and simplest technique in safety estimation is the use crash frequency as an indicator of crash incidence on a road network or certain segments of the roads (Abdulhafedh, 2016). Crash frequency is defined as the number of crashes occurring at a particular site, facility or road network in a given period, often in one-year period (AASHTO, 2010). However, the use of crash frequency in safety analyses is undermined by a number of limitations including those related to natural fluctuation in crash data due to the random nature of crash events, changes in roadway characteristics, land uses and other exposure variables (AASHTO, 2010). To deal with these shortcomings, many researchers in the field of road safety use statistical methods to estimate expected crash frequency as a function of a variety of potential contributing factors.

Crash modelling was initially based on the standard ordinary least-squares (OLS) regression model which assume a normal distribution of errors. However, researchers realised that the application of OLS regression is not appropriate to model count data (Lord & Mannering, 2010). Crash data is known to be discrete, random and non-negative events. Modelling this type of data requires the application of regression methods or other approaches that are appropriate in handling count data or the integer nature of the data such as the Poisson regression model (Lord & Mannering, 2010). Researchers thus began applying the Poisson regression model and its derivatives including the negative binomial regression model and zero-inflated model, all being part of an advanced modelling technique called the Generalised Linear Models (GLMs) (Caliendo *et al*., 2007; Ukkusuri *et al*., 2011). The GLMs are the most common modelling techniques (Lord & Mannering, 2010; Noland & Oh, 2004; Washington *et al*., 2010) to have been used in traffic safety research (e.g. Hadayeghi *et al*., 2010; Li *et al*., 2013; Pirdavani *et al*., 2014).

Parameter estimates in the traditional Generalised Linear Modelling (GLM) are estimated to quantify the average associations between the explanatory variables and the outcome variables for the entire study area, with the assumption that these associations do not vary across the study area (Amoh-Gyimah *et al*., 2017; Hadayeghi *et al*., 2010). However, research in traffic safety recognises that spatial data such as crash counts are not generally independent (Hadayeghi *et al*., 2010). Using geospatial analysis techniques, numerous studies conducted with a particular focus on spatial patterns have confirmed the presence of spatial dependencies among different variables across geographic areas (Flahaut, 2004; Flahaut *et al*., 2003; Geurts *et al*., 2005; Huang *et al*., 2010; Moons *et al*., 2009). To address the issue of spatial dependence among variables, these studies highlighted the importance of integrating spatial autocorrelation in the modelling process of spatial data (Aguero-Valverde & Jovanis, 2008; Flahaut, 2004; Guadamuz-Flores & Aguero-Valverde, 2017).

In the process of modelling spatial data, spatial variations in explanatory variables may be observed, especially when the study area is relatively large (Pirdavani et *al*., 2014). Research has confirmed the existence of the spatial variation in numerous variables such as population, employment, road characteristics and other environmental characteristics across geographic areas (LaScala *et al*., 2000; Levine *et al*., 1995). Spatial variations in relationships are referred to as spatial heterogeneity or spatial non-stationary (Fotheringham *et al*., 2002; LeSage & Pace, 2009). The inability to capture spatial heterogeneity could lead to inconsistent and parameter estimates (Mannering *et al*., 2016; Washington *et al*., 2010). Parameter estimates in the traditional GLMs are a set of fixed coefficients generated globally over the entire study area. However, crash frequency is influenced by a number of zonal variables (such as land use, population characteristics, traffic volumes, pedestrian volumes, speed, roadway and environmental variables) that vary over the study area. Accordingly, the influence of some of these variables may be more pronounced in certain locations but this influence may be weak in other locations (Hadayeghi *et al*., 2010).

To account for heterogeneity across observations, several modelling techniques were developed by researchers and are currently being applied in crash modelling. Among them, the Geographically Weighted Regression (GWR) modelling is the most commonly used technique (Li *et al*., 2013) and has been reported to generate more accurate parameter estimates than the traditional GLM (Delmelle & Thill, 2008; Erdogan, 2009; Hadayeghi, Shalaby & Persaud, 2003; Li *et al*., 2013; Zhao & Park, 2004). The parameter estimates in GWR modelling are allowed to vary over the study area to account for spatial variations in relationships among observations (Li *et al*., 2013; Zhang *et al*., 2015).

The basic GWR modelling technique assumes that errors are normally distributed (Zhang *et al*., 2015). However, this assumption is often not supported in crash modelling (Hadayeghi *et al*., 2010; Zhang *et al*., 2015). To increase the accuracy of parameter estimates, the GWR technique has been adapted to generalised linear models (Poisson regression and Negative Binomial models) to form geographically weighted Generalised linear models (GWGLMs) (Fotheringham *et al*., 2002). The GWR modelling used in conjunction with a Poisson regression (i.e. with a Poisson distribution for errors) is referred to as Geographically Weighted Poisson Regression (GWPR). The model used in conjunction with Negative Binomial regression (i.e. incorporating the over-dispersion of count data or with negative binomial distribution for errors) is termed the Geographically Weighted Negative Binomial (GWRNBR) (Fotheringham *et al*., 2002). The GWPR approach has been applied more frequently than the GWRNBR approach in safety research, owing to the claim that the GWPR does generate accurate estimates (Hadayeghi *et al*., 2010). Nonetheless, a recent study by Gomes, Cunto and da Silva (2017) has reported that the GWPR model performs better than the GWPR model in capturing the spatial heterogeneity of crash frequency. Another reason explaining the frequent use of the GWPR approach is the availability of software tools (such as the GWR 4.08 software) that support GWR with a Poisson regression structure (Amoh-Gyimah *et al*., 2017; Li *et al*., 2013; Zhang *et al*., 2015).

2.7 Concluding notes on the literature survey

This chapter involves a review of existing literature relevant to the research questions and the theoretical framework adopted in this study. The key implications of the literature review chapter can be summarised as follows:

- The chapter provided an enhanced understanding on the risk factors of pedestrian crashes internationally and in the context of South Africa.
- The literature reviewed in this chapter guided the development of the methods applied in this study to investigate the research questions.
- The findings from the reviewed works helped to validate, compare and discuss the results produced in this study.

Chapter 3: Research Methodology

3.1 Introduction

This chapter describes the methodological approach used to investigate the relationships between the built environment and the patterns of pedestrian crashes in urban spaces. The chapter provides a description of the procedures followed to achieve the research objectives, including the research design, data acquisition, data sampling, data processing and data analysis. It also provides an outline of research tools and software packages used to gather, process and analyse research data.

The first purpose of procedures presented in this chapter was to obtain reliable quantitative and qualitative data on the built environment and pedestrian crashes within the study area, which was aggregated at the level of the suburb. The second purpose was to identify relationships between the attributes of the built environment and pedestrian crash incidence through statistical methods.

3.2 Methodology

3.2.1 Research instrumentation

The study was conducted by using research tools including hardware, software applications and web applications, accessed via the Stellenbosch Smart Mobility Lab (SSML) at Stellenbosch University. Tools such as hardcopy and electronic books, databases, journals and thesis were accessed via the university library and other local as well as international online databases. These tools served as a means of gathering research ideas and for accessing information on existing literature relevant to the research topic. The web-based application "Refworks" enabled the management of references and the creation and organization of the researcher's reference database specific to this study. The data collection process required two web database applications - Google Maps, and Online Zoning Viewer. The Online Zoning Viewer is an open source database containing property details and other spatial features such as streets, boundaries of wards and suburbs for the City of Cape Town. The data processing and analysis were made possible by the use of 4 software applications; MS Excel 2013, ArcGIS version 10.3.1, IBM-SPSS Version 24.0 and STATISTICA version 13.3.

3.2.2 Data collection

3.2.2.1 Data types and sources

The study made use of two categories of data: primary and secondary data. Primary data is directly gathered by the researcher, whereas secondary data includes information gathered by someone else (Mouton, 2001). Secondary data was collected by various institutions at national, provincial and local levels. Observations were the source of information for primary data and were conducted on transport facilities to either supplement information on secondary data or gather information on different variables under investigation.

1. **Primary data**

   i. **<u>Data related to crash locations</u>**

Primary data was collected using google (street view and satellite) images and aerial photographs. The data consists of a list of design features associated with the location of a pedestrian crash and the geographical coordinates of the pedestrian crash location. An additional 15 columns were added to the Excel spreadsheet of pedestrian crash details obtained from the Cape Town city's Transport and Urban Development Authority (TDA) to accommodate the additional primary data. These additional columns were labelled as (1) Accuracy; (2) Facility Type; (3) Control Type; (4) Number of lanes; (5) Node; (6) Node Type; (7) Intersecting Roads; (8) Crossing Distance_Minor; (9) Crossing Distance_Major; (10) Crosswalks; (11) Refuges; (12) Sidewalks; (13) Street lighting; (14) Longitude and (15) Latitude. Dropdown lists were created in a separate spreadsheet to allow for consistent coding of information and to facilitate data entry. In addition, a unique identification number (I.D) was assigned to an individual pedestrian crash to distinguish it from others.

Coding of design features for intersection-related pedestrian crashes

A pedestrian crash for which the location is described as a "node" was considered to be an intersection-related pedestrian crash. As previously noted, crash locations were provided with reference to a number of features including node description, name of a suburb, SAPS station and other pertinent road features such as bridges, kilometre marker and so forth. The researcher used a combination of these details to identify a pedestrian crash location on Google maps. Once the location had been determined, design features were confirmed using Google Maps imagery (Street View, 2D and 3D Earth's satellite) and aerial photographs, and this data was then coded into the Excel spreadsheet of pedestrian crashes. Geographical coordinates (longitude and latitude) of the central point of the intersection were also retrieved from Google Maps and reproduced into the pedestrian crash database. A list of codes and their interpretation is given in Table 3-1.

Table 3-1: List of codes used in collecting data on crash locations and their interpretations

Data Item	Codes	Interpretation	Example/Comment
Accuracy	Yes	Crash location with accurate descriptions such as node name, kilometre marker, or a unique feature	"STRAND ST X LOOP ST, CBD" "N2, KM 8.3 MARK" "N2, UNDER R300 BRIDGE"
	Close	Crash location described only by street name or adjacent nodes or interchange with ambiguous directions	"RIEBEEK STR, CBD" "VANGUARD DR, HIGHLANDS DR // MORGENSTER RD" "VANGUARD DR X N2, UNKNOWN DIRECTION
	No	Vague or ambiguous location of crash	"N1, UNKNOWN LOCATION" "PRIVATE PROPERTY, GOODWOOD" "HOUT BAY MAIN RD, UNKNOWN NODE"
Facility Type	Freeway	Road with full access control	N2, N1, N7, M7, M5, M3, R300
	Intersection	A node or a junction	3-legged, 4-legged, staggered and traffic circle/roundabout
	Mid-block	A link or between two adjacent nodes	"N2, MEW WAY // R300"
	Private property	Location away from a public road such as Driveway, open parking areas, parking in buildings	Reported information
	Shoulder	Emergency stopping lane adjacent to travelled way	Reported information
	Sidewalk/Verge	A sidewalk (separated from a roadway) or a verge of the road. A verge is an area between the roadway edge and the road reserve boundary	Reported information
Mid-block control	Signalised	Intersection or mid-block location controlled by traffic signals	"M4, PED XING AT ROSEHOPE, ROSEBANK"
	Unsignalised	Intersection without controls or with control device other than signals	Uncontrolled intersection and intersection controlled by YIELD OR STOP signs, traffic circles, roundabouts

Lane Number	1 to 10 lanes	Number of lanes at a mid-block location in both directions	Applicable only for non-intersection crashes
Node Type	3Leg	Three-legged intersection	T-intersection
	4Leg	Four-legged intersection	Cross intersection
	Mleg	Multi-legged intersection	Intersection with more than four legs
	Staggered	Staggered intersection	Staggered T-intersection
	Roundabout/Minicircle	Traffic circle	"BHUNGA AVE X NDABENI ST (CIRCLE), LANGA"
	Gravel Intersection	Intersection on unpaved roads	"RADIO RD X BLUEGUM ST, KLIPHEUWEL"
Intersection control	Signal	Traffic lights or "Robots" (South African Context)	
	1Way Traffic	Intersection with one-way traffic on one or more legs	Intersection controlled by No entry (R3) sign
	1Way STOP	Three-legged intersection where a single minor road approach is controlled by a STOP sign	Intersections where a minor street is a private driveway are excluded in this study
	2Way STOP	Four-legged intersection where a major road is uncontrolled and two minor road approaches are controlled by a STOP sign	
	3way STOP	Four-legged intersection where one approach is uncontrolled while three other approaches are controlled by a STOP sign	In most cases, the uncontrolled approach has one-way traffic
	4Way STOP	Four-legged intersection where all four approaches are controlled by a STOP sign and the principle of "first come, first served" is applied	
	YIELD	Intersection controlled by YEILD sign	
	Uncontrolled	Intersection without any control device	

Access types	FI	Full intersection	It allows all possible movement of travel
	PI	Partial intersection	It allows left-in, left-out and right-in movements of travel
	MI	Marginal intersection	It allows left-in and left-out movements of travel
Intersecting Roads	Minor & Major	Roads with different functional classification	"KOEBRG RD X FREEDOM WAY,MILNERTON"
	Same size	Roads with the same functional classification	"KLIPFONTEIN RD X DUINEFONTEIN RD"
CrossDist_Minor	1 to 10 lanes	Width of a portion of an intersection leg of a minor road where pedestrians are exposed to traffic in both directions	"4 lanes" denotes a crossing distance on a minor road of four lanes excluding medians and refuges
CrossDist_Major	1 to 10 lanes	Width of a portion of an intersection leg of a major road where pedestrians are exposed to traffic in both directions	"8 lanes" denotes a crossing distance on a major road of eight lanes excluding medians and refuges
Crosswalks	C/L	C= Number of crosswalks L= Number of intersection legs	1/3 denotes presence of only one designated crosswalk at a three-legged intersection
Refuges	R/L	R = Number of refuges L= Number of intersection legs	2/4 denotes presence of two refuges (splitter islands or medians) at a four-legged intersection
Slipway	S/L	C= Number of slipway L= Number of intersection legs	A slipway is a roadway that passes to the left of the main junction without intersecting the junction
Sidewalks	S/E	S= Number of available sidewalks E= Number of expected sidewalks	6/8 denotes presence of six sidewalks at an intersection where eight sidewalks are expected (four-legged)
Street lighting	Yes	Presence of at least one street light pole	Presence was recorded regardless the number of street lights and their status
	No	Absence of street lights	
Coordinates	Longitude & Latitude	Geographical coordinates of a crash location	Information obtained from Google Maps

Intersection design features were checked for each intersection-related pedestrian crash. These included alignment and number of legs (three legs, four legs, multi legs, staggered, roundabout or mini-circle); type of control (traffic signals, STOP sign, YIELD sign, and uncontrolled); access types (full, partial and marginal intersections); geometric features of road approaches (number of lanes, major road versus minor road, splitter island, slipway), pedestrian facilities (number of crosswalks, crossing distances, refuges and sidewalk) and road amenities (e.g. street lights). The availability of details and ability to visualize them on both Google imagery and aerial photographs determined the nature of design features to include in the intersection safety analysis.

Based on the design layout, intersections were categorized into three-legged, four-legged, multi-legged, staggered, roundabout/mini-circle and gravel intersections. A multi-legged intersection is defined as an intersection having more than four legs. An intersection is considered as staggered when access spacing (distance between the centre lines of consecutive intersection) was less than 50 metres. Otherwise, the two adjacent intersections were treated as two isolated T-junctions. The access spacing of 50 metres is the minimum access spacing recommended on priority-controlled local streets in South African urban areas (Committee of Transportation Officials, 2014). Staggered intersections were recorded as either "left-right" (LR) or "right-left" (RL) according to the order of the turning movements of a vehicle travelling on the minor road and crossing the major road. Figure 3-1 illustrates the main steps followed in the coding process of intersection-related pedestrian crashes.

Coding of design features for midblock-related pedestrian crashes

Inaccurate crash locations for midblock-related pedestrian crashes significantly limited the level of detail regarding road design features that were recorded at midblock locations. The number of lanes, the type of control (signalised or unsignalised) and the road class are the only design features which were recorded for this category of pedestrian crashes.

Midblock-related pedestrian crashes which were reported to have occurred at a pedestrian crossing or within 50 metres from a pedestrian crossing (pedestrian crashes containing the record "At crossing" and "Within 50 m from" in the column labelled "Pedestrian Location") were filtered out for special treatment. A pedestrian crash of this category is deemed to have occurred at an at-grade designated crosswalk located at a specific link reported in the crash database.

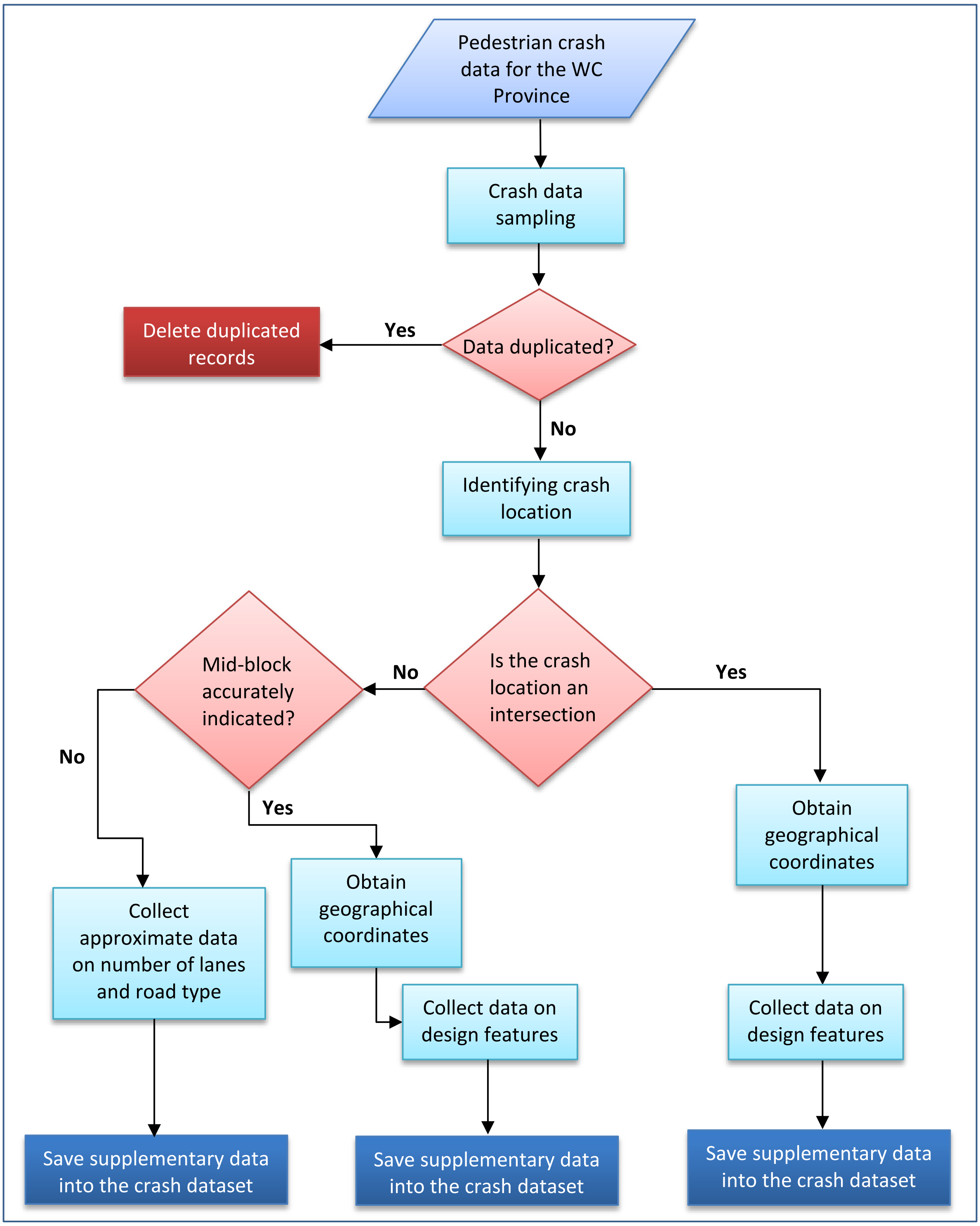


Figure 3-1: Flowchart for the coding process of pedestrian crash locations

The researcher used the name of the road to identify the link on Google Map and a meticulous search was performed along the link using Google Street View and aerial photographs to identify the accurate location of the crosswalk. Considerable care was taken to ensure that the scanned link was contained within the boundaries of the reported SAPS station. When a unique designated crosswalk was detected, the researcher confirmed with certainty the pedestrian crash location, and the corresponding geographical coordinates were retrieved from Google Maps. Following this, the number of lanes and type of controls were recorded and an entry “Yes” was recorded in the data field labelled “Accuracy”.

When more than one designated crosswalk was found on the link, the design features, together with the geographical coordinates of the predominant type of crosswalk at the link, were captured and the accuracy was referred to as “Close”. In circumstances where no crosswalk was found, the design features at the central point of the street segment and the crash location was reconsidered as a non-designated (or informal) midblock crossing. Interestingly, the crosswalk scanning revealed significant cases where speed humps/bumps were recorded as designated crosswalks in the crash database. For these particular cases, the crash location was also reconsidered as informal pedestrian crossings except in the case of a raised pedestrian crossing.

In a few cases, crash locations on freeway facilities were indicated using kilometre markers which were helpful for identifying the crash location using Google Street View as shown in Figure 3-2 and Figure 3-3.

Figure 3-2: Google image of a roadside kilometre marker on a freeway facility

Figure 3-3: Google image of kilometre marker on a concrete median barrier

Pedestrian crashes for which the data entry in the “Pedestrian Location” column was “Unknown” and “Not at crossing” were treated separately. For convenience purposes, the central point of the street segment represented the crash location. The corresponding geographical coordinates were captured to allow further analyses despite imprecise crash location details. A summary of the main steps followed in the process of coding midblock-related pedestrian crashes is provided in Figure 3-1 on Page 63.

ii. <u>Data on street connectivity</u>

As documented in the literature review section, street connectivity refers to the number of connections on a road network, including road segments, walking and cycling paths linking people to their destinations (Marshall, 2005). The literature review chapter presented a variety of techniques adopted by previous studies to measure street connectivity. Among the reviewed measures of street connectivity, four proxy measures were applied in this study and these comprise intersection density, the number of intersections with more than three legs, street density and the ratio of intersections to cul-de-sacs. These street connectivity measures are the most widely used in traffic safety research and their estimation requires data which is easily available at a citywide scale. The number of each type of intersections and culs-de-sac is crucial information for the determination of intersection density, the number of intersections with more than three legs and the ratio of intersections to culs-de-sac. These counts were aggregated at the census suburb level and further calculations were performed on aggregated counts to determine the four proxy measures of street connectivity.

Street density was determined using spatial data on the road network of the City of Cape Town. Intersection and cul-de-sac counting was performed manually by type of intersection control and configuration for each census suburb. Intersections were classified into 11 categories, which are: (1) four-legged signalised, (2) four-legged two way stop, (3) four-legged four way stop, (4) three-legged signalised, (5) three-legged one way stop, (6) three-legged one way stop, (7) three-legged two way stop, (8) three-legged three way stop, (9) staggered, (10) roundabout/mini-circle and (11) gravel road/uncontrolled. Both Google maps and aerial photographs (processed in ArcMap) were used to visualise the type of control and the intersection configuration. Further details on the determination of the proxy measures of street connectivity are provided in the data processing section.

To facilitate the counting process, a macro (i.e. a computer code written for Excel using the Visual Basic for Applications (VBA) programming language) was written in Excel such that each click on a cell added 1 to another cell specified in the code. This is to say that the number of clicks on a specific cell was counted in another separate cell. Each of the 12 categories of intersections was assigned a button named after the type of intersection or cul-de sac and a cell above the button was designated to display the number of clicks performed. Each time a particular type of intersection or a cul-de sac was identified on a suburb street network (extracted from the national road network using ArcMap); a click was performed on the corresponding button. At the end of the process, the total number of clicks performed on the button was displayed in the cell above the button and this number corresponded to the number of intersections of a particular type or the number of cul-de-sacs available within the boundary of the suburb of interest. A screenshot of the worksheet buttons with an assigned macro is illustrated in Figure 3-4.

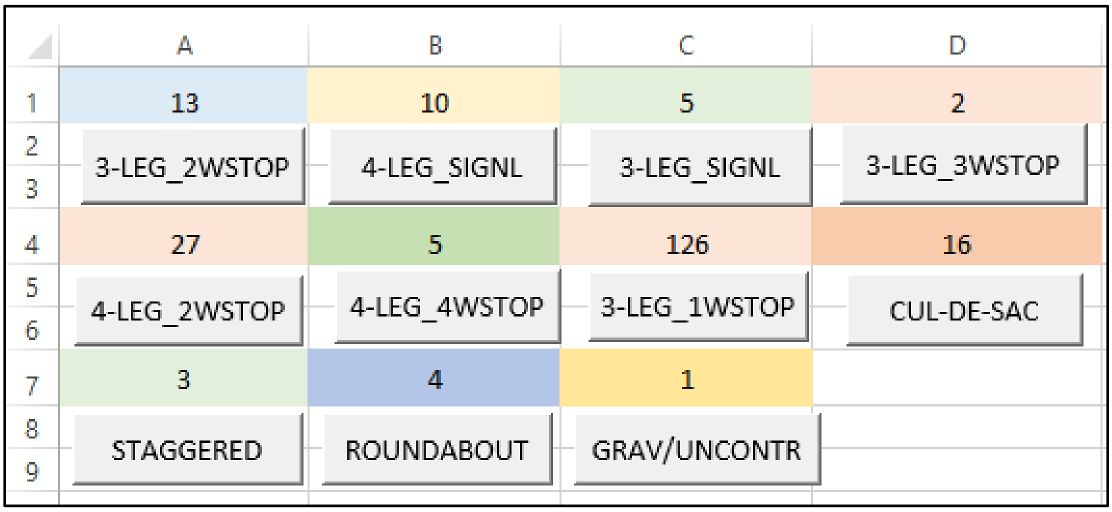


Figure 3-4: An example of VBA buttons with an assigned macro to count the number of intersections and culs-de-sac

2. Secondary data

i. Demographic data

The City of Cape Town through the Strategic Development Information and GIS Department (SDI & GIS) provided demographic data compiled from the 2011 population census. The data consists of two formats; data in MS Excel format and a GIS-related database. Data in MS Excel format is aggregated at the census suburb level with a similar level of detail for all 190 census suburbs that constitute the City of Cape Town. A census suburb as defined by Statistics South Africa is "a spatial division into which the country is demarcated for the purpose of census enumeration, as well as to facilitate data processing and analysis" (Statistics South Africa, 2010). The demographic data includes information about population, age, adult education, employment status, number of households, average household size and dwelling type. Each information type is provided in separate spreadsheets for each census suburb and summary statistics are provided. Spatial data (i.e. GIS-related database) consists of a digital map of Cape Town and its 190 census suburbs as well as demographic data (population, household size, average household size and area) provided in the attribute table.

ii. Pedestrian crash data

The City of Cape Town, through the Transport and Urban Development Authority (TDA), provided data on pedestrian crashes that occurred in the entire city from January 2005 until December 2014. A set of 13 characteristics are available to describe each pedestrian casualty included in the database. These details include node description, Police station, crash date, day of week, time of crash occurrence, severity of injury, population group, gender, age, pedestrian position, pedestrian location, pedestrian manoeuvre, and pedestrian action. The data is recorded in Excel format and contain monthly summary statistics organised in different spreadsheets using Excel pivot table tools.

iii. <u>Data on the built environment</u>

Land use data

The City of Cape Town's Strategic Development Information and GIS Department supplied digital spatial data and maps of land use available within the City of Cape Town. The data is in the form of a GIS-related dataset containing both spatial data (identifying the geographic location of each land parcel) and attribute data (data in tabular format describing characteristics of each land parcel).

Data on transportation systems

Geospatial data on transportation systems of the City of Cape Town were obtained from the Strategic Development Information and GIS Department. The data consists of GIS layers of street network and public transport routes. More specifically, data on street network includes the overall street network of the city and GIS layers of the City's freeways, expressways, local distributors, primary arterials, secondary arterials and roads reclassified to secondary arterials. Data on public transport systems consists of railway lines, railway stations, bus routes, mini-bus taxi routes, routes and stations for integrated rapid transit (IRT).

Aerial photographs

Aerial photographs covering the whole area of the City of Cape Town were also obtained from the Strategic Development Information and GIS Department. The majority of these photographs were taken in 2013 and a small proportion of them were taken in 2014. They were provided in the Enhanced Compression Wavelet (ECW) format. Aerial photographs were used as to collect supplementary information on land uses and to gather more detailed information on transportation systems.

Online open data

Additional sources of information freely available online was used either as a supplement to the information provided by the City of Cape Town or where information is missing, found incorrect or available but requires verifications. A number of online sources accessed are summarised in the following sections.

Geographical data on SAPS boundaries

Geospatial data on boundaries of South African Police Service (SAPS) stations located within the City of Cape Town were downloaded from the website of the SAPS. The information included relates to the SAPS stations existing until the year 2015. SAPS boundaries were used in conjunction with information provided in crash dataset (node descriptions for example) to easily locate a pedestrian crash on the map.

Google images

Google images were used in conjunction with aerial photographs, especially when more detailed information was required. Two Google map-related products were used in this study: street view images were utilized to gather data on characteristics of existing transport facilities and on design features of crash locations. Google Maps were used to display satellite and street-level imagery. Google mapping service also assisted in locating pedestrian crashes and obtaining geographical coordinates (latitude and longitude) of targeted locations.

Online Zoning Viewer of the City of Cape Town

The Online Zoning Viewer is an open source containing information similar to that of the GIS layer of land use. The viewer allows the user to access and view details of all properties within the City of Cape Town (City of Cape Town, n.d.). The information in this open source database is more recent than the one in the GIS layer and this advantage facilitated the researcher to access information that was missing or incorrectly gathered in the dataset for land use. The user interface of the Online Zoning View can be viewed in Figure 3-9 presented later in Section 3.2.5 (on Page 89).

Online open sources for crash location identification

In circumstances where the name of the street provided in the crash dataset was not found using the Google search engine, an advanced search was undertaken by using other online databases such as www.geographic.org (Photius Coutsoukis and Information Technology Associates, n.d.). This online database was useful in instances where a pedestrian crash was reported using an old name of the street which cannot be found using the Google search engine or using the Google Maps tool. This case was frequently encountered while identifying crash locations in informal settlements of the city or neighbourhoods that have been recently upgraded from informal to formal settlements.

3.2.3 Data quality and limitations

3.2.3.1 Minimum data requirements

In general, road crash investigations require a comprehensive, detailed and up-to-date crash database as well as skilled personnel with the ability to conduct scientifically sound analyses and interpret the results (Ogden, 1996). A good database requires a set of accurate and comprehensive information related to road crashes enabling the analyst to undertake meaningful and statistically reliable analyses. For these purposes, prerequisite information such as crash data, facility data and traffic flow data are paramount (AASHTO, 2010).

Crash data encompasses records including narratives describing the overall aspects of the crash occurrence. Although the level of detail may differ from jurisdiction to jurisdiction, the most basic crash dataset should generally describe the following elements (AASHTO, 2010; Ogden, 1996; PIARC Technical Committee on Road Safety, 2007):

- Accurate crash location: this data element provides information on where the crash occurred by the use of a reference localization system including administrative entity number/name, geographical coordinates, node/link identifier, road name, kilometre post markers, road layout, type of control, and other pertinent road features.
- Date and time: information on when the crash occurred in terms of year, month, day of month, day of week and time of day.
- Information pertaining to people involved in the crash with their characteristics (e.g. age, gender, alcohol test and result, etc.), vehicles in the crash, animals or roadside objects.
- Crash consequences in terms of crash severity and casualty class: a crash may result in varying levels of injury (fatal, serious and slight) or property damage. Crash severity refers to a person while casualty class refers to the most severe injury sustained by any victim of the crash.
- Crash type classification: according to traffic movements of conflicting road users before the crash or (e.g. rear-end, head-on, sideswipe, angle, run off road, turning, etc.).
- Environmental conditions: lighting conditions, weather, pavement surface conditions, etc.
- Narratives of how the accident occurred and subjective contributory factors of the crash (if the coding system allows the inclusion).

Road facility data should describe the most basic physical features of the crash site such as roadway classification, geometric details (number of lanes, number of legs, presence of medians or refuges, shoulder width, curve, grade, sag, crest, etc.) type of control, speed limits, road surface conditions, etc. This information should enable a basic safety assessment which aims at identifying locations with an unacceptable level of crash rate within the road network, unveiling possible circumstances and factors contributing to the crash occurrence and guiding the development of appropriate countermeasures (AASHTO, 2010; PIARC Technical Committee on Road Safety, 2007).

Traffic flow data are common information required to estimate the potential probability of a traffic crash occurring. Traffic volumes are the most frequently used form of exposure measures and they are usually expressed in terms of average annual daily traffic (AADT). Other exposure measures which are used in crash investigations include total number of vehicles entering the intersection (EV), vehicle-kilometres travelled (VKT) and pedestrian volumes (AASHTO, 2010). Nevertheless, traffic flow data are not always available and this sometimes forces safety analysts to resort to using a range of proxy measures of exposure. Exposure data such as traffic volumes are usually obtained from other sources including transportation agencies or local road authorities.

To some extent, and more specifically for this study, supplementary data are desirable in certain circumstances for in-depth examination of particular variables and relationships. In general, minimum data requirements for this particular set of data depend on the nature and the purposes of the study. From the reviewed literature, studying the influence of the built environment on pedestrian safety at an area-wide scale has required a range of information related to population, vehicle fleet, transportation systems, land use, urban design and so on. In addition to police-reported crash data, information applicable to this type of study can be obtained from various sources such as reports from local authorities, hospitals, local knowledge of the area, interviews with road users, surveys, focus groups, traffic conflict studies, site investigations and so on.

3.2.3.2 Data deficiencies

1. **Socio-demographic data**

For certain census suburbs, the area is completely covered by an institution (e.g. hospital, learning institution), or a facility (e.g. airport). These census suburbs are inhabited by population mainly living in collective living quarters such as hotels, hostels, students' residences, hospitals, military camps, prisons and so forth. In these instances, data on population, household size and average household size is reported as zero in the socio-demographic data received from the Strategic Development Information and GIS Department (SDI & GIS), City of Cape Town. The same variables are also reported as zero for certain census suburbs regarded as industrial or commercial areas. This was the case for the following census suburbs: Cape Town International Airport, Epping Industria, Killarney Gardens, Silvermine, Tygerberg Hospital and University of Cape Town. A large number of the population living in collective living quarters or industrial/commercial areas led to inflated average household size (ranging from 7 to 170) for census suburbs such as University of Western Cape/Peninsula Technikon, College, Castle Rock, Bellville Teachers' Training, Stikland Hospital, and V & A Waterfront, among others.

2. Land use data

The dataset of land-use for the City of Cape Town comprises data on zonings and subzonings ascribed to each land parcel located within the boundaries of the City of Cape Town Metropolitan Municipality. All land parcels, buildings and structures within the area of jurisdiction of the City of Cape Town Metropolitan Municipality are designated for a particular development or land use category or zoning according to zoning scheme regulations of the city. According to bylaws for planning and building development, zoning (when used as a noun) is defined as a land use category identified by means of a specific notation and prescribed by rules regulating the purposes for which land may be used and rules governing how land may be developed for that particular land-use category (City of Cape Town, 2015b)

The zoning scheme for the City of Cape Town acknowledges nine main zonings which are (1) single residential; (2) general residential; (3) community; (4) local business; (5) general business and mixed use; (6) industrial; (7) utility, transport and national port; (8) open space and (9) agricultural, rural and limited use. Certain zonings such as general residential, general business and mixed use zonings and industrial zonings are subdivided into subzonings which are regulated by different development rules. Detailed information about different zonings and subzonings adopted in the City of Cape Town is found in the planning bylaws of the City of Cape Town (City of Cape Town, 2015b). A list of zonings and subzonings adopted in the zoning scheme of the City of Cape Town are provided in Table 3-2 and descriptions of different zonings and subzonings are provided in APPENDIX A. The zonings and subzonings are considered in this study as types of land use.

Table 3-2: Descriptions of zonings and subzonings applicable to the City of Cape Town (City of Cape Town, 2015b)

1. Single residential zonings
Single residential zoning 1: Conventional housing (SR1)
Single residential zoning 2: Incremental housing (SR2)
2. General residential zonings
General residential subzoning 1: Group housing (GR1)
General residential subzonings (GR2, GR3, GR4,GR5 & GR6)
3. Community zonings
Community zoning 1: Local (CO1)
Community zoning 2: Regional (CO2)
4. Local business zonings
Local business zonings 1: Intermediate business (LB1)
Local business zoning 2: Local business (LB2)
5. General business and mixed use zonings
General business subzonings (GB1, GB2, GB3, GB4, GB5, GB6 & GB7)
Mixed use subzoning (MU1, MU2 & MU3)
6. Industrial zonings
General industrial subzonings (GI1 & GI2)
Risk industry zoning (RI)
7. Utility, transport and national zonings
Utility zonings (UT)
Transport zoning 1:Transport Use (TR1)
Transport zonings 2 : Public road and public parking (TR2)
National port zoning (NP)
8. Open space zonings
Open space zoning 1: Environmental conservation (OS1)
Open space zoning 2: Public open space (OS2)
Open space zoning 3: Special open space (OS3)
9. Agricultural, rural and limited use zonings
Agricultural zoning (AG)
Rural zoning (RU)
Limited use zoning (LU)

The dataset of land use includes 751 128 land parcels designated for various developments or land use categories. Each land parcel is spatially represented by a polygon and its dimensions (length and area) are provided in the attribute table. Of the 751 128 land parcels, land-use types or zonings were not yet assigned for 14 716 land parcels which constitute nearly 2 percent of all land parcels (see APPENDIX B).

3. Data on transportation systems

Data on transportation systems in the City of Cape Town consists of 13 separate GIS layers of street network and public transport routes and stations. The GIS layers include a layer for a national coverage of street network, layers for five road classes (i.e. freeways, expressways, local distributors, primary arterials, secondary arterials) available in the city with the exception of the access road class, a layer of minibus-taxi routes, a layer for bus routes (Golden Arrow Bus Services), a layer for Integrated Rapid Transit (IRT) bus routes, a layer for IRT bus stops, a layer for railway lines and a layer for railway stations.

The attribute table for the street network layer contains a range of items for each road segment including posted speed limit, road name, class, and the length of segment, to name a few. Attribute tables of route layers for public transport systems contain items such as route name, origin, destination and length for each segment. For IRT and bus stations, the attribute tables contain information such as the name of the station, the identification number (if available), street name, shelter type, and so on. The name of the railway station is the only information provided in the attribute table of the railway stations layer. The spatial data of transportation systems was obtained in March 2014 and is the most recent data available to the researcher. As a result, the data set does not include upgrades or new developments that occurred after the reception of the data. The completeness of the datasets is provided in APPENDIX B.

4. Pedestrian casualty data

Pedestrian casualty data was obtained in MS Excel format and this consists of information on police-reported traffic crashes involving pedestrians in the Western Cape Province from the 2005-2014 period. The datasets comprises 73 785 pedestrian casualty records (each record holds information concerning one specific pedestrian casualty). Of these pedestrian casualties, 534 cases were found to have been duplicated and the removal of duplicated cases led to a dataset of 73 251 pedestrian casualties. Pedestrian casualty records are organized in Excel worksheet rows while different characteristics of pedestrian causalities are arranged in columns. The crash worksheet contains 13 field names (i.e. column headings) coded as follows: node description, Police station, crash date, day of the week, time of crash occurrence, severity of injury, population group, gender, age, pedestrian position, pedestrian location, pedestrian manoeuvre, and pedestrian action. Data description and the quality of information provided for different variables are shown in Table 3-3.

Table 3-3: Quality of pedestrian casualty data in the Western Cape: Period 2005-2014

Data fields	Number of variables	Missing variables	Unknown variable	Total	% Completeness
Node description	73243	8	0	73251	100.0
Police station	73251	0	0	73251	100.0
Accident date	73251	0	0	73251	100.0
Day of the week	73251	0	0	73251	100.0
Time	73251	0	0	73251	100.0
Injury severity	67439	0	5812	73251	92.1
Population Group	55063	0	18188	73251	75.2
Gender	54500	0	18751	73251	74.4
Pedestrian position	41934	0	31317	73251	57.2
Pedestrian location	40270	0	32981	73251	55.0
Pedestrian manoeuvre	40441	0	32810	73251	55.2
Pedestrian action	42054	0	31197	73251	57.4

Of the 73 251 pedestrian casualties recorded in the Western Cape Province, 54 744 pedestrian casualties occurred within the boundaries of the City of Cape Town for a 10-year period extending from 2005 to 2014. Data description of the 10-year casualty data in Cape Town and the quality of casualty records are provided in Table 3-4.

Table 3-4: Quality of pedestrian casualty data in the Cape Town area: Period 2005-2014

Data fields	Number of variables	Missing variables	Unknown variable	Total	% Completeness
Node description	54742	2	0	54744	100.0
Police station	54744	0	0	54744	100.0
Accident date	54744	0	0	54744	100.0
Day of the week	54744	0	0	54744	100.0
Time	54744	0	0	54744	100.0
Injury severity	49642	0	5102	54744	90.7
Population Group	40103	0	14641	54744	73.3
Gender	39979	0	14765	54744	73.0
Pedestrian position	27147	0	27597	54744	49.6
Pedestrian location	25898	0	28846	54744	47.3
Pedestrian manoeuvre	26493	0	28251	54744	48.4
Pedestrian action	27740	0	27004	54744	50.7

A comparison of data completeness from Table 3-3 and Table 3-4 shows that the casualty dataset for the entire Western Cape Province exhibits a higher level of data completeness than that of the City of Cape Town.

Pedestrian crash locations are described by referring to street intersections. Names of intersecting streets are used to describe a node. Although the Excel column is labelled "node description", the descriptions provided include both street nodes and mid-block or link locations. In some cases, crash locations are described by using further descriptions such as kilometre marker (although this is rare), suburb name, type of facility (e.g. circle, interchange bridge, railway bridge, pedestrian bridge, transit terminal), a name of a popular property or business (e.g. estate name, supermarket, petrol station etc.) or the nearest cross streets or nodes for non-intersection locations. Symbols are also used as a way to describe a crash location. For example, the node of Imperial Street and Alpine Street in the suburb of Mitchells Plain is described by the following notation: "IMPERIAL STRX ALPINE STR, MITCHELLS PLAIN". A link or midblock location on R300, between Stock road and New Eisleben road is denoted by: "R300, STOCK RD//NEW EISLEBEN RD". The latter notation is sometimes utilized when describing a crash location between two adjacent suburbs, like in the following example: "KLOOF RD, SEA POINT//CAMPS BAY".

The name of the local South African Police Service (SAPS) station whose officers attended and reported a crash is provided for every pedestrian injury. The column of injury severity consists of five categories, which are, "Killed", "Serious", "Slight", "No injury" and "Unknown". The population group column contains ethnic groups of injured pedestrians, which are recorded as "Black", "White", and "Coloured", "Asian", "Other and "Unknown". Pedestrian position prior to a crash is described in terms of six locations: "Median", "Roadway", "Shoulder of the road", "Sidewalk/verge", "Other", and "Unknown". The pedestrian location column contains information regarding the use of formal crossing points by the pedestrian and this information is recorded in six notation categories: "At crossing", "Jaywalking", "Not at crossing", Within 50 m from", "Somewhere" and "Unknown". The pedestrian manoeuvre column indicates the pedestrian movement direction in relation to moving traffic and this is denoted as "Back to traffic", "Crossing road", Facing traffic", "Other" and "Unknown". The last column of the dataset includes pedestrian actions during a crash and these are categorized in ten records: "None", "Lying down", "Playing", "Running", "Sitting", "Standing", "Walking", "Working", "Other" and "Unknown".

Nearly 50 percent of the information is recorded as "Unknown" in the columns labelled as pedestrian position, pedestrian location, pedestrian manoeuvre and pedestrian action. Pedestrian gender and ethnicity is also recorded as unknown in nearly a quarter of reported pedestrian injuries. Injury severity is unknown in about 8 percent of recorded injuries and the column of node descriptions contains 8 blank cells. Regarding the age of injured pedestrian, "0 age" is reported for 38 522 pedestrians, representing 52.6 percent of all reported pedestrian casualties. Ages deemed unrealistic (ages greater than 100 years old) are also reported for 137 pedestrian casualties as shown in Table 3-5. Of these 137 casualties, age greater than 120 years old was observed in 16 cases.

Table 3-5: Incorrect age records in the pedestrian casualty dataset

Age outliers	Cape Town 2012	Cape Town 2013	Cape Town 2014	Cape Town 2012-2014	Western Cape 2005-2014
0 Age	2806	2862	2651	8319	38522
Age>100	0	1	1	2	137

The above-mentioned data quality deficiencies are a component of crash data limitations and are extensively reviewed in the literature review. These limitations and their potential sources are summarized in Figure 3-5. Road safety assessment in this study was carried out with awareness that crash records used as source of information are inevitably imperfect. However, historical crash records still remain the main source of information on crash occurrence in South Africa as well as in many other countries. Furthermore, they are the only source of information on crash events which is available for the study area in question.

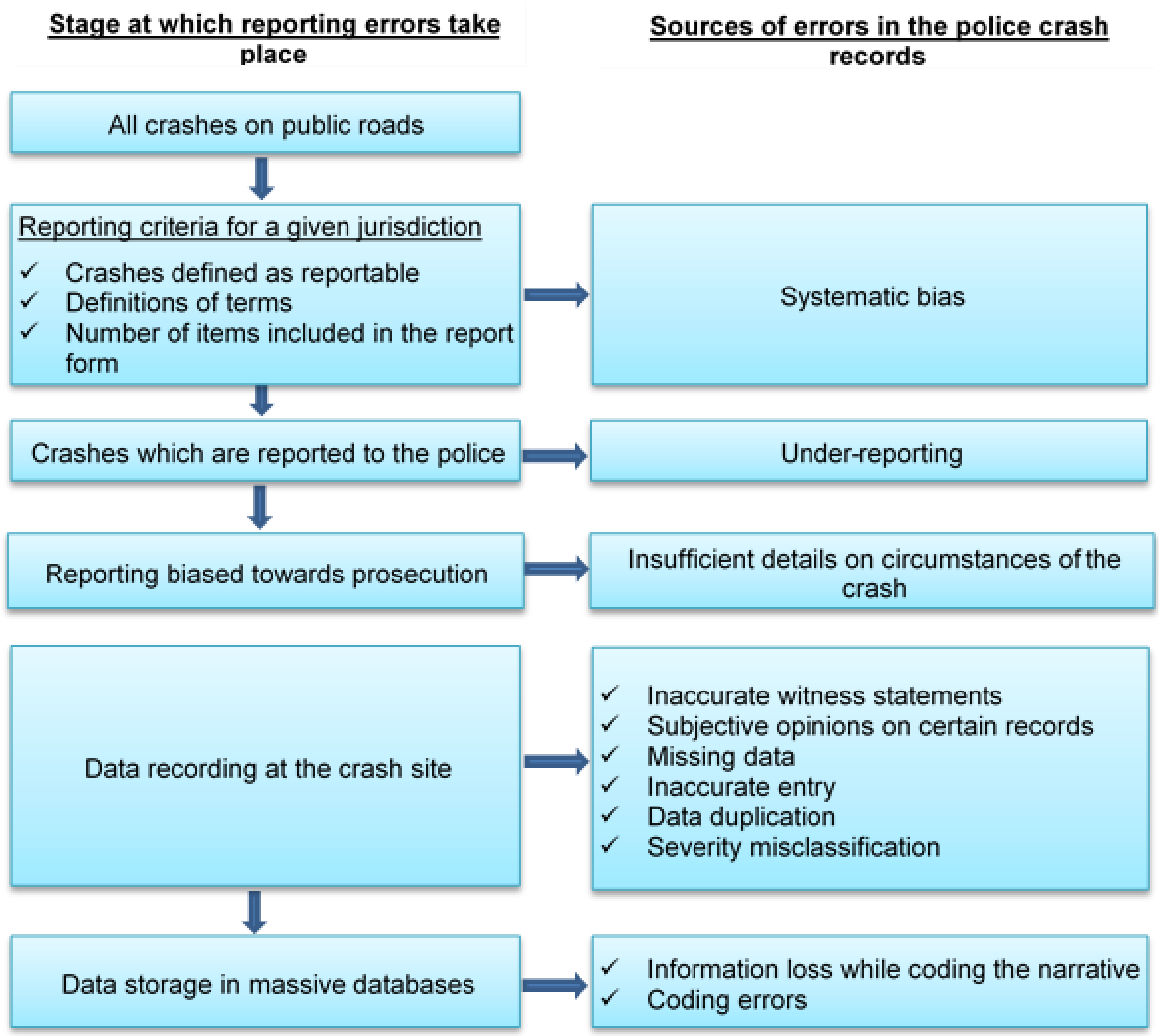


Figure 3-5: Data quality deficiencies associated with historical crash records (adopted from Elvik *et al*. (2009)).

3.2.4 Sampling of pedestrian casualty data

The target population for this study are pedestrians involved in road traffic crashes in the urban spaces. The City of Cape Town Metropolitan Municipality (CPT) was chosen as the study area for this study. The choice of the study area was based on the scale of safety concerns affecting pedestrians in this municipality. According to many reports and studies, more than half of road traffic deaths on the city's road network affect pedestrians, a figure which is above the average value for the Western Cape Province and for the whole country (Mitullah *et al*., 2017; Provincial Government of The Western Cape, 2007). Furthermore, the choice was motivated by a number of subjective reasons such as spatial proximity to the University of Stellenbosch at which the current study was conducted, the ease for the researcher to access data for the study, familiarity with the study area and personal interest and connection with the study area.

The dataset of pedestrian casualties comprises 73 251 pedestrian casualties recorded within the Western Cape Province for the time period extending from 2005 to 2014. Since the study could not include all pedestrian casualties observed over a 10-year period, there was a need to select an appropriate sample size representing the target population of pedestrian casualties. In fact, crash investigations carried out over long periods of study hold a main advantage of being statistically significant ensuring that the limitations due to natural fluctuation of crash frequencies are minimized.

In practice, five years of observed crash data is the most suitable analysis period for statistical reliability (Nicholson, 1987). However, shorter analysis periods ranging from 1 to 3 years are very common in research and practice especially when early identification of immediate safety problems is the target (Ogden, 1996). Crash investigations over short periods are also considered in instances where significant changes in explanatory variables may take place within the study period. As a rule of thumb, a study period of three to five years is recommended (Golembiewski & Chandler, 2011).

With respect to the recommendations above-mentioned, a decision was taken to focus on a three-year crash period, starting from 2012 to 2014. The choice of this study period emerged from a number of reasons: Firstly, reported crashes in this period were the most recent at the time they were obtained (year 2015). Secondly, the time frame corresponds to the date range of the aerial photographs (captured in year 2013) and the majority of imagery available on Google Street View (captured from 2009 to 2015) for the City of Cape Town at the time the data was collected and processed. Thirdly, it was chosen due to considerations of potential within-period variation in explanatory variables. There have been some changes in the transportations system of the City of Cape Town, notably the expansion of the bus rapid transit (BRT) system. Accompanying this is a rapid population growth mainly as a result of internal and international migration inflows, requiring new land development in the city. Lastly, the study faced a serious challenge of data quality which required significant effort on the part of the researcher to undertake remedial strategies to minimize the negative impact on the study's results and inferences. Collecting additional information on crash locations for a large metropole like Cape Town and treating missing data in the land-use dataset was a laborious process which was constrained by time limitations

For the 2012-2014 study period, there were 13 868 pedestrians involved in road traffic crashes in the City of Cape Town. Of these casualties, the crash location could not be identified for only 15 cases due to a vague description of a node or mid-block (e.g. “Cape Town”), a crash location was reported as “unknown” and a description using a street name was given which could not be identified on Google Maps and other search engines. The removal of these 15 casualties led to a final sample of 13 853 pedestrian casualties. Of these casualties, 4 672 (33.72 percent) were reported to have occurred in 2012; 4 529 (32.70 percent) were reported to have occurred in 2013 and 4 652 (33.58 percent) were reported to have occurred in 2014.

To evaluate whether the sample size of 13,853 pedestrian casualties is big enough to enable sound statistical results and appropriate inferences, the technique of statistical power analysis was performed in STATISTICA. The objectives of this technique was first to evaluate the smallest sample size required to detect effect at the desired level of confidence (Field, 2013; Murphy *et al*., 2009). Secondly, given the time constraints and laborious processes involved in crash data preparation, the technique helped the researcher to select an appropriate sample size to avoid the selection of a larger sample that could be costlier for minimal gain.

Sample size calculation usually makes use of information on the entire population to determine the size of a small subset of the population that can be used to draw inferences about the whole population. As explained previously in this section, it is often inappropriate to use crash data collected over long analysis periods. For this reason, an entire statistical population of road traffic crashes could not be used in this study to determine an appropriate sample size of pedestrian crashes to use in the analysis. Nevertheless, it is still crucial to evaluate the sample size that will have a statistical effect and to determine the size of that effect. The ability of a test to detect an effect is known as its statistical power (Field, 2013).

The inferential statistics distinguishes two types of error that can arise while testing statistical hypotheses: a Type I and a Type II error (Field, 2013; Montgomery & Runger, 2014). The Type I error rejects the null hypothesis H_0 when in fact it is true, while the Type II error fails to reject the null hypothesis H_0 when in reality it is false (Montgomery & Runger, 2014). Decisions in hypothesis testing are often explained using a statistical decision matrix illustrated in Table 3-6.

Table 3-6: Statistical Test Decision Matrix (Rossi, 2012)

	True State of the Population	
Test Decision	Effect Absent H_0 is True	Effect Present H_0 is False
Test result: $p < \alpha$ Test decision: Reject H_0 Conclusion: “Effect present"	Type I Error $p = \alpha$	Power $p= 1-\beta$
Test result: $p \geq \alpha$ Test decision: Do not reject H_0 Conclusion: “Effect absent"	Correct decision $p= 1-\alpha$	Type II Error $p = \beta$

Where α is the probability of a Type I Error and β is the probability of a Type II error.

$$\alpha = P(Type\ I\ error) = P(reject\ H_0\ when\ H_0\ is\ true)$$
$$\beta = P(Type\ II\ error) = P(fail\ to\ reject\ H_0\ when\ H_0\ is\ false)$$

In statistical power analysis, the power of a statistical test is defined as the probability of rejecting the null hypothesis H_0 when in fact it is false (i.e. when the alternative hypothesis is true)(Cohen, 1992a; Montgomery & Runger, 2014). Put in another way, it is the measure of how good a test is or the probability that a given test will find an effect assuming that one exists in the population (Field, 2013). The technique of statistical power analysis mainly deals with Type II error. As illustrated in Table 3-6, the power is estimated as $1 - \beta$ and can be interpreted as the probability that a statistical test will correctly reject a false null hypothesis (Montgomery & Runger, 2014; Rossi, 2012). Cohen (1992a) suggested that the maximum acceptable probability of a Type II error (i.e.β-level) would be 0.2 (or 20 percent). This implies that ideally the power of a statistical test $(1 - \beta)$ should be at least 0.80 to detect a reasonable effect.

Simply put, the power of a statistical test is the probability of obtaining a statistically significant result and the power depends on the significance criterion (α) , the sample size (N), and the population effect size (ES) (Cohen, 1992a). Conventionally, the probability of a Type I error (α-level) is 0.05 (Cowles & Davis, 1982; Nickerson, 2000). By setting the value of α-level and the desired power $(1 - \beta)$, it is possible to estimate the necessary sample size that will have a statistically significant effect. The null hypothesis tested is that there is no effect in the population as illustrated in Table 3-6.

Table 3-7: Summary output of the power and sample size calculation

	Sample Size Calculation One Proportion, Z, Chi-Square Test H_0: Pi = Pi_0
	Value
Null Proportion (Pi0)	0.5000
Population Proportion (Pi)	0.5138
Alpha (Nominal)	0.0500
Actual Alpha (Exact)	0.0501
Power Goal	0.9000
Actual Power (Normal Approx.)	0.8999
Actual Power (Exact)	0.9000
Required Sample Size (N)	13783

The statistical power analysis was carried out in STATISTICA software tool. By setting the α-level at 0.05 and the desired power at 0.90 (i.e. 90 percent chance of detecting an effect if one genuinely exists), the required sample size was found to be 13 783 pedestrian casualties which is slightly smaller than the actual sample size used in this study (i.e. 13 853 pedestrian casualties). This finding suggests that the power for the sample size of 13 853 pedestrian casualties is slightly greater than 0.90. Furthermore, it can be observed on the plot of power against sample size illustrated in Figure 3-6 that the sample size required to achieve the power of 0.80 (considered as the minimum acceptable level) is smaller than the actual sample size used in this study. Therefore, from a statistical point of view, the sample size of 13 783 pedestrian casualties is sufficiently large to produce accurate and reliable inferences.

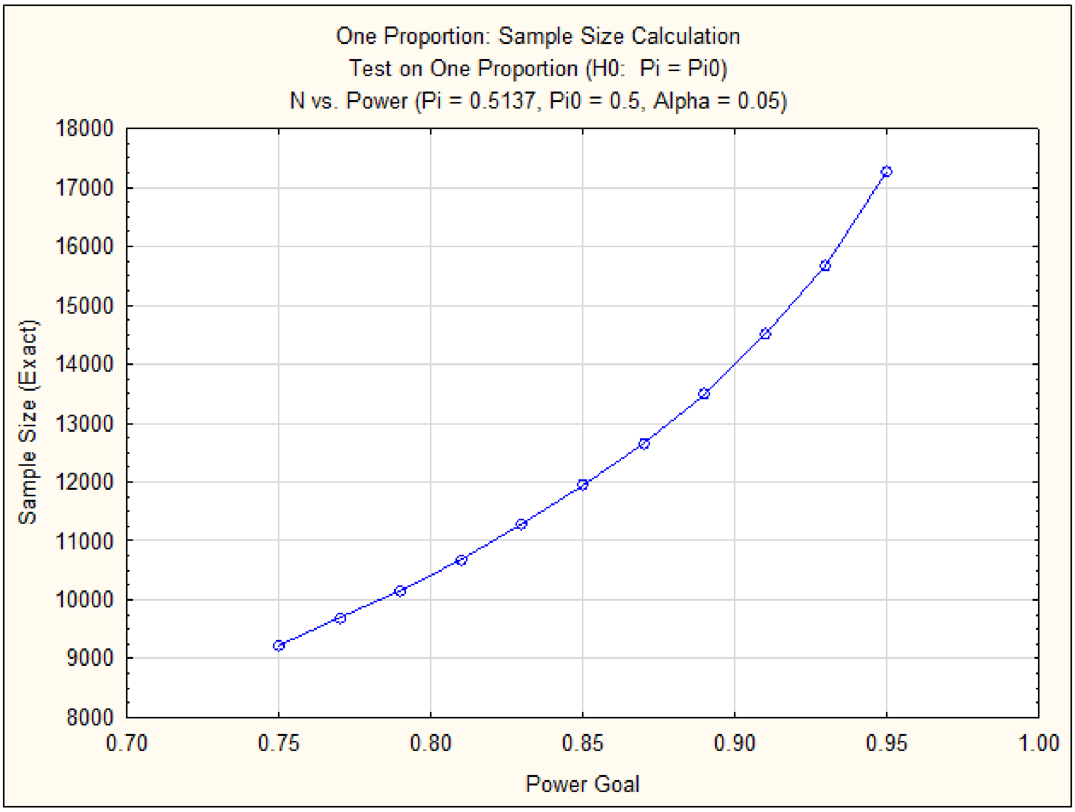


Figure 3-6: Plot of power versus sample size

Regarding other datasets (i.e. land-use data, transportation system data and population census data), sampling was not necessary since the data was collected for the entire target population, that is, the target study area.

3.2.5 Data processing

Data processing in this study denotes the process of manipulating raw data (both primary and secondary data) with the intention of converting it into meaningful information which is easier to analyse. In general, data processing involves 4 steps which are (1) raw data aggregation; (2) data screening; (3) data transformation and (4) data organization for statistical analysis. Raw data aggregation involves extracting data for a particular unit of analysis from the main dataset. Data screening in this study denotes a process of identifying and correcting errors or abnormalities in the dataset prior to performing data analysis. The main intention of this process is to minimize the negative impacts on the results of the study. Data transformation entailed procedures to convert variables into new variables through computation. As an example, variables “population number” and “size of the area” were converted into the variable “population density” through the process of data transformation. The last step of “data organization for statistical analysis" involved data summation and ordering data into databases to facilitate analysis.

3.2.5.1 Processing data on land use

1. **Raw data aggregation**

The objective of processing data for land use patterns was to acquire land zoning features (or land use categories) aggregated at suburb level. Zoning features for a given suburb were extracted from the overall zoning spatial data in ArcMap and the attribute table of extracted features was copied into an Excel spreadsheet for further processing. This procedure was performed in 28 steps grouped into four stages: (i) Selecting a census suburb from a shapefile of the 2011 census suburbs of the City of Cape Town; (ii) extracting a layer of the suburb from the shapefile of the 2011 census suburbs; (iii) selecting zoning features contained within the boundaries of the suburb and (iv) creating a layer of zoning features for the suburb.

2. **Data screening**

Screening data on land use patterns was performed in four main steps: (i) selecting railway lines and public road coded as "TR1" and "TR2" from the zoning dataset; (ii) removing "TR1" and "TR2" from the zoning dataset; (iii) excluding zoning features which are not contained within the boundaries of the suburb and (iv) treating missing data in the zoning dataset. These steps can be viewed on a flowchart presented in Figure 3-7.

As mentioned in previous sections, zoning data consists of characteristics of land which is used for diverse purposes. Land designated for transport purposes is classified into two zoning categories: "TR1" and "TR2". The "TR1" category includes land used for railway lines and bus routes, while the "TR2" category comprises land dedicated for public streets and roads among other use. However, these types of land use (railway lines, bus routes, public streets and roads) are already encapsulated in the transportation system datasets in the form of line segments. Accordingly, it was decided to remove them from the zoning data to avoid data duplication. Nevertheless, the land used for transport purposes, other than those mentioned above, were kept in the zoning dataset since they were not included in the transportation system categories. This was land used for (a) airports, (b) harbours, (c) public transport terminals, ranks and holding areas, (d) cable car stations and (e) premises for public parking.

In general, extracting features for a unit of analysis in ArcMap requires the use of the "Select by Location" tool/option. Among various methods supported in ArcMap for spatial selection, two methods, "intersect the source layer feature" and "are within the source layer feature", appeared to be the most appropriate for the purpose of the procedure. The spatial query "intersect" allows the selection of features which either fully or partially overlaps the source layer while "are within" is a spatial query allowing the selection of features completely contained inside the geometry of the source layer (ESRI, n.d.). Initially, the spatial query "are within" was expected to provide the best selection results but the method failed to include zoning features that touch the boundaries of the suburb (i.e. source layer). An example of the selection outcome for both the "intersect" and "are within" spatial queries for the suburb of Observatory is illustrated in Figure 3-8.

It can be observed from Figure 3-8 (on Page 88) that the zoning selection by the "are within" spatial query results in missing zoning features (mostly those that touch the boundaries of the suburb) which need to be identified and then added to the dataset of extracted features for the suburb. On the other hand, the zoning selection by the "intersect" spatial query does include a number of zoning features extending outside the boundaries of the suburb, i.e. zoning features contained within the boundaries of adjacent suburbs. For these reasons, both selection methods are susceptible to introducing errors in the analysis if no remedial measure is taken ascertain that zoning features are completely contained inside the suburb boundaries. Two options of remedial measures were anticipated; (1) using "intersect" as a spatial query, identifying zoning features extending outside the boundaries of the suburb and removing them manually or (2) using "are within" spatial query, identifying missing zoning features and adding them to the dataset of extracted zoning features for the suburb. The evaluation of the remedial measures qualified the use of "intersect" as the most appropriate spatial query and this approach introduced a third stage of data processing cited earlier in this section.

The fourth stage of data processing entailed the treatment of missing variables in the individual dataset of extracted zoning features for a given suburb. For the entire City of Cape Town, the type of zoning was not assigned for 14 716 land parcels out of a total of 751 128 land parcels (i.e. approximately 2 percent of all land parcels) included in the overall zoning dataset. As such there was a need for a strategy to deal with the incompleteness of the spatial dataset of land use. Attempts to obtain a more complete land use dataset were unsuccessful and a decision to obtain information on missing data from other sources was considered. The Online Zoning

Viewer appeared to be the most complete and updated land use dataset for the City of Cape Town.

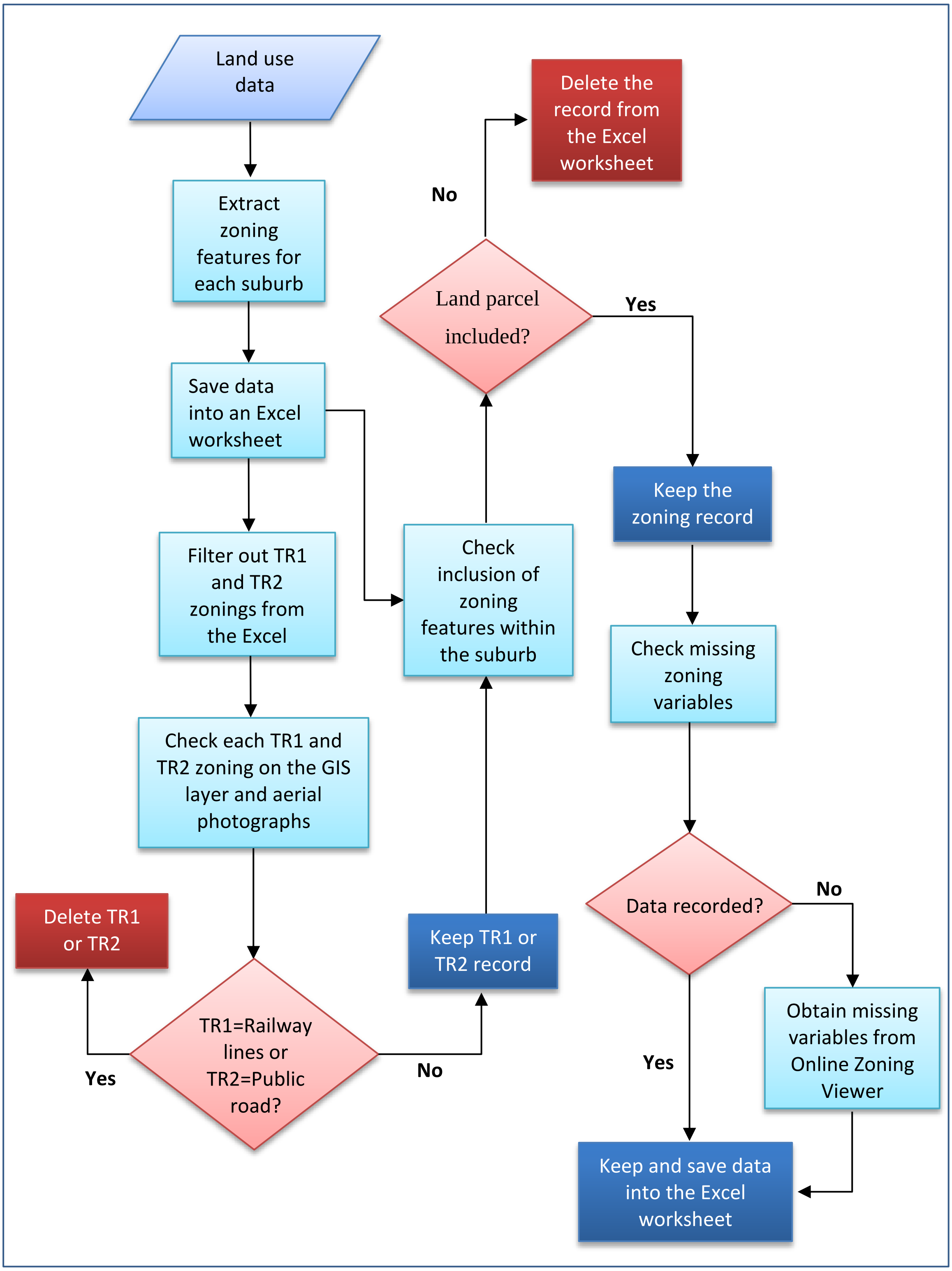


Figure 3-7: Aggregating and screening processes for land use data

The Zoning Viewer is an online tool for viewing maps and accessing detailed information on individual properties. The zoning viewer of the City of Cape Town consists of a toolbar including functionalities such as zoom in, zoom out, pan, zoom full extent, identify mode and clear map graphics. The planning viewer area comprises a number of features including layers, legend, search, print and measure. Two layers are available, one for base data and another for aerial photographs. The legend provides a list of symbols and colour codes used for the identification of different zonings or land uses as well as other features displayed in the application. The search function allows four searching options; by property number, by farm number, by street number and by ward number.

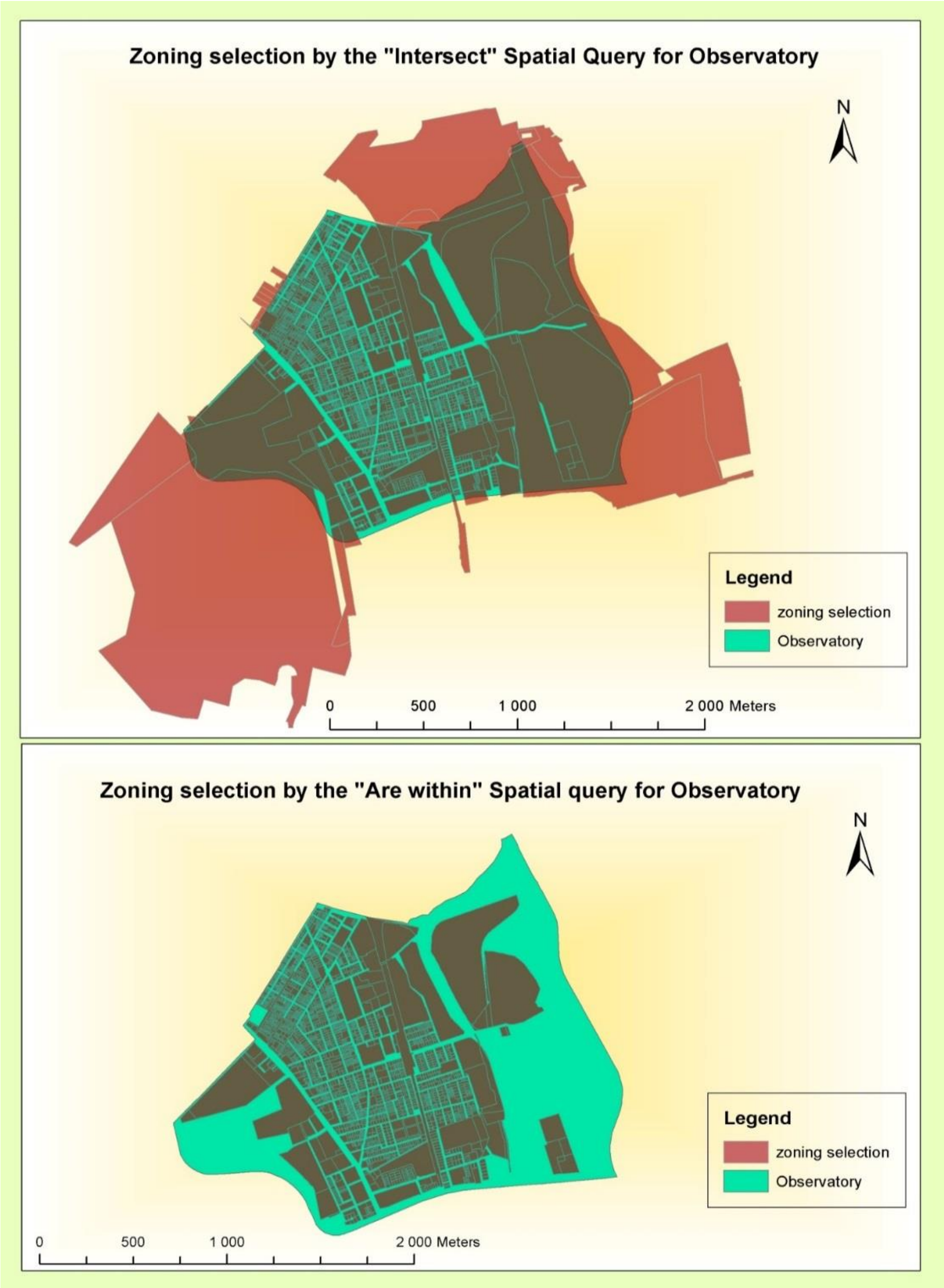


Figure 3-8: Differences in spatial selection methods: "Intersect" versus "Are within" spatial queries

The measure function enables three different measuring options: area measurement, distance measurement and measurement by means of geographical coordinates (longitude and latitude). Base data displayed in the map area includes detailed information (zoning description, ward number, subcouncil, street address and property number) for each individual property, boundaries for properties, wards, sub councils and centrelines of public and private roads. The user interface of the Online Zoning Viewer for the City of Cape Town is illustrated in Figure 3-9.

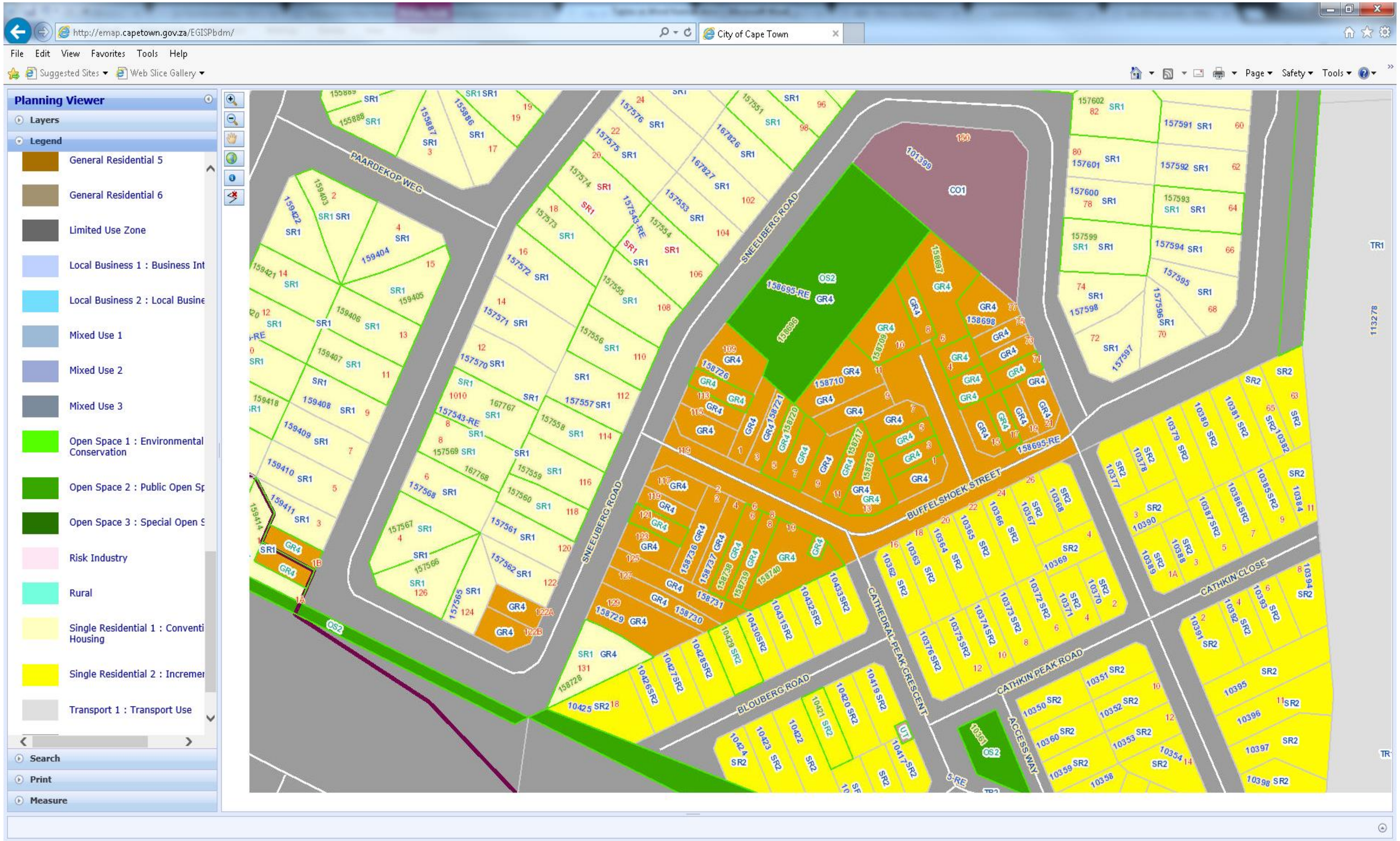


Figure 3-9: User interface of the Online Zoning Viewer for the City of Cape Town

3. Data transformation

With reference to the reviewed measures of land use interaction, this study included two categories of land use measure; (i) intensity-based measure expressed as a proportion of a particular land use to total land area of the unit of analysis and (ii) pattern-based measures expressed in terms of entropy index (ENT) and Herfindahl-Hirschman Index (HHI). The inclusion of these two measures was guided by recommendations from the study by Song *et al*. (2013) who underlined three contextual factors governing the choice of the measure: the number of land use types incorporated in the measure, the relative ease of calculation and the scale of the unit of analysis. In addition to these recommendations, there are research claims that these two measures are strongly correlated and can be hence used interchangeably as an indicator of mixed land use Song *et al*. (2013). For these reasons, it was decided to include these measures into the current study to test the measure that seems to be the most appropriate for the context of the study area.

In contrast to the intensity-based measure applicable to this study, land use mix could not be determined directly from the zoning data received from the City of Cape Town. Quantifying land use mix required a transformation of land use data which was obtained after the fourth stage of data processing (i.e. treating missing variables) into indices defined below.

i. <u>Entropy Index</u>

The entropy index is defined as a measure of variation, dispersion or diversity which takes into account the relative percentage of two or more types of land use within an area (Turner *et al*., 2001). The Entropy Index is the most widely accepted and commonly used measure of land use mix (Bordoloi, Mote, Sarkar & Mallikarjuna, 2013; Gehrke & Clifton, 2015; Manaugh & Kreider, 2013; Yue, Zhuang, Yeh, Xie, Ma & Li, 2017) and is generally calculated using the following equation:

$$ENT = \frac{-\left[\sum_{j=1}^{k} P_j \ln\left(P_j\right)\right]}{\ln(k)} \tag{1}$$

Where P_j denotes the percentage of each land use j in the area and k is the number of different land use types in the area. Assuming that:

C =Area $(m^2)$of land designated for commercial purposes

R =Area $(m^2)$of land designated for residential purposes

O =Area $(m^2)$of land designated for office purposes

T =Total area $(m^2)$of land use in a region

The entropy index for these types of land use is calculated as follows:

$$ENT = \frac{-\left[\left(\frac{C}{T} \times \ln\left(\frac{C}{T}\right)\right) + \left(\frac{R}{T} \times \ln\left(\frac{R}{T}\right)\right) + \left(\frac{O}{T} \times \ln\left(\frac{O}{T}\right)\right)\right]}{\ln(3)} \quad (2)$$

The entropy index ranges from 0 to 1, with 0 implying homogenous land use or a single use while 1 implies a perfect mix of land use types (i.e. all types of land use are equally distributed) within a unit of analysis. Three percentiles were used to rank the Entropy index following common practice from existing literature and other studies (Surjono & Ridhoni, 2017). These percentiles are entropy indices between 0-0.33, 0.34-0.66 and 0.67-1 and were classified into low, medium and high, respectively.

ii. Herfindahl-Hirschman Index (HHI)

The Herfindahl-Hirschman Index (HHI) originated in the field of economics and was initially used to measure market concentration (Ordover *et al*., 1982). The US Department of Justice (DOJ) ranks market concentrations by the use of the HHI: below 1500 the market is unconcentrated, between 1500 and 2500 the market is moderately concentrated and above 2500 the market rated to be highly concentrated (Hisrich & Kralik, 2016). When applied to quantifying a mixture of land use, the HHI is simply the sum of squares of the share of each individual land use type in a particular area. This definition is expressed mathematically by the following formula:

$$HHI = \sum_{j=1}^{k} (100 \times P_j)^2 \quad (3)$$

Where P_j denotes the percentage of each land use j in the area and k is the number of different land use types in the area The HHI ranges from 0 to 10000, with 0 denoting a perfect mix of land use types while 10 000 implies a single land use.

iii. Addressing limitations associated with measuring land use mix

Although the entropy index and the Herfindahl-Hirschman index are relatively easy to calculate and comprehend, these indicators are subjected to a number of inherent drawbacks (Manaugh & Kreider, 2013; Song *et al*., 2013). First, they are sensitive to the size of the geographic entity under analysis. Large areas tend to have a wider variety of land use types than smaller ones, implying that scores obtained by these two indicators may be influenced by the scale of the unit of analysis (Dark & Bram, 2007; Song *et al.*, 2013). Second, they fail to capture the arrangement or spatial separation of land use types within the unit of analysis (Manaugh & Kreider, 2013; Song *et al.*, 2013). Third, these two measures do not distinguish types of land use under analysis (Manaugh & Kreider, 2013; Song *et al.*, 2013). As an example, an area with 55% residential use, 35% commercial use and 10% office use will produce the same scores of HHI and entropy index as an area with 10% agricultural use, 35 % residential use and 55% commercial use. Lastly, as the HHI is the sum of squares of individual percentages of land use types, it is sensitive to the size of the most prevalent land use (Song *et al.*, 2013).

To compensate for these drawbacks, this study adopted an approach proposed in the study by Song *et al*. (2013). Using this approach, the geography of reference is considered as well balanced and the percentage of individual land use type within the unit of analysis is determined relative to that of the same land use type in the reference geography. It follows then that a smaller area with a land use distribution similar to that of the reference geography will score an entropy index close to 1 while an area with a land use distribution deviating substantially from that of the reference geography will score an entropy measure close to 0 (Song *et al.*, 2013). By extending this approach to HHI, an area with land use distribution similar to that of the reference geography will score a HHI close to 0, while an area with land use distribution deviating substantially from that of the reference geography will score a HHI close to 10,000.

The modified formula proposed by Song *et al*. (2013) was applied to calculate the percentage of each land use type (P_j) relative the distribution of the same land use type within the reference geography.

Assuming that:

- Each land use type available in the geography of reference is denoted by j
- Each unit of analysis is denoted by i
- The total amount of land use types j in the geography of reference is denoted by Z
- The amount of each land use j in the unit of analysis i is xij
- The total amount of all land use types j in the unit of analysis i is denoted by Xi
- The amount of each land use type j in the geography of reference is denoted by Xj
- The total number of units of analysis is denoted by n

- The total number of all land use types available in the geography of reference is K

Mathematically, it follows that:

- The total area of land use types in a given unit of analysis is expressed as:

$$\sum_{j=1}^{k} x_{ij} = X_i \tag{4}$$

- The total area of a particular land use type within the geography of reference is expressed as:

$$\sum_{i=1}^{n} x_{ij} = X_j \tag{5}$$

- The total area of all land use types within the geography of reference can be expressed as:

$$\sum_{i=1}^{n} X_i = \sum_{j=1}^{k} X_j = Z \tag{6}$$

- The percentage of a particular land use type in the unit of analysis is:

$$r_{ij} = \frac{x_{ij}}{X_i} \tag{7}$$

The sum of these percentages in each unit of analysis should be equal to 1:

$$\sum_{j=1}^{k} r_{ij} = 1 \tag{8}$$

The percentage of each land use type in a particular unit of analysis relative to reference geography is expressed in two steps:

i. By creating quotients q_{ij}:

$$q_{ij} = \frac{r_{ij}}{t_j} \tag{9}$$

These quotients are no longer percentages and can take any value less or greater than 1. For each unit of analysis the sum of these quotients is expressed as:

$$\sum_{j=1}^{k} q_{ij} \tag{10}$$

ii. By creating adapted land use percentages P_{ij} which will be used in the mathematical expression to calculate the entropy index and the HHI:

$$P_{ij} = \frac{q_{ij}}{\sum_{j=1}^{k} q_{ij}} \tag{11}$$

In this study, the entropy index proposed by Song *et al.* (2013) is referred to as the "Relative Entropy" (coded as "RelENT"). The reference geography is the Cape Town Metropolitan Municipality and the unit of analysis is a census suburb. Accordingly, the number of all census suburbs of the study area is 190 ($n = 190$). The two forms of entropy index (i.e. ENT and RelENT) and the HHI were determined for the entire City of Cape Town and individual census suburbs which constitute the City of Cape Town. Initially, land use mix was measured using all 34 zoning and subzoning categories available in the City of Cape Town. In an effort to achieve the most appropriate measure of land use mix, 34 zonings and subzonings/ subzones were grouped into nine main zoning categories, namely, (1) single residential zonings (SR); (2) general residential zonings (GR); (3) community zonings (CO); (4) local business zonings (LB); (5) general business and mixed use zonings (GB_MU); (6) industrial zonings (GI); (7) utility and transport zonings (UT_TR); (8) open space zonings (OS) and (9) agricultural, rural and limited use zonings (AG_RU_LU). A further grouping of these nine zoning categories resulted in four categories based on whether the main activity allocated is residential, commercial/service, recreational and agricultural use. The four main groups of land use identified by this categorization include (i) residential use (i.e. SR & GR), (ii) service and commercial use (i.e. CO, LB, GB_MU, GI & UT_TR), (iii) open space use (OS) and (iv) agricultural, rural and limited use (AG_RU_LU).

Land use mix for the entire city of Cape Town and individual census suburbs was measured firstly using 34 land use categories ($k = 34$) then using nine land use categories ($k = 9$) and lastly by considering the four land use categories ($k = 4$). Measuring land use mix using three aggregation levels of land use types was intended to help select the appropriate measure of land use mix to include in the modelling process. In addition, it was also intended to gain insight regarding the influence the numbers of land use types have on the outcome of land use mix measure. The computational process for the modified version of the entropy index and the HHI involved 5 steps which are outlined in APPENDIX C.

3.2.5.2 Processing street connectivity data

Street connectivity data (i.e. counts of different types of intersections, cul-de-sacs and total length of street network) was aggregated at the suburb level and then transformed into four standard measures of connectivity. These measures include intersection density, percentage of intersections with more than three legs, ratio of intersections to cul-de-sacs and street density. Two measures of intersection density were calculated; the intersection density as a ratio of intersection counts to the area of the suburb and as a ratio of intersection counts to the total length of the street network within the boundaries of the suburb. The number of intersections with more than three legs was determined as the sum of all four-legged and multi-legged intersections, staggered intersections and roundabouts or traffic circles with more than three legs. The percentage of intersections with more than three legs was calculated as the ratio of the sum of the intersection of the types above-mentioned to the total number of all intersections within the boundaries of the suburb. The ratio of intersections to cul-de-sacs was determined by dividing the total number of all intersection types in the suburb by the number of dead end streets or cul-de-sacs present in the same suburb. Street density was determined as the total length of the street network per unit of area. It is important to note that spatial data on the street network represents each carriageway by a line segment. It follows that all dual carriageways, freeways or divided highways are represented by double line segments. In this sense, the length of the street network should be understood as the length of carriageways (i.e. roadways separated by a central reservation or other physical barriers).

3.2.5.3 Processing data for transportation systems

1. Raw data aggregation

Data on transportation systems was initially extracted from spatial data (i.e. road network and public transport systems) and then aggregated at suburb level in Excel spreadsheets. The procedure followed the same steps as those applied to land use patterns which are: selecting a census suburb, extracting a layer of the suburb, selecting features and creating a layer of selected features.

2. Data screening

The road network for the City of Cape Town includes roads of different classes with walking trails being the lowest class. This class of road is coded as “RUR” and “RWW” in the spatial dataset. The inclusion of this class of roads into the analysis has the potential to inflate the length of the extracted street network of suburbs, especially those regarded as rural suburbs with a significant coverage of roads of this class. With this in mind, it was decided to exclude this class from the analysis. After the removal of the trail class, road segments extending outside the boundaries of the suburb of interest were removed from the extracted dataset. Road segments which were partially included inside the boundaries of the suburb were trimmed to the boundaries and a new measurement of the trimmed segment was taken and saved into the Excel dataset. A flowchart of the screening process for transportation system data is illustrated in Figure 3-10. The last step within the data screening process entailed the summation of segment lengths for each class of road and the summation of data points such as railway and bus stations.

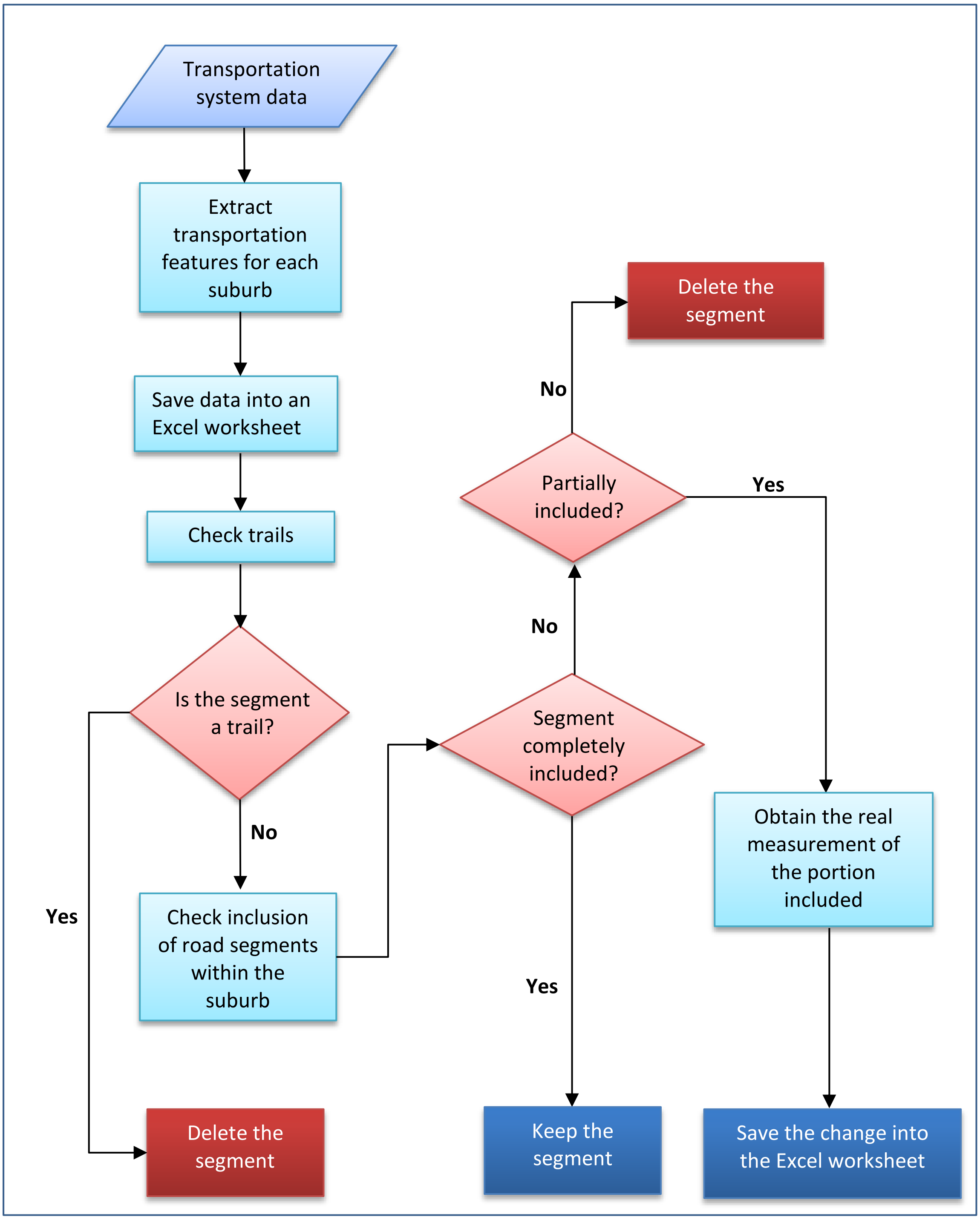


Figure 3-10: Flowchart of activities performed during processing data on transportation systems

3.2.5.4 Processing pedestrian casualty data

The objective of this step in the process was to create a shapefile of georeferenced pedestrian crash data and aggregate them at suburb level. Pedestrian crash data in Excel format was imported into ArcGIS through its integrated application, namely ArcCatalog and ArcMap. The imported Excel dataset was then converted into a shapefile in ArcMap to allow data visualization and facilitate spatial queries on the data. The aggregation of pedestrian crashes with particular characteristics was carried out by using GIS functionality of spatial joins using the polygon map layer for the 2011 census suburbs as the input data. Counts of pedestrian casualties aggregated at both metropolitan and suburban level were then converted into frequencies, then normalized to facilitate a comparison of crash frequencies across the census suburbs.

3.2.6 Data analysis

3.2.6.1 Univariate analysis

Univariate analysis involves a description of a dataset by reviewing the distribution of observations and providing a measure of central tendency (i.e. mean, mode and median) and dispersion (i.e. range, variance, standard deviation, maximum, minimum and quartiles) (Field, 2013; Kent, 2015).

3.2.6.2 Bivariate analysis

Bivariate analysis involves looking at associations between pairs of variables and trying to understand how those associations work. Inferential statistics is an example of bivariate analysis commonly used to draw conclusions about the population based on evidence from a sample (Field, 2013; Kent, 2015). Hypothetical testing and confidence interval procedures depend on whether the test is parametric or nonparametric. Generally, parametric tests make certain assumptions about parameters (or characteristics) of the population distribution upon which a test is based while nonparametric tests make few or less stringent assumptions about the distribution of the underlying population (Montgomery & Runger, 2007; Sheskin, 2011).

Traditionally, parametric tests assume that the distribution of the underlying population follows a normal distribution and their application are restricted to (i) normally distributed data, (ii) data with homogeneity of variance, (iii) continuous data (i.e. interval or ratio data) and (iv) data that are independent of one another (Field, 2013). Whether data is normally distributed or not affects a number of procedures including parameter estimates, confidence intervals, hypothetical testing, and error estimates while fitting the model (Field, 2013).

1. **Assumptions applicable in bivariate analysis**

Before analysing data in STATISTICA software tool, the four assumptions mentioned above were tested to ensure that the right statistics has been applied. This was done in five steps which are (1) checking outliers; (2) checking additivity and linearity; (3) checking normality of the distribution; (4) checking homogeneity of variance and (5) checking independence (Field, 2013).

Step 1: Checking outliers

Detecting outliers in a distribution of data was done with the use of STATISTICA software and MS Excel. The processing of detecting outliers followed the "outlier labelling rule" (Hoaglin *et al.*, 1986). The first step of this process was to build histograms and boxplots (or box-whisker diagrams) and to display percentiles and outliers in the output summaries of STATISTICA. The next step was to look at the plotted histogram of the sample and identify observations at the tails of the distribution that appear differently from others (i.e. observations that may look like outliers). The 25th percentile (or lower quartile) value (denoted here as "Q1") and the 75th percentile (or the upper quartile) value (denoted here as "Q3") were two of the most useful indicators required to apply the outlier labelling rule. These values were displayed in the output table of percentiles. An Excel spreadsheet was then created to calculate lower and upper limits using the outlier labelling rule. Lastly, observations that fell outside the calculated lower and upper limits were labelled as outliers.

To compute the lower and upper limits, the interquartile range (IQR) was determined for the distribution. The IQR is the difference between the lower quartile (Q1) value and the upper quartile (Q3) value (i.e.$IQR = Q3 - Q1$). The IQR was multiplied by 2.2 to obtain demarcation points determining the outliers. The lower limit was calculated by $Q1 - (2.2 \times IQR)$ and the upper limit by $Q3 + (2.2 \times IQR)$. Any observation that is lower than $Q1 - (2.2 \times IQR)$ or greater than $Q3 + (2.2 \times IQR)$ was labelled as an outlier. The 2.2 multiplier was recommended by Hoaglin and Iglewicz (1987) based on simulation research and is claimed to have greater validity than the more commonly used 1.5 multiplier introduced by (Tukey, 1977). It was reported that the 1.5 multiplier leads to inaccurate outlier spotting in approximately 50 percent of applied cases (Hoaglin *et al.*, 1986).

Step 2: Checking linearity

The assumption of linearity predicts that there is a linear relationship between the explanatory variables and the outcome variables. This assumption was tested in STATISTICA by the use of scatter plots. For this assumption to be met, the scatter plot should follow a linear pattern. Otherwise, a non-linear model is appropriate to describe the relationship.

Step 3: Checking normality of the distribution

The normality of the distribution was checked in STATISTICA by using a combination of methods including (i) visual inspection of the shape of the histogram and the boxplots, (ii) the "probability-probability plot" (P-P plot), (iii) measures of shape (kurtosis and skewness), (iv) the Kolmogorov-Smirnoff test and (v) the Shapiro-Wilk test.

With the help of STATISTICA, it can be visualized whether the plotted histogram are bell-shaped (i.e. normally distributed) or not. In the same way, a visual examination of the boxplots can indicate whether they are approximately symmetrical (i.e. Q1 and Q3 are located approximately at the same distance away from the median and the whiskers of the plot are approximately of the same height) and do not appear too sharp or too flat (Field, 2013). The normal P-P plot is another graphical method of inspecting normality. For a normal distribution, the data points should be close to the ideal diagonal line. Normality is assessed visually by the closeness of data points to the diagonal line. A significant deviation of data points from the diagonal line indicates that the normal distribution is not adequate to describe the data (Montgomery & Runger, 2007). According to Field (2013), the distribution is viewed as deviating from normality when data points sag consistently below or above the diagonal line (i.e. indication of kurtosis) or when data points deviate from the diagonal line in the 'S' shape (i.e. indication of skewness).

Skewness refers to where data is piled up on the distribution (i.e. whether data is heavily weighted towards the left or the right) while kurtosis indicates how flat or peaked the distribution is. For a normal distribution, STATISTICA produces skewness and kurtosis measures that are close to zero. A positive value of skewness indicate that data is heavily weighted towards the left of the distribution (i.e. higher probability for low scores in the distribution) whereas the negative value indicates a pile up of data on the right side of the distribution (i.e. higher probability of high scores in the distribution). A positive kurtosis value depicts a sharp distribution with heavy tails while a negative value is indicative of a flat distribution with light tails (Field, 2013). Skewness and kurtosis values were converted into z-values to test whether these values differ significantly from 0 (i.e. expected value for a perfect normal distribution). The computation of z-scores was carried out in Excel by dividing the skewness or kurtosis measure by the corresponding standard error:

$$z_{skewness} = S/SE_{skewness};\ z_{kurtosis} = K/SE_{kurtosis} \tag{12}$$

where S denotes skewness measure and K denotes kurtosis measure. It was concluded that skewness and kurtosis measures were significantly different from 0 if z-scores were found greater than -1.96 or 1.96 at 95% confidence interval or greater than -2.58 or 2.58 at 99% confidence interval or greater than -3.29 or 3.29 at 99.9 % confidence interval.

The Kolmogorov-Smirnov (Kolmogorov, 1933; Smirnov, 1936) and Shapiro-Wilk tests (Shapiro & Wilk, 1965) are based on hypothetical tests of whether the distribution of the sample deviates significantly from a normal distribution. The null hypothesis for these tests is that where the data are normally distributed and if the p-value is below 0.05, the null hypothesis is rejected and the inference is that the distribution of the sample is significantly different from a normal distribution. However, these tests were excluded as they are considered inappropriate for large samples because they are likely to produce p-values that are below 0.05 even when the interpretation of the skewness and kurtosis values justifies the normality of the distribution (Field, 2013).

Step 4: Checking variance homogeneity

The assumption of homogeneity of variance means that the variance of the outcome variable should be equal for all levels (or groups) of the predictor variable (Field, 2013). Homogeneity of variance can be tested for both normally distributed data (i.e. parametric data) and non-normally distributed data (non-parametric data). The assumption of homogeneity of variance is assessed using Levene's test (Levene, 1960) which tests the null hypothesis that different groups have equal variance. If Levene's test is non-significant (i.e. p-value greater than 0.05), the null hypothesis is accepted and the assumption of homogeneity of variance is approved.

2. Inferential statistics applied in the analysis

Two methods of inferential statistics were applied in this study to draw inferences on the statistical population. These are *t*-test, to test hypotheses about the means of two normal distributions, and analysis of variance (ANOVA), to test whether there are any statistically significant between the means of three or more groups. When the ANOVA test is significant, it indicates that the means of various groups are statistically different from each other. However, it does not tell which groups are different (Montgomery & Runger, 2014). To investigate this issue, post-hoc tests were used to perform some follow-ups test to identify the specific differences. There are many post-hoc procedures and these differ based on their assumptions of variance and sample size, and depending whether the test is parametric or non-parametric (Shingala & Rajyaguru, 2015). Two post-hoc procedures applied in in this study are the Games-Howell and Bonferroni post-hoc procedures, and both were selected based on criteria illustrated in Figure 3-11.

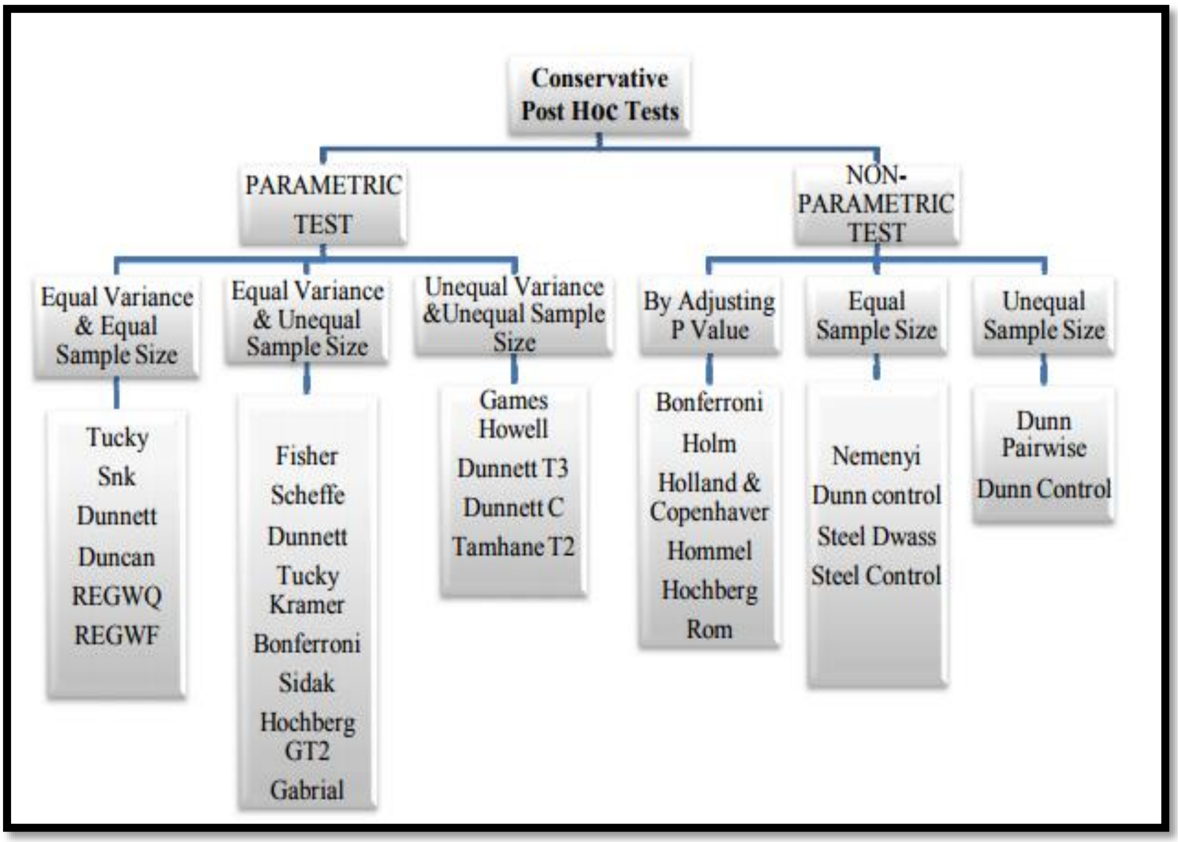


Figure 3-11: Criteria for selecting appropriate Post Hoc Tests (Shingala & Rajyaguru, 2015)

3.2.6.3 Geospatial analysis

Generally, geospatial analysis refers to an approach of processing geographical or spatial data and applying statistical analyses to it. In this study, geospatial analysis was carried out with the use of ArcMap and following an analytical procedure proposed by Mitchel (2005). This procedure is illustrated in Figure 3-12 presented below.

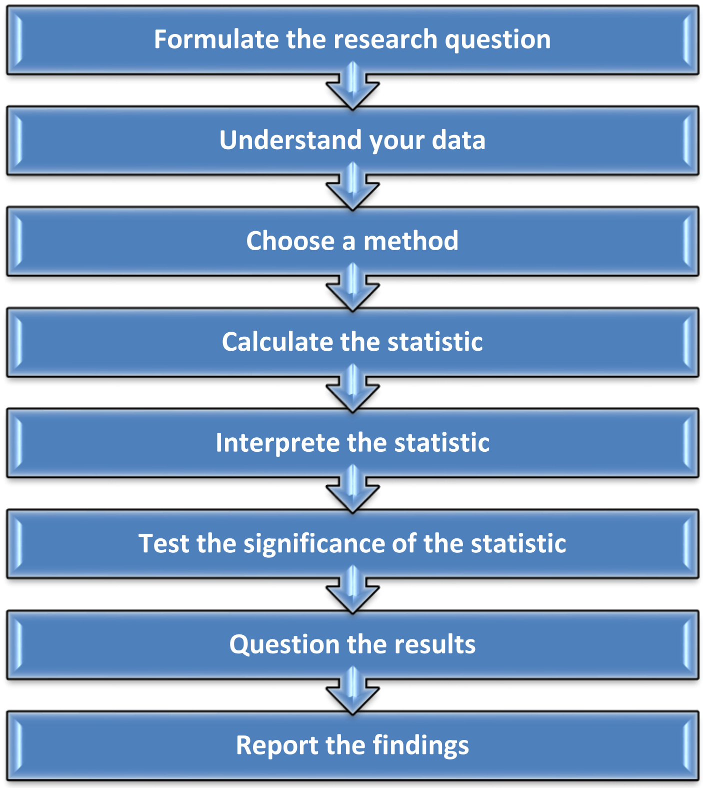


Figure 3-12: Geospatial analysis: Analytic procedure (Mitchel, 2005)

1. **Formulating the research question**

Each process of geospatial analysis was performed with the intention of responding to a specific research question as formulated in the first chapter of this dissertation. Geospatial analysis enabled the examination of the location of pedestrian crashes with particular characteristics and made use of collected information to study the relationship between attribute variables and the locations of pedestrian crashes. The following research questions were addressed through this process: (i) where are pedestrian crashes located? (ii) How are pedestrian crashes spatially distributed? (iii) What are the environmental characteristics of pedestrian crash locations? In addition to addressing the research questions, spatial analysis can also allow for the examination of where and how factors affecting pedestrian crashes are spatially distributed and concentrated. Analysing these explanatory variables in their spatial context can in turn generate understanding of the relationships between pedestrian crash location and the built environment.

2. **Understanding the nature of spatial data**

Geographical data consists of two main categories: vector data and raster data (Bossler *et al*, 2010; Pandey, 2014; De Smith *et al.*, 2015). Vector data comprises 3 types of data: points (i.e.

with x and y coordinates), line or arc (series of points between two nodes) and area polygons or lines with the same start and end points (Lloyd, 2010). Raster data is made up of discrete cells (or pixels) of equal size, arranged in columns and rows. Each cell is associated with a numeric value or class and positional information. Groups of grid cells sharing the same attribute value characterize geographic features of the same type (Wade & Sommer, 2006). The analytic tasks and statistical techniques applicable to spatial data often depend on the type of the data, its quality, strengths and weakness. In addition, it will be shown later in this chapter that the choice of the statistic largely depended on whether there was an attribute value attached to the feature or not (see Figure 3-15 on Page 118). Therefore, an understanding of the nature of the dataset and its qualitative aspects is an essential step in the overall process of geospatial analysis. For this study, spatial data comprises all types of data, namely, point, line, polygons and raster data.

3. The choice of method for geospatial analysis

The literature contains a wide variety of approaches to geospatial analysis performed with the intention to study spatial location and distribution of phenomena of interest. These approaches apply a range of statistical techniques designed to analyse and predict the values attached to spatial phenomena. For this reason, they are commonly referred to as spatial statistics or geostatistics (Wade & Sommer, 2006).

Geostatistics is a sub-branch of the statistics discipline and makes use of standard statistical techniques such as exploratory data analysis, descriptive and inferential statistics, and modelling techniques. Descriptive statistics for spatial data summarizes a sample of spatial data by producing quantitative measures of central tendency, frequency of spatial distributions and measures of spread or dispersion of spatial data. Geostatistical techniques concerned with the central tendency are also referred to as centrographic techniques (Wheeler *et al*., 2013). Inferential statistics makes use of hypothetical tests from observed patterns of a sample of spatial data to draw conclusions about the general population from which the sample is drawn. Inferential statistics involves two tasks: estimation of parameters and hypothetical testing (Lloyd, 2010; Oyana & Margai, 2016).

With regards to this study, geospatial analysis involved two main tasks: creating and manipulating maps layers and running exploratory spatial data analysis (ESDA). The first tasks involved activities such as creating map layers in ArcMap from available spatial data, navigating the created maps and checking attribute data connected to features, adding layers to the group layers, editing attribute tables (creating new attributes and joining tables), aggregating data (selecting and exporting a subset of a feature class), and performing spatial queries (i.e. selecting and extracting certain data based on its geographical location). Exploratory spatial data analysis (ESDA) encompasses a range of techniques to (i) visualize data in a spatial framework through maps and other graphics; (ii) identify patterns of spatial association and spatial clustering through spatial correlation and regression analysis, (iii) detect remarkable and significant patterns such as atypical locations (i.e. spatial outliers) and (iv) suggest different spatial regimes or other forms of spatial heterogeneity (Anselin & Bao, 1997). The measure of global and local spatial autocorrelation is the central focus of the ESDA (Anselin & Bao, 1997; Anselin, 1998). This approach of spatial analysis makes use of descriptive statistics and suggests hypotheses about spatial autocorrelation and features clustering to quantify spatial patterns. Techniques of spatial autocorrelation which were applied to this study are described in the following sections.

4. Visual inspection of mapped data

After creating map layers and performing spatial queries in ArcMap, data visualisation was the starting point for the exploratory spatial data analysis. Data visualization involves techniques of data graphing and mapping using a combination of visual elements such as choropleth maps, graphs, histograms, charts, boxplots, scatter plots, and 3D maps (Kim, 2009; Wade & Sommer, 2006). A choropleth map is a thematic map portraying differences in the phenomenon being mapped (or categorized classes of the mapped phenomenon) by the use of divided geographical areas or polygons that are coloured, patterned or shaded in relation to the attribute value attached to the mapped phenomenon (De Smith *et al.*, 2015; Wade & Sommer, 2006). The main purpose of these visualization techniques was to present spatial patterns and depict information in a way that is easily understandable.

GIS-related techniques applied in this study for geospatial visualization include categorizing spatial data and designing symbology for each category, controlling selected feature classes or values to be displayed, mapping quantities associated with mapped phenomena, choosing appropriate classification methods (i.e. natural breaks, quantile, equal interval and standard deviation), checking and dealing with spatial outliers, creating a map series, mapping density values (i.e. dot densities or surface densities), creating and mapping buffers, clipping features, mapping percent changes in value, creating map layouts, adding graphs to a layout and print map outputs.

5. Pattern analysis

i. Concepts and definitions

The techniques of spatial data visualization, especially those utilizing classification schemes (i.e. natural breaks, quantile, equal interval and standard deviation) introduces subjectivity in the analysis since the responsibility to choose an appropriate procedure lies with the analyst (De Smith *et al*., 2015). These techniques serve a purpose for visual inspection of spatial data but they cannot prove statistically whether there is a spatial pattern in the mapped data. To answer this question, geospatial analysts have resorted to using spatial statistical analyses and inferential statistics to analyse spatial dependence and spatial heterogeneity. Spatial heterogeneity simply refers to the uneven distribution of observations across the geographical area (Anselin, 2010) while spatial dependence refers to the similarity of attribute values of a single variable in spatial proximity or closeness (Griffith, 2017).

The First Law of Geography, according to Waldo Tobler, is that "all things are related, but nearby things are more related than distant things" (Tobler, 1970). Nevertheless, as seen previously in Section 3.2.6.2, most classical statistical theory and practice assume that observations are independent of one another, identically distributed (i.e. homogeneity of variance) and usually conform to a normal distribution or a bell-shaped curve (Chun & Griffith, 2013; Field, 2013). These assumptions are in contrast with Tobler's First Law of Geography which states that observations are often spatially dependent due to their relatively close locations on the earth's surface (Brus *et al*., 2014; Chun & Griffith, 2013; Malczewski, 1999). This situation is commonly referred to as "spatial autocorrelation" in the literature of geospatial analysis (Griffith, 2017). It introduces a deviation from the assumption of independent observations applied by classic statistical analysis (Chun & Griffith, 2013; Griffith, 2017). Spatial autocorrelation explores how an attribute value of a variable at one location in space is related to the attribute value of that same variable in a nearby location (Dixon *et al.*, 2016; Rogerson, 2001). The prefix "auto-" attached to "correlation" is indicative of correlated attribute values within a single variable. The analysis of spatial autocorrelation produces indices measuring the way observations are spatially distributed and quantifying the magnitude of spatial association (or correlation) between neighbouring observations.

The analysis of spatial autocorrelation can be carried out from either a global perspective or a local perspective. Global autocorrelation statistics provide a single summary value for the entire study area, depicting whether spatial observations display clustering or not, regardless of the presence of local dependence in the attributes values within the study area (Loo & Yao, 2012; Wong & Wang, 2017). These statistics allow an assessment of overall clustering patterns across the entire study area but are unable to identify locations where clustering occurs (Zhu, 2016). In contrast to global autocorrelation measures, local autocorrelation statistics are concerned with detecting clustering tendency (such as hot spots, cold spots and spatial outliers) at the local level (Anselin & Bao, 1997; Loo & Yao, 2012; Wong & Wang, 2017; Zhu, 2016).

ii. Global spatial autocorrelation measures

The theory of spatial autocorrelation was first pioneered by Moran (1948) and Geary (1954) and popularized later by a number of other scholars (Chun & Griffith, 2013; Cliff & Ord, 1973, 1981; Dixon *et al.*, 2016; Krige, 1966; Matheron, 1971). The Moran's index (or Moran's I), the Geary ratio (GR) also known as Geary's contiguity ratio or Geary's C ratio and the Getis-Ord G statistic are the most widely used measures of global spatial autocorrelation (Loo & Yao, 2012; Wong & Wang, 2017).

The Global Moran's Index

The global Moran's index is a measure of covariance that relates directly to the Pearson product-moment correlation coefficient "r", calculated for two variables, X and Y, as follows (Chun & Griffith, 2013; Griffith, 2003):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})/n}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n}\ \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/n}} \tag{13}$$

Where x_i and y_i are paired values of two variables for observation i; $\bar{x}$ and $\bar{y}$ are the respective means of these two variables. To measure global spatial autocorrelation, the spatial relationship between all pairs of nearby georeferenced values is captured using a $n \times n$ binary spatial weight or spatial matrix generally denoted by C, with its cells being denoted by c_{ij}. If two locations are designated as neighbours (i.e. areal units or cells that are immediately adjacent or areal units or cells sharing a common non-zero length boundary), then the value of c_{ij} .is 1, otherwise the value of f c_{ij} is equal to 0 (Chun & Griffith, 2013; Griffith, 2003; O'sullivan & Unwin, 2010). Accordingly, the 2n data values are treated as a set of n spatial data values for a variable y, with the covariance being weighted by the sum of the elements of the weight matrix C (Chun & Griffith, 2013). The equation (13) then becomes:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\,(y_i - \bar{y})(y_j - \bar{y}) / \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 / n}\ \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 / n}} \tag{14}$$

Where I is the Moran's index; c_{ij} are the elements (i.e. 0 or 1) of the weight matrix C; the term $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$ is the sum of all the elements of the weight matrix C; y_i and y_j are the attribute values for the aerial unit or cell i and j ; $\bar{y}$ is the mean attribute value for the n areal units; and the term $(y_i - \bar{y})$ denotes the deviation from the mean attribute value within the same variable. The equation for the Moran's I is usually simplified as follows:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}\ \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\ (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{15}$$

The Moran's I can be positive or negative and generally ranges between +1 and -1. Positive spatial autocorrelation implies a tendency towards clustering, that is to say that nearby geographical areas tend to have observations with similar attribute values (i.e. high values tend to be located near high values, medium values near medium values and low values near low values). Negative spatial autocorrelation implies that observations with dissimilar attribute values occur near one another (i.e. high values tend to be surrounded by nearby low values and low values tend to be surrounded by nearby high values) or the tendency towards dispersion (Dixon *et al*., 2016; Griffith, 2003). A Moran's I that has a value of zero is indicative of independent observations in the dataset, a random pattern or the absence of spatial autocorrelation between attribute values of the same variable y (Dixon *et al*., 2016; Griffith, 2003; Zhu, 2016).

The statistical inference on the global Moran's I uses the calculated value of Moran's I and both z-score and p-value to evaluate whether the spatial pattern of observations is clustered, dispersed or random (see Figure 3-13).

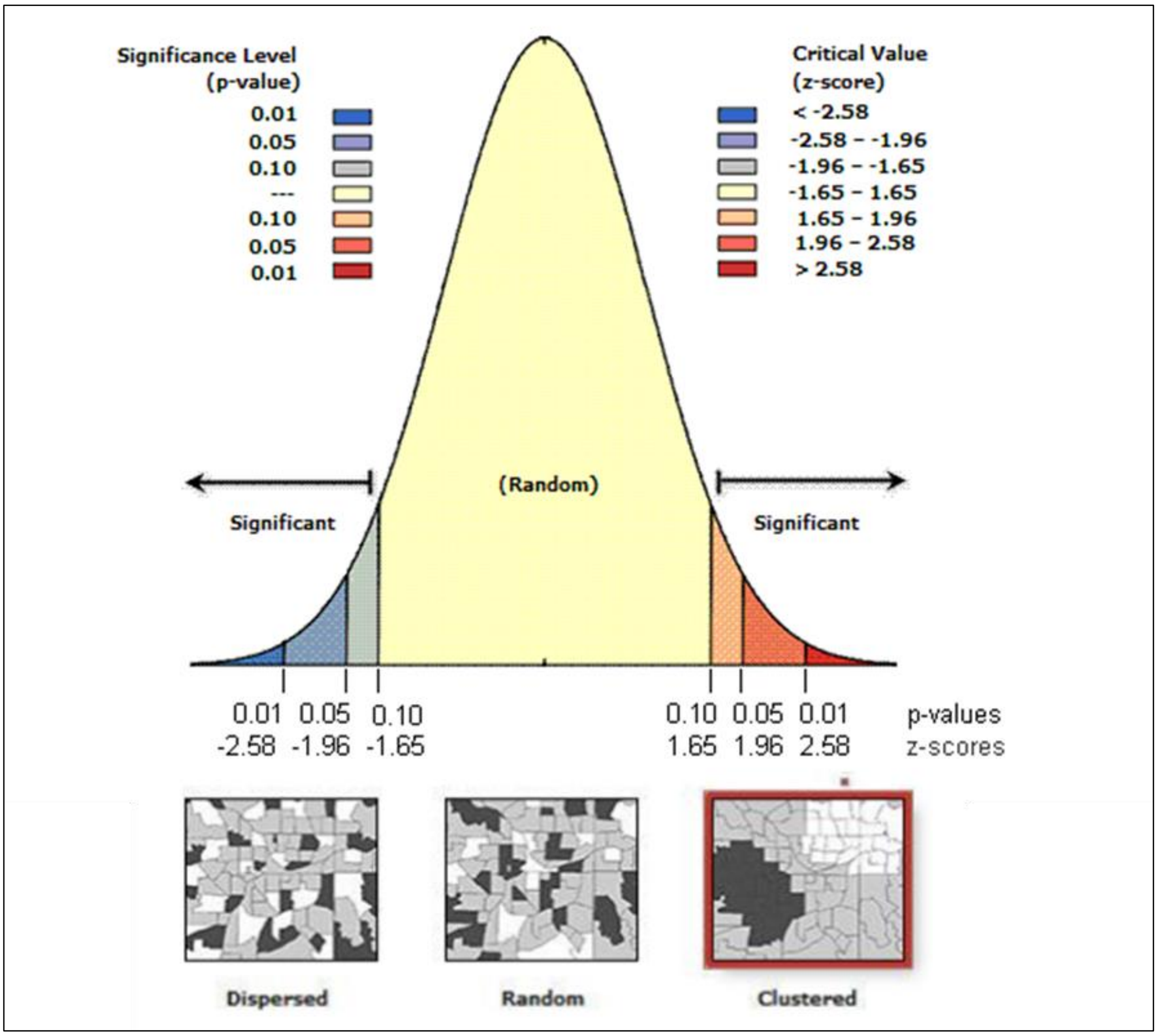


Figure 3-13: Interpretation of z-scores and p-values for the Global Moran's I statistic

The z-score for the global Moran's I statistics is computed as follows:

$$z_I = \frac{I - E(I)}{\sqrt{V(I)}} \tag{16}$$

Where $E(I)$ is the expected value of I when the null hypothesis H_O is true, and $V(I)$ is the variance of I . $E(I)$ is expressed as follows:

$$E(I) = \frac{1}{n-1} \tag{17}$$

Where n is the sum of areal units within the study area. The null hypothesis Ho states that the spatial pattern is random (i.e. spatial independence) with zero spatial autocorrelation. When the calculated p-value is statistically significant, the null hypothesis is rejected, implying that the spatial pattern is either clustered or dispersed (Mitchell, 2005; Zhu, 2016).

The Geary Ratio

The Geary ratio (GR) or Geary's C ratio is the second popular measure of global spatial autocorrelation and an alternative measure to the Moran's index (Griffith, 2003; O'Sullivan & Unwin, 2010; Wong & Wang, 2017). While the Moran's I is based upon the measure of covariance (i.e. mean deviations) between attribute values, the GR uses the sum of squared differences between paired values of the variable y instead of the product (as it is in the case of Moran's I) to assess the similarity of attribute values of a given variable y (Acevedo, 2012; Wong & Wang, 2017; Zhou & Lin, 2008). The GR is based upon the unbiased estimates (i.e. division by division by n-1 instead of n) and can be expressed as follows (Griffith, 2003; Wong & Wang, 2017):

$$GR = \frac{n-1}{2\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}} \frac{\sum_{i=1}^{n}\left(y_i - y_j\right)^2\left(\sum_{j=1}^{n} c_{ij}\right)}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{18}$$

The appearance of the number 2 in the denominator of the first term stems from the calculation of the Moran's I statistic which involves both c_{ij} and c_{ji} (Chun & Griffith, 2013). The GR is more sensitive to local variations than the Moran's I and the sensitivity of the GR can be attributed to the denominator of first term (i.e. $2\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$) and the numerator term $\sum_{i=1}^{n}\left(y_i - y_j\right)^2\left(\sum_{j=1}^{n} c_{ij}\right)$ because it gives smaller values for similar values in nearby locations and accentuates large deviations, since differences in attribute values of variable y are squared (Chun & Griffith, 2013; O'sullivan & Unwin, 2010; Wong & Wang, 2017). The interpretation of GR can be sometimes confusing; the GR values ranges from 0 to 2 (or $0 < GR < 2$). A low value of GR less than 1 (i.e. $0 \leq GR < 1$) indicates a positive spatial correlation, a value of 1 indicates the absence of spatial autocorrelation, and a high value greater than 1 indicates a negative spatial autocorrelation (O'Sullivan & Unwin, 2010; Zhou & Lin, 2008).

The drawbacks of sensitivity to local variations coupled with those related to the interpretation of the GR makes the Moran's index the most appealing and preferred index of spatial autocorrelation (Chun & Griffith, 2013). In addition, The Moran's I offers computational ease and facilitates a number of extensions that increase its application and usefulness (De Smith *et al.*, 2015). Accordingly, the Moran's I is the most popular index of spatial autocorrelation frequently implemented in most spatial analytical tools, such as ArcMap Spatial Statistics Toolbox, the AUTOCORR function in Idrisi and others (de Smith *et al.*, 2015).

The Getis-Ord General G statistic

The Getis-Ord General G statistic or simply G-statistic (Getis & Ord, 1992) is another popular measure of global spatial autocorrelation, formally expressed as follows (Wong & Wang, 2017):

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} y_i y_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j} \tag{19}$$

Where y_i and y_j are attribute values for aerial units or cell i and j; and the term $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}$ is the sum of all the elements of the spatial weight matrix C between i and j, carrying a value of 0 or 1. Following the equation above, the G-statistic can be regarded as a ratio of the association of attribute values within nearby locations (depicted by c_{ij}) to the association of attribute values over the entire study area (Wong & Wang, 2017). As seen previously, when two locations are designated as neighbours, the weight matrix c_{ij} carries a value 1, otherwise c_{ij} takes a value of 0. When relatively large attribute values are within neighbouring locations, the value of G-statistic becomes relatively large (owing to the sum of products). When relatively small attribute values are designated as neighbours, the value of G-statistic becomes relatively small. This implies that higher values of G-statistic are indicative of a cluster of higher attribute values or the presence of a hot spot (i.e. high values next to high values) and lower values of G-statistic indicate the presence of a cold spot (i.e. low values next to low values) (Wong & Wang, 2017). The statistical inference on the G-statistic is calculated as follows:

$$z_G = \frac{G - E(G)}{\sqrt{V(G)}} \tag{20}$$

Where $E(G)$ is the expected value of G-statistic and $V(G)$ is the variance of G. $E(G)$ is calculated as follows:

$$E(G) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}{n(n-1)} \tag{21}$$

Where n is the sum of areal units within the study area. The variance of G-statistic is expressed as $V(G) = E(G)^2 - [E(G)]^2$.

The expected value of G-statistic corresponds to the null hypothesis H_O which assumes that the y values are randomly distributed in space (or there is no spatial clustering of y values). When the p-value appears small and statistically significant, the null hypothesis is rejected and the alternative hypothesis is that the spatial pattern is not random. In this, the sign of the z-score becomes important in determining whether the clustering of y values consists of high values (i.e. hot spots) or low values (i.e. cold spots). A positive value of z-score (i.e. $G > E(G)$) indicates the presence of a hot spot or high-high cluster while a negative z-score indicates the presence of a cold spot or low-low cluster (Getis & Ord, 1992; Mitchell, 2005; Wong & Wang, 2017). However, it is worth noting that the interpretation of z-scores for the G-statistic differs from that of z-scores for the global Moran's I and results from these two statistics are often not comparable (Mitchell, 2005; Wong & Wang, 2017).

iii. Local spatial autocorrelation measures

Local spatial autocorrelation statistics are used when the intention is to detect geographic variation in events or phenomena at a local scale within the study area (Anselin *et al*., 2013; O'sullivan & Unwin, 2010). Local statistics of autocorrelation were developed in the effort to deal with general concerns associated with the computation of global statistics. These concerns include the assumption of homogeneity over the entire region, the presence of localized clusters of similar or dissimilar values within the study region and a situation wherein a mixture of positive and negative spatial autocorrelation tend to cancel each other out, leading to a failure to detect a spatial autocorrelation (Chun & Griffith, 2013; O'sullivan & Unwin, 2010). However, the concept of local statistics introduces computational difficulties, requiring substantial simulation-based methods to test simultaneously multiple localized patterns within the study area (Chun & Griffith, 2013; O'sullivan & Unwin, 2010).

Local indices of spatial association (LISA) introduced by Anselin (1995) and the Getis-Ord local statistics, Gi and Gi* developed by Getis and Ord (1992) and Ord and Getis, (1995) are the two statistics used in the study of spatial autocorrelation. Global statistics of spatial autocorrelation are regarded as average of n local statistics, implying that local statistic can be derived by a spatial disaggregation of the global statistics at a small geographic unit (Chun and Griffith, 2013; Wong & Wang, 2017). The widespread application of these statistics in recent years have been prompted by their implementation in commercial GIS software and their ability to detect significant local patterns which can be visually displayed on a map (Anselin *et al*., 2013).

The Local Indices of Spatial Association (LISA)

The local indices of spatial associations (LISA) comprise both local versions of global Moran's I and Geary ratio (Wong & Wang, 2017) and relate to the Moran scatterplot (Anselin, 1995, 1996). The Moran scatterplot is one of the graphical methods used to depict and qualitatively measure the spatial autocorrelation. Other graphical methods used for the same purpose include a semi-variogram, a covariogram and a correlogram (Chun & Griffith, 2013; De Smith *et al.*, 2015; Griffith, 2017). The Moran scatterplot portrays a bivariate relationship between the y value of the i-th areal unit and its corresponding lagged value (i.e. the sum of spatially weighted values of neighbours) on a two-dimensional diagram using Cartesian coordinates (Anselin *et al.*, 2013; Griffith, 2017). The value of y_i is first converted to a z-score (that is $z_i = (y_i - \bar{y})/\sqrt{V(y)}$, with V(y) denoting the variance of y) and this is plotted against its corresponding sum of spatially weighted z-scores of neighbours (i.e. $\sum_{j=1}^{n} c_{ij} z_j$). Following this, each local Moran (I_i) is the product of the pair of values on the horizontal (i.e. z_i) and the vertical axis (i.e. $\sum_{j=1}^{n} c_{ij} z_j$). Hence, the local Moran can be expressed as follows (Wong & Wang, 2017):

$$I_i = z_i \sum_{j=1}^{n} c_{ij} z_j \tag{22}$$

The Moran scatterplot is partitioned into four quadrants distinguished by a trend line (i.e. zero value on the horizontal and vertical axis). These four quadrants portray the four types of local spatial association; the upper right quadrant represents high values surrounded by high values (HH), the lower left quadrant represents low values surrounded by low values (LL), the upper left quadrant represents low values surrounded by high values (LH) and the lower right quadrant represents high values surrounded by low values (HL). Furthermore, the HH and LL quadrants reflect a positive spatial autocorrelation while both LH and HL quadrants correspond to a negative spatial autocorrelation (Chun & Griffith, 2013; Griffith, 2017). Figure 3-14 illustrates an example of the Moran scatterplot with the four quadrants.

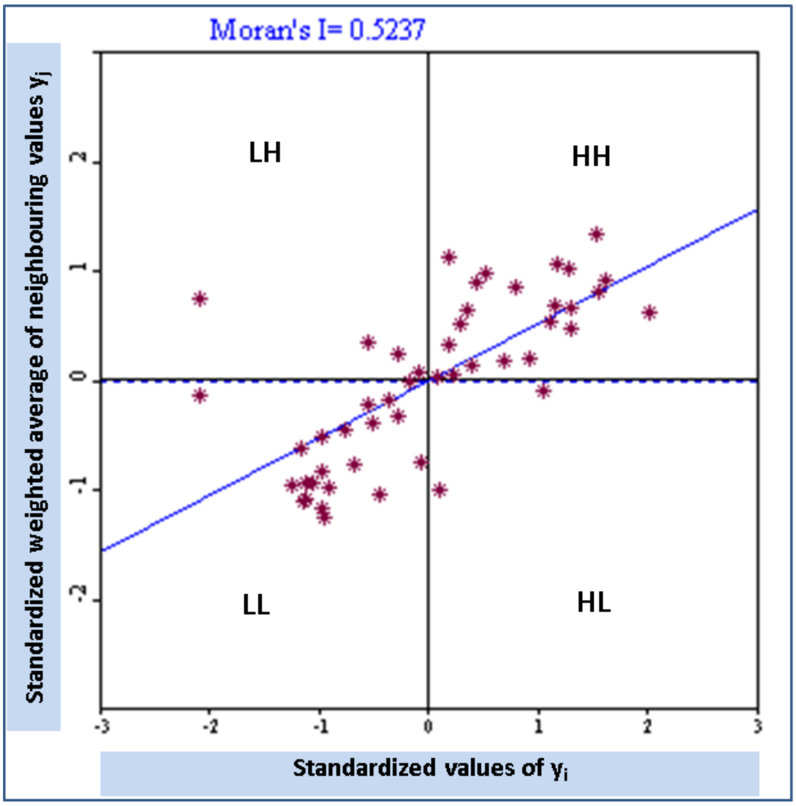


Figure 3-14: An example of the Moran scatterplot with its four quadrants

The statistical inference on the local Moran's I_i is conducted by calculating the z-scores from the expected value and the variance of the local Moran's I_i under the assumption of normal distribution of attribute variables:

$$z_{I_i} = \frac{I_i - E(I_i)}{\sqrt{V(I_i)}} \tag{23}$$

Where: $$E(I_i) = {-\sum_{j=1}^{n} c_{ij}} \big/ (n-1) \tag{24}$$

The measure of variance of the local Moran's index $V(I_i)$ is relatively complex but computational details can be found in Anselin (1995).

The Getis-Ord, local statistics, Gi and Gi*

The Getis-Ord local statistics, Gi and Gi* (Getis & Ord, 1992; Ord & Getis, 1995) are other local versions of spatial autocorrelation statistics. While the local Moran's I relates to the Moran scatterplot, the Gi and Gi* relates more to a semi-variogram (Chun & Griffith, 2013). A semi-variogram is a plot depicting spatial dependence by plotting distance variability between two geographic locations and fitting a model through the plotted pairs of observations (De Smith *et al.*, 2015; Wade & Sommer, 2006). The Gi and Gi* are defined as the ratio of the sum of values in neighbours of areal unit i , within a distance band d , to the sum of values in all areal units (excluding the value in areal unit i for the Gi statistics, but including it for the Gi*statistic) (Fischer & Wang, 2011). It follows then that the distinction between the Gi and Gi* statistics depends on whether a focal attribute value (y_i) is included in calculations or not (Chun & Griffith, 2013; Griffith, 2017). Mathematically, these local statistics can be expressed by the following equations (Fischer and Wang, 2011; Wong & Wang, 2017):

$$G_i = \frac{\sum_{j \neq i}^{n} c_{ij}(d) z_j}{\sum_{j \neq i}^{n} z_j} \tag{25}$$

$$G_i{}^* = \frac{\sum_{j=1}^{n} c_{ij}(d) z_j}{\sum_{j=1}^{n} z_j} \tag{26}$$

where z_j is the deviation from the mean of the neighbouring value. Further development of the $G_i{}^*$ formula leads to the following mathematical equation:

$$G_i{}^* = \frac{\sum_{j=i}^{n} c_{ij}(d) y_j - \bar{Y} \sum_{j=1}^{n} c_{ij}(d)}{S_Y \sqrt{\left[n \sum_{j=1}^{n} c_{ij}{}^2(d) - \left(\sum_{j=1}^{n} c_{ij}\right)^2\right] / (n-1)}} \tag{27}$$

where y_j denotes the attribute value of areal unit j; $c_{ij}(d)$ is the spatial weight linking areal unit i and j; and n is the total number of all areal units and:

$$\bar{Y} = \frac{\sum_{j=1}^{n} y_j}{n} \tag{28}$$

$$S_y = \sqrt{\frac{\sum_{j=1}^{n} y_j{}^2}{n} - (\bar{Y})^2} \tag{29}$$

The Getis-Ord Gi* statistic is the most widely used local version of Getis-Ord statistic and is implemented in GIS applications. The Gi* values are z-scores and are used to test statistical significance (Zhu, 2016). A statistically significant positive z-score indicates a clustering of high values or hot spot (high values surrounded by high values), a statistically significant negative z-score refers to a clustering of low values or cold spot (i.e. low values surrounded by low values) and z-scores close to zero may be indicative of a random pattern or clustering of moderate values (Griffith, 2017; Wong & Wang, 2017).

iv. Kernel density estimation (KDE)

KDE is a widely used technique for spatial analysis and a powerful tool for visualization of spatial patterns of features (Bíl *et al*., 2013; Thakali *et al*., 2015; Xie & Yan, 2013). The method uses a search radius distance (or bandwidth) defined by a kernel function. Several kernel density functions are provided in literature. The Bandwidth and the grid cell size are two main parameters which affect the results for the hotspot analysis(Hashimoto *et al*., 2016; Thakali *et al*., 2015). In this study, a grid cell of 30m X 30m was chosen as appropriate given the size of the study area and the processing time required for the identification of hot spot in ArcMap. Different bandwidths ranging to 200m to 1,895m were tested in this study to achieve a better visualisation of hot spots of pedestrian casualties within the study area and to study the sensitivity of hot spot patterns with respect to the bandwidth size.

6. Exploratory spatial data analysis (ESDA) using ArcMap

This study applied three local statistics of spatial autocorrelation and the planar kernel density estimation (KDE) to identify clusters of pedestrian casualties within the City of Cape Town. The applied statistics of spatial autocorrelation are the Anselin Local Moran's I (also denoted in ArcMap as "Cluster and Outlier Analysis"), the Getis-Ord Gi* (or the "Hot Spot Analysis" in ArcMap) and the Optimized Hot Spot analysis. These statistics produce z-scores and p-values that help to ascertain the presence of spatial associations between features under the study. Unlike the local statistics of spatial autocorrelation, the kernel density estimation (KDE) does not generate statistical inference on the existence of spatial associations between features.

The Anselin Local Moran's I is able to identify both statistically significant clusters and spatial outliers of features. The tool displays two forms of clustering, the HH (i.e. features of high values surrounded by other features of high values or simply hot spots) and the LL (i.e. features of low values in close proximity to other features of low values also denoted as cold spots). In addition, the statistic also has the ability to highlight two forms of spatial outliers: the HL (i.e. features of high values surrounded by other features of low values) and the LH (i.e. features of low values surrounded by those of high values). The Anselin Local Moran's was applied to features with an attribute value attached to them such as polygons (e.g. census suburbs with the associated count of pedestrian casualties that occurred within the boundaries of the suburbs) or point data (e.g. intersection point with the assigned index score for the design and the provision of pedestrian facilities)

The Hot Spot Analysis tool or Getis-Ord Gi* was applied in a similar way as the Anselin Local Moran's I to features with an attribute value attached to them. The tool allows the identification of statistically significant hot and cold spots by the use of the Getis-Ord G* Statistic. With this statistic, clusters are identified at different confidence levels (e.g. 90%, 95% and 99% confidence intervals).

The Optimized Hot Spot Analysis tool was applied to incident data (i.e. feature with no attribute value attached to them such as crash points) or weighted features. The tool provides three options of incident data aggregation; (1) count of incident point within fishnet grids, (2) count of incident points within existing polygons (e.g. census suburbs in the context of this study) and (3) creation of weighted points by snapping nearby incident points. After running one of these aggregation procedures, the tool runs the Getis-Ord Gi* statistic using default parameters set or calculated by the tool.

In addition to local statistics of spatial autocorrelation, the plan kernel density estimation (KDE) was applied to visualize where clusters of pedestrian casualties appear using six different bandwidth values: 200m; 400m, 500m; 800m, 1 000m and 1 895m. Different bandwidth values tested in this study were used in previous studies and the adoption of these bandwidth values was intended to allow a comparison between the findings of the current study and those from other scholars' works. The density surface produced by the Optimized Hot Spot analysis tool was based on the optimal fixed distance of 1895 meters (value automatically calculated by the tool based on the characteristics of the input pedestrian crash dataset). As a result, kernel density estimation (KDE) with 1895 m bandwidth was included in the analysis to allow a comparison between hotspots identified by the KDE and the Optimized Hot Spot Analysis tool. Figure 3-15 presents the flowchart for the exploratory data analysis in ArcgGIS.

7. Comparison of methods of cluster analysis

The prediction accuracy index (PAI) was used in this study to evaluate the performance of each method used to detect hot spots. The method was developed by Chainey *et al*. (2008) and its application has spread to research in road (e.g. Thakali *et al*., 2015). The prediction accuracy index is calculated as follows:

$$PAI = \frac{\left(\frac{n}{N}\right) \times 100}{\left(\frac{m}{M}\right) \times 100} \tag{30}$$

where n is the count of events (i.e. casualties in this studies) identified in hot spot regions; *N* is the total number of events; *m* is the length of the road network covered in hot spot regions; and *M* is the total length of road network in the study area. Higher values of PAI indicate a better performance of the technique for hot spot detection (e.g. Thakali *et al*., 2015).

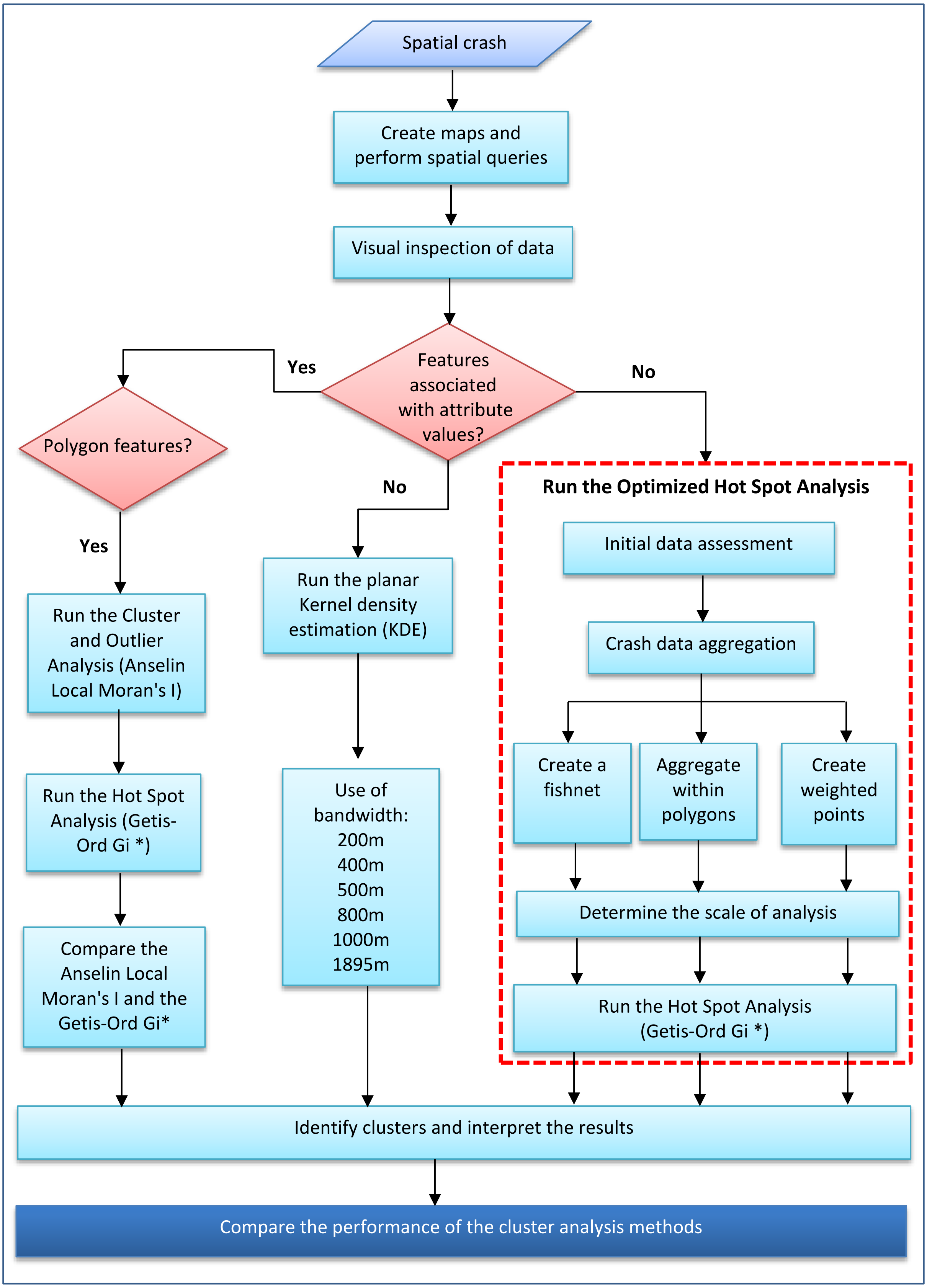


Figure 3-15: Flowchart of exploratory spatial data analysis in ArcMap

3.2.6.4 Multivariate analysis

This study applied two modelling techniques to investigate associations between the attributes of the built environment and the incidence of pedestrian crashes. These techniques are the Generalised linear modelling (GLM) and the geographically weighted regression (GWR) modelling. A description of these modelling techniques and the goodness of fit measures for the model comparison is provided in this section. A flowchart of exploratory spatial data analysis in ArcMap

1. **Generalised linear models (GLMs)**

Generalised Linear Models (GLMs) are the modelling techniques widely used in the context of traffic safety (Amoh-Gyimah *et al.*, 2017). Crash data are count data (i.e. On-negative integers) for which the distribution often follows a Poisson or negative binomial (NB) distribution (Washington *et al.*, 2010). A GLM usually comprises three components: a random component; a systematic component; and a link function that connects the random and systematic components to represent a linear prediction (Lord & Persaud, 2000). For the Poisson distribution, the underlying assumption is that the variance is equal to the mean of observations (Xu & Huang, 2015). The Poisson distribution is the most common starting point to model crash outcomes (Lord & Mannering, 2010; Washington *et al.*, 2010).

For the Poisson regression model applied in this study, the probability of a census suburb *i* (i.e. spatial unit of analysis) having y_i number of pedestrian casualties per unit of time period (i.e. three years in this study) is given by the following formula:

$$P(y_i) = \frac{exp(-\lambda_i){\lambda_i}^{y_i}}{y_i!} \tag{31}$$

where $P(y_i)$ is the probability of a census suburb i having y_i number of pedestrian casualties over a three-year period with the underlying Poisson mean λ_i. The term λ_i denotes the expected number of pedestrian casualties in a census suburb i over a three-year period. The Poisson parameter λ_i is modelled as a function of explanatory variables x_i, of which the most widely used functional form is written as follows:

$$log(\lambda_i) = \beta_0 + \beta x_i + \varepsilon_i \tag{32}$$

where β_0 is the intercept, x_i is a row vector of explanatory variables for census suburb i, β is a row vector of coefficient estimate of model covariates x_i and ε_i is a random residual term or the error term which reflects heterogeneity that accounts for unobserved factors and other random disturbances such as omitted explanatory variables and intrinsic randomness (Huang, Zhou, Wang, Chang & Ma, 2017).

However, the underlying assumption that the variance is equal to the mean is not often upheld when modelling count data such as traffic crashes. In many instances, the variance exceeds the mean of crash counts and this phenomenon is referred to as over-dispersion of the distribution. To account for the issue of over-dispersion, the negative binomial (NB) regression modelling has widely been applied. To estimate the NB regression model, Poisson parameter λ_i is specified as a function of explanatory variables plus a gamma-distributed error term (Amoh-Gyimah *et al.*, 2017). Using a log-linear function, the NB regression model can be specified as:

$$log(\lambda_i) = \beta_0 + \beta x_i + \varepsilon_i + \theta_i \tag{33}$$

Where θ_0 is a gamma-distributed error term with mean 1 and variance α; λ_i is the parameter of a Poisson distribution (i.e. expected number of pedestrian casualties in a census suburb i); β is a row vector of coefficient estimate of model covariates. The addition of the gamma-distributed error term allows the variance to differ from the mean to the extent that $var(n_i) = \lambda_i + \alpha{\lambda_i}^2$ (Amoh-Gyimah *et al.*, 2017; Xu & Huang, 2015).

2. Geographically Weighted Regression (GWR) models

In addition to the Generalised Linear Models, this study employed another modelling approach, the Geographically Weighted Regression (GWR) Model to quantify the associations between the aspects of the built environment and pedestrian crash incidence. The particularity of this modelling approach lies in its ability to address the issue of spatial correlation and stationary relationships between explanatory variables and outcome variable (Amoh-Gyimah *et al.*, 2017; Fotheringham *et al.*, 2002). This study applied the geographically weighted regression (GWR) modelling approach to develop models that are able to allow parameter estimates to vary across the study area.

The conventional approach to the empirical analyses of spatial data is to build a global model that assumes homogeneous (stationary) cross-spatial relationships between dependent and independent variables (Lewandowska-Gwarda, 2018). The global regression model is expressed mathematically in the following form:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \tag{34}$$

where y_i is the explanatory variable is observed in census suburb i; β_0 is the intercept; k is the total number of explanatory variables; β_k is the parameter of the kth explanatory variable observed in census suburb i; ε_i is the error term for the estimation in census suburb i. The row vector of coefficient estimate β_k is estimated globally (for the entire study area) and does not change across the study area. Therefore, this model is called a "global" model (Zhang *et al.*, 2015).

In the GWR modelling approach, local rather than parameters β are estimated by extending this traditional regression equation as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{35}$$

where (u_i, v_i) means the two-dimensional coordinates of the centroid of the ith location (i.e. census suburb in this study) in space; $\beta_k(u_i, v_i)$ is unknown functions of geographical locations and $\varepsilon_i$is the error term with mean zero and common variance σ^2. $\beta_k(u, v)$ are locally estimated at each location (u_i, v_i) by the weighted least-squares procedure in which some distance-decay weights are used. Each set of the estimated coefficients at n locations can produce a map of variation which may give useful information on non-stationarity of the regression relationship.

Spatial heterogeneity of relationships between explanatory variables and the outcome variable was explored through a comparison of estimates of local models and those in the global model developed for the entire study area. The global models were estimated by the use of Ordinary Least Square (OLS) tool implemented in ArcMap 10.3.1. With OLS, a single (i.e. global) coefficient is estimated for each explanatory variable and a number of other parameters (e.g. Standard errors, t-statistic, the probability (p-value) indicating statistical significance, Variance Inflation Factor (VIF) and robust probability) are generated in the output report as well. Unlike the global parameters estimated by OLS, GWR Models enables local variations in parameter estimates across the study area. This means that the regression coefficients β_k take different value for each geographic unit of analysis.

The initial step for running GWR Models and interpreting their results was to estimate parameters of the global models using candidate variables suggested by the Exploratory Regression tool. The global models were run by the use of the OLS tool implemented in ArcMap 10.3.1.

3.2.6.5 Crash data modelling process

This section presents steps followed in the Generalised Linear and Geographically Weighted Regression (GWR) Modelling processes as illustrated in Figure 3-17.

1. **Importing data into the modelling tools**

The first step of modelling process was importing prepared data into the statistical analysis tool. For the Generalised Linear Modelling procedure, data were imported from Excel spreadsheets into STATISTICA software tool. For the GWR modelling procedure, data were imported from Excel spreadsheets into ArcMap.

2. **Running the exploratory data analysis**

For the Generalised Linear Modelling, Exploratory Data Analysis (EDA) involved the following steps:

- Checking missing data and other mistakes;
- Univariate visualization of each variable of the dataset, with summary statistics;
- Bivariate visualization (e.g. correlational analysis) and summary statistics to study relationships between each exploratory variable and the outcome variable as well as relationships among exploratory variables (i.e. assessment of multicollinearity);
- Checking assumptions associated with model fitting (e.g. equality of mean and variance of data);
- Identifying the most influential variables;

For GWR modelling, Exploratory Spatial Data analysis (ESDA) was applied to spatial data and this entailed the following steps:

- Visualize spatial data through maps and graphs (e.g. choropleth maps, histograms, charts, boxplots, etc.);
- Identify patterns of spatial association and spatial clustering through spatial autocorrelation and regression analysis;

- Detect spatial outliers;
- Check assumptions associated with model fitting (e.g. a normal distribution of errors for GWR models);
- Examine other spatial patterns such as spatial heterogeneity across the study area.

3. Selecting candidate variable to include into the modelling process

i. <u>Selecting variables for Generalised Linear Models</u>

Factor analysis and correlational analysis were used to explore a set of 110 variables included in the overall dataset comprising information on pedestrian casualties and the aspects of the built environment all aggregated at the census suburb level. Factor Analysis was supplemented by Pearson correlation coefficient and Spearman test firstly to assess how each explanatory variable is correlated to the outcome variable and secondly to detect potential predictors with significant associations in the models. This section describes specifically the technique of Factor Analysis applied in this study to select the most influential variables to include into the Generalised Linear Modelling process.

<u>Definition of Factor Analysis</u>

Factor analysis is a mathematical technique used to reduce a large number of variables into a smaller set of latent variables (known as "factors") based on shared variance (Child, 2006; Loehlin & Beaujean, 2017; Yong & Pearce, 2013). "Factors" (also referred to as "latent variables") represent clusters of variables that correlate highly with each other (Field, 2013). Factor analysis consists of two main factor analysis techniques: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) (Child, 2006; Yong & Pearce, 2013). While the EFA attempts to uncover complex patterns of a large set of variables by exploring the underlying structure of the dataset and testing predictions, the CFA involves hypothesis testing techniques and uses path analysis diagrams to represent variables and factors (Child, 2006; Cudeck, 2000; Field, 2013; Hoyle, 2000).

<u>Choosing the Factor Analysis method</u>

Factor analysis was used in this study for two purposes: (1) to explore interrelationships among explanatory variables (i.e. identify variables that have more in common with each other) and (2) to reduce the number of explanatory variables into fewer variables that are manageable for the modelling process. Hence, the Exploratory Factor Analysis (EFA) was selected as the factor analysis technique appropriate for the purposes of the analysis.

Theoretical framework of Factor Analysis

The total variance of a particular variable consists of three components: variance that is shared with other variables (common variance); variance that is specific to that variable (unique variance); and error or random variance (sometimes referred to as unreliability of variance) which is also specific to one variable (Field, 2013; Yong & Pearce, 2013). The proportion of common variance present in a variable is referred to as the "communality". In such a way, a variable with no unique variance and error variance would have a commonality of 1 while a variable that shares none of its variance with other variables would have a communality of 0 (Field, 2013; Yong & Pearce, 2013). The concept of communality is very important in factor analysis as the analysis is concerned by finding common variance (i.e. communalities between variables). Accordingly, variables with low communalities (less than 0.20 so that 80% variance is unique) are usually eliminated from the analysis.

Graphical representation

Factors are represented in a coordinate system by the axes, and variables are lines or vectors (Cattell, 1973). When a variable is in close proximity to a certain factor, this means that the variable is associated with that particular factor. The strength of the relationship between the variable and each factor is portrayed by the coordinates of that variable along each axis. Variables that have large coordinates on the same axis are assumed to measure different aspects of some common underlying dimension. The coordinate of a variable along a classification axis is known as a "factor loading" (Field, 2013). For instance, Figure 3-16 illustrates factor analysis of untransformed demographic variables included in the modelling process in this study. All 13 variables can reduced to two factors. The circles represent clusters of variables that have higher loading factor loadings (i.e. a strong correlation) which can be represented by a single variable (or a factor). The first cluster loads onto factor 1 and consists of (1) population number; (2) number of households; (3) population younger than 15 years old; (4) population in the 15-24 age group; and (5) population in the 25-54 age group. The second cluster loads onto factor 2 and includes: (1) Coloured population; (2) population with other ethnicity; and (3) population aged 55 years and older. When there are three factors within the data, a third factor is represented by a third axis, creating thus a 3-D graph.

Variable	Factor Loadings (Varimax normalized)	
	Factor 1	Factor 2
Popu	0.956279	0.278028
Popu_dens	0.392989	-0.172718
Popu_Black	0.904108	-0.233149
Popu_Coloured	0.461616	0.634548
Popu_Asian	0.064957	0.684542
Popu_White	-0.037284	0.537824
Popu_Other	0.659728	0.577128
Households	0.963505	0.207644
Avg_HH	-0.000005	-0.156662
Age_less15	0.967557	0.220568
Age15_24	0.976515	0.188694
Age25_54	0.961014	0.259051
Age55_plus	0.687924	0.671987
Expl.Var	6.755139	2.325658
Prp.Totl	0.519626	0.178897

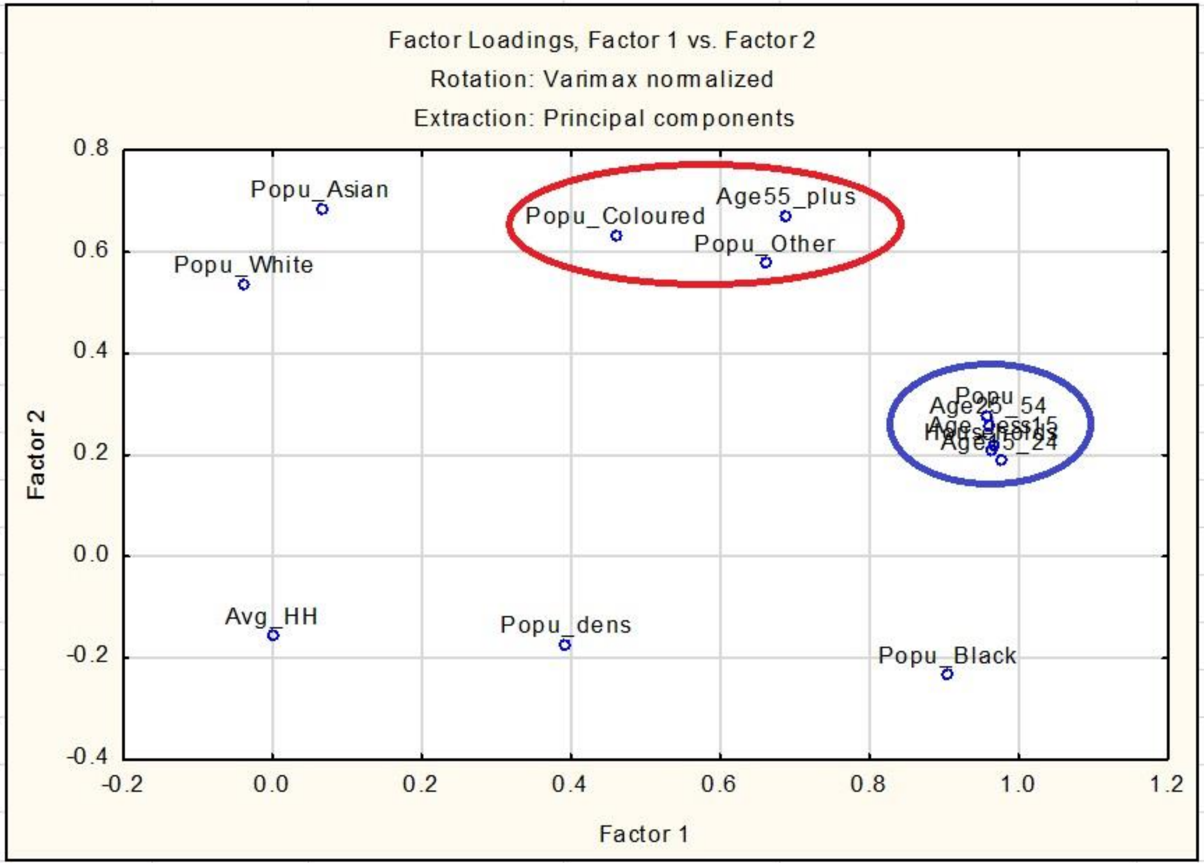


Figure 3-16: Example of graphical representation of factor loading for demographic variables

Procedure for Factor Analysis

Without going into the detail about the computational aspects of factor analysis, 5 main steps involved in factor analysis are (1) descriptive statistics; (2) factor extraction; (3) factor rotation; (4) factor scores; and (5) Options. Further computational details for these steps can be found in Field (2013) and in Yong and Pearce (2013). Factor Analysis was run by using STATISTICA software tool.

ii. Selecting variables for GWR Models

Before the modelling process, a test for multicollinearity between candidate explanatory variables was run by using the Exploratory Regression tool implemented in ArcMap. This tool runs diagnostic tests on the candidate explanatory variables and provides a summary which is useful in choosing variables with greater performance (highest adjusted R-squared values). The tool also highlights variables which exhibit severe multicollinearity. The Ordinary Least Square (OLS) tool implemented in ArcMap was applied to the best passing models to provide global models (with a single coefficient for each variable included).Subsequently, the same models were run using the Geographically Weighted Regression tool also implemented in ArcMap. Using the t-test, the local estimates from GWR models were compared with global estimates from OLS models to evaluate spatial heterogeneity of relationship across the study area. The absence of spatial heterogeneity was confirmed when the t-test showed non-significant results at 95 percent confidence interval (i.e. $p>0.05$) for a particular explanatory variable.

4. Data transformation

The Exploratory Data Analysis (ESDA), Exploratory Spatial Data Analysis (ESDA) and Factor Analysis helped to identify variables with distributional problem (problems with normality), outliers, lack of linearity or unequal variance and multicollinearity among exploratory variables. Data transformation was adopted to correct these problems and to improve model results by allowing models to include a bigger number of explanatory variables. Two types of transformation were applied to explanatory variables: Log transformation for the variable "population" and the use of proportion for variables that can be broken down into categories. However, certain variables were not transformed using proportions if they represent a single category of an aspect (e.g. roundabouts/mini-circles, ratio of intersections to culs-de-sacs, population density street density and entropy index) or if the variable is a specific measure or proxy of an a particular aspect (e.g. number of intersections with more than three legs is used as a proxy of urban design). As certain census suburbs have no population (zero population), a log value of zero could not be computed. Subsequently, a constant of 1 was added to all population data (adding 1 to the population number of each census suburb). Therefore, the log transformation used the following formula:

$$Log(population) = log(population\ number + 1) \tag{36}$$

5. Selecting the best performing models

For the Generalised Linear Modelling, numerous test trials were conducted in order to find out the best model for each dataset of pedestrian casualties. For each candidate model, the prediction performance was assessed by using a number of goodness-of-fit measures which tell how well the model fits the data. For the Poisson Regression model and the Negative Binomial model, the STATISTICA software tool generated a number of goodness-of-fit measures such as the ratio of the deviance to the degrees of freedom (deviance/df), the Pearson chi-Square, the Akaike's Information Criterion (AIC), the Bayesian Information criterion (BIC) and the corrected version of the Akaike's Information Criterion (AICc). Competing models were compared using these measures of goodness-of-fit. A model that provided a ratio of the deviance to the degree of freedom close to one was chosen as a better fit to the observed data (El-Basyouny & Sayed, 2009; Kim *et al*., 2007). When comparing the performance of many models, AIC, AICc and BIC are the measures often used to compare the performance of models. The model with lowest values of AIC, AICc or BIC is regarded as the best model since higher values are indicative of a failure of model to fit the data (Burnham, Anderson & Huyvaert, 2011; Washington *et al.*, 2010). Following this, the final GLM models presented in this studies are the ones that had the lowest values of AIC, AICc or BIC or simply put, the best performing models.

With respect to GWR modelling, the goodness-of-fit measures generated by the GWR tool include the local R^2 and the corrected Akaike's Information Criterion (AICc). The GWR model performance can also be assessed by visualizing mapped residuals (e.g. raw residuals or standardised residuals).

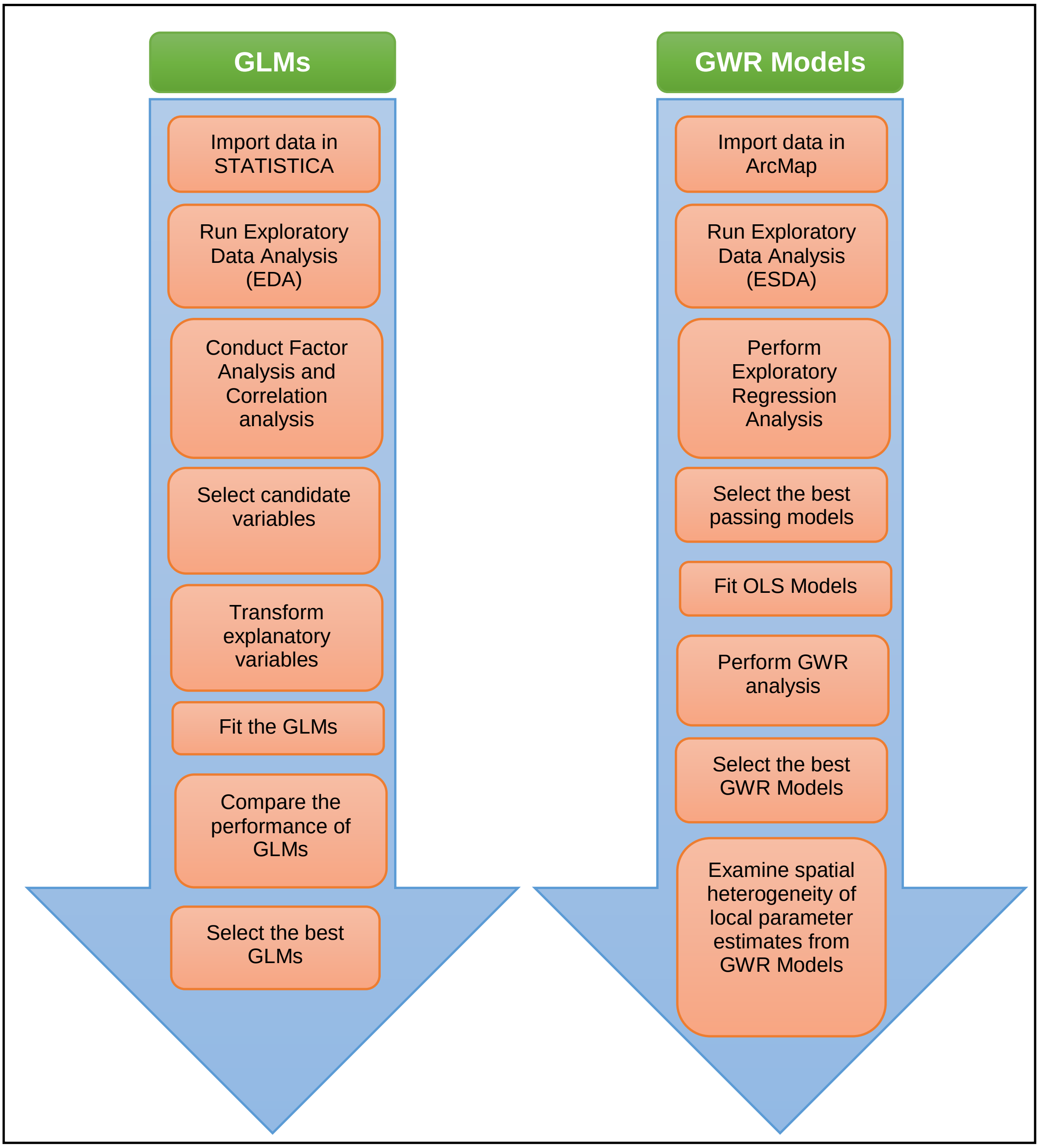


Figure 3-17: Steps of modelling process followed in the study

Chapter 4: Results and discussions

This section presents the results and discussions from the three types of analysis applied in this study. The analysis involved three analytical approaches which are:

- Univariate and bivariate analyses
- Geospatial analysis and
- Multivariate analysis.

4.1 Univariate and bivariate analyses of pedestrian casualties

4.1.1 Analysis of pedestrian casualty frequency

4.1.1.1 Temporal variation of pedestrian casualty frequency

Crash data analysed in this study comprises 13 853 pedestrian casualties (i.e. killed, injured and those whose injury severity was unknown or not recorded) collected by the police in Cape Town between 2012 and 2014. Of these pedestrian casualties, 4 672 casualties (33.7 percent) were recorded in 2012, 4 529 casualties (32.7 percent) were recorded in 2013 and 4 652 casualties (33.6 percent) occurred in 2014. In terms of annual crash frequency, these figures are equivalent to an average crash frequency of 4 618 pedestrians involved in collisions with vehicles each year in the Cape Town area. An analysis of annual frequencies of pedestrian causalities recorded in Cape Town between 2005 and 2014 shows a decrease of 28.56 percent (from 6 512 pedestrian casualties in 2005 to 4 652 pedestrian casualties in 2014) However, trends in pedestrian casualties did not vary significantly over the last 3-year period selected for this study (see Figure 4-1). By using the Cape Town's population size (i.e. the population size in 2013) as a proxy measure of pedestrian exposure, an annual rate of 123.5 pedestrian casualties per 100 000 population is found for the entire study area.

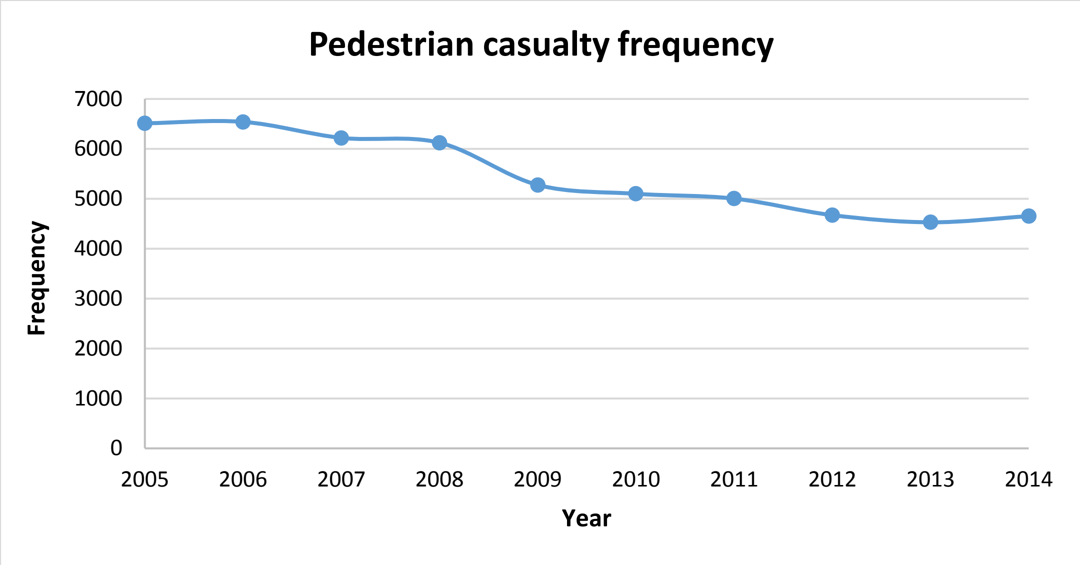


Figure 4-1: Frequency of pedestrian casualties in Cape Town for the 2005-2014 period

4.1.1.2 Pedestrian casualty frequency by gender and age

As expected, male pedestrians were found overrepresented in pedestrian crashes. The analysed crash data comprise 6 274 (45.3 percent) male pedestrian victims and 3 761 female victims. These figures indicate a male-to-female ratio of 1.67 to 1. The gender of the victim was recorded as "unknown" for 3 761 (27.0 percent) pedestrians involved in road traffic crashes. The distribution of pedestrian casualties by gender is illustrated in Figure 4-2.

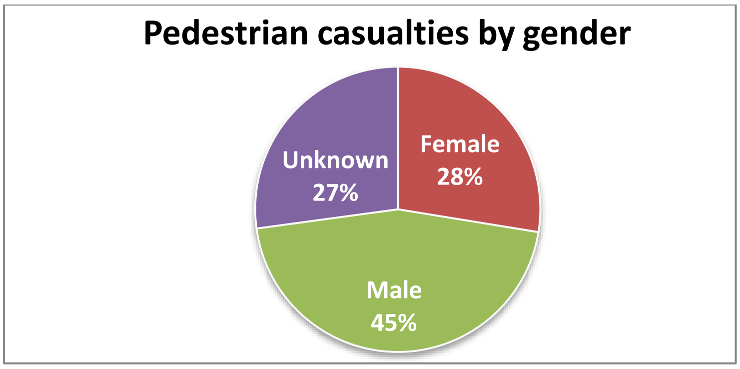


Figure 4-2: Pedestrian casualties by gender

The analysed crash data comprises 8 310 pedestrian casualty cases with age of 0 years records and these constitute 60 percent of all reported pedestrian casualties. With the exclusion of these cases (since they appear to be unrealistic), the mean age for the remaining 5 543 cases is 29.5 years (SD=18.532). The lowest age observed for the analysed crash data is 1 year and the highest age is 108 years. It is not certain whether this is accurate- more likely it is an error and should have been either 18 years or 10 years. The analysis of casualties according to the age of pedestrians is carried out at three levels of analysis: by analysing the entire sample of pedestrian casualties, by restricting the analysis to killed and seriously injured (KSI) pedestrians and by considering only cases in which injuries were reported as fatal (i.e. coded as "Killed"). In addition, gender was also included in the analysis of these three datasets. This section presents only the analysis of the entire sample of pedestrian casualties and further analyses related to injury severity are presented later in Section 4.1.2.

Figure 4-3 presents the distribution of pedestrian casualties according to age groups and gender. It is important to note that the overall figures include cases with unknown gender which are not represented in Figure 4-3. The results show disproportionate casualty frequencies across age groups and genders. Looking at the overall figures, the group with the highest casualty frequency are pedestrians aged between 6 and 10 years, followed by those aged between 26 and 30 years and then the 31-35 year group. Interestingly, the casualty frequencies for 11-15 and 16-20 age groups are far lower than that of adjacent age groups in which pedestrian casualty frequency peaks. The frequency of casualties reduces dramatically among pedestrians older than 55 years old and part of this reduction may be due to the lower rate of exposure they experience.

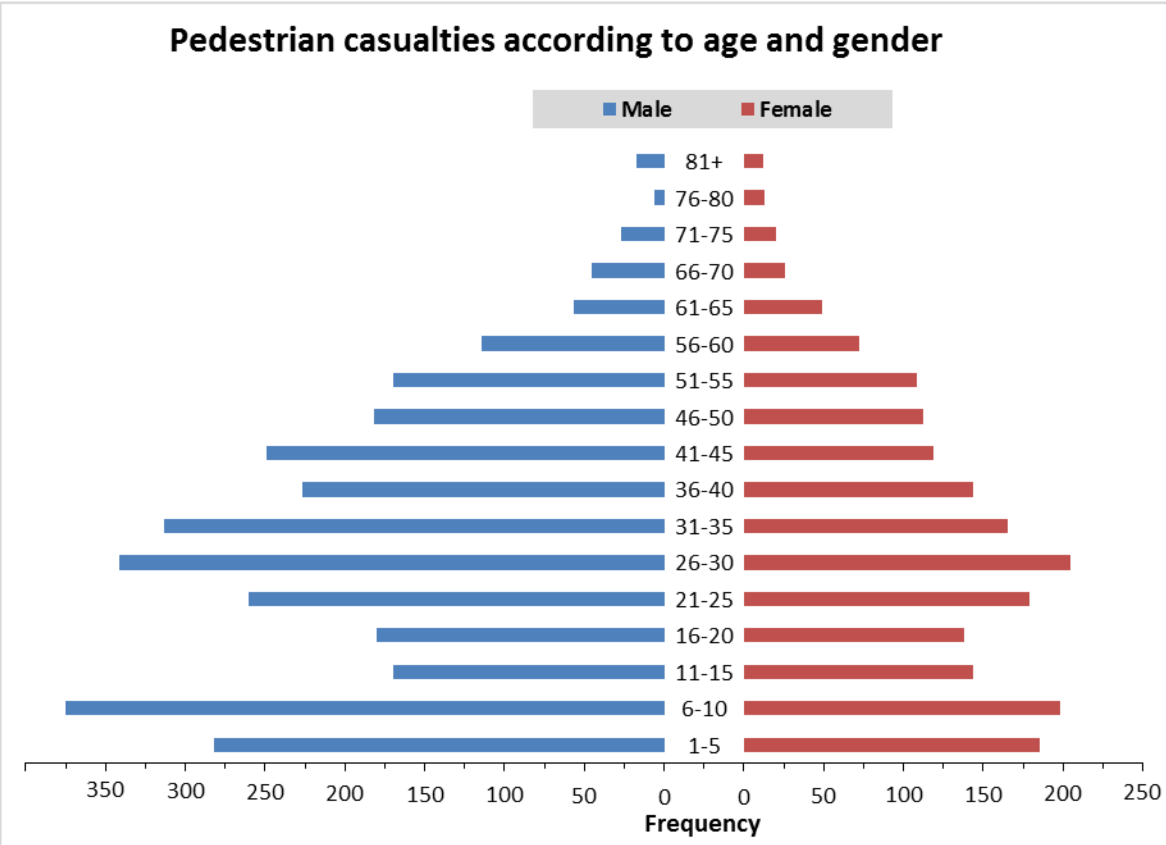


Figure 4-3: Pedestrian casualties by age and gender

Figure 4-3 and Table 4-1 show a disproportionate casualty frequency between male and female pedestrians. The male-to-female ratio for pedestrian casualties varies across the age groups (see Table 4-11) with values greater than one in almost all age groups. Examining the top five age groups with the highest male-to-female ratio for pedestrian casualties (41-45 age group), the casualty frequency among males is nearly double that of females. Males are also at a higher casualty risk than females in the 31-35 age group (M:F=1.90), the 6-10 age group (M:F=1.89), the 66-70 age group (M:F=1.73) and the 26-30 age group (M:F=1.66). Interestingly, the casualty frequency among females emerges approximately twice as high as that of males in the 76-80 age group, however here the actual numbers are very small.

Table 4-1: Pedestrian casualties by age and gender

Age group	Female	Male	Unknown	Total	M:F
1-5	185	282	66	533	1.52
6-10	198	375	89	662	1.89
11-15	144	170	44	358	1.18
16-20	138	180	30	348	1.30
21-25	179	260	48	487	1.45
26-30	205	341	84	630	1.66
31-35	165	313	66	544	1.90
36-40	144	227	50	421	1.58
41-45	119	249	44	412	2.09
46-50	112	182	40	334	1.63
51-55	108	170	28	306	1.57
56-60	72	114	21	207	1.58
61-65	49	57	18	124	1.16
66-70	26	45	3	74	1.73
71-75	20	27	5	52	1.35
76-80	13	6	2	21	0.46
81+	12	17	1	30	1.42

4.1.1.3 Pedestrian casualty frequency by time of day

The analysis of pedestrian casualties by time of crash occurrence indicates that pedestrians are at the highest risk of being involved in road crashes during peak traffic times. As illustrated in Figure 4-4, the highest incidence of pedestrian crashes is observed between 7:00 AM and 8:00 AM. Seven percent of the recorded pedestrian casualties occurred during this time period. Another peak of pedestrian casualties emerges during early evening hours, stretching over the time period from 4:00 PM until 7:00 PM. The lowest incidence of pedestrian crashes is observed during early hours of the morning from 3:00 AM until 5:00 AM (see Figure 4-4).

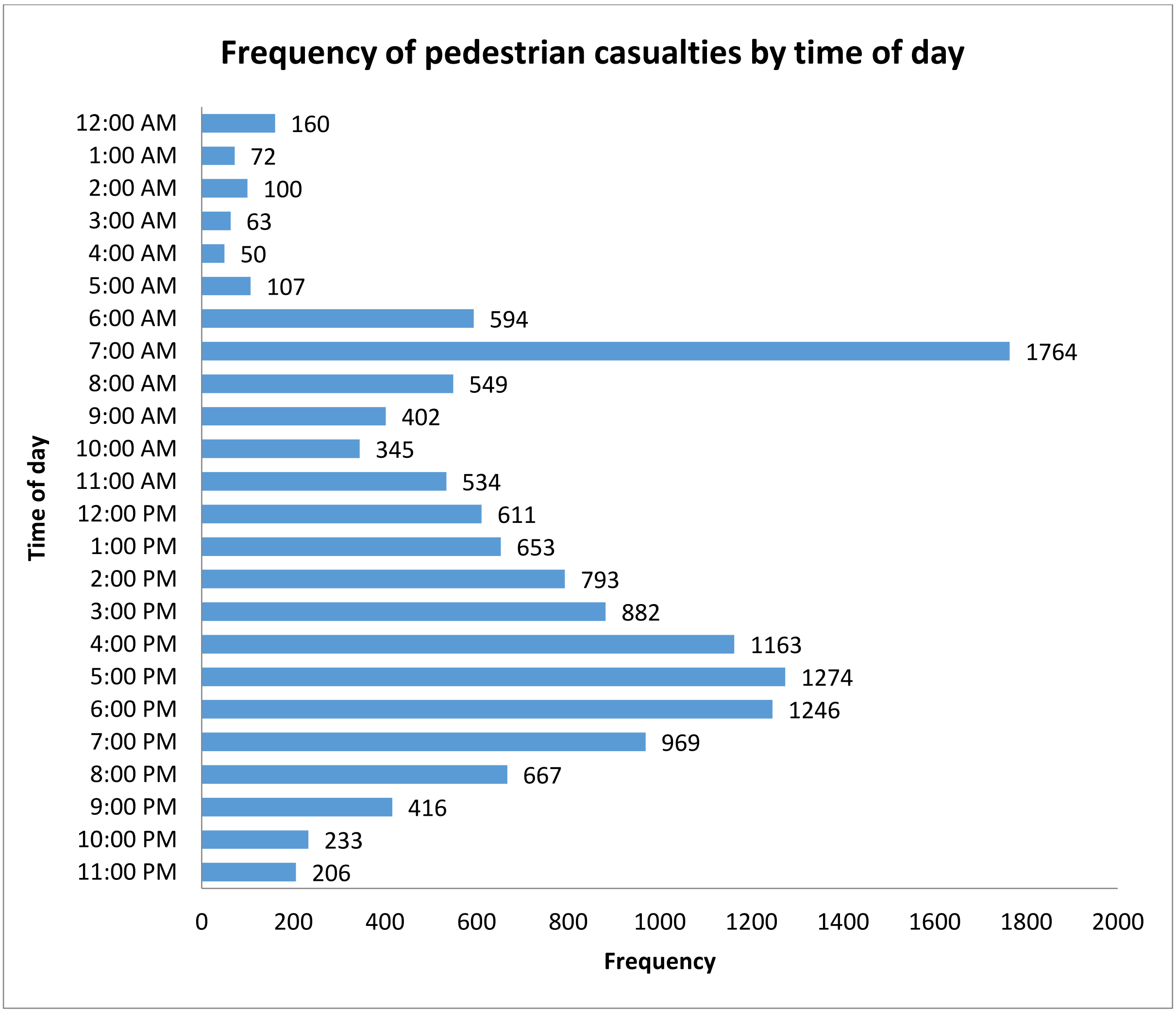


Figure 4-4: Pedestrian casualties by time of day in Cape Town

4.1.1.4 Pedestrian casualty frequency by day of week

Figure 4-5 shows the distribution of daily pedestrian casualty frequency over a virtual week, with a peak being identified over Saturday (2 340 pedestrian casualties) and Friday (2 335 pedestrian casualties). The lowest frequencies is identified over Wednesday (1 751 pedestrian casualties) and over Thursday and Sunday (1 821 pedestrian casualties).

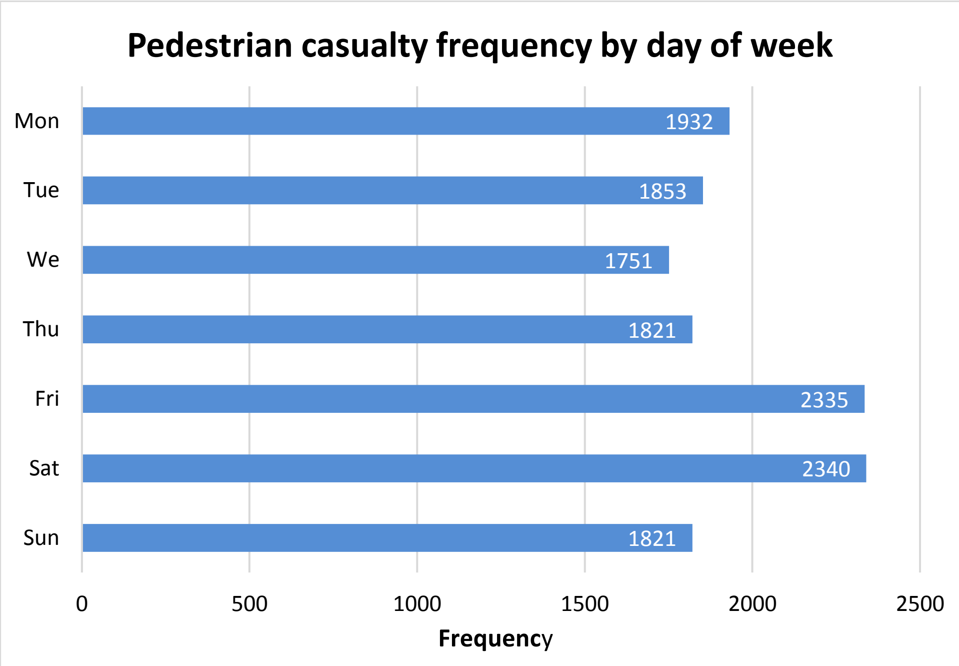


Figure 4-5: Pedestrian casualty frequency by day of week

After identifying a disproportionate distribution of pedestrian casualties over the week, the daily frequencies of pedestrian casualties were generated from the 3-year casualty dataset and mean daily frequencies were calculated. The descriptive statistics of the daily counts of pedestrian casualties is presented in Table 4-2 and Figure 4-6.

Table 4-2: Descriptive statistics of daily pedestrian casualty counts

Day of week	Mean	Std. Deviation	N
Monday	12.58	4.798	146
Tuesday	11.74	3.569	151
Wednesday	11.05	3.746	149
Thursday	11.64	4.372	151
Friday	15.05	4.429	149
Saturday	15.01	5.375	153
Sunday	11.54	4.347	153
Holidays	12.16	7.123	44
Total	12.64	4.790	1096

Holiday dates were included in the analysis to allow more insights on crash risk on these particular days. It is worth noting that this analysis included only 12 national public holidays for each year of the 2012-2014 period as determined by the Public Holidays Act (Act no36 of 1994) (South African Government, n.d.). For instance, the analysis did not include all dates of school holidays or Christmas holiday season. Data on holiday dates was collected from the official website of the South African Government (South African Government, n.d.) and were checked using other websites such as www.timeanddate.com.

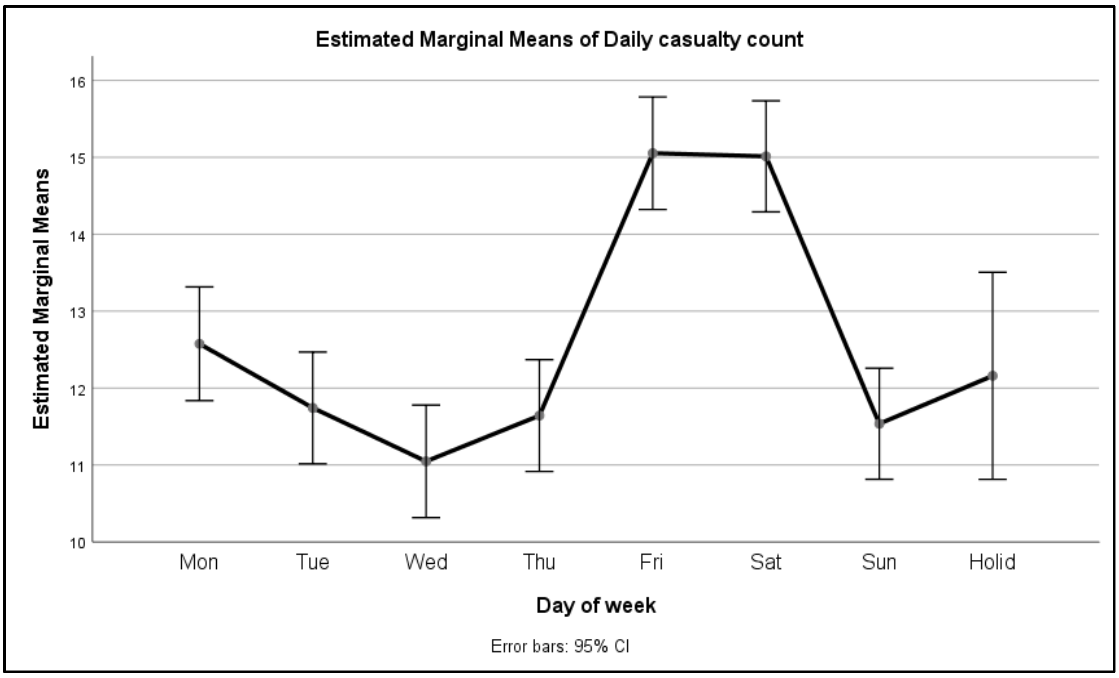


Figure 4-6: Estimated Marginal Means of daily casualty count

Individual mean differences were analysed using the Analysis of Variance (ANOVA) test to ascertain whether there a statistically significant difference among the mean values of daily pedestrian casualties. Two underlying assumptions of the ANOVA test, namely the normality of distributions and homogeneity of variance, were tested in IBM-SPSS Statistics and the test results determined the choice of the Post-hoc test applied in the analysis. The Post-hoc test provided specific information on which means are statistically different form each other. Results from the Levene's test for homogeneity of variance are presented in Table 4-3.

Table 4-3: Levene's Test for homogeneity of variance

Levene's Test of Equality of Error Variances[a]					
		Levene Statistic	df1	df2	Sig.
Daily casualty count	Based on Mean	6.399	7	1088	0.000
	Based on Median	5.372	7	1088	0.000
	Based on Median and with adjusted df	5.372	7	861.627	0.000
	Based on trimmed mean	5.892	7	1088	0.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Daily casualty count

The results in Table 4-3 indicate that the Levene's test is significant at the 5% level (i.e. $p<0.05$) which implies that the null hypothesis (i.e. the variance is equal across different days of week) is rejected and the assumption of homogeneity of variance is negated. As presented previously in Section 3.2.6.2 and in Figure 3-11 (on Page 101), the choice of an appropriate Post Hoc test depends on the assumptions of equal variances and equal group sample sizes. For this analysis, the sample size differs (i.e. numbers in the column labelled "N" in Table 4-2) across different groups which are days of week in this particular context. Following the approach presented in Figure 3-11, the Games-Howell test was chosen as an appropriate post hoc procedure to test mean differences when variances and group sample sizes are not equal. The results from the Games-Howell post hoc test are presented in Table 4-4. The Games-Howell post hoc procedure compared means of all groups (days of week) with each other. Marked p-values (labelled as "Sig.") are significant at the 5% level ($p<0.05$).

Overall, the mean values of daily pedestrian casualty counts were found consistent over Mondays, Tuesdays, Wednesdays, Thursdays, Sundays and holidays (i.e. mean differences over these days are not significant at the 5% level). Pedestrian casualties occurred more frequently over Fridays and Saturdays and the mean daily frequencies on these days were similarly higher compared with other days of week ($p<0.05$). The results from the Games-Howell post hoc test are illustrated in Table 4-4. The descriptive statistics and the post hoc test on the dataset of daily pedestrian casualty counts show apparently two distinct groups; the Friday and Saturday group and the rest of other days of week. The frequency of pedestrian casualties emerges to be particularly higher over Fridays and Saturdays. Pedestrian casualty frequency is seen to be consistent through the rest of other days of week.

Table 4-4: Results from the Games-Howell Post Hoc Test

(I) Day of week		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Mon	Tue	0.83	0.492	0.691	-0.67	2.34
	Wed	1.53	0.502	0.051	0.00	3.06
	Thu	0.93	0.533	0.654	-0.69	2.56
	Fri	-2.48*	0.538	0.000	-4.12	-0.84
	Sat	-2.44*	0.589	0.001	-4.23	-0.64
	Sun	1.04	0.530	0.511	-0.58	2.66
	Holid	0.42	1.145	1.000	-3.19	4.02
Tue	Mon	-0.83	0.492	0.691	-2.34	0.67
	Wed	0.69	0.423	0.723	-0.60	1.98
	Thu	0.10	0.459	1.000	-1.30	1.50
	Fri	-3.31*	0.465	0.000	-4.73	-1.89
	Sat	-3.27*	0.523	0.000	-4.87	-1.67
	Sun	0.21	0.456	1.000	-1.19	1.60
	Holid	-0.42	1.112	1.000	-3.94	3.10
Wed	Mon	-1.53	0.502	0.051	-3.06	0.00
	Tue	-0.69	0.423	0.723	-1.98	0.60
	Thu	-0.60	0.470	0.910	-2.03	0.84
	Fri	-4.01*	0.475	0.000	-5.46	-2.56
	Sat	-3.97*	0.532	0.000	-5.59	-2.34
	Sun	-0.49	0.467	0.967	-1.91	0.94
	Holid	-1.11	1.117	0.973	-4.64	2.42
Thu	Mon	-0.93	0.533	0.654	-2.56	0.69
	Tue	-0.10	0.459	1.000	-1.50	1.30
	Wed	0.60	0.470	0.910	-0.84	2.03
	Fri	-3.41*	0.508	0.000	-4.96	-1.86
	Sat	-3.37*	0.562	0.000	-5.09	-1.66
	Sun	0.11	0.500	1.000	-1.42	1.63
	Holid	-0.52	1.131	1.000	-4.09	3.05
Fri	Mon	2.48*	0.538	0.000	0.84	4.12
	Tue	3.31*	0.465	0.000	1.89	4.73
	Wed	4.01*	0.475	0.000	2.56	5.46
	Thu	3.41*	0.508	0.000	1.86	4.96
	Sat	0.04	0.566	1.000	-1.69	1.77
	Sun	3.52*	0.505	0.000	1.98	5.06
	Holid	2.89	1.134	0.196	-0.68	6.47
Sat	Mon	2.44*	0.589	0.001	0.64	4.23
	Tue	3.27*	0.523	0.000	1.67	4.87
	Wed	3.97*	0.532	0.000	2.34	5.59
	Thu	3.37*	0.562	0.000	1.66	5.09
	Fri	-0.04	0.566	1.000	-1.77	1.69
	Sun	3.48*	0.559	0.000	1.77	5.18
	Holid	2.85	1.158	0.232	-0.79	6.50
Sun	Mon	-1.04	0.530	0.511	-2.66	0.58
	Tue	-0.21	0.456	1.000	-1.60	1.19
	Wed	0.49	0.467	0.967	-0.94	1.91
	Thu	-0.11	0.500	1.000	-1.63	1.42
	Fri	-3.52*	0.505	0.000	-5.06	-1.98
	Sat	-3.48*	0.559	0.000	-5.18	-1.77
	Holid	-0.62	1.130	0.999	-4.19	2.94
Holid	Mon	-0.42	1.145	1.000	-4.02	3.19
	Tue	0.42	1.112	1.000	-3.10	3.94
	Wed	1.11	1.117	0.973	-2.42	4.64
	Thu	0.52	1.131	1.000	-3.05	4.09
	Fri	-2.89	1.134	0.196	-6.47	0.68
	Sat	-2.85	1.158	0.232	-6.50	0.79
	Sun	0.62	1.130	0.999	-2.94	4.19

Based on observed means.
The error term is Mean Square (Error) = 20.718.
*. The mean difference is significant at the .05 level.

4.1.1.5 Pedestrian casualty frequency by week of month

Three categorical variables termed "pay week", "second week after pay week" and "other week" were included in the analysis of weekly pedestrian casualty counts to assess fluctuations of pedestrian casualty frequency depending on financial situation. Weekly pedestrian casualty frequency is defined as a sum of pedestrian casualties occurring from Monday at 00:00 AM to Sunday at 11:59 PM. "Pay week" is defined as the week (from Monday to Sunday) that contains the first date on a month (e.g. 1st of March). The "second week after pay week" is defined as the week following the pay week. "Other week" denotes the remaining weeks of a month other than "pay week" and "second week after pay week". The descriptive statistics of weekly pedestrian casualty frequencies is presented in Table 4-5.

Table 4-5: Descriptive statistics of weekly pedestrian casualty count

Dependent Variable: weekly count of pedestrian casualties			
Weekly financial status	Mean	Std. Deviation	N
Pay week	98.19	12.801	36
2nd week after pay week	87.29	13.503	35
Other week	85.09	12.186	85
Total	88.61	13.638	156

The mean differences among the three groups of categorical variables were tested using the ANOVA test. The results from the Levene's test for homogeneity of variance demonstrate that the test is not significant ($p>0.05$), suggesting that the null hypothesis (i.e. equal variance across the groups) is valid (see Table 4-6).

Table 4-6: Levene's Test for homogeneity of variance

Levene's Test of Equality of Error Variances[a]					
		Levene Statistic	df1	df2	Sig.
weekly count of pedestrian casualties	Based on Mean	0.245	2	153	0.783
	Based on Median	0.225	2	153	0.799
	Based on Median and with adjusted df	0.225	2	151.094	0.799
	Based on trimmed mean	0.241	2	153	0.786

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Dependent variable: weekly count of pedestrian casualties

Following the approach of post hoc procedure presented in Figure 3-11, the Bonferroni post hoc test was chosen for its suitability to test individual mean differences across groups with equal variances and unequal sample sizes.

Table 4-7: Results from the Bonferroni Post Hoc Test

Multiple Comparisons						
Dependent Variable: weekly count of pedestrian casualties						
Bonferroni						
(I) Weekly financial status		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Pay week	2nd week after pay week	10.91*	2.998	0.001	3.65	18.17
	Other week	13.10*	2.512	0.000	7.02	19.18
2nd week after pay week	Pay week	-10.91*	2.998	0.001	-18.17	-3.65
	Other week	2.19	2.537	1.000	-3.95	8.33
Other week	Pay week	-13.10*	2.512	0.000	-19.18	-7.02
	2nd week after pay week	-2.19	2.537	1.000	-8.33	3.95

Based on observed means.
The error term is Mean Square (Error) = 159.530.
*. The mean difference is significant at the .05 level.

The Bonferroni post hoc test indicates that the mean value of weekly casualty counts over "pay weeks" is statistically different from that of other weeks of the month (i.e. "2nd week after pay week" and "other weeks") at the 5% level. The test also shows that the individual mean differences between "2nd week after pay week" and "other weeks" are not statistically significant at the 5% level. This finding suggests that pedestrian casualties occurred more frequently over pay weeks compared with other weeks of the month. The mean differences across the three groups are illustrated in Figure 4-7.

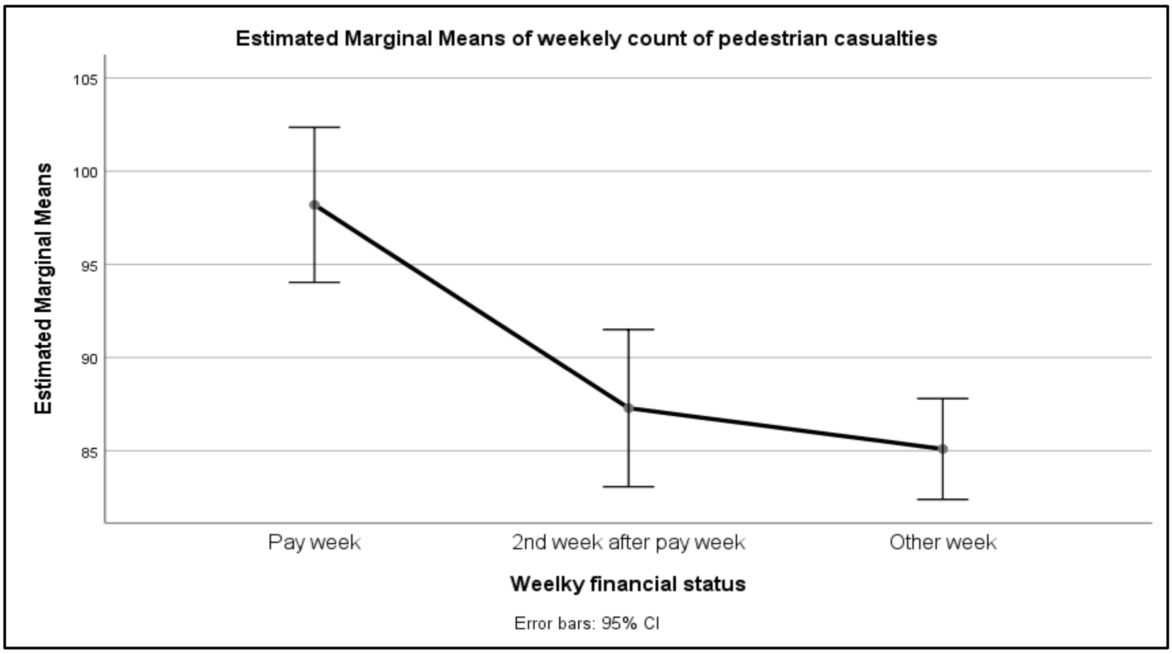


Figure 4-7: Estimated Marginal Means of weekly count of pedestrian casualties

4.1.1.6 Pedestrian casualty frequency by quarters of calendar year

The 3-year dataset of weekly pedestrian casualties was broken down into four quarters of the calendar year to assess seasonal trends of pedestrian casualty frequencies. As illustrated in Figure 4-8, each quarter contains 13 weeks and each year comprises four quarters (labelled "1st quarter", 2nd quarter" "3rd quarter" and "4th quarter"). For the 3-year study period, the sample size for each quarter comprises 39 variables which are weekly pedestrian casualty counts (i.e. the sample size for each individual quarter is 39).

A plot of weekly counts of pedestrian casualties over different quarters of calendar year and for each year included in the analysis is displayed in Figure 4-8. The three plots displayed in this figure show weekly fluctuations in pedestrian casualties. On average, it can be noticed that higher weekly counts of pedestrian casualties are predominantly observed in the third quarter of calendar year. A glance at the three plots of the weekly frequencies of pedestrian casualties shows that overall fluctuations (i.e. difference between the least and highest weekly frequencies) emerge to be more pronounced over the third quarter of calendar year. However, when the plots for each year are analysed separately, the highest weekly frequencies of pedestrian casualties as well as more marked fluctuations are identified over the last quarter of the year 2014.

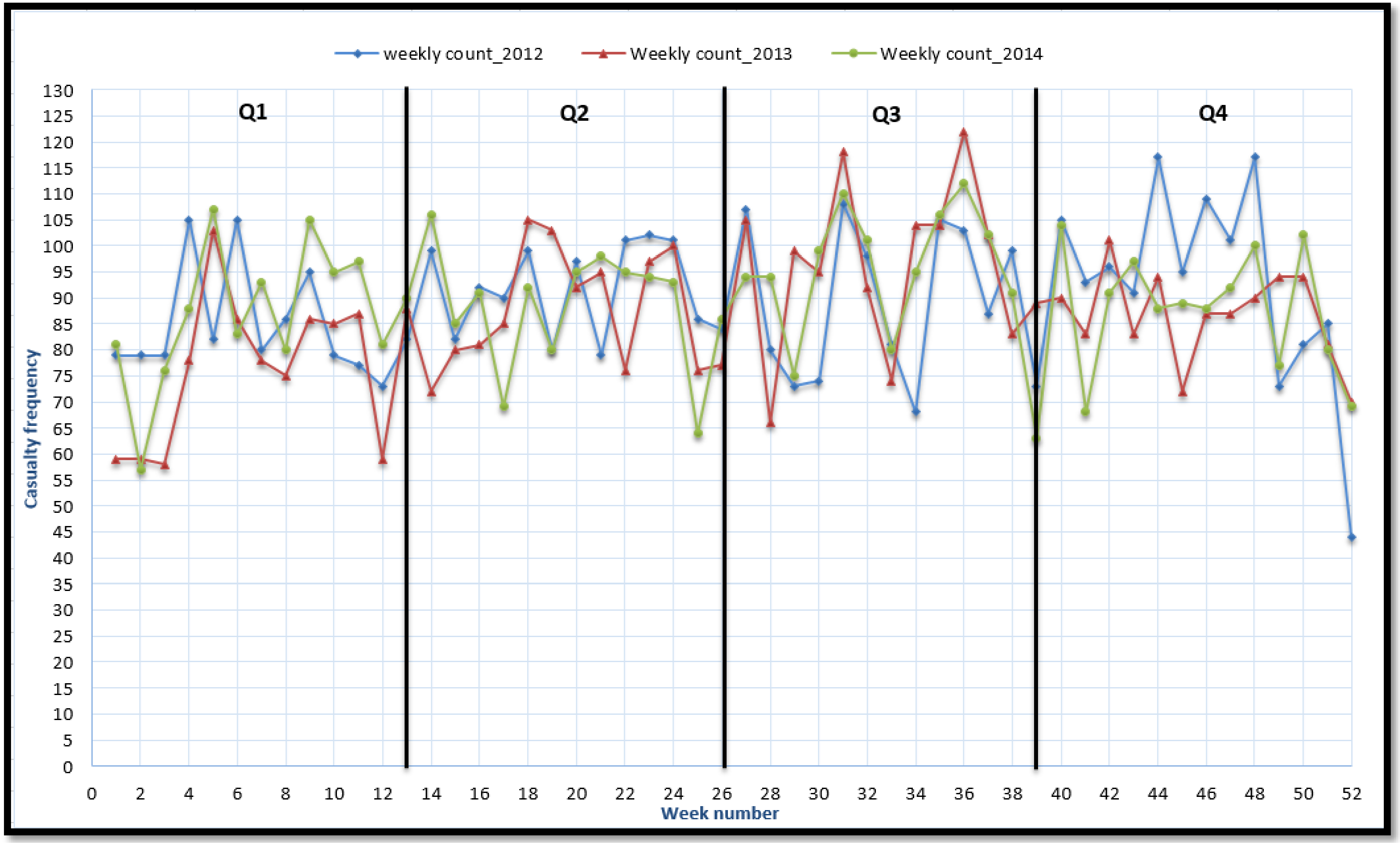


Figure 4-8: Weekly counts of pedestrian casualties across the quarters of calendar year

Table 4-8: Descriptive statistics of quarterly pedestrian casualty count

Dependent Variable: Weekly count of pedestrian casualties			
Quarter of calendar year	Mean	Std. Deviation	N
1st quarter	82.95	13.157	39
2nd quarter	89.21	10.556	39
3rd quarter	93.10	14.906	39
4th quarter	89.18	14.056	39
Total	88.61	13.638	156

The descriptive statistics of quarterly casualty data is presented in Table 4-8. The mean weekly counts of pedestrian casualties peaks over the 3rd quarter of calendar year (i.e. from July to September). The 3rd quarter includes two winter season months (July and August) with the month of September falling into the spring season in South Africa. The mean values of quarterly pedestrian casualties plotted in Figure 4-9 demonstrate a temporal variation in pedestrian casualty frequencies over different quarters of calendar year, with a peak being detected over the third quarter and a minimum mean value over the first quarter.

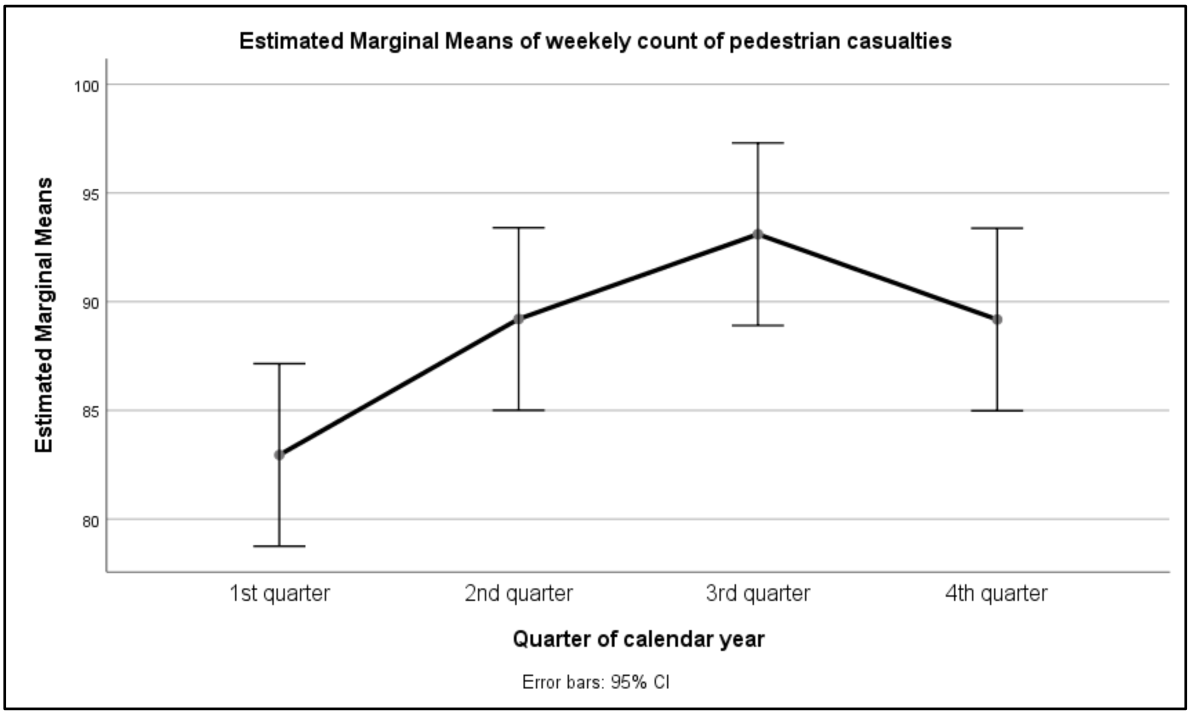


Figure 4-9: Estimated Marginal Means of Weekly count of pedestrian casualties

The Levene's test for homogeneity of variance is presented in Table 4-9. The test is not significant at the 5% level and the null hypothesis that variances across different quarters of the calendar year are equal is accepted.

Table 4-9: Levene's test for homogeneity of variance

Levene's Test of Equality of Error Variances[a]					
		Levene Statistic	df1	df2	Sig.
weekly count of pedestrian casualties	Based on Mean	1.253	3	152	0.293
	Based on Median	1.090	3	152	0.355
	Based on Median and with adjusted df	1.090	3	142.369	0.356
	Based on trimmed mean	1.254	3	152	0.292

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Dependent variable: weekly count of pedestrian casualties

The Bonferroni post hoc procedure was carried out to test whether individual means across different quarters of calendar year are statistically significant. The results from this test are indicated in Table 4-10.

Table 4-10: Results from the Bonferroni Post Hoc test for quarterly pedestrian casualty data

Multiple Comparisons						
Dependent Variable: Weekly count of pedestrian casualties						
Bonferroni						
(I) Quarter of calendar year		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1st quarter	2nd quarter	-6.26	3.005	0.234	-14.29	1.78
	3rd quarter	-10.15*	3.005	0.006	-18.19	-2.12
	4th quarter	-6.23	3.005	0.239	-14.26	1.80
2nd quarter	1st quarter	6.26	3.005	0.234	-1.78	14.29
	3rd quarter	-3.90	3.005	1.000	-11.93	4.14
	4th quarter	0.03	3.005	1.000	-8.01	8.06
3rd quarter	1st quarter	10.15*	3.005	0.006	2.12	18.19
	2nd quarter	3.90	3.005	1.000	-4.14	11.93
	4th quarter	3.92	3.005	1.000	-4.11	11.96
4th quarter	1st quarter	6.23	3.005	0.239	-1.80	14.26
	2nd quarter	-0.03	3.005	1.000	-8.06	8.01
	3rd quarter	-3.92	3.005	1.000	-11.96	4.11

Based on observed means.
The error term is Mean Square (Error) = 176.076.

*. The mean difference is significant at the .05 level.

The results from the Bonferroni post hoc test demonstrate that the mean difference is statistically significant ($p<0.05$) only between the 1st quarter and the 3rd quarter of calendar year. The difference between the mean values for the 1st quarter and the 3rd quarter is also visually apparent in Figure 4-9.

4.1.2 Description of pedestrian casualties by injury severity

4.1.2.1 Overall description of pedestrian injury severity

Injury severity was assigned for almost all pedestrian casualties with the exception of only 60 cases for which injury severity was recorded as unknown. The sample comprises 500 pedestrian fatalities (3.6 percent); 3 502 pedestrian casualties (25.3 percent) with injuries rated as serious; 6 525 pedestrian casualties (47.1 percent) with injuries assigned as slight and 3 266 cases (23.6 percent) in which the pedestrian did not sustain any injury resulting from a road crash (see Figure 4-10). Applying the 2013 population size as a proxy of pedestrian exposure to KSI and fatality figures, an annual rate of 35.7 KSI pedestrian casualties per 100 000 population and an annual fatality rate of 4.5 per 100 000 population are found in this study.

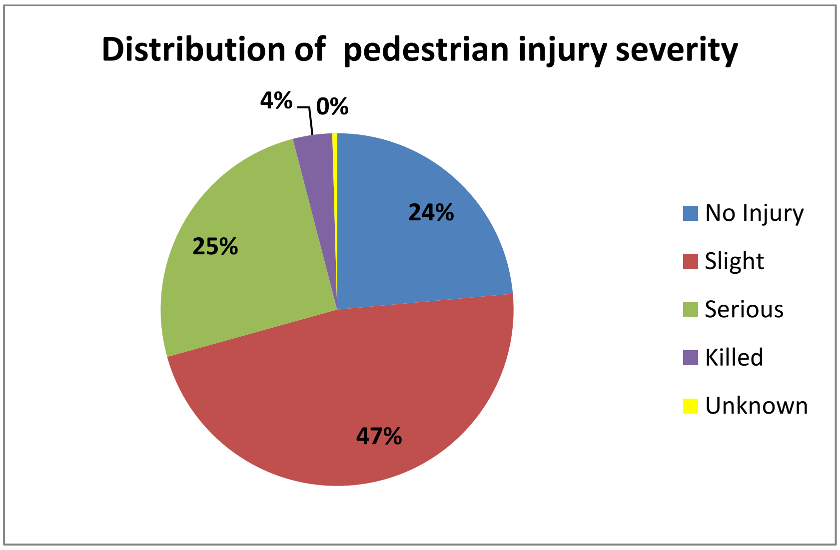


Figure 4-10: Distribution of pedestrian casualties by injury severity

1. **Injury severity by gender**

Table 4-11 and Figure 4-11 present the distribution of injury severity disaggregated by gender. The results show that male pedestrians are more likely to sustain severe injuries than females. While the proportion of male pedestrians in the sample is 62.2 percent (with the exclusion of cases in which gender is not known), male pedestrians represent 74.2 percent, 65.6 percent and 59.1 percent of pedestrian fatalities, serious injuries and slight injuries, respectively. On the contrary, while their proportion in the sample is only 37.8 percent, female pedestrians represent 25.8 percent, 34.4 percent and 40.9 percent of fatalities, serious injuries and slight injuries, respectively. These figures suggest that male pedestrians are more likely to be overrepresented in more severe injuries (i.e. fatal and serious injuries) while female pedestrians tend to be overrepresented in slight injuries.

Table 4-11: Distribution of injury severity by gender

			Injury severity					Total
			No Injury	Slight	Serious	Killed	Unknown	
Gender of pedestrian	Female	Count	467	2163	1063	120	5	3818
		% within Injury severity	37.8%	40.9%	34.4%	25.8%	41.7%	37.8%
	Male	Count	768	3130	2024	345	7	6274
		% within Injury severity	62.2%	59.1%	65.6%	74.2%	58.3%	62.2%
Total		Count	1235	5293	3087	465	12	10092
		% within Injury severity	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

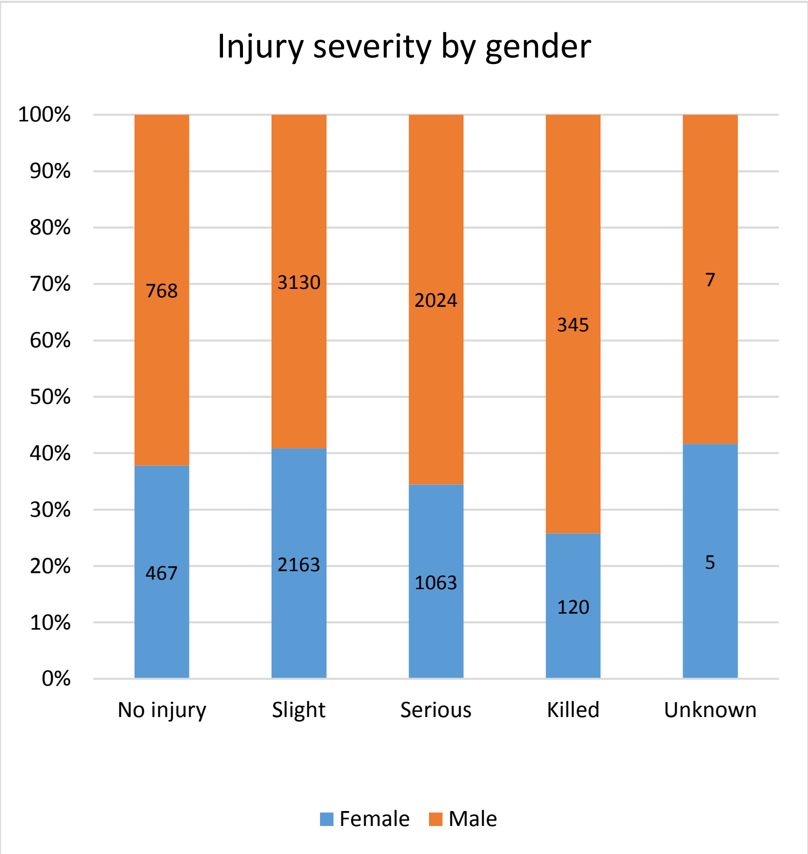


Figure 4-11: Distribution of pedestrian injury severity by gender

2. Injury severity by ethnicity

With respect to the ethnic group of pedestrians, the results show that Black African and Coloured pedestrians are more frequently involved in road crashes as pedestrians. Black African and Coloured casualties represent 36.8 percent and 30.8 percent of all pedestrian casualties, respectively. Pedestrian casualties among White and Asian people account for 3.1 percent and 0.4 percent, respectively. The sample includes a significant proportion (28.7 percent) of pedestrian casualty cases for which the ethnic group is not known (see Table 4-12).

Table 4-12: Pedestrian casualties by ethnicity

Ethnic group			Injury severity					Total
			No Injury	Slight	Serious	Killed	Unknown	
	Asian	Count	16	25	13	1	0	55
		% within Ethnic group	0.5%	0.4%	0.4%	0.2%	0.0%	0.4%
	Black	Count	503	2420	1894	274	6	5097
		% within Ethnic group	15.4%	37.1%	54.1%	54.8%	10.0%	36.8%
	Coloured	Count	445	2504	1134	173	5	4261
		% within Ethnic group	13.6%	38.4%	32.4%	34.6%	8.3%	30.8%
	White	Count	69	246	104	9	2	430
		% within Ethnic group	2.1%	3.8%	3.0%	1.8%	3.3%	3.1%
	Other	Count	6	17	5	1	0	29
		% within Ethnic group	0.2%	0.3%	0.1%	0.2%	0.0%	0.2%
	Unknown	Count	2227	1313	352	42	47	3981
		% within Ethnic group	68.2%	20.1%	10.1%	8.4%	78.3%	28.7%
Total		Count	3266	6525	3502	500	60	13853
		% within Ethnic group	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 4-12 and Table 4-12 present the distribution of injury severity according to ethnic group of pedestrian casualties. An uneven distribution of injury severity across the ethnic groups is apparent particularly for the KSI casualties and for the "no injury category". The highest frequencies of KSI pedestrian casualties as well as fatalities were identified among Black African and Coloured pedestrians.

Of all pedestrian fatalities recorded in the sample, Black African pedestrians represent 54.8 percent and Coloured pedestrians represent 34.6 percent (see Table 4-12). The proportions of KSI casualties for the respective ethnic groups are 54.1percent and 32.7 percent. Furthermore, an interesting finding has been higher proportions of injury severity recorded as "unknown" among Black African and Coloured pedestrians; 10.0 percent and 8.3 percent of casualties were recorded as unknown among these two groups, respectively (see Table 4-12).

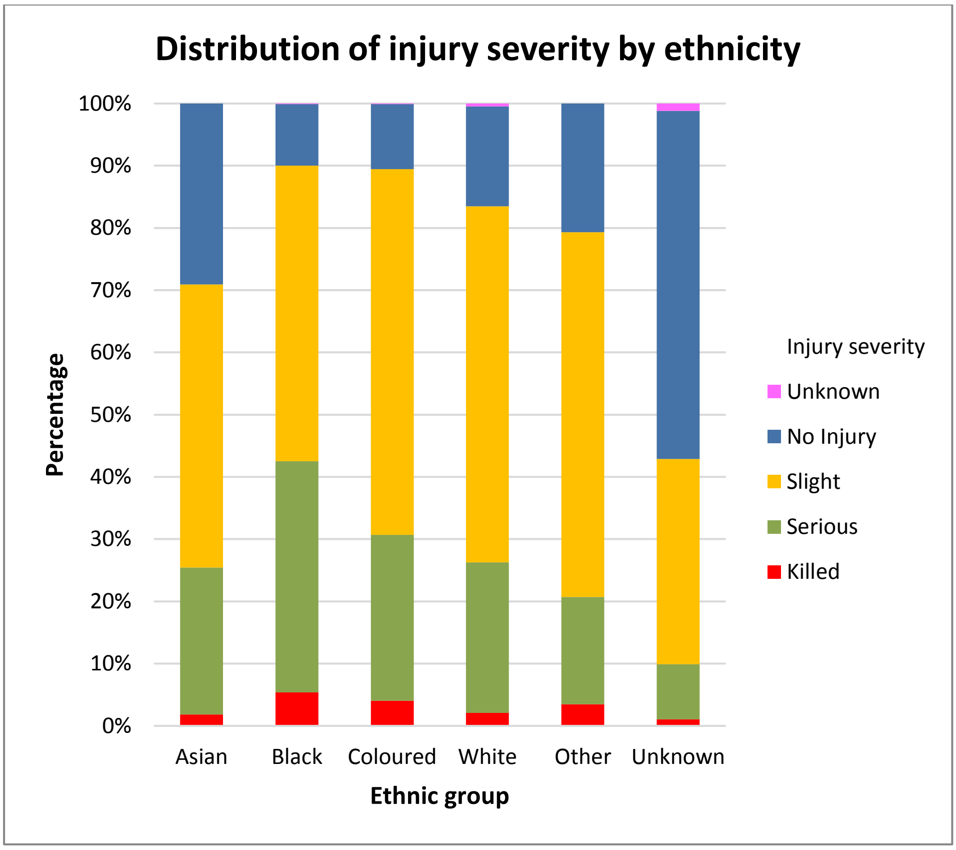


Figure 4-12: Distribution of injury severity by ethnicity

Having no injury reported as a result of a pedestrian crash was found to be related to a higher level of underreporting of other characteristics of the victim such age, gender and the ethnic group. Of pedestrian casualties with a no injury record, 68.2 percent also had no ethnicity recorded (see Table 4-12). Furthermore, 54 percent of all cases in which the gender of pedestrians was unknown had injury severity recorded as “No injury” and the gender of pedestrians was not known for 62.2 percent of all “No Injury” cases (see Table 4-13). Further investigation into the “no injury” records by personal communication with the Data Analyst within the Transport and Urban Development Authority (TDA) of the City of Cape Town revealed that “No injury” cases were classified by data capturers when there was no indication of the injury severity sustained by the pedestrian victim but there was an indication that the vehicle was damaged. This led to conclusion that the “no injury” records correspond to cases in which injury severity was not recorded by the police officer at the crash scene often because the alleged pedestrian victim vanished from the crash scene before the crash was recorded.

Table 4-13: Cross-tabulation of injury severity and pedestrian gender

Injury severity * Pedestrian gender Cross tabulation						
			Gender of pedestrian			
			Female	Male	Unknown	Total
Injury severity	No Injury	Count	467	768	2031	3266
		% within Injury severity	14.3%	23.5%	62.2%	100.0%
		% within Gender	12.2%	12.2%	54.0%	23.6%
	Slight	Count	2163	3130	1232	6525
		% within Injury severity	33.1%	48.0%	18.9%	100.0%
		% within Gender	56.7%	49.9%	32.8%	47.1%
	Serious	Count	1063	2024	415	3502
		% within Injury severity	30.4%	57.8%	11.9%	100.0%
		% within Gender	27.8%	32.3%	11.0%	25.3%
	Killed	Count	120	345	35	500
		% within Injury severity	24.0%	69.0%	7.0%	100.0%
		% within Gender	3.1%	5.5%	0.9%	3.6%
	Unknown	Count	5	7	48	60
		% within Injury severity	8.3%	11.7%	80.0%	100.0%
		% within Gender	0.1%	0.1%	1.3%	0.4%
Total		Count	3818	6274	3761	13853
		% within Injury severity	27.6%	45.3%	27.1%	100.0%
		% within Gender	100.0%	100.0%	100.0%	100.0%

3. Pedestrian injury severity by day of week

The examination of the distribution of injury severity across the days of week reveals that pedestrians are more likely to sustain severe injuries (i.e. KSI casualties) over the weekend (i.e. Saturday and Sunday). As indicated in Table 4-14 and Figure 4-13, the number of pedestrian fatalities peaks over the weekend, with 23 percent and 19.6 percent of all pedestrian fatalities occurring over Saturdays and Sundays, respectively. Significant proportions of pedestrian fatalities are also observed over Fridays and Thursdays representing 14.4 percent and 12.4 percent, respectively. The lowest number of pedestrian fatalities (7.6 percent) is identified on Wednesdays.

Table 4-14: injury severity by day of week

			Injury severity					
			No Injury	Slight	Serious	Killed	Unknown	Total
Day of week	Mon	Count	448	989	428	56	11	1932
		% within Injury severity	13.7%	15.2%	12.2%	11.2%	18.3%	13.9%
	Tue	Count	482	889	414	59	9	1853
		% within Injury severity	14.8%	13.6%	11.8%	11.8%	15.0%	13.4%
	Wed	Count	424	874	408	38	7	1751
		% within Injury severity	13.0%	13.4%	11.7%	7.6%	11.7%	12.6%
	Thu	Count	422	901	430	62	6	1821
		% within Injury severity	12.9%	13.8%	12.3%	12.4%	10.0%	13.1%
	Fri	Count	561	1126	564	72	12	2335
		% within Injury severity	17.2%	17.3%	16.1%	14.4%	20.0%	16.9%
	Sat	Count	549	961	706	115	9	2340
		% within Injury severity	16.8%	14.7%	20.2%	23.0%	15.0%	16.9%
	Sun	Count	380	785	552	98	6	1821
		% within Injury severity	11.6%	12.0%	15.8%	19.6%	10.0%	13.1%
Total		Count	3266	6525	3502	500	60	13853
		% within Injury severity	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

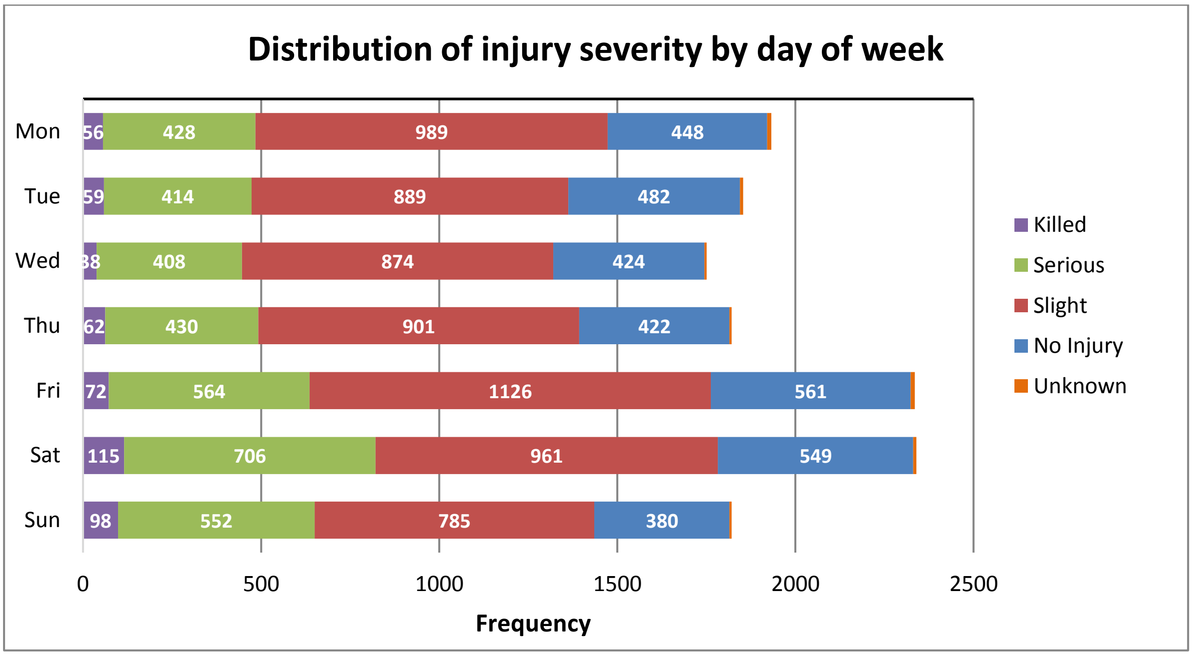


Figure 4-13: Distribution of injury severity by day of week

4. Pedestrian injury severity by month of year

The month of the year in which pedestrian crashes occurred was recorded for all 13 853 pedestrian casualties analysed in this study. The analysis of monthly frequencies of pedestrian casualties shows a peak in the month of August, with two minor peaks emerging in May and November. The lowest incidences of pedestrian casualties is observed over the period extending from December until February (see Table 4-15 and Figure 4-14).

Table 4-15: Pedestrian injury severity by month of year

			Injury severity					
			No Injury	Slight	Serious	Killed	Unknown	Total
Month of crash occurrence	January	Count	205	523	274	25	5	1032
		% within Month	19.9%	50.7%	26.6%	2.4%	0.5%	100.0%
	February	Count	227	509	266	30	11	1043
		% within Month	21.8%	48.8%	25.5%	2.9%	1.1%	100.0%
	March	Count	271	565	263	38	6	1143
		% within Month	23.7%	49.4%	23.0%	3.3%	0.5%	100.0%
	April	Count	254	509	288	54	7	1112
		% within Month	22.8%	45.8%	25.9%	4.9%	0.6%	100.0%
	May	Count	296	578	311	52	2	1239
		% within Month	23.9%	46.7%	25.1%	4.2%	0.2%	100.0%
	June	Count	265	538	296	41	6	1146
		% within Month	23.1%	46.9%	25.8%	3.6%	0.5%	100.0%
	July	Count	301	549	283	41	8	1182
		% within Month	25.5%	46.4%	23.9%	3.5%	0.7%	100.0%
	August	Count	308	594	313	48	3	1266
		% within Month	24.3%	46.9%	24.7%	3.8%	0.2%	100.0%
	September	Count	271	565	311	55	3	1205
		% within Month	22.5%	46.9%	25.8%	4.6%	0.2%	100.0%
	October	Count	316	562	298	34	0	1210
		% within Month	26.1%	46.4%	24.6%	2.8%	0.0%	100.0%
	November	Count	298	577	313	44	2	1234
		% within Month	24.1%	46.8%	25.4%	3.6%	0.2%	100.0%
	December	Count	254	456	286	38	7	1041
		% within Month	24.4%	43.8%	27.5%	3.7%	0.7%	100.0%
Total		Count	3266	6525	3502	500	60	13853
		% within Month	23.6%	47.1%	25.3%	3.6%	0.4%	100.0%

The distribution of injury severity across the months of the year is presented in Figure 4-14. The results indicate that injury severity is fairly evenly distributed across the months of the year. Nonetheless, April appears to be the month with the highest incidence of pedestrian fatalities (4.9 percent), followed by September (4.6 percent) and May (4.2 percent). The lowest incidence of pedestrian fatalities is detected in January (2.4 percent), followed by October (2.8 percent) and lastly February (2.9 percent).

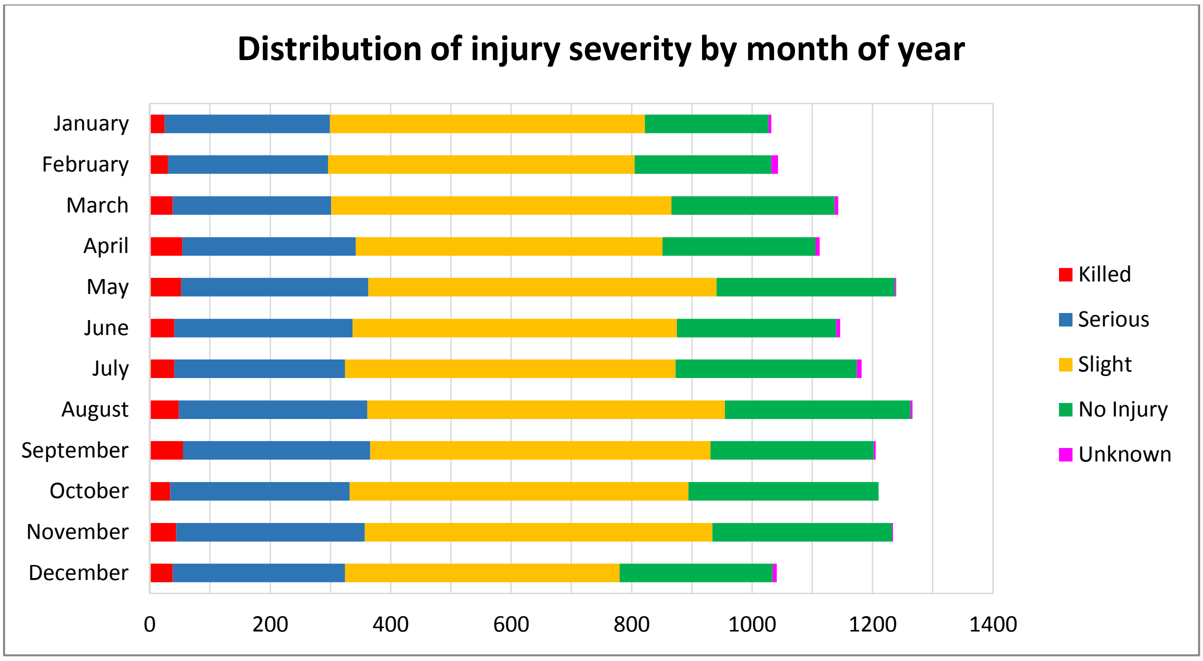


Figure 4-14: Distribution of injury severity by month of year

4.1.2.2 Description of pedestrian fatalities

1. **Pedestrian fatalities by age and gender**

By restricting the analysis to only fatally injured pedestrians, the highest incidence of pedestrian fatalities is observed in the 31-35 age group, followed by the 1-5 age group, the 26-30 age group, the 21-25 age group and the 36-40 age group (see Figure 4-15).

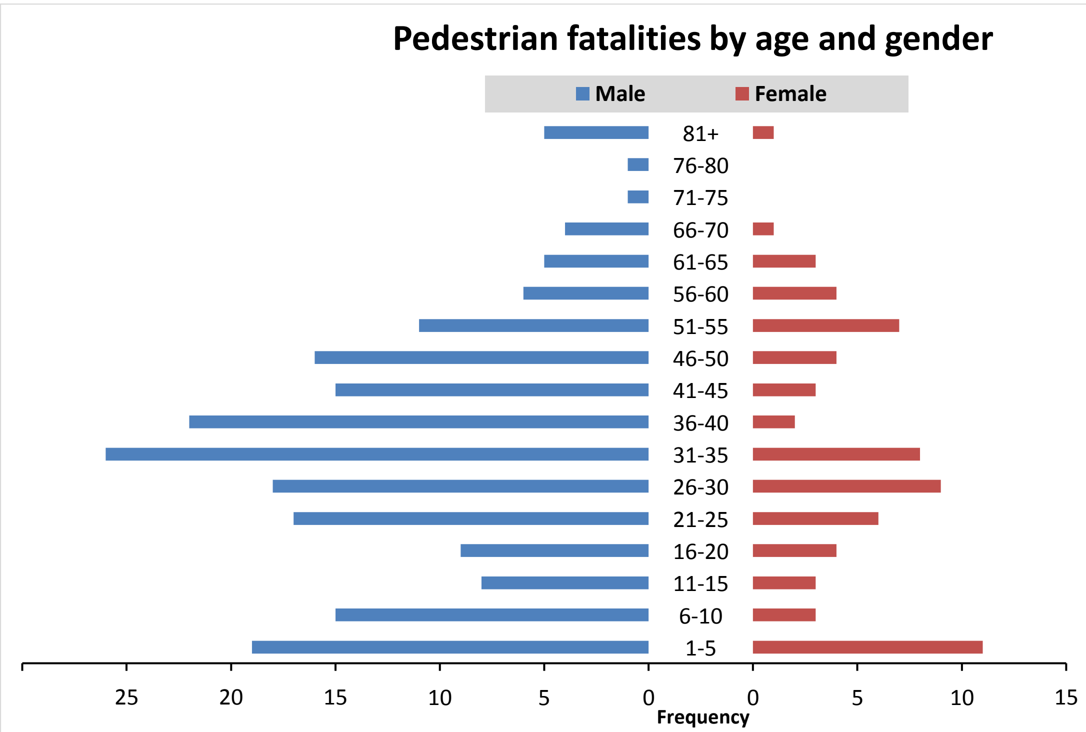


Figure 4-15: Distribution of pedestrian fatalities by age and gender

Further insight into the male-to-female fatality ratios indicates that all the ratios are greater than one, suggesting that male pedestrians are always overrepresented in fatal pedestrian crashes across all the age groups (see Table 4-16). As illustrated in Figure 4-16, the highest male-to-female ratios are identified in the 36-40 age group (M:F=11), the 6-10 age group (M:F=5), the 41-45 age group (M:F=5), the 81+ age group (M:F=5) and the 66-70 age group which has the same ratio as the age group 46-50 (M:F=4).

Table 4-16: Pedestrian fatalities by age and gender

Age group	Female	Male	Unknown	Total	M:F
1-5	11	19	2	32	1.73
6-10	3	15	0	18	5.00
11-15	3	8	0	11	2.67
16-20	4	9	0	13	2.25
21-25	6	17	1	24	2.83
26-30	9	18	2	29	2.00
31-35	8	26	1	35	3.25
36-40	2	22	0	24	11.00
41-45	3	15	0	18	5.00
46-50	4	16	1	21	4.00
51-55	7	11	0	18	1.57
56-60	4	6	0	10	1.50
61-65	3	5	0	8	1.67
66-70	1	4	0	5	4.00
71-75	0	1	0	1	
76-80	0	1	0	1	
81+	1	5	0	6	5.00

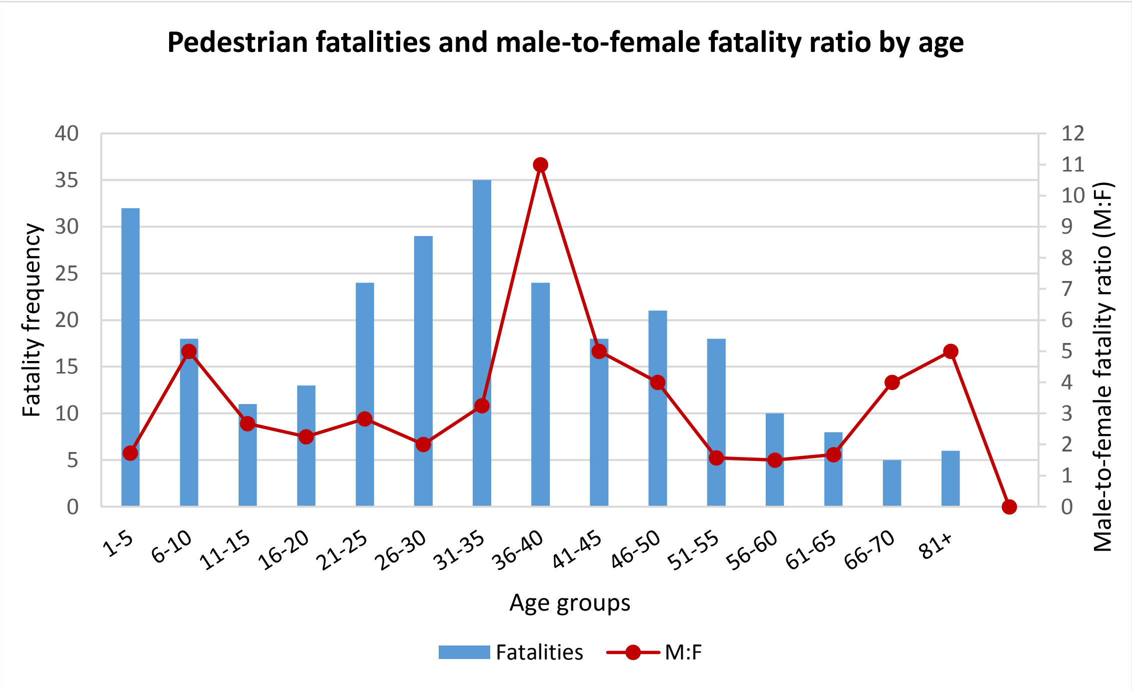


Figure 4-16: Pedestrian fatalities and male-to-female fatality ratio by age

2. **Pedestrian fatalities by time and gender**

The distribution of pedestrian fatalities by gender and the time of crash occurrence is displayed in Figure 4-17. It can be seen from this figure that male pedestrians are more frequently involved in fatal crashes than female pedestrians.

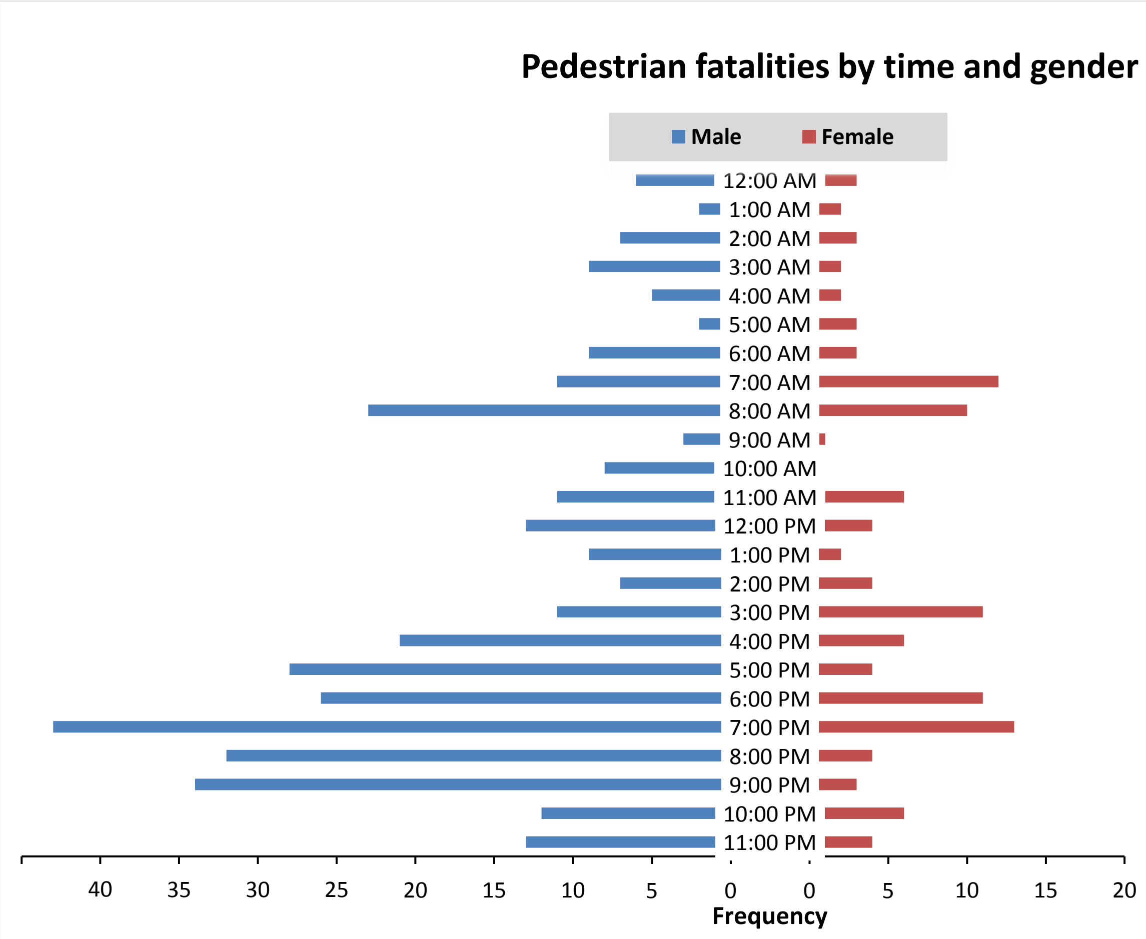


Figure 4-17: Distribution of pedestrian deaths by time and gender

Pedestrian fatality frequencies by time of crash occurrence as well as the corresponding male-to-female ratios of fatality figures are plotted in Figure 4-18. A peak of hourly frequency of pedestrian fatalities is identified between 07:00 PM and 08:00 PM and another minor peak is observed between 8:00 AM and 9:00 AM. A close examination of male-to-female ratios of pedestrian fatalities by time of crash occurrence indicates that the incidence of fatalities for male pedestrians is about 11 times greater than that for female pedestrians between 09:00 PM and 10:00 PM (M:F=11.33). Other times of the day during which males are more likely to sustain fatal injuries than females are between 08:00 PM and 09:00 PM (M:F=8.00); between 10:00 AM and 11:00 AM (M:F=8.00); between 5:00 PM and 06:00 PM (M:F=7.00); between 01:00 PM and 02:00 PM (M:F=4.50); and between 03:00 AM and 04:00 AM (M:F=4.50).

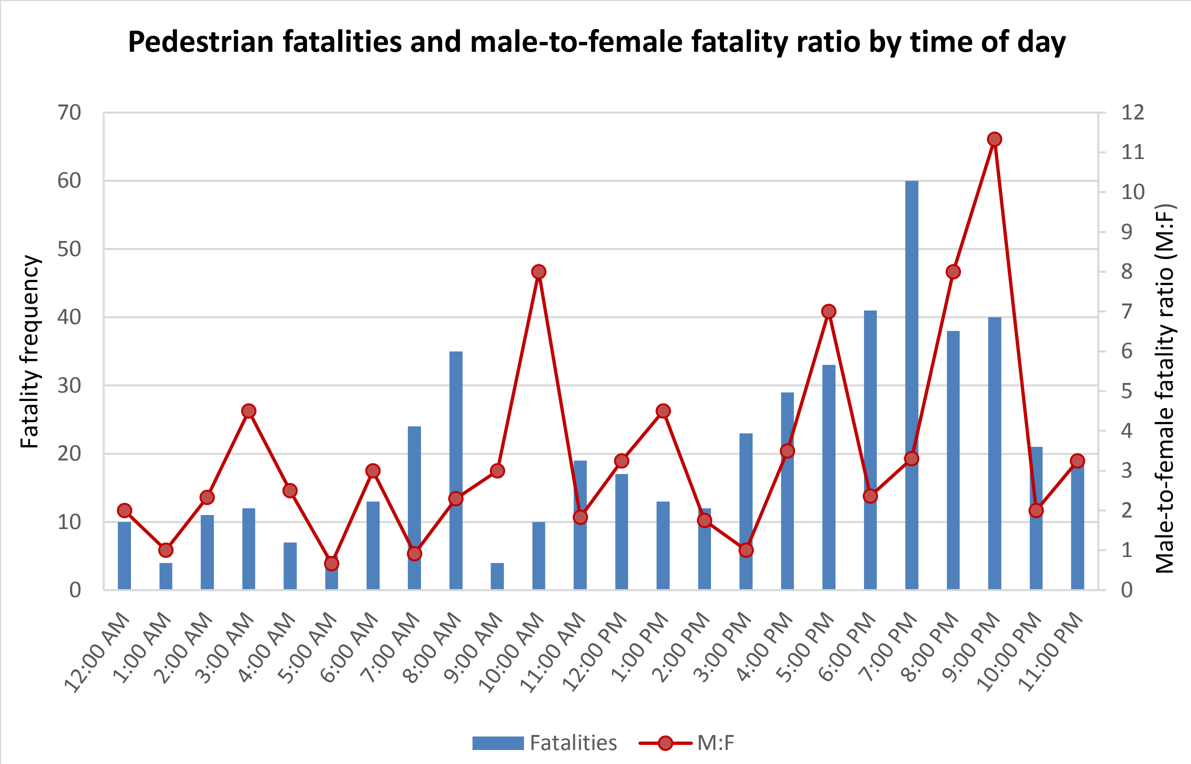


Figure 4-18: Pedestrian fatalities and male-to-female ratios by time of day

3. Pedestrian fatalities by day of week

A dataset of pedestrian fatalities was generated from the total sample of pedestrian casualties Descriptive and inferential statistics were applied to this dataset to assess temporal variations of pedestrian fatalities across the days of week. The descriptive statistics of daily counts of pedestrian fatalities is presented in Table 4-17 and visually illustrated in Figure 4-19. The results of the descriptive statistical analysis demonstrates that pedestrian fatalities are more frequently observed over Saturday, followed by Sunday and Friday.

Table 4-17: Descriptive statistics of daily counts of pedestrian deaths

Dependent Variable: Daily count of pedestrian deaths			
Day of week	Mean	Std. Deviation	N
Monday	0.35	0.533	146
Tuesday	0.38	0.587	151
Wednesday	0.25	0.544	149
Thursday	0.41	0.646	151
Friday	0.47	0.758	149
Saturday	0.74	0.841	153
Sunday	0.62	0.811	153
Holidays	0.32	0.639	44
Total	0.46	0.700	1096

Individual mean differences were evaluated using the ANOVA test. The assumption of homogeneity of variance was tested by the Levene's test. The results from the Levene's test presented in Table 4-18 indicates that the test is not significant at the 5% level, implying that the variance is not equal across the days of week.

Table 4-18: Levene's test for homogeneity of variance

Levene's Test of Equality of Error Variances[a]					
		Levene Statistic	df1	df2	Sig.
Killed	Based on Mean	8.898	7	1088	0.000
	Based on Median	6.533	7	1088	0.000
	Based on Median and with adjusted df	6.533	7	992.578	0.000
	Based on trimmed mean	10.117	7	1088	0.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Killed

The Games-Howell post hoc test was applied to test differences in mean values of daily pedestrian fatality counts across different days of week. The results from the Games-Howell test are presented in Table 4-19. Marked p-values are statistically significant at the 5% level.

Differences in mean values emerged as not significant ($p>0.05$) between:

- Saturday and Sunday
- Saturday and Friday and
- Weekdays (Monday to Friday) and holidays.

Differences in mean values were found to be statistically significant between:

- Saturday and weekdays from Monday to Thursday
- Saturday and holidays
- Sunday and Monday
- Sunday and Wednesday.

Table 4-19: Results from the Games-Howell post hoc test

(I) Day of week		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Mon	Tue	-0.03	0.065	0.999	-0.23	0.16
	Wed	0.10	0.063	0.744	-0.09	0.29
	Thu	-0.06	0.069	0.987	-0.27	0.15
	Fri	-0.12	0.076	0.761	-0.35	0.11
	Sat	-.39*	0.081	0.000	-0.64	-0.14
	Sun	-.27*	0.079	0.016	-0.51	-0.03
	Holid	0.03	0.106	1.000	-0.30	0.36
Tue	Mon	0.03	0.065	0.999	-0.16	0.23
	Wed	0.14	0.065	0.432	-0.06	0.34
	Thu	-0.03	0.071	1.000	-0.24	0.19
	Fri	-0.09	0.078	0.958	-0.33	0.15
	Sat	-.35*	0.083	0.001	-0.61	-0.10
	Sun	-0.24	0.081	0.073	-0.48	0.01
	Holid	0.07	0.108	0.999	-0.27	0.40
Wed	Mon	-0.10	0.063	0.744	-0.29	0.09
	Tue	-0.14	0.065	0.432	-0.34	0.06
	Thu	-0.16	0.069	0.268	-0.37	0.05
	Fri	-0.22	0.076	0.077	-0.46	0.01
	Sat	-.49*	0.081	0.000	-0.74	-0.24
	Sun	-.37*	0.079	0.000	-0.61	-0.13
	Holid	-0.07	0.106	0.998	-0.40	0.26
Thu	Mon	0.06	0.069	0.987	-0.15	0.27
	Tue	0.03	0.071	1.000	-0.19	0.24
	Wed	0.16	0.069	0.268	-0.05	0.37
	Fri	-0.06	0.081	0.996	-0.31	0.19
	Sat	-.33*	0.086	0.004	-0.59	-0.07
	Sun	-0.21	0.084	0.198	-0.47	0.05
	Holid	0.09	0.110	0.990	-0.25	0.44
Fri	Mon	0.12	0.076	0.761	-0.11	0.35
	Tue	0.09	0.078	0.958	-0.15	0.33
	Wed	0.22	0.076	0.077	-0.01	0.46
	Thu	0.06	0.081	0.996	-0.19	0.31
	Sat	-0.27	0.092	0.073	-0.55	0.01
	Sun	-0.15	0.090	0.705	-0.43	0.12
	Holid	0.15	0.115	0.887	-0.20	0.51
Sat	Mon	.39*	0.081	0.000	0.14	0.64
	Tue	.35*	0.083	0.001	0.10	0.61
	Wed	.49*	0.081	0.000	0.24	0.74
	Thu	.33*	0.086	0.004	0.07	0.59
	Fri	0.27	0.092	0.073	-0.01	0.55
	Sun	0.12	0.094	0.918	-0.17	0.41
	Holid	.42*	0.118	0.013	0.05	0.79
Sun	Mon	.27*	0.079	0.016	0.03	0.51
	Tue	0.24	0.081	0.073	-0.01	0.48
	Wed	.37*	0.079	0.000	0.13	0.61
	Thu	0.21	0.084	0.198	-0.05	0.47
	Fri	0.15	0.090	0.705	-0.12	0.43
	Sat	-0.12	0.094	0.918	-0.41	0.17
	Holid	0.30	0.117	0.171	-0.06	0.66
Holid	Mon	-0.03	0.106	1.000	-0.36	0.30
	Tue	-0.07	0.108	0.999	-0.40	0.27
	Wed	0.07	0.106	0.998	-0.26	0.40
	Thu	-0.09	0.110	0.990	-0.44	0.25
	Fri	-0.15	0.115	0.887	-0.51	0.20
	Sat	-.42*	0.118	0.013	-0.79	-0.05
	Sun	-0.30	0.117	0.171	-0.66	0.06

Based on observed means.
The error term is Mean Square (Error) = .468.
*. The mean difference is significant at the .05 level.

Generally, the results show that pedestrian fatalities are more likely to occur over Fridays, Saturdays and Sundays and this finding is in line with that presented previously in Figure 4-13 on Page 149. Trends of temporal variations in daily frequencies of pedestrian fatalities as well as individual mean differences are illustrated Figure 4-19.

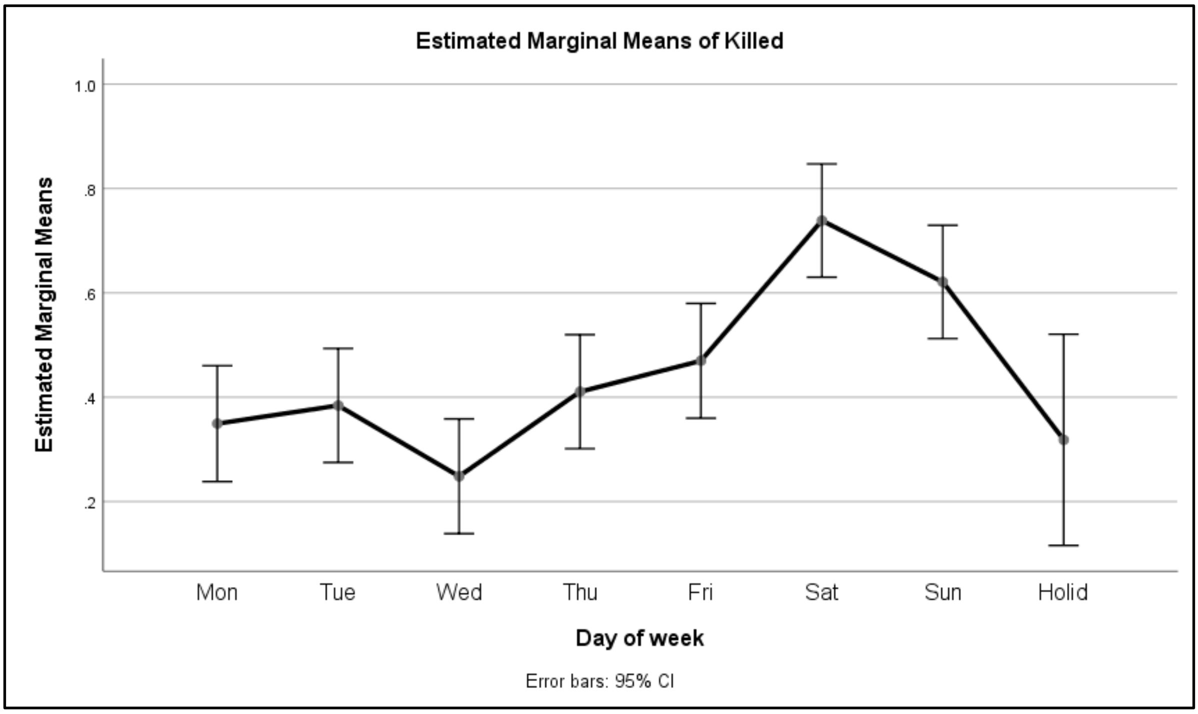


Figure 4-19: Estimated Marginal Means of daily count of pedestrian deaths

4.1.2.3 Description of KSI pedestrian casualties

1. **KSI pedestrian casualties by age and gender**

The KSI category was recorded for 4,002 cases representing 28.9 percent of all pedestrian casualties. The analysis of KSI pedestrian casualties by age and gender was carried out on a sub-dataset of 2 135 pedestrian casualties obtained after excluding cases with zero age records from the dataset of 4 002 KSI pedestrian casualties. The distribution of KSI pedestrian casualties by age and gender is presented Table 4-20 and displayed graphically in Figure 4-20. It can be seen from Figure 4-20 that two apparent peaks of KSI pedestrian casualties emerge in the 1-10 and 26-35 age groups among male pedestrians while KSI cases peak in the 21-30 and 1-10 age groups among female pedestrians. The distribution for both males and females shows a symmetric dip in KSI pedestrian casualties in the 11-20 age group. As expected, the lowest frequencies for both genders emerge among elderly pedestrians aged over 60 years old.

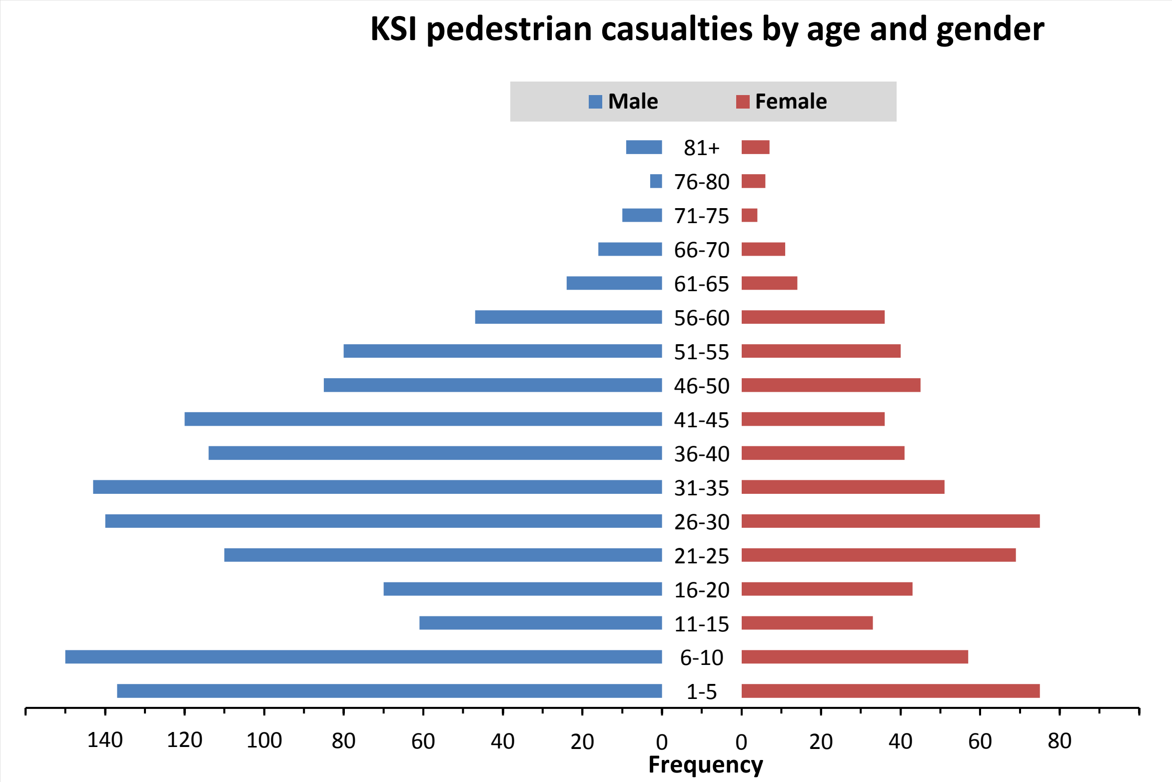


Figure 4-20: Distribution of KSI pedestrian casualties by age and gender

Table 4-20 presents frequencies of KSI pedestrian casualties and the corresponding male-to-female ratios by age group. In all age groups, males are overrepresented in KSI pedestrian casualties, except for the 76-80 age group with a male-to-female ratio of 0.50 (see Table 4-20).

Table 4-20: KSI pedestrian casualties by age and gender

Age group	Female	Male	Unknown	Total	M:F
1-5	75	137	17	229	1.83
6-10	57	150	26	233	2.63
11-15	33	61	11	105	1.85
16-20	43	70	8	121	1.63
21-25	69	110	19	198	1.59
26-30	75	140	19	234	1.87
31-35	51	143	13	207	2.80
36-40	41	114	15	170	2.78
41-45	36	120	7	163	3.33
46-50	45	85	11	141	1.89
51-55	40	80	12	132	2.00
56-60	36	47	9	92	1.31
61-65	14	24	4	42	1.71
66-70	11	16	1	28	1.45
71-75	4	10	0	14	2.50
76-80	6	3	1	10	0.50
81+	7	9	0	16	1.29

The total KSI casualty counts and the male-to-female KSI ratio for each of the 17 age groups are illustrated in Figure 4-21. The figure shows two apparent peaks in the two age groups ranging from 1 to 10 years old and in the three age groups ranging from 21 to 35 years old. The top five age groups in which the male-to-female KSI ratio emerges to be the highest are the 41-45 age group (M:F=3.33); the 31-35 age group (M:F=2.80); the 36-40 age group (M:F=2.78); the 6-10 age group (M:F=2.63); and the 71-75 age group (M:F=2.50). It is worth noting that that the male-to-female ratio values are more pronounced for the KSI casualties than those for the sample of overall pedestrian casualties (see Table 4-1 on Page 132). In general, this suggests that male pedestrians are more likely to sustain more severe injuries (i.e. fatal or serious injuries) when involved in road traffic crashes.

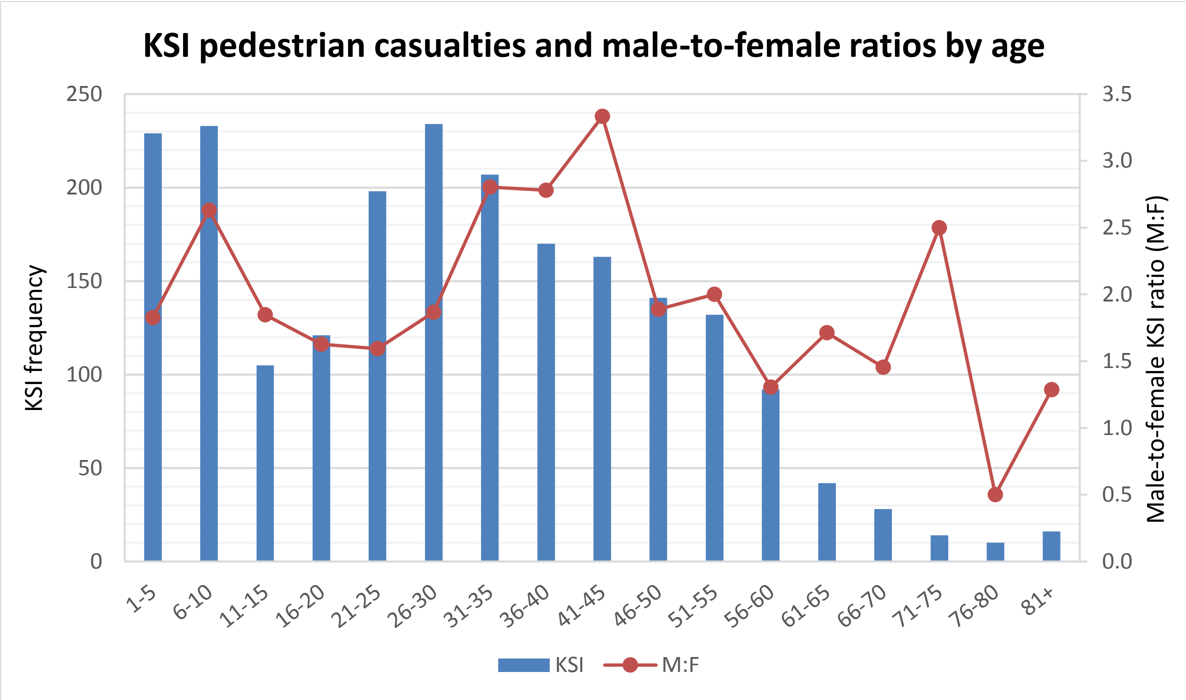


Figure 4-21: KSI pedestrian casualties and male-to-female KSI ratio by age

2. KSI pedestrian casualties by time and gender

Figure 4-22 shows temporal variations of daily counts of KSI casualties according to gender. Again, male pedestrians emerge always to be at a higher risk of being fatally or seriously injured in road traffic crashes than females and the risk appears more pronounced during certain times of the day. For both female and male pedestrians, the highest frequencies of KSI pedestrian casualties are observed between 8:00 AM and 9:00 AM and between 6:00 PM and 7:00 PM. KSI pedestrian casualties are least frequently observed between 4:00 AM and 5:00 AM for both genders.

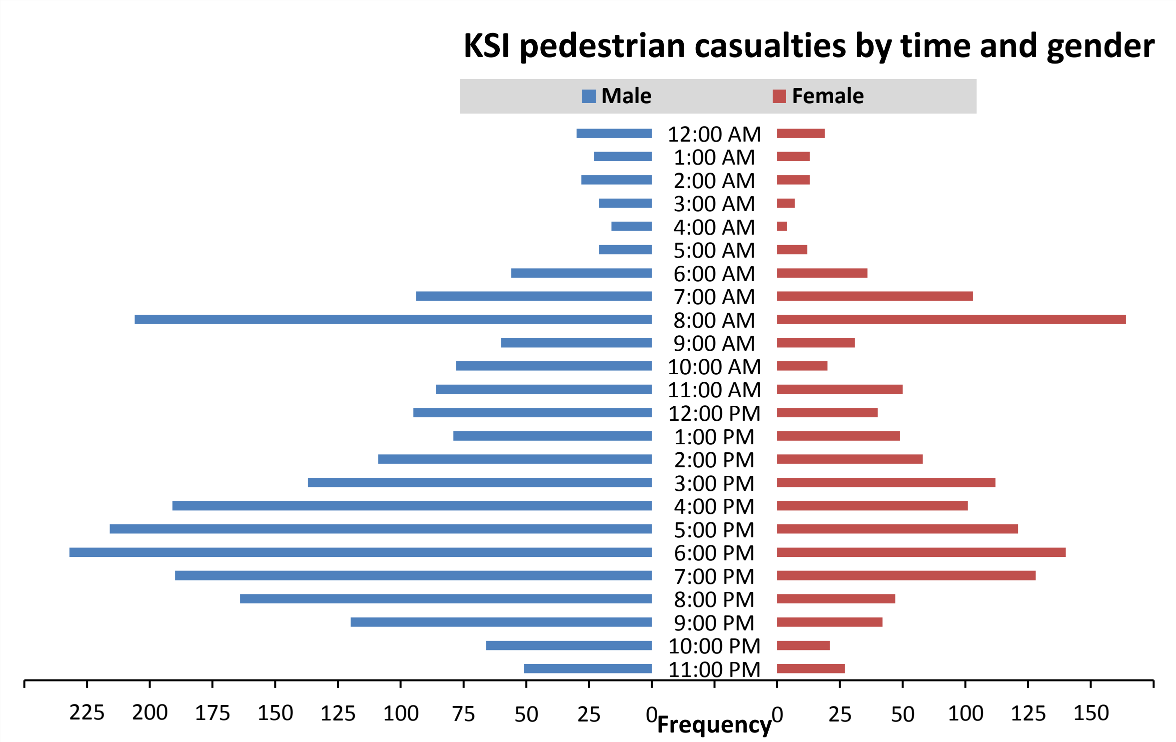


Figure 4-22: Distribution of KSI pedestrian casualties by time and gender

The distribution of KSI pedestrian casualties by time of crash occurrence and the corresponding male-to female KSI ratios are displayed in Figure 4-23. The top five male-to-female KSI ratios presented in descending order are observed between: 4:00 AM and 5:00 AM (M:F=4.00); 10:00 AM and 11:00 AM (M:F=3.90); 08:00 PM and 09:00 PM (M:F=3.49); 03:00 AM and 04:00 AM (M:F=3.00); and 10:00 PM and 11:00 PM (M:F=3.14).

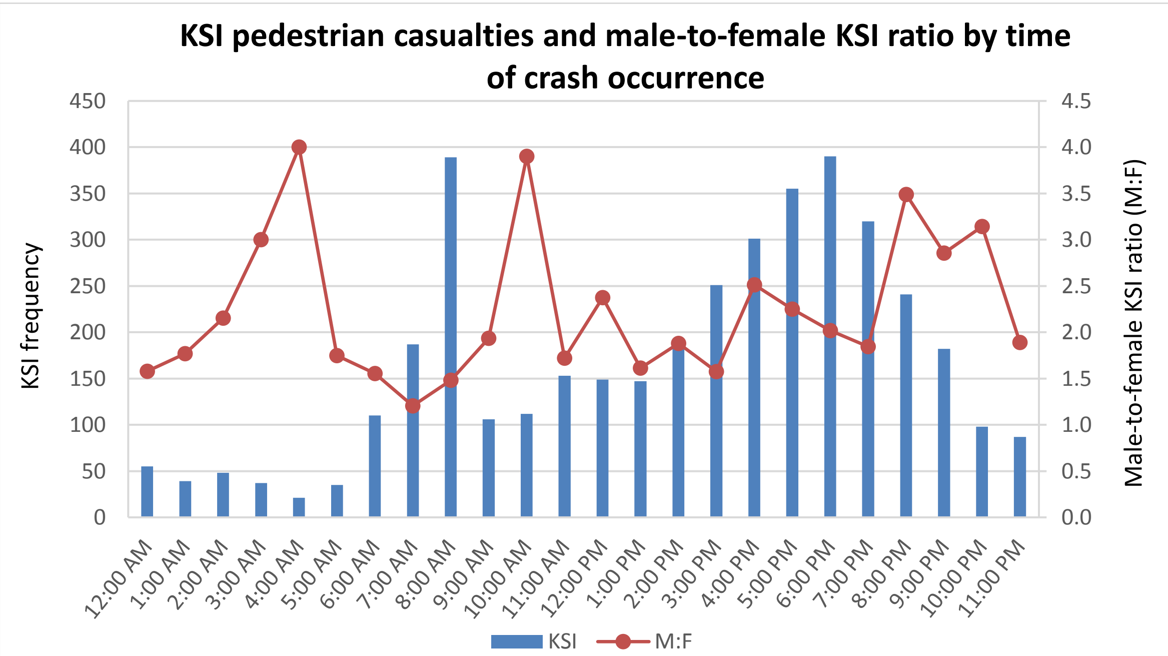


Figure 4-23: KSI pedestrian casualties and male-to-female KSI ratio by time of day

3. KSI pedestrian casualties by day of week

The results from the descriptive statistical analysis of daily counts of KSI pedestrian casualties are presented in Table 4-21 and mean values are plotted in Figure 4-24. A peak of daily KSI casualty counts is observed on Saturday (5.30; S.D=2.47). Higher values of daily casualty counts are also identified on Sunday (4.12; S.D=2.27), Friday (4.09; S.D=2.15) and holidays (3.82; S.D=2.73).

Table 4-21: Descriptive statistics of daily counts of KSI pedestrian casualties

Dependent Variable: Daily counts of KSI pedestrian casualties			
Day of week	Mean	Std. Deviation	N
Monday	3.08	1.876	146
Tuesday	2.98	1.655	151
Wednesday	2.72	1.709	149
Thursday	3.17	2.077	151
Friday	4.09	2.151	149
Saturday	5.30	2.468	153
Sunday	4.12	2.272	153
Holidays	3.82	2.730	44
Total	3.65	2.237	1096

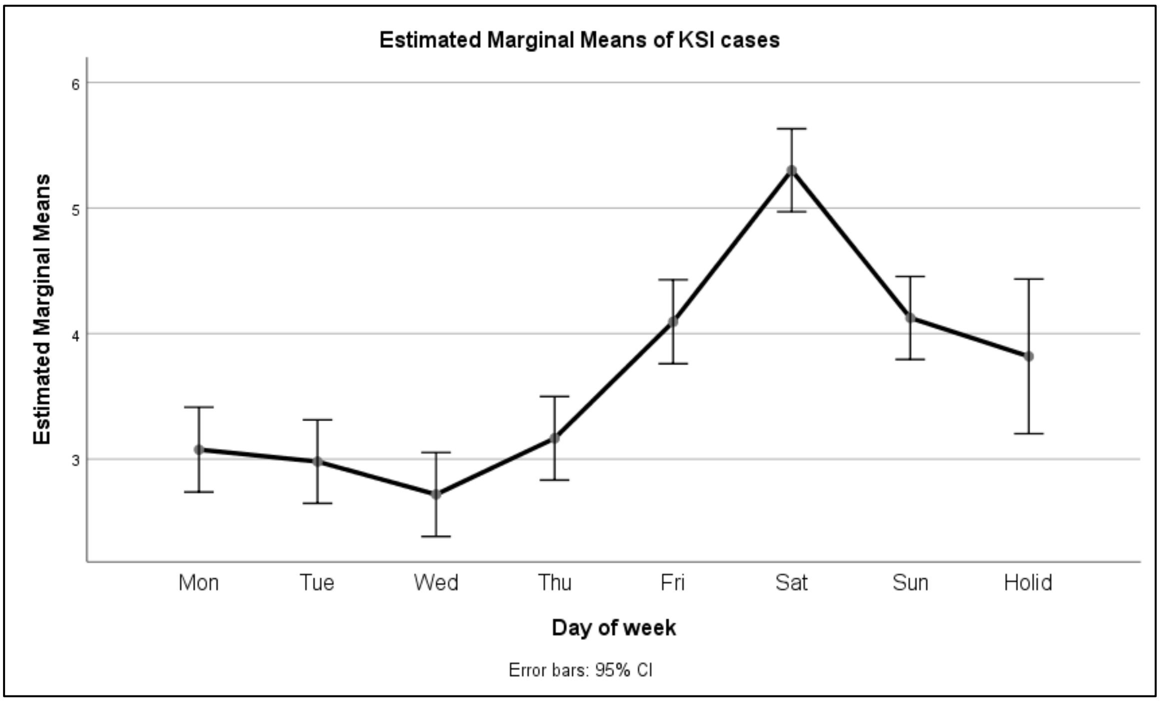


Figure 4-24: Means of daily KSI pedestrian casualties

The results presented in Table 4-22 indicate that the Levene's test is significant at the 5% level and this leads to conclusion that the assumption of homogeneity of variances across the groups is violated. Subsequently, the Games-Howell post hoc test was used to investigate individual mean differences in situations of unequal variances and group sample sizes.

Table 4-22: Levene's test for homogeneity of variance

Levene's Test of Equality of Error Variances[a]					
		Levene Statistic	df1	df2	Sig.
KSI cases	Based on Mean	6.617	7	1088	0.000
	Based on Median	5.368	7	1088	0.000
	Based on Median and with adjusted df	5.368	7	1020.313	0.000
	Based on trimmed mean	6.270	7	1088	0.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Dependent variable: KSI cases

The results from the Games-Howell post hoc test are presented in Table 4-23. The results demonstrate that the individual mean values of daily KSI pedestrian casualties are quite similar (i.e. mean differences are not significant at the 5% level) between the following groups:

- Friday and Sunday
- Sunday and holidays
- Weekdays (Monday to Friday) and holidays.

On the contrary, individual mean differences are statistically significant (i.e. mean values are statistically different) between the following groups:

- Weekdays from Monday to Thursday and Friday
- Weekdays from Monday to Friday and Saturday
- Weekdays from Monday to Thursday and Sunday
- Saturday and other days of week
- Holidays and Saturdays.

Generally, it can be summarised that pedestrians are mostly likely to sustain fatal and serious injuries as a result of road traffic crashes on Saturdays, and the likelihood of the incidence of KSI pedestrian casualties is generally higher over weekends, holidays and Fridays (see Figure 4-24).

Table 4-23: Results from the Games-Howell post hoc test

(I) Day of week		Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Mon	Tue	0.10	0.206	1.000	-0.53	0.72
	Wed	0.36	0.209	0.682	-0.28	1.00
	Thu	-0.09	0.229	1.000	-0.79	0.61
	Fri	-1.02*	0.235	0.001	-1.74	-0.30
	Sat	-2.23*	0.253	0.000	-3.00	-1.45
	Sun	-1.05*	0.240	0.000	-1.78	-0.31
	Holid	-0.74	0.440	0.694	-2.13	0.64
Tue	Mon	-0.10	0.206	1.000	-0.72	0.53
	Wed	0.26	0.194	0.879	-0.33	0.86
	Thu	-0.19	0.216	0.989	-0.85	0.47
	Fri	-1.11*	0.222	0.000	-1.79	-0.44
	Sat	-2.32*	0.241	0.000	-3.06	-1.58
	Sun	-1.14*	0.228	0.000	-1.84	-0.45
	Holid	-0.84	0.433	0.534	-2.20	0.53
Wed	Mon	-0.36	0.209	0.682	-1.00	0.28
	Tue	-0.26	0.194	0.879	-0.86	0.33
	Thu	-0.45	0.219	0.457	-1.12	0.22
	Fri	-1.38*	0.225	0.000	-2.06	-0.69
	Sat	-2.58*	0.244	0.000	-3.33	-1.84
	Sun	-1.41*	0.231	0.000	-2.11	-0.70
	Holid	-1.10	0.435	0.205	-2.47	0.27
Thu	Mon	0.09	0.229	1.000	-0.61	0.79
	Tue	0.19	0.216	0.989	-0.47	0.85
	Wed	0.45	0.219	0.457	-0.22	1.12
	Fri	-.93*	0.244	0.004	-1.67	-0.18
	Sat	-2.14*	0.262	0.000	-2.93	-1.34
	Sun	-.96*	0.250	0.004	-1.72	-0.20
	Holid	-0.65	0.445	0.822	-2.05	0.75
Fri	Mon	1.02*	0.235	0.001	0.30	1.74
	Tue	1.11*	0.222	0.000	0.44	1.79
	Wed	1.38*	0.225	0.000	0.69	2.06
	Thu	.93*	0.244	0.004	0.18	1.67
	Sat	-1.21*	0.266	0.000	-2.02	-0.39
	Sun	-0.03	0.255	1.000	-0.81	0.75
	Holid	0.28	0.448	0.999	-1.13	1.68
Sat	Mon	2.23*	0.253	0.000	1.45	3.00
	Tue	2.32*	0.241	0.000	1.58	3.06
	Wed	2.58*	0.244	0.000	1.84	3.33
	Thu	2.14*	0.262	0.000	1.34	2.93
	Fri	1.21*	0.266	0.000	0.39	2.02
	Sun	1.18*	0.271	0.001	0.35	2.00
	Holid	1.48*	0.457	0.038	0.05	2.92
Sun	Mon	1.05*	0.240	0.000	0.31	1.78
	Tue	1.14*	0.228	0.000	0.45	1.84
	Wed	1.41*	0.231	0.000	0.70	2.11
	Thu	.96*	0.250	0.004	0.20	1.72
	Fri	0.03	0.255	1.000	-0.75	0.81
	Sat	-1.18*	0.271	0.001	-2.00	-0.35
	Holid	0.31	0.451	0.997	-1.11	1.72
Holid	Mon	0.74	0.440	0.694	-0.64	2.13
	Tue	0.84	0.433	0.534	-0.53	2.20
	Wed	1.10	0.435	0.205	-0.27	2.47
	Thu	0.65	0.445	0.822	-0.75	2.05
	Fri	-0.28	0.448	0.999	-1.68	1.13
	Sat	-1.48*	0.457	0.038	-2.92	-0.05
	Sun	-0.31	0.451	0.997	-1.72	1.11

Based on observed means.
The error term is Mean Square (Error) = 4.335.
*. The mean difference is significant at the .05 level.

4.1.3 Description of pedestrian behavioural aspects

4.1.3.1 Distribution of pedestrian behavioural aspects

In South Africa, a police officer who attends a road traffic crash involving one or more pedestrians is required to fill in the accident report form with particulars describing pedestrian behaviour or actions before the incidence of a road crash. These particulars include position, location, manoeuvres, pedestrian action and colour of clothing. The accident report form provides four options for the position record. These are (1) roadway, (2) sidewalk/verge, (3) shoulder of the road and (4) median. Particulars describing pedestrian location are (1) within marked crossing, (2) within 50 meters of crossing and (3) not at crossing. Pedestrian manoeuvres are describes using three options which are (1) facing traffic, (2) back to traffic and (3) crossing road. Pedestrian actions are broken down into eight categories: (1) Walking, (2) running, (3) standing, (4) playing, (5) standing, (6) lying down, (7) working and (8) other. The options provided to record the colour of clothing are (1) light, (2) dark, (3) light & dark, (4) reflective and (5) other. Of these particulars, only the colour of clothing was not provided in the pedestrian casualty dataset obtained from the City of Cape Town. A copy of the accident report form used in South Africa to gather information on road crashes is provided in APPENDIX D.

1. **Analysis of casualties by pedestrian position**

As illustrated in Table 4-24, a large proportion (85.6 percent) of pedestrian crashes occurred when pedestrians were on the roadway. Approximately 10 percent of pedestrian casualties occurred when pedestrians were on sidewalks, verges or hard shoulders.

Table 4-24: Pedestrian casualties by pedestrian position

		Frequency	Percent
Pedestrian position	Median	133	1.0
	Roadway	11856	85.6
	Shoulder of road	573	4.1
	Sidewalk/Verge	1074	7.8
	Unknown	217	1.6
	Total	13853	100.0

2. Analysis of casualties by pedestrian location

The distribution of pedestrian casualties by pedestrian location is illustrated in Figure 4-25. The proportions presented in this figure indicate that the vast majority of pedestrian casualties occurred when pedestrians were located outside designated crossing points on the road network. Of all pedestrian casualties included in the study, 12 192 casualty cases (88.0 percent) are reported to have occurred outside the crossing points. Only 5.7 percent (790 cases) occurred at designated crossing facilities and 4.3 percent (596 cases) are reported to have occurred within 50 metres from a designated crossing point.

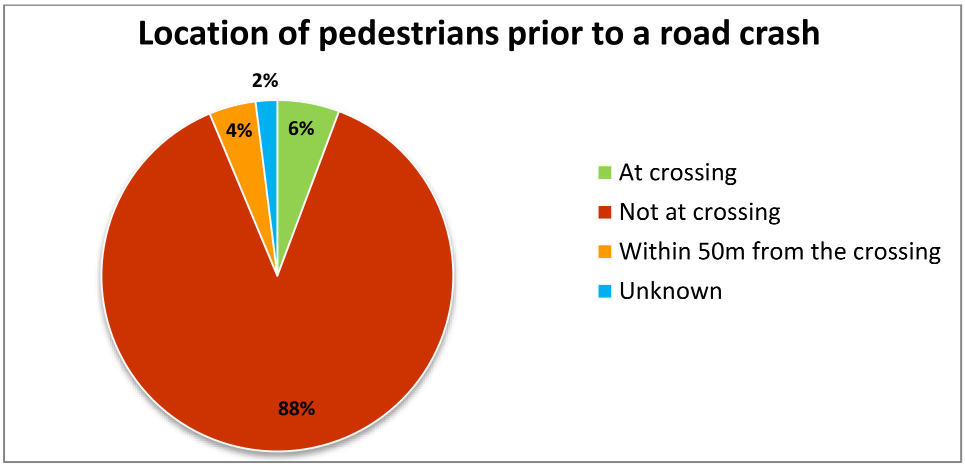


Figure 4-25: Location of pedestrians on a road facility prior to a road crash

3. Analysis of casualties by pedestrian manoeuvres

The distribution of pedestrian casualties by type of manoeuvres is presented in Table 4-25. The results demonstrate that 81.3 percent of pedestrians were involved in road crashes while crossing the road. This finding clearly shows that crossing the road is task associated with a higher crash risk for pedestrians. The results also indicate that in 8.6 percent of pedestrian casualties, pedestrians were walking with their back to the traffic when a crash occurred. Pedestrians who were hit by vehicles while walking facing the oncoming traffic represent 7.8 percent of all pedestrian casualties.

Table 4-25: Pedestrian casualties by type of manoeuvres

		Frequency	Percent
Pedestrian manoeuvres	Back to traffic	1194	8.6
	Crossing road	11258	81.3
	Facing traffic	1082	7.8
	Other	12	0.1
	Unknown	307	2.2
	Total	13853	100.0

4. **Analysis of casualties by pedestrian actions**

The most frequent action in which pedestrians were engaged during the time of a road crash was found to be walking and this was reported in 61.3 percent of pedestrian casualty cases (see Table 4-26). A quarter of all pedestrian casualties occurred when pedestrians were running. Eight percent of pedestrians were standing, three percent were sitting and 1.8 percent of pedestrians were playing when they were hit by vehicles. The sample also includes 68 pedestrian casualties (0.5 percent) who were working on the road (most likely they were road construction or maintenance workers).

Table 4-26: Pedestrian actions prior to a road crash

		Frequency	Percent
Pedestrian action	Lying down	84	0.6
	Playing	244	1.8
	Running	3457	25.0
	Sitting	418	3.0
	Standing	1094	7.9
	Walking	8487	61.3
	Working	68	0.5
	None	1	0.0
	Total	13853	100.0

4.1.3.2 Gender differences in pedestrian behavioural aspects

In this section, gender is analysed in relation to the behavioural aspects previously presented. The intention here is to investigate whether there are gender differences in behaviour performed by pedestrians before the incidence of a road crash.

1. **Pedestrian location by gender**

With respect to pedestrian location, the results show similar behaviour patterns among females and male pedestrians (see Table 4-27 and Figure 4-26). Of all female pedestrian casualties, 86.6 percent (3 292 cases) were involved in road crashes while being outside a designated crossing point. For male pedestrians, crossing outside a designated crossing point is found in 88.5 percent of pedestrian casualty cases. However, the proportion of females who were hit by vehicles at designated crossing locations is shown to be greater than that of (7. 7 percent versus 5.1 percent).

Table 4-27: Pedestrian location by gender

			Location of pedestrian				
			At crossing	Not at crossing	Within 50m from	Unknown	Total
Gender	Female	Count	294	3292	164	68	3818
		% within Gender	7.7%	86.2%	4.3%	1.8%	100.0%
	Male	Count	320	5552	284	118	6274
		% within Gender	5.1%	88.5%	4.5%	1.9%	100.0%
	Unknown	Count	176	3348	148	89	3761
		% within Gender	4.7%	89.0%	3.9%	2.4%	100.0%
Total		Count	790	12192	596	275	13853
		% within Gender	5.7%	88.0%	4.3%	2.0%	100.0%

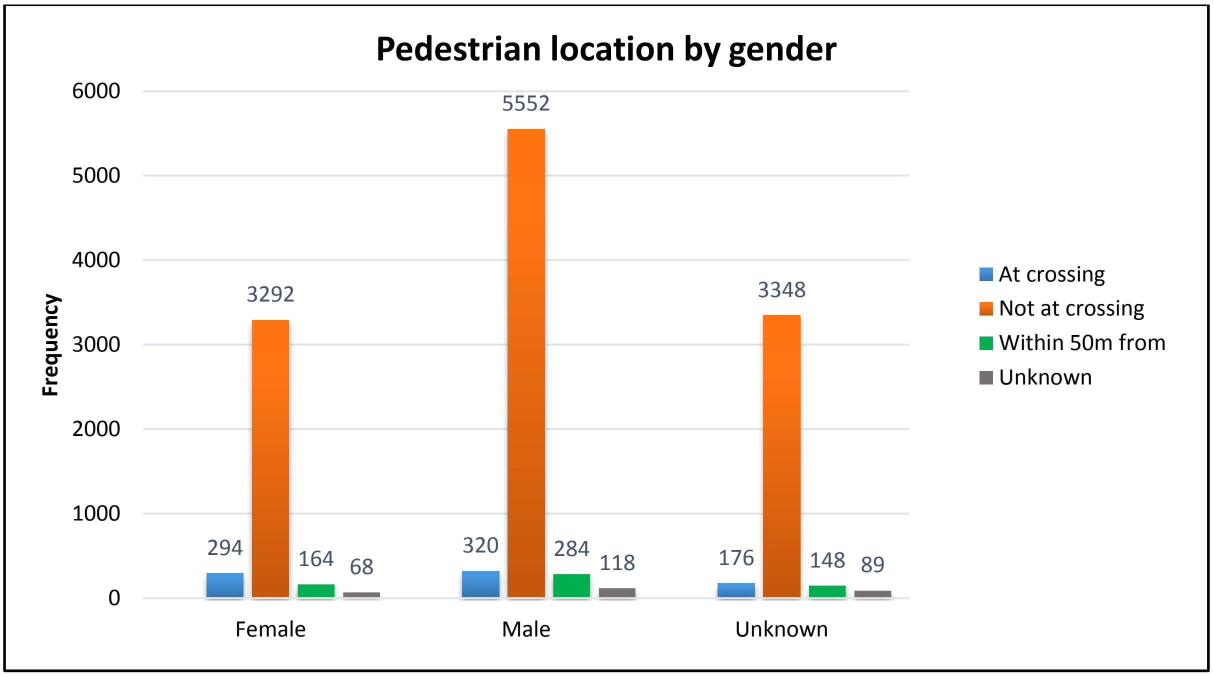


Figure 4-26: Pedestrian location by gender

2. Pedestrian manoeuvres by gender

The distribution of pedestrian manoeuvres by gender is shown in Table 4-28. The analysis reveals similar patterns of pedestrian manoeuvres between females and males except for walking with the back to the traffic. This particular behaviour was found to be slightly more common among males than among female (9.9 percent versus 9.2 percent).

Table 4-28: Pedestrian manoeuvres by gender

			Pedestrian manoeuvres					
			Back to traffic	Crossing road	Facing traffic	Other	Unknown	Total
Gender	Female	Count	377	3046	313	2	80	3818
		% within Gender	9.9%	79.8%	8.2%	0.1%	2.1%	100.0%
	Male	Count	580	5044	510	7	133	6274
		% within Gender	9.2%	80.4%	8.1%	0.1%	2.1%	100.0%
	Unknown	Count	237	3168	259	3	94	3761
		% within Gender	6.3%	84.2%	6.9%	0.1%	2.5%	100.0%
Total		Count	1194	11258	1082	12	307	13853
		% within Gender	8.6%	81.3%	7.8%	0.1%	2.2%	100.0%

3. Pedestrian actions by gender

A breakdown of pedestrian actions by gender is presented in Figure 4-27. The results show slight differences in proportions of pedestrian actions between males and females. Gender differences in terms of performed actions arose in running (21.1 percent for females versus 25.6 percent for males), playing (1.2 percent for females versus 2.3 percent males), sitting (2.6 percent for females versus 3.3 for males), standing (9.3 percent for females versus 7.8 percent for males) and walking (65.0 percent for females versus 59.9 percent for males).

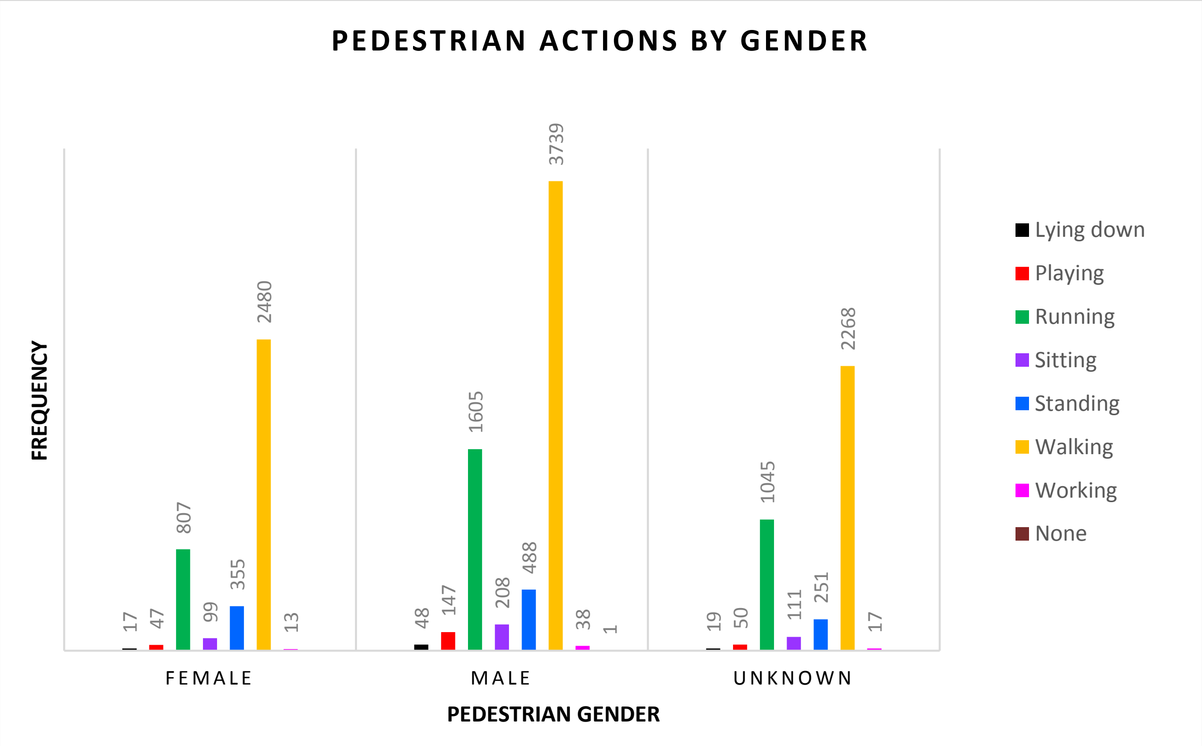


Figure 4-27: Pedestrian actions by gender

4.1.3.3 Age differences in pedestrian behavioural aspects

1. **Pedestrian locations by age group**

Figure 4-28 shows the distribution of pedestrian locations by age group. Pedestrian casualties occurring outside a designated crossing facility are most frequently observed among child pedestrians aged 10 years and younger, followed by the middle aged group between 21 and 35 years old. The peaks of pedestrian casualties taking place both at a crossing facility and within 50 metres from a crossing point are identified among middle aged groups (i.e. from 21 to 50 years old) and among child pedestrians in the 6-10 age group. Overall, pedestrians who were hit by vehicles while at designated crossing points represent less than 10 percent in all age groups, except for the 56-60 age group (10.1 percent) and the 81+ age group (10.0 percent). Furthermore, the highest proportion (20 percent) of pedestrians who were hit by vehicles within 50 metres from a formal crossing point is found among pedestrian casualties aged 81 years and older.

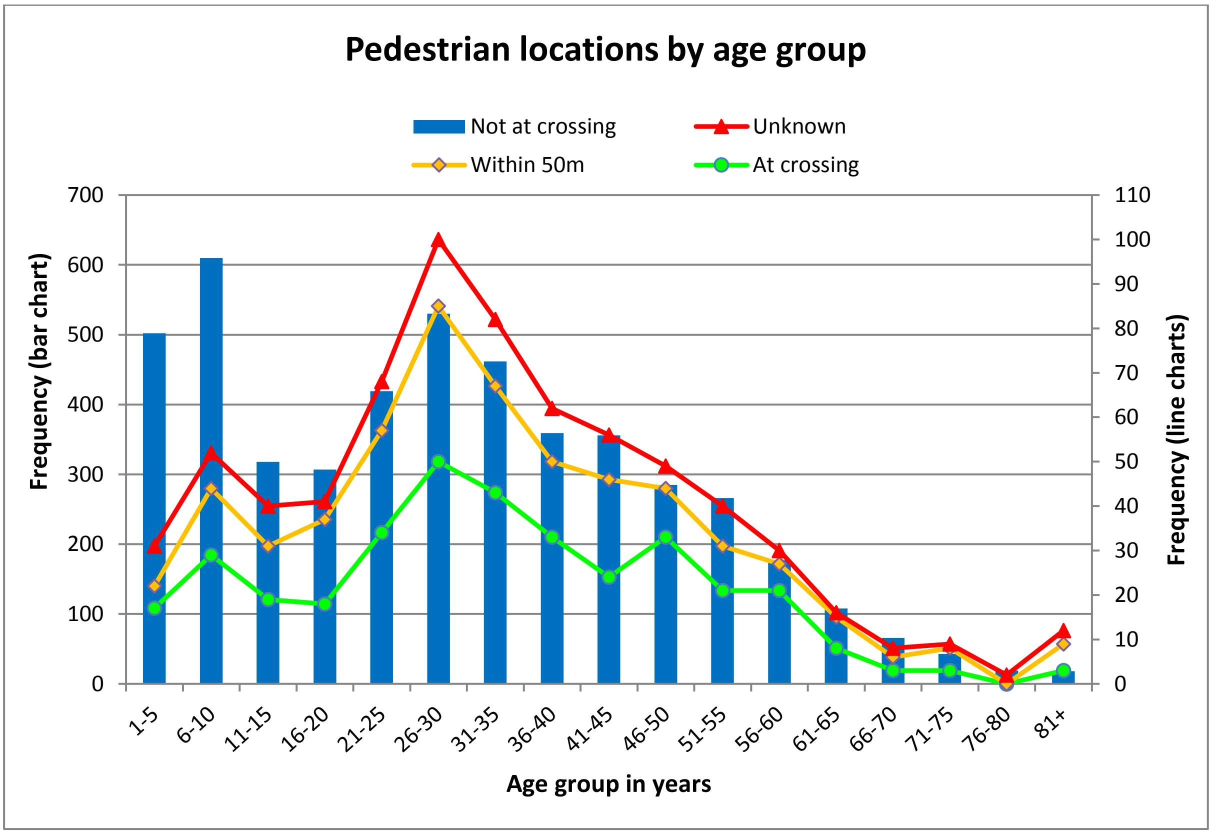


Figure 4-28: Distribution of pedestrian location by age group

2. Pedestrian manoeuvres by age group

Figure 4-29 presents the distribution of pedestrian manoeuvres by age group. In all age groups, the results show that pedestrian crashes in which pedestrians were walking facing the oncoming traffic are more frequent than those in which pedestrians were walking in the same direction as the traffic (i.e. with their back to the traffic). When manoeuvres are analysed in terms of proportions in each age group, walking in the same direction as the traffic during a pedestrian crash is most frequently observed among elderly pedestrians aged over 65 years old, with proportions ranging from 16 to 22 percent. The lowest proportions (less than 8 percent) of pedestrians who were hit by vehicles while walking facing the oncoming traffic emerge among child pedestrians aged between 1 and 10 years and elderly pedestrians aged 81 years and older.

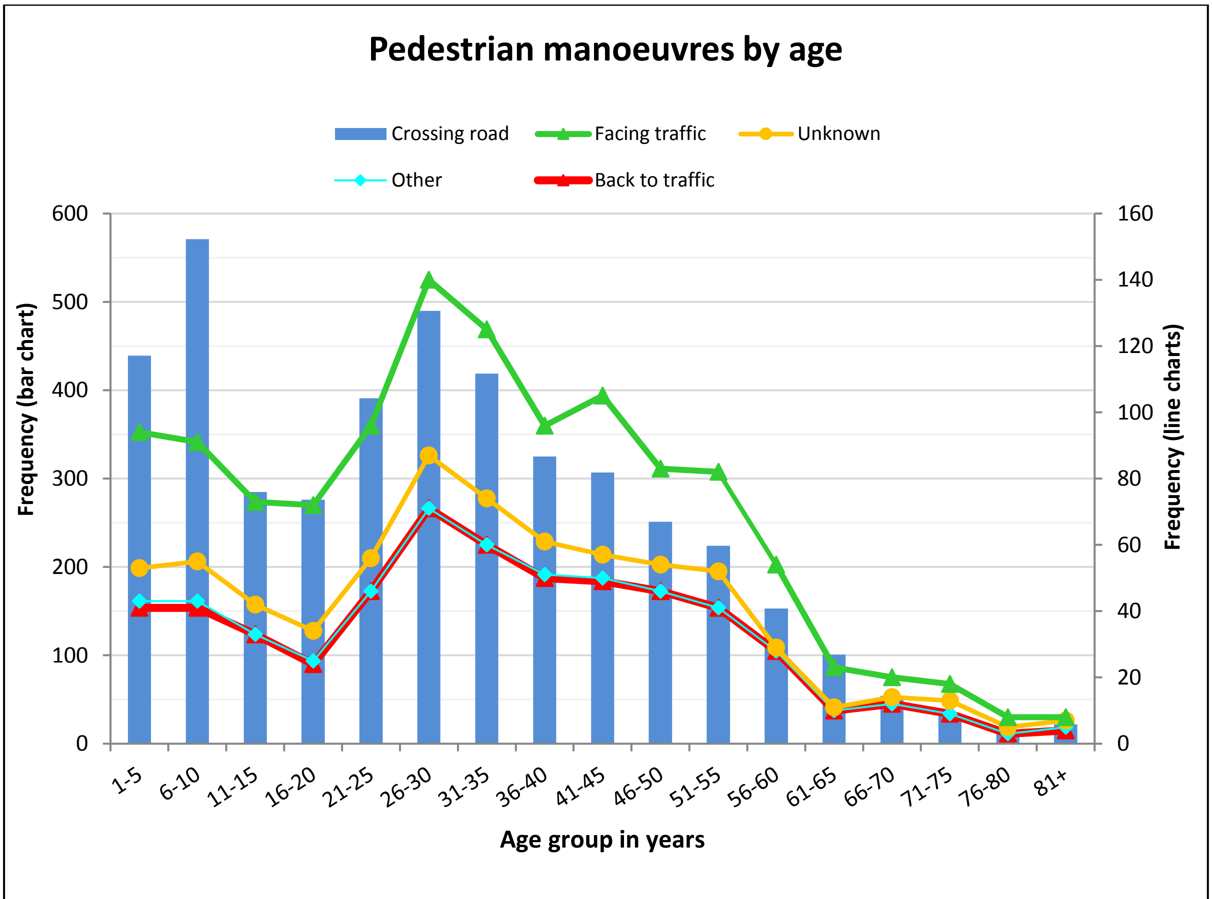


Figure 4-29: Distribution of pedestrian manoeuvres by age group

3. Pedestrian actions by age group

Figure 4-30 illustrates the distribution of pedestrian actions according to age groups. Running, playing, sitting, standing and walking are the actions most predominantly observed among child pedestrians in the 1-10 age group. The frequencies of lying down and working peak among the middle aged groups, in the 26-35 and the 31-50 age ranges, respectively.

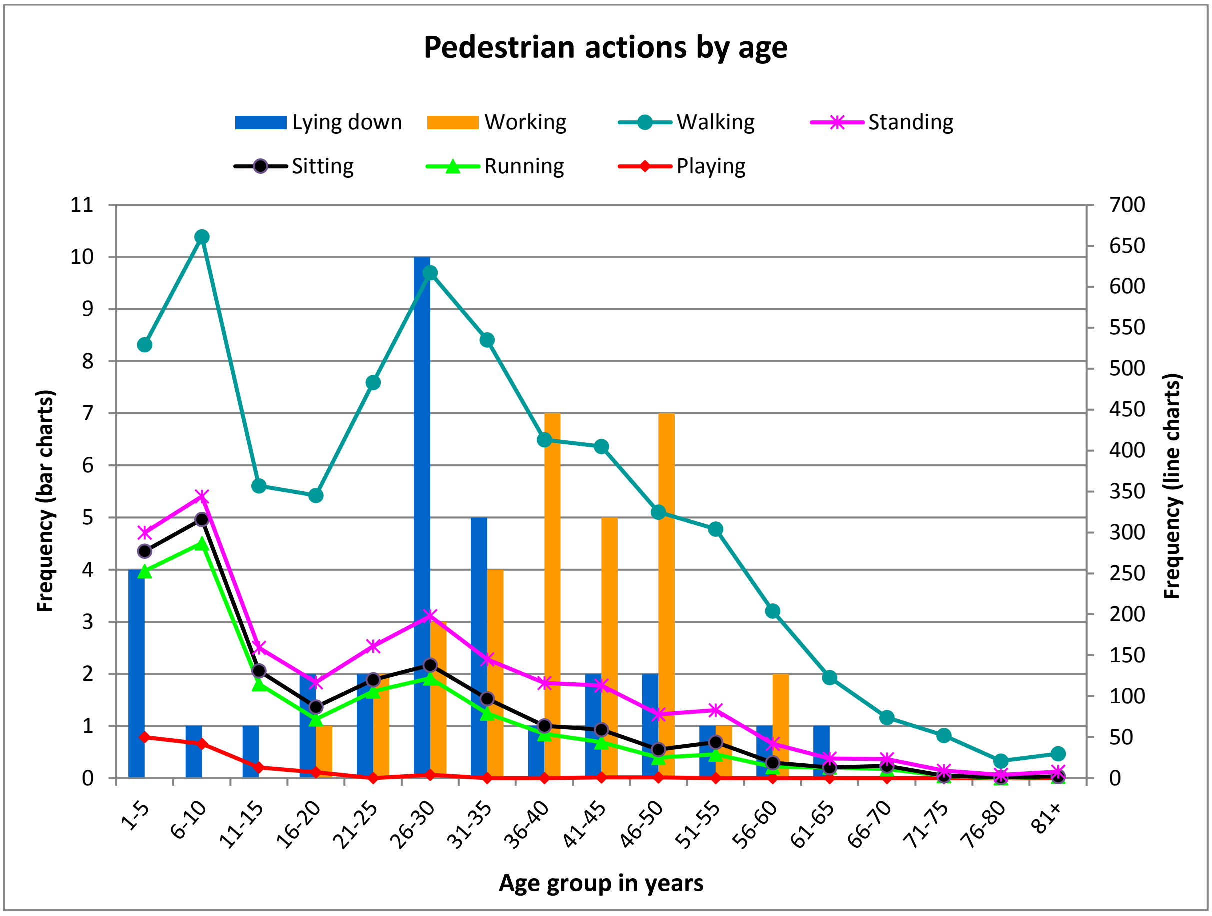


Figure 4-30: Distribution of pedestrian actions by age group

4.1.4 Analysis of intersection-related pedestrian casualties

In addition to data recorded by the police regarding pedestrian behaviour, information on design characteristics of intersections where pedestrian crashes occurred is examined in this study. The analysis of this information allows the identification of high-risk locations for intersection-related pedestrian casualties and enables greater insights into intersection design features that may have influenced the incidence of pedestrian crashes. The analysis involves pedestrian casualties that occurred at road junctions (or nodes) for which the exact location was successfully identified and mapped in ArcMap. The descriptive analysis presented in this section is broken down into three separate sub-sections: (1) profile of pedestrian casualties by type of road facility; (2) an analysis of pedestrian casualties by intersection configuration type; and (3) an analysis of pedestrian casualties by intersection control type.

4.1.4.1 Profile of pedestrian casualties by type of road facility

The types of road facilities involved in the analysis are intersection locations (or nodes) and non-intersection locations (i.e. midblock locations or links). The analysis of additional information collected on crash locations reveals that 10 313 pedestrian casualties occurred at non-intersection locations, representing nearly three quarters (74.4 percent) of all pedestrian casualties in the sample. Intersection-related pedestrian casualties represent 25.6 percent (i.e. 3 540 casualties) of the total number of pedestrian casualties (see Figure 4-31).

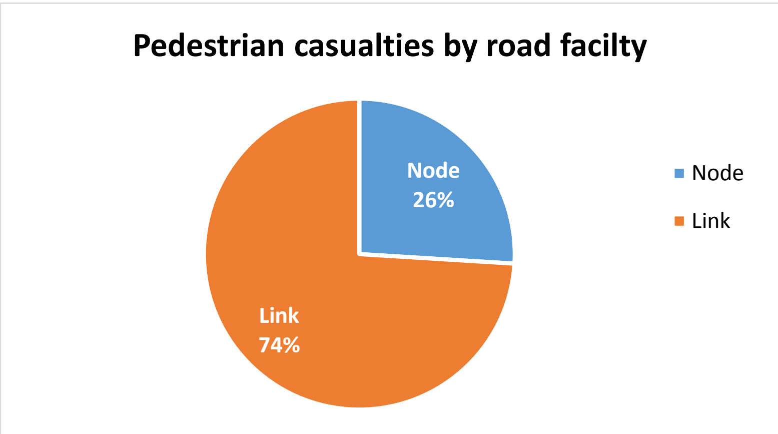
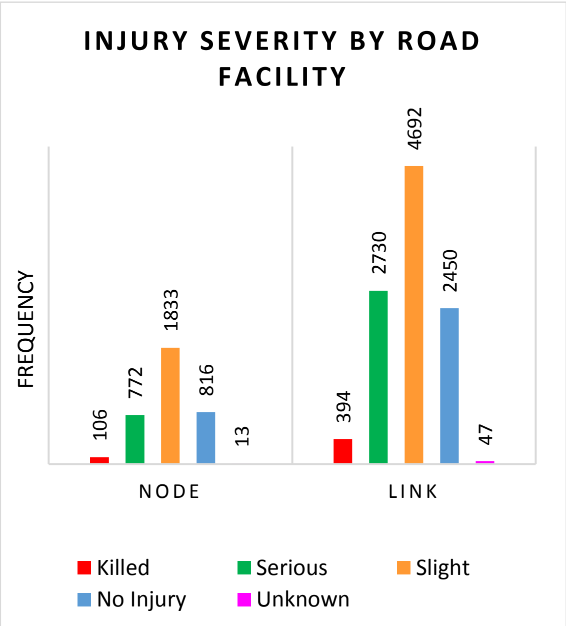


Figure 4-31: Distribution of pedestrian casualties by road facility

The distribution of injury severity according to the type of road facility (i.e. node or link) is presented in Figure 4-32. While fatalities represent 3.6 percent of all pedestrian casualties in the sample, fatalities account for 3.0 percent and 3.8 percent of pedestrian casualties occurring at intersections and links, respectively. While KSI cases account for 28.9 percent of the total number of pedestrian casualties, KSI cases that are recorded at intersections represent 24. 8 percent and those identified at midblock locations represents 30.3 percent. The results from the univariate analysis show that KSI pedestrian casualties occur more frequently at midblock locations than at intersections.



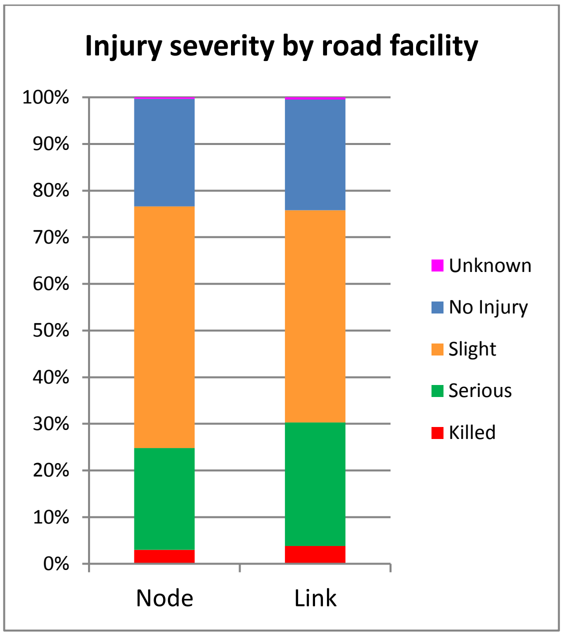


Figure 4-32: Distribution of injury severity by road facility

4.1.4.2 Pedestrian casualties by intersection configuration type

Of pedestrian casualties categorised as intersection-related, 2 021 casualty cases (56.3 percent) are identified at four-legged intersections while 1 398 casualty cases (38.9 percent) are found at three-legged intersections. Roundabouts and mini-circles are locations for 129 pedestrian casualties (3.6 percent) and 43 casualties (1.2 percent) are detected at intersections classified as staggered (see Figure 4-33).

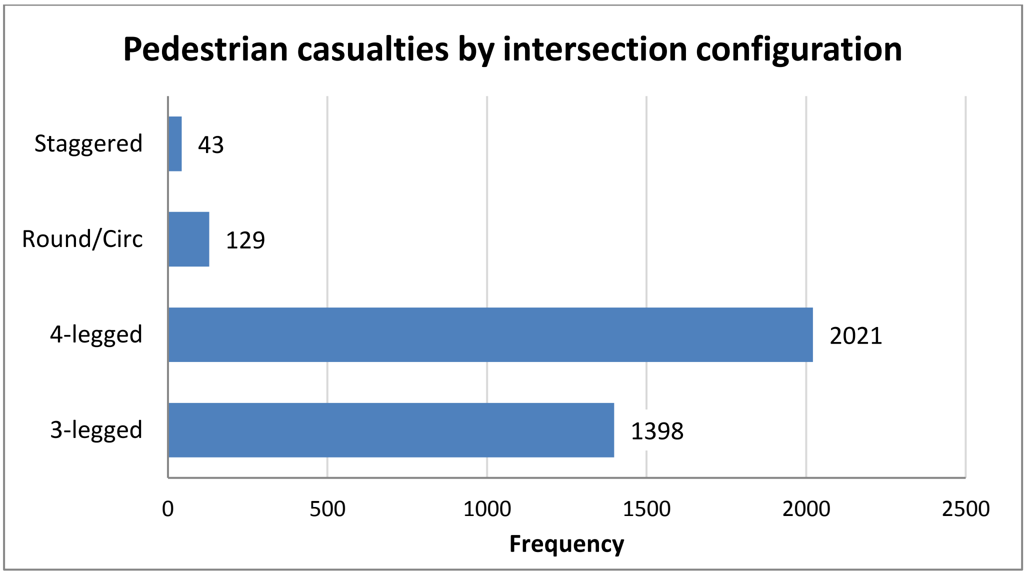


Figure 4-33: Pedestrian casualties by type of intersection configuration

To achieve a better understanding of crash risk associated with each type of intersection configuration, pedestrian casualty frequencies are normalised with respect to the total number of each intersection types in the study area. This procedure enables the estimation of pedestrian casualty rate (expressed as the number of pedestrian casualties per 100 intersection types) by assuming equal pedestrian traffic and pedestrian volumes. Pedestrian casualty rates estimated at each intersection configuration type enable a rough comparison of pedestrian crash risk at different types of intersection. The results of this analysis are presented in Table 4-29.

Table 4-29: Pedestrian casualty rate by intersection configuration type

Intersection configuration type	Pedestrian casualties	Number of intersections	Ped. casualty rate by intersection type [per 100 intersections]
3-legged	1398	45217	3.09
4-legged	2021	10432	19.37
Roundabout/Mini-circle	129	649	19.88
Staggered	43	518	8.30

The highest rate of pedestrian casualties is found at roundabouts and mini-circles (19.88 pedestrian casualties per 100 intersections), followed by four-legged intersections (19.37 pedestrian casualties per 100 intersections). The lowest pedestrian casualty rate is identified at three-legged intersections.

4.1.4.3 Pedestrian casualties by intersection control type

Figure 4-34 illustrates the distribution of pedestrian casualties according to the type of intersection control. With the exclusion of the intersection configuration, intersections with traffic signals are found to have the highest proportion (43.6 percent) of pedestrian casualties, followed by 1-way stops (28.3 percent). Another significant proportion is exhibited by 2-way stops (17.6 percent). The lowest proportion of pedestrian casualties was found at intersections controlled by 2-way yield sign at which only four pedestrian casualties were recorded.

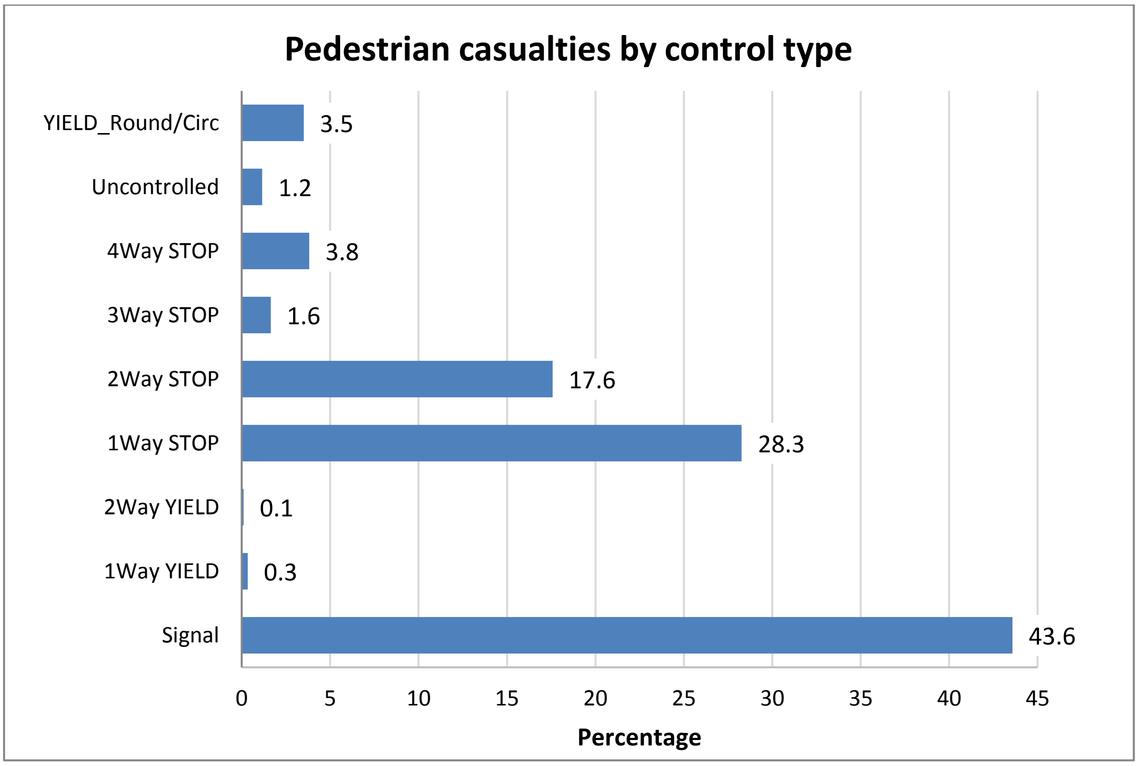


Figure 4-34: Pedestrian casualties by type of intersection control

Pedestrian casualty rates by the type of intersection control are estimated by normalising pedestrian casualty counts against the number of intersection control types in the study area. Again, this is done under the assumption of equal pedestrian and traffic volumes at different types of controls. In this way, rates are estimated in terms of the number of pedestrian casualties recorded per 100 intersections of a particular control type. This analysis excludes pedestrian casualties that occurred at three- and four-legged intersection controlled by the YIELD sign, due to the difficulty in distinguishing the YIELD sign from the STOP sign on aerial google photographs. The highest casualty rate is found at signalised intersections (120.14 pedestrian casualties per 100 intersections), followed by roundabouts and min-circles controlled by the YIELD sign (19.20 pedestrian casualties per 100 roundabouts/mini-circles) and intersections controlled by the 4-Way STOP sign at the third place, with a rate of 11.81 pedestrian casualties per 100 intersections (see Table 4-30). The lowest pedestrian casualty rate is identified at intersections controlled by the 1-Way STOP sign with 2.35 pedestrian casualties per 100 intersections.

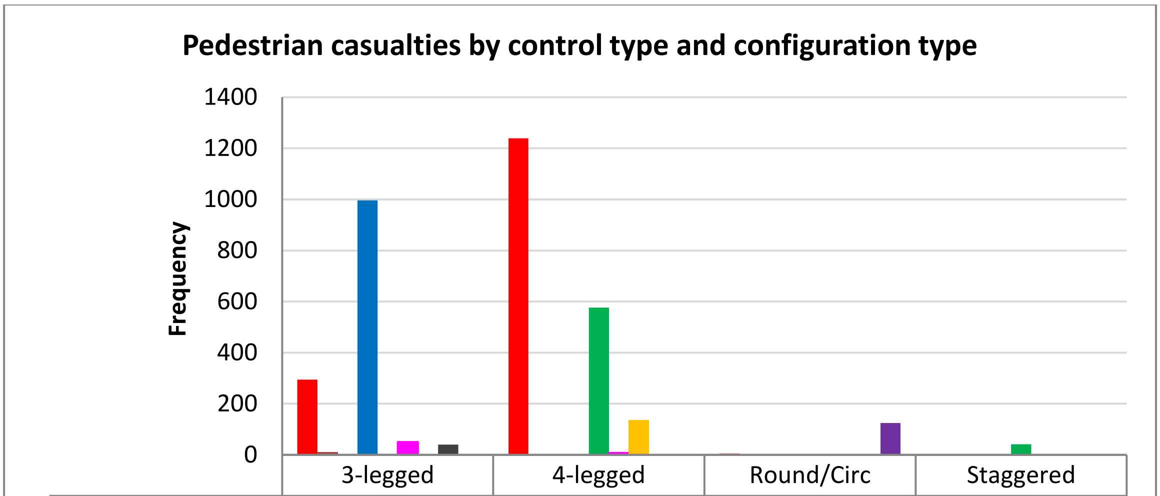
Table 4-30: Pedestrian crash rate by intersection control type

Control type	Pedestrian casualties	Number of intersections	Ped. casualty rate by intersection type [per 100 intersections]
Signal	1539	1281	120.14
1-Way STOP	1011	43087	2.35
2-Way STOP	625	8523	7.33
3-Way STOP	58	973	5.96
4-Way STOP	135	1143	11.81
YIELD_Roundabouts/Mini-Circles	124	646	19.20
Uncontrolled	41	502	8.17

4.1.4.4 Pedestrian casualties by intersection configuration and control type

The results from a bivariate analysis of pedestrian casualties by both control type and intersection configuration are presented in Figure 4-35. Intersection configuration is analysed in four categories: three-legged intersections; four-legged intersections, roundabouts and mini-circles; and staggered intersections.

At three-legged intersections, pedestrian casualties are most frequently identified at intersections controlled by 1-Way STOP (996 casualties), followed by those controlled by traffic signals (295 casualties). At four-legged intersections, the highest frequency of pedestrian casualties is found at intersections controlled by traffic signals (1 239 casualties), followed by those controlled by the 2 Way STOP sign (576 casualties). At roundabouts and mini-circles, five pedestrian casualties are found at signalised roundabouts and 124 pedestrian casualties are detected at roundabouts/mini-circles controlled by the YELD sign (see Figure 4-35).



	3-legged	4-legged	Round/Circ	Staggered
Signal	295	1239	5	0
1Way YIELD	12	0	0	0
2Way YIELD	0	4	0	0
1Way STOP	996	2	0	1
2Way STOP	4	576	0	41
3Way STOP	50	8	0	0
4Way STOP	0	135	0	0
Uncontrolled	40	0	0	1
YIELD_Round/Circ	0	0	124	0

Figure 4-35: Pedestrian casualties by type of intersection control and configuration

4.1.4.5 Injury severity by intersection configuration type

The frequency distribution of injury severity at various intersection configuration types is illustrated in Figure 4-36. Both KSI casualties and pedestrian fatalities are predominantly identified at four-legged and three-legged intersections.

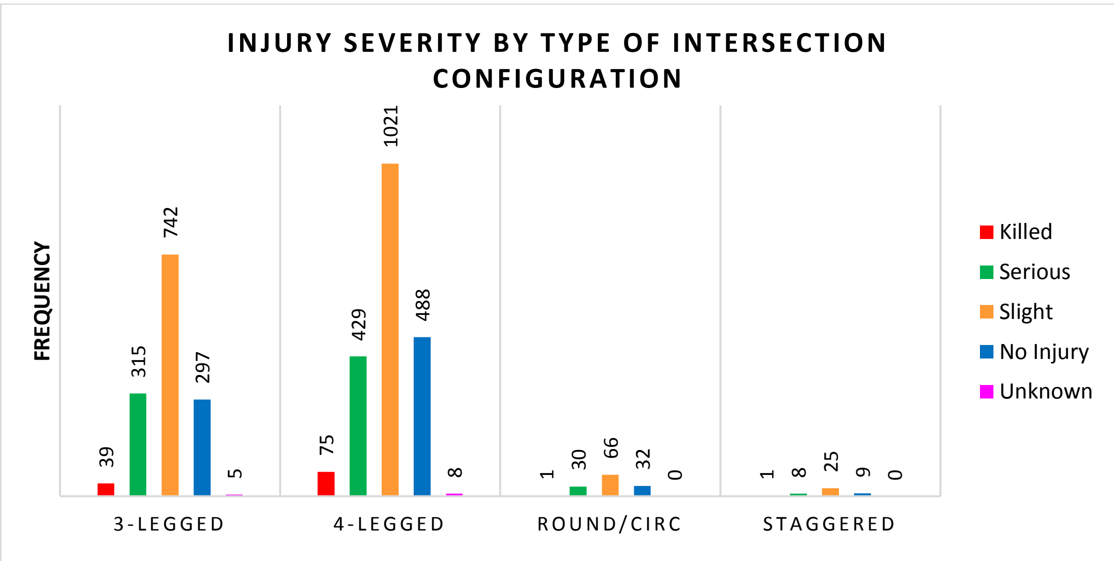


Figure 4-36: Frequency distribution of injury severity by intersection configuration type

Of all pedestrian casualties identified at four-legged intersections, 24.9 percent are KSI casualties and 3.7 percent are fatal injuries. Of all pedestrian casualties observed at three-legged intersections, 25.3 percent are KSI casualties and 2.8 percent are fatal injuries. At roundabouts and mini-circles, 24.0 percent and 0.8 percent of casualties are KSI casualties and fatal injuries, respectively. Of pedestrian casualties recorded at staggered intersections, 29.9 percent are KSI casualties and 2.3 percent are fatal injuries. All pedestrian casualties with unknown injury severity are observed only at three- and four-legged intersections. Another worth noting observation from this analysis has been higher proportions of “no injury” cases at four-legged intersections and roundabouts/mini-circles. Overall, nearly half of pedestrian casualties at the four types of intersection configuration resulted in slight injuries.

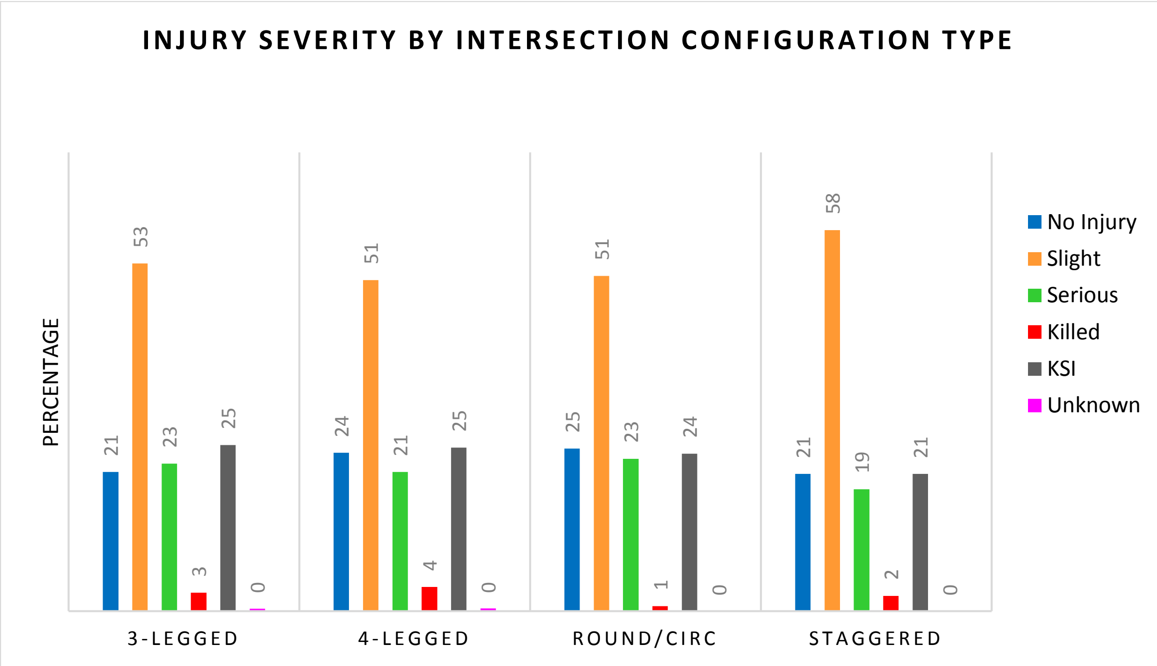


Figure 4-37: Percentage distribution of injury severity by intersection configuration type

In terms of fatality and KSI casualty rates per 100 intersections, three-legged intersections emerge as the safest facilities for pedestrians- the lowest rates of fatalities and KSI pedestrian casualties are identified at these facilities (see Table 4-31). The highest rates of fatalities and KSI pedestrian casualties are experienced at four-legged intersections. Roundabouts and mini-circles are also identified as locations of high KSI casualty rates after the four-legged intersection type.

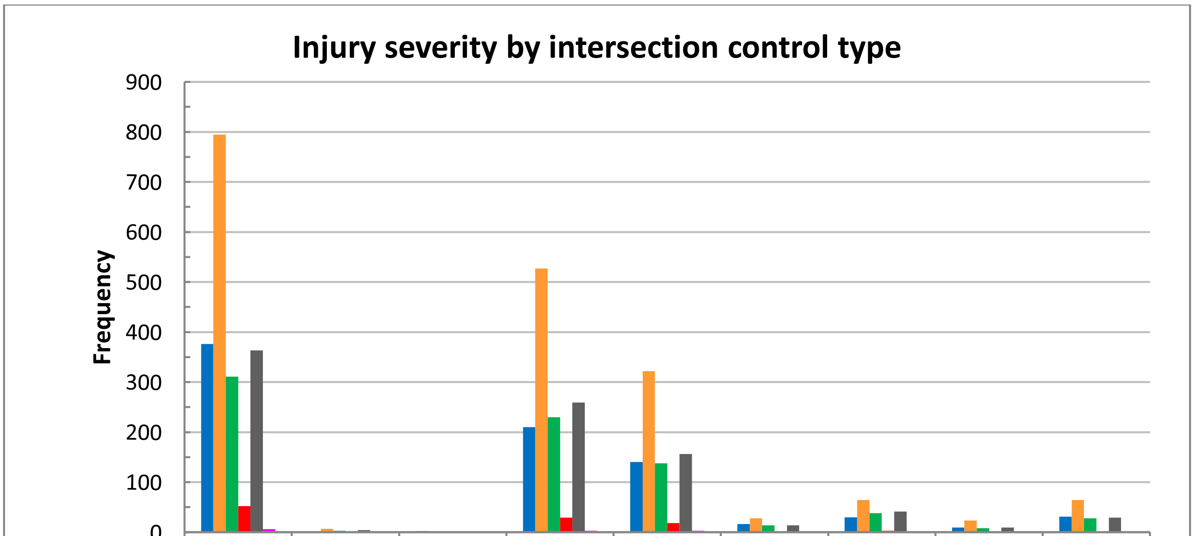
Table 4-31: Rates of fatalities and KSI pedestrian casualties per 100 intersections

Configuration type	Pedestrian fatalities	KSI pedestrian casualties	Intersection number	Ped. fatality rate per 100 intersections	KSI ped. casualty rate per 100 intersections
3-legged	39	354	45217	0.09	0.78
4-legged	75	504	10432	0.72	4.83
Round/Circ	1	31	649	0.15	4.78
Staggered	1	9	518	0.19	1.74

4.1.4.6 Injury severity by intersection control type

Figure 4-38 illustrates the frequency distribution of injury severity according to intersection control type. Large numbers of pedestrian casualties are predominantly observed at signalised intersections and intersections controlled by the 1-Way STOP and 2-Way STOP signs. Large numbers of pedestrian fatalities and KSI pedestrian casualties are also found at these three types of intersection controls as displayed in

Figure 4-38.



	Signal	1Way YIELD	2Way YIELD	1Way STOP	2Way STOP	3Way STOP	4Way STOP	Uncont rolled	YIELD_ Round/ Circ
No Injury	376	1	2	210	140	16	30	9	31
Slight	795	7	1	527	322	28	64	23	64
Serious	311	3	1	230	138	14	38	8	28
Killed	52	1	0	29	18	0	3	1	1
KSI	363	4	1	259	156	14	41	9	29
Unknown	6	0	0	3	3	0	0	0	0

Figure 4-38: Frequency distribution of injury severity by intersection control type

The percentage distribution of injury severity at each type of intersection control is displayed in Figure 4-39. Unexpectedly, the highest proportions of both fatalities and KSI pedestrian casualties are observed at intersections controlled by the YIELD signs. Of all pedestrian casualties observed at intersections controlled by the 1-Way YIELD sign, a quarter of them are serious injuries, 8.3 percent are fatal injuries and KSI pedestrian casualties represent 33.3 percent. Moreover, the highest proportion of “no injury” cases is detected at the intersections with 2-Way YIELD sign. However, these findings are based on a small number of crash events (only16 pedestrian casualties) that occurred at intersections controlled by the YIELD sign. An analysis of a sample containing a large number of pedestrian crashes at intersections controlled by the YIELD sign may lead to more meaningful findings. The second highest proportions of pedestrian fatalities are found at intersection controlled by traffic signals and STOP signs.

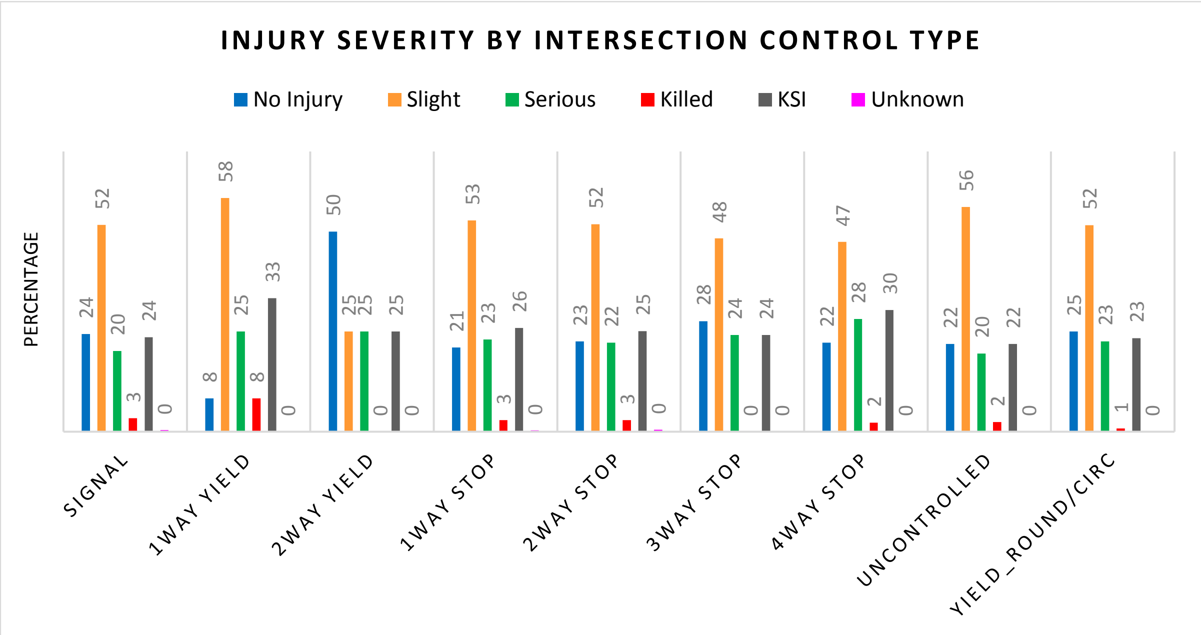


Figure 4-39: Percentage distribution of injury severity by intersection control type

Numbers of pedestrian fatalities and KSI casualties normalised against the number of intersection types are presented in Table 4-32. With the exclusion of pedestrian casualties recorded to have occurred at intersections controlled by the YIELD sign, the highest rates of pedestrian fatalities and KSI casualties emerge at intersections controlled by traffic signals. Under the assumption of equal traffic and pedestrian volumes, the fatality rate at signalised intersections is six times as high as that at roundabouts and mini-circles. In similar way, the fatality rate at signalised intersection is approximately 50 times and 10 times higher than that at intersections controlled by 1-Way STOP and 4-Way STOP signs, respectively.

Table 4-32: Rates of fatalities and KSI casualties by intersection control type

Control type	Ped. fatalities	KSI ped. casualties	Intersection number	Ped. fatality rate per 100 intersections	KSI ped. casualty rate per 100 intersections
Signals	1539	363	1281	120.14	28.34
1Way STOP	1011	259	43087	2.35	0.60
2Way STOP	625	156	8523	7.33	1.83
3Way STOP	58	14	973	5.96	1.44
4Way STOP	135	41	1143	11.81	3.59
YIELD_Round/Circ	124	9	646	19.20	1.39
Uncontrolled	41	29	502	8.17	5.78

4.2 Geospatial analyses of pedestrian causalities

The results presented in this section were obtained from geospatial analyses applied to four datasets of pedestrian casualties. These datasets include:

- The entire sample of pedestrian casualties
- The dataset of intersection-related pedestrian casualties
- The dataset of pedestrian fatalities and KSI casualties that occurred at intersections.
- The dataset of intersection-related pedestrian casualties who sustained slight injuries.

4.2.1 Geospatial analysis for the entire dataset of pedestrian crashes

4.2.1.1 Spatial distribution of pedestrian casualties across the City of Cape Town

Figure 4-40 presents the spatial distribution of pedestrian casualty counts aggregated at the census suburb level. In this figure, it can be observed that the frequency of pedestrian casualties is highest in the Khayelitsha/Mitchell's Plain regions, Stand, Delft, Bellville, Elsies Rivier, and the CBD of Cape Town. The spatial distribution of pedestrian casualties normalised against the population number is illustrated in Figure 4-41. These are exposure-based rates expressed in terms of annual average pedestrian casualties per 100,000 population for the 2012-2014 period. Based on this measure, the highest pedestrian casualty rates were found in Epping Industria, Paarden Eiland, Gardens Beam, Marconi Beam, Cape Town CBD and Table Mountain Nature Reserve suburbs. However, it is important to note that these census suburbs are less populated which led to inflated pedestrian casualty rates.

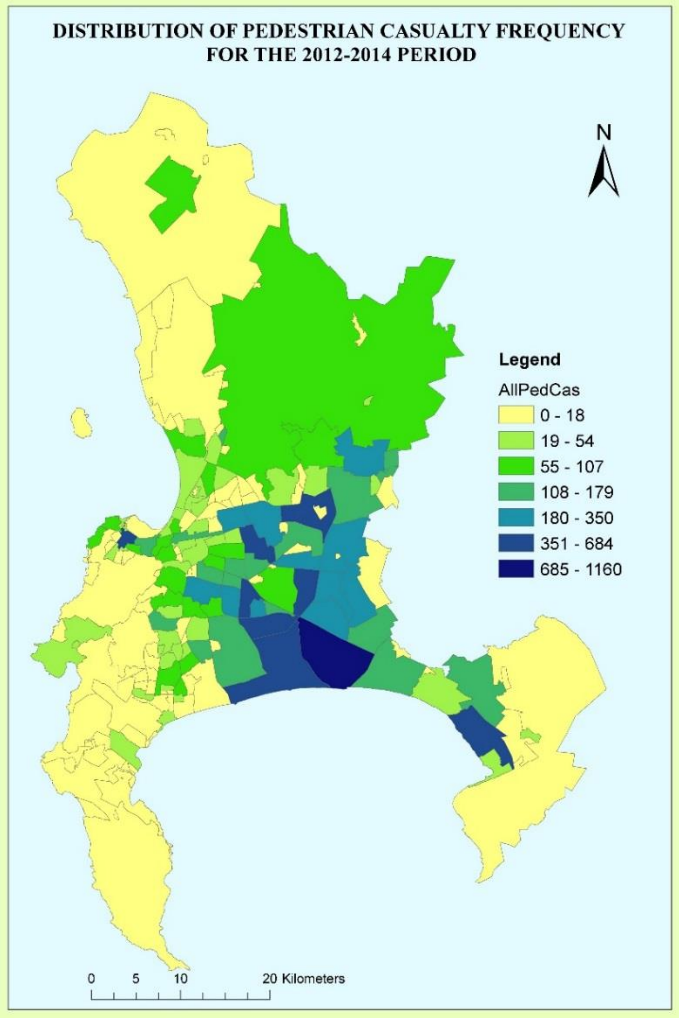


Figure 4-40: Spatial distribution of pedestrian casualty counts for the 2012-2014 period

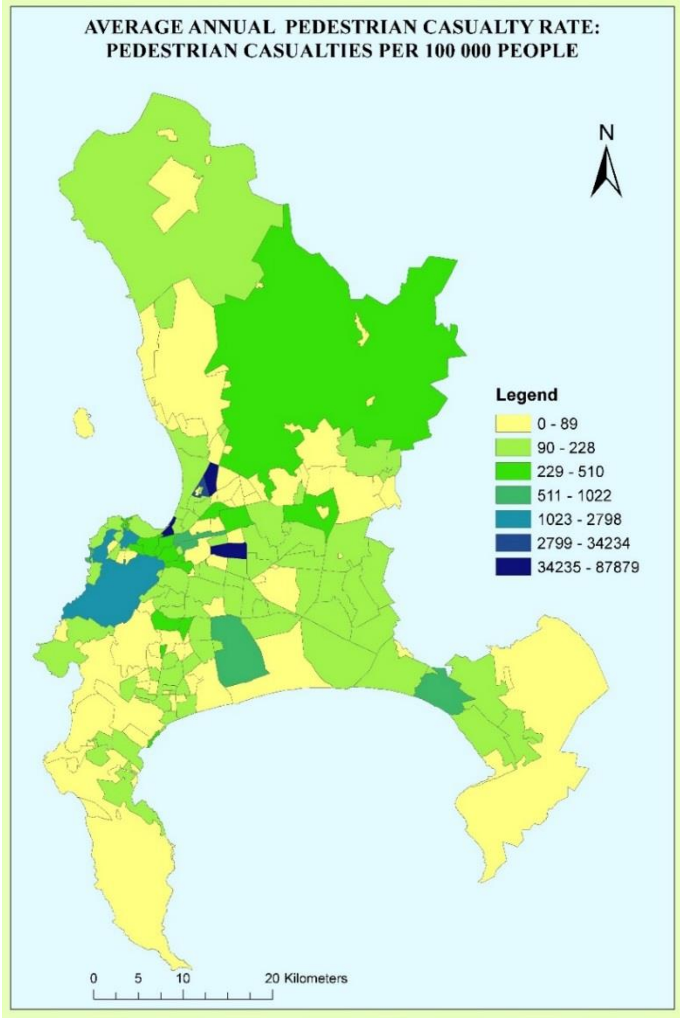


Figure 4-41: Spatial distribution of average annual pedestrian casualties per 100,000 people

4.2.1.2 Geostatistical analysis of pedestrian casualties across the study area

Three geostatistical methods were applied to the dataset of all pedestrian casualties aggregated at the census suburb level. These geostatistical methods include the Anselin Local Moran's I (or Local Moran's I statistic), the Getis-Ord Gi* and the Optimized Hot Spot analysis. The expected outcome of these analyses was the identification (with levels of statistical significance) of census suburbs regarded as hot spots or cold spots of pedestrian casualties as well as those in which the spatial pattern of pedestrian casualties is random (i.e. locations with no spatial autocorrelation among crash locations).

The spatial distribution of all pedestrian casualties analysed in this study is illustrated in Figure 4-42. The Hot Spot analysis (Getis-Ord Gi*) highlights one suburb identified as a cold spot at 95% confidence level and 27suburbs detected as hot spot locations of pedestrian casualties at different significance levels- 17 suburbs at 99%, 5 suburbs at 95% and 5 suburbs at 90% confidence levels (see Figure 4-43 and Table 4-33). With the use of the Local Moran's I tool, 14 suburbs are identified as hot spots. However, no cold spot or outlier was spotted in the study area by the Local Moran's I statistic (see Figure 4-44 and Table 4-33). The Optimized Hot Spot Analysis tool detects 13 census suburbs as hot spot locations of pedestrian causalities at 99 %, 2 suburbs at 95 % and 5 suburbs at 90% confidence levels (see Figure 4-45 and Table 4-33).

The three local statistics produce consistent results with regard to certain census suburbs, and this confirms that pedestrian casualties are clustered in those areas of the city. However, mixed results emerge for a number of suburbs which are qualified as hot spot locations of pedestrian casualties by one statistical method while other methods fails to spot casualty clustering in those census suburbs (see Table 4-33). Suburbs which were detected as being hot spot or cold spot locations by at least two statistical methods are confirmed in this study as being hot spot or cold spot locations of pedestrian casualties in Cape Town. In this way, 21 suburbs out of 190 are identified as being hot spot locations for pedestrian casualties in the City of Cape Town. These census suburbs are (1) Mitchells Plain, (2) Khayelitsha, (3) Mfuleni, (4) Philippi, (5) Delft, (6) Crossroads, (7) Nyanga, (8) Gugulethu, (9) Cape Town International Airport, (10) Bishop Lavis, (11) Freedom Park Airport, (12) Belhar, (13) Philippi Small Holdings, (14) Blackheath, (15) Blue Downs, (16) Bellville South, (17) Manenberg, (18) Kuils River, (19) Parow, (20) Eerste River and (21) Elsies River. For the entire study area, only one suburb, the Cape Peninsula National Park is found to be a cold spot for pedestrian casualties (see Figure 4-43).

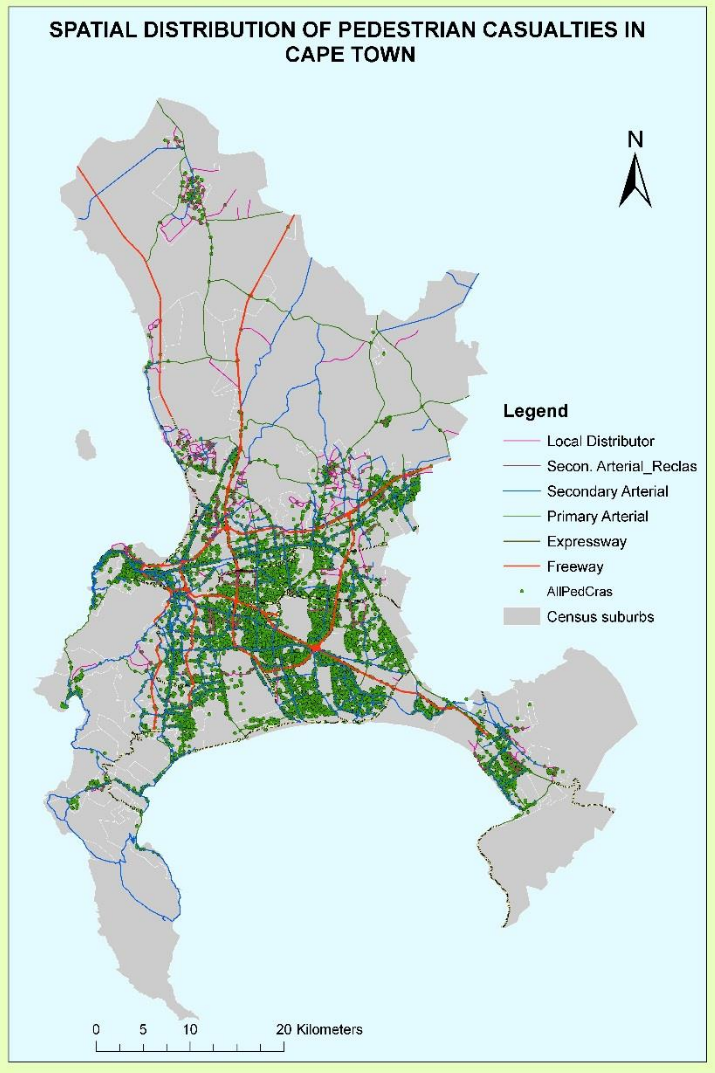


Figure 4-42: Spatial distribution of pedestrian casualty locations

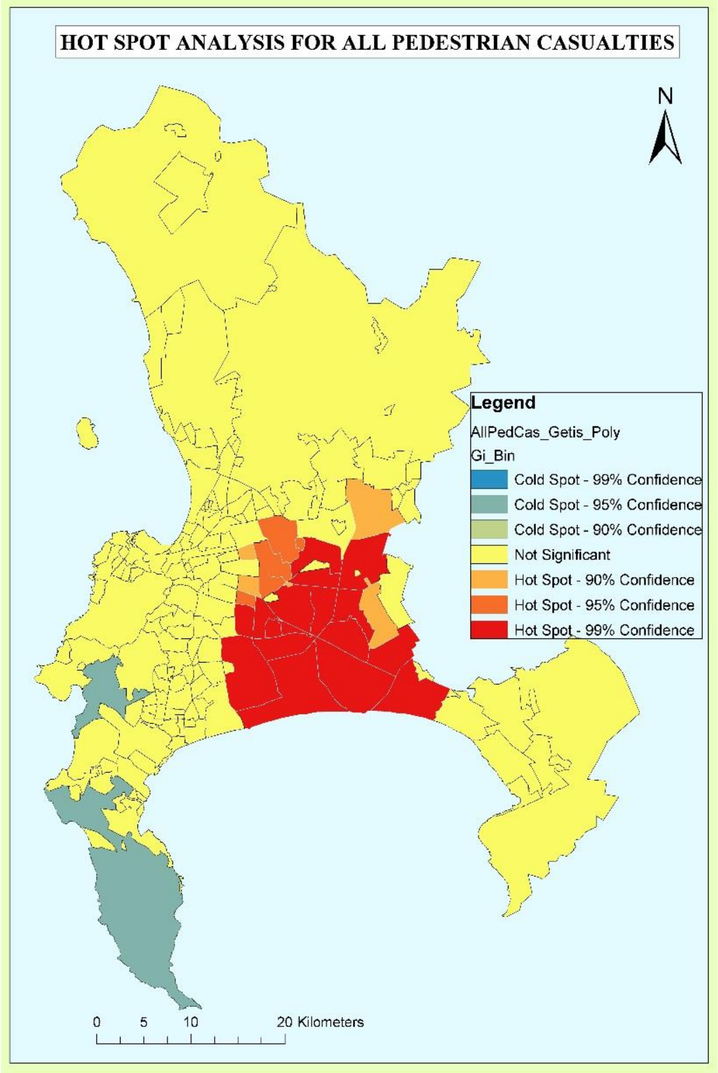


Figure 4-43: Cluster analysis of pedestrian casualties by the Getis-Ord Gi* tool

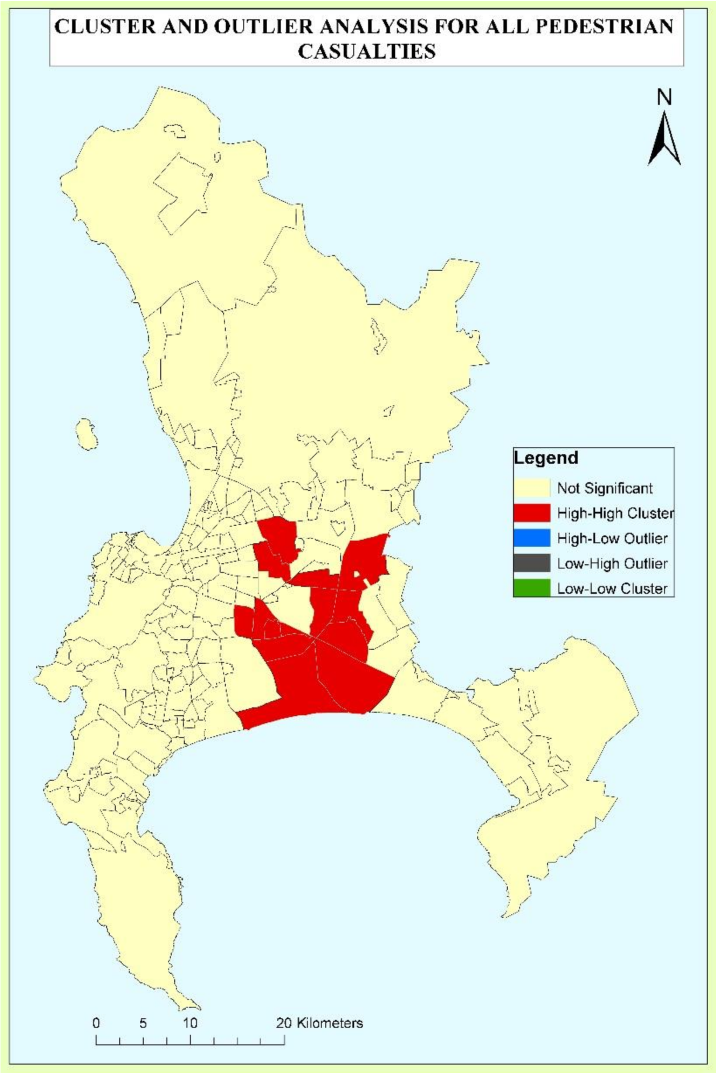


Figure 4-44: Cluster analysis of pedestrian casualties by Moran I tool

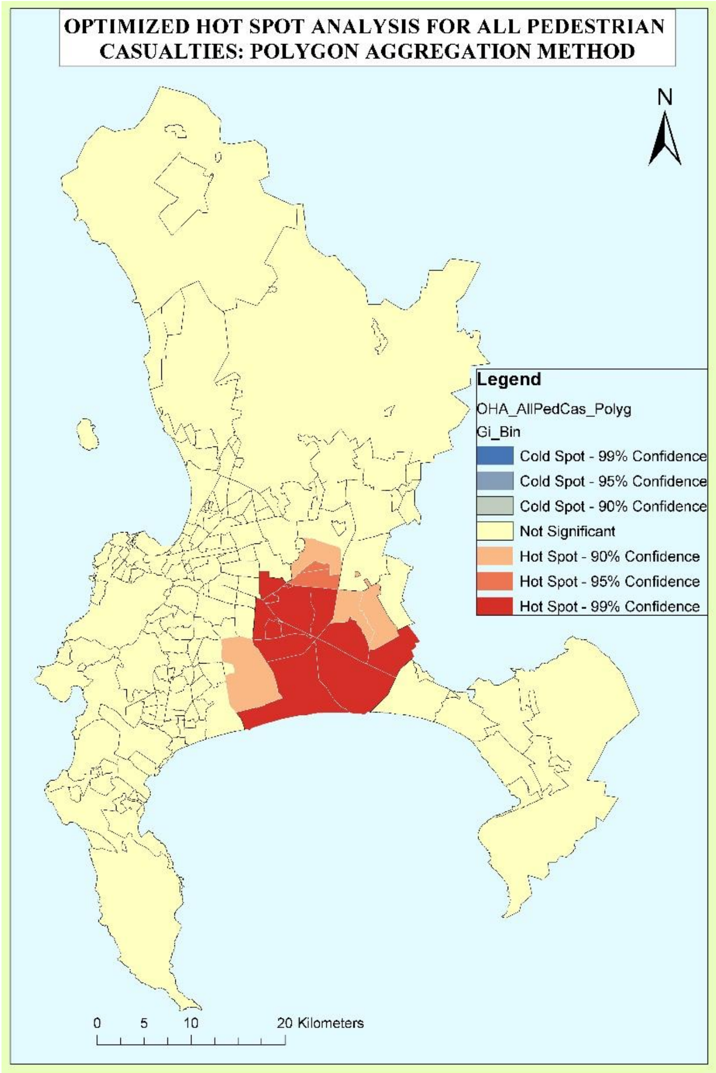


Figure 4-45: Cluster analysis of pedestrian casualties by the OHA tool

Table 4-33: Hot and cold spot suburbs of pedestrian casualties in Cape Town

OHA	Getis-Ord Gi*	Local Moran's I statistic
Mitchells Plain	Mitchells Plain	Mitchells Plain
Khayelitsha	Khayelitsha	Khayelitsha
Eerste River	Eerste River	
Mfuleni	Mfuleni	Mfuleni
Philippi	Philippi	Philippi
Delft	Delft	Delft
Crossroads	Crossroads	Crossroads
Nyanga	Nyanga	Nyanga
Gugulethu	Gugulethu	Gugulethu
Cape Town International Airport	Cape Town International Airport	
Montevideo		
Bishop Lavis	Bishop Lavis	
Freedom Park Airport	Freedom Park Airport	
Belhar	Belhar	Belhar
UWC/CPUT		
Philippi Small Holdings	Philippi Small Holdings	
Blackheath	Blackheath	
Blue Downs	Blue Downs	Blue Downs
Bellville Teachers' Training College		
Bellville South	Bellville South	
	Macassar	
	Manenberg	Manenberg
	Kuils River	Kuils River
	Parow	Parow
	Tygerberg Hospital	
	Elsies River	Elsies River
	Heideveld	
	Bonteheuwel	
	Brackenfell	
	Ruyterwacht	
	Cape Peninsula National Park	

LEGEND					
OHA		**Getis-Ord Gi***		**Local Moran I**	
(gold)	Hot spot at 99% CI	(gold)	Hot spot at 99% CI	(gold)	HH cluster
(blue)	Hot spot at 95% CI	(blue)	Hot spot at 95% CI		
(green)	Hot spot at 90% CI	(green)	Hot spot at 90%CI		
		(pink)	Cold spot at 95 CI		

4.2.2 Geospatial analyses of intersection-related pedestrian casualties

4.2.2.1 Geospatial analysis by the use of local statistics of spatial autocorrelation

The spatial distribution of intersection-related pedestrian casualties in Cape Town is illustrated in Figure 4-46. At a glance, a visual assessment of the spatial pattern of pedestrian casualties in Figure 4-46 can help to detect the presence of localised casualty clusters in the study area. All three local statistics – the Anselin Local Moran's I, the Getis-Ord Gi*, and the Optimized Hot Spot Analysis (OHA) – have succeeded in identifying clusters of pedestrian casualties. However, some variations in the shape and size of clusters detected by the three geospatial analysis tools are apparent. The hot spot identified by the Anselin Local Moran's I tool is very close in terms of size and shape to that identified by the Getis-Ord Gi* tool at 99% confidence level (see Figure 4-47 and Figure 4-48). As it was noticed previously for the entire sample of all pedestrian casualties, the Getis-Ord Gi* always outperforms other local statistics in identifying relatively larger hot spot areas than other statistics.

For a dataset of intersection-related pedestrian casualties aggregated at the suburb level, hot spot areas detected by the Local Moran's I consist of 14 census suburbs. One census suburb, Tygerberg Hospital, is identified by the tool as being a "Low-High" outlier. This finding suggests that while Tygerberg Hospital (as a census suburb) has fewer intersection-related pedestrian casualties, the neighbouring suburbs are characterised by higher numbers of intersection-related pedestrian casualties. This is not a surprising finding given the fact that Tygerberg Hospital suburb is totally covered by a public hospital and a learning institution where traffic patterns and risk exposure in general differ from those of the neighbouring suburbs.

At 99% confidence level, the hot spot identified by the Getis-Ord Gi* statistic consists of 14 census suburbs with 11 of these matching the hot spot region detected by the Local Moran's I. The size of the hot spot highlighted by the Getis-Ord Gi* extends on 40 census suburbs. In addition, two census suburbs, Silvermine and Cape Peninsula National Park, are detected by the Getis-Ord Gi* tool as cold spot areas of intersection-related pedestrian casualties at 90% and 95% confidence levels, respectively (see Figure 4-48).

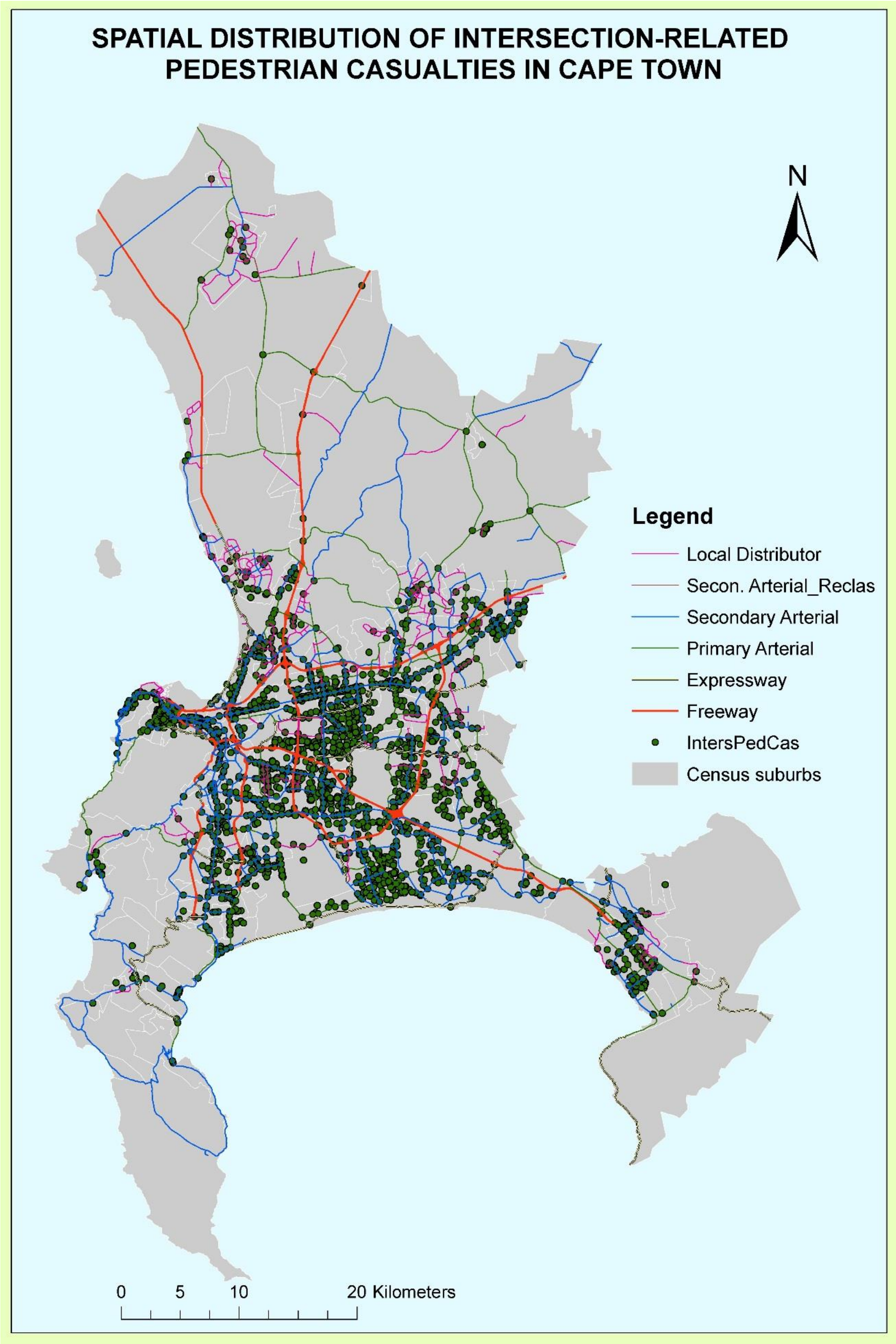


Figure 4-46: Spatial distribution of intersection-related pedestrian casualties in Cape Town

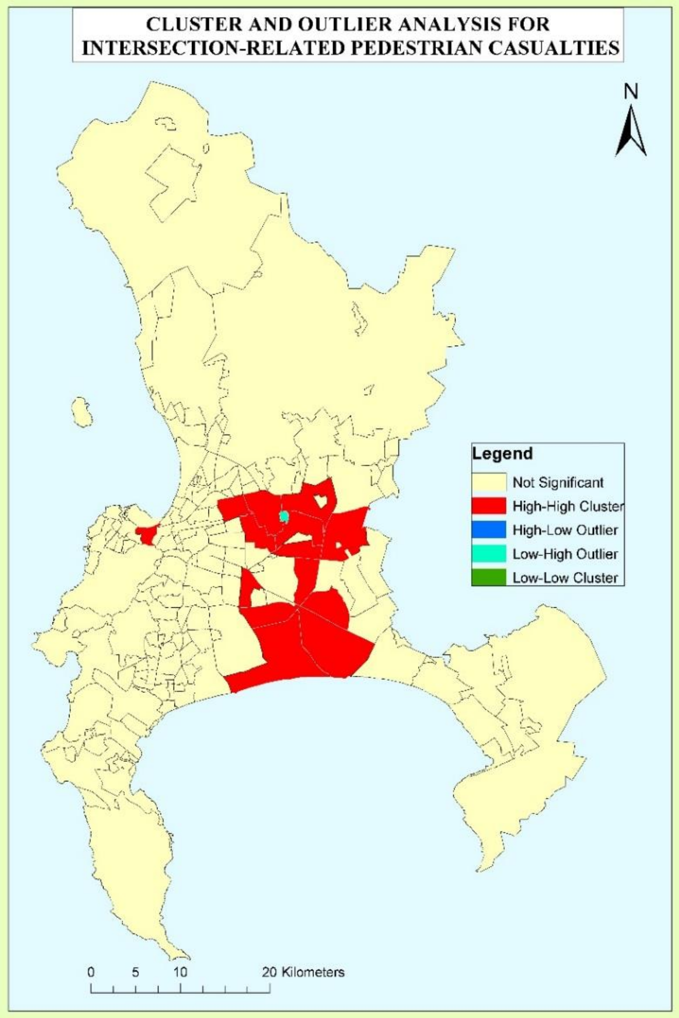


Figure 4-47: Clusters of intersection-related pedestrian c asualties by the Local Moran's I tool

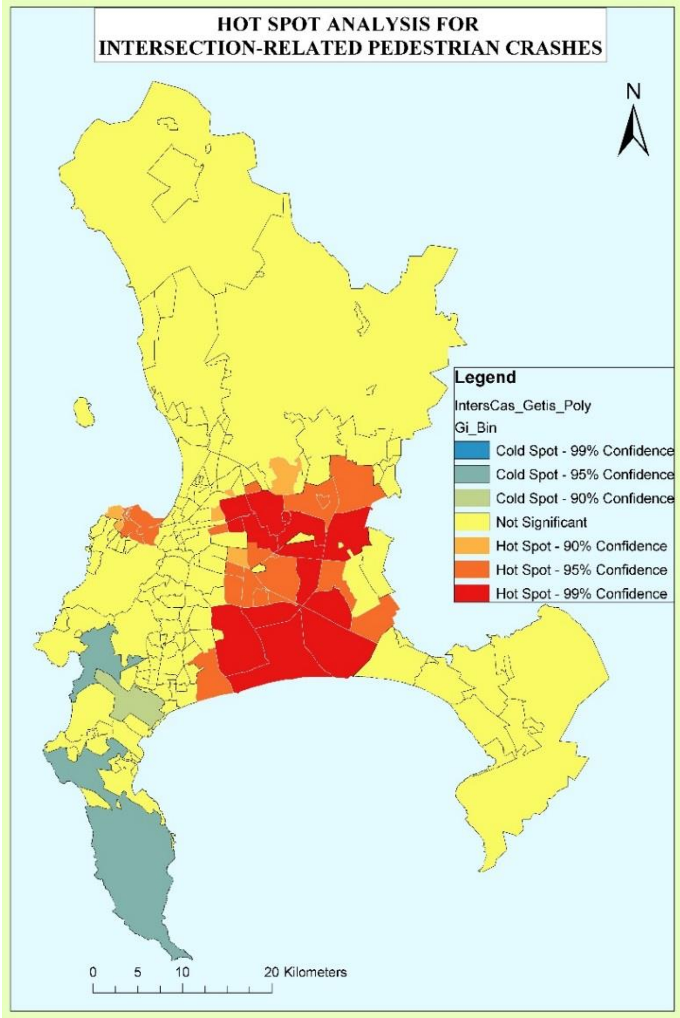


Figure 4-48: Clusters of intersection-related pedestrian casualties by the Getis-Ord Gi* tool

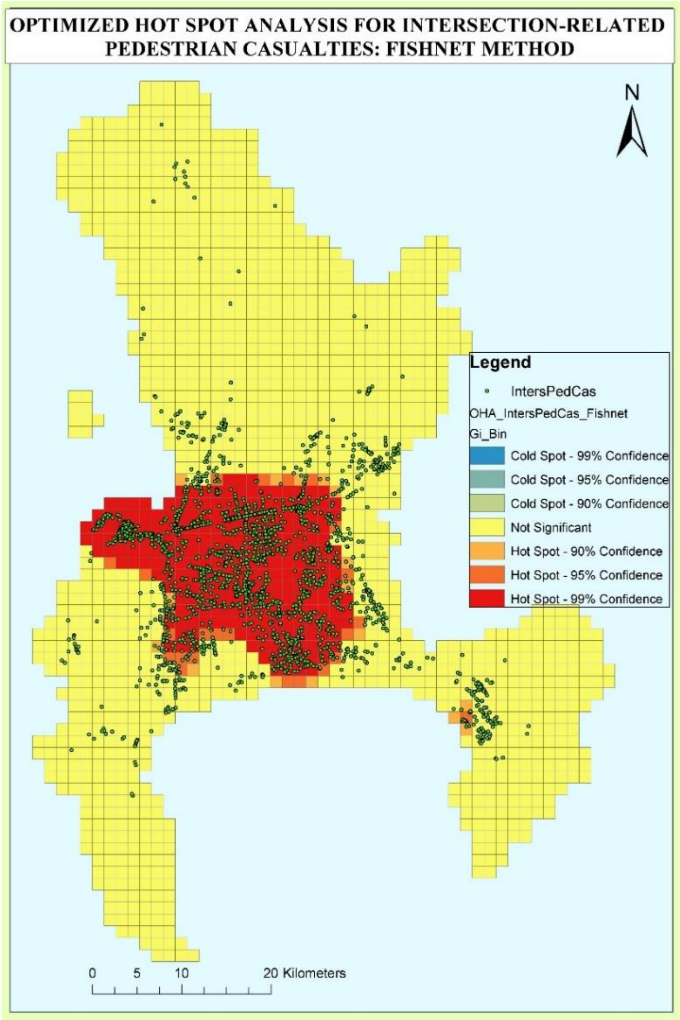


Figure 4-49: Clusters of intersection-related pedestrian casualties by the OHA Fishnet method

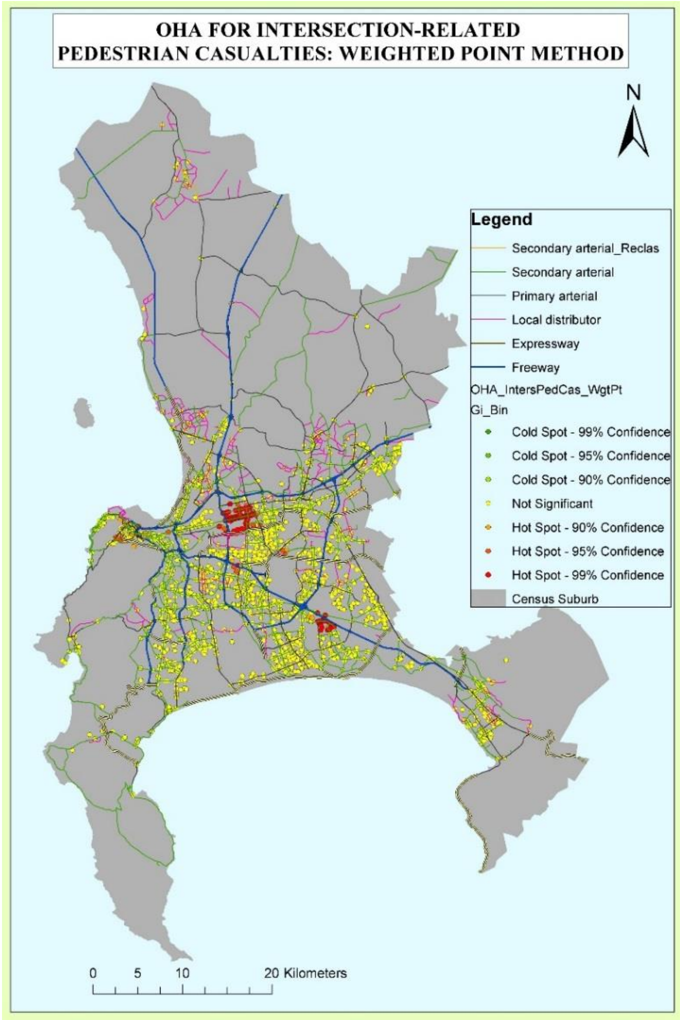


Figure 4-50: Clusters of intersection-related pedestrian casualties by the OHA Weighted Point method

The two techniques of the Optimized Hot Spot Analysis tool which are applied to intersection-related pedestrian casualties are the fishnet grid and the weighted point techniques. The hot spots detected by the respective tools are illustrated in Figure 4-49 and Figure 4-50.

With the use of the OHA weighted point technique, proximal pedestrian casualty locations are aggregated into a single point which is given a weight reflecting the number of pedestrian casualty points that were aggregated together. With the application of this technique to the dataset of intersection-related pedestrian casualties, a number of weighted points are detected as hot spots at 90 % to 99 % confidence. The weighted points (i.e. intersection-related casualty locations) are mapped in Figure 4-50 and a closer look at these points is displayed in Figure 4-51. The weighted points identified as hot spot locations of intersection-related pedestrian casualties show localised spatial clustering on the transportation system of the study area. Five spatial clusters of pedestrian casualties have been identified and these consist of intersection locations highlighted in black circles in Figure 4-51.

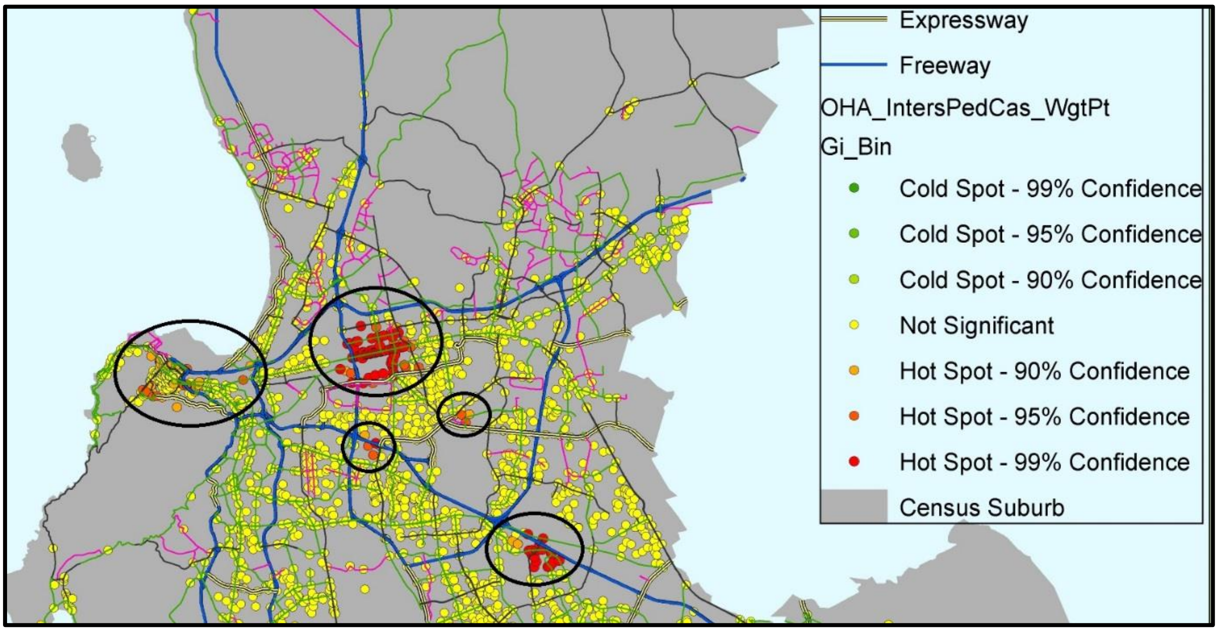


Figure 4-51: Hot spot locations detected by the OHA Weighted Point technique

The most significant clustering emerges in the area extending over four census suburbs which are Goodwood, Thornton, Ruyterwacht and Elsies River. The majority of the hot spot locations in this region are found on a section of Voortrekker Road (R102) between Jakes Gerwel Drive (M7) and Giel Basoon Drive (M12). Another significant clustering is apparent in Khayelitsha and Mfluleni suburbs between the R300 and Spine Road (M32). Furthermore, several hot spot locations are detected in four census suburbs of Cape Town Central. These are Cape Town CBD, Zonnebloem and Woodstock. A few hot spot locations are noticeable in the suburbs of

Heideveld and Bonteheuwel between Jakes Gerwel Drive (M17) and Robert Sobukwe Road (M10). Lastly, three weighted points located in Belhar between Robert Sobukwe Road (M10) and Stellenbosch Arterial (M12) are also detected as hot spot locations of intersection-related pedestrian casualties.

As seen in Figure 4-49, the hot spot locations identified by the OHA fishnet method extend on a larger area compared to those detected by the Getis-Ord Gi* tool displayed in Figure 4-48 and the Local Moran’s I tool shown in Figure 4-47. Moreover, unlike the Getis-Ord Gi*and the Local Moran’s I tools, both techniques of OHA, the fishnet and weighted point techniques, are unable to detect cold spot locations of intersection-related pedestrian casualties in the study area.

4.2.2.2 Geospatial analysis by the use of planar kernel density estimation (KDE)

The planar kernel density estimation (KDE) technique is another geospatial analysis tool applied in this study to identify hotspots of pedestrian casualties. With the KDE technique, hot spots are detected using a cell size of 30 metres and different bandwidths ranging from 200 metres to 1895 metres. The latter bandwidth was adopted from the OHA density surface method and the exact value of this bandwidth is applied to the KDE technique for comparison purposes.

The hot spot locations detected by KDE using the bandwidth sizes of 200 m; 400 m; 500 m; 100 m; and 1895 m are respectively mapped in Figure 4-52; Figure 4-53; Figure 4-54; Figure 4-55; Figure 4-56; and Figure 4-57. One important observation from these analyses is that an increase of bandwidth size offers a better visualisation of mapped hotspots on a road network. However, larger bandwidth sizes result in wider hot spot regions with less variability between areas. Therefore, it is recommended to apply an optimal bandwidth for Kernel Density Estimation, which is often a subjective matter.

.

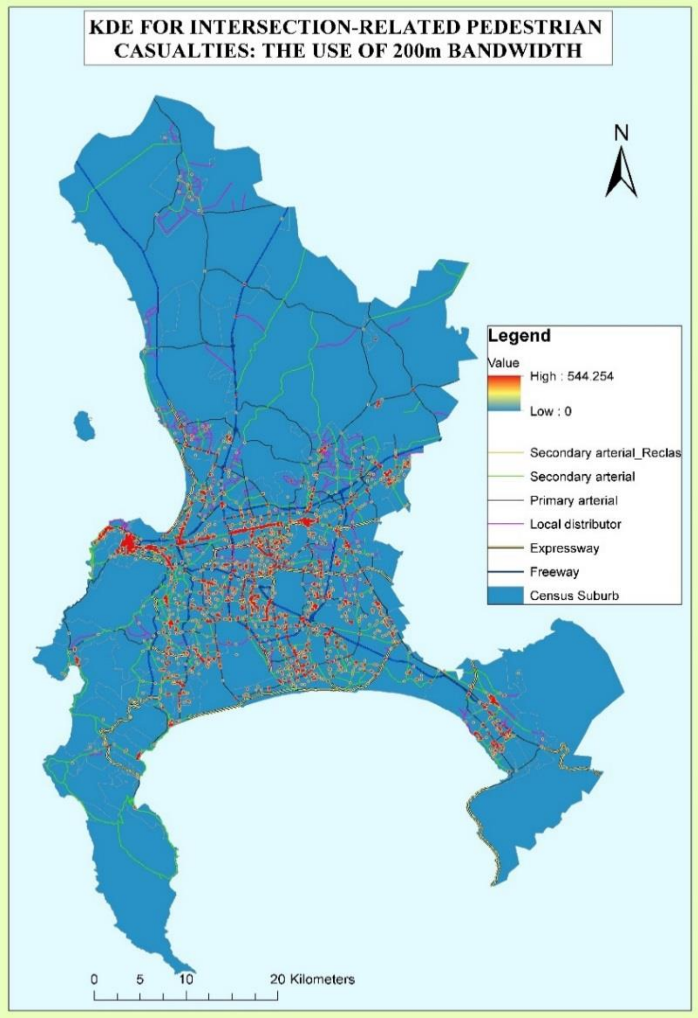


Figure 4-52: Hot spots of intersection-related pedestrian casualties by KDE 200m bandwidth

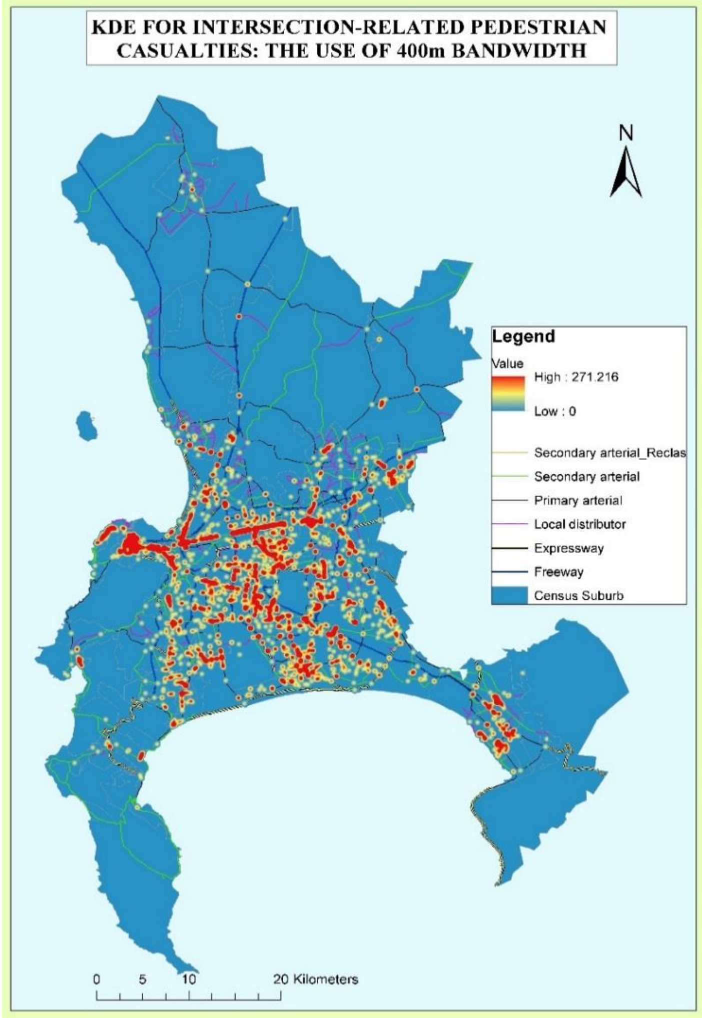


Figure 4-53: Hot spots of intersection-related pedestrian casualties by KDE 400m bandwidth

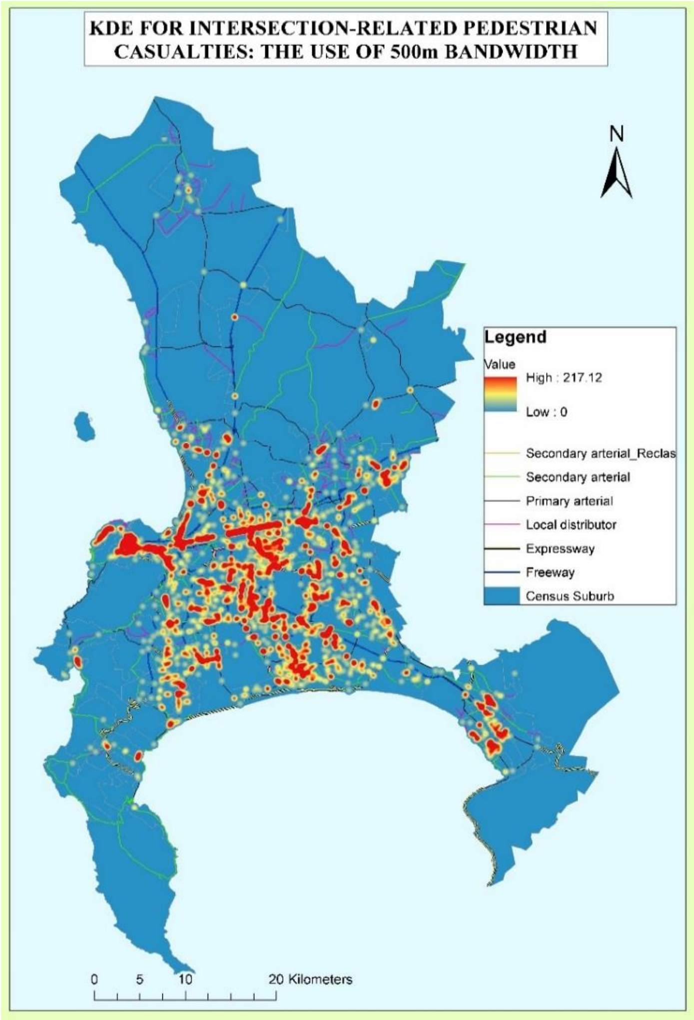


Figure 4-54: Hot spots of intersection-related pedestrian casualties by KDE 500m bandwidth

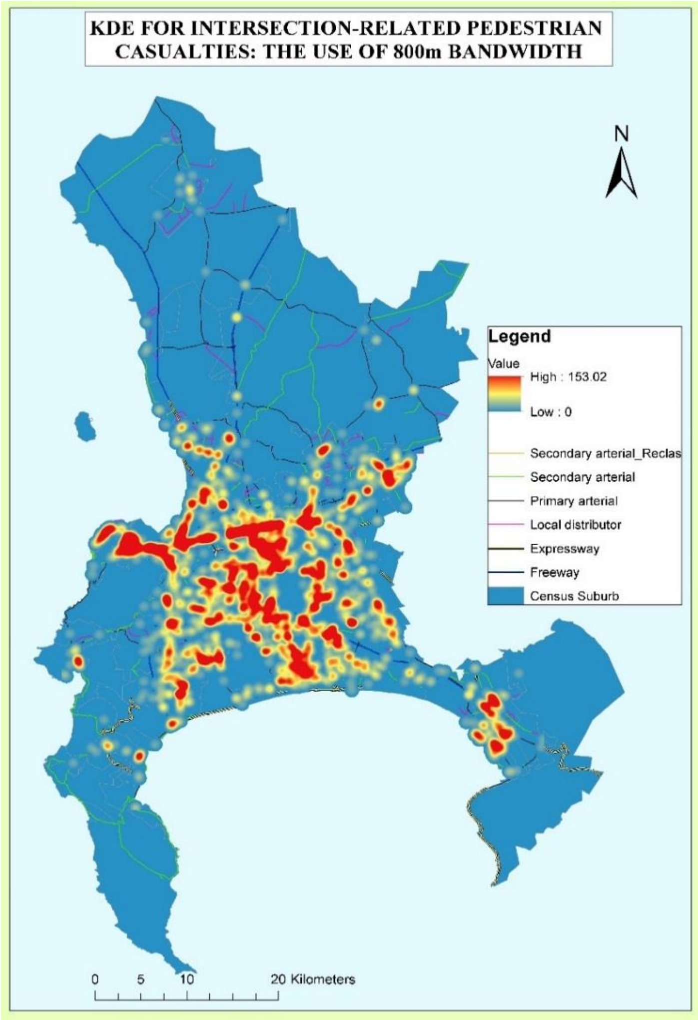


Figure 4-55: Hot spots of intersection-related pedestrian casualties by KDE 800m bandwidth

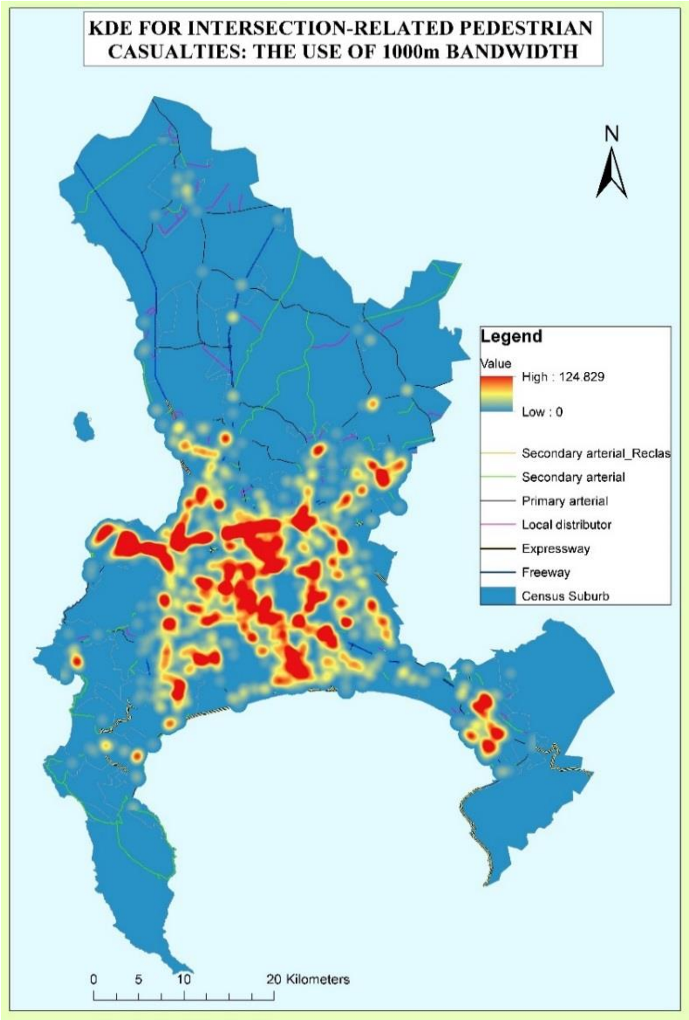


Figure 4-56: Hot spots of intersection-related pedestrian casualties by KDE 1000m bandwidth

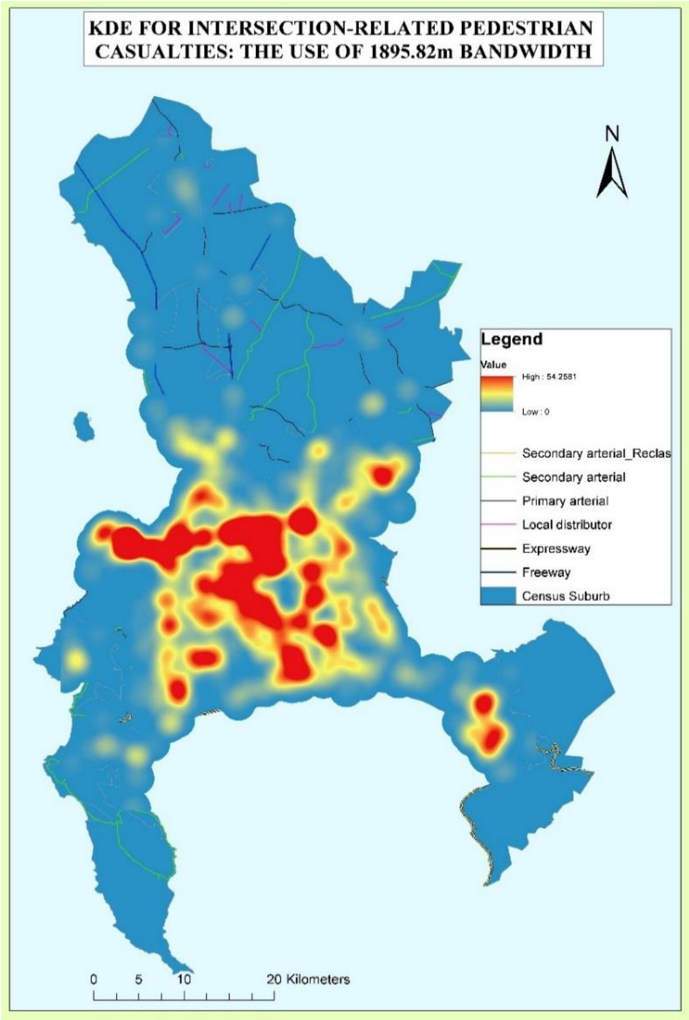


Figure 4-57: Hot spots of intersection-related pedestrian casualties by KDE 1895.82m bandwidth

A close examination of hot spot locations of pedestrian casualties can easily detect a relationship between the incidence of pedestrian casualties and the road network structure. For instance, as seen in Figure 4-58, the application of Kernel Density Estimation with the bandwidth of 500 metres to the dataset of intersection-related pedestrian casualties clearly shows that the large majority of hot spot locations are identified on two classes of road: arterial roads (i.e. primary and secondary arterial roads) and urban freeways (i.e. freeways and expressways). Hot spot locations detected on urban freeways consist mainly of at-grade junctions where two freeways or a freeway and another road of a lower class (mostly arterial roads) intersect.

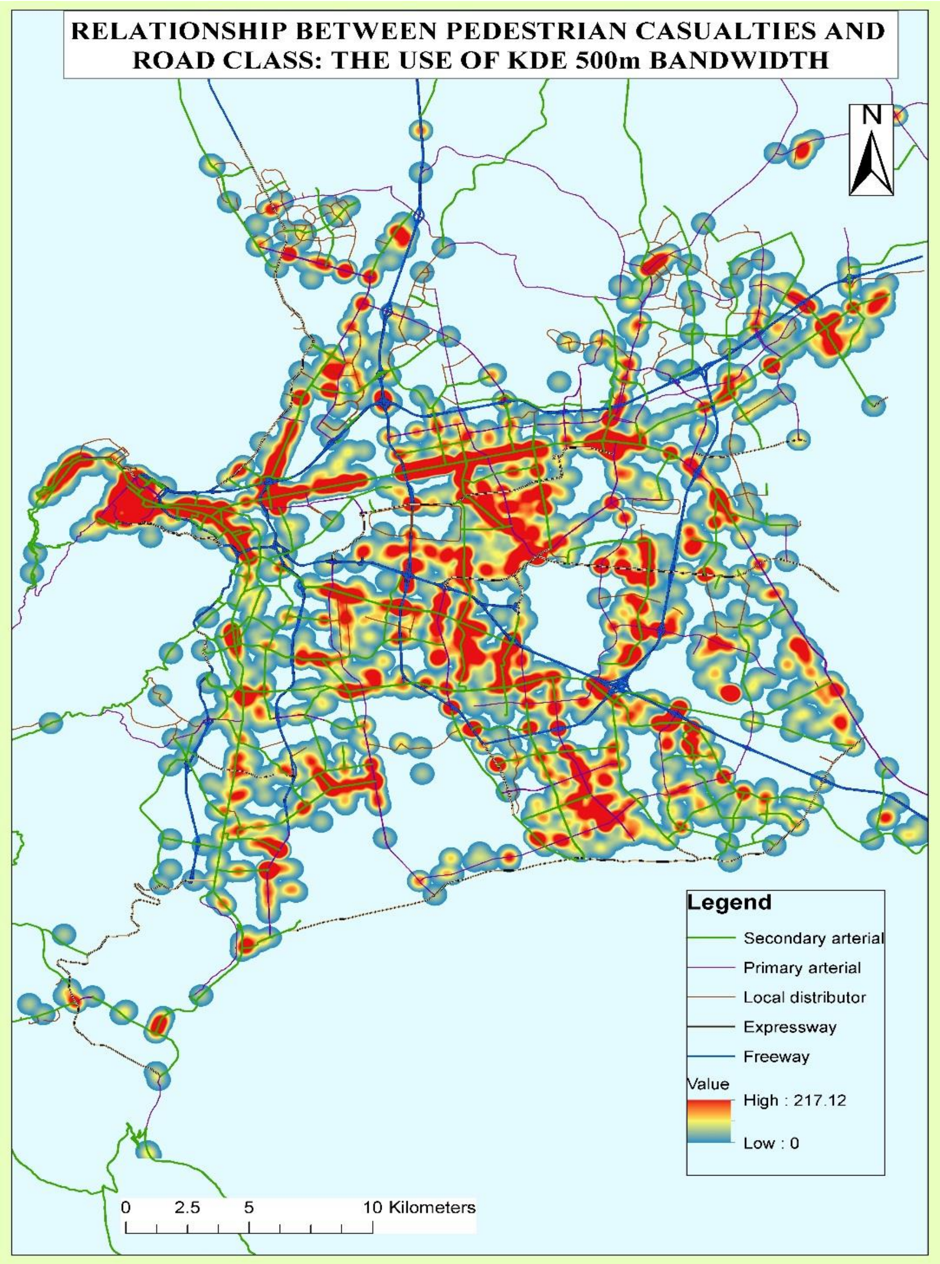


Figure 4-58: Relationship between pedestrian casualties and the road network structure

4.2.3 Geospatial analyses of pedestrian casualties by injury severity

The results from geospatial analyses carried out on pedestrian injuries, KSI pedestrian casualties and pedestrian casualties who sustained slight injuries are presented in Figure 4-59 to Figure 4-62. The results clearly show that the hot spot locations of KSI pedestrian casualties are more clustered than those of pedestrians who sustained slight injuries.

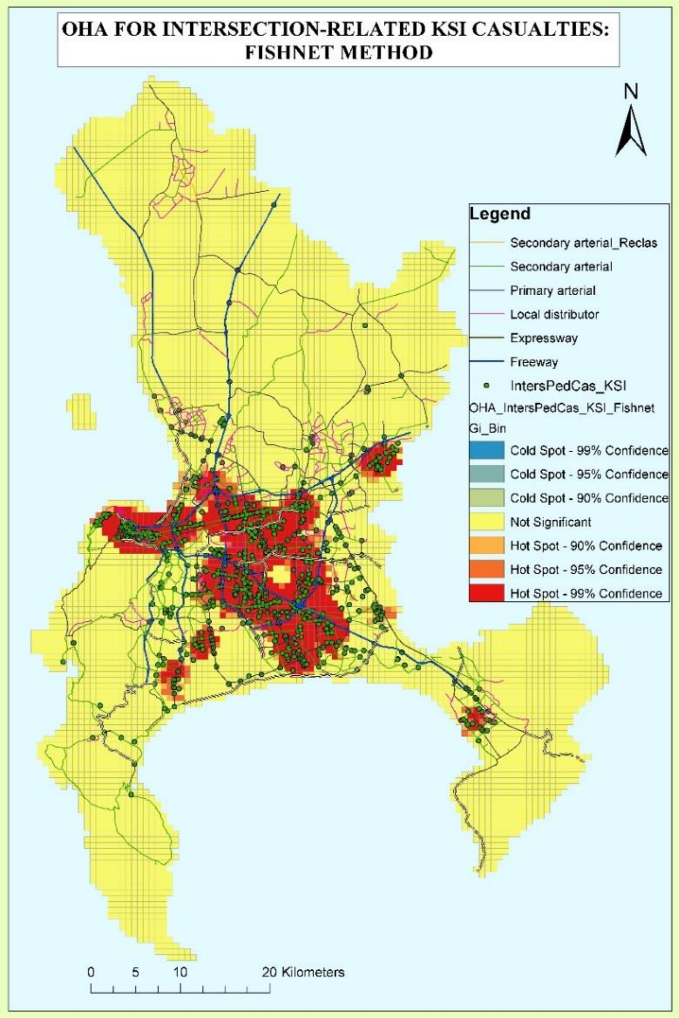


Figure 4-59: Cluster analysis of KSI pedestrian casualties by the OHA Fishnet method

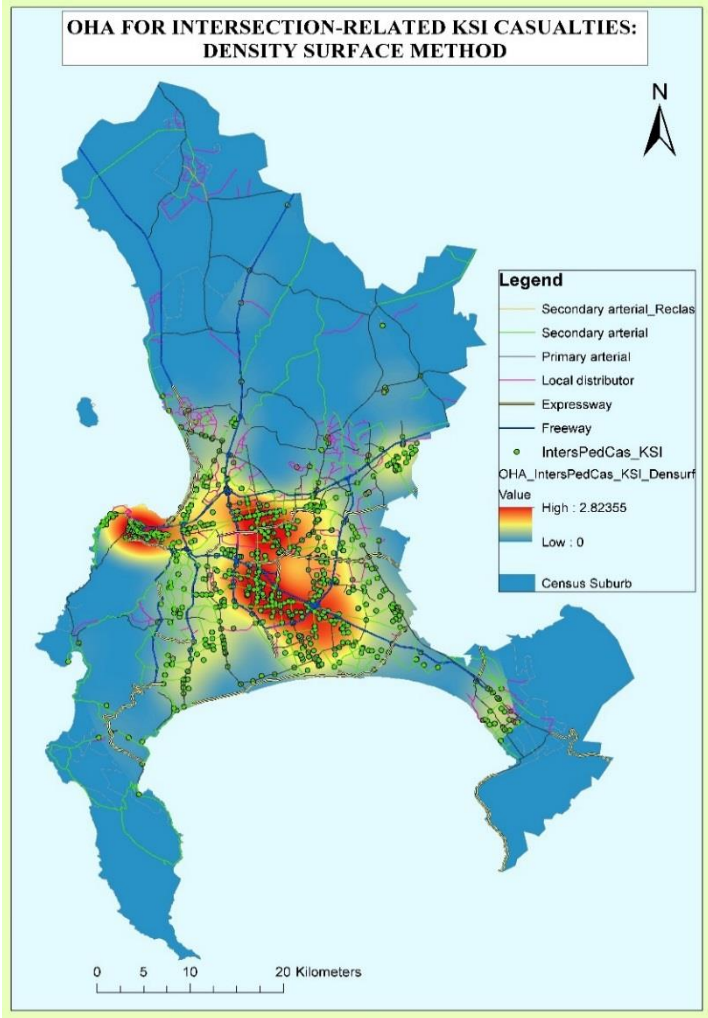


Figure 4-60: Cluster analysis of KSI pedestrian casualties by the OHA Density Surface method

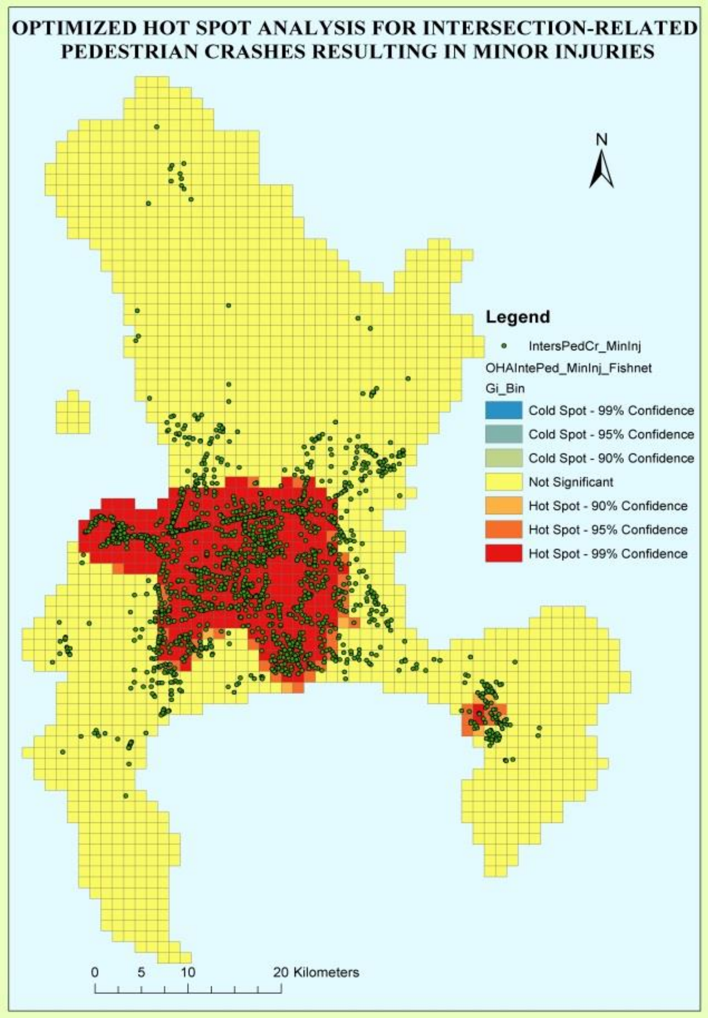


Figure 4-61: Cluster analysis of slight injuries by the OHA Fishnet method

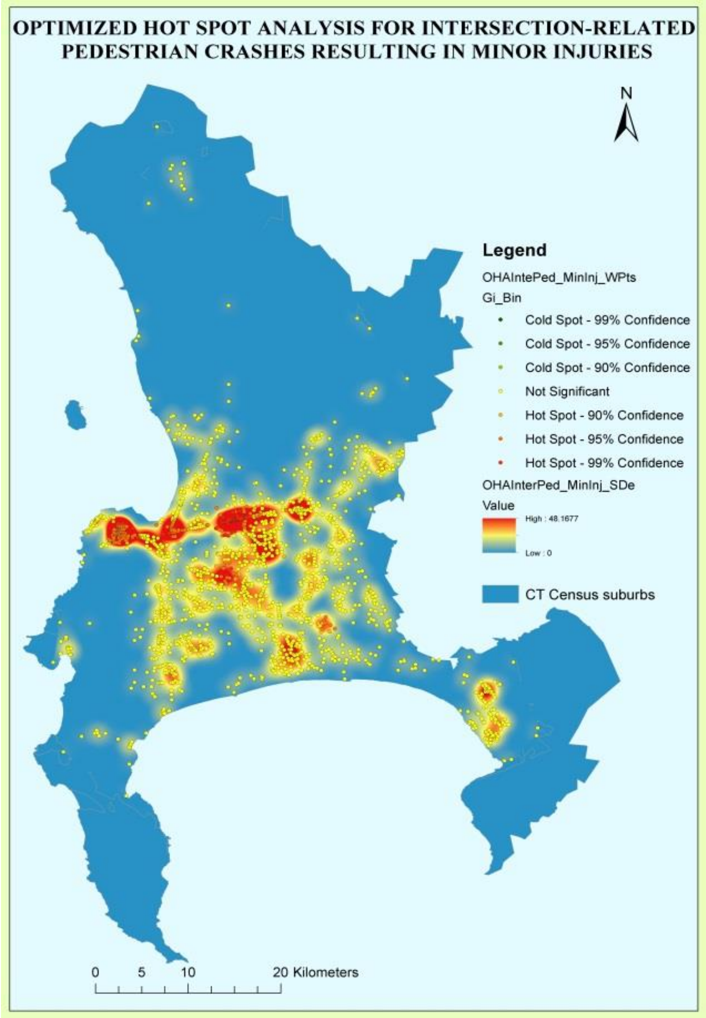


Figure 4-62: Cluster analysis of slight injuries by the OHA weighted point method

4.2.4 Design characteristics of intersections locations for pedestrian casualties

An index score ranging from 0 to 100 is calculated for each intersection identified as a pedestrian casualty location. This index reflects the characteristics of pedestrian facilities provided at an intersection and the extent to which an intersection accommodates pedestrians and facilitates them to negotiate it safely. The index is calculated using four proxy indicators including: the length of a pedestrian crossing (expressed in number of travel lanes); the number of pedestrian refuges available at an intersection; the availability of designated pedestrian crossings; and availability of sidewalks. A score of 0 is indicative of the lowest level of pedestrian accommodation at an intersection while the opposite is indicated by a score of 100. The overall index is an arithmetic mean of index scores of the four proxy indicators of pedestrian accommodation at an intersection. Although all intersections concerned with this analysis are locations of at least one pedestrian casualties, it should be pointed out that index scores assigned to intersections do not reflect the level of pedestrian safety in terms of casualty frequency or injury severity.

4.2.4.1 Descriptive statistics for intersection index scores across the study area

The distribution of index scores assigned to the dataset of intersection-related pedestrian casualties is illustrated in Figure 4-63. Index scores are estimated for intersection locations where a total of 3 533 pedestrian casualties occurred. The mean index score for the entire sample is found to be 71.1 (SD=13.73). The lowest index score emerges to be 32.63 and intersections with the highest level of pedestrian accommodation have a score of 97.23. Furthermore, percentile results indicate that a quarter of all analysed pedestrian casualty locations have an index score greater than 82.41.

Intersection index for pedestrian facilities		
N	Valid	3533
Mean		71.1080
Median		70.8417
Std. Deviation		13.73247
Skewness		-0.329
Std. Error of Skewness		0.041
Kurtosis		-0.547
Std. Error of Kurtosis		0.082
Minimum		32.63
Maximum		97.23
Percentiles	25	62.1361
	50	70.8417
	75	82.4134

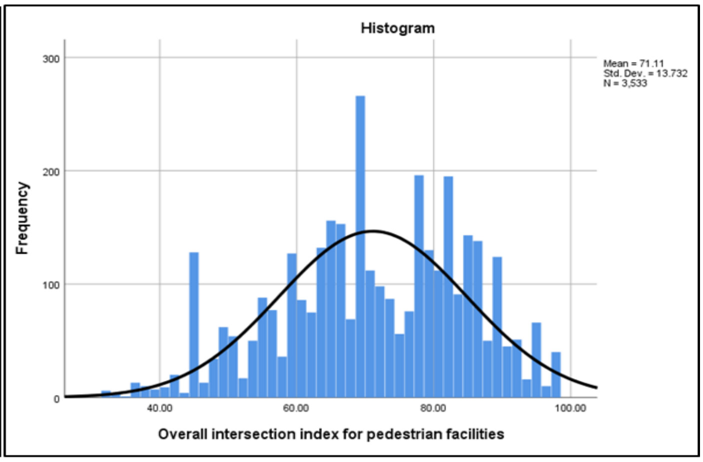


Figure 4-63: Descriptive statistics for intersection index scores

4.2.4.2 Geospatial analysis of intersection index scores

The Local Moran's I statistic was applied to the dataset of intersection-related pedestrian casualties with index scores as attribute values attached to pedestrian casualty locations. Intention behind the use of this tool is to assess spatial patterns of pedestrian casualty locations with respect to the magnitude of index scores attached to them. Simply put, the analysis is intended to determine whether there is a statistical clustering tendency among: (1) casualty locations with lower index scores (i.e. Low-Low Cluster); (2) casualty locations with higher index scores (i.e. High-High Cluster); (3) casualty locations with higher index scores surrounded by those with lower index scores (i.e. High-Low Outlier); and (4) casualty locations with lower index scores surrounded by those with higher index scores (i.e. Low-High Outlier).

The first step of this analysis was to ascertain whether there is a clustering tendency in the entire dataset. This was done by running the Global Moran's I index in ArcMap. The summary report generated by the Global Moran's I statistic is provided in Figure 4-64. According to this report, the Moran's Index is found to be 0.908215. This is a positive value which suggests that there is a positive autocorrelation in the dataset. In other words, the value of Moran's Index implies that there is a tendency for casualty locations with similar index scores to cluster together (i.e. High-High and Low-Low Clusters). The produced z-score informs that the spatial autocorrelation is statistically significant.

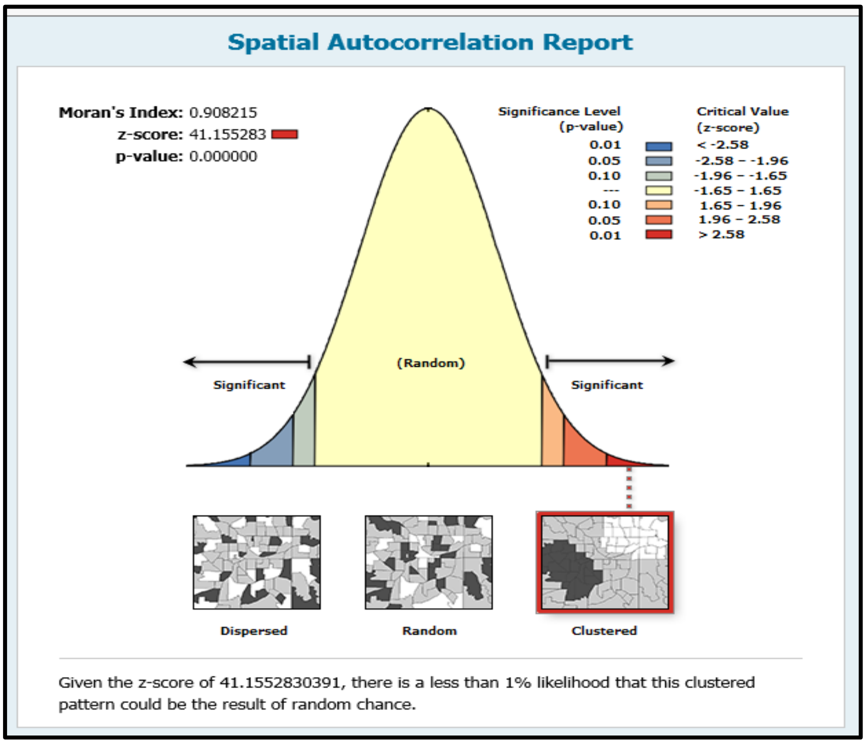


Figure 4-64: Summary report generated by the Global Moran's I statistic

After testing the spatial association in the dataset, the Local Moran's I statistic was applied to identify localised clustering in the dataset. This process involves first the determination of average distance required for any given pedestrian casualty location to have at least one neighbouring pedestrian casualty location. This was carried out in ArcMap using the tool "Calculate Distance Band from Neighbour count". The results generated by this tool indicate that the average critical distance (or distance bandwidth) of 120.54 meters is required for any given casualty location to have at least one neighbour. The next step involves the determination of a scale at which the spatial autocorrelation is the most pronounced. This was done in ArcMap by using the tool "Incremental Spatial Autocorrelation". The critical distance of 200 metres was used as the starting distance to search the maximum autocorrelation with an increment of 100 metres. The results from this process are presented in APPENDIX E. According to this test, the clustering of pedestrian casualty locations is maximum at a distance of 900 metres. This is subsequently the distance value that was used in the input of the Cluster and Outlier Analysis (i.e. Local Moran's I) tool to detect clusters and outliers in the dataset of intersection-related pedestrian casualties. Mapped clusters (cold spots and hot spots) and outliers resulting from the aforementioned processes are illustrated in Figure 4-65.

The most interesting clusters are the Low-Low clusters (i.e. "Cluster: Low" according to Figure 4-65) which reflect clusters of pedestrian casualty locations with lower values of intersection index scores. The Cluster and Outlier Analysis tool identifies 302 pedestrian casualty locations with lower intersection index scores. These are the locations that need remedial treatments to improve pedestrian safety at these facilities. Significant clusters of casualty locations with poor intersection index scores are identified in a number of suburbs including Strand, Macassar, Eerste River, Blackheath, Khayeltisha, Philippi, Kuils River, Dunoon, Fisantekraal, Hout Bay, to name a few. Low-High Outliers are also locations of a great concern as they represent pedestrian casualty locations with poor level of pedestrian accommodation in comparison to that of the neighbouring locations. The presence of High-High Clusters (or hot spot locations) indirectly implies that pedestrian crashes may be influenced by contributory factors other than the intersection design elements involved in the determination of index scores.

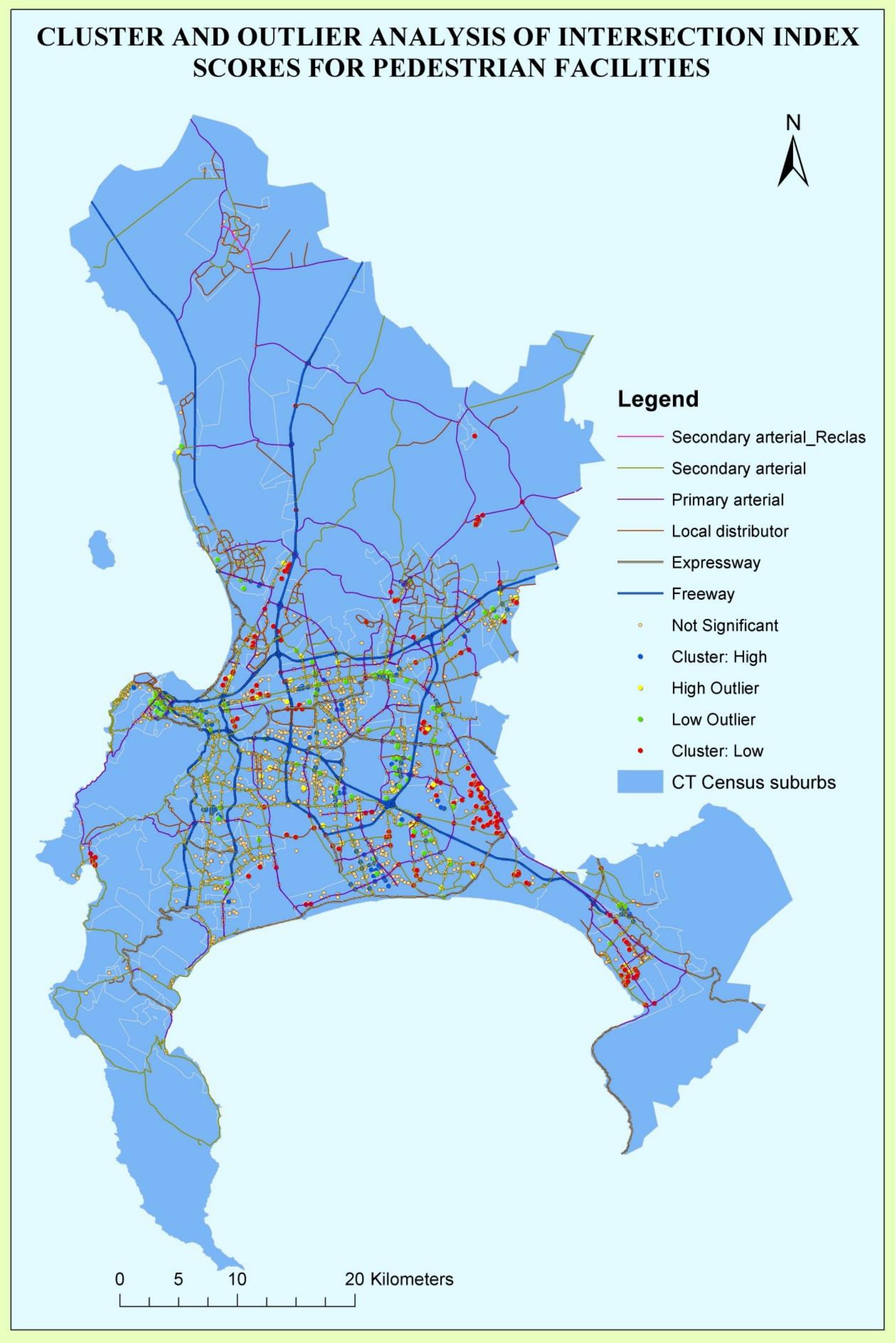


Figure 4-65: Clusters and outliers for locations of intersection-related pedestrian casualties

Google images taken at several intersections which scored the lowest index scores (i.e. score less than 50) are presented in Figure 4-66; Figure 4-67; Figure 4-68; Figure 4-69 and Figure 4-70. A visual inspection of these intersections clearly shows two common characteristics, the lack of pedestrian crossing facilities and longer crossing distances. A thorough examination also reveals that sidewalks are not provided at all approaches of the intersections.

Figure 4-66: Intersection: Francie Van Zijl Dr X Boulevard Park X Tienie Meyer Bypass (Source: Google)

Figure 4-67: Intersection: M7 (Jakes Gerwel Dr) X Milton Rd X Road to Wingfield Aerodrome (Source: Google)

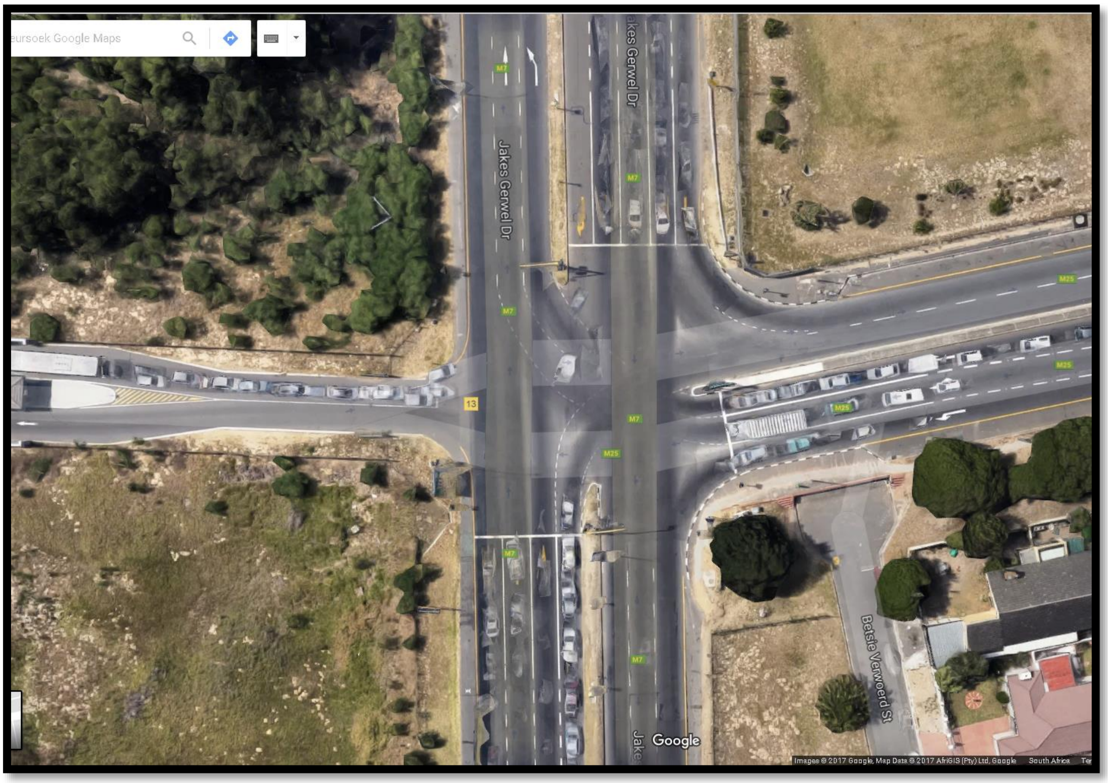


Figure 4-68: Intersection: M7 (Jakes Gerwel Dr) X Frans Conradie Dr X Rd to Wingfield House (Source: Google)

Figure 4-69: Intersection: Bottelary Rd X Kruis St X Langverwacht Rd (Source: Google)

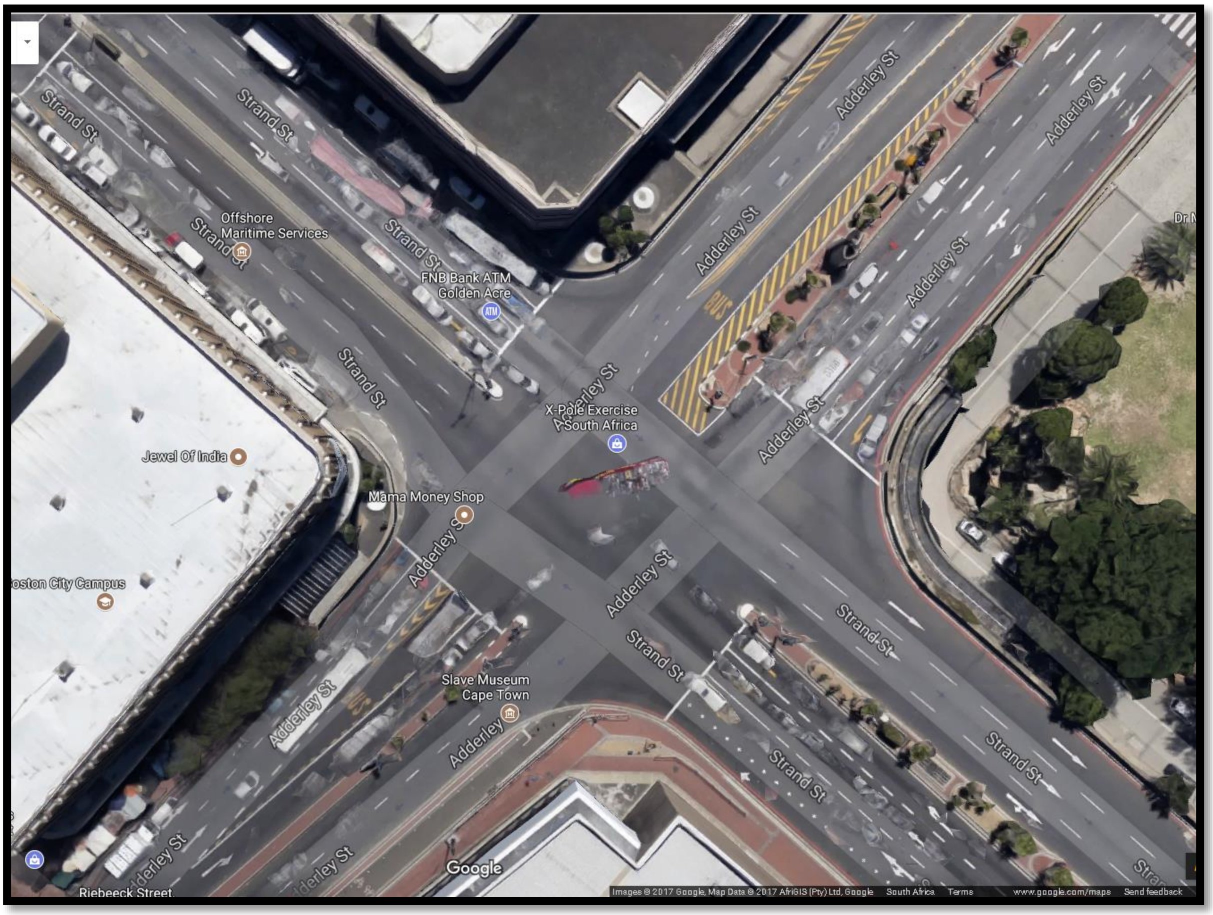


Figure 4-70: Intersection: Adderley St X Strand St (source: Google)

4.3 Comparison of the methods of cluster analysis

Table 4-34 presents the calculated vales of Prediction Accuracy Index (PAI) for the planar KDE with different bandwidth sizes ranging from 400 metres to 1895 metres and two techniques of Optimized Hot Spot Analysis. The PAI values presented in Table 4-34 indicate the performance of the analysis tool in predicting hot spots of events (i.e. pedestrian casualties). The larger the PAI value the better the performance of the geospatial analysis tool. An example of clipped hot spot regions which were analysed to estimate the PAI values for each tool is provided in Figure 4-71.

Table 4-34: PAI for cluster analysis methods applied in the study

	OHA Density Surface	OHA Fishnet	KDE 400m Bandwidth	KDE 500m Bandwidth	KDE 800m Bandwidth	KDE 1000m Bandwidth	KDE 1895.82m Bandwidth
The number of pedestrian casualties in hotspots regions	2005	2746	3204	3104	2683	2492	2007
Length of the road network in hotspot regions [m]	2603547	6866060	2876973.61	2982414.87	2991129.42	2922941.9	2603658.7
Area of hotspot regions [m^2]	1.42E+08	5.08E+08	145290600	152874000	157570200	155487600	141626700
Total length of the road network [m]	13695025	13695025	13695025	13695025	13695025	13695025	13695025
Total area of the study area [m^2]	2.46E+09	2.46E+09	2459962106	2459962106	2459962106	2.46E+09	2.46E+09
PAI	**2.931**	**1.522**	**4.239**	**3.961**	**3.414**	**3.245**	**2.934**

According to PAI values calculated for each geospatial analysis technique, the kernel density estimation (KDE) with 400 m bandwidth emerges as the best performing tool when compared with other analysis tools. The Optimized Hot Spot Analysis (OHA) tool which aggregates incident data into fishnet grids shows the lowest performance in predicting hot spots of pedestrian casualties when compared with the rest.

Figure 4-71: An example of clipped KDE hot spots with casualty points and the road network

4.4 Results from multivariate analysis of pedestrian casualties

This study applied three modelling approaches to uncover associations between the attributes of the built environment and the incidence of pedestrian casualties. The three modelling approaches tested in this study are the Poisson Regression Model, the Negative Binomial (NB) Regression Model and the Geographically Weighted Regression (GWR) Model. The three modelling techniques were applied to three datasets of pedestrian casualties. The first dataset includes a sample of all pedestrian casualties, the second dataset consists of intersection-related pedestrian casualties and the third datasets comprises killed and seriously injured (KSI) pedestrian casualties. Models developed based on the respective datasets of pedestrian casualties are referred to as Models 1, Models 2, and Models 3.

4.4.1 Description of dependent variables

The descriptive statistics for the three outcome variables: (1) All pedestrian casualties; (2) intersection-related pedestrian casualties and (3) Killed and Seriously Injured (KSI) pedestrian casualties is described in Figure 4-72, Figure 4-73 and Figure 4-74, respectively. In addition, these figures also provide the results of the test for normality performed on the distribution of the outcome variables by using a combination of methods such as visual inspection of the shape of the histogram, boxplots and the normal P-P plot, kurtosis and skewness measures, Kolmogorov-Smirnoff and Shapiro-Wilk tests. These tests indicate that the distributions of the three samples do not follow a normal distribution. For the three distributions, the values of the variance values are significantly higher than the mean values, implying that the three datasets of pedestrian casualties are subjected to over-dispersion. Accordingly, the application of the Poisson regression model may not be appropriate as the basic underlying assumption of this modelling procedure is that the mean should be equal to the variance.

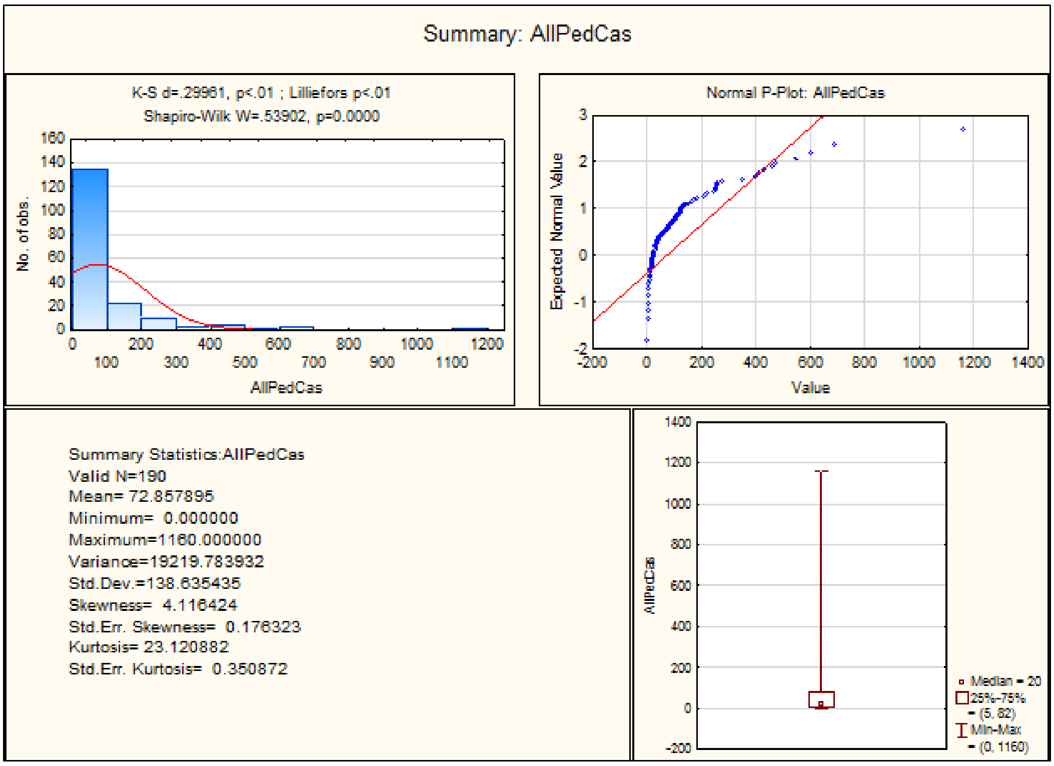


Figure 4-72: Summary statistics for the entire sample of pedestrian casualties

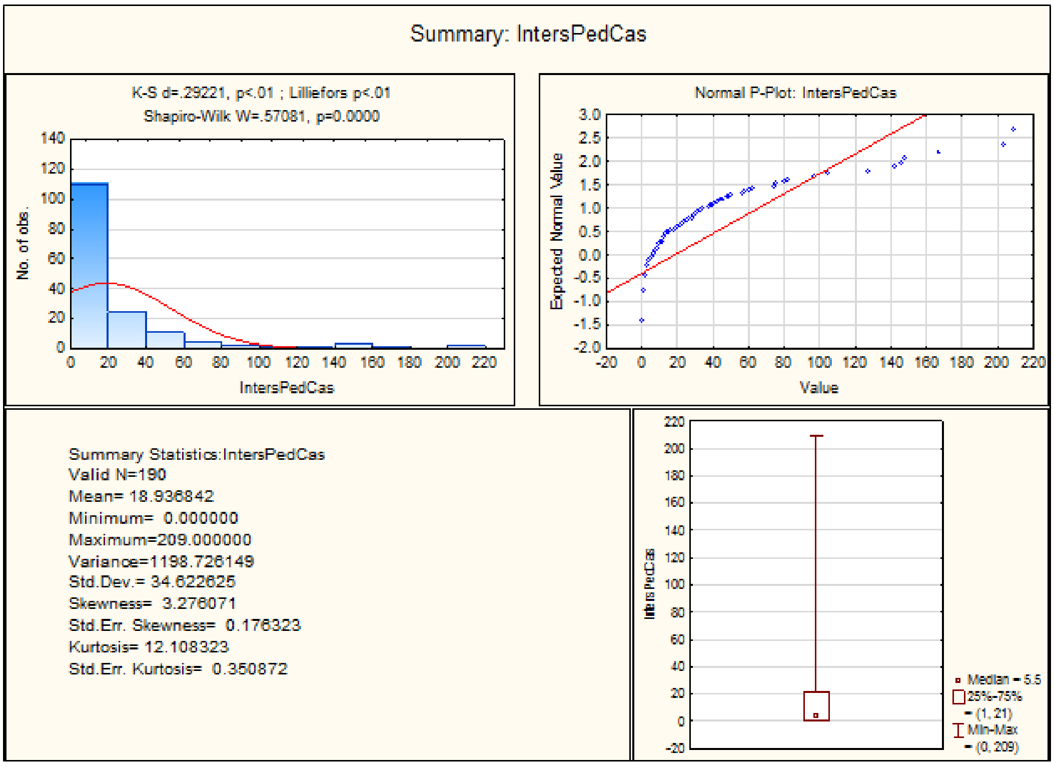


Figure 4-73: Summary statistics for intersection-related pedestrian casualties

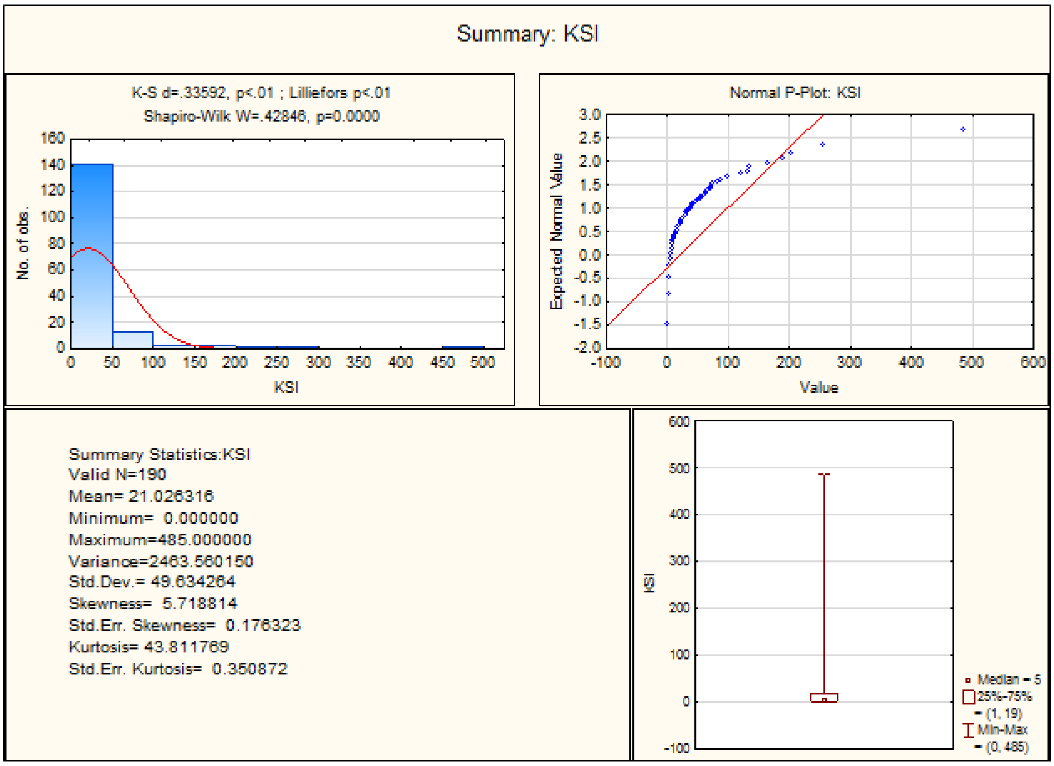


Figure 4-74: Summary statistics for Killed and Seriously Injured (KSI) pedestrian casualties

4.4.2 Description of explanatory variables

Table 4-35 presents summary statistics of explanatory variables tested in the modelling procedures. In total, 42 candidate explanatory variables were included in this process. Of these variables, 10 variables describe socio-demographic characteristics (i.e. population number, age, and ethnicity) and a further 10 variables reflect the socio-economic status (education level, employment, income level, and dwelling type) of the population in the study area. The variables describing the attribute of the built environment include three types of entropy scores (coded as ENT_AllCat, ENT_9Cat and ENT_4Cat) for land use mix; seven variables relating to land use patterns; three variables defining urban design (coded as Ratio_inters-cds, Inters_grt3leg and StrDens); and eight variables relate to elements of the transportation system in the study area.

Table 4-35: Descriptive statistics of explanatory variables

Variable	Definition	Mean	Min	Max	S.D.
Log_Popu	Logarithm of the population	3.698	0.000	5.593	0.893
Prop_Black	Proportion of the Black population	26.102	0.000	99.285	27.164
Prop_Coloured	Proportion of the Coloured population	34.709	0.000	97.323	31.489
Prop_Asian	Proportion of the Asian population	2.098	0.000	21.985	3.001
Prop_White	Proportion of the White population	32.721	0.000	96.518	32.676
Prop_Other	Proportion of other population	2.792	0.000	22.310	3.149
Prop_AgeLess15	Proportion of children (younger than 15 years)	19.570	0.000	31.299	7.566
Prop_Age15_24	Proportion of young adults (15-24 years)	17.425	0.000	97.028	10.081
Prop_Age25_54	Proportion of middle-age (25-54 years)	44.927	0.000	81.818	9.888
Prop_Age55_plus	Proportion of elderly (55 year and older)	16.499	0.000	54.475	10.405
Prop_LowEd	Proportion of the population with low education level	10.977	0.000	72.644	11.331
Prop_AvgEd	Proportion of the population with average education level	57.037	0.000	100.000	18.666
Prop_HighEd	Proportion of the population with higher education level	29.355	0.000	86.957	22.741
Prop_NotWork	Proportion of non-workers	16.063	0.000	80.144	14.974
Prop_Work	Proportion of workers	81.305	0.000	100.000	19.920
Prop_LowInc	Proportion of the population with low income level	43.325	0.000	100.000	27.619
Prop_MidInc	Proportion of the population with middle income level	31.060	0.000	100.000	17.265
Prop_UpperInc	Proportion of the population with upper income level	22.457	0.000	78.947	18.768
Prop_FormalDwe	Proportion of formal dwellings	84.491	0.000	100.000	26.553
Prop_InformalDwe	Proportion of informal dwellings	9.979	0.000	99.085	20.137
ENT_AllCat	Entropy score measured by the use of 34 land use types	0.354	0.000	0.677	0.161
ENT_9Cat	Entropy score measured by the use of 9 land use types	0.484	0.000	0.876	0.225
ENT_4Cat	Entropy score measured by the use of 4 main land use types	0.563	0.000	0.979	0.270
Prop_SR9Cat	Proportion of single residential use	29.003	0.000	89.017	24.220
Prop_GR9Cat	Proportion of general residential use	7.439	0.000	64.119	10.745
Prop_CO9Cat	Proportion of community use	8.837	0.000	100.000	17.084
Prop_GI9Cat	Proportion of general industrial use	5.786	0.000	97.462	16.814
Prop_UT.TR9Cat	Proportion of utility and transport use	7.474	0.000	100.000	20.486
Prop_OS9Cat	Proportion of open space use	15.775	0.000	94.840	17.356
Prop_AG.RU.LU9Cat	Proportion of agricultural, rural and limited use	20.742	0.000	100.000	30.347
Ratio_inters-cds	Ratio of intersections to culs-de-sac	4.282	0.469	16.833	2.878
Inters_grt3leg	Number of intersections with more than 3 legs	60.311	0.000	583.000	92.606
StrDens	Street density	13.742	0.620	31.813	6.277
Prop_Freeways	Proportion of freeway roads	5.554	0.000	46.617	8.422
Prop_Expressways	Proportion of expressway roads	2.017	0.000	44.165	4.449
Prop_PrimaryArter	Proportion of primary arterial roads	4.515	0.000	33.953	5.986
Prop_SecondArter	Proportion of secondary arterial roads	8.096	0.000	59.120	8.401
Prop_LocalDistr	Proportion of local distributor roads	2.178	0.000	43.409	5.033
Prop_LocalStr	Proportion of local streets	77.640	35.363	100.000	12.626
Round_Circ	Number of roundabouts and mini-circles	3.416	0.000	38.000	6.042
Prop_Signal	Proportion of signalised intersections	2.813	0.000	23.011	3.174

4.4.3 Results from Generalised Linear Models

4.4.3.1 GLM model performance: Goodness-of-fit measures

Table 4-36 summarises the goodness-of-fit measures of the two forms of Generalised Linear Model (GLM) developed in this study, the Poisson Regression model and the Negative Binomial Regression model (NB model). A comparison of the performance of the Poisson regression model and the NB model shows that the latter model has always produced lower values of different types of residuals (e.g. deviance, scaled deviance, Pearson Chi², scaled Pearson Chi² and likelihood residuals) than those produced by the Poisson regression model (see Table 4-36). The three Poisson regression models generated values of the ratio of the deviance to the degree of freedom (Dev/Df) which are 15 times higher, 4.9 times higher and 5.2 times higher than those generated by the NB model, for Model 1, Model 2 and Model 3, respectively. As seen in Table 4-36, the goodness-of-fit measures demonstrate that the Negative Binomial model performed better than the Poisson Regression model in fitting pedestrian casualty data.

Table 4-36: Model comparisons using goodness-of-fit measures

	Poisson Regression Models			Negative Binomial Regression Models		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Degree of freedom (Df)	172	172	175	172	172	175
Deviance (Dev)	3250.1874	1000.1563	1115.2777	215.711	204.2907	214.9342
Dev/Df	18.8964	5.8149	6.373	1.2541	1.1877	1.2282
Scaled Deviance	3250.1874	1000.1563	1115.2777	215.711	204.2907	214.9342
Scaled Deviance/Df	18.8964	5.8149	6.373	1.2541	1.1877	1.2282
Pearson Chi²	4465.6003	1171.766	1319.5221	198.6804	230.333	245.4777
Pearson Chi²/Df	25.9628	6.8126	7.5401	1.1551	1.3391	1.4027
Scaled P. Chi²	4465.6003	1171.766	1319.5221	198.6804	230.333	245.4777
Scaled P. Chi²/Df	25.9628	6.8126	7.5401	1.1551	1.3391	1.4027
AIC	4181.6336	1673.1636	1782.989	1618.3324	1151.243	1148.8588
AICc	4185.6336	1677.1636	1785.7476	1622.803	1155.7136	1152.0033
BIC	4240.0801	1731.61	1831.6943	1680.0258	1212.9365	1200.8112
Loglikelihood	-2072.8168	-818.5818	-876.4945	-790.1662	-556.6215	-558.4294

Following the assessment of the performance of the two GLM models, the analysis and discussion presented in this section is based on the results generated by the Negative Binomial models and the Geographically Weighted Regression models.

4.4.3.2 Parameter estimates from the Negative Binomial Regression Model 1

Table 4-37 presents the parameter estimates for the best Negative Binomial model developed based on the entire sample of pedestrian casualties. The model comprises 17 explanatory variables that were shown to have a significant effect on the frequency of pedestrian casualties. The sign and the magnitude of the coefficient estimates (B) are indicative of the effect each explanatory variable has on the outcome variable (i.e. the number of pedestrian casualties). A coefficient B with a positive sign implies that the variable is associated with an increase in the number of pedestrian casualties while the variable with a negative sign is associated with a decreased number of pedestrian casualties. To make the interpretation of the coefficients (B) more straightforward, the exponentiated coefficient "Exp (B)" was calculated in Excel spreadsheets and added to the estimate output generated by STATISTICA software tool. All variables in the final model are significant at the 5% level. However, the parameter estimates of the intercept are not statistically significant (p>0.05). Values marked in blue in Table 4-37 are not statistically significant at the 5% level.

Table 4-37: Model estimates for NB Model 1

Variables	AllPedCas - Parameter estimates Distribution : NEGATIVE BINOMIAL Link function: LOG						
	Estimate (B)	Standard Error	Wald Stat.	Lower CL 95.0%	Upper CL 95.0%	p	Exp(B)
Intercept	0.2826	0.4518	0.3912	-0.6029	1.1682	0.5317	1.3266
Log_Popu	1.3675	0.1438	90.4719	1.0857	1.6492	0.0000	3.9253
Prop_White	-0.0132	0.0032	17.1956	-0.0195	-0.0070	0.0000	0.9869
Prop_AgeLess15	-0.0426	0.0124	11.8643	-0.0669	-0.0184	0.0006	0.9583
Prop_Age15_24	-0.0405	0.0076	28.4604	-0.0554	-0.0257	0.0000	0.9603
Prop_Age25_54	-0.0222	0.0074	9.1091	-0.0367	-0.0078	0.0025	0.9780
Prop_AvgEd	-0.0129	0.0051	6.3629	-0.0229	-0.0029	0.0117	0.9872
Prop_UpperInc	-0.0204	0.0064	10.3044	-0.0328	-0.0079	0.0013	0.9798
ENT_9Cat	1.1577	0.3595	10.3721	0.4532	1.8623	0.0013	3.1826
Prop_GI9Cat	0.0242	0.0041	35.0876	0.0162	0.0323	0.0000	1.0245
Inters_grt3leg	0.0030	0.0008	12.2772	0.0013	0.0046	0.0005	1.0030
StrDens	0.0216	0.0108	4.0305	0.0005	0.0427	0.0447	1.0218
Prop_Freeways	0.0323	0.0072	19.9158	0.0181	0.0464	0.0000	1.0328
Prop_Expresways	0.0585	0.0132	19.6145	0.0326	0.0844	0.0000	1.0603
Prop_PrimaryArter	0.0239	0.0103	5.3240	0.0036	0.0441	0.0210	1.0241
Prop_SecondArter	0.0160	0.0072	4.8441	0.0017	0.0302	0.0277	1.0161
Round_Circ	0.0352	0.0106	11.0020	0.0144	0.0560	0.0009	1.0358
Prop_Signal	0.0729	0.0218	11.1239	0.0300	0.1157	0.0009	1.0756
Dispersion	0.4426	0.0534		0.3380	0.5472		

Marked values are not statistically significant at 95% confidence interval (i.e. p>0.05)

The top five variables with the highest absolute values of the coefficient "B" are (1) the log of population (B=1.3675); (2) the entropy index (B=1.1577); (3) the proportion of signalised intersection (B=0.0729); (4) the proportion of expressways (B=0.0585) and (5) the proportion of the population younger than 15 years old (B= -0.0426).

The NB Model 1 includes 11 explanatory variables with positive associations with the number of pedestrian casualties. The coefficient estimates of each of the 11 variables can be interpreted using the exponentiated coefficients "Exp (B)" presented in Table 4-37. The parameter "Exp (B)" tells that, if other explanatory variables in the model are held constant:

- An increase of one unit in Log of population (i.e. a 10-fold increase in population number) would result in an increase in the number of pedestrian casualties by a factor of 3.925;
- An increase of one unit in entropy index measured using nine land-use categories ("ENT_9Cat") would get the number of pedestrian casualties increased by a factor of 3.183;
- A one percent increase in the number of signalised intersections would result in an increase of 7.56 percent of the total number of pedestrian casualties;
- An increase of one percent in the proportion of expressways would cause an increase of 6.03 percent in the number of pedestrian casualties;
- An addition of one extra roundabout or mini-circle to the total number of these facilities would increase the number of pedestrian casualties by 3.58 percent;
- An increase of one percent in the proportion of freeway facilities would result in an increase of 3.28 percent in the number pedestrian casualties;
- An increase of one percent in the proportion of primary arterial roads would contribute to an increase of 2.41 percent in the number of pedestrian casualties;
- An increase of one percent in the proportion of secondary arterial roads would get the number of pedestrian casualties increased by 1.61 percent;
- A one unit in the proportion of land used for industrial purposes (i.e. General Industry land use) would be associated with an increase of 2.45 percent in the number of pedestrian casualties;
- Increasing street density by one unit (i.e. one kilometre road per square kilometre of land area) would result in 2.18 percent increase in the number of pedestrian casualties;

- An addition of one four- or multi-legged intersection would result in 0.3 percent increase in the number of pedestrian casualties.

The model results point out six demographic variables which are negatively associated with the number of pedestrian casualties. The coefficient estimates of these variables explain that:

- An increase of one percent in the proportion of the population with an average education level (i.e. some secondary studies and Grade 12) would reduce the number of pedestrian casualties by 1.28 percent;
- A one percent increase in the proportion of the White population would be associated with a decrease of 1.31 percent in the number of pedestrian casualties;
- An increase of one percent in the proportion of the population with upper income (i.e. earning a monthly income higher than R25,601) would lead to a reduction of 2.02 percent in the number of pedestrian casualties;
- An increase of one percent in the proportion of the population in the 15-24 age range would result in 3.97 percent decrease in the number of pedestrian casualties;
- A one percent increase in the proportion of the population in the 25-54 age range would contribute to a decrease of 2.2 percent in the number of pedestrian casualties;
- An increase of one percent in the proportion of the population younger than 15 years old would get the number of pedestrian casualties decreased by 4.17 percent.

4.4.3.3 Parameter estimates from Negative Binomial Regression Model 2

Table 4-38 presents the parameter estimates for the NB model developed based on the dataset of intersection-related pedestrian casualties. This model is referred to as NB Model 2. Similar to NB Model 1, 17 explanatory variables that are included in NB Model 2 are all statistically significant (i.e. $p<0.05$). The top five variables with the highest absolute values of the coefficient “B” are: (1) the log of population (B=1.9470); (2) the entropy index (B=1.6102); (3) the proportion of signalised intersections (B=0.1234); (4) the proportion of expressways (B=0.0795); and (5) the proportion of freeways (B=0.0692). Of the 17 explanatory variables of NB Model 2, 11 variables have positive associations with the number of intersection-related pedestrian casualties while six demographic variables are shown to have negative associations with the number of intersection-related pedestrian casualties.

Table 4-38: Model estimates for NB Model 2

Variables	IntersPedCas - Parameter estimates Distribution : NEGATIVE BINOMIAL Link function: LOG						
	Estimate (B)	Standard Error	Wald Stat.	Lower CL 95.0%	Upper CL 95.0%	p	Exp(B)
Intercept	-5.0606	1.5921	10.1029	-8.1811	-1.9401	0.0015	0.0063
Log_Popu	1.9470	0.1510	166.2123	1.6510	2.2430	0.0000	7.0074
Prop_AgeLess15	-0.0460	0.0158	8.4439	-0.0771	-0.0150	0.0037	0.9550
Prop_Age15_24	-0.0562	0.0103	29.6206	-0.0765	-0.0360	0.0000	0.9453
Prop_Age55_plus	-0.0277	0.0093	8.8287	-0.0460	-0.0094	0.0030	0.9727
Prop_AvgEd	-0.0245	0.0056	19.3274	-0.0354	-0.0136	0.0000	0.9758
Prop_NotWork	-0.0350	0.0072	23.5936	-0.0491	-0.0209	0.0000	0.9656
Prop_UpperInc	-0.0573	0.0063	82.5747	-0.0696	-0.0449	0.0000	0.9444
ENT_9Cat	1.6102	0.3862	17.3878	0.8534	2.3671	0.0000	5.0038
Prop_GI9Cat	0.0245	0.0045	29.9836	0.0157	0.0333	0.0000	1.0248
Ratio_inters-cds	0.0529	0.0225	5.5256	0.0088	0.0969	0.0187	1.0543
Prop_Freeways	0.0692	0.0165	17.6544	0.0369	0.1014	0.0000	1.0716
Prop_Expressways	0.0795	0.0200	15.7909	0.0403	0.1188	0.0001	1.0828
Prop_PrimaryArter	0.0533	0.0189	7.9638	0.0163	0.0903	0.0048	1.0547
Prop_SecondArter	0.0365	0.0165	4.8580	0.0040	0.0689	0.0275	1.0371
Prop_LocalStr	0.0330	0.0155	4.5344	0.0026	0.0634	0.0332	1.0336
Round_Circ	0.0297	0.0105	8.0416	0.0092	0.0503	0.0046	1.0302
Prop_Signal	0.1234	0.0231	28.4753	0.0781	0.1688	0.0000	1.1314
Dispersion	0.4108	0.0615		0.2902	0.5314		

Using the exponentiated coefficients "Exp (B)" presented in Table 4-38, the coefficient estimates "B" of the 11 explanatory variables (with positive associations with the frequency of pedestrian casualties) suggest that, when other explanatory variables in NB Model 2 are held constant:

- A one unit increase in Log of population would result get the number pedestrian casualties increased by a factor of 7.00;
- An increase of one unit in entropy index measured using nine land-use categories ("ENT_9Cat") would result in an increase in the number of pedestrian casualties by a factor of 5.00;
- Increasing the proportion of signalised intersections by one percent would result in 13.14 percent increase in the number of pedestrian casualties;
- A one percent increase in the proportion of expressways would result in 8.28 percent increase in the number of pedestrian casualties;

- Elevating the proportion of freeways by one percent would contribute to an increase of 7.16 percent in the number of pedestrian casualties;
- An increase of one percent in the proportion of primary arterial roads would lead to an increase of 5.47 percent in the number of pedestrian casualties;
- An increase of one percent in the proportion of secondary arterial roads would result in 3.71 percent increase in the number of pedestrian casualties;
- Increasing the proportion of local streets by one percent would contribute to a rise of 3.36 percent in the number of pedestrian casualties;
- A one unit increase in the ratio of intersections to cul-de-sacs would result in 5.43 percent increase in the number of pedestrian casualties;
- One additional roundabout or mini-circle to the total number of these facilities would result in an increase of 3.02 percent in the number of pedestrian casualties;
- An increase of one percent in the proportion of land zoned as General Industry (GI) would result in 2.48 percent increase in the number of pedestrian casualties.

With respect to the six explanatory variables shown to have negative relationships with the frequency of intersection-related pedestrian casualties, their parameter estimates indicate that, while holding the other explanatory variables constant in the model:

- An increase of one percent in the proportion of the population with upper income would contribute to a decrease of 5.56 percent in the number of pedestrian casualties;
- An increase of one percent in the proportion of the population in the 15-24 age group would result in a decrease of 5.47 percent in the number of pedestrian casualties;
- A one percent increase in the proportion of the population younger than 15 years old would reduce the number of pedestrian casualties by 4.50 percent;
- A one percent increase in the proportion of the population classified as not in the labour force (i.e. unemployed and discouraged workers) would result in 3.44 percent decrease in the number of pedestrian casualties;
- A one percent increase in the proportion of the population aged 55 years and older would result in 2.73 percent decrease in the number of pedestrian casualties;
- An increase of one percent in the proportion of the population with an average education level would contribute to a decrease of 2.4 percent in the number of pedestrian casualties.

4.4.3.4 Parameter estimates from Negative Binomial Model 3

The parameter estimates for Negative Binomial Model 3 (NB Model 3) fitted to KSI pedestrian casualties are described in Table 4-39. NB Model 3 comprises 14 explanatory variables shown to be statistically significant. The top five variables which show a prevailing influence on the number of KSI pedestrian casualties are: (1) the log of population (B=1.5511); (2) the entropy index (B=0.8852); (3) the proportion of signalised intersections (B=0.0829); (4) the proportion of the population aged 55 years and older (B= -0.0496) and (5) the proportion of the population in the 15-24 age group (B= -0.0453). Of the 14 explanatory variables in NB Model 3, seven demographic variables emerge with negative values of the coefficient B while the other half of the variables are shown with positive values of the coefficient B.

Table 4-39: Model estimates for NB Model 3

Variables	KSI - Parameter estimates Distribution : NEGATIVE BINOMIAL Link function: LOG						
	Estimate (B)	Standard Error	Wald Stat.	Lower CL 95.0%	Upper CL 95.0%	p	Exp(B)
Intercept	-0.2856	0.4791	0.3555	-1.2246	0.6533	0.5510	0.7515
Log_Popu	1.5511	0.1687	84.5105	1.2204	1.8818	0.0000	4.7166
Prop_AgeLess15	-0.0354	0.0151	5.5136	-0.0650	-0.0059	0.0189	0.9652
Prop_Age15_24	-0.0453	0.0100	20.4844	-0.0649	-0.0257	0.0000	0.9557
Prop_Age55_plus	-0.0496	0.0092	28.8671	-0.0677	-0.0315	0.0000	0.9516
Prop_AvgEd	-0.0159	0.0061	6.7725	-0.0280	-0.0039	0.0093	0.9842
Prop_NotWork	-0.0181	0.0072	6.3818	-0.0322	-0.0041	0.0115	0.9821
Prop_MidInc	-0.0129	0.0049	6.9497	-0.0225	-0.0033	0.0084	0.9872
Prop_UpperInc	-0.0418	0.0064	42.2392	-0.0544	-0.0292	0.0000	0.9591
ENT_9Cat	0.8852	0.3583	6.1022	0.1829	1.5875	0.0135	2.4235
Prop_GI9Cat	0.0186	0.0041	20.2964	0.0105	0.0267	0.0000	1.0188
Inters_grt3leg	0.0032	0.0007	18.7940	0.0018	0.0047	0.0000	1.0032
Prop_Freeways	0.0345	0.0077	19.9692	0.0193	0.0496	0.0000	1.0351
Prop_Expresways	0.0337	0.0133	6.3744	0.0075	0.0598	0.0116	1.0342
Prop_Signal	0.0829	0.0212	15.3167	0.0414	0.1244	0.0001	1.0864
Dispersion	0.3975	0.0614		0.2772	0.5178		

Marked values are not statistically significant at 95% confidence interval (i.e. p>0.05)

Using the exponentiated coefficients “Exp (B)”, the parameter estimates with positive values indicate that, when the other variables in NB Model 3 are held constant:

- A one unit increase in Log of population would increase the number of KSI pedestrian casualties by a factor of 4.72;
- A one unit increase in entropy index measured using nine land-use categories (“ENT_9Cat”) would get the number of KSI pedestrian casualties increased by a factor of 2.42;
- Raising the proportion of signalised intersections by one percent would contribute to an increase of 8.64 percent in the number of KSI pedestrian casualties;
- An increase of one percent in the proportion of freeways would result in 3.51 percent increase in the number of KSI pedestrian casualties;
- An increase of one percent in the proportion of expressways would contribute to a rise of 3.42 in the number of KSI pedestrian casualties;
- A one unit increase in the proportion of the general industrial use (GI) would be associated with an increase of 1.88 percent in the number of KSI pedestrian casualties;
- An addition of one extra four-or multi-legged intersection would be associated with an increase of 0.32 percent in the number of KSI pedestrian casualties.

The model results for the explanatory variables with negative values of the coefficient B suggest that, while holding the other variables constant in NB Model 3:

- A one percent increase in the proportion of the population aged 55 years and older would contribute to 4.84 percent decrease in the number of KSI pedestrian casualties;
- An increase of one percent in the proportion of the population in the 15-24 age group would result in a decrease of 4.43 percent in the number of KSI pedestrian casualties;
- An increase of one percent in the proportion of the population with upper income would be associated with a reduction of 4.09 percent in the number of KSI pedestrian casualties;
- An increase of one percent in the proportion of the population younger than 15 years old would result in 3.48 percent decrease in the number of KSI pedestrian casualties;
- A one percent increase in the proportion of the population classified as not in the labour force (i.e. unemployed and discouraged population) would result in 1.79 percent decrease in the number of KSI pedestrian casualties;

- A one percent increase in the proportion of the population with an average education level would be associated with a decrease of 1.58 percent in the number of KSI pedestrian casualties;
- A rise of one percent in the proportion of the population with middle income (earning a monthly income ranging from R6 401 to R25 600) would result in a reduction of 1.28 percent in the number of KSI pedestrian casualties.

4.4.3.5 Sensitivity of variables over different days of a week

In addition to the three models developed in this study (i.e. Model 1, Model 2 and Model 3), three additional models were included in the analysis to test the sensitivity of parameter estimates over different days of the week. One model was developed for pedestrian casualties that occurred on weekdays (Model 4). Another model was developed for Saturday pedestrian casualties (Model 5) and the last one was developed for Sunday pedestrian casualties (Model 6). Public holidays were excluded from this analysis since exposure on holidays may differ depending on the holiday being celebrated. For instance, exposure over long weekends may differ from that of a single public holiday, as long weekends are associated with increased trip distances (inter- and intra-provincial travels) and an increase in social and recreational activities involving travelling in rural and unfamiliar road environments.

The three additional models were developed using the Negative Binomial regression modelling and with reference to the 17 explanatory variables of NB Model 1, to enable a better basis for the comparison of parameter estimates. The coefficients B for NB Model 4, NB Model 5 and NB Model 6 are demonstrated in Table 4-40 and the full outputs for the three models are provided in APPENDIX F. The estimates in Table 4-40 with red font colour are found to be not statistically significant ($p>0.05$). The rows highlighted in blue illustrate variables with decreasing values of coefficient B, while those in yellow represent variables with increasing values of coefficient B.

Table 4-40: Estimates for NB Model 4, NB Model 5 and NB Model 6

Variables	Parameter estimates Distribution : NEGATIVE BINOMIAL Link function: LOG		
	NB Model 4 Estimate (B)	**NB Model 5** Estimate (B)	**NB Model 6** Estimate (B)
Intercept	-0.2127	-1.0284	-1.2010
Log_Popu	1.4715	1.4657	1.4108
Prop_White	-0.0113	-0.0116	-0.0203
Prop_AgeLess15	-0.0489	-0.0349	-0.0093
Prop_Age15_24	-0.0408	-0.0454	-0.0693
Prop_Age25_54	-0.0252	-0.0216	-0.0122
Prop_AvgEd	-0.0152	-0.0149	-0.0217
Prop_UpperInc	-0.0202	-0.0333	-0.0297
ENT_9Cat	1.0756	0.8049	0.6821
Prop_GI9Cat	0.0295	0.0171	0.0051
Inters_grt3leg	0.0027	0.0028	0.0031
StrDens	0.0279	0.0062	0.0076
Prop_Freeways	0.0357	0.0181	0.0352
Prop_Expressways	0.0607	0.0318	0.0224
Prop_PrimaryArter	0.0153	0.0161	0.0302
Prop_SecondArter	0.0099	0.0068	0.0188
Round_Circ	0.0368	0.0260	0.0195
Prop_Signal	0.0912	0.0808	0.0494
Dispersion	0.4197	0.3131	0.3274

Marked values are not statistically significant at 95% confidence interval (i.e. p>0.05)

The test for sensitivity of estimates over different days of the week was conducted by comparing the coefficients B of the three models- Model 4, Model 5 and Model 6. The sensitivity analysis intends to test possible mediating effects of exposure variables- pedestrian volumes, traffic volumes and vehicular speed- which are not captured in the analysis due to lack of data. Variables such as population, land use mix and road class are considered in this study as rough proxy variables for exposure. The number of population and land use mix are considered as rough proxy variables of pedestrian activity or pedestrian trip density, while road class is considered as a rough proxy of vehicular speed and traffic volumes. The sensitivity analysis is based on the hypothesis that exposure varies over different days of the week. For instance, research in South Africa has demonstrated that traffic volumes on weekdays are higher than those on weekends (Jongh & Bruwer, 2017; Sampson, 2017). However, with respect to pedestrian volumes, not much is known about temporal variation in pedestrian activity in South Africa.

Although not based on research evidence, one would presuppose that pedestrian activity or pedestrian trip density is reduced on Saturdays and that the lowest activity is expected on Sundays. Reduced pedestrian activity on Saturdays may be justified by decreased density of both commuting trips and walking trips to access services (e.g. schools, governmental services etc.). In addition to the reduced number of trips for commuting and accessing services, a decrease in the number of shopping trips would contribute to further reductions in pedestrian activity on Sundays. Many shops in South Africa are closed and some are opened for limited hours on Sundays. Therefore, the expectation that Sunday is the day of week with minimum pedestrian activity may be a reasonable claim. However, it is worth noting that trends in pedestrian activity may also depend on the location. For example, places of leisure might experience more pedestrian activity on weekends than weekdays (e.g. beaches; shopping malls such as Canal Walk Shopping Centre, Waterfront and to name a few; theme parks and water amusement parks, etc.) while the situation for places of business (e.g. industrial areas, CBD) would tend to be the inverse. Therefore, the magnitude of the contribution (measured by the coefficient B) of proxy variables of exposure (e.g. land-use mix, road class, population) to the number of pedestrian casualties is expected to vary over different days of week and across different locations within the study area.

The estimates (coefficient B) of the three models developed for the test of sensitivity are presented in Table 4-40. A decreasing contribution to the outcome variable across the three models (from Model 4 to Model 6) is apparent for 10 explanatory variables (variables colour-coded in blue in Table 4-40). These variables are:

- The proportion of the White population
- The proportion of the population in the 15-24 age group
- The proportion of the population with an average education level
- The proportion of the population with upper income
- Entropy index
- The proportion of land zoned as General Industry (GI)
- Street density
- The proportion of expressways
- The number of roundabouts and mini-circles and
- The proportion of signalised intersections.

The model results clearly show that the highest influence of the aforementioned variables on the number of pedestrian casualties is expected on weekdays. Reduced influence is expected on Saturdays and the least contribution of these variables to the pedestrian casualty frequency is expected on Sundays. The entropy index which reflects the extent to which land use types are mixed emerged among these variables. The varying influence of entropy index suggests that pedestrian casualties mainly related to pedestrian activity are more likely to occur on weekdays than on weekends. In a similar way, the results indicate that pedestrian casualties mostly influenced by population characteristics such the number of the population with upper income, average education level, and in the 15-24 age group are more likely to occur on weekdays than weekends. The reason for this finding could be reduced trip density among commuters (who earn more income and are more likely to be educated than unemployed and discouraged workers) and young adults (mostly learners and young workers) over weekends. The same reason could explain why industrial use is more associated with pedestrian casualties on weekdays than weekends: industrial activity is usually more intense on weekdays than weekend. With respect to roadway characteristics that are concerned by this analysis, temporal variations in traffic volume could be the reason for varying influence of roadway characteristics on pedestrian casualty frequency.

In contrast, four variables (colour-coded in yellow) have an increasing influence on pedestrian casualty counts across the days of week, with greater influences being shown during the weekends. These variables include:

- The proportion of the population younger than 15 years
- The proportion of the population in the 25-54 age group
- The number of four- and multi-legged intersections and
- The proportion of primary arterial roads.

These findings explain that the associations between pedestrian casualty frequency and two age groups (population in the 25-54 age group and those younger than 15 years) are more pronounced on weekends than weekdays. The explanation for this could be that children younger than 15 years who are predominantly school-age children are free from school on weekends and consequently, the level of children’s activity such as playing and wandering off the street is greater. Moreover, alcohol involvement may play an important role in greater associations between pedestrian casualty frequency and the number of people in the 25-54 age range on weekends. For instance, a recent study conducted in South Africa has reported that the prevalence of binge drinking is highest among males and females aged 25-34 years (Vellios & Van Walbeek, 2017). It is also documented that rates of heavy alcohol consumption in South Africa are higher on weekends than on weekdays (Econometrix, 2013; National Department of Health, 2007).

Alcohol involvement may also be the possible reason for greater relationships between pedestrian casualty frequency and roadway characteristics such as four- or multi-legged intersections and arterial roads. Although traffic volume is recognised as the main determinant in pedestrian crashes, evidence has emerged in several studies that in certain circumstances low traffic is associated with increased number of pedestrian crashes (Yao *et al.*, 2015). The main reason that could explain this relationship may be reduced pedestrians' attention when traffic is low, which may lead to collisions with motorists (Harrell, 1991; Yao *et al.*, 2015). This might also be another possible explanation of stronger relationships between pedestrian casualty frequency and the two roadway factors (four- or multi-legged intersections and arterial roads).

4.4.4 Model results from Geographically Weighted Regression (GWR) models

As with the Generalised linear models (GLM) used in this study, six geographically weighted regression (GWR) models were developed to investigate associations between the aspects of the built environment and pedestrian casualty counts at the census suburb level. The specific advantage of the GWR modelling procedure over the GLM is the ability to capture non-stationary or spatially varying relationships between pedestrian casualty counts and the explanatory variables representing the built environments. Six models were developed using the GWR tool implemented in ArcMap 10.3.1 software. The analysis with the GWR models is intended to respond to the second and the third research questions investigated in this study (i.e. "If the link between pedestrian crashes and the built environment exists, what is its extent and how is it spatially distributed?")

4.4.4.1 Parameter estimates for GWR Models

Parameter estimates in the GWR model are allowed to vary spatially and this ability allows for an exploration of local variations of a range of parameter estimates across the study area. With the GWR models, local parameter estimates were produced for each census suburb and these include the (1) condition number; (2) the local R^2; (3) predicted values (i.e. predicted pedestrian casualty count for the context of this study); (4) intercept.; (5) coefficients of explanatory variables; (6) residuals; (7) standard error; (8) coefficients of standard error; (9) and standardised residuals.

Each GWR model comprises 190 sets of parameter estimates and each explanatory variable included in the model has 190 coefficients corresponding to the 190 census suburbs of the City of Cape Town. A model with lower value of the Corrected Akaike's Information (AICc) was selected to study the relationship between the explanatory variables and the outcome variable. Three outcome variables analysed using GWR models include: (1) all pedestrian casualties (GWR Model 1); (2) intersection-related pedestrian casualties (GWR Model 2); (3) and killed and seriously injured (KSI) pedestrian casualties (GWR model 3).

For each outcome variable, two GWR models were developed to allow for a greater range of explanatory variables as some important variables of the built environment could not all be included in a single model because of the issue of spatial correlation among variables (i.e. multicollinearity issue). The multicollinearity was assessed by the variance inflation factor (VIF). For each outcome variable, the model with the best goodness of fit measure (i.e. the lowest value of AICc) was selected and denoted by the letter "A". The second model (denoted by the letter "B") was selected based on a trade-off between goodness of fit measures and inclusion of explanatory variables of interest in the model (especially those describing urban form land use and transportation system) but which were not captured in the models A. In total, six GWR models were developed and parameters reflecting the goodness of fit measures for these six models are summarised in Table 4-41. By examining the values of AICc, it can be noticed that all Models A have minimum values of AICc, indicating that they perform better than Models B.

Table 4-41: Output of goodness of fit parameters for the GWR models

Indicators	GWR Model 1A	GWR Model 1B	GWR Model 2A	GWR Model 2B	GWR Model 3A	GWR Model 3B
Bandwidth	16671.88	801924.95	22753.21	273412.97	19897.58	18314.71
Residual Squares	469723.60	715569.02	47695.47	130105.28	119394.01	114275.91
Effective Number	32.23	15.02	22.25	16.21	19.14	33.89
Sigma	54.56	63.95	16.86	27.36	26.43	27.06
AICc	2082.07	2138.79	1628.99	1817.54	1797.06	1818.43
R^2	0.87	0.80	0.79	0.43	0.74	0.75
R^2Adjusted	0.85	0.79	0.76	0.38	0.72	0.70

1. **Explanatory variables for GWR Model 1**

Initially, the 11 explanatory variables included in the GLMs to model pedestrian casualty counts at the census suburb level were considered for the GWR Model 1. When there were issues of either global or local multicollinearity among variables in the model, ArcMap could not compute the results for the GWR model and an error "no 040038" that states that: "results cannot be computed because of severe model design problems" appeared. To address this issue, the Exploratory Regression tool implemented in ArcMap software was used as a starting point to select the best model with maximum performance indicators (e.g. Adjusted R^2, AICs) and minimum level of multicollinearity measured by the Variance Inflation Factor (VIF). Models with higher values of VIF were not selected since higher VIF values imply that two or more variables are redundant (i.e. may be telling the same story).

For the entire sample of pedestrian casualties, GWR Model 1A and GWR model 1B were selected for the analysis. The first model consists of seven explanatory variables and the second comprises 14 explanatory variables. Explanatory variables included in GWR Models 1 are summarized in Table 4-42.

Table 4-42: Explanatory variables for GWR models 1

Var No	GWR Model 1A	GWR Model 1B
1	Prop_Black	Log_Popu
2	Prop_AgeLess15	Prop_AgeLess15
3	Prop_MidInc	Prop_AvgEd
4	Prop_GB.MU9Cat	Prop_UpperInc
5	Inters_grt3leg	ENT_9Cat
6	Round_Circ	Prop_GI9Cat
7	Prop_Signal	Inters_grt3leg
8		StrDens
9		Prop_Freeways
10		Prop_Expressways
11		Prop_PrimaryArter
12		Prop_SecondArter
13		Round_Circ
14		Prop_Signal

2. Local parameter estimates for GWR Model 1

This section provides a description of local parameter estimates for the best GWR Model 1 based on the goodness of fit measure (i.e. GWR Model 1A). The GWR Model 1A was calibrated based on the explanatory variables presented in Table 4-42. The spatial distribution of local estimates for each explanatory variable included in GWR model 1A is illustrated in Figure 4-75 and Figure 4-76. On these figures, a pattern of spatial non-stationarity can be seen by examining variations in the coefficients of explanatory variables. Summary statistics of the coefficient values behind the choropleth maps (i.e. thematic maps patterned in proportion to values of variable displayed on the map) in Figure 4-75 and Figure 4-76 are provided in Table 4-43. Local estimates with a positive sign indicate that the corresponding explanatory variable is associated with increased number of pedestrian casualties whereas a negative sign implies that the variable is associated with a reduced number of pedestrian casualties.

The *t*-test indicates variables that show significant varying influence on the number of pedestrian casualties (i.e. spatial heterogeneity in the relationship between those variables and pedestrian casualty counts). Unlike the GLMs, the number of roundabouts and mini-circles is found to be associated with a decreased number of pedestrian casualties in GWR Model 1A. Contradictory findings also arise with respect to the influence of the proportion of population younger than 15 years. In GWR Model 1 A, this variable shows a positive influence on the number of pedestrian casualties while the opposite is found in the NB Model 1.

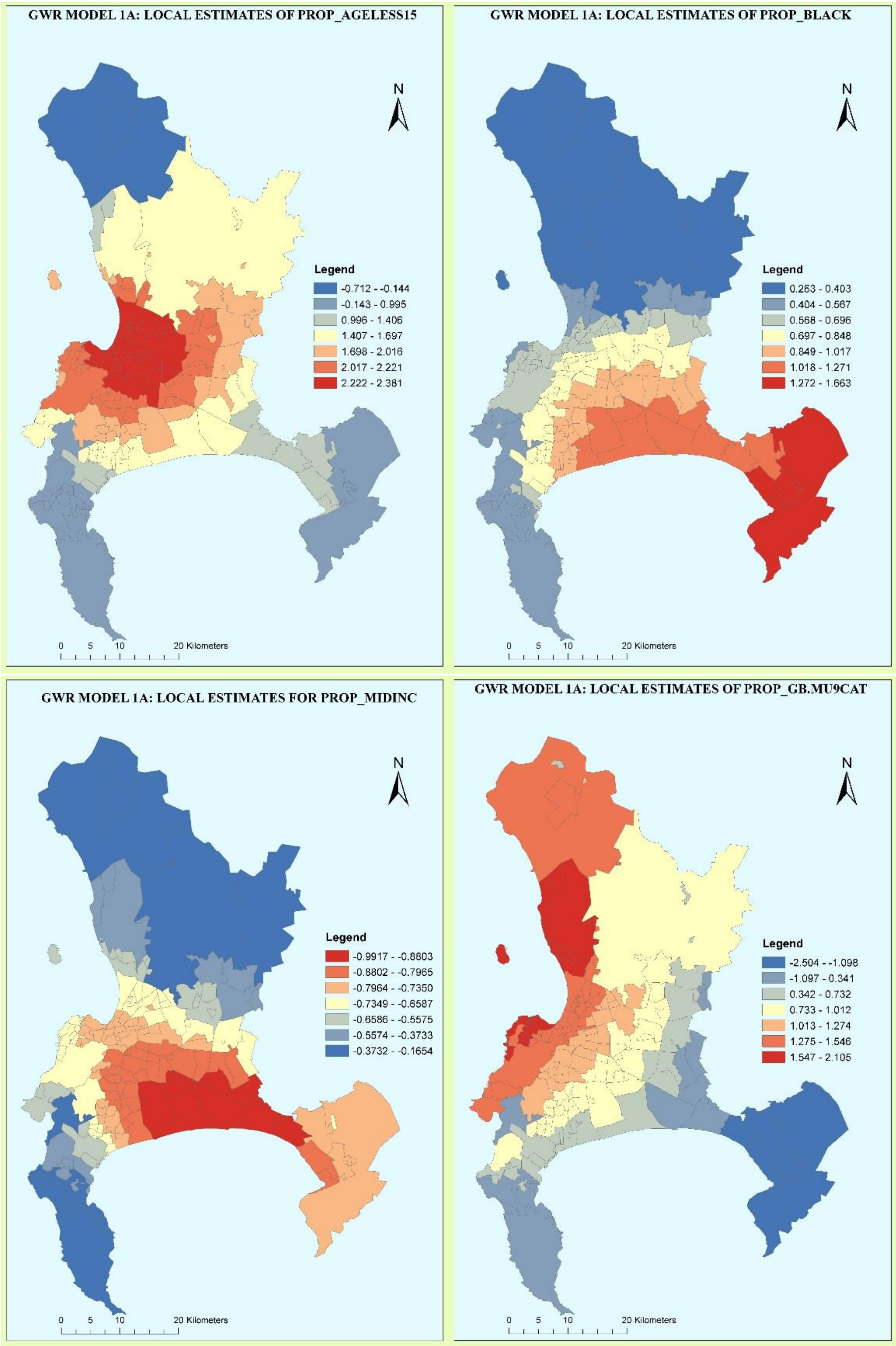


Figure 4-75: Local estimates for GWR Model 1A: (1) Prop_AgeLess15; (2) Prop_Black; (3)Prop_MidInc; (4) Prop_GB.MU9Cat

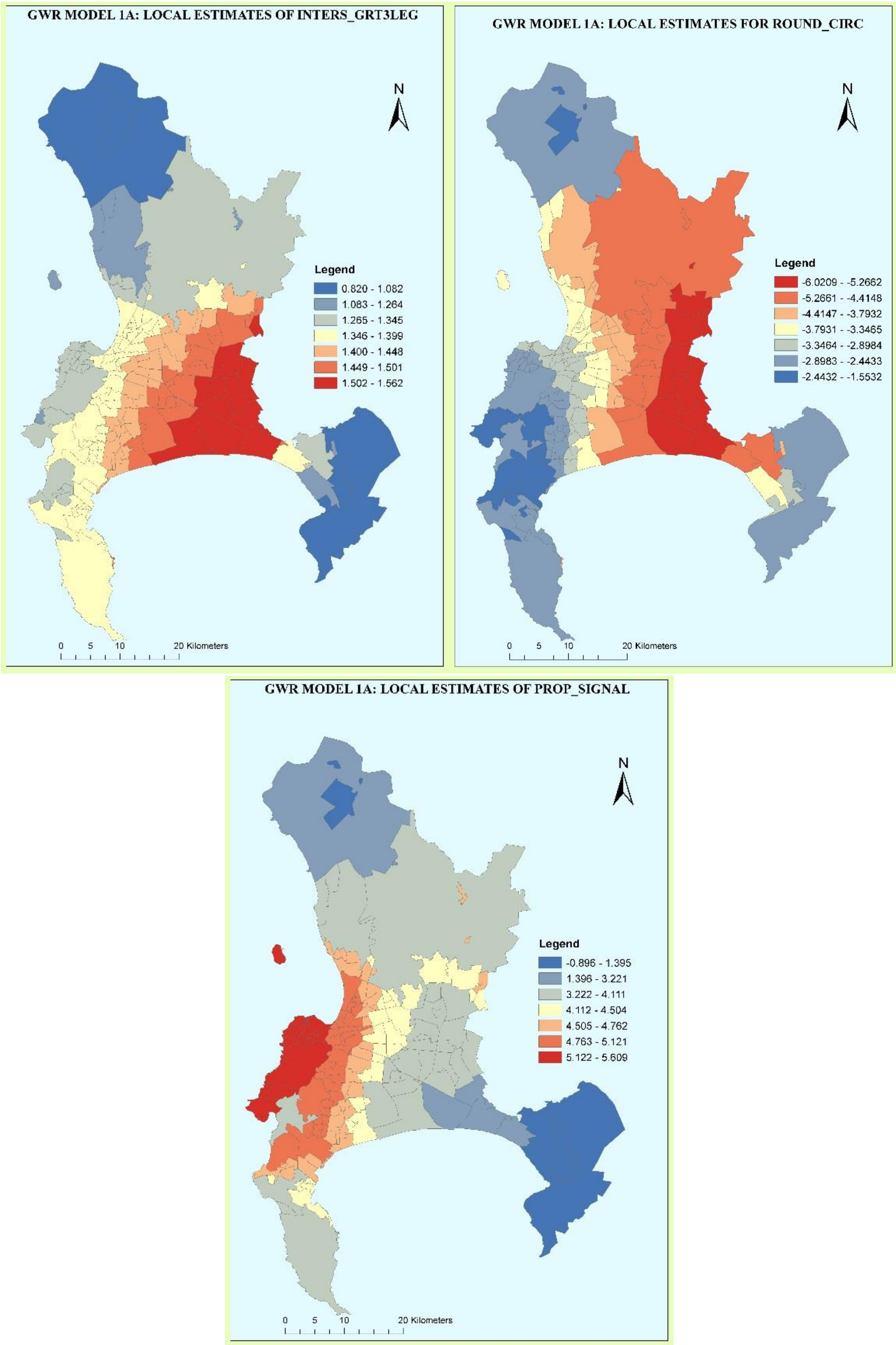


Figure 4-76: Local estimates for GWR Model 1A: (5) Inters_grt3leg; (6) Round_Circ; (7) Prop_Signal

The *t*-test computed to detect significant variations in local estimates demonstrates that five variables of the model are statistically significant at the 5% level (i.e. p<0.05). The proportion of the Black population and the proportion of signalised intersections are statistically non-significant (i.e. p>0.05), suggesting that the local coefficients for these two variables are not significantly different from those of the global model (OLS model). In other words, this means that there are no significant variations in the local estimates of these variables across the study area. The results of the *t*-test are presented in Table 4-43.

Table 4-43: Summary statistics of local estimates for GWR Model 1A

Local parameters	Summary statistics of local estimates for GWR Model 1A							
	Mean	Minimum	Maximum	Std.Dev.	Global Model	t-value	df	p
Coeff Intercept	-46.032	-58.573	16.840	13.236	-42.499	-3.679	189	0.000
Coeff Prop_Black	0.743	0.263	1.663	0.255	0.716	1.463	189	0.145
Coeff Prop_AgeLess15	1.796	-0.712	2.381	0.587	1.452	8.064	189	0.000
Coeff Prop_MidInc	-0.682	-0.992	-0.165	0.172	-0.634	-3.869	189	0.000
Coeff Prop_GB.MU9Cat	0.862	-2.504	2.105	0.786	1.227	-6.397	189	0.000
Coeff Inters_grt3leg	1.374	0.820	1.562	0.115	1.417	-5.057	189	0.000
Coeff Round_Circ	-3.497	-6.021	-1.553	1.053	-4.536	13.596	189	0.000
Coeff Prop_Signal	4.284	-0.896	5.609	1.131	4.384	-1.212	189	0.227
Local R^2	0.811	0.760	0.958	0.033	0.825	-5.722	189	0.000
Residual	-1.619	-145.540	221.686	49.826	0.000	-1.271	189	0.205
Std. Residual	-0.047	-5.873	5.792	1.117	0.000	-2.023	189	0.044

Marked values are not statistically significant at 95% confidence interval (i.e. p>0.05)

The spatial patterns of local estimates are described in reference to the eight planning districts of the City of Cape Town illustrated in Figure 2-4 on Page 42. The spatial patterns of local estimates illustrated in Figure 4-75 and Figure 4-76 demonstrate that:

- The influence of the proportion of the Black population on pedestrian casualties is greatest in the south eastern suburbs (Helderberg district);
- The influence of the proportion of the population younger than 15 years old on the pedestrian casualty frequency is most pronounced in the central parts of the City of Cape Town (City Bowl, Table bay and Tygerberg districts);
- The proportion of the population with a middle income level has the greatest negative influence on pedestrian casualty counts in Khayelitsha/Mitchells Plain and Cape Flats district In other words, the proportion of the population earning a middle income level is greatly associated with a reduced number of pedestrian casualties in these regions;

- The proportion of mixed use and general business use has a more pronounced influence on the number of pedestrian casualties in the western parts of the city (Blaauwberg and Table Bay districts);
- A positive relationship between the number of four- and multi-legged intersections and the frequency of pedestrian casualties is most significant in the south eastern parts of Cape Town (Khayelitsha/Mitchells Plain, Cape Flats and Tygerberg district);
- A negative relationship between the number of roundabouts/mini-circles and the frequency of pedestrian casualties is most marked in the eastern regions of Cape Town (Northern district, Tygerberg and Khayelitsha districts).

For GWR Model 1B, summary statistics of local estimates are presented in APPENDIX G (APPENDIX G1 on Page 345) and the spatial distribution of local estimates of the predictors is illustrated in APPENDIX H (APPENDIX H1 on Page 365).

3. Local parameter estimates for GWR Model 2

Explanatory variables included in GWR Models 2 are shown in Table 4-44. GWR Model 2A comprises seven explanatory variables while GWR Model 2B consists of 15 explanatory variables.

Table 4-44: Explanatory variables for GWR Models 2

Var No	GWR Model 2A	GWR Model 2B
1	Log_Popu	Log_Popu
2	Prop_White	Prop_AgeLess15
3	Prop_SR9Cat	Prop_Age15_24
4	Prop_GB.MU9Cat	Prop_AvgEd
5	Inters_grt3leg	Prop_NotWork
6	Round_Circ	Prop_UpperInc
7	Prop_Signal	ENT_9Cat
8		Prop_GI9Cat
9		Ratio_inters-cds
10		Prop_Freeways
11		Prop_Expresways
12		Prop_PrimaryArter
13		Prop_SecondArter
14		Round_Circ
15		Prop_Signal

Local parameter estimates for GWR Model 2A are presented in Table 4-45. The model includes seven explanatory variables to estimate the number of intersection-related pedestrian casualties. Of the seven explanatory variables, three variables emerge with negative values of local coefficients. These variables are: (1) the proportion of the White population; (2) the proportion of the single residential use (Prop_SR9Cat) and (3) the number of roundabouts and mini-circles.

Table 4-45: Summary statistics of local estimates for GWR Model 2A

Local parameters	Summary statistics of local estimates for GWR Model 2A							
	Mean	Minimum	Maximum	Std.Dev.	Global Model	t-value	df	p
Coeff Intercept	-15.686	-21.780	5.364	3.839	-15.378	-1.107	189	0.270
Coeff Log_Popu	4.134	-3.023	5.772	1.171	4.055	0.940	189	0.349
Coeff Prop_White	-0.077	-0.100	0.044	0.021	-0.065	-7.915	189	0.000
Coeff Prop_SR9Cat	-0.096	-0.134	-0.007	0.027	-0.092	-2.459	189	0.015
Coeff Prop_GB.MU9Cat	0.716	0.093	0.920	0.147	0.718	-0.129	189	0.897
Coeff Inters_grt3leg	0.348	0.247	0.395	0.026	0.325	12.257	189	0.000
Coeff Round_Circ	-1.447	-1.949	-0.388	0.298	-1.318	-5.990	189	0.000
Coeff Prop_Signal	1.960	1.140	2.134	0.179	2.018	-4.497	189	0.000
Local R^2	0.756	0.613	0.857	0.027	0.759	-1.829	189	0.069
Residual	-0.142	-35.973	77.518	15.885	0.000	-0.638	189	0.524
Std. Residual	-0.010	-2.361	5.440	1.052	0.000	-0.830	189	0.408

Marked values are not statistically significant at 95% confidence interval (i.e. p>0.05)

The *t*-test shows that five explanatory variables are statistically significant at the 5% level, with the exception of logarithm of population (Log-Popu) and the proportion of mixed use and general business use (Prop_GB.MU9Cat). This explains that spatial heterogeneity (i.e. variation in relationships across the study area) is not statistically significant for these two variables. Local estimates of each explanatory variable in GWR Model 2A are mapped in Figure 4-77 and Figure 4-78. An examination of the spatial patterns of the mapped local estimates shows that:

- Population number (Log_Popu) is more positively related to the frequency of intersection-related pedestrian casualties in the south eastern parts of the city (Helderberg district) and east-central parts (Khayelitsha/Mitchells Plain and Tygerberg districts) ;
- The proportion of the White population is greatly associated with a reduced number of intersection-related pedestrian casualties in the east-central parts of the city (Tygerberg and Khayelitsha /Mitchells Plain districts);

- The proportion of the single residential use is greatly related to a reduced number of intersection-related pedestrian casualties in the western regions of Cape Town (Table Bay district, certain suburbs of South Peninsula and Blaauwberg districts);
- A positive influence of the proportion of mixed use and general business use on the frequency of intersection-related pedestrian casualties is most pronounced in the western parts of Cape Town (Blaauwberg and Table Bay districts);
- The number of four- and multi-legged intersections is greatly associated with increased number of intersection-related casualties in northern and western parts of the city (Northen, Blaauwberg and Table Bay districts);
- A negative relationship between the number of roundabouts/mini-circles and intersection-related pedestrian casualties is most prominent in Northen and Tygerberg districts;
- The proportion of signalised intersection is most distinctly associated with increased number of intersection-related pedestrian casualties in Table Bay, Khayelitsha /Mitchells Plain, Cape Flats districts as well as in certain suburbs of Tygerberg and Southern districts.

Summary statistics of local estimates for GWR Model 2B are shown in APPENDIX G (APPENDIX G2 on Page 346) and the spatial patterns of local estimates of the predictors in GWR Model 2B are illustrated in APPENDIX H (APPENDIX H2 on Page 370).

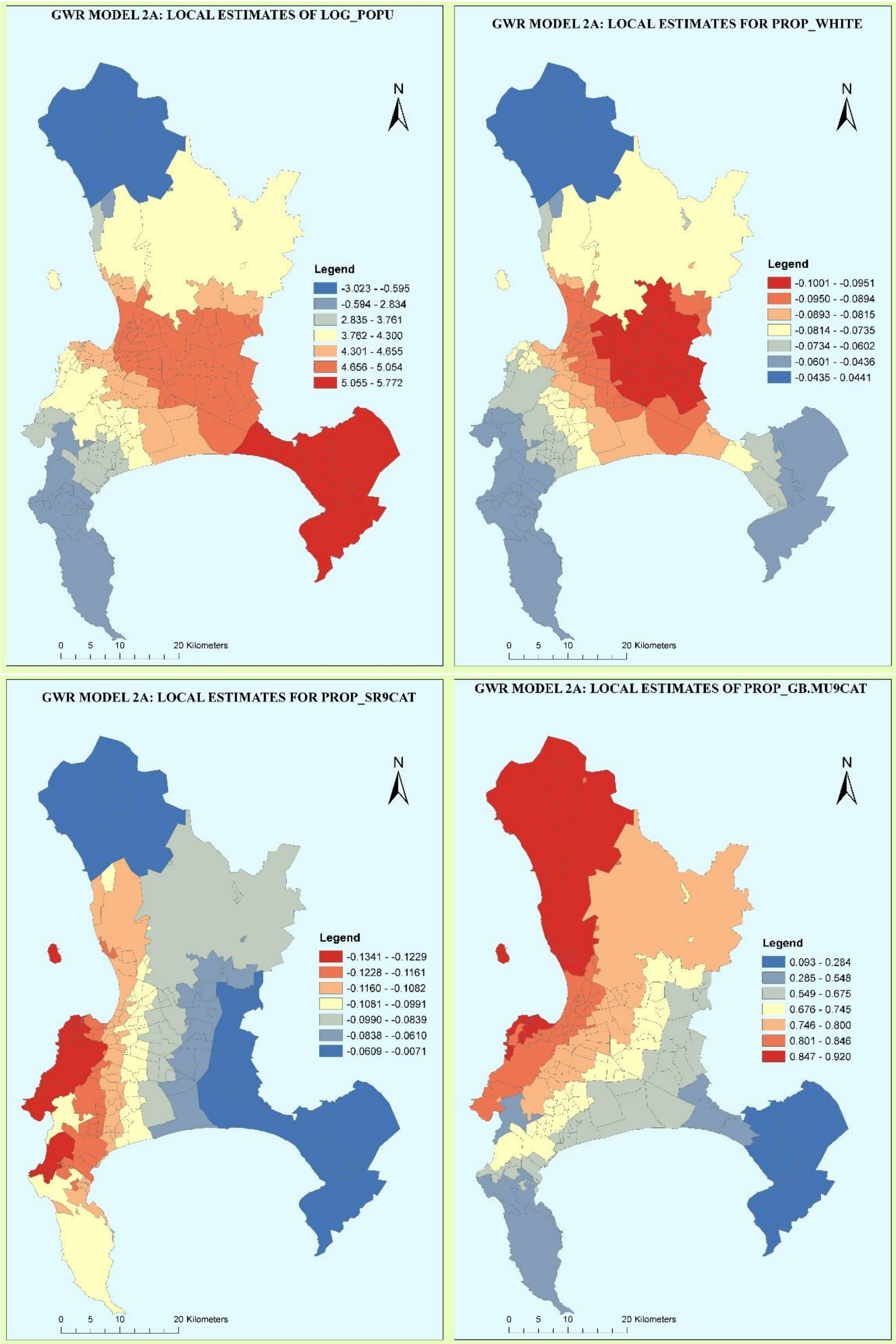


Figure 4-77: Local estimates for GWR Model 2A: (1) Log_Popu; (2) Prop_White; (3) Prop_SR9Cat; (4) Prop_GB.MU9Cat

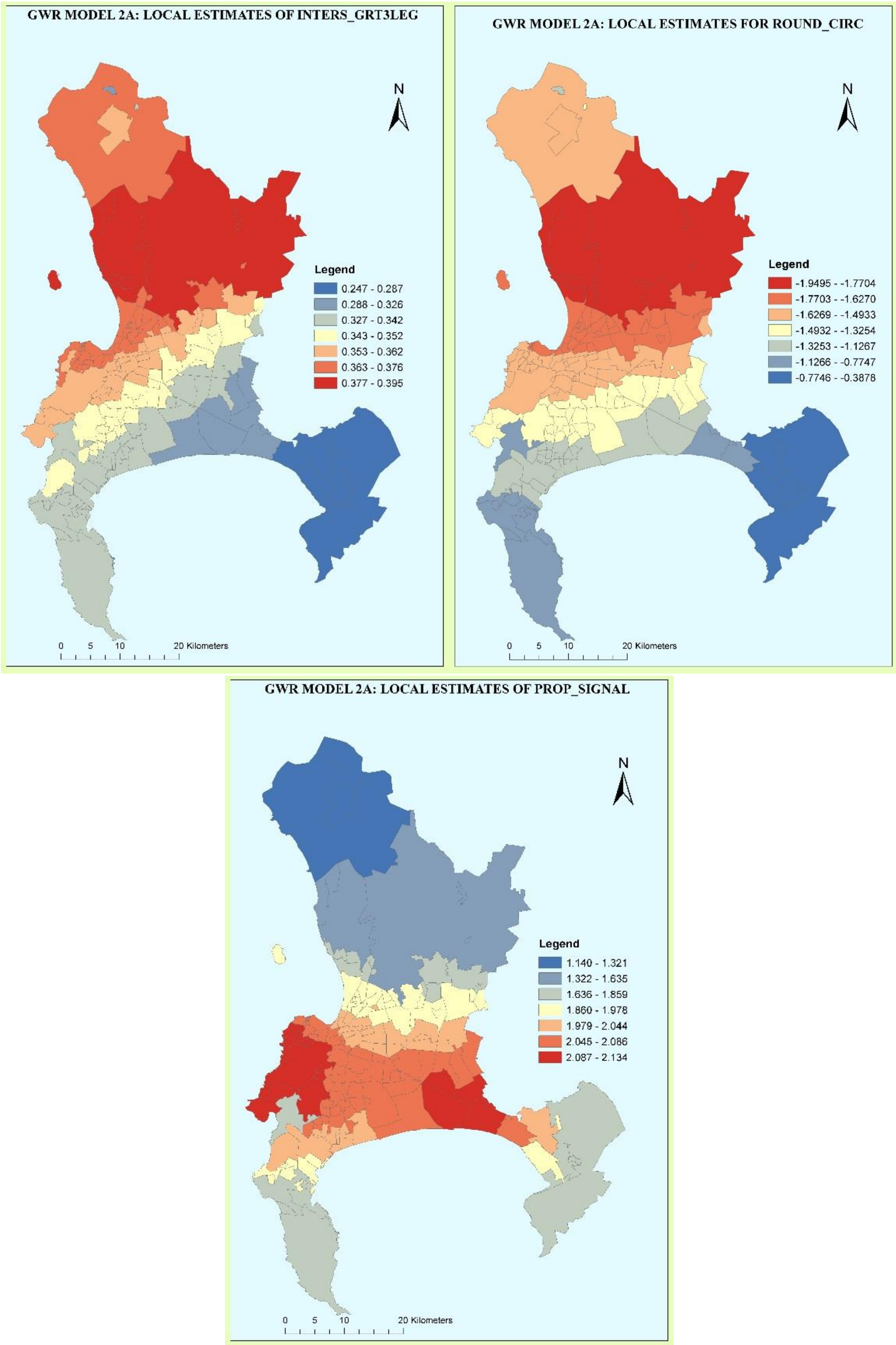


Figure 4-78: Local estimates for GWR Model 2A: (5) Inters_grt3leg; (6) Round_Circ; (7) Prop_Signal

4. Local parameter estimates for GWR Model 3

Table 4-46 presents explanatory variables included in GWR Models 3. In GWR Model 3A, KSI pedestrian casualty counts are estimated by five explanatory variables while the same outcome variable is estimated by eight predictors in GWR Model 3B.

Table 4-46: Explanatory variables for GWR Models 3

Var No	GWR Model 3A	GWR Model 3B
1	Prop_Black	Log_Popu
2	Prop_Age15_24	Prop_Coloured
3	Prop_Age25_54	Prop_White
4	Prop_NotWork	Inters_grt3leg
5	Inters_grt3leg	Prop_Freeways
6		Prop_Expressways
7		Prop_PrimaryArter
8		Prop_SecondArter

Summary statistics of local parameter estimates for GWR Model 3A are presented in Table 4-47. In this model, five explanatory variables were included to predict the number of KSI pedestrian casualties. Among these variables, the proportion of the population in the 15-24 age group and the proportion of the population in the 25-54 age group are negatively associated with the frequency of KSI pedestrian casualties. The analysis of spatial homogeneity of local estimates by the use of the *t*-test demonstrates that the proportion of the population in the 25-54 age group is statistically non-significant at the 5% level.

Table 4-47: Summary statistics of local estimates for GWR Model 3A

Local parameters	Summary statistics of local estimates for GWR Model 3A							
	Mean	Minimum	Maximum	Std.Dev.	Global Model	t-value	df	p
Coeff Intercept	7.135	-0.997	18.777	3.587	7.550	-1.597	189	0.112
Coeff Prop_Black	0.542	0.124	1.220	0.182	0.516	1.982	189	0.049
Coeff Prop_Age15_24	-0.382	-0.804	0.318	0.163	-0.358	-2.005	189	0.046
Coeff Prop_Age25_54	-0.523	-0.969	-0.200	0.155	-0.524	0.103	189	0.918
Coeff Prop_NotWork	0.477	0.005	0.837	0.123	0.438	4.438	189	0.000
Coeff Inters_grt3leg	0.384	0.217	0.434	0.042	0.378	2.174	189	0.031
Local R^2	0.680	0.552	0.775	0.042	0.668	3.740	189	0.000
Residual	-1.355	-90.005	177.811	25.097	0.000	-2.641	189	0.009
Std. Residual	-0.068	-4.977	8.396	1.103	0.000	-3.436	189	0.001

Marked values are not statistically significant at 95% confidence interval (i.e. p>0.05)

Local estimates of the predictors in GWR Model 3A are mapped in Figure 4-79 and Figure 4-80. An examination of the spatial patterns of local coefficients for each explanatory variable shows that:

- The contribution of the proportion of the Black population to the frequency of KSI pedestrian casualties is most significant in the south eastern parts of the city (Helderberg, Khayelitsha /Mitchells Plain and Cape Flats districts and partially in Tygerberg district);
- The proportion of the population in the 15-24 age group and that of the population in the 25-54 age group are most significantly associated with reduced numbers of KSI pedestrian casualties in the south eastern parts of Cape Town (Helderberg, Khayelitsha /Mitchells Plain and Cape Flats districts as well as certain suburbs of Tygerberg district);
- The proportion of the population who are not working (unemployed and discouraged workers) is greatly associated with increased number of KSI pedestrian casualties in the south eastern and central parts of Cape Town (Helderberg, Khayelitsha /Mitchells Plain, Cape Flats and Tygerberg districts);
- A positive relationship between the number of four- and multi-legged intersections and the frequency of KSI pedestrian casualties is most marked in the southern part of the city (Southern district, Khayelitsha /Mitchells Plain district, Cape Flats district as well as a part of Tygerberg district).

Summary statistics of local estimates for GWR Model 3B are provided in APPENDIX G (APPENDIX G3 on Page 347) and the spatial distribution of local estimates for the predictors in GWR Model 3B is shown in APPENDIX I (APPENDIX I3).

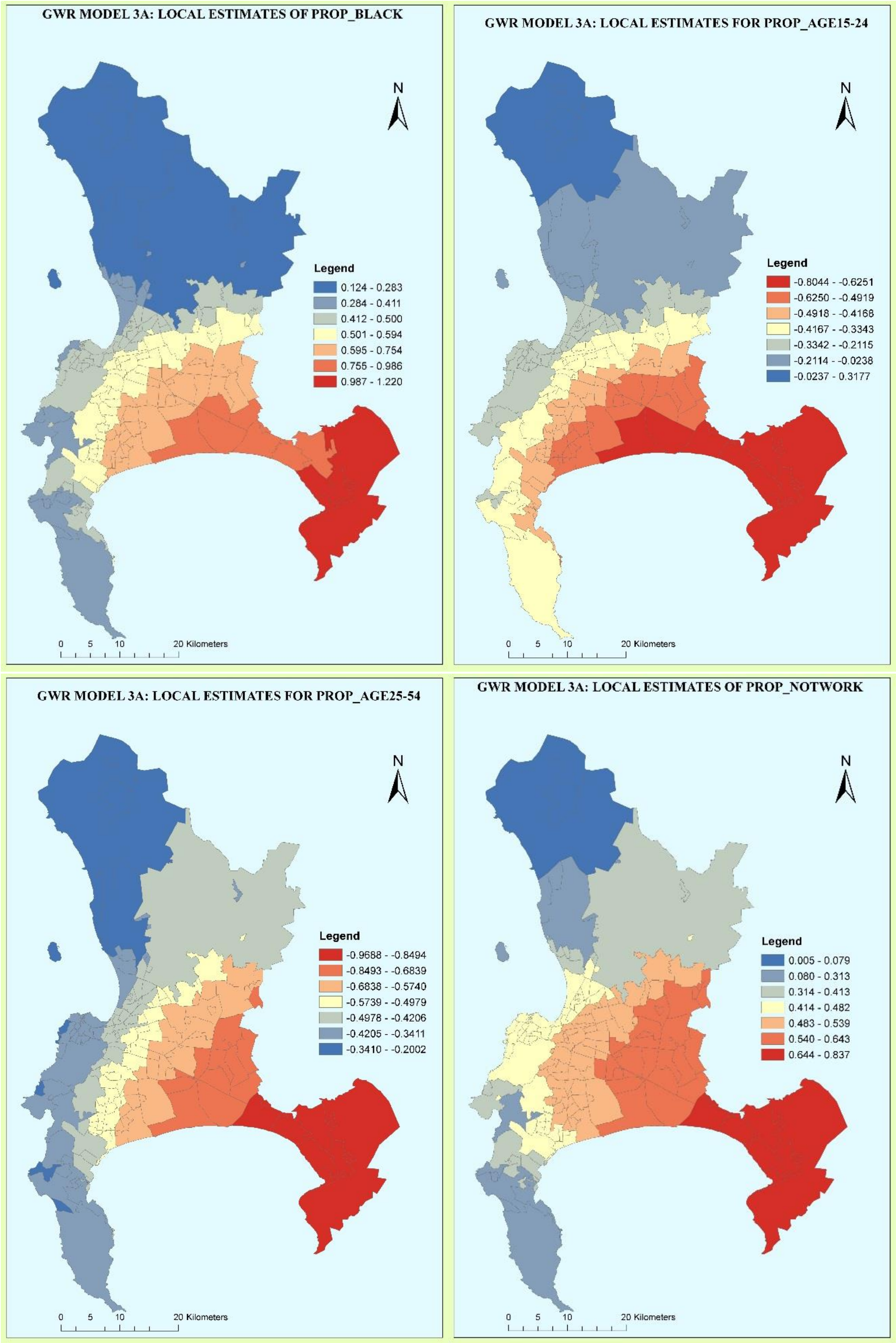


Figure 4-79: Local estimates for GWR Model 3A: (1) Prop_Black; (2) Prop_Age15-24; (3) Prop_Age 25-54; (4) Prop_NotWork

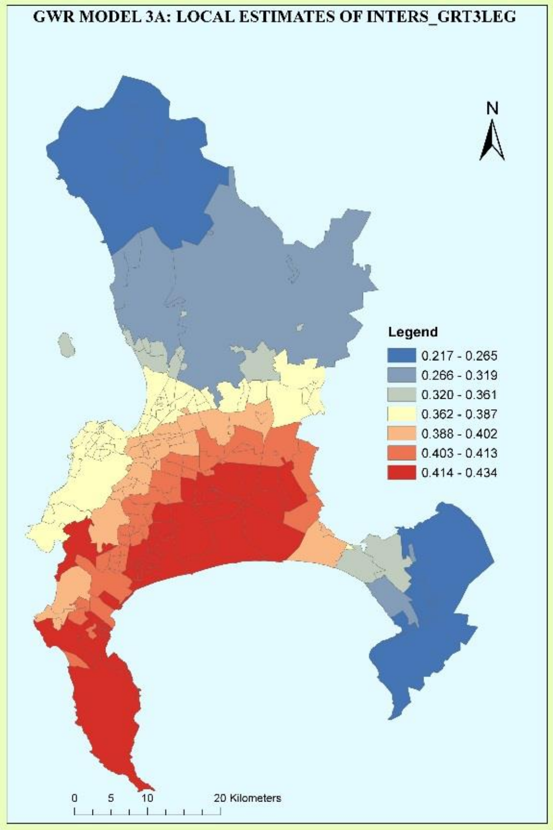


Figure 4-80: Local estimates for GWR Model 3A: (5) Inters_grt3leg

4.4.4.2 Evaluation of GWR model performance

The performance of GWR models was assessed by examining the goodness-of-fit measures previously provided in Table 4-41 on Page 228. In addition to these measures, an examination of raw residuals (i.e. the difference between the predicted and the observed number of pedestrian casualties for each census suburb) can tell how good a model fits the data. A comparison of goodness-of-fit measures such as R^2, Adjusted R^2 and AICc shows that GWR Models A have a better fitting performance than GWR Models B. For a better visualisation, raw residuals were standardised and mapped in ArcMap. Standardised residuals have a mean of zero and a standard deviation of 1 (ESRI, n.d.). A negative residual value (either raw or standardised) means that the predicted value is greater than the observed value (i.e. over-prediction) while a positive residual value explains the opposite (under-prediction). A residual value close to zero implies that the predicted value is very close to the actual value. The spatial distribution of the standardised residuals for the six GWR models is illustrated from Figure 4-81 to Figure 4-86. An analysis of mapped residuals helps to identify locations where the model performs well or poorly in predicting the outcome variable.

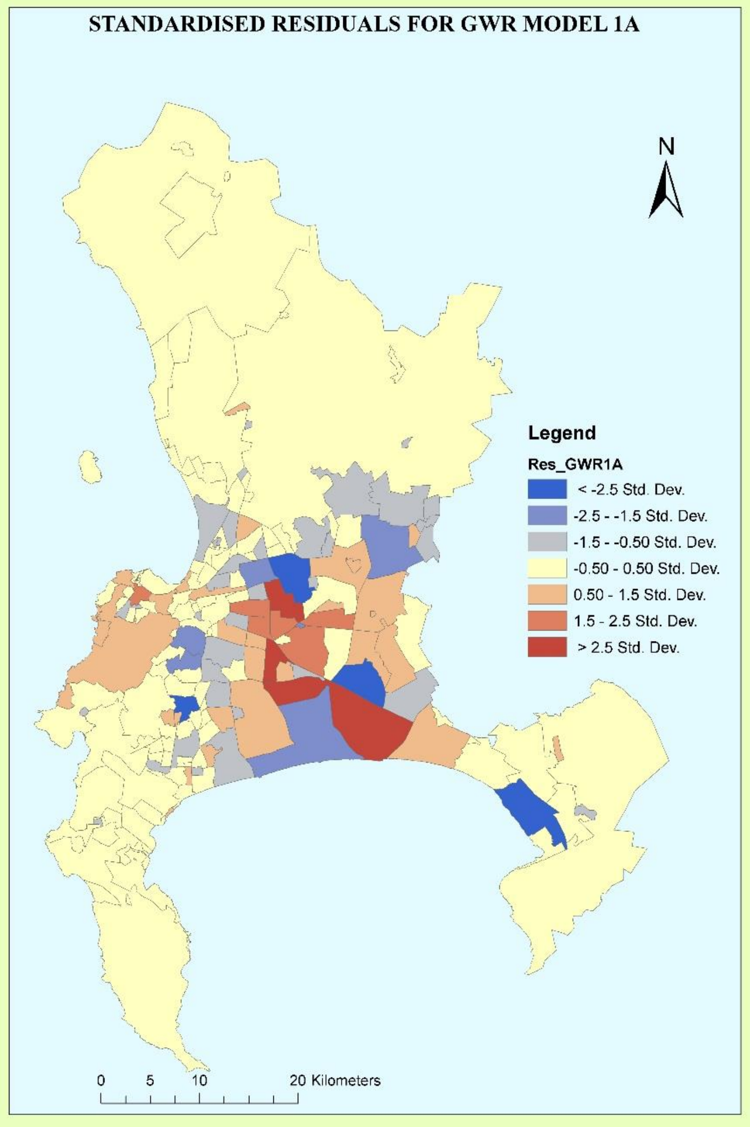


Figure 4-81: Standardised residuals for GWR Model 1A

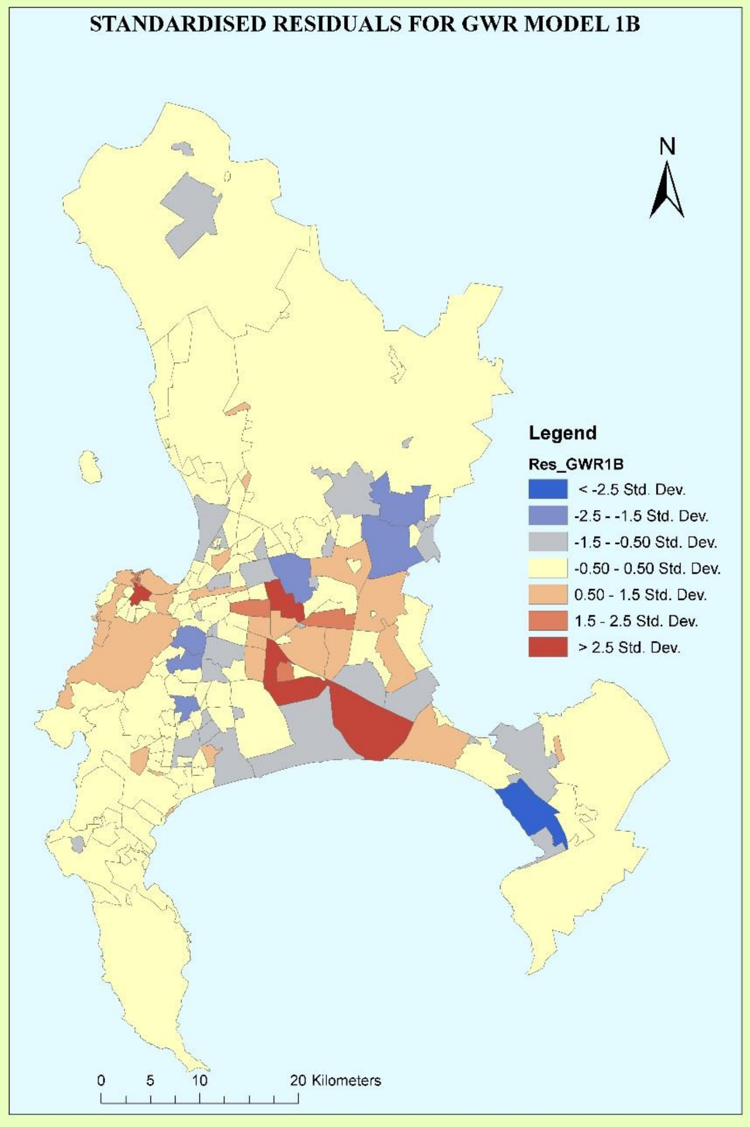


Figure 4-82: Standardised residuals for GWR Model 1B

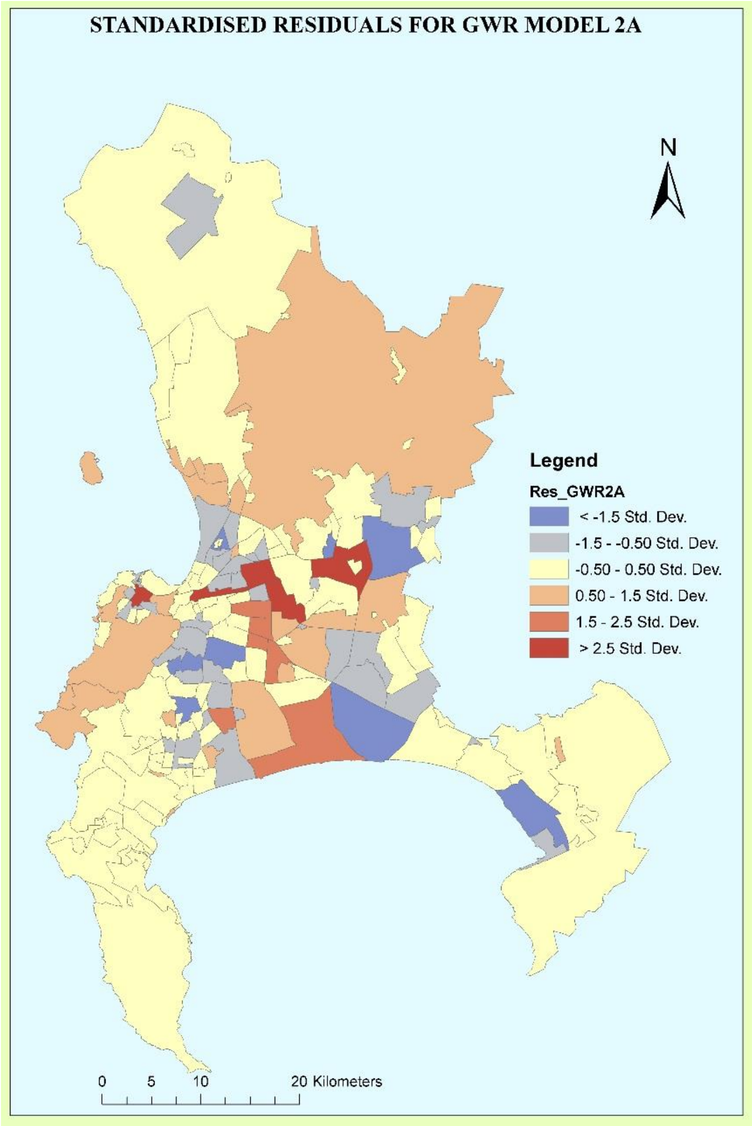


Figure 4-83: Standardised residuals for GWR Model 2A

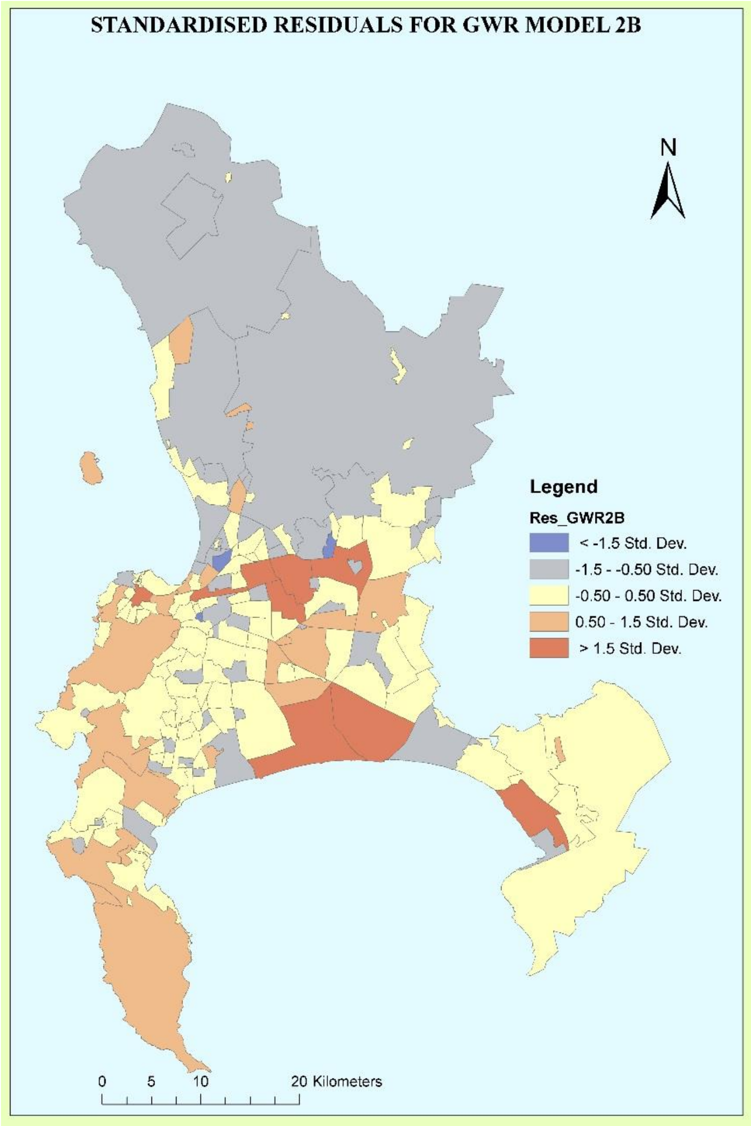


Figure 4-84: Standardised residuals for GWR Model 2B

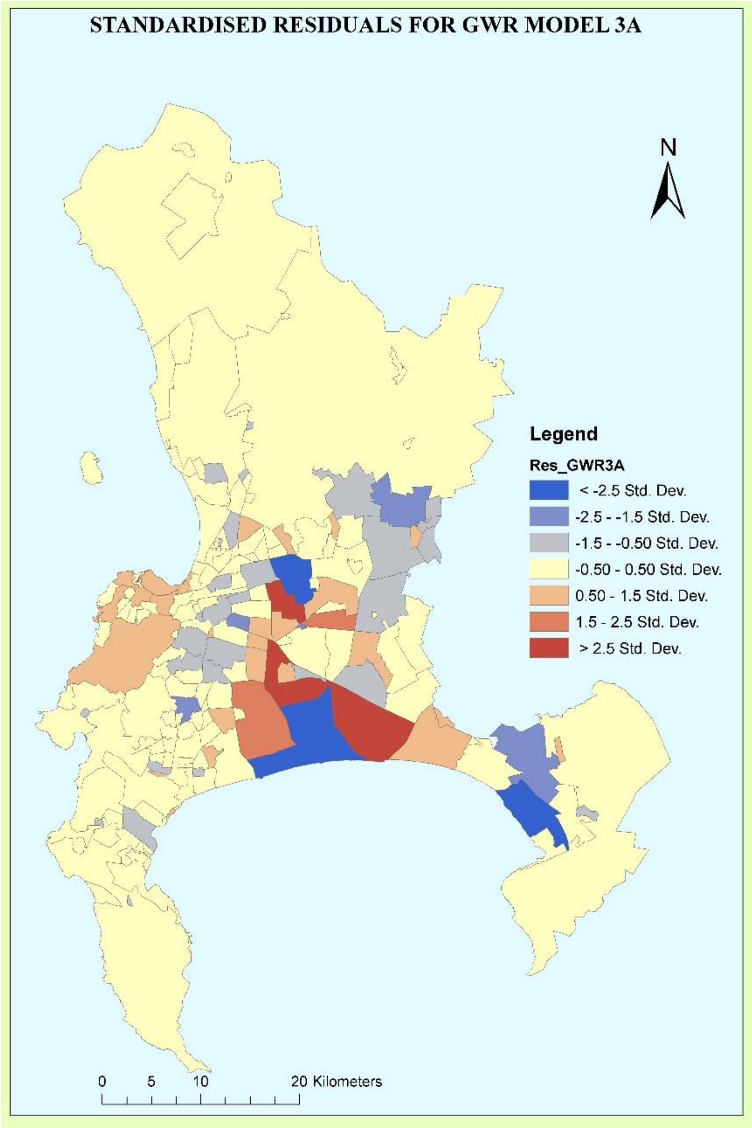


Figure 4-85: Standardised residuals for GWR Model 3A

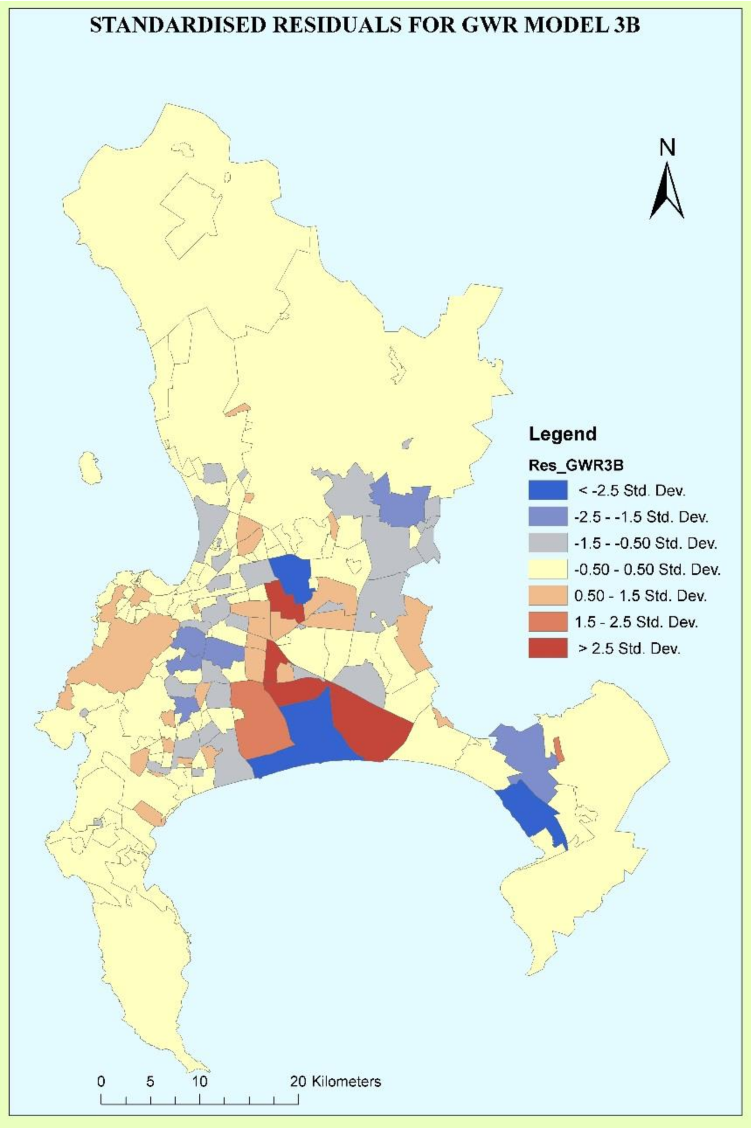


Figure 4-86: Standardised residuals for GWR Model 3B

4.4.5 Comparison of the model results

4.4.5.1 Parameter estimates comparison

Although the two modelling approaches (GLMs and GWR Models) applied in this study did not include the same number of explanatory variables, it is worth examining differences and similarities of parameter estimates generated by the two modelling techniques. The estimates (coefficients B) for GLMs and GWR Models applied in this study to fit the three datasets of pedestrian casualties (all pedestrian casualties, intersection-related pedestrian casualties and KSI pedestrian casualties) are summarised in Table 4-48, Table 4-49 and Table 4-50, respectively. In the two techniques of GLMs (NB models and Poisson regression models), all the demographic predictors are shown to be negatively associated with the frequency of pedestrian casualties, apart from the logarithm of population. The results from the best GWR Models (i.e. Models A) show the dominance of variables describing the population characteristics (socio-demographic and socio-economic variables) in the models, with the exception of GWR Model 2A. Of the seven variables in GWR Model 1A, three are demographic and the remaining variables describe land use, urban design and the transportation system. Of the five variables in GWR Model 3A, four are demographic whereas GWR Model 2A contains only two demographic variables. These findings are indicative of the powerful influence that demographic factors have on the frequency and the severity of pedestrian crashes.

In GWR Models, three explanatory variables do not maintain a negative sign in all models. These variables are: (1) the proportion of the population younger than 15 years (in GWR Models 1); (2) the proportion of the Black population (in GWR Models 1 and GWR Models 3); and (3) the proportion of the population who are not working (in GWR Models 3). The logarithm of population retains a positive sign in both GLMs and GWR Models. The coefficients for the logarithm of population in all three NB Models vary from 1.368 to 1.947 while the mean values of local coefficients for the same variable vary from 4.134 to 23.772 in GWR Models. .

Table 4-48: Estimates for GLMs and GWR Models: Models 1

Variables	Model 1: All pedestrian casualties							
	Parameter estimates							
	NB	Poisson	GWR Model 1A			GWR Model 1B		
			Mean	Minimum	Maximum	Mean	Minimum	Maximum
Intercept	0.283	-0.436	-46.032	-58.573	16.840	-33.446	-33.468	-33.426
Log_Popu	1.368	1.537				8.437	8.424	8.457
Prop_Black			0.743	0.263	1.663			
Prop_White	-0.013	-0.008						
Prop_AgeLess15	-0.043	-0.036	1.796	-0.712	2.381	1.589	1.589	1.590
Prop_Age15_24	-0.041	-0.035						
Prop_Age25_54	-0.022	-0.001						
Prop_AvgEd	-0.013	-0.021				-0.369	-0.370	-0.369
Prop_MidInc			-0.682	-0.992	-0.165			
Prop_UpperInc	-0.020	-0.032				-0.888	-0.888	-0.888
ENT_9Cat	1.158	0.823				-27.864	-27.892	-27.838
Prop_GI9Cat	0.024	0.015				-0.032	-0.033	-0.031
Prop_GB.MU9Cat			0.862	-2.504	2.105			
Inters_grt3leg	0.003	0.002	1.374	0.820	1.562	1.374	1.374	1.374
StrDens	0.022	0.021				0.590	0.589	0.592
Prop_Freeways	0.032	0.026				0.406	0.405	0.406
Prop_Expresways	0.059	0.032				0.396	0.396	0.397
Prop_PrimaryArter	0.024	0.019				-0.194	-0.196	-0.194
Prop_SecondArter	0.016	0.017				0.406	0.406	0.406
Round_Circ	0.035	0.013	-3.497	-6.021	-1.553	-3.492	-3.495	-3.491
Prop_Signal	0.073	0.078	4.284	-0.896	5.609	5.037	5.034	5.040

Marked values are not statistically significant at 95% confidence interval (i.e. p>0.05)

Table 4-49: Estimates for GLMs and GWR Models: Models 2

Variables	Model 2: Intersection-related pedestrian casualties							
	Parameter estimates							
	NB	Poisson	GWR Model 2A			GWR Model 2B		
			Mean	Minimum	Maximum	Mean	Minimum	Maximum
Intercept	-5.061	-4.926	-15.686	-21.780	5.364	-48.186	-48.250	-48.140
Log_Popu	1.947	2.020	4.134	-3.023	5.772	23.772	23.641	23.933
Prop_White			-0.077	-0.100	0.044			
Prop_AgeLess15	-0.046	-0.045				-0.337	-0.346	-0.332
Prop_Age15_24	-0.056	-0.069				-0.376	-0.381	-0.373
Prop_Age55_plus	-0.028	-0.039						
Prop_AvgEd	-0.025	-0.024				-0.194	-0.197	-0.190
Prop_NotWork	-0.035	-0.041				-0.211	-0.212	-0.210
Prop_UpperInc	-0.057	-0.060				-0.605	-0.607	-0.604
ENT_9Cat	1.610	1.583				5.281	5.110	5.553
Prop_SR9Cat			-0.096	-0.134	-0.007			
Prop_GI9Cat	0.025	0.014				0.120	0.117	0.124
Prop_GB.MU9Cat			0.716	0.093	0.920			
Inters_grt3leg			0.348	0.247	0.395			
Ratio_Inters-cds	0.053	0.074				0.625	0.614	0.635
Prop_Freeways	0.069	0.060				0.179	0.177	0.181
Prop_Expressways	0.080	0.080				0.351	0.347	0.354
Prop_PrimaryArter	0.053	0.062				0.316	0.313	0.318
Prop_SecondArter	0.037	0.063				0.102	0.100	0.104
Prop_LocalStr	0.033	0.033	-1.447	-1.949	-0.388			
Round_Circ	0.030	0.015	1.960	1.140	2.134	0.999	0.990	1.005
Prop_Signal	0.123	0.101				2.546	2.523	2.564

Table 4-50: Estimates for GLMs and GWR Models: Models 3

Variables	Model 3: KSI pedestrian casualties							
	Parameter estimates							
	NB	Poisson	GWR Model 3A			GWR Model 3B		
			Mean	Minimum	Maximum	Mean	Minimum	Maximum
Intercept	-0.286	0.023	7.135	-0.997	18.777	1.081	-14.763	64.421
Log_Popu	1.551	1.934				8.955	-0.317	14.686
Prop_Black			0.542	0.124	1.220			
Prop_Coloured						-0.492	-1.273	-0.084
Prop_White						-0.624	-1.314	-0.152
Prop_AgeLess15	-0.035	-0.054						
Prop_Age15_24	-0.045	-0.079	-0.382	-0.804	0.318			
Prop_Age25_54			-0.523	-0.969	-0.200			
Prop_Age55_plus	-0.050	-0.065						
Prop_AvgEd	-0.016	-0.019						
Prop_NotWork	-0.018	-0.012	0.477	0.005	0.837			
Prop_MidInc	-0.013	-0.013						
Prop_UpperInc	-0.042	-0.046						
ENT_9Cat	0.885	0.419						
Prop_GI9Cat	0.019	0.009						
Inters_grt3leg	0.003	0.001	0.384	0.217	0.434	0.343	0.201	0.426
Prop_Freeways	0.035	0.037				-0.035	-0.192	0.930
Prop_Expressways	0.034	0.021				0.209	-0.377	0.622
Prop_PrimaryArter						0.227	-0.045	0.641
Prop_SecondArter						0.336	-0.256	0.739
Prop_Signal	0.083	0.094						

Marked values are not statistically significant at 95% confidence interval (i.e. p>0.05)

With respect to the attributes of the built environment, the number of four- and multi-legged intersections are shown to be positively associated with the frequency of pedestrian casualties in two of the three NB Models (Model 1 and Model 3), with the coefficients B ranging from 0.0030 to 0.0032. The number of four- and multi-legged intersections emerges statistically significant in five of the six GWR Models, in which the variable is positively related to the frequency of pedestrian casualties. In these models, the mean values of local coefficients vary from 0.348 to 1.384.

The model results indicate that the proportion of signalised intersections is positively related to the frequency of pedestrian casualties in all GLMs, GWR Models 1 and GWR Models 2. The coefficients B in NB Models vary from 0.073 to 0.123. The mean values of local coefficients B are in the range of 2.546 to 5.037. However, this variable does not emerge statistically significant in GWR Models 3.

Two types of land use are shown to be statistically significant in the best GWR Models (Models A). These are the proportion of the single residential use (Prop_SR9Cat) which demonstrates a negative mean value of local estimates in GWR Model 2A and the proportion of mixed use and general business use (Prop_GB.MU9Cat) with positive mean values of local estimates in both GWR Model 1A and GWR Model 2A. The proportion of the general industrial use (Prop_GI9Cat) is significant in GWR Model 1B with a negative mean value of local estimates (mean value of B= -0.032). However, local estimates of this variable demonstrate a positive mean value of local estimates (mean value of B=0.120) in GWR Model 2B. For GLMs, the proportion of the general industrial use (Prop_GI9Cat) is the only land use type that has a statistically significant influence in these models. For the NB Models, the coefficient B is in the range from 0.019 to 0.026.

The number of roundabouts and mini-circles (Round_Circ) demonstrates a negative relationship with the frequency of pedestrian casualties in two of four GWR Models that have this variable as a predictor. Negative mean values of local estimates range from -3.492 to -3.497 in GWR Model 1A and GWR Model 1B. However, this variable has positive values of local estimates in GWR Model 2A and GWR Model 2B, ranging from 0.999 to 1.960. In GLMs, this variable is shown to be statistically significant in two models (Model 1 and Model 2), with positive relationships with the frequency of pedestrian casualties. The coefficient B is 0.035 and 0.030 in NB Model 1 and NB Model 2, respectively.

Another intersecting observation is the absence of the entropy index (ENT_9Cat) in the best GWR Models. This variable is statistically significant in GWR Model 1B and GWR Model 2B with mean values of local coefficients of -27.86 and 5.28, respectively. However, it is important to note that the explanatory variables in Models B were selected with the intention to have similar variables to those included in GLMs for comparison purposes. For instance Model 1B and GWR Model 2B were selected based not necessarily on their performance, but on their inclusion of entropy index and the variables describing functional road class. In all GLMs, the entropy index is the second most powerful predictor of pedestrian casualties after the logarithm of population (Log_Popu). The coefficients B in NB Models are in the range from 0.885 to 1.610.

The exploratory regression analyses carried out in ArcMap to find the best models demonstrated that the entropy index emerges in a very limited number of models that passed all of the necessary ordinary least of square (OLS) diagnostic tests. In other words, models containing the entropy index as an explanatory variable performed poorly in predicting the frequency of pedestrian casualties. For instance, of 135 models that include seven to 10 explanatory variables, only six models with entropy index passed the diagnostic tests of the exploratory regression tool for the entire sample of pedestrian casualty (Model 1). In addition, these six passing models are also characterised by higher values of maximum variance inflation factor (VIF), ranging from 2.55 to 6.41. The magnitude of these VIF values suggests the presence of a higher level of multicollinearity among explanatory variables. In summary, land use mix measured by entropy index was not found to be a reliable predictor of pedestrian casualties using the GWR modelling technique.

Similarly to the entropy index, none of variables describing functional road classes (i.e. the proportions of freeways, expressways, primary arterial roads and secondary arterial roads) emerged in the GWR Models A. The inclusion of these variables in the model was one of the selection criteria for the three GWR Models B. The GWR Models B intend to examine the spatial variation of the influence of the variables of functional road class on pedestrian casualties. Generally, the three GWR Models containing four variables of functional road class demonstrate positive associations with the frequency of pedestrian casualties, apart from the proportion of primary arterial roads in GWR Model 1B and the proportion of freeways in GWR Model 3B. A further analysis of *t*-test confirms that the spatial heterogeneity for all four variables is statistically significant. Therefore, the GWR models successfully capture the spatial heterogeneity of the relationship between road class and the frequency of pedestrian casualties. Among the four variables of road class, only the proportion of freeways and proportion of expressways are shown to have significant relationships with the frequency of pedestrian casualties in all GLMs. Two other variables (the proportion of primary arterial roads and the proportion of secondary arterial roads) emerged statistically significant only in Model 1 and Model 2 of GLMs. A significant influence of the proportion of local streets on the frequency of pedestrian casualties is shown only in Model 2 of the Generalised Linear Modelling technique.

Other variables of the built environment which are shown to be related to the frequency of pedestrian casualties include street density (Model 1 of GLMs and GWR Model 1B) and the ratio of intersection to culs de sac (Model 2 of GLMs and GWR Model 2B). Street density has a coefficient B of 0.022 in NB Model 1 while the local coefficients B in GWR Model 1B have a mean value of 0.590. In NB Model 2, the ratio of intersections to culs de sac has a coefficient B of 0.053, while the mean value of local coefficients B in GWR Model 2B is 0.625.

Another important observation concerning local parameter estimates from the GWR Models is that, while the mean value of local coefficients might be either positive or negative, some census suburbs in study area have demonstrated parameter estimates with a counterintuitive sign. This phenomenon is identified for a total number of 10 explanatory variables in four GWR Models as illustrated on mapped results from Figure 4-75 to Figure 4-80 (GWR Models A) and APPENDIX H (GWR Models B) as well as in the summary statistics of parameter estimates provided from Table 4-48 to Table 4-50. Parameter estimates with a counterintuitive sign have been identified for the following variables:

1) The proportion of the population younger than 15 years, the proportion of mixed use and general business use and the proportion of signalised intersections in GWR Model 1A;
2) Log of population and the proportion of the White population in GWR Model 2A;
3) The proportion of the population in the 15-24 age range in GWR Model 3A;
4) Log of population and the proportions of the four road class types (i.e. freeways, expressways, primary arterial roads and secondary arterial roads) in GWR Models 3B.

The issue of counterintuitive signs is common for the GWR modelling technique and has been reported in previous studies that used this modelling techniques (Amoh-Gyimah *et al.*, 2017; Guo, Ma & Zhang, 2008; Hadayeghi *et al.*, 2010; Li *et al.*, 2013; Pirdavani *et al.*, 2014; Zhang *et al.*, 2015). The possible reason explaining the issue of counterintuitive signs is the presence of multicollinearity among certain variables in some spatial units of analysis even though multicollinearity among variables may not exist globally (Pirdavani *et al.*, 2014). The issue of counterintuitive signs could also arise as a result of some explanatory variables that may be less significant or even insignificant in some spatial units of analysis producing thus unexpected signs (Amoh-Gyimah *et al.*, 2017).

In total, 13 explanatory variables describing the built environment are statistically significant in GWR Models. For each explanatory variable, the intensity and spatial distribution of local estimates for the three datasets (all pedestrian casualties, intersection-related pedestrian casualties and KSI pedestrian casualties) were compared. The comparison also encompasses mapped quantities aggregated at census suburb level to assess the relevance of the spatial relationships depicted by the parameter estimates. The comparison is illustrated in APPENDIX H.

From the mapped model results, it can be seen that the intensity and the direction of the influence of the variables describing the built environment and population characteristics on the frequency of pedestrian casualty generally vary spatially across the study area. However, for a small number of predictors, the spatial heterogeneity is not statistically significant. These predictors include the proportion of the Black population and the proportion of signalised intersections for GWR Model 1A; the logarithm of population and the proportion of mixed use and general business use for GWR Model 2A; and the proportion of the population aged 25-54 years old for GWR Model 3A.

4.4.5.2 Model performance comparison

The goodness-of-fit measures for GLMs summarized in Table 4-36 and those in Table 4-41 clearly demonstrate that the NB models have the lowest value of Corrected Akaike's Information (AICc), suggesting that the NB model is the most suitable to fit pedestrian casualty data in this study. The values of AICc for GWR models are found to be lower than those produced by the Poisson regression model, making the GWR models the second best models to fit pedestrian casualty data. Nevertheless, the ability of GWR Models to explore local variations of parameter estimates makes this modelling technique the most suitable when the interest is to evaluate whether the relationships are spatially consistent across the study area.

Another approach when assessing the predictive quality of models is to look at the value of model residuals. An evaluation of residuals indicates how far the predicted values deviate from the observed ones. Raw residuals (the difference between the observed and the predicted values) were produced in STATISTICA while fitting the Poisson Regression and Negative Binomial models to pedestrian casualty data. A dataset of raw residuals was then imported in ArcMap and standardised residuals were mapped in ArcMap for each GLM procedure (NB and Poisson regression models) to allow for a comparison with those produced in ArcMap from the GWR models. GLMs are compared with GWR Models A as these are considered as the best models in terms of performance assessed by goodness-of-fit measures provided in Table 4-41. A comparison of the spatial pattern of the standardised residuals is shown in Figure 4-87 (Model 1 for GLMs and GWR Model); Figure 4-88 (Model 2 for GLMs and GWR Model); and Figure 4-89 (Model 3 for GLMs and GWR Model). Negative standardised residuals which significantly deviate from 0 indicate over-prediction of the outcome variable whereas those with higher positives values reflect under-prediction of outcome variable by the model.

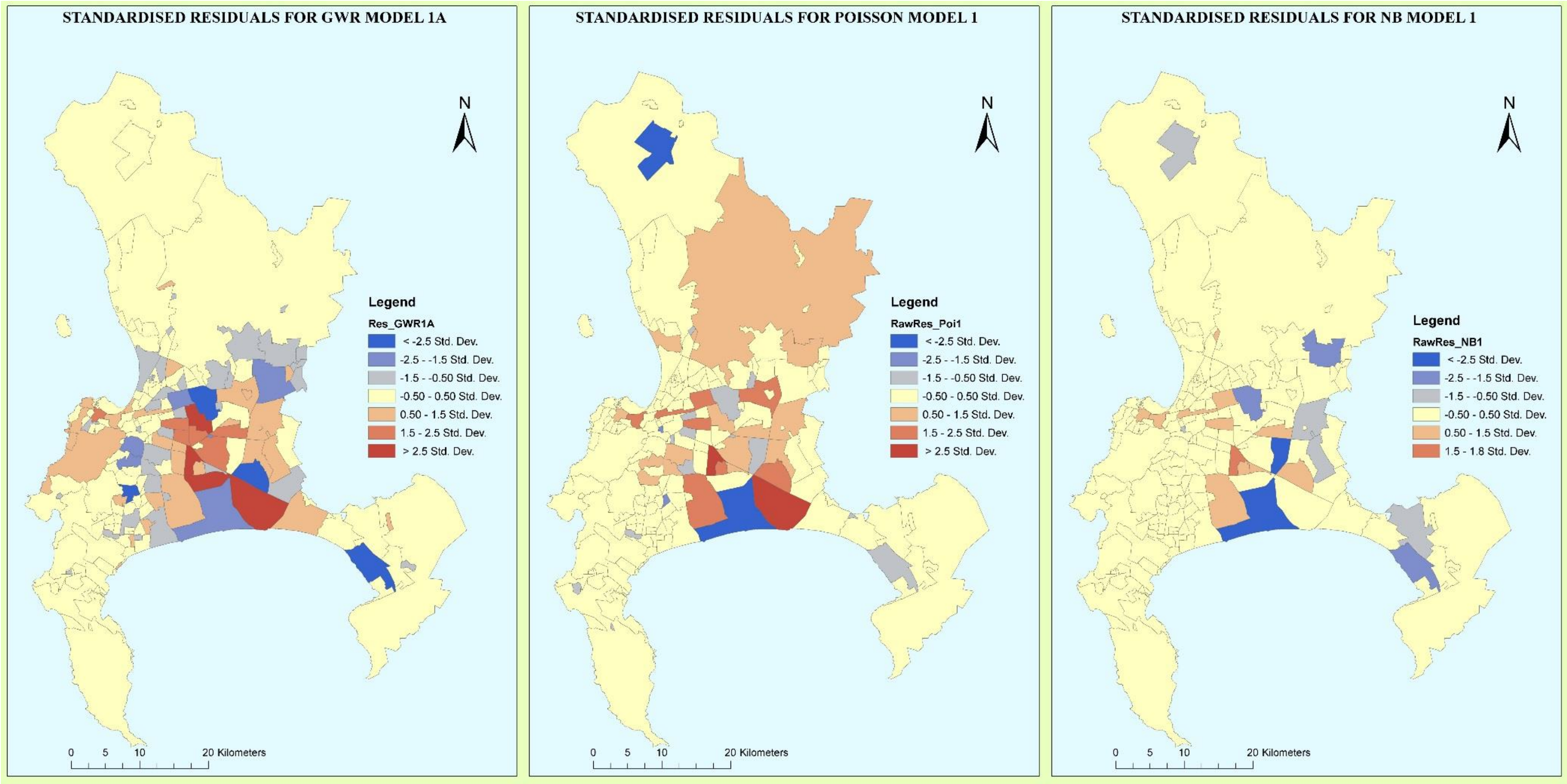


Figure 4-87: Spatial distribution of residuals for Models 1: GWR, Poisson and NB

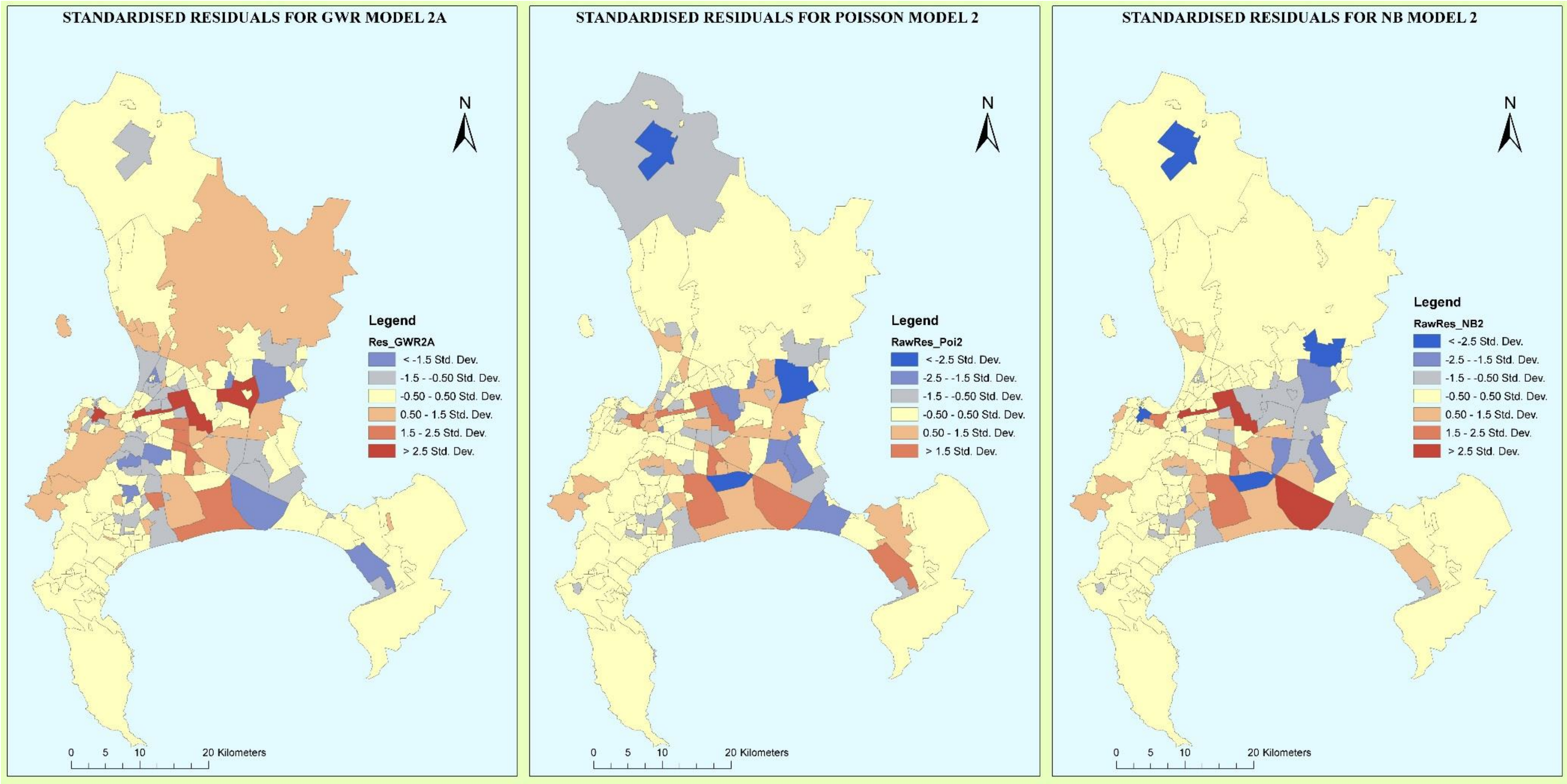


Figure 4-88: Spatial distribution of residuals for Models 2: GWR, Poisson and NB

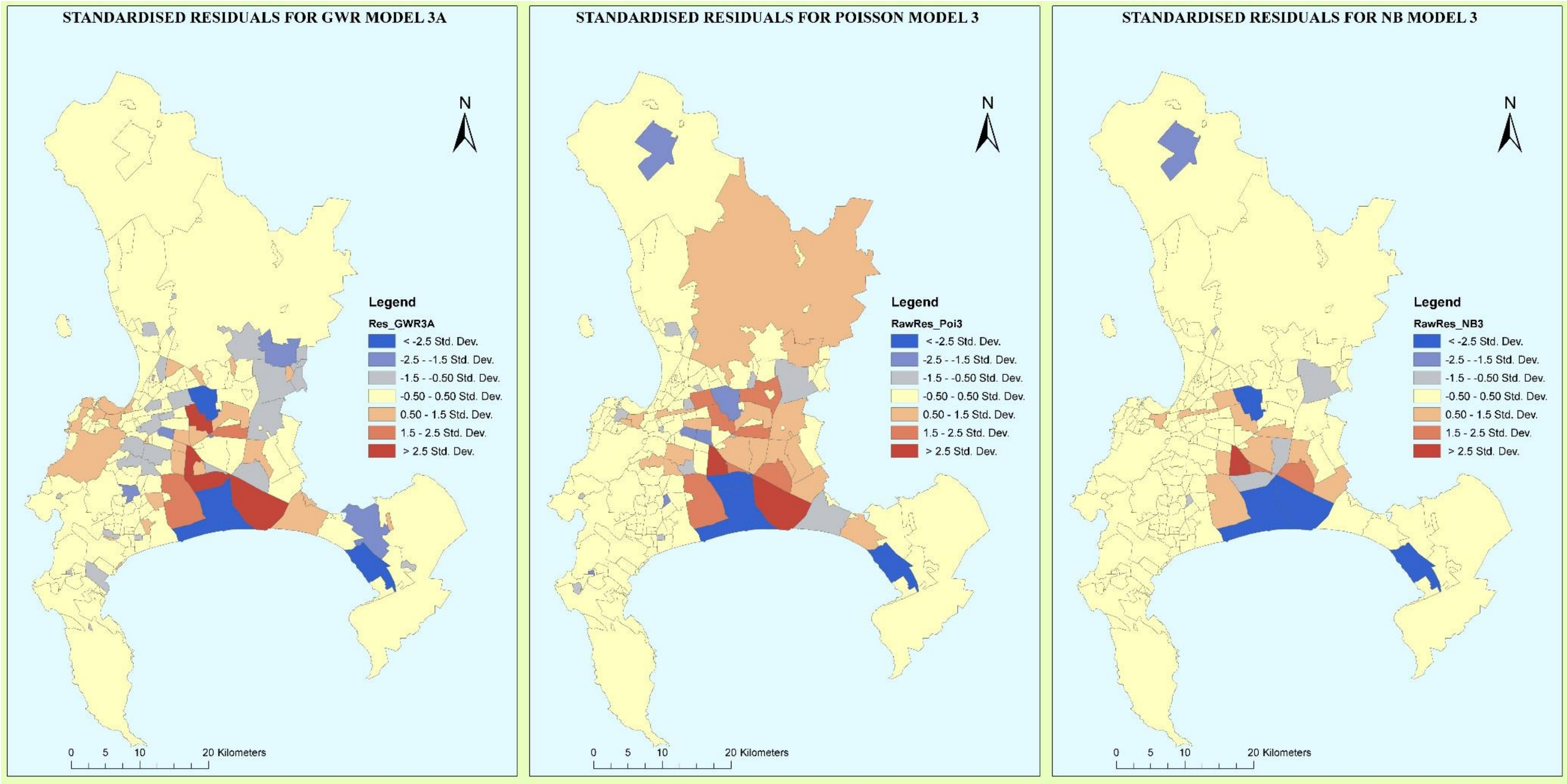


Figure 4-89: Spatial distribution of residuals for Models 3: GWR, Poisson and NB

4.5 Summary of key results

This section presents a summary of the results obtained from the mix of analysis methods applied to pedestrian casualty data to investigate the link between the built environment and pedestrian crashes. The analysis techniques include descriptive analysis, inferential analysis, geospatial and statistical modelling techniques.

4.5.1 Results from univariate and bivariate analyses of pedestrian casualties

4.5.1.1 Pedestrian casualty frequency

- On average, 4,618 pedestrians were involved in road traffic crashes annually in the City of Cape Town.
- Of all recorded pedestrian casualties, males represent 45 percent while females account for 28 percent. The remaining proportion (27 percent) consists of pedestrian casualties for which the gender of the victim was unknown.
- The age of the victim was recorded as “0” in sixty percent (8, 310 cases) of pedestrian casualties.
- The highest frequency of pedestrian casualties is found among child pedestrians in the 6-10 age group, followed by the 26-30 age group for both male and female victims.
- Male pedestrians are overrepresented in more severe injuries (i.e. fatal and serious injuries) while female pedestrians are overrepresented in slight injuries.
- There is an apparent dip in pedestrian casualty trends for both males and females in the 11-20 age group.
- Disproportionate casualty frequencies among female and male pedestrians are apparent in certain age groups: the first three highest male-to-female ratio emerges in the 41-45 age group, followed by the 31-35 age group and lastly in the 6-10 age group. In contrast, the frequency of female pedestrian casualties in the 76-80 age group is found to be twice as high as that of male casualties in the same age group. The profile of underlying population in the study area justifies this unexpected finding.
- The highest frequencies of pedestrian casualties are observed during morning and evening peak traffic hours (7:00 to 8:00 AM and 4:00 pm to 6:00 PM).
- The mean daily counts of pedestrian casualties peaks on Friday (15.05±4.43) and Saturday (15.01±5.38) and the lowest mean daily counts is observed on Wednesday (11.05±3.74).

- The frequency of pedestrian casualties is highest during the pay week (i.e. the week of the month that contains the first date on a month), followed by the second week after the pay week and the lowest frequency is observed in other remaining weeks of a month. The respective mean weekly counts are found to be 98.19 (±12.80); 87.29 (±13.50) and 85.09 (±12.19).
- Weekly casualty counts analysed according to the quarters of calendar year peak over the third quarter (93.10 ±14.90) and the lowest casualty counts are observed over the first quarter (82.95 ±13.16).

4.5.1.2 Injury severity description

- Fatally injured pedestrians represent 4 percent of the total sample; seriously injured pedestrians account for 25 percent; slightly injured pedestrians represent 47 percent; and "no injury cases" are reported for 24 percent of the total sample.
- The annual casualty rate is found to be 123.5 pedestrian casualties per 100,000 population for the entire study area. The annual KSI rate stands at 35.7 KSI pedestrian casualties per 100,000 population and the annual fatality rate is found to be 4.5 pedestrian fatalities per 100,000 population.
- Males are found overrepresented in more severe injuries than females: while males account for 62.6 percent of the total sample, they represent 74.2 percent of fatal pedestrian injuries.
- The risk of being involved in both fatal crashes and KSI casualties is found to be highest among Black African pedestrians followed by Coloured pedestrians.
- The number of pedestrian fatalities peaks over the weekend (Saturdays and Sundays) and another minor peak is observed on Fridays. The lowest number of pedestrian fatalities is observed on Wednesdays.
- Injury severity is fairly evenly distributed across the months of the year.
- Pedestrian fatalities are most predominant among middle-aged groups (age groups between 26-40 years) and children aged 5 years old and younger.
- There are three apparent dips in female fatalities: in the 6-15 age range, 36-45 age range and the 71-80 age range.
- There are three apparent dips in male fatalities: in the 11-10 age range and the 71-80 age range.

- Pedestrian fatalities peak between 7:00 PM and 8:00 PM and a second peak emerges between 8:00 AM and 9:00 AM.
- Pedestrian fatalities are more likely to occur during late afternoon hours and night times (between 3:00 PM and 12:00 PM).
- There is a dip in KIS casualties in the 11-20 age range for both females and males and another dip is found in the 41-45 age group for female KSI casualties.
- The highest frequency of KSI pedestrian casualties is observed on Saturdays (mean daily count: 5.30±2.47), followed by Sundays (mean daily count: 4.12±2.27) and Fridays (mean daily count: 4.09±2.15).
- Two peaks of KSI pedestrian casualties are observed between 6:00 PM and 7:00 PM and between 8:00 AM and 9:00 AM.
- Pedestrian fatalities are found to occur more frequently during late afternoon hours and night times (between 2:00 PM and 10:00 PM).

4.5.1.3 Pedestrian behaviour and actions prior to the incidence of crashes

- The vast majority (88 percent) of pedestrian casualties are found to have occurred outside of a designated crossing point. Only 5.7 percent of pedestrian casualties are observed at a designated crossing point and 4.3 percent of the total sample are found to have occurred within 50 metres of a pedestrian crossing facility.
- The majority (81.3 percent) of pedestrian casualties occurred when pedestrians were crossing the road.
- Before the incidence of pedestrian crashes, pedestrians who were walking with their back facing the traffic account for 8.6 percent and those who were walking facing the traffic represent7.8 percent of the total sample.
- The most common actions performed by pedestrians prior to a pedestrian crash are: walking (61.3 percent); running (25.0 percent); standing (7.9 percent), sitting (3.0 percent) and playing (1.8 percent).
- Crossing outside a designated crossing point is most frequently observed among young children of 5 years of age and younger (94.2 percent); children in the 6-10 age group (92.1 percent) and elderly pedestrian in the 76-80 age group (90.5 percent).
- Crossing within 50 metres of a crossing facility is observed more frequently among elderly pedestrians aged 81 and older (20 percent) and those in the 71-75 age group (9.6 percent).

- Running is commonly observed among the three age groups of children and adolescents: the 1-5 age group (38.1 percent); the 6-10 age group (37.0 percent) and the11-15 age group (28.5 percent). Likewise, playing is most frequently observed in the same age groups, representing 9.4 percent, 6.3 percent and 2.0 percent in the respective age groups.

4.5.1.4 Locations of pedestrian crashes

- Approximately three quarters (74 percent) of pedestrian casualties occurred at non-intersection locations (links) and 26 percent of pedestrian casualties took place at intersection locations.
- Of intersection-related pedestrian casualties;
    - 56.3 percent occurred at four-legged intersections
    - 38.9 percent occurred at three-legged intersections
    - 3.6 percent are found at roundabouts and mini-circles
    - 1.2 percent are observed at staggered intersections.
- Pedestrian casualty rates (casualties per 100 intersection configuration types) at different intersection types are found as follows:
    - Four-legged intersection (19.37)
    - Roundabouts/mini-circle (19.88)
    - Staggered intersections (8.30)
    - Three-legged (3.09).
- The following proportions of the total sample of pedestrian casualties are found according to the type of intersection control:
    - 43.6 percent are found at signalised intersections
    - 28.3 percent were observed at 1-Way Stop intersections
    - 17.6 percent occurred at 2-Way Stop intersections
    - 3.8 percent took place at 4-Way Stop intersections
    - 1.6 percent were found at 3-Way Stop intersections
    - 1.2 percent occurred at uncontrolled intersections
    - 0.3 percent were reported at 1-Way Yield intersections and
    - 0.1 percent took place at 2-Way Yield intersections.
- Pedestrian casualty rates (casualty counts per 100 intersection control types) by intersection control type are found as follows:
    - Signalised intersections (120.14)

    - 4-Way Stop intersections (11.81)
    - Uncontrolled intersections (8.17)
    - 2-Way Stop intersections (7.33)
    - 3-Way Stop intersections (5.96)
    - 1-Way Stop intersections (2.35).
- Injury severity by intersection configuration type:

    The highest frequencies of both KSI casualties and pedestrian fatalities are found at four-legged and three-legged intersections.
- Injury severity by intersection control type:
    - Higher proportions of pedestrian fatalities are found at 1-Way Yield intersections and signalised intersections;
    - Higher proportions of KSI pedestrian casualties are observed at 1-Way Yield intersections and 4-Way Stop intersections.

4.5.2 Results from geospatial analyses

- The highest frequency of pedestrian casualties are found in the Khayelitsha/Mitchells Plain regions, in Stand, Delft, Bellville, Elsies Rivier, and the CBD of Cape Town.
- Geospatial analyses by Moran's I , Getis-Ord Gi, and OHA tools demonstrate that pedestrian casualties are concentrated in the south-eastern regions of the City of Cape Town:
    - Khayelitsha/Mitchells Plain districts
    - Cape Flats districts
    - Certain suburbs of Tygerberg district
- Similarly, hot spots of intersection-related pedestrian casualties are found in:
    - Khayelitsha/Mitchells Plain districts
    - Certain suburbs of Tygerberg district
    - Table Bay district (CBD of Cape Town, Woodstock, Foreshore Green Point, V&A Waterfront and Schotschekloof suburbs).
- Cold spots of intersection-related pedestrian casualties are identified in Silvermine and Cape Peninsula National Park suburbs which are part of Southern district.
- Cluster analysis of intersection-related pedestrian casualties by Weighted Point Method:
    - The most significant clustering is observed in the area extending over four census suburbs which are Goodwood, Thornton, Ruyterwacht and Elsies River

  - Another significant clustering is identified in Khayelitsha and Mfluleni between the R300 and Spine Road (M32)
  - Several other hot spots are detected in three census suburbs of Table Bay district: Cape Town CBD, Zonnebloem and Woodstock
  - A few noticeable hot spots are evident in Heideveld and Bonteheuwel suburbs between Jakes Gerwel Drive (M17) and Robert Sobukwe Road (M10).
- Hotspot analysis by kernel density estimation (KDE): Hot spots of intersection-related pedestrian casualties are identified:
  - On arterial roads
  - At junctions of arterial roads and urban freeways and
  - On local roads of the CBD of Cape Town.

4.5.3 Results from multivariate analyses

4.5.3.1 Results from Generalised Linear Models (GLMs)

- Negative binomial (NB) models have demonstrated a better prediction performance than Poisson regression models.
- NB Model 1 (i.e. the model fitted to the entire sample of pedestrian casualties) consists of 17 explanatory variables.
  - The following are variables shown to have significant influence (absolute value of the coefficient B greater than 0.03) on the frequency of pedestrian casualties in NB Model 1, in descending order:
    1) Log of population
    2) Entropy index
    3) The proportion of signalised intersections
    4) The proportion of expressways
    5) The proportion of the population younger than 15 years (negative associations)
    6) The proportion of the population in the 25-54 age range (negative associations)
    7) The number of roundabouts and mini-circles
    8) Proportion of freeways.
  - All demographic variables, apart from the logarithm of population, are shown to be negatively associated with the frequency of pedestrian casualties.

- NB Model 2 (i.e. the model fitted to the dataset of intersection-related pedestrian casualties) includes 17 explanatory variables.
    - Variables with significant influence (absolute value of the coefficient B greater than 0.03) on the frequency of pedestrian casualties in NB Model 2 are:
        1) Log of population
        2) Entropy index
        3) The proportion of signalised intersections
        4) The proportion of expressways
        5) The proportion of freeways
        6) The proportion of the population with upper income
        7) The proportion of the population in the 15-24 age range
        8) The proportion of primary arterial roads
        9) The ratio of intersections to culs-de-sacs
        10) The proportion of the population younger than 15 years;
        11) The proportion of secondary arterial roads
        12) The proportion of not working population;
        13) The proportion of local streets.
    - All demographic variables except for the logarithm of population are negatively associated with the frequency of intersection-related pedestrian casualties.

- NB Model 3 (i.e. the model fitted to the dataset of KSI pedestrian casualties) includes 14 explanatory variables.
    - Variables with significant influence (absolute value of the coefficient B greater than 0.03) in NB Model 3 in descending order are:
        1) Log of population
        2) Entropy index;
        3) The proportion of signalised intersections
        4) The proportion of the population aged 55 years and older
        5) The proportion of the population in the 15-24 age range
        6) The proportion of the population with upper income
        7) The proportion of the population younger than 15 years
        8) The proportion of freeways
        9) The proportion of expressways.

- All demographic variables apart from the logarithm of population are negatively associated with the frequency of intersection-related pedestrian casualties.
- Sensitivity analysis of model estimates over different days of week:
    - The following variables have demonstrated a greater influence on the frequency of pedestrian casualties during weekdays than on weekends:
        - Entropy index
        - The general industrial use
        - The number of roundabouts and mini-circles;
        - The proportion of signalised intersections
        - The proportion of expressways
        - Log of population
        - The proportion of the White population
        - The proportion of the population in the 15-24 age range.
    - The following variables have demonstrated a reduced influence on the frequency of pedestrian casualties on weekdays than on weekends:
        - The proportion of the population younger than 15 years old
        - The proportion of the population in the 25-54 age range
        - The number of four- and multi-legged intersections
        - The proportion of primary arterial roads.

4.5.3.2 Results from Geographically Weighted Regression (GWR) Models

- Two GWR models (GWR Model A and GWR Model B) were developed for each of the three datasets of pedestrian casualties.
- The best GWR models (Models A) are dominated by demographic variables as follows:
    - Three out of seven explanatory variables are demographic in GWR Model 1A
    - Two out of seven explanatory variables are demographic in GWR Model 2A
    - Four out of five explanatory variables are demographic in GWR Model 3A.
- The following variables describing the built environment variables are statistically significant in GWR Models A:
    - GWR Model 1A:
        - The proportion of general business use and mixed use
        - The number of four- and multi-legged intersections
        - The number of roundabouts and mini-circles and

        - The proportion of signalised intersections.
    - GWR Model 2A:
        - The proportion of the single residential use
        - The proportion of general business use and mixed use
        - The number of four- and multi-legged
        - The proportion of local streets and
        - The number of roundabouts and mini-circles.
    - GWR Model 3A:
        - The number of roundabouts and mini-circles.
- GWR Models B were developed to assess the influence of land use mix (entropy index) and functional road class
- Generally, demographic variables are found to be negatively associated with the frequency of pedestrian casualties, with the exception of:
    - The proportion of the population younger than 15 years old (GWR Model 1A and GWR Model 1B)
    - Log of population (GWR Model 1B; GWR Model 2A; GWR Model 2B and GWR Model 3B) and
    - The proportion of the Black population (GWR Model 1A; GWR Model 3A).
- Generally, the built environment variables are found to be positively associated with the frequency of pedestrian casualties, apart from:
    - Entropy index (GWR Model 1B)
    - The proportion of primary arterial roads (GWR Model 1B)
    - The number of roundabouts and mini-circles old (GWR Model 1A and GWR Model 1B)
    - The proportion of the single residential use (GWR Model 2A)
    - The proportion of local streets (GWR Model 2A) and
    - The proportion of freeways (GWR Model 3B).
- The following variables have demonstrated local parameter estimates with counterintuitive signs in GWR Models;
    - In GWR Model 1A:
        - The proportion of the population younger than 15 years old
        - The proportion of general business use and mixed use and
        - The proportion of signalised intersections.

- In GWR Model 2A:
  - Log of population and
  - The proportion of the White population.
- In GWR Model 3A:
  - The proportion of the population in the 15-24 age range.
- In GWR Model 3B:
  - Log of population
  - The proportion of freeways
  - The proportion of expressways
  - The proportion of primary arterial roads and
  - The proportion of secondary arterial roads.

- The *t*-test has indicated that spatial heterogeneity is not significant for the following variables:
  - GWR Models 1:
    - The proportion of the Black population and
    - The proportion of signalised intersections.
  - GWR Models 2:
    - Log of population
    - The proportion of general business use and mixed use and
    - The number of roundabouts and mini-circles.
  - GWR Models 3:
    - The proportion of the population in the 25-54 age group
    - The proportion of the Coloured population and
    - The number of four- and multi-legged intersections.

4.5.3.3 Model comparison

- Comparison of model performance:
  - NB Models have shown the best performance in fitting pedestrian casualty data
  - GWR Models emerge as the second best performing models
  - GWR Models perform well in capturing spatial heterogeneity of relationships across the study area.
- Comparisons of model estimates:
  - The following explanatory variables have emerged in all four Models 1(NB Model 1, Poisson Model 1, GWR Model 1A and GWR Model 1B):

  - ✓ The proportion of the population younger than 15 years old
  - ✓ The number of four- and multi-legged intersections
  - ✓ The number of roundabouts and mini-circles and
  - ✓ The proportion of signalised intersections
- Two explanatory variables have emerged in all four Models 2 (NB Model 2, Poisson Model 2, GWR Model 2A and GWR Model 2B):
  - ✓ Log of population and
  - ✓ The number of roundabouts and mini-circles.
- One explanatory variable is found in all four Models 3 (NB Model 3, Poisson Model 3, GWR Model 3A and GWR Model 3B): The number of four- and multi-legged intersections.
- In all three modelling procedures (NB, Poisson Regression , and GWR):
  - ✓ Log of population is found to be significantly associated with an increased number of pedestrian casualties, with the coefficient B ranging from 1.37 to 23.77;
  - ✓ Entropy index has positive associations with the frequency of pedestrian casualties, with the exception of GWR Model 1B;
  - ✓ The variables describing the road network structure are generally positively associated with the number of pedestrian casualties, except for:
    1) The proportion of primary arterial roads in GWR Model 1B
    2) The proportion of local streets in GWR Model 2A and
    3) The proportion of freeways in GWR Model 3B.
  - ✓ The number of roundabouts and mini-circles is found to be positively associated with the frequency of pedestrian casualties, except for GWR Model 1A and GWR Model 1B;
  - ✓ The proportion of signalised intersections has positive associations with the frequency of pedestrian casualties in all three modelling procedures;
  - ✓ The variables describing urban design (i.e. the number of four- and multi-legged intersections, street density and the ratio of intersections to culs-de-sacs) are all positively related to the frequency of pedestrian casualties.
  - ✓ Two variables of land use have positive associations with the frequency of pedestrian casualties:

1) The proportion of general business use and mixed use
2) And the proportion of the general industrial use.

- ✓ The majority of demographic variables are found to be negatively related to the frequency of pedestrian casualties.

4.6 Result discussions

4.6.1 Discussing the results from univariate and bivariate analyses

The results from this study show that an average of 4 618 pedestrians were involved in road traffic crashes each year in the City of Cape Town for the 2012-2014 period. For the entire study area, the annual casualty rate stands at 123.5 pedestrian casualties per 100 000 population. The annual KSI rate reported in this study stands at 35.7 KSI pedestrian casualties per 100 000 population while the annual fatality rate stands at 4.5 pedestrian fatalities per 100 000 population. The figures of KSI and pedestrian fatality rates found in this studies for the City of Cape Town are lower than those usually reported in previous works (e.g. RTMC, 2016, 2017). This discrepancy may be explained by higher levels of injury misclassification among pedestrian casualties in the police-reported crash records

In this study, male pedestrians are found overrepresented in the analysed sample of pedestrian casualties and male dominance is observed in all injury severity types. This is not a new finding as previous studies have consistently reported the same findings (Gårder, 2004; Hunter *et al.*, 1996; Mabunda *et al.*, 2008). The predominance of male pedestrians in road traffic crashes may be attributed in part to their inherent risk taking behaviour. Risk-taking behaviour among male pedestrians has been confirmed in a number of previous studies. For instance, numerous studies that examined pedestrian crossing behaviour found that male pedestrians are more likely to be risk takers when they cross streets than females (Holland & Hill, 2007; Keegan & O'Mahony, 2003; Latrémouille, Thouez, Ranou, Bergeron, Bourbeau & Bussière, 2004; Moyano Díaz, 2002; Rosenbloom, 2009; Rosenbloom, Nemrodov & Barkan, 2004; Tom & Granié, 2011; Yagil, 2000). Other factors that exacerbate male crash risk may be a greater level of alcohol consumption compared with that of females. International research has consistently reported that male pedestrians are overrepresented in alcohol-related crashes (Holubowycz, 1995a,b; Ortiz & Ramnarayan, 2017; Öström & Eriksson, 2001; Prijon & Ermenc, 2009). Research in South Africa has also confirmed that alcohol consumption is more predominant among male pedestrians than female pedestrians (Mabunda *et al.*, 2008; Peden *et al.*, 1996; Van der Spuy, 1991).

An examination of pedestrian casualties by age demonstrates that the highest casualty frequency emerges among child pedestrians in the 6-10 age group. The explanation for this could be the fact that children in this age group tend to travel more independently unlike the younger ones who are more likely to be accompanied by adults when they travel (Peden, Oyegbite, Ozanne-Smith, Hyder, Branche, Rahman, Rivara & Bartolomeos, 2008). Children are also more likely to spend more time away from home and spend more time playing. Children living in socioeconomically disadvantaged areas are particularly predisposed to the greatest crash risk as children may spend more time playing in streets owing to the lack of safe places to play (Toroyan & Peden, 2007). In addition, school-going children in the 6-10 age group often travel independently to school on foot in economically deprived areas. They tend to walk with their friends who are also in the same age range and peer pressure can influence their behavioural and direct them to behave in a risky manner on the road. As novice pedestrians, child pedestrian in this age range lack experience and skills to understand the road environment and this affect their ability to perceive hazards and to respond appropriately to them.

Another reason behind the higher crash risk among child pedestrians is their small stature that limits their visibility in road environment (Wilson, Baker, Teret, Schock & Garbarino, 1991). Furthermore, it is well documented that developmental factors such as limited attention to navigate the road environment, difficulty seeing over vehicles, lack of the knowledge and skills of traffic movement, difficulty in judging vehicular speed and discerning appropriate gaps between traffic streams all have a great impact on children involvement in road crashes (Thomson, Tolmie, Foot & McLaren, 1996; Toroyan & Peden, 2007). When hit by vehicles, children are more likely than adult to sustain a head or neck injury (Peden *et al.*, 2008; Wilson *et al.*, 1991), which suggests a greater likelihood of more severe injury.

The study results reveal a dip in casualty profile among children aged between 11 and 20 years old for both females and males. The dip in the same age group is also apparent when casualties are analysed by considering the injury severity. This dip arguably has its origin in the underlying population pyramid of Cape Town created using the 2011 population census data (see Figure 2-6 on Page 43). The analysis of pedestrian casualties by gender also demonstrates another dip for female casualties in the 36-45 age group. Two possible reasons that could explain this finding may be a reduced exposure to traffic environment or/ and lower level of risk-taking behaviour among females of this age group. However, research into levels of exposure and age differences in risk-taking behaviour particularly among female pedestrians is inexistent in South Africa in order to substantiate these claims.

The time of day that pedestrian crashes are most likely to occur are found to be morning peak traffic hours (7:00 AM to 8:00 AM) and evening peak traffic hours (4:00 PM to 6:00 PM).

There is no surprise for this finding as these time periods host the heaviest traffic volumes and highest levels of pedestrian flow rates. The finding is thus indicative of the influence of traffic and pedestrian volumes on the incidence of pedestrian crashes, which has been reported by many researchers (e.g. Wier *et al.*, 2009; Miranda-Moreno, Morency and El-Geneidy, 2011; Yao, Loo and Lam, 2015; Zhang *et al.*, 2015).

The results also demonstrate a peak of pedestrian casualties on Fridays and Saturdays and the lowest frequency of pedestrian casualties on Wednesdays. For fatally injured casualties, the peak emerges over the weekend (Saturdays and Sundays) and this is line with the findings from a number of previous local studies (Burstein, Fauteux-Lamarre & As, 2016; Mabunda *et al.*, 2008). One crucial contributing factor reported in many studies is increased travel speeds over the weekend due mainly to less traffic on the road. Tom Tom historical traffic data accessed through Stellenbosch Smart Mobility Lab shows that average speeds on the Cape Town's road network are generally higher on weekends than weekdays. Moreover, alcohol impairment among both motorists and pedestrians has been pointed out in several studies as the predominant reason of higher rates of crashes involving pedestrians. Research in road safety acknowledges a correlation between alcohol and speeding: several studies found that alcohol-impaired motorists are more likely to be involved in speeding-related crashes (Liu, Chen, Subramanian & Utter, 2005; Road Safety Directorate, 2008).

The results of this study also demonstrate variations in the frequency of pedestrian casualties across the weeks of a month. Pedestrian casualties are found most likely to occur during the pay week (i.e. week contains the first data on a month) and during the following week (second week after pay week). The explanation of these trends could be that pedestrian activity is the highest during pay week. Survey data in Cape Town indicates that walking is the main mode of transport for low- and middle-income households, accounting for 61 percent and 43 percent, respectively (Behrens, 2002). Of these walking trips, the vast majority are undertaken for shopping, social and recreational activities. Behrens (2002) reported that 73 percent and 47 percent of all shopping trips are made on foot by low-income households and middle-income households, respectively. For social activities, 92 percent are made on foot by low-income households and 53 percent of trips are undertaken by foot by middle-income households. Finally, 73 percent and 44 percent of all trips done for recreational purposes are made on foot by low-income households and middle-income households, respectively (Behrens, 2002). Although not supported by research evidence, it may not be wrong to expect that the pay week hosts a greater number of walking trips made for leisure, social and shopping activities, increasing thus pedestrian exposure to traffic environment. Higher pedestrian crash risk could be also influenced by alcohol consumption as drinking behaviour may be related to financial status (World Health Organization, 2011).

The examination of pedestrian behaviour and actions prior to a road crash reveals that the vast majority (88 percent) of pedestrians were involved in road crashes while being outside of designated crossing points. Only 5.7 percent of pedestrians were at designated crossing points prior to the crash occurrence and 4.3 percent were within 50 metres of a pedestrian crossing facility. These findings highlight a significant role played by pedestrian behaviour (i.e. spatial complaint behaviour) in the incidence of pedestrian crashes. Higher frequencies of pedestrian crashes at non-designated crossing points could be attributed to higher levels of spatial non-compliant behaviour which has been reported in several studies conducted in South Africa (Behrens, 2010; Nteziyaremye & Sinclair, 2013; Ribbens, 1996). Several qualitative studies pointed out a number of motivations that determine unsafe pedestrian crossing/walking choices, including time saving, distance saving (i.e. shortest route), fear from crime, minimal effort, habits, traffic volumes, presence of alternative choice, effects of group dynamics (e.g. crossing where most people cross) and location of pedestrian desire lines (Behrens, 2010; Sinclair and Zuidgeest, 2016; Behrens and Makajuma, 2017). However, it is important to recognise that unsafe crossing behaviour is sometimes enabled by a lack of alternative facilities (i.e. absence of crossing facilities) and poorly located or insufficient pedestrian crossing facilities. For instance, pedestrians are required to walk longer distances of about 5 kilometres to reach the nearest grade-separated pedestrian facility at certain locations on freeway facilities in Cape Town (Sinclair and Zuidgeest, 2016).

The results of this study also highlight a number of crossing styles that predispose pedestrians to a higher crash risk. A significant proportion of pedestrians were hit by vehicles while running (25 percent) or standing in the middle of the roadway to find an appropriate gap (7.9 percent). These are common crossing strategies among pedestrians and similar findings are reported in a small number of previous studies. For instance, an observational study on pedestrian crossing behaviour in Stellenbosch (South Africa) by way of videotaped images revealed that crossing the road while running and standing in the middle of the roadway to select a gap in the traffic stream are common crossing styles adopted by many pedestrians, particularly at signalised and wider intersections with four-lane approaches (Nteziyaremye & Sinclair, 2013).

With respect to pedestrian casualty locations, the results from univariate analyses demonstrate that higher rates of pedestrian casualties are found at four-legged intersections, roundabouts and mini-circles, signalised intersections and intersections controlled by the four-way stop signs, when pedestrian casualties are normalised against the number of these facilities. With regard to injury severity, more severe pedestrian injuries are most frequently observed at signalised intersections and intersections controlled by the Yield sign. However, the finding regarding the influence of the Yield sign is based on a very small number of cases (16 casualties only) reported at these intersections and the paucity of casualty data at this intersection type may affect the reliability of this finding. Nevertheless, the fact that motorists are not forced to stop completely at an intersection controlled by the Yield sign may predispose pedestrians to more severe injuries if they collide with vehicles.

4.6.2 Discussing the results from multivariate analyses

The model results confirm the findings of univariate and bivariate analyses, with the number of four- and multi-legged intersections, the number of roundabouts and mini-circles, and the proportion of signalised intersections all emerging as significant predictors of pedestrian casualties in almost all the developed models.

Pedestrian safety at roundabouts has not been studied extensively locally as well as internationally. Praises attributed to the modern roundabout as being safer than other methods of intersection control originate from a number of studies that included a very small sample of pedestrian crash data or no pedestrian crash data at all (Retting, Persaud, Garder & Lord, 2001; Robinson, Rodegerdts, Scarborough, Kittelson, Troutbeck, Brilon, Bondzio, Courage, Kyte, Mason & Flannery, 2000). Improved pedestrian safety associated with the modern roundabout is often based on reduced number of pedestrian-vehicle conflicts and reduced vehicular speed (Stone, Chae & Pillalamarri, 2002). Although few studies confirmed that the modern roundabout is safer for pedestrians (Jordan, 1985; Ulf & Jorgen, 1999) this claim is not always supported, especially in the South African context as it has been shown in this study.

A number of factors could explain a greater crash risk for pedestrians at roundabouts and mini-circles. When approaching these facilities, the attention of left-turning motorists is much focused to the right to judge an appropriate gap in order to merge into the traffic stream, and this could lead to a failure to detect pedestrians who are crossing from the left side of the approach. Navigating roundabouts and mini-circles can be a challenging task for pedestrians as traffic is always moving and pedestrians have to choose an appropriate gap in the traffic stream. Confusion about the right-of-way at roundabouts may also cause conflicts between pedestrian and motorists. Non-yielding behaviour of motorists may also be a cause of frustration and can lead to longer delays for pedestrians. Pedestrian crosswalks at the exit points of roundabouts or traffic circles are positioned sometimes in a way that enables exiting motorists to obstruct the traffic movement in the circulatory roadway when yielding to pedestrians. The design of roundabout exits can thus discourage drivers from yielding to pedestrians who are crossing at roundabouts or mini-circles. In addition, roundabouts and mini-circles pose safety and comfort problems to visually impaired pedestrians who may find it more difficult to choose a gap by judging traffic sound (Stone *et al.*, 2002). In conclusion, roundabouts and mini-circles are proved to be safer for motorists but pedestrian safety may be compromised at these facilities.

Another important finding from this study is the influence of the number of four- and multi-legged intersections on the frequency of pedestrian casualties. The number of possible traffic conflicts between motorists and pedestrians increases for each increase of intersection approach, suggesting that pedestrian exposure to risk is greater at intersections with more approaches. This may be the possible reason behind the higher crash risk observed at four-legged and multi-legged intersections. Another possible reason that may explain this finding is the higher intensity of pedestrian activity in locations with greater concentrations of these type of intersections. The findings of this study with regard to the contribution of the number of four-legged and multi-legged intersections to the frequency of pedestrian casualties are consistent with findings reported in previous studies (Dumbaugh & Li, 2010; Gårder, 2004; Ukkusuri *et al.*, 2012; Zhang *et al.*, 2015).

The results from multivariate analyses have indicated that the proportion of signalised intersections is associated with increased numbers of pedestrian casualties. A number of factors may explain these findings. Traffic signals are usually provided based on a number of factors related to the existing operation and safety at a particular intersection. According to the South African Road Traffic Signs Manual, traffic signals are provided based mainly on the queue length warrant (Department of Transport, 2012). The queue length is indirectly indicative of the presence of high traffic volumes and pedestrian flows as well as longer delays caused to road users at an intersection. Simply put, traffic signals are mainly warranted at intersections where heavy traffic and pedestrian flows exist. The installation of traffic signals at an intersection is aimed at primarily improving traffic flow and facilitating access by distributing priority amongst road users (Department of Transport, 2012). It does not always guarantee an increase in road safety and this is emphasised in the South African Road Traffic Signs Manual. Although the effect of traffic volume on pedestrian safety was not directly investigated in this study, there is evidence in the reviewed literature that traffic volume is an important risk factor for pedestrian crashes (Wier *et al.*, 2009; Miranda-Moreno *et al.*, 2011; Zhang *et al.*, 2015). The presence of heavy traffic volumes and pedestrian flows at signalised intersections increases pedestrian exposure to risk which may lead to the deterioration of pedestrian safety. Traffic signals are often installed at wide intersections with multi-lane approaches to accommodate high traffic volumes. To negotiate these facilities, pedestrians are often exposed to longer crossing distances and longer waiting times. When waiting time increases, pedestrians are more likely to take chances, increasing the risk of a crash occurring.

Behavioural aspects of both pedestrian and motorists also contribute significantly to pedestrian crash occurrence at signalised intersections. It is a common occurrence in South Africa that turning motorists fail to yield to pedestrians who are crossing at signalised intersections during the pedestrian green time. Non-yielding behaviour of turning motorist has been reported in several studies as being the motive of both spatial and temporal non-complying behaviour among pedestrians. For instance, on-street personal surveys conducted in Stellenbosch in the Western Cape Province revealed that non-yielding behaviour of turning motorists is among the motives for crossing outside a designated crossing point and for violating the red man signal (Nteziyaremye, 2013). Speeding is another driver-related unsafe behaviour that can influence the incidence and severity of pedestrian crashes. The South African Road Traffic Signs Manual stipulates that speed limit on any approach of a signalised intersection or signalised pedestrian/cyclist crossing shall not exceed 80 km/h. However, vehicular speeds higher than this speed limit are often observed at signalised facilities where pedestrian flows exist.

Pedestrian unsafe crossing behaviour at signalised intersections and signalised mid-block crossings are also an important factor contributing to higher crash rates observed at these locations. In South Africa, non-compliance with traffic signals is very common among pedestrians. It is reported that more than 80 percent of pedestrians crossing at signalised facilities do not comply with traffic signals (Nteziyaremye & Sinclair, 2013). Temporal non-complying behaviour is influenced by a number of motives including among others, low traffic volumes (i.e. no traffic is present), time saving, beliefs that it is safe to cross when vehicles are stopped, failure to notice pedestrian signals at intersections, driver non-yielding behaviour, mistrust of pedestrian signals (e.g. green man signal is too short), longer waiting times, perception that pedestrian signal controls are defective (e.g. pushbutton are not working) and familiarity and experience with crossing during red man signal (Nteziyaremye, 2013).

Perception of comfort and safety together with the actual crash risk observed at signalised intersection in this study are compelling evidence that the installation of traffic signals improve the efficiency and capacity of the intersection but on the other hand it leads to diminished safety for pedestrians.

The model results have demonstrated two types of land use that are positively related to the frequency of pedestrian casualties. These are the general industrial use (GI) and a combination of general business (GB) and mixed use (MU). These findings suggest that areas with greater intensity of these two type of land use are likely to experience more pedestrian crashes. However, one land use type, the single residential use (SR) is found to be negatively associated with intersection-related pedestrian casualties using the GWR modelling technique.

According to the Cape Town Planning By-Law, the general business use includes business premises, places of instructions, place of assembly, place of entertainment, hotel, conference facility, service trades, authority use, utility services, and shopping malls, among others. The industrial use is ascribed to any property developed primarily for manufacturing and related processes. The mixed land use (as a type of land use) is simply defined as a mixture of business, industrial and residential development either horizontally or vertically (e.g. multi-storey developments consisting of ground level retail/business and residential use above it). Industrial use applies for instance to industry, restaurants, service stations, motor repair garage, scrap yard, agricultural industry and so on (City of Cape Town, 2015b). Areas with greater intensity of these types of land use are more likely to attract many walking trips, increasing thus the intensity of pedestrian activity in these locations. Therefore, a greater intensity of pedestrian activity and more walking trips attracted in areas with greater intensity of industrial, business and mixed use are the underlying cause of pedestrian crashes.

With regard to the influence of land use on the frequency of pedestrian crashes, similar findings to those reported in this study exist in literature. For instance, a number of studies have found that a number of land use types including business, the number of megastores, retail or commercial industrial use, are associated with elevated numbers of pedestrian crashes (Dumbaugh & Li, 2010; Kim *et al.*, 2010; Ukkusuri *et al.*, 2011, 2012, Wedagama *et al.*, 2008, 2006; Wier *et al.*, 2009; Zhang *et al.*, 2015).

The single residential use (SR), according to the Cape Town Planning By-Law, encompasses two categories of land use based on socio-economic circumstances of the city: the conventional housing (SR1) and the incremental housing (SR2). The conventional housing includes primarily single-family dwelling houses located in low- to medium-density neighbourhoods providing a safe and pleasant living environment. On the other hand, the incremental housing land use is ascribed to dwellings located in an informal settlement or those in an area that has been upgraded from an informal to a formal settlement, but the upgrading has not yet reached an appropriate stage required to be rezoned to SR1 or another appropriate zoning (City of Cape Town, 2015b).

Usually, land use and urban activities create opportunities for land use-transport interaction (Rietveld & Bruinsma, 1998). This interaction can happen at two levels: between zones (i.e. inter-zonal) or within a zone (i.e. intra-zonal) (Wedagama *et al.*, 2008). There is a possibility that areas with a greater intensity of the residential use tend to be mono-functional neighbourhoods. A census suburb with predominantly residential use creates opportunity for inter-zonal trips. Inter-zonal trips tend to cover long distance trips which require the use of motorised modes (mainly private cars or public transport). The negative association between the residential land use and the frequency of pedestrian casualties suggests that there is a lower likelihood of experiencing pedestrian crashes in areas dominated by the residential land use. This is perhaps due to fewer walking trips between zones (i.e. inter-zonal walking trips) and lesser pedestrian activity in these areas. Another possible explanation of this finding may be a lower density of high-class roads (i.e. high-speed roads) within residential neighbourhoods, suggesting these areas attract low traffic volumes and motorists in these areas travel at lower speeds.

Similar findings on the influence of the single residential use on the frequency of pedestrian crashes are reported in a small number of previous studies (e.g. Pulugurtha *et al.*, 2013; Ukkusuri *et al.*, 2012) while contradicting findings (i.e. positive associations between the single residential use and the incidence of pedestrian crashes) are reported in many other studies (e.g. Amoh-Gyimah *et al.*, 2016; Guo *et al.*, 2017; Loukaitou-Sideris *et al.*, 2007; Siddiqui *et al.*, 2012; Wier *et al.*, 2009). Mixed results suggest that the direction of the association (negative or positive) between the single residential use and the frequency of pedestrian casualties may be context-specific.

Land use mix measured by the entropy index emerged among the explanatory variables which are strongly associated with increased numbers of pedestrian casualties. This finding is quite consistent with those reported by many previous studies (Amoh-Gyimah *et al.*, 2016; Guo *et al.*, 2017; Pulugurtha *et al.*, 2013; Tian, Ewing, White, Hamidi, Walters, Goates & Joyce, 2015; Zahabi, Strauss, Manaugh & Miranda-Moreno, 2011). The literature indicates that a greater level of land use mix tends to reduce trips length, subsequently encouraging walking trips and the use of other non-motorised modes (Cervero & Kockelman, 1997a; Clifton, Ewing, Knaap & Song, 2008; Ewing, Greenwald, *et al.*, 2011; Frank, Greenwald, Kavage & Devlin, 2011; Gehrke & Clifton, 2017; Handy, 2005; Tian *et al.*, 2015). Moreover, land use mix promotes public transport use (Cervero & Kockelman, 1997b; Ewing, Greenwald, *et al.*, 2011; Tian *et al.*, 2015) which is often used in conjunction with walking trips. In light of this research, strong associations between land use mix and the frequency of pedestrian casualties observed may be attributed to increased numbers of walking trips or a greater level of pedestrian activity which in turns leads to higher pedestrian exposure to risk. It is important to note that this finding does not imply that pedestrian crashes are caused by a mix of land use types – it is likely that the cause itself is the increased pedestrian activity associated with a mixture of land use types.

Looking at the magnitude of the associations revealed by Generalized Linear Models (GLM), the results demonstrate a reduced contribution of land use mix on the frequency of KSI pedestrian casualties (coefficient B=0.885 in NB Model 3) compared with that for the entire sample (coefficient B=1.158 in NB Model 1) and that of intersection-related pedestrian casualties (coefficient B=1.610 in NB Model 2). These findings suggest that land use mix is still associated with increased numbers of KSI pedestrian casualties but the magnitude of this association is less marked compared with that observed for the entire sample and intersection-related pedestrian casualties. Again, a higher degree of land use mix is indicative of greater pedestrian activity which is also a factor related to exposure variables such as traffic and pedestrian volumes. As traffic volumes and pedestrian numbers increase, speed decreases and the amount of kinetic energy transferred during a vehicle-pedestrian crash is less likely to cause serious or fatal injuries. Therefore, the reduced contribution of land use mix on the frequency of KSI pedestrian casualties could be explained by lower vehicular speeds in areas of greater intensity of pedestrian activity. The findings on the influence of land use mix on pedestrian injury severity are consistent with those reported in the study by Amoh-Gyimah *et al*. (2016) who found that .pedestrian crash risk is higher in urban areas with a greater level of land use mix but the likelihood for a pedestrian crash to result in more severe injuries is somewhat reduced.

A cross-comparison of estimates from the three models developed using separate datasets of pedestrian casualties that occurred on weekdays (Model 4), Saturdays (Model 5) and Sundays (Model 6) demonstrate temporal variations of the influence of land use mix on the frequency of pedestrian casualties. As land use mix and other attributes of the built environment are not subjected to daily variations, varying associations may be attributed to other confounding variables which are not captured in the link between the built environment and pedestrian crashes. The most plausible variables which may be at play are exposure variables, namely traffic volume, pedestrian activity and vehicular speed. Hence, daily variations in traffic volumes, pedestrian activity and vehicular speed may be the underlying cause of varying contribution of land use mix to pedestrian casualties. Consequently, it can be presumed that traffic volumes, pedestrian activity and vehicular speed are the main exposure variables that play the mediating role in the link between the built environment and pedestrian crashes. The interpretation given to this finding affirms the conceptual framework adopted in this study and is supported by several other studies (Ewing & Dumbaugh, 2009; Miranda-Moreno *et al.*, 2011; Stoker, Garfinkel-Castro, Khayesi, Odero, Mwangi, Peden & Ewing, 2015; Ukkusuri *et al.*, 2012).

As it was expected, variables describing the road network structure (i.e. classes of road) are generally shown to have positive associations with the frequency of pedestrian casualties. In the majority of the developed models, two types of road class- urban freeways and arterial roads- demonstrate significant positive associations with the frequency of pedestrian casualties. These findings are consistent with those reported in many other studies (Dumbaugh & Li, 2010; Gårder, 2004; Mohan *et al.*, 2017; Theofilatos, Yannis, Kopelias & Papadimitriou, 2016; Ukkusuri *et al.*, 2012; Wier *et al.*, 2009). The underlying reason behind the higher pedestrian crash risk may be higher levels of pedestrian exposure in terms of traffic volumes, pedestrian flows and vehicular speed.

The variables describing the population characteristics have also emerged significant in all developed models. This is not a surprising finding as the population number is a potential proxy for pedestrian activity and is often regarded as an important risk factor for pedestrian crashes. This justification is consistent with the positive association shown by the percentage of workers. High risk is expected for commuters since they tend to travel more often, which elevates their risk exposure in the road environment. These findings are consistent with those reported in previous research (Amoh-Gyimah *et al*., 2016; Lee, Abdel-Aty & Jiang, 2015; Miranda-Moreno *et al*., 2011; Wang, Yang, Lee, Ji & You, 2016; Wier *et al*., 2009).

The results also show that high-income areas are more likely to experience fewer pedestrian casualties. Given the historical context of South Africa, it is not surprising that race was related to pedestrian casualties since it is often regarded as a proxy for socio-economic status. The Apartheid policies of land use planning forced the poor people (mostly the Back population) to live in sprawling and overcrowded settlements deprived of economic opportunities and basic services (Turok, 1994). Vehicle ownership in these areas is lower compared to that of wealthier ones. This explains why residents of the poorer communities are the most affected by pedestrian crashes as a large number of their trips are made on foot in road environments with inadequate walking facilities. The findings on the effect of socio-economic deprivation are consistent with those reported by previous research studies (Cottrill & Thakuriah, 2010; Graham & Glaister, 2003; Loukaitou-Sideris, Liggett & Sung, 2007; Siddiqui, Abdel-Aty & Choi, 2012).

Chapter 5: Conclusions

This chapter provides a summary of key findings, the original contribution of the study, transferability of model results, the practical implications of the study, the limitations that challenged the development of this study, and considerations for future research.

5.1 Key findings of the study

This section presents the key findings from this study and these are presented with reference to the research questions explored in this study. The main findings of the study are summarised below:

1) The variety of analytical methods applied in this studies demonstrated that pedestrian crashes are related to the attributes of the built environment and population characteristics. This finding responds to the first research question investigated in this study which was concerned with finding a measurable link between the built environment and pedestrian crashes ("Is there a measurable link between the built environment and pedestrian crashes?"). It has been found in this study that a number of variables describing population characteristics, land use patterns, urban design, and transportation systems are associated with the incidence of the three types of pedestrian casualties included in the analysis (the entire sample of pedestrian casualties, intersection-related and KSI pedestrian casualties). The revealed relationships can be described in four components:

   i. <u>The relationships between population characteristics and pedestrian crashes</u>
   The incidence of pedestrian crashes and the resulting injury severity are significantly influenced by six population characteristics (socio-demographic and socio-economic variables), which are population number, age, income level, education level, employment status and race. All these socio-economic factors are interrelated, resulting in comparable findings. A noteworthy comment about the influence of race in the context of South Africa is that race is related to socio-economic status and can used as an indicator of socio-economic status. The four population groups identified as being at greater crash risk are male pedestrians, child pedestrians (younger than 10 years old), middle-aged pedestrians (especially those in the 26-35 age group), and pedestrians living in socio-economically disadvantaged areas.

ii. <u>The relationships between land use patterns and pedestrian crashes</u>

The study has revealed powerful associations between land use mix and pedestrian crashes. These statistical associations imply that land use mix increases pedestrian activities thereby increasing pedestrian exposure to road crashes. The magnitude of the statistical associations between pedestrian crashes and land use mix infers that the latter may be used as a crude proxy of pedestrian activity or pedestrian volumes for the study area. In addition, positive associations were found between pedestrian casualty counts and two types of land use: general industrial use; and a combination of general business and mixed use. In summary, the findings suggest that higher frequencies of pedestrian crashes are experienced in areas with higher levels of land use mix and areas with greater intensity of business use or industrial use.

iii. <u>The relationships between urban design features and pedestrian crashes</u>

Three proxy variables of urban design have showed positive associations with the frequency of pedestrian casualties. These are street density, intersections having at least four legs and the ratio of intersections to culs-de-sacs.

iv. <u>The relationships between transportation system features and pedestrian crashes</u>

Functional road class, the type of intersection configuration and the type of intersection control have all showed significant influence on the incidence and the severity of pedestrian crashes. More specifically, elements of the transportation system including urban freeways and arterial roads, roundabouts and mini-circles, intersections having at least four approaches as well as intersections controlled by traffic signals were found to be associated with higher frequencies of pedestrian crashes.

The revealed statistical relationships are indicative of the existence of a measurable link between the built environment and pedestrian crashes, confirming the thesis investigated in this study. The first and second research questions have been addressed through a thorough description of the relationships (i.e. nature and magnitude of associations) between the built environment and pedestrian crashes.

2) While the associations investigated through the use of Generalised Linear Models (GLM) are static (i.e. homogenous), those explored through the use of Geographically Weighted Regression (GWR) showed geographical variations across the study area. These variations

were in terms of direction (i.e. positive or negative associations) and intensity. Further analysis by the use of *t*-test successfully highlighted variables subjected to significant spatial variations (i.e. spatial heterogeneity of associations) and those that are not. The third research question ("If the link exist, does it vary across space?") investigated in this study has been addressed through the examination of spatial heterogeneity of the associations. Generally, the findings suggest that the link between the built environment and pedestrian crashes is subjected to spatial variations in both direction and intensity.

3) The findings on characteristics of pedestrian casualties in Cape Town (i.e. the fourth research question investigated in this study) are summarised in four components: casualty trends, behavioural aspects, temporal factors and locations of pedestrian crashes:

    i. An annual casualty rate of 123.5 pedestrian casualties per 100,000 population was found for the entire study area. The annual KSI rate found in this study stands at 35.7 KSI pedestrian casualties per 100,000 population and the annual fatality rate was 4.5 pedestrian fatalities per 100,000 population. The figures of KSI and pedestrian fatality rates found in this studies for the City of Cape Town are lower than those reported in previous works, suggesting that injury misclassification may be the underlying cause of the discrepancy in pedestrian casualty rates.

    ii. In addition to population characteristics and the attributes of the built environment explored through univariate, bivariate and multivariate analyses, the study shed light on behavioural and temporal factors that contribute to the incidence of pedestrian crashes. With respects to pedestrian behavioural aspects, spatial non-compliant behaviour (i.e. crossing outside a designated pedestrian crossing point and within 50 metres from a designated crossing point) was found to be the major contributing factor to pedestrian crashes. Running and midway standing emerged as unsafe crossing styles that were contributory factors in a significant number of pedestrian crashes. Moreover, the incidence of a number of pedestrian crashes was found to be linked to certain pedestrian activities taking place in the road environment, such as playing and working.

    iii. Concerning temporal patterns, the findings demonstrated that the time of day, the day of week, the week of month and the quarter of year have all a significant influence on the incidence of pedestrian crashes. The frequency of pedestrian casualties was the highest during traffic peak times, on Fridays, Saturdays and Sundays, during the pay week, and over the third quarter of calendar year.

iv. Looking at pedestrian casualty locations on the Cape Town's transportation system, approximately three quarters of all pedestrian casualties were identified at non-intersection locations (i.e. links) and 26 percent took places at intersections. Two intersection configuration types were found to be associated with the highest casualty rates. These are four-legged intersections and roundabouts/mini-circles. Similarly to multivariate analyses, univariate analysis also highlighted traffic signals as the intersection control type significantly associated with higher frequencies of pedestrian casualties.

v. Briefly, pedestrian crashes were found more likely to occur: in socioeconomically disadvantaged areas which accommodate higher proportions of poorer population, and where many trips are made on foot; at midblock locations; at signalised intersections; at four-legged intersections; and roundabouts/ mini-circles locations. These findings have addressed the fifth research question investigated in this study ("Where pedestrian crashes are more likely to occur?").

4) The study investigated the sixth research question ("Where are hot spots for pedestrian crashes located in the study area?") by applying a variety of geospatial analysis methods. Geospatial analysis methods that used polygons (i.e. census suburbs in this study) as *Incident Data Aggregation Method* were applied to the entire sample of pedestrian casualties. These methods highlighted a number of census suburbs considered as hot spots of pedestrian casualties. The regions identified as hot spots of pedestrian casualties are the districts of Khayelitsha, Mitchell's Plain, Cape Flats and certain census suburbs of Tygerberg district. The same regions and a number of census suburbs of Table Bay district were also identified as hot spots of intersection-related pedestrian casualties. Furthermore, other geospatial analysis techniques that use point density detected individual hot spots of intersection-related pedestrian casualties. These hot spots are mainly located on arterial roads, at junctions of arterial roads and urban freeways and on local roads of the CBD of Cape Town.

5) Geospatial analysis techniques applied in this study to detect clusters of pedestrian casualties across the study area include the three local statistics of spatial autocorrelation

(the Anselin Local Moran's I, the Getis-Ord Gi* and the Optimized Hot Spot analysis) and the planar kernel density estimation (KDE). Although the three local statistics of spatial autocorrelation use different methods of *Incident Data Aggregation*, they generally produced comparable results on high-risk locations for pedestrian crashes. Looking at the size of the identified hot spot regions, the largest hot spot region was detected by the OHA which uses the fishnet method, followed by the Getis-Ord Gi* and lastly the Anselin Local Moran's I. However, the latter technique showed one advantage over other techniques –the ability to identify outliers (High-Low outliers and Low-High outliers). The Prediction Accuracy Index (PAI) was used to assess the performance of three geospatial analysis tools (OHA that uses density surface, OHA that uses fishnet aggregation methods and KDE) applied to the dataset of intersection-related pedestrian casualties. The KDE with 400 m bandwidth was confirmed to be the best performing tool to detect hot spots of pedestrian casualties in this study.

6) The goodness-of-fit measures qualified the Negative Binomial regression modelling as the best performing modelling procedure in fitting pedestrian casualty data. The Geographically Weighted Regression (GWR) modelling came out at second place in terms of performance. However, the GWR models proved to be a useful tool for their particular ability to detect spatial variations of relationships between the explanatory variables and pedestrian casualties, and to draw conclusion concerning local determinants of the incidence of pedestrian crashes. The existence of over-dispersion in the pedestrian casualty data rendered the Poisson regression modelling inappropriate to represent the data. Overall, the analytical methods applied in this study for road safety investigations have produced consistent results, and the conclusion can be drawn that they are generally suitable for the context of South Africa, with the exception of Poisson regression modelling. This inference addresses the seventh research question investigated in this study.

7) The findings from this study have been discussed in reference to the existing literature relevant to the research questions. In the 'Result Discussion' section, parallels between the findings of this study and those reported in the reviewed previous works have been drawn. The similarity of findings has been highlighted and a discussion of contrasting findings have been provided. This was the concern of the last research question investigated in this study. The general conclusion on this point is that despite certain contextual differences

regarding the determinants of pedestrian safety, the study produced findings that are in line with those in the existing international literature.

5.2 Original contributions of the study

This research is novel in the context of South Africa in four aspects. Firstly, the study developed a methodology of improving and supplementing poor quality secondary data on pedestrian crashes. Secondly, the study applied a wide variety of geospatial and statistical modelling methods to validate the results and to test the performance of the methods (the appropriateness of the methods to the South African context). Thirdly, the study included a large number of variables describing the built environment in the safety analysis. This type of research is scarce around the world, especially in South Africa. Lastly and most importantly, the study proposes new predictive models which are useful in a twofold way: (a) the models are useful in understanding the nature and the magnitude of relationships between the built environment and pedestrian crashes; (b) the models can be used to predict future pedestrian crashes using information that is easily available at the city level.

5.3 Transferability of model results

The proposed models, when tested and calibrated locally, are expected to be replicable in other South African cities and other developing countries, as the fundamental relationships between land use, intersection design and roadway design are all proxies for pedestrian activity which is a direct cause of pedestrian crashes. The models utilised information that is easily available in most cities across the world, where city planning offices are functioning. They remove the need for traffic flow data and for the more difficult collection of pedestrian flows and volumes. They allow for the identification of pedestrian crash hotspots at the disaggregated level of the census suburb, which is again the unit of analysis that most cities are familiar with. More work is needed to test the easy replicability of these models within South Africa and Africa, but for now they offer a method of pedestrian risk assessment that relies solely on data on the attributes of the built environment and population characteristics. However, it is worth noting that GWR models provide a set of local estimates that are specific to each geographic unit. As a result, the GWR models are not spatially transferable, suggesting that cities need to develop their own models using local data on road crashes and the built environment.

5.4 Practical implications of the study

This study has enhanced the understanding of the influence of attributes of the built environment on pedestrian safety in urban areas. The role of aspects including population characteristics, land use patterns, intersection design and control types as well as road structure have been explored through a variety of analytical methods with statistical modelling being the most essential method. The new insights from this study have significant and long-lasting implications for the practice in South Africa and other developing countries. The findings of this study highlight priority areas that needs urgent attention to make walking safer in urban environments. This goal is in accordance with the commitments of the National Road Safety Strategy (NRSS) adopted in South Africa for the 2016-2030 period. Accordingly, the practical implications of the study are presented in this section with reference to priority areas identified in the NRSS 2016-2030 for interventions.

The NRSS 2016-2030 adopted the principles of the Safe Systems Approach and was developed in line with commitments and a framework of actions proposed in the *United Nations Decade of Action for Road Safety 2011-2020 (UNDA)* (Department of Transport, 2014). The UNDA framework for actions consists of the five pillars that guide strategic plans and activities over the Decade of Action. The five pillars include: (1) Road safety management; (2) Safer roads and mobility; (3) Safer vehicles; (4) Safer road users; and (5) Post-crash response (World Health Organization, 2013). The strategies of the NRSS which are in connection with the findings of this study are outlined in red frames in Figure 5-1. The way this study can assist with the achievement of Pillar 1, 2 and 4 is addressed in the section that follows.

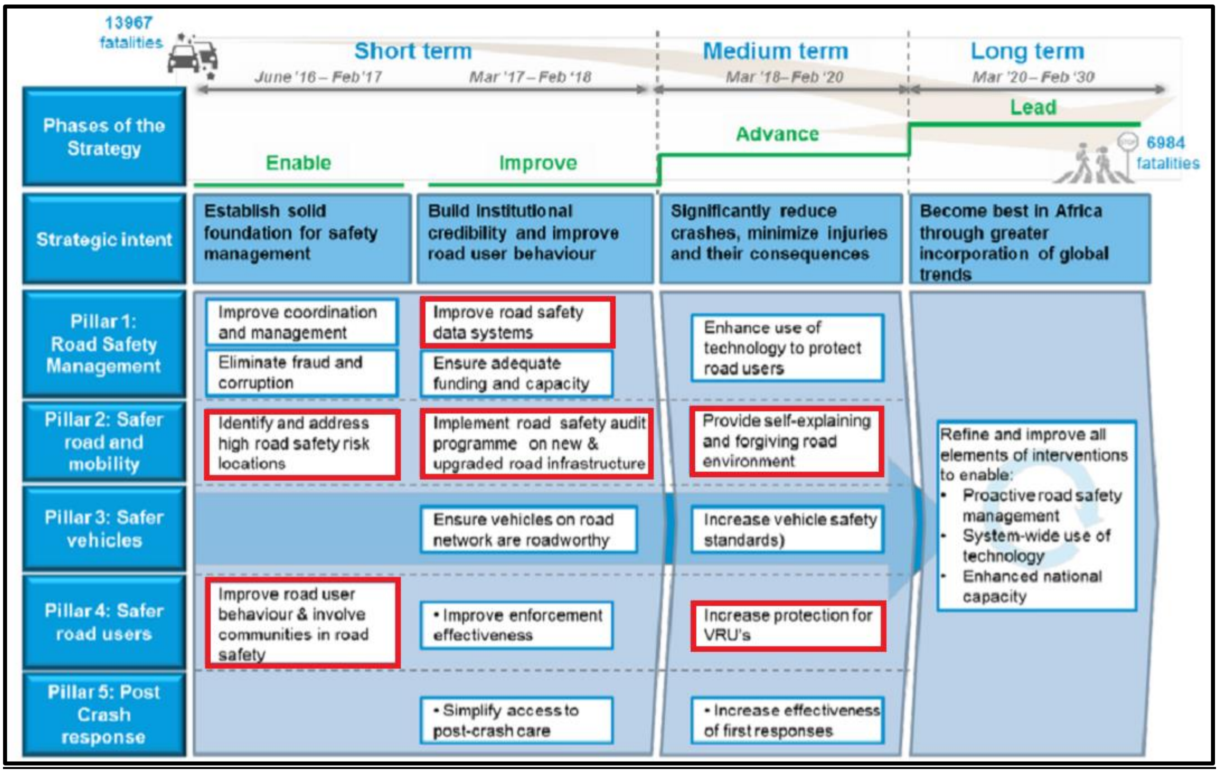


Figure 5-1: Strategic plan for interventions adopted in the South African National Road Safety Strategy 2016-2030 (Department of Transport, 2014)

5.4.1 Pillar 1: Road Safety Management: Improve road safety data systems

Limitations in the National Crash Data are one of the eight key challenges identified in the NRSS under Pillar 1. This study created an opportunity to explore one type of crash data system, which is police-reported crash database. Throughout the investigation of the research questions, the study identified a number of crash data deficiencies that need to be addressed. These deficiencies include duplication of records, missing records, inaccurate records, imprecise crash locations and omission of other crash records critical for the establishment of trends relating to pedestrian crashes and other types of crashes as well. It is therefore crucial to direct interventions on these data limitations to improve both the quality and management of crash data. Improved crash data quality would foster research in road safety as crash data limitations are often a source of frustrations among road safety researchers in South Africa. The research outcomes of this study came from an extensive effort geared towards improving the existing crash database and have demonstrated that high-quality crash database holds tremendous benefits for local research in road safety, countermeasure design and monitoring of interventions.

5.4.2 Pillar 2: Safer road and mobility:

The safer roads and mobility pillar focus on road design and the environment that offer protection to all road users. As they are the most vulnerable road users, pedestrians require special protection in road environments. To achieve safer road and pedestrian movement, road design practices should give priority to pedestrian needs and should ensure that risks faced by pedestrians in road environments are minimised. In addition, high-risk locations for pedestrians should be identified through road safety risk assessments and road safety audits.

In connection with these key strategies under the Pillar 2, the findings from cluster analysis identified high-risk locations for pedestrians on the Cape Town’s road network that need thorough investigations both in terms of conventional traffic safety studies and road safety audits. In addition, the findings on hot spots locations are of a great relevance as they can guide where road safety interventions should be prioritised. The identified hot spots of the different pedestrian casualty types communicate, albeit indirectly, the type of road safety remedial scheme that is needed. For instance, hot spots of KSI pedestrian casualties may necessitate speed reduction plans such as traffic calming measures (where they are appropriate), speed enforcement initiatives, or spatial separation of motorised and non-motorised modes. In the same way, hot spots of intersection-related pedestrian casualties may be addressed by improvements in intersection design that prioritises pedestrian safety and caters for pedestrian needs. In addition to guiding remedial treatments, findings on hot spots of pedestrian crashes can guide the allocation of funding for safety improvement initiatives at local level. For instance, greater priority should be given to the census suburbs identified as hot spot regions of pedestrian crashes while allocating funding for pedestrian safety improvements.

The findings of this study show that pedestrian safety is significantly affected by the aspects of the built environment including road structure, intersection design and land use planning. These findings suggest that changes in the design and the planning of these aspects should be part of the strategies geared towards addressing the pedestrian safety problem in South African urban spaces. Key strategies should include rethinking how streets and highways are designed and modifying land use planning and urban design practices to accommodate and encourage walking. The majority of high-crash locations for pedestrians identified in this study are located in socioeconomically disadvantaged areas which host mostly poorer communities living in informal settlements, or those living in areas that were upgraded from an informal to a formal settlement. As these areas were not initially in the planning schemes of the city, they lack basic urban infrastructure such as roads, pedestrian paths, street lights, and access to main roads (to name a few) and where infrastructure is provided it is often in poor condition (e.g. roads are not maintained regularly, or inadequately provided). In addition, the risk exposure is greater for pedestrians living in these areas as they spend a great deal of time walking to reach urban services located far from their homes. The findings of this study reflect the fact that walking is the primary mode of transport for the poorer communities and the associated safety problems should be given much consideration in informal settlement upgrading programmes. A concerted effort is also needed from both urban planners and transportation engineers to accommodate pedestrians in the management of future spatial growth of urban environments.

Providing self-explaining and forgiving roadway environments for all road users is one of the strategic themes under Pillar 2 of the NRSS 2016-2030. In the light of the findings of this study on the role played by the built environment in the incidence of pedestrian crashes, self-explaining and forgiving road environment for pedestrians could be achieved by the implementing the following interventions:

1) Minimising pedestrian exposure on roads with high traffic volumes and high speed roads (e.g. arterial roads and urban freeways). This can be achieved by using measures such as :
    a. Spatial separation of pedestrians from motorised traffic by means of sidewalks, separated walking paths, overpasses and underpasses;
    b. Temporal separation at intersections and mid-blocks (i.e. by means of pedestrian signals) to eliminate or reduce conflicts between pedestrians and motorists;
    c. Provision of minimum crossing distances by means of raised medians, pedestrian refuges and lane narrowing or kerb extensions.
2) Managing speed by using strategies such as:
    a. Traffic calming measures (vertical deflections and horizontal deflections) on roads where these measure are warranted;
    b. School zones to limit speed during certain hours in the vicinity of a school;
    c. Shared zones where both pedestrians and motorized traffic utilised the same road space that has been adapted for very low vehicular speed, but vehicles must always give way to pedestrians (Austroads, 2008). However, special attention should be given to a number of negative aspects associated with the

implementation of shared zones. These include: (1) the relative high cost of implementing shared zones; (2) drivers' failure to observe speed restrictions when pedestrian volumes are low; (3) need for enforcement and educational initiatives to encourage understanding and compliance among road users; and (4) concerns that pedestrian safety might be compromised by non-complying motorists (Austroads, 2008);

   d. Pedestrianisation (e.g. creating pedestrian only zones) in areas of high pedestrian activity by restricting vehicular access;
   e. Speed enforcement measures (speed enforcement officers and speed cameras).

3) Integrating pedestrian needs into transport planning: this can be attained by:
   a. Providing pedestrian facilities where pedestrian movements are expected;
   b. Addressing missing links (non-continuous walking routes) to ensure that walking environments are connected;
   c. Removing physical obstacles from walking paths;
   d. Providing roadway environments with universal design features to accommodate pedestrians with special needs (children, pedestrian with disabilities and elderly pedestrians);
   e. Addressing pedestrian needs at major generators (e.g. mixed use with high pedestrian activity, urban centres, schools, transit termini, hospitals, commercial areas, parks and recreational places);
   f. Minimising waiting times and provide adequate time for pedestrians to cross at signalised intersections;
   g. Providing pedestrian amenities that encourage walking (e.g. pedestrian-scale lighting, trees, landscaping, shelters etc.);
   h. Providing safer walking environments in socioeconomically disadvantaged areas that hosts higher proportions of population who relies on walking as a primary transport mode.

4) Enhancing pedestrian conspicuity on the roadway by:
   a. Providing street lighting for night-time visibility;
   b. Using colours and textures of materials (e.g. road pavement) to emphasize the presence of pedestrian movement in pedestrian zones or at pedestrian crossings;
   c. Providing adequate sight distances at intersections and roundabouts;
   d. Keeping signs in good conditions (must be reflectorized or illuminated and with an adequate vertical clearance);

    e. Keeping markings for pedestrian crossings in good conditions.

5) Integrating land use and transport planning by:
    a. Designing cities in such a way that residences, workplaces, schools, shops and other facilities are in close proximity so that trips to reach human activities can be made by walking;
    b. Providing safe play areas and other recreational facilities in poorer communities to curb the problem of children who are hit by cars while playing in streets.

In light of the role played by traffic signals and the size of an intersection (i.e. number of intersection legs) in pedestrian crash occurrence, engineering countermeasures directed at these locations comprise those requiring site-specific remedial actions and those implemented at a regional or city level. Site-specific engineering interventions include those that have the potential to impact pedestrian crossing behaviour at an individual intersection. Certain of these interventions hold the potential to encourage spatial compliant behaviour (crossing at designated crossing locations) among pedestrians. These are, for instance, installing guard railing to channel pedestrian movement at designated crossing points and spatial separation by means of pedestrian underpasses or overpasses at wider intersections where crossing is most challenging. Another set of interventions can be deployed to encourage pedestrian compliance with traffic signal. These may include, for instance, the adoption of signal strategies that shorten waiting times for pedestrians (e.g. pedestrian actuated traffic signals such as Pelican and Puffin), those that inform about the waiting times (e.g. countdown timer), provisions of pedestrian phases, and all-red periods (Martin, 2006). However, it should be recognised that these countermeasures tend to increase delays to motorised traffic and signal timing should be optimised with proper consideration of a trade-off between pedestrian safety and intersection efficiency.

Engineering interventions implementable at a regional or city level may entail changes in the design of road networks aiming at enhancing pedestrian comfort and safety. According to the South African Road Traffic Signs Manual, a reduced number of intersections requiring traffic signals on a road network could be achieved by channelizing traffic to alternative routes that have fewer intersections, or alternatively to intersections that can handle traffic signals more adequately (Department of Transport, 2012). These changes would reduce the number of intersections requiring traffic signals and the size of intersection (in terms of the number of travel lanes per approach). The adoption of network management strategies such as redistribution of traffic on the road network by means of street closures, one-way systems and traffic calming measures can help to prevent wider signalised intersections which pose major comfort and safety problems to pedestrians. Local research is needed to evaluate safety benefits of these countermeasures and to ensure that their provision matches pedestrian needs and characteristics.

5.4.3 Pillar 4: Safer road users

The safer road user pillar entails strategies aiming at improving road user behaviour and increasing protection for vulnerable road users. The finding of this study highlighted pedestrian behaviours and actions that should be addressed by behavioural change interventions. For instance, educational programmes should be directed to pedestrians to instruct them how to operate as pedestrians on the road legally and safely. These programmes should target not only elementary school students but also other category of pedestrians who are at a greater crash risk, such as middle-aged groups and pedestrians living in poorer communities who depends on walking and public transport as primary modes of travel. Educational initiatives should aim at increasing awareness of risk associated with certain pedestrian behaviour such as temporal and spatial non-complying behaviour (i.e. jaywalking, red light violation), unsafe crossing strategies (e.g. running, midway stopping, crossing between stopped cars), drinking and walking, adult supervision in child pedestrian safety and the risk of being inconspicuous in the road environment and how to avoid this.

Although this study has not reported on driver factors that contribute to pedestrian crash occurrence, it is well documented locally and internationally that speeding, drinking and driving, and drivers' disregard of pedestrians in the road environment are the major contributing factors to pedestrian crashes. In addition to the basic traffic safety education directed to motorists during the licensing process, effective educational programs should be directed regularly to drivers of all categories (e.g. novice drivers, experienced drivers, transit drivers, truck drivers etc.) to enhance awareness of risks associated with the unsafe driving behaviour and to instruct them about safe operation around pedestrians and other vulnerable road users.

5.5 Limitations of the study

This research is subject to several limitations:

1) **Omission in reporting certain crash information**: The crash data provided by the City of Cape Town did not have records such as alcohol level of pedestrian involved in road crashes, visibility conditions, colour of clothes and weather conditions (e.g. wet pavement, visibility, etc.). The lack of this information has significantly restricted this study from reporting on the effect of alcohol factors on pedestrian safety and carrying out further analysis on the influence of weather conditions and pedestrian visibility on the outcome of pedestrian casualties in the study area.
2) **Incorrect records in the crash database**: This was predominantly found in recording the age of pedestrians involved in car crashes. This study showed that more than 60 percent of pedestrian casualties were aged 0 years old, a finding which is unrealistic. It seems that in circumstances where age is not known age is recorded as zero age. The incorrect recording of variable age is a common challenge for road safety analysis and has been reported by other scholars in South Africa.
3) **A crash database which is not georeferenced**: Inadequate description of crash locations particularly at non-intersection locations posed a great challenge to geocoding a subsample of midblock-related pedestrian casualties. This restricted certain methods of cluster analysis (e.g. hot spot analysis by KDE and other geospatial analysis techniques that use point density as a method of *Incident Data Aggregation*) to only intersection-related pedestrian casualties. Cluster analysis that included midblock-related pedestrian casualties was limited to geospatial analysis methods that use polygons (census suburbs in the context of this study) as the method of *Incident Data Aggregation*. Subsequently, the study was not able to identify high-risk locations for non-intersection pedestrian crashes on the road network. In addition, effort to obtain a geocoded database of pedestrian casualties and the procedures involved in data screening and data preparation for analysis were laborious and time consuming. These procedures occupied a significant part of time frame work for this study.
4) **Limitation on the analytical methods used in this study**: the standard planar Kernel Density Estimation (KDE) is one of the cluster analysis methods used in this study to identify high-risk locations for pedestrians on the Cape Town's road network. This technique produces a smooth density surface of spatial point events over a 2-D geographic space. However, events such as road crashes occur on a road network which

is a 1-D linear space. As a result, the planar KDE is likely to overestimate the density values as it covers space beyond the road network space. The Network Kernel Density Estimation (NKDE) is more appropriate for hot spot analysis on a road network, unfortunately this tool is not implemented in ArcMap which was the analytical tool used in geospatial analyses in this study. Therefore, the results on hot spots obtained using the standard planar KDE could have been improved by the use of The Network Kernel Density Estimation (NKDE). In a similar way, this study used the Geographically Weighted Regression tool implemented in ArcGIS to model spatially varying relationships pedestrian casualties and a set of predictors including variables describing the aspects of the built environment and population characteristics. This tool calibrates the basic GWR models that assume that the error terms are normally distributed. Simply put, a normal distribution of error terms was assumed in the GWR Models produced in this study. However, this assumption is often violated when fitting models for count data such as crash data. In this regard, a modelling procedure that combines the GWR modelling and any form of Generalised Linear Modelling (Poisson or Negative Binomial regression models) would have been beneficial to this study. Therefore, the GWR model results would have been improved by the use a software tool that supports the calibration of Geographically Weighted Generalised Linear Models (GWGLM).

5.6 Considerations for future research

Based on delineations and limitations highlighted in this study, four future research directions for pedestrian safety analysis are suggested as below.

1. As the majority of pedestrian crashes were identified at midblock locations, much research is needed to identify high-risk locations for midblock-related pedestrian crashes on the road network, of course with the help of better quality crash data. The research effort should be coupled with investigations into micro and macro environment factors associated with midblock-related pedestrian crashes in South Africa.

2. Research is needed to investigate the effect of spatial unit of analysis on the relationships between pedestrian crash counts and explanatory variables such as the attributes of the built environment and demographic characteristics in the context of South Africa. Research of this nature may consider, for instance, micro-level safety

analysis employing data aggregated in various spatial units, such as buffers of different dimensions around a crash location, traffic analysis zones (TAZs), districts, or grid-based schemes. This would shed light on the effects of modifiable areal unit problem (MAUP) on crash modelling results in traffic safety studies.

3. In the effort to gain more insight into the circumstances that lead to pedestrian crashes in South Africa, thorough research is needed into other factors contributing to pedestrian crashes which are not covered in this study. These are, for instance, driver-related factors, vehicle factors, weather factors, vehicular speed, alcohol involvement, transit characteristics and social psychological determinants of unsafe behaviour.

4. To address the methodological limitations underlined previously in this study, further research could assist with evaluating whether the results obtained in this study could be improved by the use of alternative analytical methods such as the Network Kernel Density Estimation (NKDE) for crash cluster analysis and Geographically Weighted Generalised Linear Models (GWGLM) for crash modelling.

References

AASHTO. 2010. *Highway Safety Manual, 1st Edition*. Washington, D.C.: American Association of State Highway and Transportation Officials.

Abbas, K.A. 2004. Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt. *Accident Analysis & Prevention*. 36(2):149–163.

Abdulhafedh, A. 2016. Crash Frequency Analysis. *Journal of Transportation Technologies*. 06(04):169.

Acevedo, M.F. 2012. *Data Analysis and Statistics for Geography, Environmental Science, and Engineering*. Boca Raton, USA: CRC Press.

Adams, J.D. 2013. Theorizing a Sense of Place in a Transnational Community. *Children, Youth and Environments*. 23(3):43.

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

Albers, P.N., Wright, C. & Olwoch, J. 2010. Developing a South African Pedestrian Environment Assessment Tool: Tshwane Case Study. *South African Journal of Science*. 106(9–10):1–8.

Amoh-Gyimah, R., Saberi, M. & Sarvi, M. 2016. Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. *Accident Analysis & Prevention*. 93:147–159.

Amoh-Gyimah, R., Saberi, M. & Sarvi, M. 2017. The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models. *Analytic Methods in Accident Research*. 13:28–51.

Anselin, L. 1995. Local Indicators of Spatial Association-LISA. *Geographical Analysis*. 27(2):93–115.

Anselin, L. 1996. The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In M. Fischer, H. Scholten, & D. Unwin (eds.). London, UK: Taylor and

Francis *Spatial Analytical Perspectives on GIS in Environmental and SocioEconomic Sciences*. 111–125.

Anselin, L. 2010. Thirty Years of Spatial Econometrics. *Papers in Regional Science*. 89(1):3–25.

Anselin, L. & Bao, S. 1997. Exploratory Spatial Data Analysis: Linking SpaceStat and ArcView. In M.M. Fischer & A. Getis (eds.). Springer *Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling, and Computational Intelligence*. 35–59.

Association, A.P. 2006. *Planning and urban design standards*. John Wiley & Sons.

Austroads. 2008. *Guide to Traffic Management Part 8: Local Area Traffic Management*. Sydney: Austroads Incorporated.

Aygoren, S. 2004. Ins and Outs of Mixed-use. In Vol. 59 *Real Estate Forum*. 66–71.

Badoe, D.A. & Miller, E.J. 2000. Transportation–land-use interaction: empirical findings in North America, and their implications for modeling. *Transportation Research Part D: Transport and Environment*. 5(4):235–263.

Baltes, M. 1998. Descriptive analysis of crashes involving pedestrians in Florida, 1990-1994. *Transportation Research Record: Journal of the Transportation Research Board*. 1636:138–145.

Bashirul Haque, M., Rahman, M., Abu Sayed Khan, M. & Nahid Parvez, M. 2013. Impact of Land Use Parameters on Household Travel Behavior. *American Journal of Civil Engineering and Architecture*. 1(4):70–74.

Behrens, R. 2002. Findings of an activity-based household travel survey in Cape Town, with particular reference to walking as a travel mode. In Pretoria, South Africa: Southern African Transport Conference (SATC) *The 21st Annual South African Transport Conference*.

Behrens, R. 2010. Pedestrian arterial and freeway crossing behaviour in Cape Town: Observations and Implications. In Lisbon, Portugal *The 12th WCTR, July 11-15, 2010*.

Behrens, R. & Makajuma, G. 2017. Pedestrian crossing behaviour in Cape Town and Nairobi. In W. V. Mitullah, M. Vanderschuren, & M. Khayesi (eds.). New York: Routledge Ltd *Non-Motorized Transport Integration Into Urban Transport Planning in Africa* . 27–56.

Berrigan, D., Pickle, L.W. & Dill, J. 2010. Associations between street connectivity and active transportation. *International Journal of Health Geographics*. 9(1):20.

Bíl, M., Andrášik, R. & Janoška, Z. 2013. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. *Accident Analysis & Prevention*. 55:265–273.

Blum, U. & Gaudry, M. 2000. The Snus-2.5 Model for Germany. In M. Gaudry & S. Lassarre (eds.). Bingley, UK: Emerald Group Publishing Limited *Structural Road Accident Models: The International DRAG Family*. 67–96.

Bordoloi, R., Mote, A., Sarkar, P.P. & Mallikarjuna, C. 2013. Quantification of Land Use Diversity in The Context of Mixed Land Use. *Procedia - Social and Behavioral Sciences*. 104:563–572.

Van den Bossche, F., Wets, G. & Brijs, T. 2005. Role of Exposure in Analysis of Road Accidents: A Belgian Case Study. *Transportation Research Record: Journal of the Transportation Research Board*. 1908:96–103.

Bossler, J.D., Campbell, J.B., McMaster, R.B. & Rizos, C. 2010. *Manual of Geospatial Science and Technology, Second Edition*. 2nd ed. Boca Raton, FL. USA: CRC Press.

Brownson, R.C., Hoehner, C.M., Day, K., Forsyth, A. & Sallis, J.F. 2009. Measuring the Built Environment for Physical Activity. *American Journal of Preventive Medicine*. 36(4):S99–S123.e12.

Brus, J., Vondrakova, A. & Vozenilek, V. 2014. *Modern Trends in Cartography: Selected Papers of CARTOCON 2014*. Springer.

Burnham, K.P., Anderson, D.R. & Huyvaert, K.P. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol*. 65:23–35.

Burstein, B., Fauteux-Lamarre, E. & As, A.B. Van. 2016. Increased morbidity associated with weekend paediatric road traffic injuries: 10-year analysis of trauma registry data. *Injury*. 47(6):1236–1241.

Caliendo, C., Guida, M. & Parisi, A. 2007. A crash-prediction model for multilane roads. *Accident Analysis & Prevention*. 39(4):657–670.

Cantillo, V., Arellana, J. & Rolong, M. 2015. Modelling pedestrian crossing behaviour in urban roads: A latent variable approach. *Transportation Research Part F: Traffic Psychology and Behaviour*. 32:56–67.

Cervero, R. & Kockelman, K. 1997a. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*. 2(3):199–219.

Cervero, R. & Kockelman, K. 1997b. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*. 2(3):199–219.

Cervero, R. & Radisch, C. 1996. Travel choices in pedestrian versus automobile oriented neighborhoods. *Transport Policy*. 3(3):127–141.

Chainey, S., Tompson, L. & Uhlig, S. 2008. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal*. 21(1–2):4–28.

Chakravarthy, B., Anderson, C.L., Ludlow, J., Lotfipour, S. & Vaca, F.E. 2010. The Relationship of Pedestrian Injuries to Socioeconomic Characteristics in a Large Southern California County. *Traffic Injury Prevention*. 11(5):508–513.

Chandra, S. & Quadrifoglio, L. 2013. A new street connectivity indicator to predict performance for feeder transit services. *Transportation Research Part C: Emerging Technologies*. 30:67–80.

Chen, Y., Ravulaparthy, S., Deutsch, K., Dalal, P., Yoon, S.Y., Lei, T., Goulias, K.G., Pendyala, R.M., et al. 2011. Development of Indicators of Opportunity-Based Accessibility. *Transportation Research Record: Journal of the Transportation Research Board*. 2255(1):58–68.

Child, D. 2006. *The Essentials of Factor Analysis*. 3rd Editio ed. New York, USA: Continuum International Publishing Group.

Cho, G., Rodríguez, D.A. & Khattak, A.J. 2009. The role of the built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. *Accident Analysis & Prevention*. 41(4):692–702.

Chun, Y. & Griffith, D.A. 2013. *Spatial Statistics and Geostatistics: Theory and Applications for Geographic Information Science and Technology*. Sage.

City of Cape Town's Transport Authority. 2005. *City of Cape Town: NMT Policy and Strategy, Volume 1: Status quo assessment*. Cape Town: City of Cape Town.

City of Cape Town. 2012a. *The City of Cape Town: Annual Report 2011/12*. Cape Town, South Africa.

City of Cape Town. 2012b. *City of Cape Town-2011 census-Cape Town: Trends and Change-10 years, Census 2001-Census 2011*. Cape Town, South Africa.

City of Cape Town. 2015a. *Economic District Fact Sheet*. Cape Town, South Africa.

City of Cape Town. 2015b. *City of Cape Town Municipal Planning By-Law, 2015*. Vol. 2017. [Online], Available: http://openbylaws.org.za/za-cpt/act/by-law/2015/municipal-planning/.

Cliff, A.D. & Ord, J.K. 1973.

Cliff, A.D. & Ord, J.K. 1981.

Clifton, K., Ewing, R., Knaap, G. & Song, Y. 2008. Quantitative analysis of urban form: a multidisciplinary review. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*. 1(1):17–45.

Cohen, J. 1992a. Statistical Power Analysis. *Current Directions in Psychological Science*. 1(3):98–101.

Cohen, J. 1992b. A Power Primer. *Psychological Bulletin [PsycARTICLES*. 112(1):155–159.

Cottrill, C.D. & Thakuriah, P. (Vonu). 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis & Prevention*. 42(6):1718–1728.

Cowles, M. & Davis, C. 1982. On the origins of the .05 level of statistical significance.

American Psychologist. 37(5):553–558.

Crucitti, P., Latora, V. & Porta, S. 2006. Centrality measures in spatial networks of urban streets. *Physical Review E*. 73(3):036125.

CSIR Building and Construction Technology. 2005. *Guidelines for human settlement planning and design, Volume 2*. Pretoria, South Africa: CSIR Building and Construction Technology.

Cubbin, C. & Smith, G.S. 2002. Socioeconomic Inequalities in Injury: Critical Issues in Design and Analysis. *Annual Review of Public Health*. 23(1):349–375.

Cudeck, R. 2000. Exploratory Factor Analysis. In E.A. Tinsley, Howard & S.D. Brown (eds.). Academic Press *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. 265–296.

Dangermond, J. 1992. What is a Geographic Information System (GIS)? In ASTM International *Geographic Information Systems (GIS) and Mapping—Practices and Standards*.

Dark, S.J. & Bram, D. 2007. The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*. 31(5):471–479.

Davis, D.G. & Braaksma, J.P. 1988. Adjusting for luggage-laden pedestrians in airport terminals. *Transportation Research Part A: General*. 22(5):375–388.

Delmelle, E. & Thill, J.-C. 2008. Urban Bicyclists: Spatial Analysis of Adult and Youth Traffic Hazard Intensity. *Transportation Research Record: Journal of the Transportation Research Board*. 2074:31–39.

Department of Transport. 2012. *South African Road Traffic Signs Manual Volume:Traffic Signal Design*. 3rd Editio ed. Pretoria, South Africa: National Department of Transport.

Department of Transport. 2014. *National Road Safety Strategy 2016-2030*. Pretoria, South Africa: Department of Transport, RSA.

Dill, J. 2004. Measuring network connectivity for bicycling and walking. In Washington, DC: Transportation Research Board *The 83rd Annual Meeting of the Transportation Research*

Board. 11–15.

Dixon, B., Uddameri, V. & Ray, C. 2016. *GIS and Geocomputation for Water Resource Science and Engineering*. Chichester, UK.: John Wiley & Sons.

Dultz, L.A. & Frangos, S.G. 2013. The impact of alcohol in pedestrian trauma. *Trauma*. 15(1):64–75.

Dumbaugh, E. & Li, W. 2010. Designing for the Safety of Pedestrians, Cyclists, and Motorists in Urban Environments. *Journal of the American Planning Association*. 77(1):69–88.

Econometrix. 2013. *Economic Impact of an Advertising Ban on Alcoholic Beverages*. Econometrix (Pty) Ltd.

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

Elvik, R. 1995. Meta-analysis of evaluations of public lighting as accident countermeasure. *Transportation Research Record :Journal of the Transportation Research Board*. 1485:112–123.

Elvik, R., Vaa, T., Hoye, A. & Sorensen, M. 2009. *The handbook of road safety measures*. Emerald Group Publishing.

Erdogan, S. 2009. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *Journal of Safety Research*. 40(5):341–351.

ESRI. n.d. *Using Select By Location*. Vol. 2017.

ESRI. n.d. *Interpreting GWR results*. [Online], Available: http://resources.esri.com/help/9.3/arcgisdesktop/com/gp_toolref/spatial_statistics_tools/interpreting_gwr_results.htm [2018b, May 09].

Ewing, R. & Cervero, R. 2010. Travel and the Built Environment:A Synthesis. *Journal of the American Planning Association*. 76(March 2015):265–294.

Ewing, R. & Dumbaugh, E. 2009. The Built Environment and Traffic Safety: A Review of Empirical Evidence. *Journal of Planning Literature*. 23(4):347–367.

Ewing, R., Schmid, T., Killingsworth, R., Zlot, A. & Raudenbush, S. 2003. Relationship between urban sprawl and physical activity, obesity, and morbidity. 18(1):47–57.

Ewing, R., Meakins, G., Bjarnson, G. & Hilton, H. 2011. Transportation and Land Use. In A. Danneberg, H. Frumkin, & J. Jackson, Richard (eds.). Washington, DC: Island Press *Making Healthy Places: Designing and Building for Health, Well-being, and Sustainability*. 149–169.

Ewing, R., Greenwald, M., Zhang, M., Walters, J., Feldman, M., Cervero, R., Frank, L. & Thomas, J. 2011. Traffic Generated by Mixed-Use Developments—Six-Region Study Using Consistent Built Environmental Measures. *Journal of Urban Planning and Development*. 137(3):248–261.

Ewing, R., Meakins, G., Hamidi, S. & Nelson, A.C. 2014. Relationship between urban sprawl and physical activity, obesity, and morbidity – Update and refinement. *Health & place*. 26:118–126.

Field, A. 2013. *Discovering statistics using IBM SPSS statistics*. 4. ed. ed. Los Angeles: SAGE.

Flahaut, B. 2004. Impact of infrastructure and local environment on road unsafety: Logistic modeling with spatial autocorrelation. *Accident Analysis & Prevention*. 36(6):1055–1066.

Flahaut, B., Mouchart, M., Martin, E.S. & Thomas, I. 2003. The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach. *Accident Analysis & Prevention*. 35(6):991–1004.

Forsyth, A. & Southworth, M. 2008. Cities Afoot—Pedestrians, Walkability and Urban Design. *Journal of Urban Design*. 13(1):1–3.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. 2002. *Geographically weighted regression : the analysis of spatially varying relationships*. Wiley.

Fournier, F. & Simard, R. 2000. The DRAG-2 Model for Quebec. In M. Gaudry & S. Lassarre (eds.). Bingley, UK: Emerald Group Publishing Limited *Structural Road Accident Models: The International DRAG Family*. 37–66.

Frank, L.D., Stone Jr., B. & Bachman, W. 2000. Linking land use with household vehicle emissions in the central puget sound: methodological framework and findings.

Transportation Research Part D: Transport and Environment. 5(3):173–196.

Frank, L.D., Greenwald, M.J., Kavage, S. & Devlin, A. 2011. *An Assessment of Urban Form and Pedestrian and Transit Improvements as an Integrated GHG Reduction Strategy*. Washington, USA.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R. & Thomsen, L.K. 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis & Prevention*. 27(1):1–20.

Gårder, P.E. 2004. The impact of speed and other variables on pedestrian safety in Maine. *Accident Analysis & Prevention*. 36(4):533–542.

Geary, R.C. 1954. The contiguity ratio and statistical mapping. *The incorporated statistician*. 5(3):115–146.

Gehrke, S.R. & Clifton, K.J. 2015. Toward a spatial-temporal measure of land-use mix. *Journal of Transport and Land Use*. 9(1).

Gehrke, S.R. & Clifton, K.J. 2017. An activity-related land use mix construct and its connection to pedestrian travel. *Environment and Planning B: Urban Analytics and City Science*. 0(0):1–18.

Gerdenitsch, C. 2017. New Ways of Working and Satisfaction of Psychological Needs. In C. Korunka & B. Kubicek (eds.). Cham, Switzerland: Springer International Publishing *Job Demands in a Changing World of Work: Impact on Worker's Health and Performance and Implications for Research and Practice*.

Getis, A. & Ord, J.K. 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*. 24(3):189–206.

Geurts, K., Thomas, I. & Wets, G. 2005. Understanding spatial concentrations of road accidents using frequent item sets. *Accident Analysis & Prevention*. 37(4):787–799.

Golembiewski, G.A. & Chandler, B. 2011. *Roadway Safety Information Analysis: A Manual for Local Rural Road Owners (No.FHWA-SA-11-10)*. Washington, D.C.: Federal Highway Admnistration.

Gomes, M.J.T.L., Cunto, F. & da Silva, A.R. 2017. Geographically weighted negative binomial regression applied to zonal level safety performance models. *Accident Analysis & Prevention*. 106:254–261.

Gomez-Ibanez, D.J., Boarnet, M.G., Brake, D.R., Cervero, R.B., Cotugno, A., Downs, A., Hanson, S., Kockelman, K.M., et al. 2009. *Driving and the built environment: the effects of compact development on motorized travel, energy use, and CO2 emissions*. Oak Ridge, TN (USA): Oak Ridge National Lab. (ORNL).

Greene-Roesel, R., Diogenes, M.C. & Ragland, D.R. 2007. *Estimating Pedestrian Accident Exposure: Protocol Report*. California: University of California, Berkeley.

Greenwald, M. & Boarnet, M. 2001. Built Environment as Determinant of Walking Behavior: Analyzing Nonwork Pedestrian Travel in Portland, Oregon. *Transportation Research Record*. 1780(1):33–41.

Griffith, D.A. 2003. *Spatial Autocorrelation and Spatial filtering: Gaining Understanding Through Theory and Scientific Visualization*. Springer.

Griffith, D.A. 2017. Measuring spatial dependence. In D. Richardson, N. Castree, M.F. Goodchild, A. Kobayashi, W. Liu, & R.A. Marston (eds.). Oxford, UK: John Wiley & Sons *The International Encyclopedia of Geography: People, the Earth, Environment, and Technology*.

Guadamuz-Flores, R. & Aguero-Valverde, J. 2017. Bayesian Spatial Models of Crash Frequency at Highway–Railway Crossings. *Transportation Research Record: Journal of the Transportation Research Board*. 2608:27–35.

Guo, L., Ma, Z. & Zhang, L. 2008. Comparison of bandwidth selection in application of geographically weighted regression: a case study. *Canadian Journal of Forest Research*. 38(9):2526–2534.

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. *Accident Analysis & Prevention*. 99:114–124.

Hadayeghi, A., Shalaby, A. & Persaud, B. 2003. Macrolevel Accident Prediction Models for

Evaluating Safety of Urban Transportation Systems. *Transportation Research Record: Journal of the Transportation Research Board*. 1840:87–95.

Hadayeghi, A., Shalaby, A.S. & Persaud, B.N. 2010. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. *Accident Analysis & Prevention*. 42(2):676–688.

Hakim, S., Shefer, D., Hakkert, A.S. & Hocherman, I. 1991. A critical review of macro models for road accidents. *Accident Analysis & Prevention*. 23(5):379–400.

Handy, S. 2005. Smart Growth and the Transportation-Land Use Connection: What Does the Research Tell Us? *International Regional Science Review*. 28(2):146–167.

Handy, S., Paterson, R.G. & Butler, K.S. 2003. *Planning for street connectivity : getting from here to there*. Chicago, IL (122 S. Michigan Ave, Suite 1600, Chicago 60603): American Planning Association, Planning Advisory Service.

Handy, S.L., Boarnet, M.G., Ewing, R. & Killingsworth, R.E. 2002. How the built environment affects physical activity: Views from urban planning. *American Journal of Preventive Medicine*. 23(2, Supplement 1):64–73.

Hannan, V.A. 2013. Capturing the impacts of land use on travel behavior : comparison of modeling approaches. [Online], Available: https://dspace.mit.edu/handle/1721.1/85919 [2018, April 12].

Hanneman, R.A. & Riddle, M. 2005. *Introduction to social network methods*. Riverside, CA: University of California at Riverside.

Hanson, C.S., Noland, R.B. & Brown, C. 2013. The severity of pedestrian crashes: an analysis using Google Street View imagery. *Journal of Transport Geography*. 33:42–53.

Harrell, W.A. 1991. Factors Influencing Pedestrian Cautiousness in Crossing Streets. *The Journal of Social Psychology*. 131(3):367–372.

Hashimoto, S., Yoshiki, S., Saeki, R., Mimura, Y., Ando, R. & Nanba, S. 2016. Development and application of traffic accident density estimation models using kernel density estimation. *Journal of Traffic and Transportation Engineering (English Edition)*. 3(3):262–270.

Herbert, H.K., van As, A.B., Bachani, A.M., Mtambeka, P., Stevens, K.A., Millar, A.J.W. & Hyder, A.A. 2012. Patterns of pediatric injury in South Africa. *Journal of Trauma and Acute Care Surgery*. 73(1):168–174.

Hess, M.P., Moudon, V.A. & Logsdon, M. 2001. Measuring Land Use Patterns for Transportation Research. *Transportation Research Record: Journal of the Transportation Research Board*. 1780:17–24.

Hess, P., Moudon, A., Snyder, M. & Stanilov, K. 1999. Site design and pedestrian travel. *Transportation Research Record: Journal of the Transportation Research Board*. (1674):9–19.

Hisrich, R.D. & Kralik, D.W. 2016. *Advanced Introduction to Corporate Venturing*. Cheltenham, UK: Edward Elgar Publishing. [Online], Available: http://lib.myilibrary.com?ID=950954.

Hoaglin, D.C., Iglewicz, B. & Tukey, J.W. 1986. Performance of Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association*. 81(396):991–999.

Hodgson, F., Page, M. & Tight, M. 2004. *A review of factors which influence pedestrian use of the streets: Task 1 report for an EPSRC funded project on measuring pedestrian accessibility*. Leeds, UK.

Holland, C. & Hill, R. 2007. The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations. *Accident Analysis & Prevention*. 39(2):224–237.

Holubowycz, O. 1995a. Alcohol-involved pedestrians: the Australian experience. In C. Kloeden & A. McLean (eds.). Adelaide, Australia: Road Accident Research Unit *Proceedings of the 13th International Conference on Alcohol, Drugs and Traffic Safety, Adelaide, 13-18 August 1995*. 700–710.

Holubowycz, O. 1995b. Age, sex, and blood alcohol concentration of killed and injured pedestrians. *Accident; analysis and prevention*. 27(3):417–22.

Hoogendoorn, S.P. & Bovy, P.H.L. 2004. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*. 38(2):169–190.

Hoppenbrouwer, E. & Louw, E. 2005. Mixed-use development: Theory and practice in Amsterdam's Eastern Docklands. *European Planning Studies*. 13(7):967–983.

Hoyle, R.H. 2000. Confirmatory Factor Analysis. In H.E.A. Tinsley & S.D. Brown (eds.). Academic Press *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. 465–497.

Huang, H., Abdel-Aty, M. & Darwiche, A. 2010. County-Level Crash Risk Analysis in Florida. *Transportation Research Record: Journal of the Transportation Research Board*. 2148:27–37.

Huang, H., Zhou, H., Wang, J., Chang, F. & Ma, M. 2017. A multivariate spatial model of crash frequency by transportation modes for urban intersections. *Analytic Methods in Accident Research*. 14(Supplement C):10–21.

Humpel, N., Owen, N., Leslie, E., Marshall, A.L., Bauman, A.E. & Sallis, J.F. 2004.

Hunter, W.W., Stutts, J.C., Pein, W.E. & Cox, C.L. 1996. *Pedestrians and Bicycle Crash Types of the Early 1990's,*. Washington, DC.

Hutchinson, P., Kloeden, C. & Lindsay, V. 2009. *Accidents to intoxicated pedestrians in South Australia*. Adelaide, Australia: Centre for Automotive Safety Research.

Institute of Transportation Engineers. 2012. *Trip Generation, 9th Edition.* Washington, D.C.: Institute of Transportation Engineers (ITE).

Institute of Transportation Engineers. 2017. *Trip Gneration Handbook, 3rd Edition*. Washington, D.C.: Institute of Transportation Engineers (ITE).

Ishaque, M.M. & Noland, R.B. 2008. Behavioural Issues in Pedestrian Speed Choice and Street Crossing Behaviour: A Review. *Transport Reviews*. 28(1):61–85.

Jaeger, L. & Lassarre, S. 2000. The TAG-l Model for France. In M. Gaudry & S. Lassarre (eds.). Bingley, UK: Emerald Group Publishing Limited *Structural Road Accident Models: The International DRAG Family*. 157–184.

Jamil, R., Xiong, S., Kong, X., Zheng, S. & Fang, Z. 2015. Pedestrian Crossing Patterns Preference at a Non-signalized Crosswalk. *Procedia Manufacturing*. 3:3353–3359.

Jehle, D. & Cottington, E. 1988. Effect of alcohol consumption on outcome of pedestrian victims. *Annals of emergency medicine*. 17(9):953–6.

Jonah, B.A. & Engel, G.R. 1983. Measuring the relative risk of pedestrian accidents. *Accident Analysis & Prevention*. 15(3):193–206.

Jordan, P.W. 1985. Pedestrians and Cyclists at Roundabouts. In Melbourne, Australia: Institute of Engineers, Australia *3rd National Local Government Engineering Conference 1985*.

Kang, J. & Oh, H. 2016. Factors affecting vehicles' carbon emission in road networks. *Transportmetrica A: Transport Science*. 12(8):736–750.

Keegan, O. & O'Mahony, M. 2003. Modifying pedestrian behaviour. *Transportation Research Part A: Policy and Practice*. 37(10):889–901.

Kent, R.A. 2015. *Analysing quantitative data: Variable-based and case-based approaches to non-experimental datasets*. Sage.

Kim, C. 2009. Spatial Data Mining, Geovisualization. In R. Kitchin & N. Thrift (eds.). Oxford: Elsevier *International Encyclopedia of Human Geography*. 332–336.

Kim, D.-G., Lee, Y., Washington, S. & Choi, K. 2007. Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models. *Accident Analysis & Prevention*. 39(1):125–134.

Kim, K., Pant, P. & Yamashita, E. 2010. Accidents and Accessibility: Measuring the Influences of Demographic and Land Use Variables in Honolulu, Hawaii. *Transportation Research Record Journal of the Transportation Research Board*. 2147:9–17.

Kolmogorov, A.N. 1933. "Sulla Determinazione Empirica di Una Legge di Distributione". *Giornale dell'Istituto Italiano degli Attuari*. 4:83–91.

Koohsari, M.J., Badland, H., Sugiyama, T., Mavoa, S., Christian, H. & Giles-Corti, B. 2015. Mismatch between Perceived and Objectively Measured Land Use Mix and Street Connectivity: Associations with Neighborhood Walking. *Journal of Urban Health*. 92(2):242–252.

Krige, D.G. 1966. Two-dimensional weighted moving average trend surfaces for ore evaluation. In Johannesburg, South Africa.: South African Institute of Mining and Metallurgy *Proceedings of the Symposium on Mathematical Statistics and Computer Applications in Ore Valuation*. 13–38.

Krizek, K. 2000. Pretest-posttest strategy for researching neighborhood-scale urban form and travel behavior. *Transportation Research Record: Journal of the Transportation Research Board*. (1722):48–55.

Kröyer, H.R.G. 2015. Is 30km/h a 'safe' speed? Injury severity of pedestrians struck by a vehicle and the relation to travel speed and age. *IATSS Research*. 39(1):42–50.

Kudryavtsev, A., Stedman, R.C. & Krasny, M.E. 2012. Sense of place in environmental education. *Environmental Education Research*. 18(2):229–250.

Kwan, M.-P. 2013. Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility. *Annals of the Association of American Geographers*. 103(5):1078–1086.

Lam, W.W.., Yao, S. & Loo, B.P.Y. 2014. Pedestrian exposure measures: A time-space framework. *Travel Behaviour and Society*. 1(1):22–30.

Lam, W.W.Y., Loo, B.P.Y. & Yao, S. 2013. Towards exposure-based time-space pedestrian crash analysis in facing the challenges of ageing societies in Asia. *Asian Geographer*. 30(2):105–125.

Larsen, J. & El-Geneidy, A. 2011. A travel behavior analysis of urban cycling facilities in Montréal, Canada. *Transportation Research Part D: Transport and Environment*. 16(2):172–177.

LaScala, E.A., Gerber, D. & Gruenewald, P.J. 2000. Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis & Prevention*. 32(5):651–658.

Lassarre, S., Papadimitriou, E., Yannis, G. & Golias, J. 2007. Measuring accident risk exposure for pedestrians in different micro-environments. *Accident Analysis & Prevention*. 39(6):1226–1238.

Latrémouille, M.E., Thouez, J.P., Ranou, A., Bergeron, J., Bourbeau, R. & Bussière, Y. 2004.

Le sexe est-il une variable pertinente pour l'étude du comportement des piétons en intersection urbaine? [Is gender a relevant variable when studying the behaviours of pedestrians at urban intersections?]. *Recherche-Transports-Sécurité (RTS).* Number 84:171–188.

Leck, E. 2006. The Impact of Urban Form on Travel Behaviour: A Meta-Analysis. *Berkeley Planning Journal*. 19(1):37–58.

LeSage, J.P. & Pace, R.K. 2009. *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Levine, N., Kim, K.E. & Nitz, L.H. 1995. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis & Prevention*. 27(5):675–685.

Lewandowska-Gwarda, K. 2018. Geographically Weighted Regression in the Analysis of Unemployment in Poland. *International Journal of Geo-Information*. 7(17):1–16.

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. *Safety Science*. 58(Supplement C):89–97.

Liebenberg, K. & Garrod, K. 2005. Alleviation of the pedestrian safety crisis in the city of Cape Town. In Pretoria, South Africa: Southern African Transport Conference (SATC) *Proceedings of the 24th Southern African Transport Conference (SATC 2005)*. 537–547.

Lin, L. & Moudon, A.V. 2010. Objective versus subjective measures of the built environment, which are most effective in capturing associations with walking? *Health & Place*. 16(2):339–348.

Liu, C. & Shen, Q. 2011. An empirical analysis of the influence of urban form on household travel and energy consumption. *Computers, Environment and Urban Systems*. 35(5):347–357.

Liu, C., Chen, C.-L., Subramanian, R. & Utter, D. 2005. *Analysis of Speeding-Related Fatal Motor Vehicle Traffic Crashes*.

Lloyd, C. 2010. *Spatial data analysis: an introduction for GIS users*. Oxford University Press.

Loehlin, J.C. & Beaujean, A.A. 2017. *Latent Variable Models:An introduction to Factor, Path,* *and Structural Equation Analysis*. 5th Editio ed. New York, USA: Routledge.

Loo, B.P.Y. & Yao, S. 2012. Chapter 24: Geographical Information Systems. In G. Li & S.P. Baker (eds.). New York, USA.: Springer *Injury Research: Theories, Methods, and Approaches*. 447–463.

Lord, D. & Mannering, F. 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*. 44(5):291–305.

Lord, D. & Persaud, B. 2000. Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations Procedure. *Transportation Research Record: Journal of the Transportation Research Board*. 1717:102–108.

Loukaitou-Sideris, A., Liggett, R. & Sung, H.-G. 2007. Death on the Crosswalk. *Journal of Planning Education and Research*. 26(3):338–351.

Louw, E. & Vries, P. de. 2002. Working at Home: The Dutch Property Dimension. *Planning Practice & Research*. 17(1):17–30.

Luke, R. & Heyns, G.J. 2014. Reducing risky driver behaviour through the implementation of a driver risk management system. *Journal of Transport and Supply Chain Management*. 8(1):10.

Mabunda, M.M., Swart, L.-A. & Seedat, M. 2008. Magnitude and categories of pedestrian fatalities in South Africa. *Accident Analysis & Prevention*. 40(2):586–593.

MacKenzie, S., Seedat, M., Swart, L. & Mabunda, M. 2008. Pedestrian injury in South Africa: Focusing intervention efforts on priority pedestrian groups and hazardous places. *Crime, Violence and Injury Prevention in South Africa*. 31–48.

Malczewski, J. 1999. *GIS and multicriteria decision analysis*. John Wiley & Sons.

Manaugh, K. & Kreider, T. 2013. What is mixed use? Presenting an interaction method for measuring land use mix. *Journal of Transport and Land use*. 6(1):63–72.

Mannering, F.L., Shankar, V. & Bhat, C.R. 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*. 11:1–16.

Marcin, J.P., Schembri, M.S., He, J. & Romano, P.S. 2003. A population-based analysis of socioeconomic status and insurance status and their relationship with pediatric trauma hospitalization and mortality rates. *American journal of public health*. 93(3):461–6.

Marshall, S. 2005. *Streets and patterns*. Spon Press.

Marshall, W.E. & Garrick, N.W. 2010. Street network types and road safety: A study of 24 California cities. *URBAN DESIGN International*. 15(3):133–147.

Martin, A. 2006. *Factors influencing pedestrian safety : a literature review*. London, UK: TRL.

Matheron, G. 1971. *The theory of regionalized variables and its applications*. Vol. 5 ed. Paris, France: Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau.

Matthews, J.W. & Turnbull, G.K. 2007. Neighborhood Street Layout and Property Value: The Interaction of Accessibility and Land Use Mix. *The Journal of Real Estate Finance and Economics*. 35(2):111–141.

Matzopoulos, R. 2004. *A profile of fatal injuries in South Africa: Fifth annual report 2003 of the National Injury Mortality Surveillance System*. [Online], Available: https://scholar.google.co.za/.

Meel, B.L. 2007. Alcohol-Related Traumatic Deaths in Transki Region, South Africa. *Internet Journal of Medical Update - EJOURNAL*. 1(1).

Miller, H.J. 2005. A Measurement Theory for Time Geography. *Geographical Analysis*. 37(1):17–45.

Miranda-Moreno, L.F., Morency, P. & El-Geneidy, A.M. 2011. The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. *Accident Analysis and Prevention*. 43(5):1624–1634.

Mitchel, A. 2005. *The ESRI Guide to GIS analysis, Volume 2: Spatial measurements and statistics*. Redlands, CA, USA: ESRI Press.

Mitullah, Winnie, V., vanderschuren, Marianne, Khayesi & Meleckidzedeck. 2017. *Non-Motorized Transport Integration into Urban Transport Planning in Africa*. New York, USA: Routledge Ltd.

Moeinaddini, M., Asadi-Shekari, Z. & Zaly Shah, M. 2014. The relationship between urban street networks and the number of transport fatalities at the city level. *Safety Science*. 62.

Mohan, D., Bangdiwala, S.I. & Villaveces, A. 2017. Urban street structure and traffic safety. *Journal of Safety Research*. 62:63–71.

Montgomery, D.C. & Runger, G.C. 2007. *Applied statistics and probability for engineers*. 4. ed. ed. Hoboken, NJ: Wiley.

Montgomery, D.C. & Runger, G.C. 2014. *Applied statistics and probability for engineers*. 6. ed., SI ed. Hoboken, NJ: John Wiley & Sons, Inc.

Moons, E., Brijs, T. & Wets, G. 2009. Identifying Hazardous Road Locations: Hot Spots versus Hot Zones. In M.L. Gavrilova & C.J.K. Tan (eds.). Berlin, Heidelberg: Springer-Verlag *Transactions on Computational Science VI*. 288–300.

Moran, P.A.P. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society.Series B (Methodological)*. 10(2):243–251.

Morris, J.M., Dumble, P.L. & Wigan, M.R. 1979. Accessibility indicators for transport planning. *Transportation Research Part A: General*. 13(2):91–109.

Mouton, J. 2001. *How to succeed in your master's and doctoral studies: A South African guide and resource book*. Pretoria: Van Schaik.

Moyano Díaz, E. 2002. Theory of planned behavior and pedestrians' intentions to violate traffic regulations. *Transportation Research Part F: Traffic Psychology and Behaviour*. 5(3):169–175.

Murphy, K.R., Myors, B. & Wolach, A. 2009. *Statistical Power Analysis: A simple and General Model for Traditional and Modern Hypothesis Tests*. 3rd ed. ed. Routledge.

Musakwa, W. & van Niekerk, A. 2012. Implications of Land Use Mix on the Sustainability of African Urban Centres: A Case Study of Stellenbosch, South Africa. In Schwechat, Austira: REAL CORP *Re-mixing the City-Towards Sustainability and Resilience*. 1237–1243.

National Department of Health. 2007. *South Africa Demographic Health Survey 2003*.

Pretoria, South Africa: Department of Health.

NHTSA. 2012. *Traffic Safety Facts, 2010 data*. Washington, D.C.

Nicholson, A.J. 1987. Estimation of accident rates and countermeasure effectiveness. *Traffic engineering and control*. 28(10):518–523.

Nickerson, R.S. 2000. Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods*. 5(2):241–301.

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. *Accident Analysis & Prevention*. 36(4):525–532.

Nteziyaremye, P. 2013. Understandanding Pedestrian Crossing Behaviour: A Case Study in the Western Cape, South Africa. Stellebosch University.

Nteziyaremye, P. & Sinclair, M. 2013. Understanding pedestrian crossing behaviour: A case study in South Africa. *Proceedings of the 32nd Southern African Transport Conference (SATC 2013)*. 485–495.

Nyunt, M.S.Z., Shuvo, F.K., Eng, J.Y., Yap, K.B., Scherer, S., Hee, L.M., Chan, S.P. & Ng, T.P. 2015. Objective and subjective measures of neighborhood environment (NE): relationships with transportation physical activity among older persons. *The international journal of behavioral nutrition and physical activity*. 12:108.

O'sullivan, D. & Unwin, D. 2010. *Geographic information analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Office of the Deputy Prime Minister. 2006. *National Land Use Database: Land Use and Land Cover Classification, Version 4.4*. London, UK: Office of the Deputy Prime Minister.

Ogden, K.W. 1996. *Safer roads : a guide to road safety engineering*. Hampshire, England: Avebury Technical.

Ohm, B.W. & Sitkowski, R.J. 2004. Recent Developments in Land Use, Planning and Zoning: Integrating New Urbanism and Affordable Housing Tools. *The Urban Lawyer*. 36(4):857–866.

Ojungu-Omara, I. & Vanderschuren, M. 2006. Ways of reducing accidents on South African roads. *Proceedings of the 25th Southern African Transport Conference (SATC 2006)*. 454–464.

Ord, J.K. & Getis, A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*. 27(4):286–306.

Ordover, J.A., Sykes, A.O. & Willig, R.D. 1982. Herfindahl Concentration, Rivalry, and Mergers. *Harvard Law Review*. 95(8):1857–1874.

Orstad, S.L., McDonough, M.H., Stapleton, S., Altincekic, C. & Troped, P.J. 2017. A Systematic Review of Agreement Between Perceived and Objective Neighborhood Environment Measures and Associations With Physical Activity Outcomes. *Environment and Behavior*. 49(8):904–932.

Ortiz, N. & Ramnarayan, M. 2017. Analysis of Alcohol-Involved Pedestrian Fatalities in the United States, 2003–2015. *Journal of Transport & Health*. 5:S93.

Öström, M. & Eriksson, A. 2001. Pedestrian fatalities and alcohol. *Accident Analysis & Prevention*. 33(2):173–180.

Ouyang, Y. & Bejleri, I. 2014. Geographic Information System-Based Community-Level Method to Evaluate the Influence of Built Environment on Traffic Crashes. *Transportation Research Record: Journal of the Transportation Research Board*. 2432:124–132.

Oxley, J., Lenné, M.G. & Corben, B.F. 2006. The effect of alcohol impairment on road-crossing behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*. 9(4):258–268.

Oyana, T.J. & Margai, F. 2016. *Spatial Analysis: Statistics, Visualization, and Computational Methods*. Boca Raton, FL. USA: CRC Press.

Pafka, E. & Dovey, K. 2017. Permeability and interface catchment: measuring and mapping walkable access. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*. 10(2):150–162.

Pandey, J. 2014. *Geographic Information System*. The Energy and Resources Institute (TERI).

Papadimitriou, E. 2016. Towards an integrated approach of pedestrian behaviour and exposure. *Accident Analysis & Prevention*. 92:139–152.

Papadimitriou, E., Yannis, G. & Golias, J. 2009. A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*. 12(3):242–255.

Parks, J.R. & Schofer, J.L. 2006. Characterizing neighborhood pedestrian environments with secondary data. *Transportation Research Part D: Transport and Environment*. 11(4):250–263.

Parry, C.D.. & Bennetts, A.L. 1998. *Alcohol policy and public health in South Africa*. USA: Oxford University Press.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A.A., Jarawan, E.J. & Mathers, C. 2004. *World report on road traffic injury prevention*. Geneva: World Health Organization Geneva.

Peden, M., Oyegbite, K., Ozanne-Smith, J., Hyder, A.A., Branche, C., Rahman, A.F., Rivara, F. & Bartolomeos, K. 2008. *World Report on Child Injury Prevention*. Geneva: World Health Organization.

Peden, M.M., Knottenbelt, J.D., van der Spuy, J., Oodit, R., Scholtz, H.J. & Stokol, J.M. 1996. Injured pedestrians in Cape Town--the role of alcohol. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*. 86(9):1103–5.

Pei, X., Wong, S.C. & Sze, N.N. 2011. A joint-probability approach to crash prediction models. *Accident Analysis & Prevention*. 43(3):1160–1166.

Photius Coutsoukis and Information Technology Associates. n.d. *List of Townships in Western Cape, Republic of South Africa with Maps and Streets Views*. Vol. 2015.

PIARC Technical Committee on Road Safety. 2007. *Road Accident Investigation Guidelines for Road Engineers*. Paris, France: World Road Association (PIARC).

Pirdavani, A., Bellemans, T., Brijs, T. & Wets, G. 2014. Application of Geographically Weighted Regression Technique in Spatial Analysis of Fatal and Injury Crashes. *Journal of Transportation Engineering*. 140(8).

Plainis, S., Murray, I.J. & Pallikaris, I.G. 2006. Road traffic casualties: understanding the night-time death toll. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*. 12(2):125–8.

Du Plessis, D.J. 2015. Land-use mix in South African cities and the influence of spatial planning: Innovation or following the trend? *South African Geographical Journal*. 97(3):217–242.

Prijon, T. & Ermenc, B. 2009. Influence of alcohol intoxication of pedestrians on injuries in fatal road accidents. *Forensic Science International Supplement Series*. 1(1):33–34.

Prinsloo, M. 2001. Pedestrian Fatalities among children and adolescents in Cape Town, 2001. *MRC-UNISA Crime, Violence and Injury Lead Programme*.

Provincial Government of The Western Cape. 2007. *Road Traffic Accident Report*. Cape Town: Department of Transport and Public Works of Road Infrastructure Branch.

Pulugurtha, S.S., Duddu, V.R. & Kotagiri, Y. 2013. Traffic analysis zone level crash estimation models based on land use characteristics. *Accident Analysis and Prevention*. 50.

Quistberg, D.A., Howard, E.J., Ebel, B.E., Moudon, A. V., Saelens, B.E., Hurvitz, P.M., Curtin, J.E. & Rivara, F.P. 2015. Multilevel models for evaluating the risk of pedestrian-motor vehicle collisions at intersections and mid-blocks. *Accident Analysis and Prevention*. 84.

Rabianski, J., Gibler, K., Tidwell, O.A. & Clements III, J.S. 2009. Mixed-Use Development: A Call for Research. *Journal of Real Estate Literature*. 17(2):205–230.

Rajamani, J., Bhat, C., Handy, S., Knaap, G. & Song, Y. 2003. Assessing Impact of Urban Form Measures on Nonwork Trip Mode Choice After Controlling for Demographic and Level-of-Service Effects. *Transportation Research Record: Journal of the Transportation Research Board*. 1831:158–165.

Retting, R.A., Persaud, B.N., Garder, P.E. & Lord, D. 2001. Crash and injury reduction following installation of roundabouts in the United States. *American journal of public health*. 91(4):628–31.

Ribbens, H. 1985. *Pedestrian Casualties at Road Intersections and Suggested Engineering*

Countermeasures, NITRR Technical Report RV/14. Pretoria: CSIR.

Ribbens, H. 1996. Pedestrian Facilities in South Africa: Research and Practice. *Transportation Research Record*. (1538):10–18.

Rietveld, P. 2000. Non-motorised modes in transport systems: a multimodal chain perspective for The Netherlands. *Transportation Research Part D: Transport and Environment*. 5(1):31–36.

Rietveld, P. & Bruinsma, F. 1998. *Is Transport Infrastructure Effective? Transport Infrastructure and Accessibility: Impacts on the Space Economy*. Berlin, Germany: Springer-Verlag.

Road Safety Directorate. 2008. *Alcohol-related Crashes in Canada: Driver Characteristics and Casualty Trends*.

Robinson, B.W., Rodegerdts, L., Scarborough, W., Kittelson, W., Troutbeck, R., Brilon, W., Bondzio, L., Courage, K., et al. 2000. *Roundabouts: An Informational Guide*.

Rogerson, P. 2001. *Statistical Methods for Geography*. Sage.

Rosenbloom, T. 2009. Crossing at a red light: Behaviour of individuals and groups. *Transportation Research Part F: Traffic Psychology and Behaviour*. 12(5):389–394.

Rosenbloom, T., Nemrodov, D. & Barkan, H. 2004. For heaven's sake follow the rules: pedestrians' behavior in an ultra-orthodox and a non-orthodox city. *Transportation Research Part F: Traffic Psychology and Behaviour*. 7(6):395–404.

Rossi, J.S. 2012. Statistical Power Analysis. In Hoboken, NJ, USA: John Wiley & Sons, Inc. *Handbook of Psychology, Second Edition*. 71–108.

RTMC. 2011. *Road Traffic Report 31 March 2011*. Pretoria, South Africa: Department of Transport, South Africa.

RTMC. 2016. *Road Traffic Report Calendar: 1January-31 December 2015*. Pretoria, South Africa.

RTMC. 2017. *Road Traffic Report Calendar:1 January- 31 December 2016*. Pretoria, South Africa.

SAC Network. 2004. *State of the cities report 2004*. Johannesburg, South Africa.

SAC Network. 2011. *State of the Cities Report 2011*. Johannesburg.

Saelens, B.E. & Handy, S.L. 2008.

Sampson, J. 2017. Traffic Flow Variations in urban areas. In Pretoria: Southern African Transport Conference (SATC) *The 36th Southern African Transport Conference (SATC 2017)*. 712–727.

Sarkar, P.P. & Chunchu, M. 2016. Quantification and Analysis of Land-Use Effects on Travel Behavior in Smaller Indian Cities: Case Study of Agartala. *Journal of Urban Planning and Development*. 142(4):04016009.

Sarkar, P.P. & Mallikarjuna, C. 2013. Effect of Land Use on Travel Behaviour: A Case Study of Agartala City. *Procedia - Social and Behavioral Sciences*. 104:533–542.

Sarkar, C., Webster, C. & Gallacher, J. 2014. *Healthy cities: public health through urban planning*. Edward Elgar Publishing.

Shapiro, S.S. & Wilk, M.B. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*. 52(3/4):591–611.

Sheskin, D.J. 2011. Parametric Versus Nonparametric Tests. In M. Lovric (ed.). Springer Berlin Heidelberg *International Encyclopedia of Statistical Science*. 1051–1052.

Shinar, D. 2007. *Traffic Safety and Human Behavior*. 2nd Edition ed. Bingley, UK: Emerald Group Publishing Limited.

Shingala, M.C. & Rajyaguru, A. 2015. Comparison of Post Hoc Tests for Unequal Variance. *International Journal of New Technologies in Science and Engineering*. 2(5):2349–780.

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

Sinclair, M. & Zuidgeest, M. 2016. Investigations into pedestrian crossing choices on Cape Town freeways. *Transportation Research Part F: Traffic Psychology and Behaviour*. 42:479–494.

Smirnov, N. V. 1936. *"Sui la distribution de w^2 (Criterium de M.R.v. Mises)"*. Vol. 202. Paris: Comptes Rendus.

Smith, C.L. & Brooks, D.J. 2013. Chapter 4 - Built Environment. In C.L. Smith & D.J. Brooks (eds.). Boston: Butterworth-Heinemann *Security Science*. 81–103.

De Smith, M.J., Goodchild, M.F. & Longley, P.A. 2015. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. 5th ed. Winchelsea, UK: The Winchelsea Press.

Song, Y. & Knaap, G.-J. 2004. Measuring urban form: Is Portland winning the war on sprawl? *Journal of the American Planning Association*. 70(2):210–225.

Song, Y. & Rodriguez, D. 2005. *The measurement of the level of mixed land uses: a synthetic approach*. Chapel Hill,NC.

Song, Y., Merlin, L. & Rodriguez, D. 2013. Comparing measures of urban land use mix. *Computers, Environment and Urban Systems*. 42(0):1–13.

South African Government. n.d. *Public holidays in South Africa*. [Online], Available: https://www.gov.za/about-sa/public-holidays [2018, May 08].

Van der Spuy, J.W. 1991. Alcoho-related trauma. *Continuing Medical Education*. 9:859–868.

Van der Spuy, J.W. 2000. Trauma, alcohol and other substances. *South African Medical Journal (SAMJ)*. 90(3):244–246.

Stangl, P. 2012. The pedestrian route directness test: A new level-of-service model. *URBAN DESIGN International*. 17(3):228–238.

Statistics South Africa. 2010. *Concepts and Definitions for Statistics South Africa 2010 Version 3*. Pretoria: Statistics South Africa.

Statistics South Africa. 2015. *Census 2011: Population Dynamics in South Africa*. Pretoria, South Africa.

Statistics South Africa. n.d. *City of Cape Town*.

Steemers, K. 2003. Energy and the city: density, buildings and transport. *Energy and Buildings*.

35(1):3–14.

Stoker, P., Garfinkel-Castro, A., Khayesi, M., Odero, W., Mwangi, M.N., Peden, M. & Ewing, R. 2015. Pedestrian Safety and the Built Environment: A Review of the Risk Factors. *Journal of Planning Literature*. (AUGUST):1–16.

Stone, J., Chae, K. & Pillalamarri, S. 2002. *The effects of roundabouts on pedestrian safety*. The Southeastern Transport Center.

Stutts, J., Hunter, W. & Pein, W.E. 1996. Pedestrian crash types: 1990s update. *Transportation Research Record: Journal of the Transportation Research Board*. 1538:68–74.

Sukhai, A., Jones, A.P., Love, B.S. & Haynes, R. 2011. Temporal variations in road traffic fatalities in South Africa. *Accident Analysis & Prevention*. 43(1):421–428.

Surjono, A. & Ridhoni, M. 2017. Lessons learnt from and sustainability assessment of Indonesian urban kampong. *IOP Conference Series: Earth and Environmental Science*. 70(1):1–9.

Tegnér, G., Holmberg, I., Loncar-Lucassi, V. & Nilsson, C. 2000. The DRAG-Stockholm-2 Model. In M. Gaudry & S. Lassarre (eds.). Bingley, UK: Emerald Group Publishing Limited *Structural Road Accident Models: The International DRAG Family*. 127–156.

Thakali, L., Kwon, T.J. & Fu, L. 2015. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. *Journal of Modern Transportation*. 23(2):93–106.

Theofilatos, A., Yannis, G., Kopelias, P. & Papadimitriou, F. 2016. Predicting Road Accidents: A Rare-events Modeling Approach. *Transportation Research Procedia*. 14:3399–3405.

Thomson, J.A., Tolmie, A., Foot, H.C. & McLaren, B. 1996. *Child development and the aims of road safety education : a review and analysis*. H.M.Stationery Office Books.

Tian, G., Ewing, R., White, A., Hamidi, S., Walters, J., Goates, J.P. & Joyce, A. 2015. Traffic Generated by Mixed-Use Developments. *Transportation Research Record: Journal of the Transportation Research Board*. 2500:116–124.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region.

Economic geography. 46(sup1):234–240.

du Toit, N. & van As, A.B. 2001. Paediatric pedestrian carnage. *Injury and Safety Monitor*.

Tom, A. & Granié, M.-A. 2011. Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads. *Accident; Analysis and Prevention*. 43(5):1794–1801.

Toroyan, T. & Peden, M. (Eds). 2007. *Youth and road safety*. Geneva, Switzerland: World Health Organization.

TRB. 2005. *Does the Built Environment Influence Physical Activity? Examining the Evidence*. Washington.

Troped, P.J., Wilson, J.S., Matthews, C.E., Cromley, E.K. & Melly, S.J. 2010. The Built Environment and Location-Based Physical Activity. *American Journal of Preventive Medicine*. 38(4):429–438.

Tukey, J.W. 1977. *Exploratory data analysis*. Reading, MA.: Addison-Wesley.

Turner, M.G., Gardner, R.H. & O'Neill, R. V. 2001. *Landscape Ecology in Theory and Practice*. Vol. 401 ed. New York: Springer. [Online], Available: http://www.springer.com/la/book/9780387951225.

Turok, I. 1994. Urban Planning in the Transition from Apartheid: Part 1: The Legacy of Social Control. *The Town Planning Review*. 65(3):243–259.

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

Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai, G. & Isa-Tavarez, J. 2012. The role of built environment on pedestrian crash frequency. *Safety Science*. 50(4):1141–1151.

Ulf, B. & Jorgen, L. 1999. *Traffic Safety of Roundabouts for Cyclists and Pedestrians*. Linköping: Swedish National Road and Transport Research Institute (VTI).

Vanlaar, W. 2005. Drink driving in Belgium: results from the third and improved roadside survey. *Accident Analysis & Prevention*. 37(3):391–397.

Vellios, N.G. & Van Walbeek, C.P. 2017. Self-reported alcohol use and binge drinking in South Africa: Evidence from the National Income Dynamics Study, 2014 - 2015. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*. 108(1):33–39.

Venter, C., Minora, N., Shukrani, K. & du Toit, J. 2014. The role for GPS data in qualitative research: Exploring links between walking behaviour, the built environment and crime perception in South Africa. In S. Rasouli & H. Timmermans (eds.). Hershey PA, USA: IGI Global *Mobile Technologies for Activity-Travel Data Collection and Analysis*. 309–324.

Victoria Transport Policy Institute. 2012. *Roadway Connectivity: Creating More Connected Roadway and Pathway Network*. Victoria, BC, CA: Victoria Transport Policy Institute. [Online], Available: http://www.vtpi.org/tdm/tdm116.htm.

Vogel, L. & Bester, C.J. 2005. A relationship between accident types and causes. *Proceedings of the 24th Southern African Transport Conference (2005)*. 233–241.

Wade, T. & Sommer, S. 2006. *A to Z GIS: An illustrated dictionary of geographic information systems*. Redlands, CA, USA: ESRI Press.

Washington, S.P., Karlaftis, M.G. & Mannering, F. 2010. *Statistical and econometric methods for transportation data analysis*. 2nd ed. Boca Raton, USA: CRC press.

Wedagama, D.M.P., Bird, R. & Dissanayake, D. 2008. The Influence of Urban Land Use on Pedestrian Casualties: Case Study Area: Newcastle upon Tyne, UK. *IATSS Research*. 32(1):62–73.

Wedagama, P.D.M., Bird, R.N. & Metcalfe, A. V. 2006. The influence of urban land-use on non-motorised transport casualties. *Accident Analysis & Prevention*. 38(6):1049–1057.

WESGRO. 2016. *Sector: City of Cape Town*. Cape Town, South Africa: WESGRO Cape Town & Western Cape Research.

Wheeler, D., Shaw, G. & Barr, S. 2013. *Statistical Techniques in Geographical Analysis*. 3rd ed. David Fulton Publishers.

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. *Accident Analysis & Prevention*. 41(1):137–145.

Wilson, J., Fang, M. & LeBrun, D. 2003. Locations for collisions involving alcohol and drugs impaired pedestrians in urban areas. In *ICTCT-extra workshop (held in Vancouver),International cooperation on theories and concepts in traffic safety*.

Wilson, M.H., Baker, S.P., Teret, S.P., Schock, S. & Garbarino, J. 1991. *Saving children : a guide to injury prevention*. New York: Oxford University Press.

Witherspoon, R., Abbett, J.P. & Gladstone, R.M. 1976. *Mixed-use developments: new ways of land use*. Washington, DC: Urban Land Institute.

Wolfe, A.C. 1982. The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods. *Accident Analysis & Prevention*. 14(5):337–340.

Wong, D.W. 2009. Modifiable Areal Unit Problem. In Elsevier *International Encyclopedia of Human Geography*. 169–174.

Wong, D.W.S. & Wang, F. 2017. Spatial Analysis Methods. In B. Huang (ed.). Elsevier *Comprehensive Geographic Information Systems, 1st Edition*.

Wood, L., Frank, L.D. & Giles-Corti, B. 2010. Sense of community and its relationship with walking and neighborhood design. *Social Science & Medicine*. 70(9):1381–1390.

World Health Organization. 2011. *Global status report on alcohol and health*. Geneva: World Health Organization.

World Health Organization. 2013. *Global status report on road safety 2013:supporting a decade of action*. Geneva: World Health Organization. [Online], Available: http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/ [2017, December 30].

World Health Organization. 2015. *Global Status Report on Road Safety 2015*. Geneva, Switzerland: World Health Organization. [Online], Available: www. who.int.

Xie, Z. & Yan, J. 2013. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport*

Geography. 31:64–71.

Xu, P. & Huang, H. 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis & Prevention*. 75(Supplement C):16–25.

Xu, P., Huang, H. & Dong, N. 2018. The modifiable areal unit problem in traffic safety: Basic issue, potential solutions and future research. *Journal of Traffic and Transportation Engineering (English Edition)*. 5(1):73–82.

Yagil, D. 2000. Beliefs, motives and situational factors related to pedestrians' self-reported behavior at signal-controlled crossings. *Transportation Research Part F: Traffic Psychology and Behaviour*. 3(1):1–13.

Yang, H.J., Song, J. & Choi, M.J. 2016. Measuring the Externality Effects of Commercial Land Use on Residential Land Value: A Case Study of Seoul. *Sustainability*. 8(5):432.

Yao, S., Loo, B.P.Y. & Lam, W.W.Y. 2015. Measures of activity-based pedestrian exposure to the risk of vehicle-pedestrian collisions: Space-time path vs. potential path tree methods. *Accident Analysis & Prevention*. 75:320–332.

Yong, A.G. & Pearce, S. 2013. A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*. 9(2):79–94.

Yue, Y., Zhuang, Y., Yeh, A.G.O., Xie, J.-Y., Ma, C.-L. & Li, Q.-Q. 2017. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*. 31(4):658–675.

Zahabi, S.A.H., Strauss, J., Manaugh, K. & Miranda-Moreno, L.F. 2011. Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes. *Transportation Research Record: Journal of the Transportation Research Board*. 2247:81–90.

Zhang, L., Nasri, A., Hong, J.H. & Shen, Q. 2012. How built environment affects travel behavior: A comparative analysis of the connections between land use and vehicle miles traveled in US cities. *Journal of Transport and Land Use*. 5(3).

Zhang, Y., Wang, X., Zeng, P. & Chen, X. 2011. Centrality Characteristics of Road Network

Patterns of Traffic Analysis Zones. *Transportation Research Record: Journal of the Transportation Research Board*. 2256:16–24.

Zhang, Y., Bigham, J., Li, Z., Ragland, D. & Chen, X. 2012. Associations between road network connectivity and pedestrian-bicyclist accidents. In Washington, DC: Transportation Research Board *The 91st Annual Meeting of the Transportation Research Board*.

Zhang, Y., Bigham, J., Ragland, D. & Chen, X. 2015. Investigating the associations between road network structure and non-motorist accidents. *Journal of Transport Geography*. 42:34–47.

Zhao, F. & Park, N. 2004. Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. *Transportation Research Record: Journal of the Transportation Research Board*. 1879:99–107.

Zhou, X. & Lin, H. 2008. Geary's Index. In S. Shekhar & H. Xiong (eds.). New York, USA: Springer *Encyclopedia of GIS*.

Zhu, X. 2016. *GIS for Environmental Applications: a Practical Approach*. Routledge.

Zoni, A.C., Domínguez-Berjón, M.F., Esteban-Vasallo, M.D., Velázquez–Buendía, L.M., Blaya-Nováková, V. & Regidor, E. 2016. Socioeconomic inequalities in injuries treated in primary care in Madrid, Spain. *Journal of Public Health*. 39(1):fdw005.

Appendices

APPENDIX A

Descriptions of zonings and subzonings

ZONINGS	DESCRIPTIONS
1. Single residential zonings	The single residential zonings accommodate predominantly single-family dwelling houses in low-to medium density neighbourhoods with a safe and pleasant living environment. Two single residential zonings, one for conventional housing and another for incremental housing are considered in recognition of different socio-economic conditions of citizens in the City.
Single residential zoning 1: Conventional housing (SR1)	The SR1 zoning accommodates largely single-family dwelling houses and additional use rights in low-to medium-density residential neighbourhoods, on small or large land parcels.
Single residential zoning 2: Incremental housing (SR2)	The SR2 zoning incorporates dwelling houses that were upgraded from an informal settlement to a formal settlement. SR2 may also apply to individual land parcels or blocks containing an informal settlement.
2. General residential zonings	The general residential zonings are designed with the aim to promote healthy, safe and pleasant living environments in high-density urban settlements. Different zonings and subzonings allow diverse levels of development intensity, particularly relating to height and floor space.
General residential subzoning 1: Group housing (GR1)	The GR1 zoning is planned to accommodate group housing in medium-density residential developments where aesthetics, architectural form and inter-relationship between various components of the development are taken into consideration.
General residential subzonings (GR2, GR3, GR4,GR5 & GR6)	These subzonings encourage high-density residential development, including blocks and flats. Diverse subzones are subjected to different development rules, particularly with respect to height and floor space, to allow a variety of building forms. The GR2 subzoning includes flats of relatively low height and small floor space. The GR3 and GR4 subzonings accommodate flats of medium height and floor space. The GR5 and GR6 subzonings cater for high-rise flats. The predominant use in these subzoninds is residential but limited mixed-use development can be permitted.
3. Community zonings	Community zonings accommodate land used for social needs of communities such as educational, religious, welfare or health services. There are two community zonings depending on the size of the community served by the zoning.
Community zoning 1: Local (CO1)	The CO1 zoning serves social needs (education, worship and health) of local community. However other use with greater social impact to community can be approved by the city.
Community zoning 2: Regional (CO2)	The CO2 zoning comprises land developed to serve community social needs (health, welfare, worship and education) at a local or regional scale.

ZONINGS	DESCRIPTIONS
4. Local business zonings	Local business zonings create a suitable interface between business districts and adjacent residential areas. Office use of low impact and associated use are permitted but retail use of higher impact is controlled.
Local business zonings 1: Intermediate business (LB1)	The LB1 zoning serves as a buffer or interface between residential areas and general business zonings or other high-intensity non-residential uses. The prevailing uses should be for residential, office and associated purposes, but limited retail activities are permitted with the City's approval.
Local business zoning 2: Local business (LB2)	The LB2 zoning accommodates low-intensity commercial and mixed-use developments.
5. General business and mixed use zonings	The general business zonings include uses such as business, residential and community uses. They are planned to support economic development in business and development corridors. Industrial development in these zonings is restricted. By contrast, mixed use zonings comprise developments with a complete mixture of land uses including industrial, business and residential development.
General business subzonings (GB1, GB2, GB3, GB4, GB5 & GB7)	The GB zonings cater for general business activity and mixed-use development of a medium-to high-intensity. Different development rules apply to the different subzonings of GB1-GB7, particularly with regard to permitted height and floor space, in order to allow a variety of building forms within the city.
Mixed use subzoning (MU1, MU2 & MU3)	The MU zonings accommodate a mixture of business, industrial and residential development. These zonings are particularly suitable at the interface between general business and industrial zonings. Different development rules apply to the different subzonings of MU1, MU2 and MU3, particularly with regard to permitted height and floor space.
6. Industrial zonings	The industrial zonings are planned to accommodate manufacturing and related processes, ranging from general industrial uses which may have some impact on surrounding areas, to hazardous or noxious uses which have a potentially high impact and must be carefully managed. Two different subzonings; (GI1&GI2 and RI) are considered in this subzoning depending on associated environmental impacts.
General industrial subzonings (GI1 & GI2)	The GI zoning accommodates all forms of industry, except noxious trade and risk activity. Some allowance is made for non-industrial activities, but these should not compromise the general use of the area zoned for industry.
Risk industry zoning (RI)	The RI zoning include those industries which are noxious in terms of smell, product, waste or other objectionable consequence of their operation, or which carry a high risk in the event of fire or accident.

ZONINGS	DESCRIPTIONS
7. Utility, transport and national zonings	Certain government activities which cannot be classified into other zonings are included in the utility zonings. Transport zonings are designed to facilitate efficient operation of the various transport systems. Another zoning included in this category is the national port zoning.
Utility zonings (UT)	The UT zoning accommodates utility services such as electrical substations and water reservoirs, which may be supplied by a municipal, government or private agency. This zoning category also includes government or authority uses, such as prisons and military bases, which are not covered by another use or zoning category.
Transport zoning 1:Transport Use (TR1)	The TR1 zoning provides for transportation systems, excluding public roads and public streets, but including all other transport undertakings which serve the public such as airports, harbours, railway lines, bus, railway and other depots associated with public transport uses, public transport terminuses, ranks or holding areas, and cable car stations.
Transport zonings 2 : Public road and public parking (TR2)	The TR2 zoning includes public streets and roads, whether constructed or still to be constructed, as well as premises for the public parking of operable motor vehicles. Such parking may be provided in buildings or open parking areas, with or without the payment of a fee. On-site parking for a permitted activity in any zoning is considered to be an associated use and do not represent a separate use category that requires separate zoning or approval.
National port zoning (NP)	The NP zoning is provided as a zoning in which land use within a national port is controlled by an approved port development framework plan.
8. Open space zonings	This zoning consists of 3 different types of open space fulfilling different functions. Certain open spaces have particular importance as nature, cultural heritage or environmental areas and a separate zoning facilitates the management of these areas. Within this zoning provision is made for the development of amenities to meet the needs of tourists and visitors. Other open spaces have a more active role in addressing the sporting and recreation needs of the community.
Open space zoning 1: Environmental conservation (OS1)	The OS1 zoning provides for the conservation of environmental resources, although cultural heritage resources may also be included. Provision is made for limited, low-impact uses associated with conservation, such as environmental education, associated infrastructure and facilities for tourists and visitors with the approval of the City.
Open space zoning 2: Public open space (OS2)	The OS2 zoning accommodates active and passive recreational areas on public land, as well as protection of landscape and heritage areas including woodlands, ridges, watercourses, wetlands and the coastline.
Open space zoning 3: Special open space (OS3)	The OS3 zoning includes active or passive recreation and open spaces on land that is not designated as public open space. This land may be owned by private or public bodies, but does not have the status of public open space which requires particular protection.

ZONINGS	DESCRIPTIONS
9. Agricultural, rural and limited use zonings	This zoning is planned to protect land suitable for agriculture and to help to maintain its aesthetic and cultural value. Aside from sustaining a valuable economic sector, agricultural land can help to promote stability of the urban edge, conserve naturally sensitive areas and maintain rural characteristics which are valued by the community.
Agricultural zoning (AG)	The AG zoning promotes and protects agriculture on farms as an important economic, environmental and cultural resource.
Rural zoning (RU)	The RU zoning accommodates smaller rural properties that may be used for agriculture, but which may also be occupied as places of residence by people who seek a country lifestyle, and who view agriculture as a secondary reason for occupying their property.
Limited use zoning (LU)	The LU zoning is a transitional mechanism to deal with land that was zoned as undetermined in previous zoning schemes. The aim is to progressively phase this zoning out and so no property should be rezoned to this zoning.

APPENDIX B

Data completeness for land use data and transportation systems data

Data category	**Dataset**	**Data fields**	**Completeness of fields**	
			Available data/Total number of features	**%**
Land use	Zoning Category	Zoning /subzoning type	736412/751128	98
		area	751128/751128	100
Road network	Road segment	Road name	172447/247006	70
		Road number	20955/247006	8
		Segment length	247006/247006	100
		Speed limit	218006/247006	88
		Average speed	247006/247006	100
		Centre coordinates (X&Y)	247006/247006	100
		Segment length	247006/247006	100
Classified road network	Freeways	Road name	350/436	80
		Road number	416/436	95
		Speed limit	418/436	96
		Segment length	422/436	97
	Expressways	Road name	130/139	94
		Road number	128/139	92
		Speed limit	131/139	94
		Segment length	133/139	96
	Local distributors	Road name	392/435	90
		Road number	6/435	1
		Speed limit	81/435	19
		Segment length	129/435	30
	Primary arterials	Road name	411/482	85
		Road number	323/482	67
		Speed limit	431/482	89
		Segment length	450/482	93
	Secondary arterials	Road name	1026/1041	99
		Road number	428/1041	41
		Speed limit	977/1041	94
		Segment length	964/1041	93
	Reclassified to secondary arterial	Road name	2/2	100
		Road number	0/2	0
		Speed limit	0/2	0
		Segment length	0/2	0
	Minibus-taxi routes	Route name	1466/1466	100

Public transport systems		Origin & Destination	1447/1466	99
		Segment length	1466/1466	100
	Bus route (Golden Arrow Bus Services)	Route name	1985/1985	100
		Origin & Destination	1985/1985	100
		Segment length	1985/1985	100
	Integrated Rapid Transit (IRT)_Bus	Route name	135/135	100
		Origin & Destination	135/135	100
		Service Type	46/135	34
		Segment length	127/135	94
	Railway routes	Route name	158/158	100
		Segment length	158/158	100
	Railway stations	Station name	92/92	100
	IRT Bus stops	Stop name	632/632	100
		Stop number	238/632	38
		Road name	496/632	78
		Shelter type	248/632	39

APPENDIX C

Computational details of land-use mix by the Relative Entropy Index

Step 1: Calculation of distribution of land use types within the study area

Suburb (i)	land use type (j) SR1 1	SR2 2	.	.	AG k	$X_i=\sum x_{ij}$
1	$x_{1,1}$	$x_{1,2}$			$x_{1,k}$	$\sum x_{1,j}$
2	$x_{2,1}$	$x_{2,2}$	.	.	$x_{2,k}$	$\sum x_{2,j}$
3	$x_{3,1}$	$x_{3,2}$	.	.	$x_{3,k}$	$\sum x_{3,j}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
190(n)	$x_{190,1}$	$x_{190,2}$	.	.	$x_{190,k}$	$\sum x_{190,j}$
X_j	$\sum x_{i,1}$	$\sum x_{i,2}$	.	.	$\sum x_{i,k}$	$Z=\sum\sum x_{i,j}$
t_j	$(\sum x_{i,1})/Z$	$(\sum x_{i,2})/Z$			$(\sum x_{i,K})/Z$	1

Step 2: Calculation of percentages of land use type (r_{ij}) in the unit of analysis as specified in Equation (7)

Suburb (i)	Land use percentage $r_{ij}=(x_{ij})/X_i$ SR1 1	SR2 2	.	.	AG k	$R_i=\sum r_{ij}$
1	$r_{1,1}$	$r_{1,2}$			$r_{1,k}$	$\sum r_{1,j}$
2	$r_{2,1}$	$r_{2,2}$	.	.	$r_{2,k}$	$\sum r_{2,j}$
3	$r_{3,1}$	$r_{3,2}$	.	.	$r_{3,k}$	$\sum r_{3,j}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
190(n)	$r_{190,1}$	$r_{190,2}$	.	.	$r_{190,k}$	$\sum r_{190,j}$

Step 3: Calculation of quotients q_{ij} as specified in Equation (9)

Suburb (i)	Quotients $q_{ij}=(r_{ij})/t_j$					$Q_i=\sum q_{ij}$
	SR1 1	SR2 2	.	.	AG k	
1	$q_{1,1}$	$q_{1,2}$			$q_{1,k}$	$\sum q_{1,j}$
2	$q_{2,1}$	$q_{2,2}$	.	.	$q_{2,k}$	$\sum q_{2,j}$
3	$q_{3,1}$	$q_{3,2}$	.	.	$q_{3,k}$	$\sum q_{3,j}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
190(n)	$q_{190,1}$	$q_{190,2}$	.	.	$q_{190,k}$	$\sum q_{190,j}$

Step 4: Calculation of land use percentages P_{ij} as specified in Equation (10)

Suburb (i)	Adapted land use percentages $P_{ij}=(q_{ij})/\sum q_{ij}$					$\sum P_{ij}$
	SR1 1	SR2 2	.	.	AG k	
1	$P_{1,1}$	$P_{1,2}$			$P_{1,k}$	1
2	$P_{2,1}$	$P_{2,2}$	.	.	$P_{2,k}$	1
3	$P_{3,1}$	$P_{3,2}$	.	.	$P_{3,k}$	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
190(n)	$P_{190,1}$	$P_{190,2}$	.	.	$P_{190,k}$	1

Step 5.A: Calculation of the Relative Entropy Index as specified in Equation (1)

Suburb (i)	ln(P_{ij})					Entropy Index
	SR1 1	SR2 2	.	.	AG k	
1	$\ln(P_{1,1})$	$\ln(P_{1,2})$			$\ln(P_{1,k})$	$(-1)\times\sum[(P_{1j})\times\ln(P_{1j})]/\ln(k)$
2	$\ln(P_{2,1})$	$\ln(P_{2,2})$	.	.	$\ln(P_{2,k})$	$(-1)\times\sum[(P_{2j})\times\ln(P_{2j})]/\ln(k)$
3	$\ln(P_{3,1})$	$\ln(P_{3,2})$	.	.	$\ln(P_{3,k})$	$(-1)\times\sum[(P_{3j})\times\ln(P_{3j})]/\ln(k)$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
190(n)	$\ln(P_{190,1})$	$\ln(P_{190,2})$	.	.	$\ln(P_{190,k})$	$(-1)\times\sum[(P_{190,j})\times\ln(P_{190,j})]/\ln(k)$

Step 5.B: Calculation of Herfindahl-Hirschman Index (HHI) following Equation (3)

Suburb (i)	$(100\times P_{ij})^2$					HHI
	SR1 1	SR2 2	.	.	AG k	
1	$(100\times P_{1,1})^2$	$(100\times P_{1,2})^2$			$(100\times P_{1,k})^2$	$\sum(100\times P_{1,j})^2$
2	$(100\times P_{2,1})^2$	$(100\times P_{2,2})^2$	.	.	$(100\times P_{2,k})^2$	$\sum(100\times P_{2,j})^2$
3	$(100\times P_{3,1})^2$	$(100\times P_{3,2})^2$	.	.	$(100\times P_{3,k})^2$	$\sum(100\times P_{3,j})^2$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
190(n)	$(100\times P_{190,1})^2$	$(100\times P_{190,2})^2$	.	.	$(100\times P_{190,k})^2$	$\sum(100\times P_{190,j})^2$

APPENDIX D

Crash details collected on the Accident Report form in South Africa

2. Number of persons seriously injured: 4. Number of persons not injured:

PARTICULARS OF PASSENGERS WHO ARE NOT INJURED

Passenger number in vehicle (A, B, etc)
ID number / Telephone/Cellphone number ( ) H W

Passenger number in vehicle (A, B, etc)
ID number / Telephone/Cellphone number ( ) H W

Passenger number in vehicle (A, B, etc)
ID number / Telephone/Cellphone number ( ) H W

PARTICULARS OF KILLED OR INJURED PASSENGERS AND PEDESTRIANS

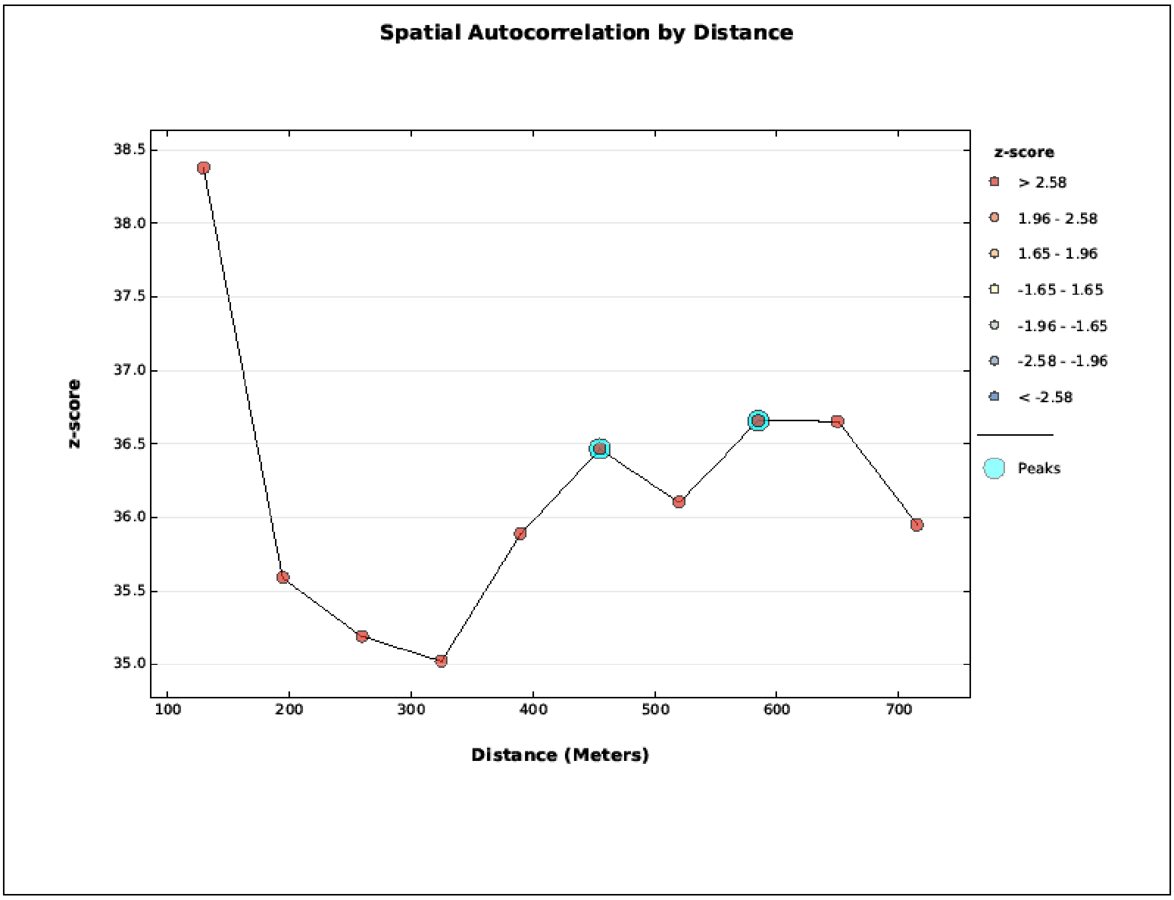
Passenger number (1, 2, etc.) in vehicle (A, B, etc.) Pedestrian (P, Q, etc.)		Passenger number (1, 2, etc.) in vehicle (A, B, etc.) Pedestrian (P, Q, etc.)
/	ID type/ ID number	/
	Country of origin of ID	
	Surname	
Initials Age		Initials Age
	Home/contact address	
( ) H W	Telephone number	( ) H W
( ) H W	Cellphone/other number	( ) H W
1. Asian 2. Black 3. Coloured 4. White 98. Other 00. Unknown	How would you describe the person?	1. Asian 2. Black 3. Coloured 4. White 98. Other 00. Unknown
1. Male 2. Female 0. Unknown	Gender	1. Male 2. Female 0. Unknown
1. Killed 2. Serious 3. Slight 4. No injury	Severity of injury	1. Killed 2. Serious 3. Slight 4. No injury
	Ambulance service, driver, case reference number & hospital	
1. Yes 2. No 0. Unknown	Seatbelt fitted/helmet present	1. Yes 2. No 0. Unknown
1. Yes 2. No 0. Unknown	Seatbelt/helmet **definitely** used	1. Yes 2. No 0. Unknown
1. Yes 2. No	Liquor/drug use **suspected**	1. Yes 2. No
1. Yes 2. No	*Liquor/drug use: **evidentiary tested**	1. Yes 2. No

Passenger number (1, 2, etc.) in vehicle (A, B, etc.) Pedestrian (P, Q, etc.)		Passenger number (1, 2, etc.) in vehicle (A, B, etc.) Pedestrian (P, Q, etc.)
/	ID type/ ID number	/
	Country of origin of ID	
	Surname	
Initials Age		Initials Age
	Home/contact address	
( ) H W	Telephone/contact number	( ) H W
( ) H W	Cellphone/other number	( ) H W
1. Asian 2. Black 3. Coloured 4. White 98. Other 00. Unknown	How would you describe the person?	1. Asian 2. Black 3. Coloured 4. White 98. Other 00. Unknown
1. Male 2. Female 0. Unknown	Gender	1. Male 2. Female 0. Unknown
1. Killed 2. Serious 3. Slight 4. No injury	Severity of injury	1. Killed 2. Serious 3. Slight 4. No injury
	Ambulance service, driver, case reference number & hospital	
1. Yes 2. No 0. Unknown	Seatbelt fitted/helmet present	1. Yes 2. No 0. Unknown
1. Yes 2. No 0. Unknown	Seatbelt/helmet **definitely** used	1. Yes 2. No 0. Unknown
1. Yes 2. No	Liquor/drug use **suspected**	1. Yes 2. No
1. Yes 2. No	*Liquor/drug use: **evidentiary tested**	1. Yes 2. No

PAGE 3

APPENDIX E

APPENDIX E1: Output report from Incremental Spatial Autocorrelation tool



APPENDIX E2: Output report from Incremental Spatial Autocorrelation tool

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
130.00*	0.703843	-0.000392	0.000337	38.380445	0.000000
195.00*	0.583490	-0.000356	0.000269	35.592724	0.000000
260.00*	0.521051	-0.000334	0.000219	35.193863	0.000000
325.00*	0.466839	-0.000317	0.000178	35.024308	0.000000
390.00*	0.441923	-0.000308	0.000152	35.892566	0.000000
455.00*	0.408617	-0.000302	0.000126	36.468493	0.000000
520.00*	0.375938	-0.000297	0.000109	36.106347	0.000000
585.00*	0.351124	-0.000293	0.000092	36.659194	0.000000
650.00*	0.325070	-0.000291	0.000079	36.655097	0.000000
715.00*	0.296909	-0.000290	0.000068	35.952486	0.000000

First Peak (Distance, Value): 455.00, 36.468493

Max Peak (Distance, Value): 585.00, 36.659194

Distance measured in Meters

* At least one distance increment resulted in features with no neighbors which may invalidate the significance of the corresponding results.

APPENDIX F

APPENDIX F1: Model results for weekday pedestrian casualties (Model 4)

Variables	WeekdayPedCas - Parameter estimates Distribution : NEGATIVE BINOMIAL Link function: LOG						
	Estimate (B)	Standard Error	Wald Stat.	Lower CL 95.0%	Upper CL 95.0%	p	Exp(B)
Intercept	-0.2127	0.4631	0.2109	-1.1203	0.6949	0.6461	0.8084
Log_Popu	1.4715	0.1482	98.6201	1.1811	1.7619	0.0000	4.3557
Prop_White	-0.0113	0.0032	12.4662	-0.0176	-0.0050	0.0004	0.9887
Prop_AgeLess15	-0.0489	0.0127	14.8253	-0.0737	-0.0240	0.0001	0.9523
Prop_Age15_24	-0.0408	0.0077	27.9716	-0.0560	-0.0257	0.0000	0.9600
Prop_Age25_54	-0.0252	0.0074	11.4901	-0.0398	-0.0106	0.0007	0.9751
Prop_AvgEd	-0.0152	0.0052	8.6228	-0.0253	-0.0050	0.0033	0.9850
Prop_UpperInc	-0.0202	0.0065	9.7594	-0.0328	-0.0075	0.0018	0.9800
ENT_9Cat	1.0756	0.3617	8.8420	0.3666	1.7845	0.0029	2.9317
Prop_GI9Cat	0.0295	0.0041	51.0211	0.0214	0.0376	0.0000	1.0299
Inters_grt3leg	0.0027	0.0008	10.2284	0.0010	0.0043	0.0014	1.0027
StrDens	0.0279	0.0109	6.5990	0.0066	0.0492	0.0102	1.0283
Prop_Freeways	0.0357	0.0073	23.8080	0.0213	0.0500	0.0000	1.0363
Prop_Expresways	0.0607	0.0132	21.0301	0.0348	0.0866	0.0000	1.0626
Prop_PrimaryArter	0.0153	0.0106	2.0751	-0.0055	0.0361	0.1497	1.0154
Prop_SecondArter	0.0099	0.0075	1.7136	-0.0049	0.0246	0.1905	1.0099
Round_Circ	0.0368	0.0105	12.4075	0.0163	0.0573	0.0004	1.0375
Prop_Signal	0.0912	0.0220	17.2345	0.0481	0.1343	0.0000	1.0955
Dispersion	0.4197	0.0521		0.3177	0.5218		

APPENDIX F2: Model results for Saturday pedestrian casualties (Model 5)

Variables	SatPedCas - Parameter estimates Distribution : NEGATIVE BINOMIAL Link function: LOG						
	Estimate (B)	Standard Error	Wald Stat.	Lower CL 95.0%	Upper CL 95.0%	p	Exp(B)
Intercept	-1.0284	0.5482	3.5187	-2.1029	0.0461	0.0607	0.3576
Log_Popu	1.4657	0.1698	74.4759	1.1328	1.7986	0.0000	4.3305
Prop_White	-0.0116	0.0039	9.0913	-0.0192	-0.0041	0.0026	0.9884
Prop_AgeLess15	-0.0349	0.0151	5.3599	-0.0644	-0.0054	0.0206	0.9657
Prop_Age15_24	-0.0454	0.0100	20.6743	-0.0649	-0.0258	0.0000	0.9557
Prop_Age25_54	-0.0216	0.0087	6.1220	-0.0387	-0.0045	0.0134	0.9786
Prop_AvgEd	-0.0149	0.0061	6.0366	-0.0268	-0.0030	0.0140	0.9852
Prop_UpperInc	-0.0333	0.0078	18.3696	-0.0485	-0.0181	0.0000	0.9672
ENT_9Cat	0.8049	0.3934	4.1857	0.0338	1.5760	0.0408	2.2365
Prop_GI9Cat	0.0171	0.0046	13.8550	0.0081	0.0262	0.0002	1.0173
Inters_grt3leg	0.0028	0.0008	11.8357	0.0012	0.0043	0.0006	1.0028
StrDens	0.0062	0.0119	0.2735	-0.0171	0.0295	0.6010	1.0062
Prop_Freeways	0.0181	0.0083	4.7518	0.0018	0.0343	0.0293	1.0182
Prop_Expresways	0.0318	0.0143	4.9144	0.0037	0.0599	0.0266	1.0323
Prop_PrimaryArter	0.0161	0.0120	1.8013	-0.0074	0.0395	0.1796	1.0162
Prop_SecondArter	0.0068	0.0090	0.5671	-0.0109	0.0244	0.4514	1.0068
Round_Circ	0.0260	0.0104	6.2916	0.0057	0.0463	0.0121	1.0263
Prop_Signal	0.0808	0.0238	11.5565	0.0342	0.1274	0.0007	1.0841
Dispersion	0.3131	0.0601		0.1954	0.4309		

APPENDIX F3: Model results for Sunday pedestrian casualties (Model 6)

Variables	SunPedCas - Parameter estimates Distribution : NEGATIVE BINOMIAL Link function: LOG						
	Estimate (B)	Standard Error	Wald Stat.	Lower CL 95.0%	Upper CL 95.0%	p	Exp(B)
Intercept	-1.2010	0.6122	3.8486	-2.4009	-0.0011	0.0498	0.3009
Log_Popu	1.4108	0.1901	55.0630	1.0381	1.7834	0.0000	4.0991
Prop_White	-0.0203	0.0044	21.5622	-0.0288	-0.0117	0.0000	0.9799
Prop_AgeLess15	-0.0093	0.0170	0.2971	-0.0427	0.0241	0.5857	0.9908
Prop_Age15_24	-0.0693	0.0148	21.8033	-0.0983	-0.0402	0.0000	0.9331
Prop_Age25_54	-0.0122	0.0103	1.4025	-0.0325	0.0080	0.2363	0.9878
Prop_AvgEd	-0.0217	0.0071	9.1942	-0.0357	-0.0077	0.0024	0.9786
Prop_UpperInc	-0.0297	0.0086	11.8281	-0.0466	-0.0128	0.0006	0.9707
ENT_9Cat	0.6821	0.4348	2.4614	-0.1700	1.5342	0.1167	1.9780
Prop_GI9Cat	0.0051	0.0052	0.9527	-0.0052	0.0154	0.3290	1.0051
Inters_grt3leg	0.0031	0.0009	13.1275	0.0014	0.0048	0.0003	1.0031
StrDens	0.0076	0.0130	0.3422	-0.0179	0.0332	0.5586	1.0077
Prop_Freeways	0.0352	0.0086	16.5398	0.0182	0.0521	0.0000	1.0358
Prop_Expresways	0.0224	0.0165	1.8481	-0.0099	0.0548	0.1740	1.0227
Prop_PrimaryArter	0.0302	0.0127	5.6400	0.0053	0.0551	0.0176	1.0307
Prop_SecondArter	0.0188	0.0101	3.4876	-0.0009	0.0386	0.0618	1.0190
Round_Circ	0.0195	0.0112	3.0104	-0.0025	0.0415	0.0827	1.0197
Prop_Signal	0.0494	0.0262	3.5563	-0.0019	0.1007	0.0593	1.0506
Dispersion	0.3274	0.0715		0.1873	0.4675		

APPENDIX G

APPENDIX G1: Summary statistics for local estimates from GWR Model 1B

Local parameters	Summary statistics for local estimates from GWR Model 1B							
	Mean	Minimum	Maximum	Std. Dev.	Global Model	t-value	df	p
Coeff Intercept	-33.446	-33.468	-33.426	0.010	-33.449	3.146	189	0.002
Coeff Log_Popu	8.437	8.424	8.457	0.006	8.435	6.580	189	0.000
Coeff Prop_AgeLess15	1.589	1.589	1.590	0.000	1.589	15.085	189	0.000
Coeff Prop_AvgEd	-0.369	-0.370	-0.369	0.000	-0.369	-6.824	189	0.000
Coeff Prop_UpperInc	-0.888	-0.888	-0.888	0.000	-0.888	-113.262	189	0.000
Coeff ENT_9Cat	-27.864	-27.892	-27.838	0.008	-27.877	21.484	189	0.000
Coeff Prop_GI9Cat	-0.032	-0.033	-0.031	0.000	-0.032	11.391	189	0.000
Coeff Inters_grt3leg	1.374	1.374	1.374	0.000	1.374	35.194	189	0.000
Coeff StrDens	0.590	0.589	0.592	0.000	0.591	-18.974	189	0.000
Coeff Prop_Freeways	0.406	0.405	0.406	0.000	0.406	-11.429	189	0.000
Coeff Prop_Expressways	0.396	0.396	0.397	0.000	0.397	-63.822	189	0.000
Coeff Prop_PrimaryArter	-0.194	-0.196	-0.194	0.000	-0.194	4.970	189	0.000
Coeff Prop_SecondArter	0.406	0.406	0.406	0.000	0.406	-13.461	189	0.000
Coeff Round_Circ	-3.492	-3.495	-3.491	0.001	-3.493	8.346	189	0.000
Coeff Prop_Signal	5.037	5.034	5.040	0.001	5.038	-5.445	189	0.000
Local R^2	0.803	0.803	0.803	0.000	0.803	-0.286	189	0.776
Residual	-0.004	-190.434	389.895	61.531	0.000	-4.303	189	0.000
Std. Residual	0.001	-3.252	6.957	1.035	0.000	0.134	189	0.893

APPENDIX G2: Summary statistics for local estimates from GWR Model 2B

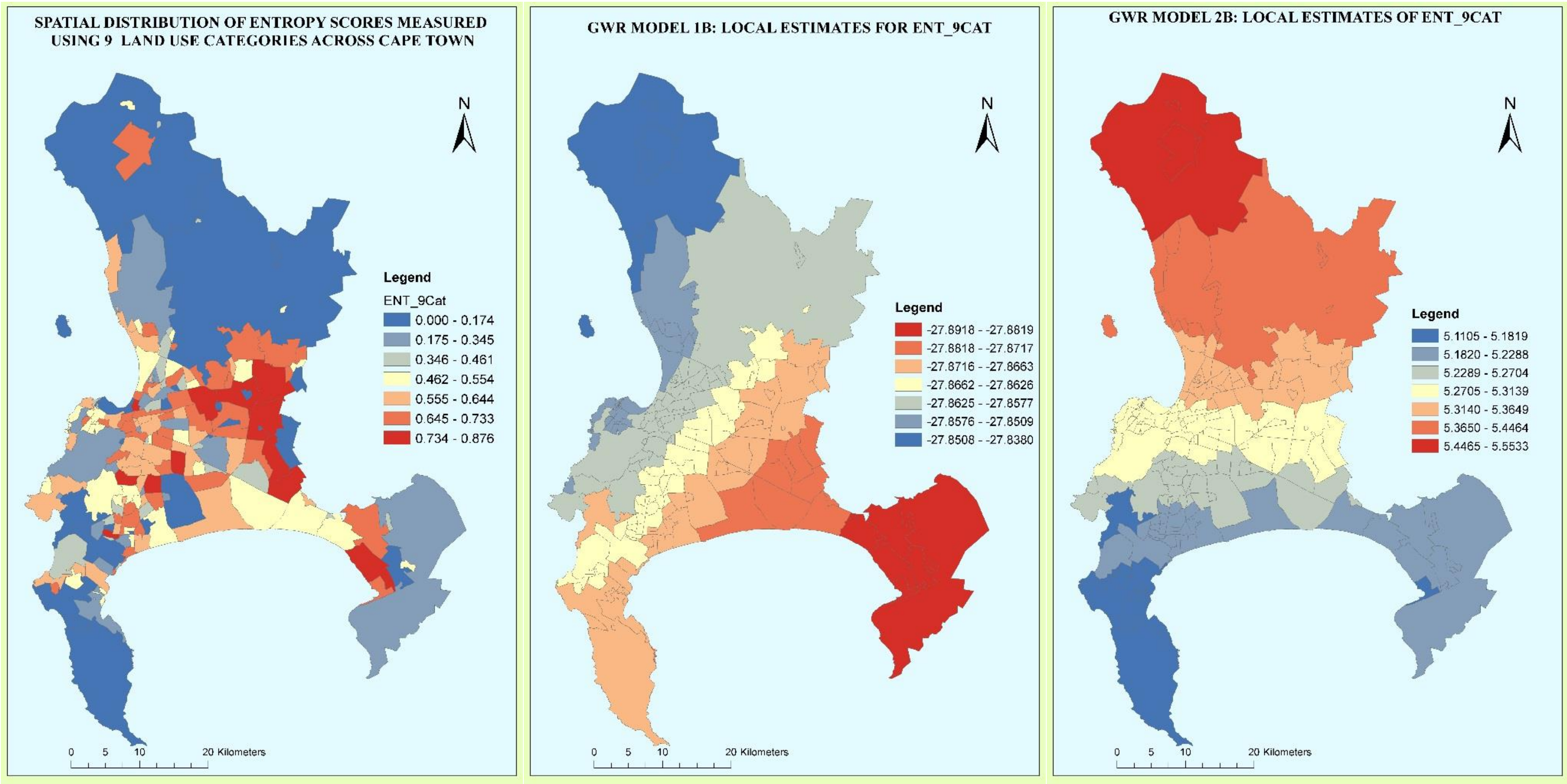
Local parameters	Summary statistics for local estimates from GWR Model 2B							
	Mean	Minimum	Maximum	Std.Dev.	Global Model	t-value	df	p
Coeff Intercept	-48.186	-48.250	-48.140	0.018	-48.054	-100.585	189	0.000
Coeff Log_Popu	23.772	23.641	23.933	0.046	23.736	10.664	189	0.000
Coeff Prop_AgeLess15	-0.337	-0.346	-0.332	0.003	-0.338	2.628	189	0.009
Coeff Prop_Age15_24	-0.376	-0.381	-0.373	0.001	-0.376	-2.433	189	0.016
Coeff Prop_AvgEd	-0.194	-0.197	-0.190	0.001	-0.194	-5.670	189	0.000
Coeff Prop_NotWork	-0.211	-0.212	-0.210	0.000	-0.210	-52.152	189	0.000
Coeff Prop_UpperInc	-0.605	-0.607	-0.604	0.001	-0.604	-21.137	189	0.000
Coeff ENT_9Cat	5.281	5.110	5.553	0.076	5.269	2.338	189	0.020
Coeff Prop_GI9Cat	0.120	0.117	0.124	0.001	0.119	13.898	189	0.000
Coeff Ratio_inters-cds	0.625	0.614	0.635	0.005	0.618	18.212	189	0.000
Coeff Prop_Freeways	0.179	0.177	0.181	0.001	0.177	15.897	189	0.000
Coeff Prop_Expressways	0.351	0.347	0.354	0.001	0.349	26.267	189	0.000
Coeff Prop_PrimaryArter	0.316	0.313	0.318	0.001	0.312	58.669	189	0.000
Coeff Prop_SecondArter	0.102	0.100	0.104	0.001	0.101	18.442	189	0.000
Coeff Round_Circ	0.999	0.990	1.005	0.003	0.999	0.751	189	0.454
Coeff Prop_Signal	2.546	2.523	2.564	0.006	2.548	-6.060	189	0.000
Local R^2	0.425	0.424	0.426	0.000	0.425	5.264	189	0.000
Residual	-0.007	-40.681	138.223	26.237	0.000	-1.576	189	0.117
Std. Residual	0.001	-1.655	5.192	1.014	0.000	0.175	189	0.861

APPENDIX G3: Summary statistics for local estimates from GWR Model 3B

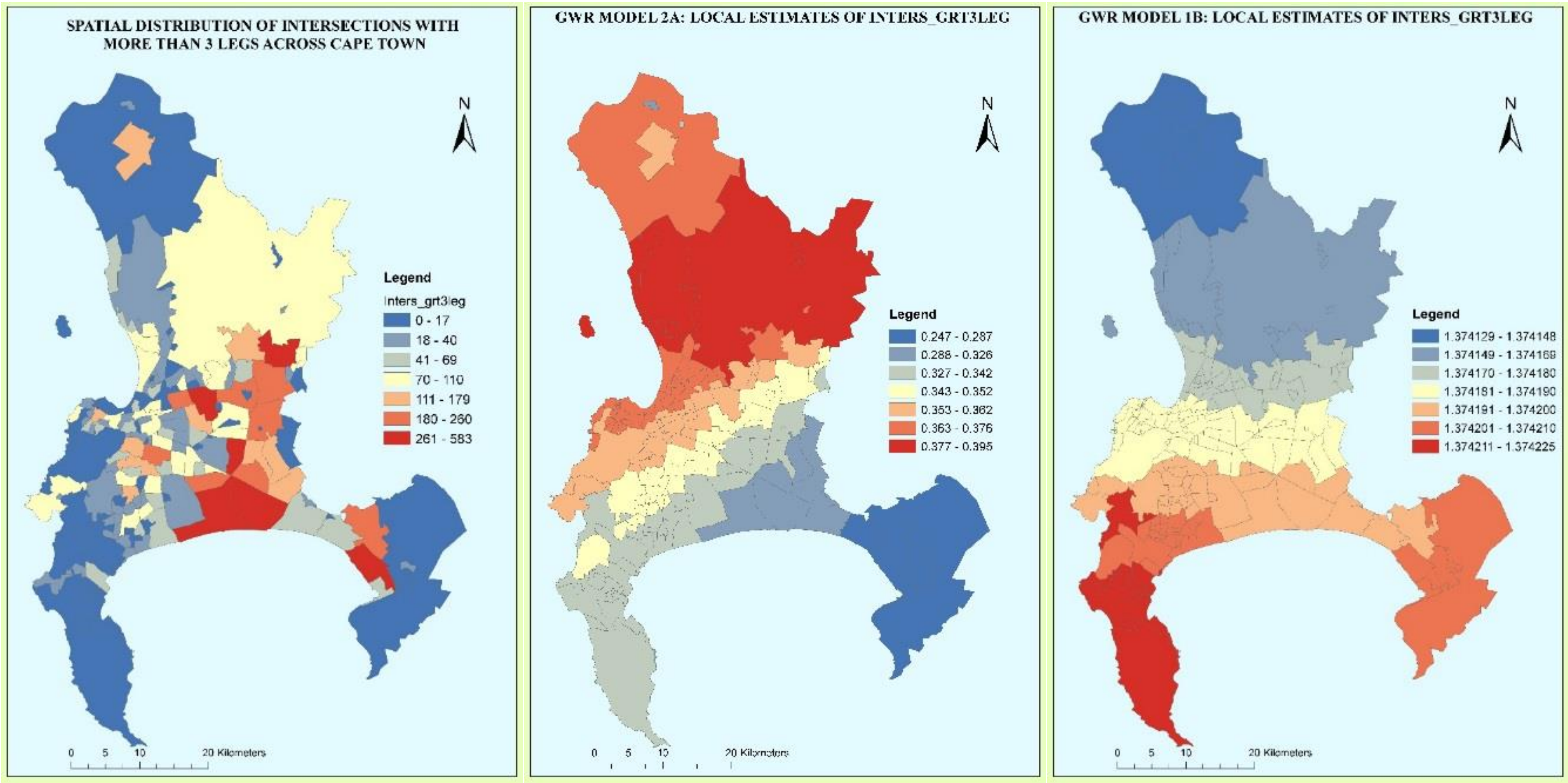
Local parameters	Summary statistics for local estimates from GWR Model 3B							
	Mean	Minimum	Maximum	Std.Dev.	Global Model	t-value	df	p
Coeff Intercept	1.081	-14.763	64.421	11.429	3.283	-2.655	189	0.009
Coeff Log_Popu	8.955	-0.317	14.686	3.117	7.823	5.003	189	0.000
Coeff Prop_Coloured	-0.492	-1.273	-0.084	0.211	-0.462	-1.911	189	0.058
Coeff Prop_White	-0.624	-1.314	-0.152	0.226	-0.571	-3.249	189	0.001
Coeff Inters_grt3leg	0.343	0.201	0.426	0.047	0.349	-1.815	189	0.071
Coeff Prop_Freeways	-0.035	-0.192	0.930	0.177	-0.002	-2.604	189	0.010
Coeff Prop_Expresways	0.209	-0.377	0.622	0.096	0.250	-5.814	189	0.000
Coeff Prop_PrimaryArter	0.227	-0.045	0.641	0.100	0.120	14.590	189	0.000
Coeff Prop_SecondArter	0.336	-0.256	0.739	0.181	0.178	12.127	189	0.000
Local R^2	0.658	0.564	0.863	0.044	0.649	2.897	189	0.004
Residual	-1.080	-77.690	177.204	24.565	0.000	-1.636	189	0.103
Std. Residual	-0.055	-5.086	8.396	1.112	0.000	-2.122	189	0.035

APPENDIX H

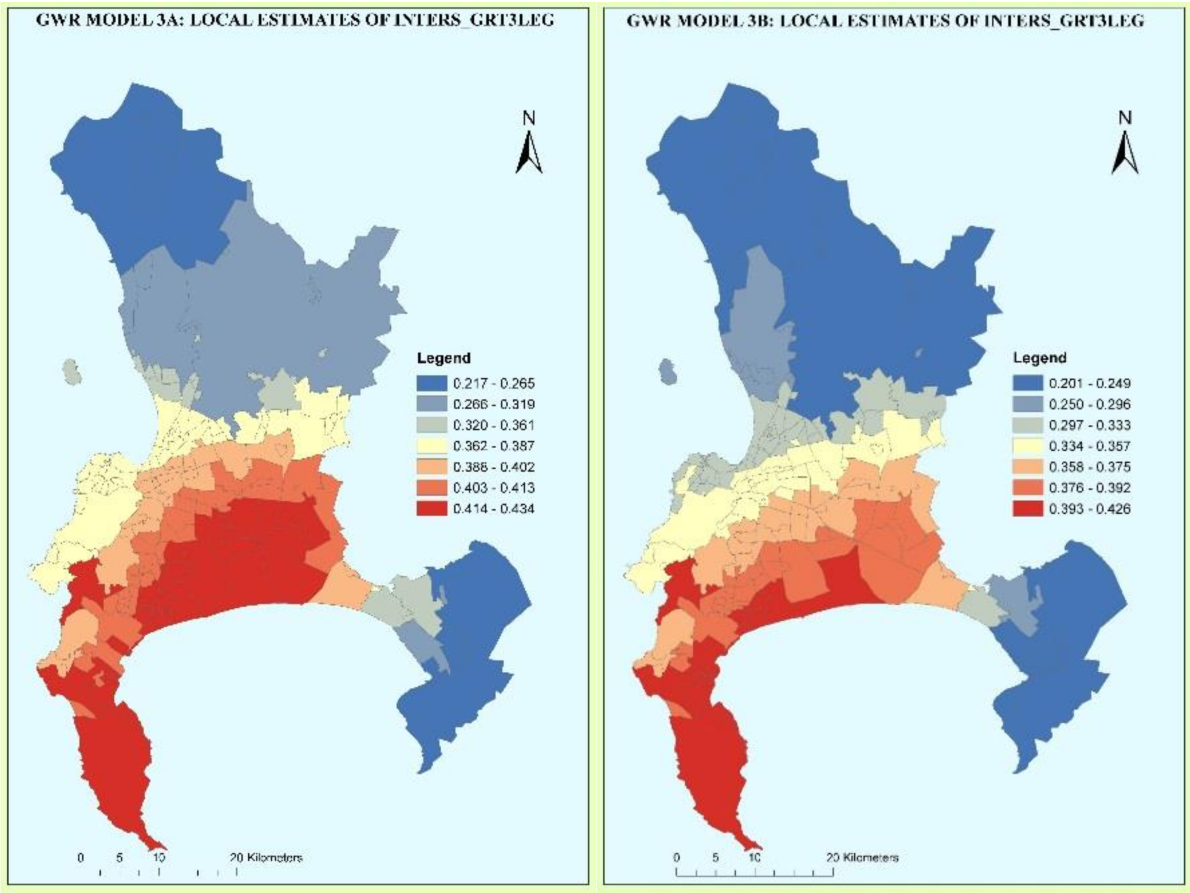
Comparison of local estimates of ENT_9Cat produced by GWR Models



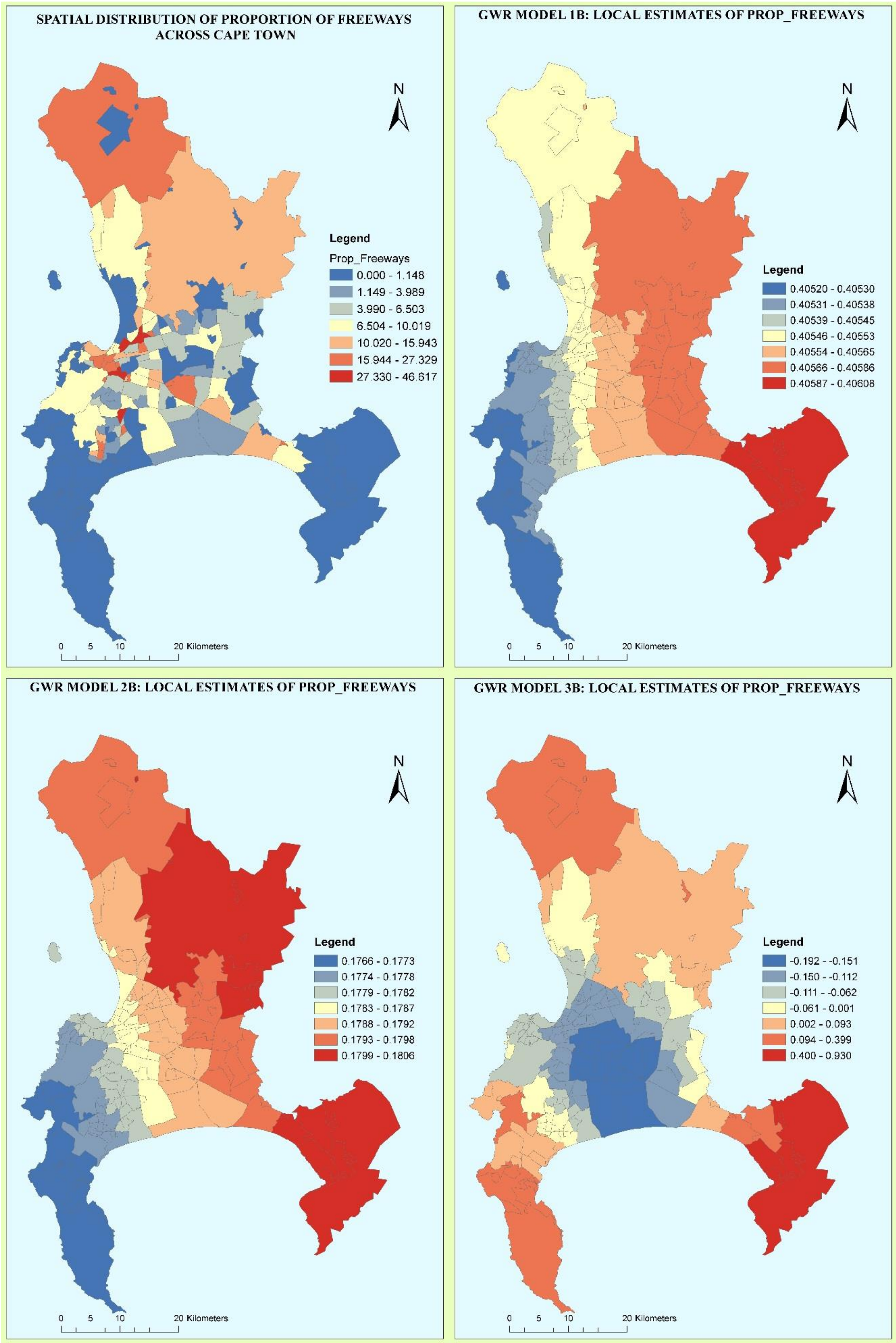
Comparison of local estimates of Inters_gr3Leg: GWR Model 2A and GWR Model 1B



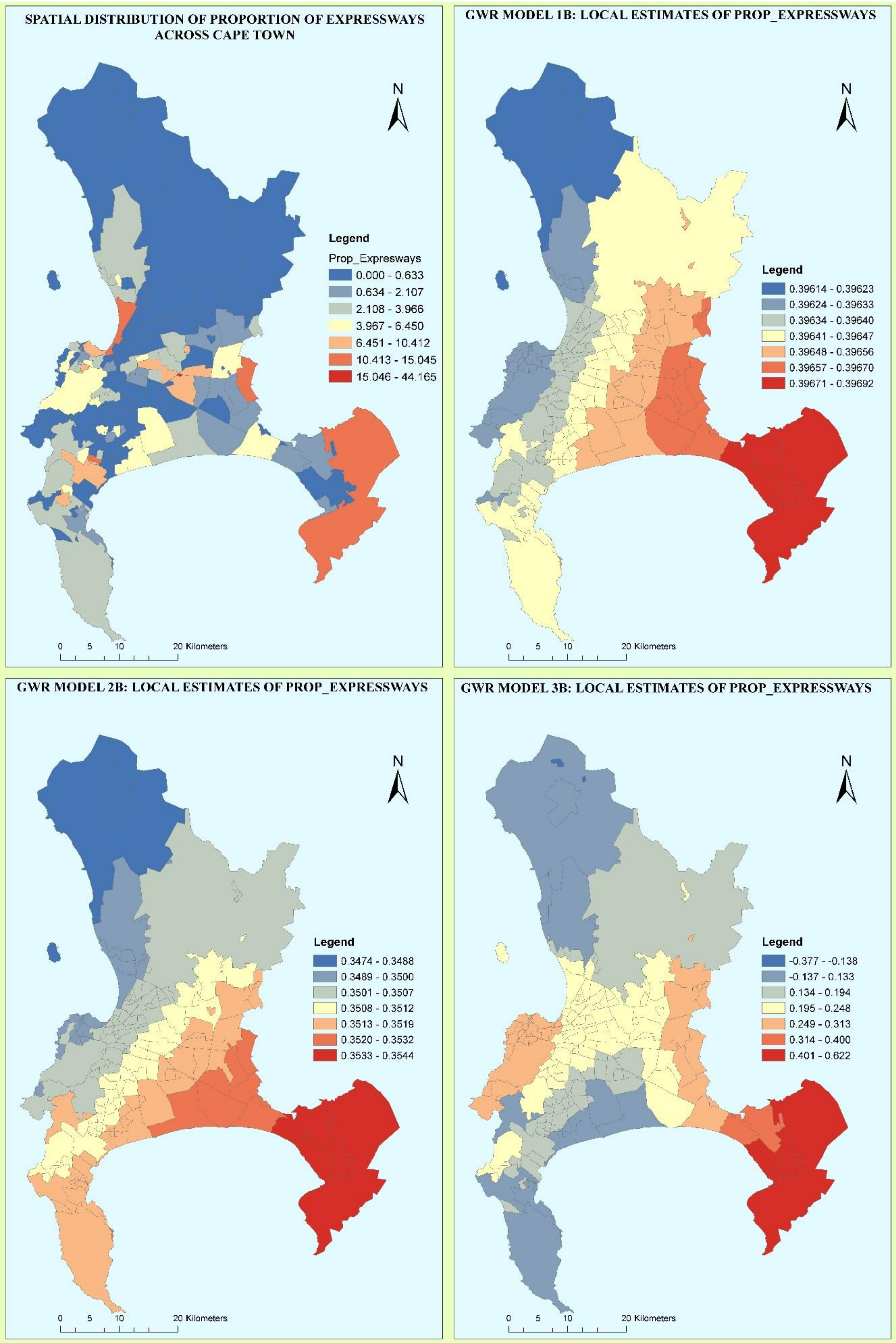
Comparison of local estimates of Inters_gr3Leg: GWR Models 3A&B



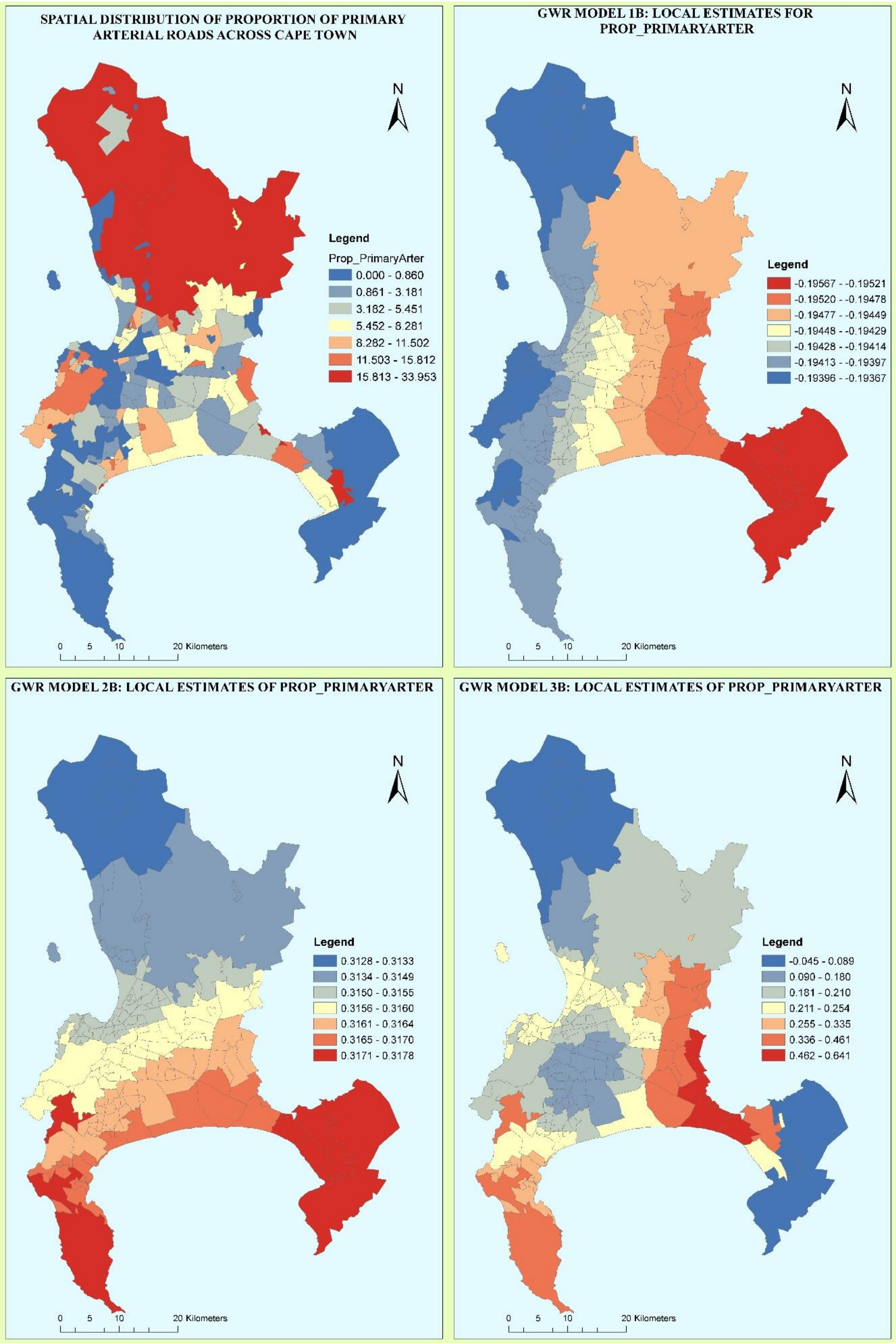
Comparison of local estimates of Prop_Freeways for GWR Models



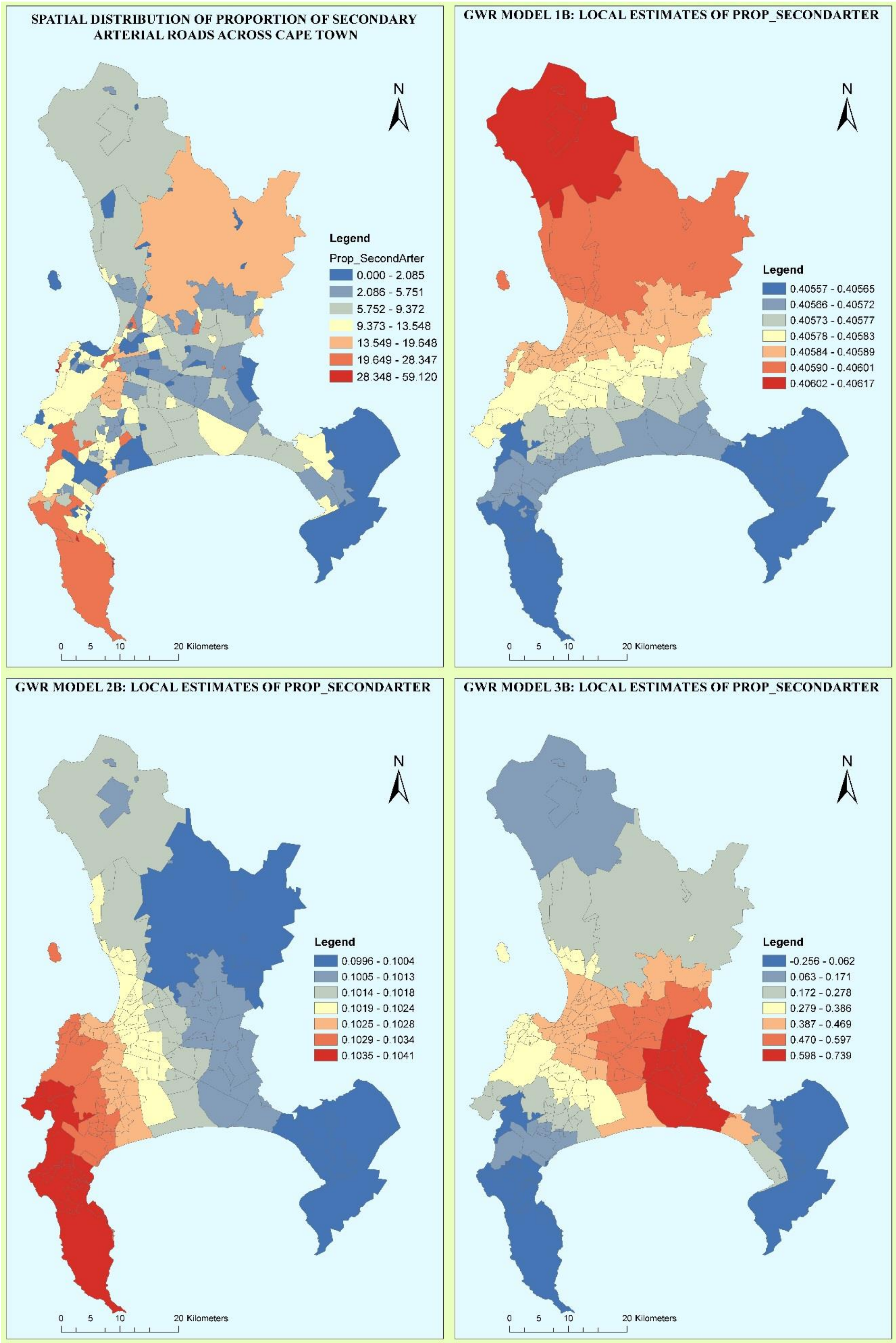
Comparison of local estimates of Prop_Expressways for GWR Models



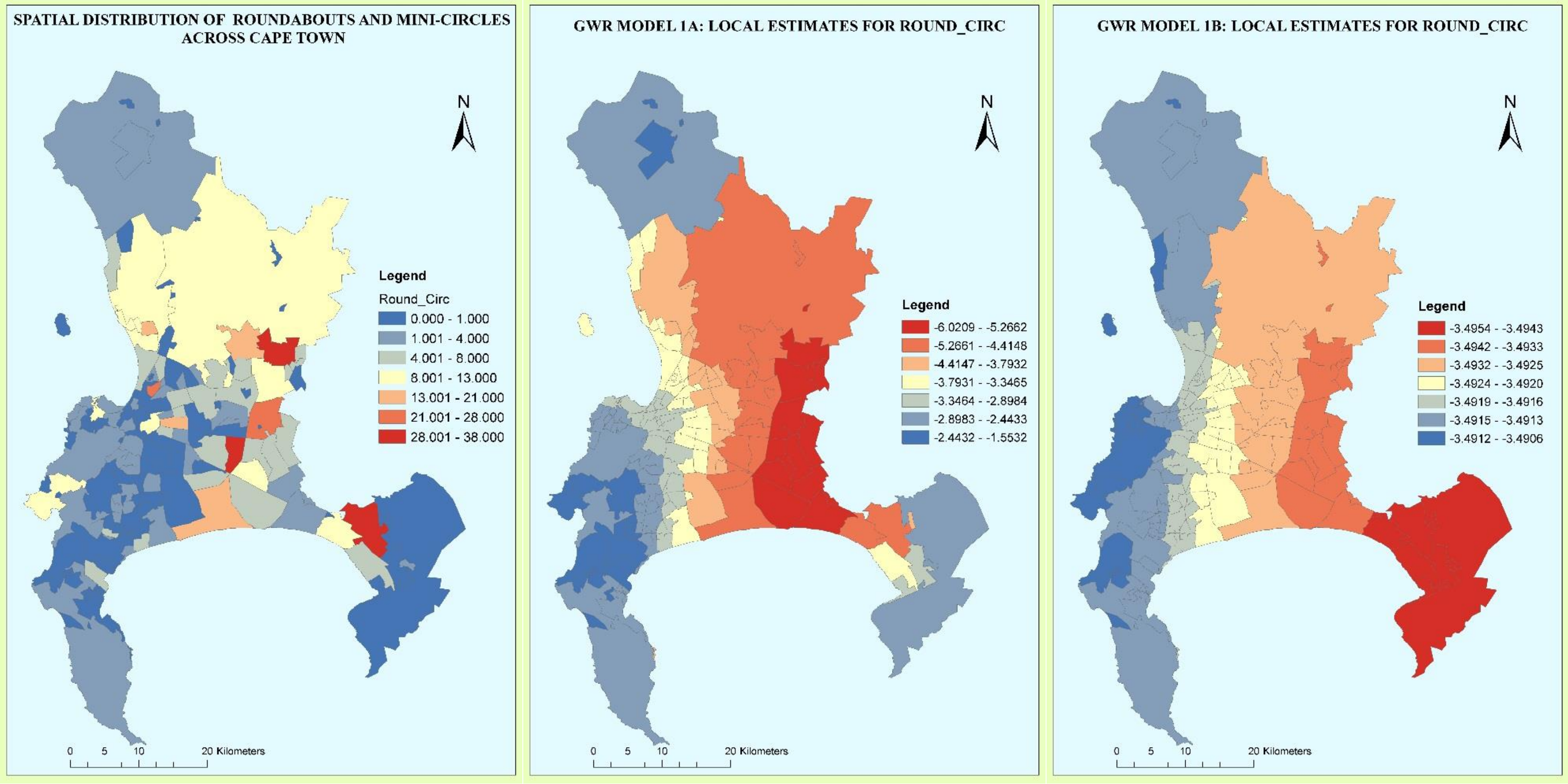
Local estimates of Prop_PrimaryArter for GWR Models



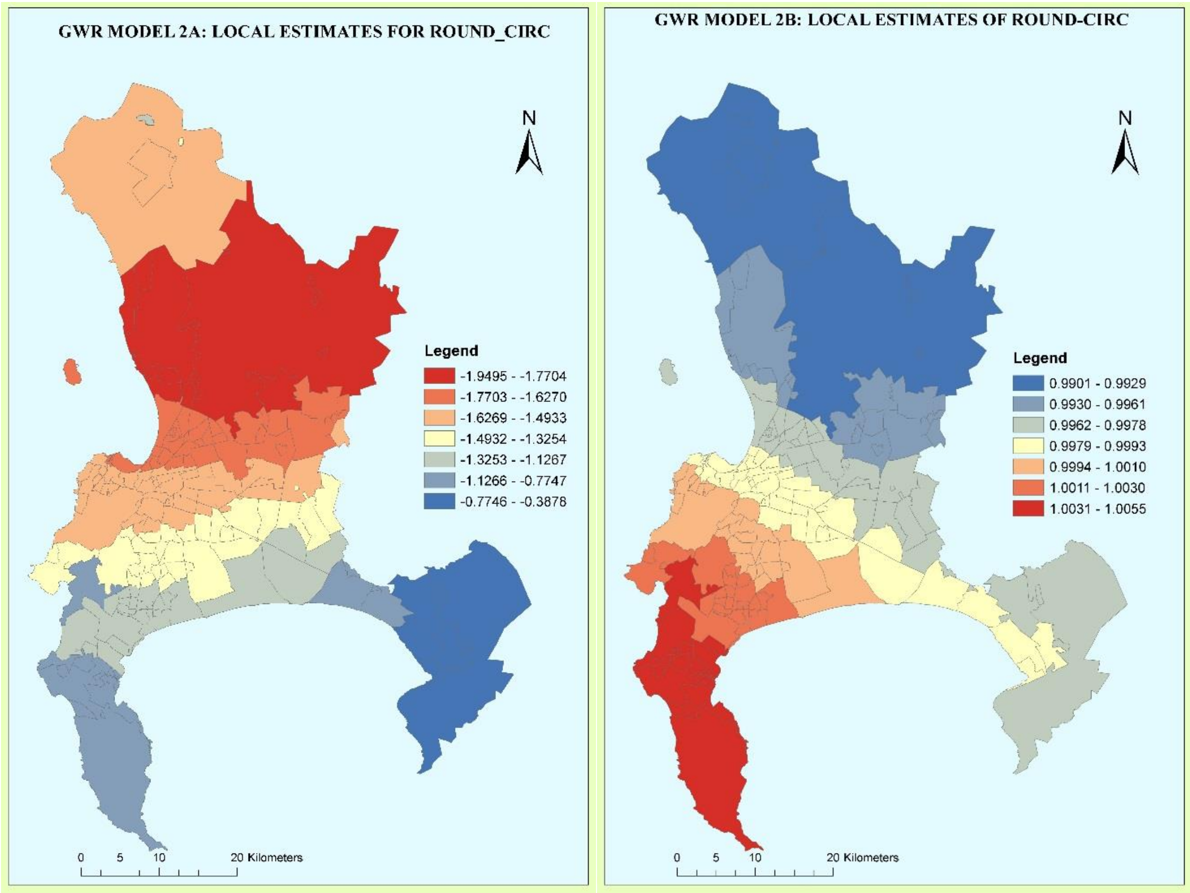
Local estimates of Prop_SecondArter for GWR Models



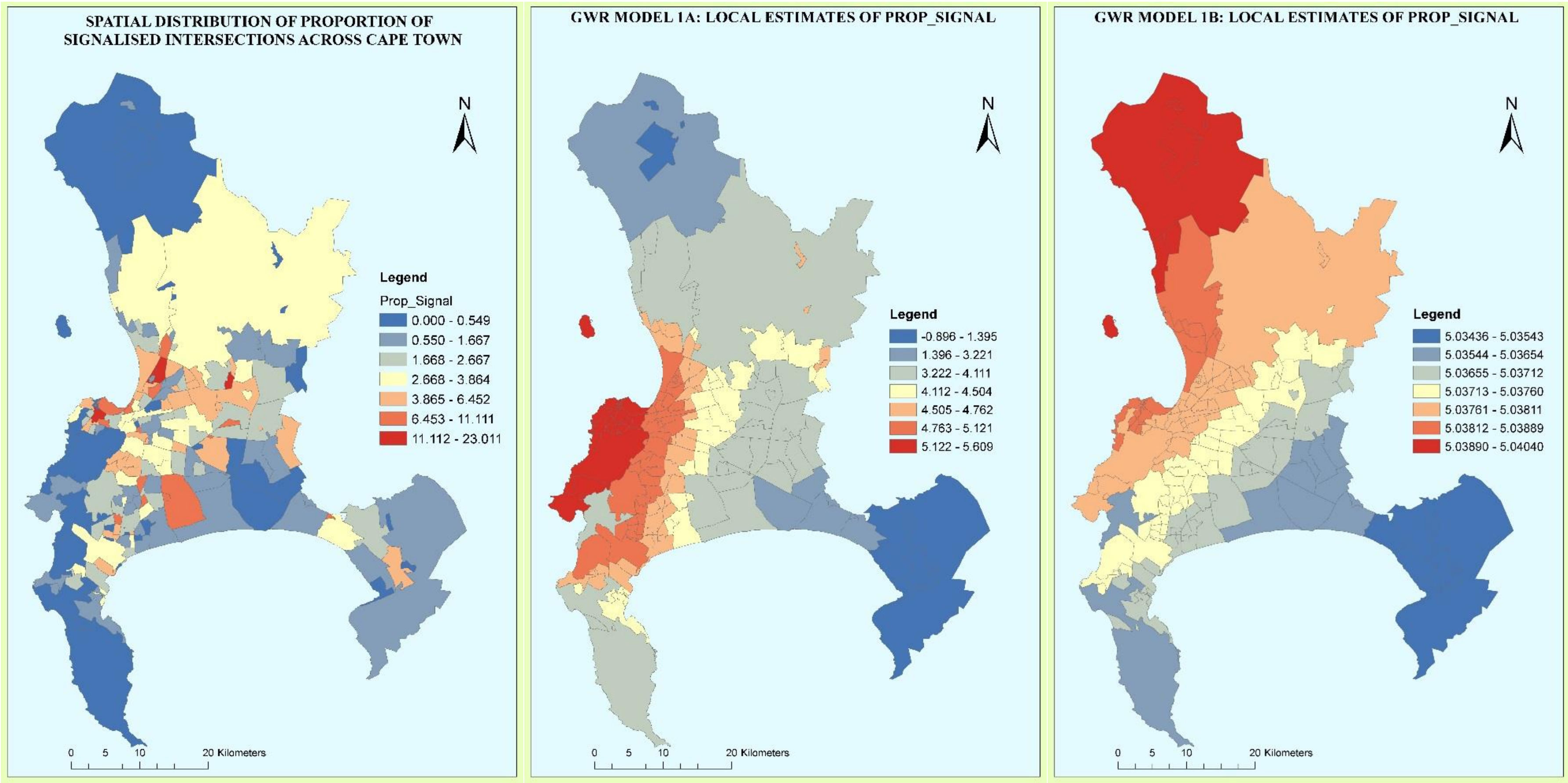
Local estimates of Round_Circ for GWR Models 1A and GWR Model 1B



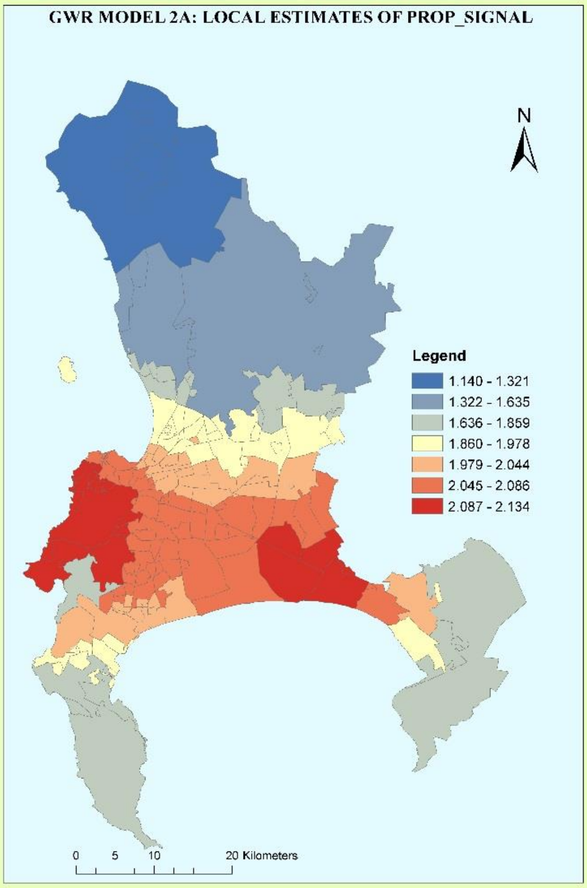
Local estimates of Round_Circ for GWR Models 2A and GWR Model 2B

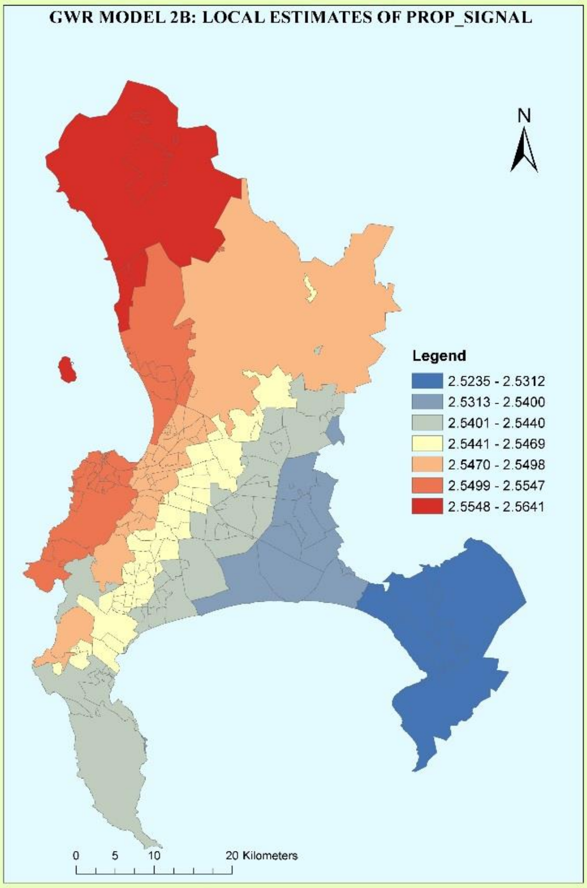


Local estimates of Prop_Signal for GWR Models 1A and GWR Model 1B

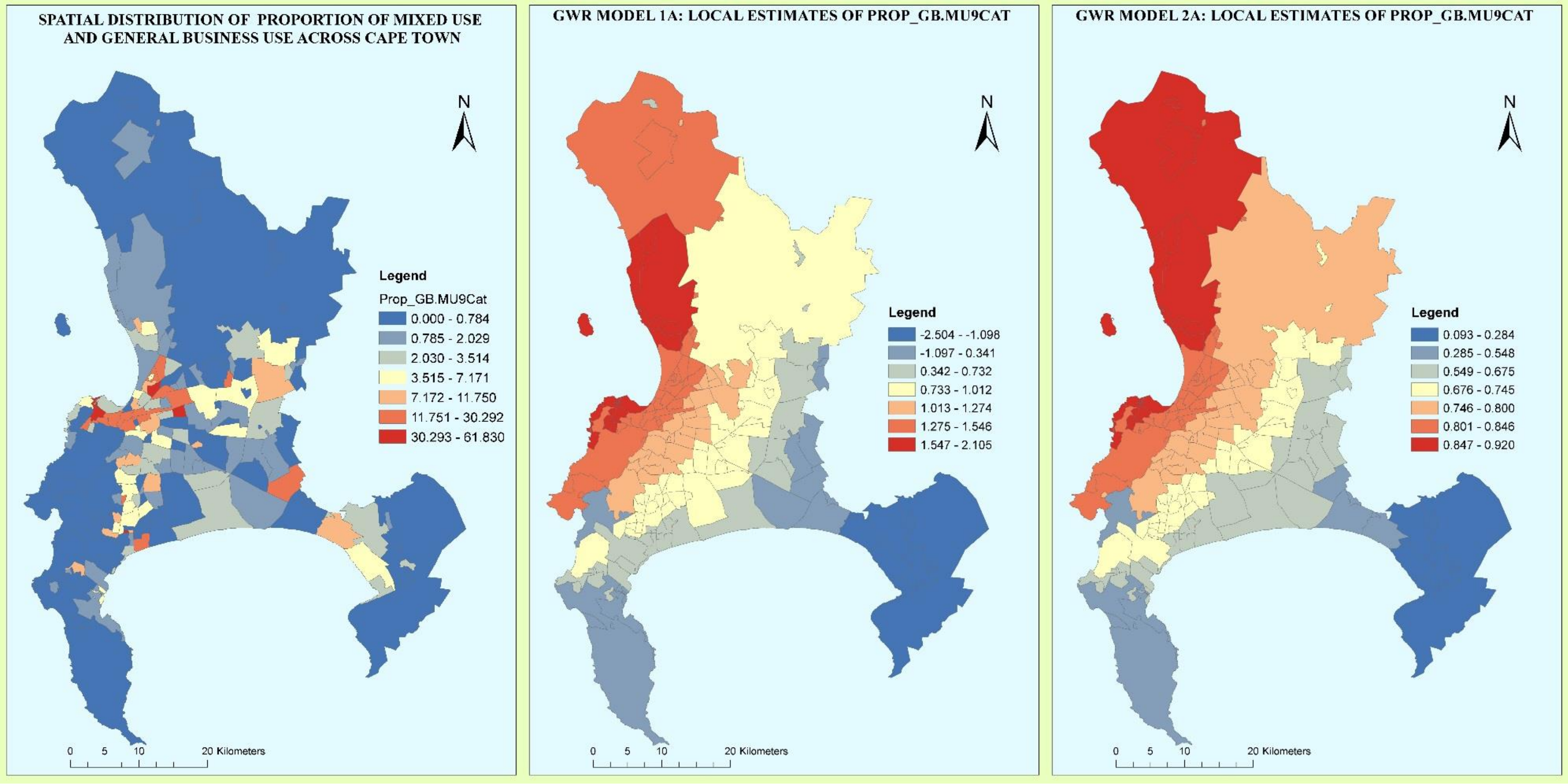


Local estimates of Prop_Signal for GWR Models 2A and GWR Model 2B

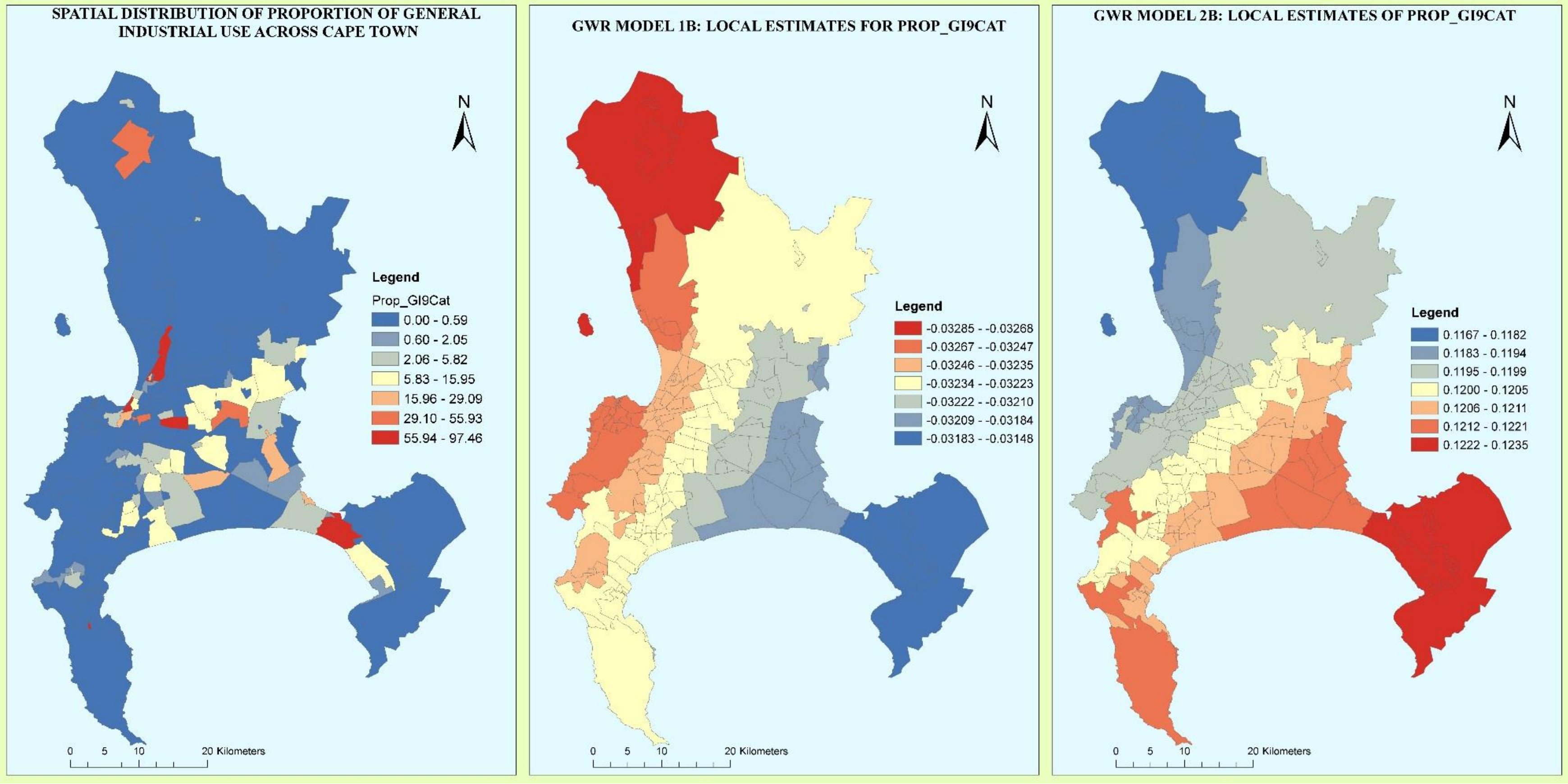




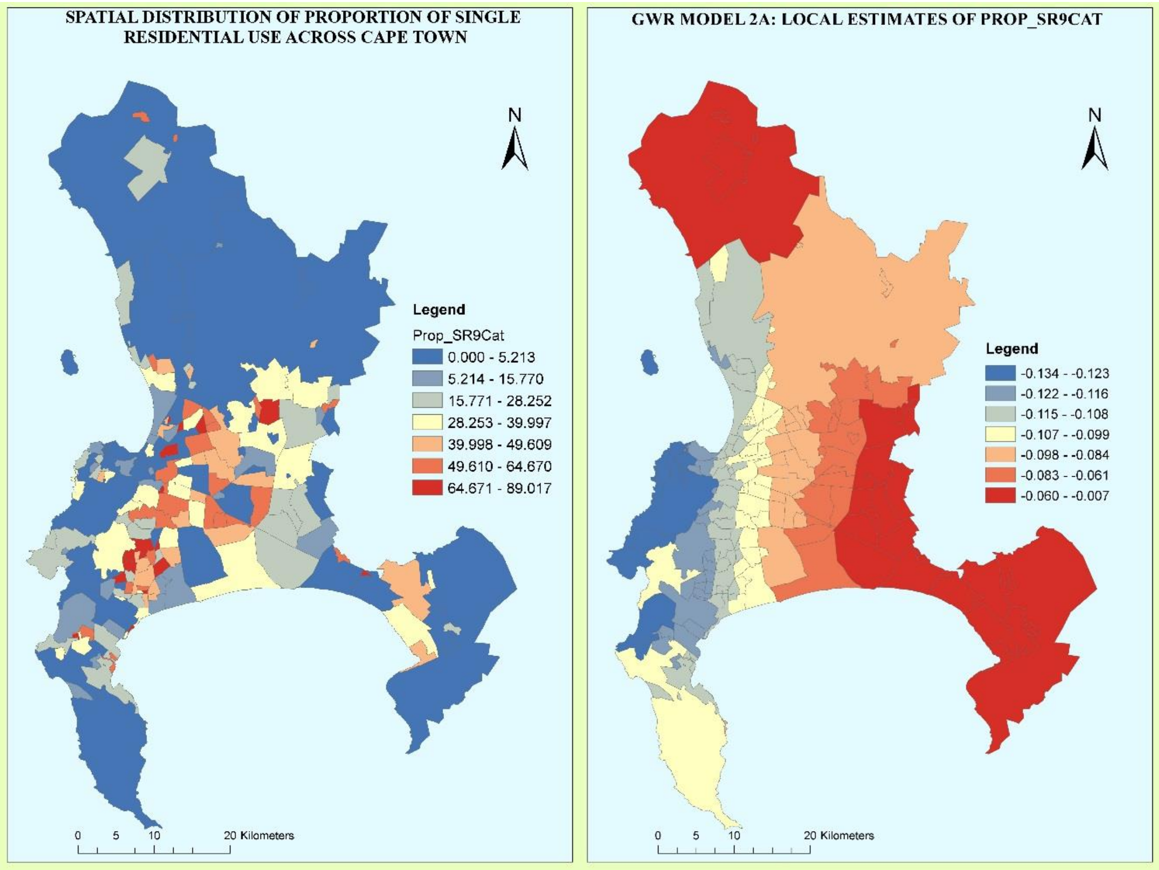
Local estimates of Prop_GB.MU9Cat for GWR Models 1A and GWR Model 2A



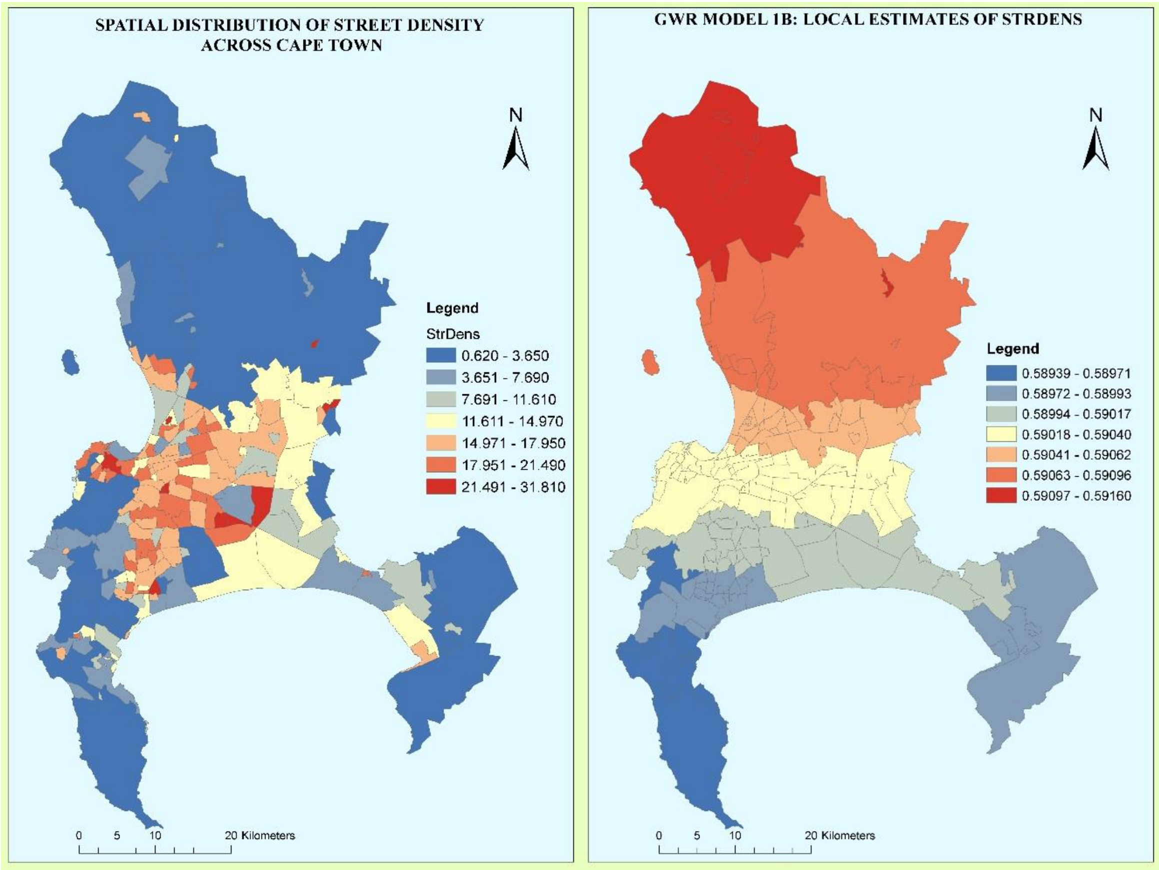
Local estimates of Prop_GI9Cat for GWR Models 1B and GWR Model 2B



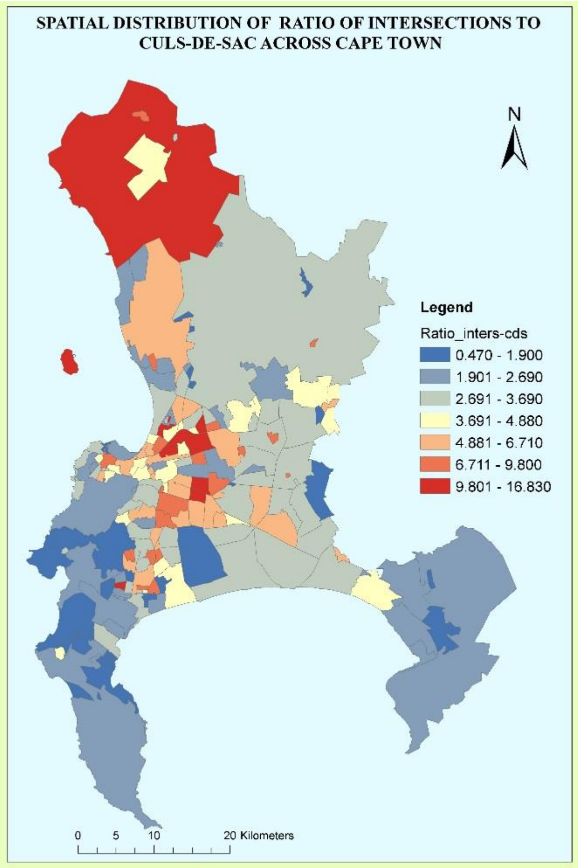
Local estimates of Prop_SR9Cat for GWR Models 2A

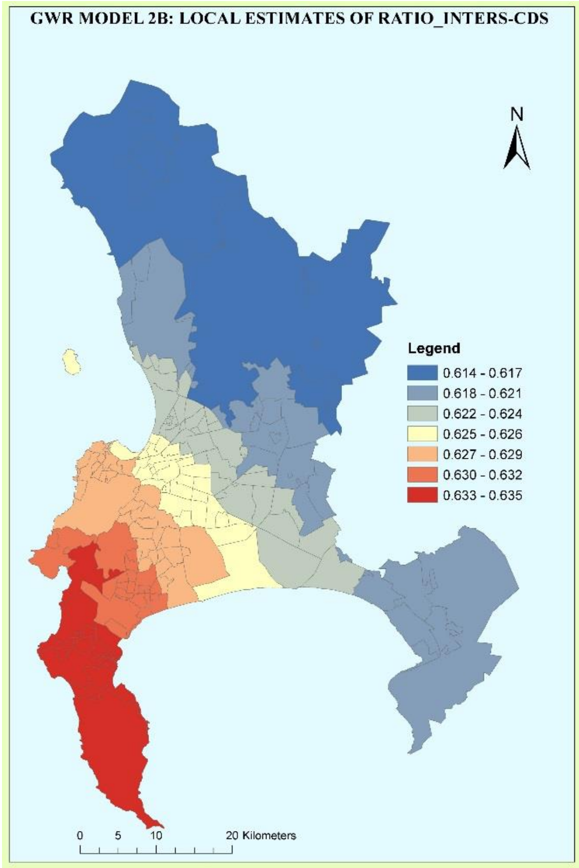


Local estimates of StrDens for GWR Model 1B



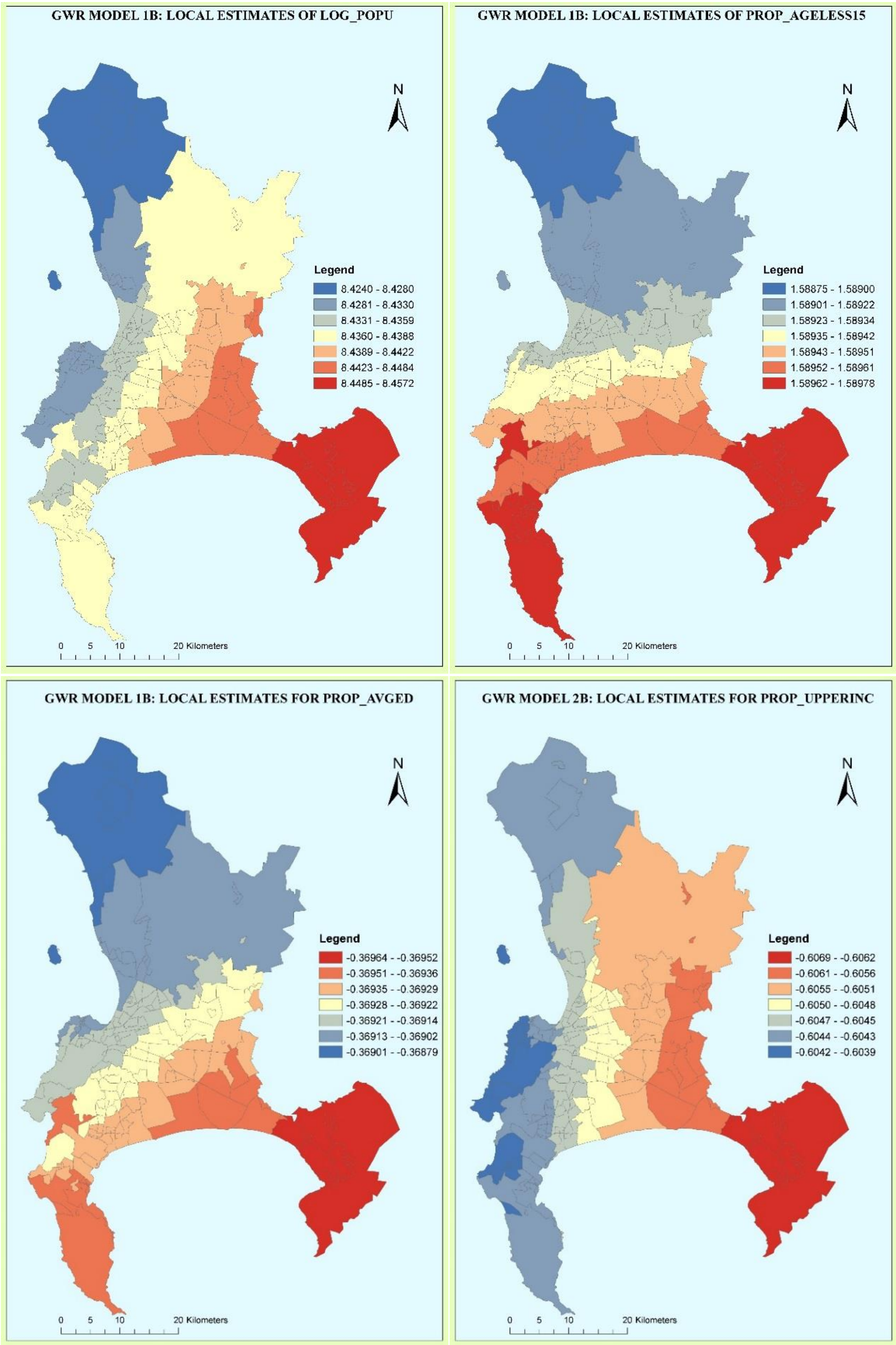
Local estimates of Ratio_Inters-cds for GWR Models 2

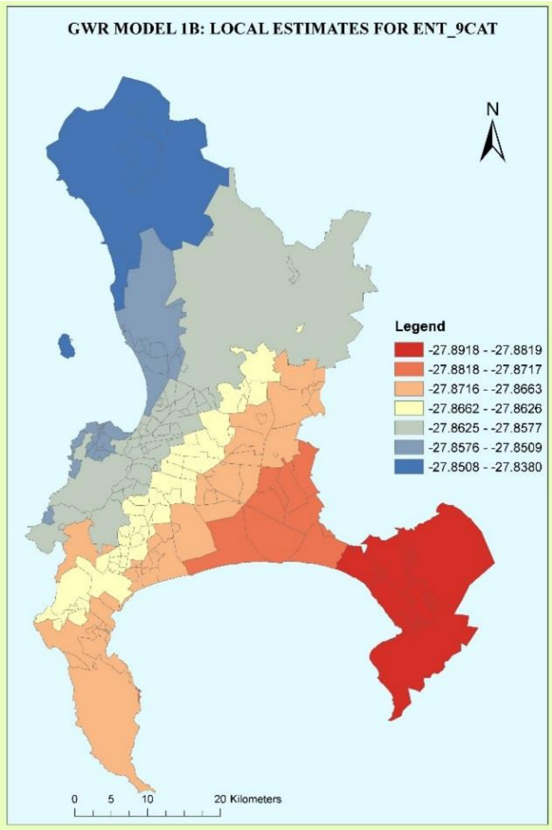


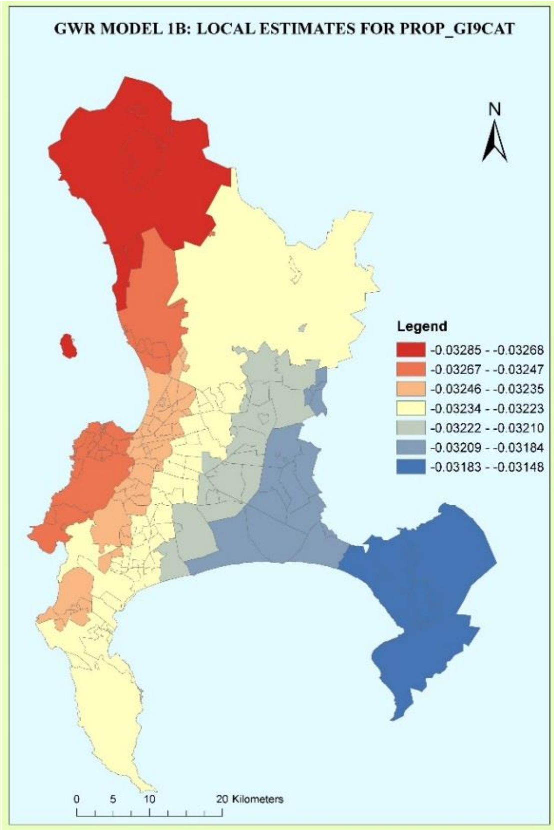


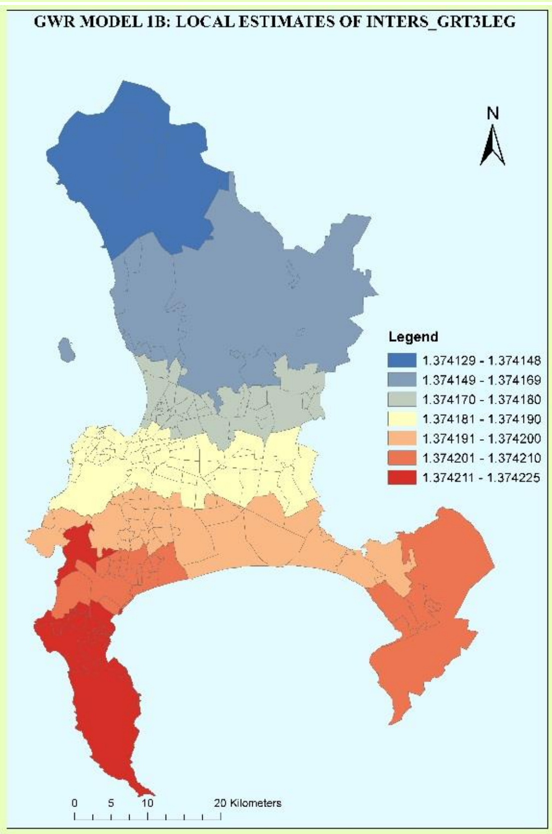
APPENDIX I

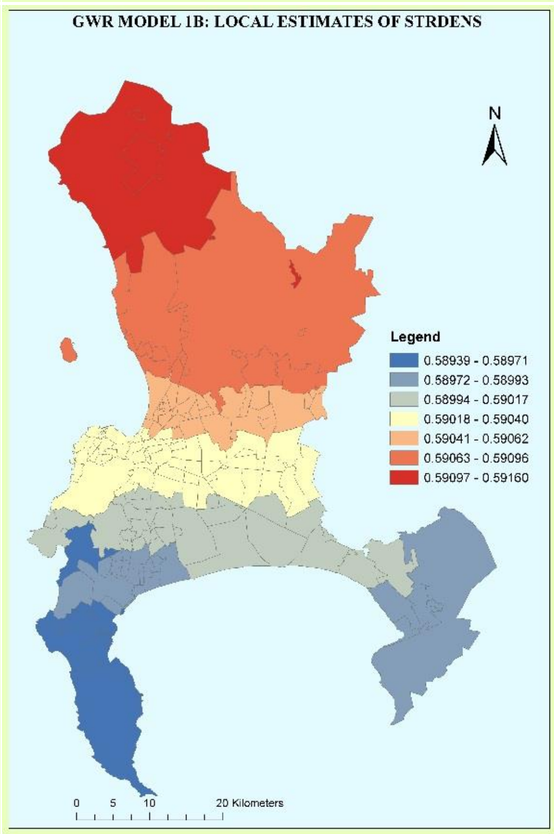
APPENDIX I1: Local estimates for predictors in GWR Model 1B

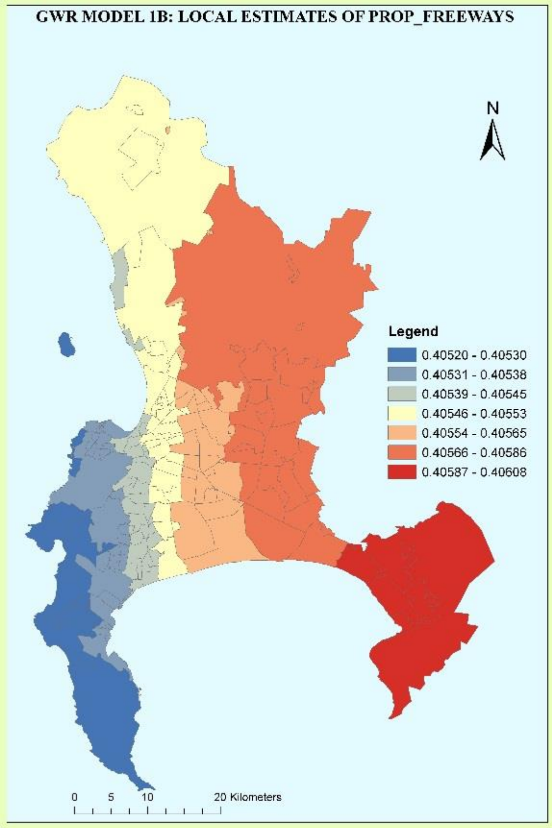


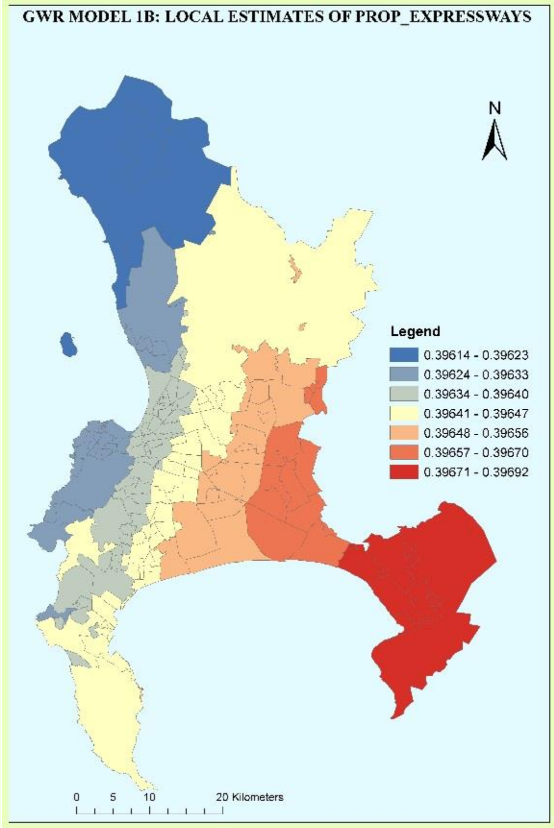

GWR MODEL 1B: LOCAL ESTIMATES FOR ENT_9CAT
N
Legend
-27.8918 - -27.8819
-27.8818 - -27.8717
-27.8716 - -27.8663
-27.8662 - -27.8626
-27.8625 - -27.8577
-27.8576 - -27.8509
-27.8508 - -27.8380
0 5 10 20 Kilometers

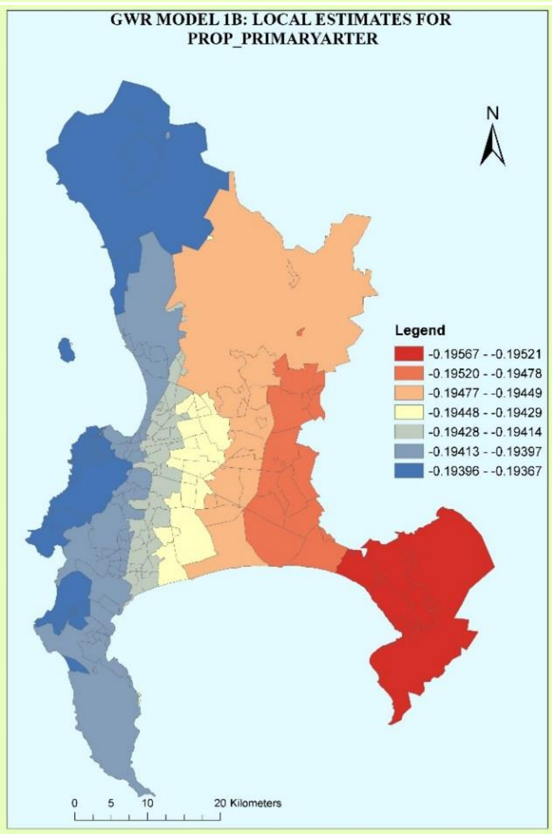

GWR MODEL 1B: LOCAL ESTIMATES FOR PROP_GI9CAT
N
Legend
-0.03285 - -0.03268
-0.03267 - -0.03247
-0.03246 - -0.03235
-0.03234 - -0.03223
-0.03222 - -0.03210
-0.03209 - -0.03184
-0.03183 - -0.03148
0 5 10 20 Kilometers

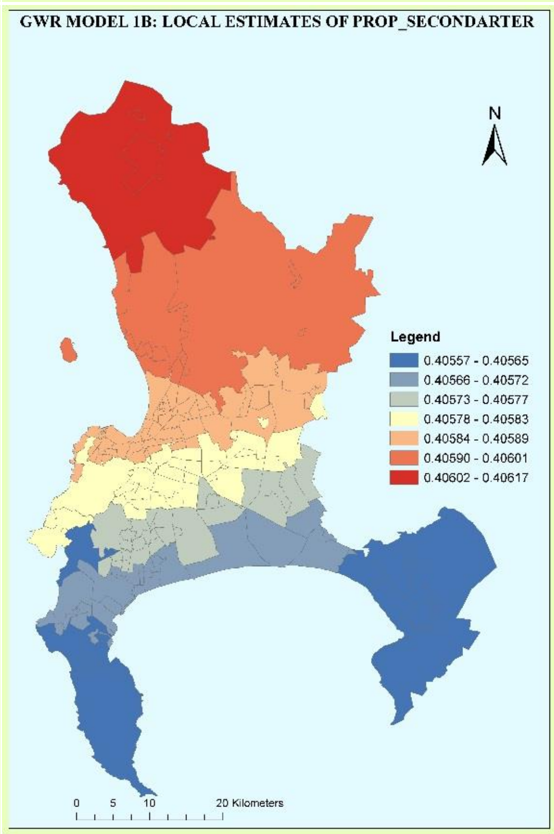

GWR MODEL 1B: LOCAL ESTIMATES OF INTERS_GRT3LEG
N
Legend
1.374129 - 1.374148
1.374149 - 1.374169
1.374170 - 1.374180
1.374181 - 1.374190
1.374191 - 1.374200
1.374201 - 1.374210
1.374211 - 1.374225
0 5 10 20 Kilometers

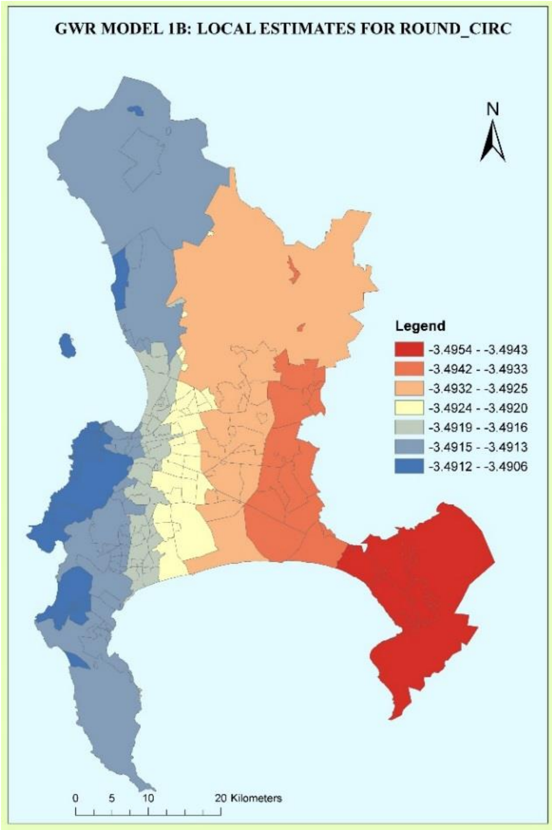

GWR MODEL 1B: LOCAL ESTIMATES OF STRDENS
N
Legend
0.58939 - 0.58971
0.58972 - 0.58993
0.58994 - 0.59017
0.59018 - 0.59040
0.59041 - 0.59062
0.59063 - 0.59096
0.59097 - 0.59160
0 5 10 20 Kilometers

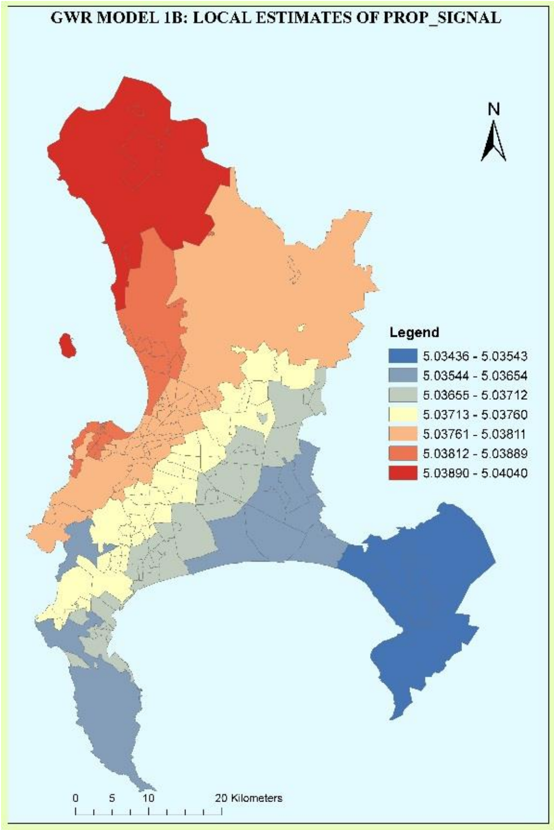

GWR MODEL 1B: LOCAL ESTIMATES OF PROP_FREEWAYS
N
Legend
0.40520 - 0.40530
0.40531 - 0.40538
0.40539 - 0.40545
0.40546 - 0.40553
0.40554 - 0.40565
0.40566 - 0.40586
0.40587 - 0.40608
0 5 10 20 Kilometers


GWR MODEL 1B: LOCAL ESTIMATES OF PROP_EXPRESSWAYS
N
Legend
0.39614 - 0.39623
0.39624 - 0.39633
0.39634 - 0.39640
0.39641 - 0.39647
0.39648 - 0.39656
0.39657 - 0.39670
0.39671 - 0.39692
0 5 10 20 Kilometers

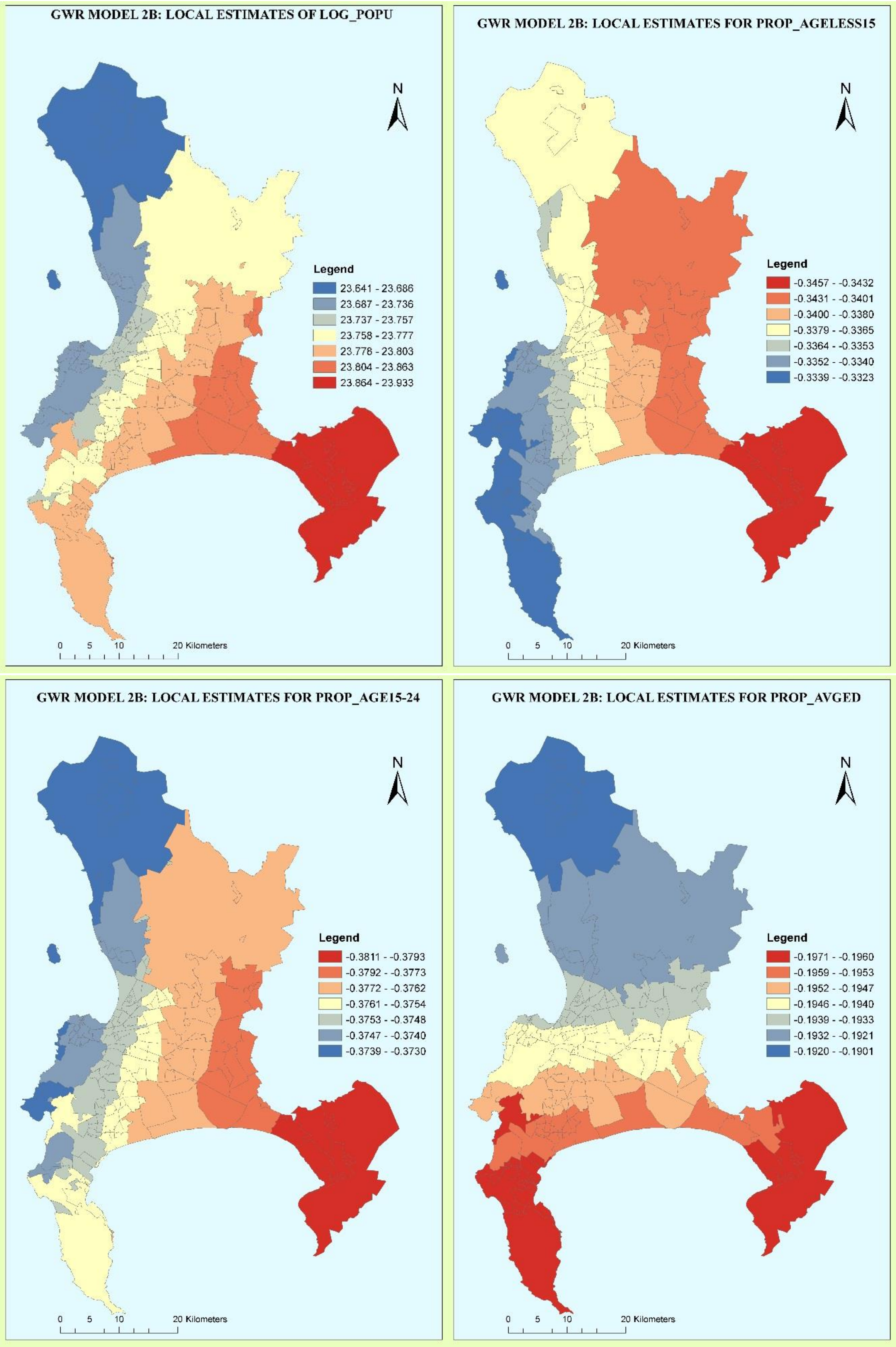

GWR MODEL 1B: LOCAL ESTIMATES FOR PROP_PRIMARYARTER
N
Legend
-0.19567 - -0.19521
-0.19520 - -0.19478
-0.19477 - -0.19449
-0.19448 - -0.19429
-0.19428 - -0.19414
-0.19413 - -0.19397
-0.19396 - -0.19367
0 5 10 20 Kilometers

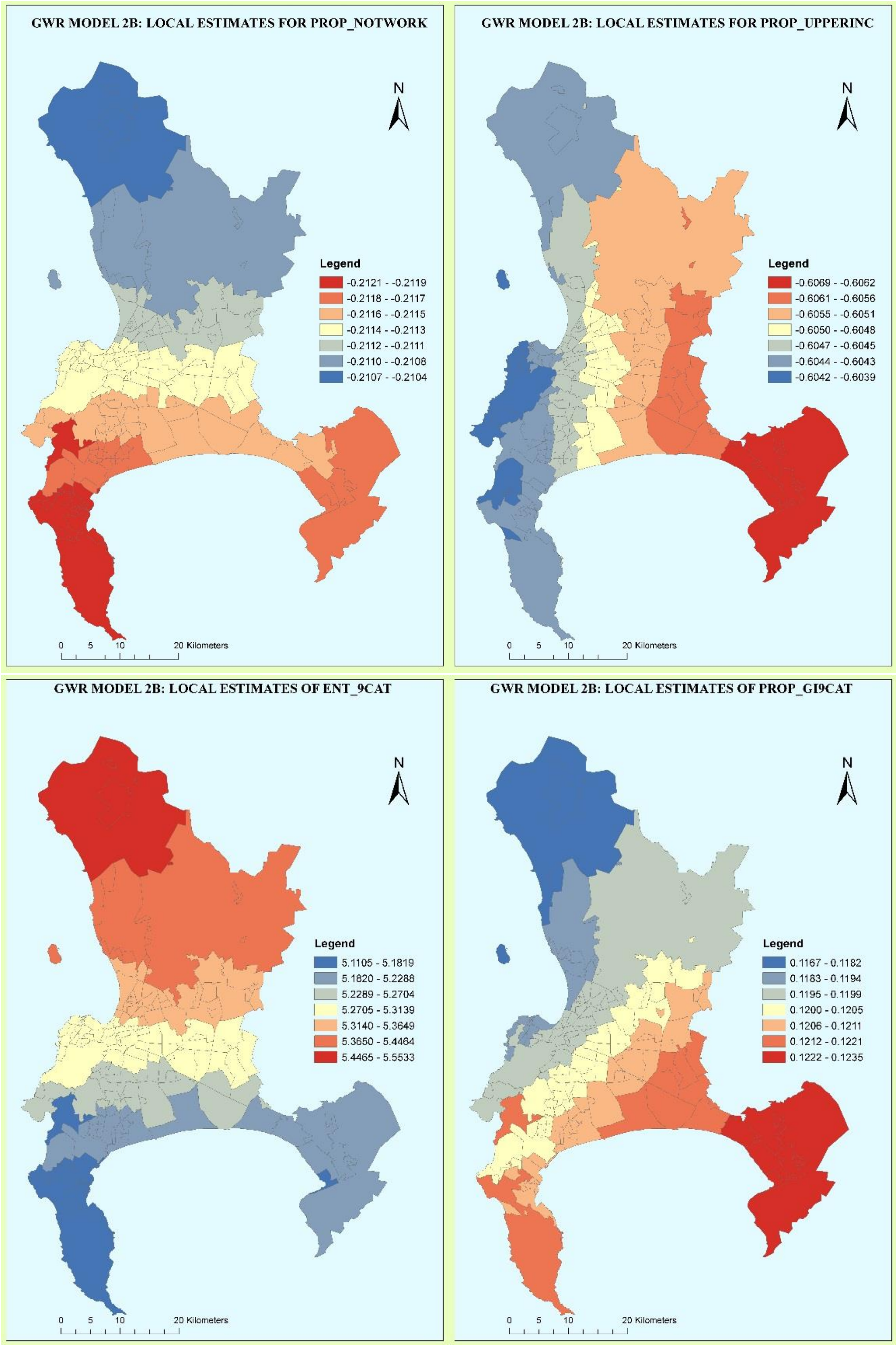

GWR MODEL 1B: LOCAL ESTIMATES OF PROP_SECONDARTER
N
Legend
0.40557 - 0.40565
0.40566 - 0.40572
0.40573 - 0.40577
0.40578 - 0.40583
0.40584 - 0.40589
0.40590 - 0.40601
0.40602 - 0.40617
0 5 10 20 Kilometers

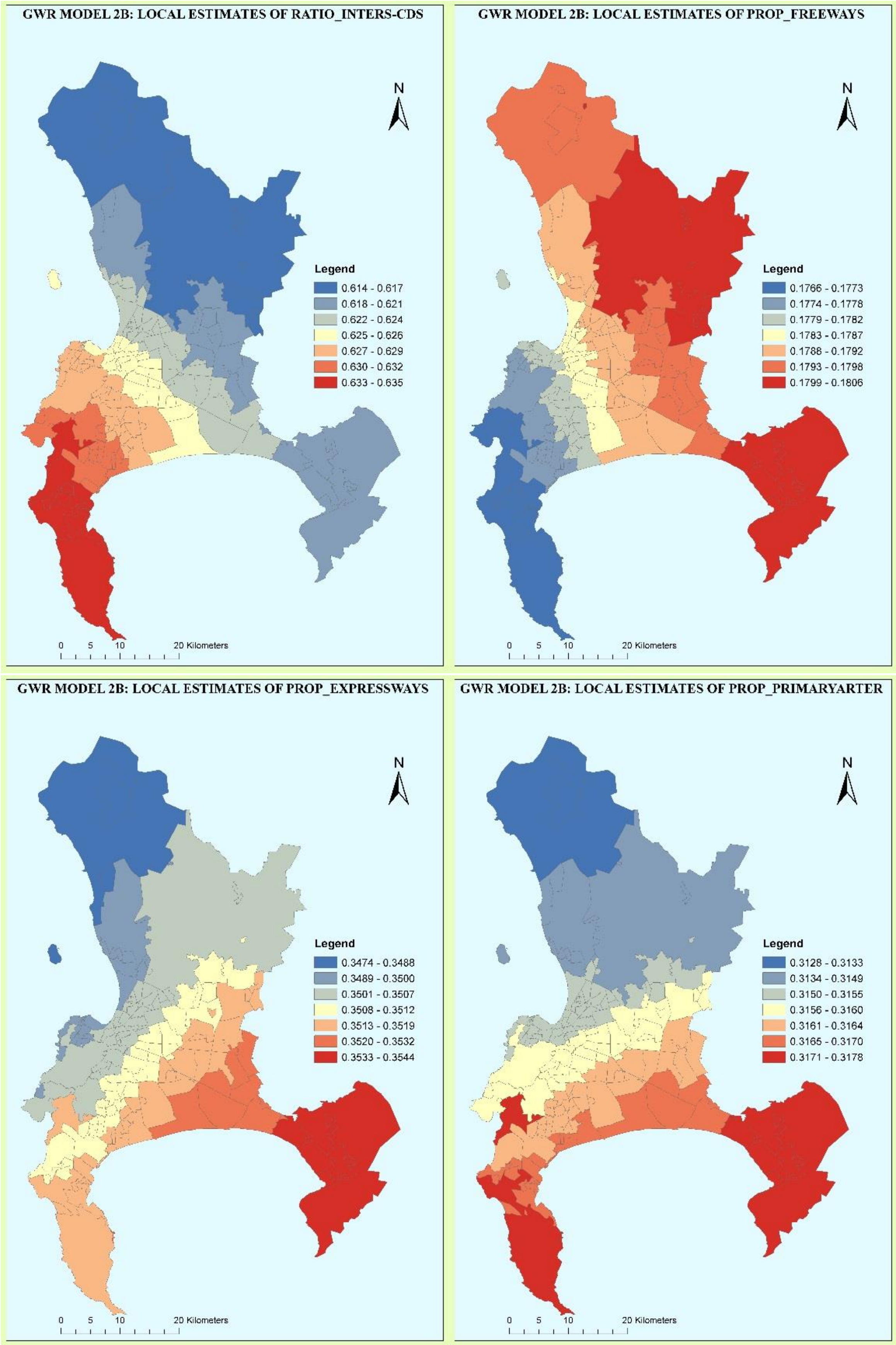

GWR MODEL 1B: LOCAL ESTIMATES FOR ROUND_CIRC
N
Legend
-3.4954 - -3.4943
-3.4942 - -3.4933
-3.4932 - -3.4925
-3.4924 - -3.4920
-3.4919 - -3.4916
-3.4915 - -3.4913
-3.4912 - -3.4906
0 5 10 20 Kilometers

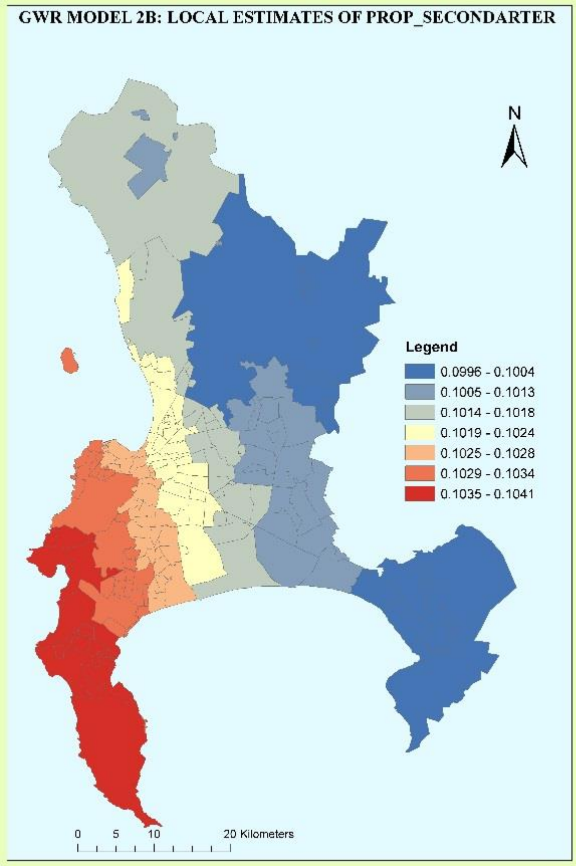

GWR MODEL 1B: LOCAL ESTIMATES OF PROP_SIGNAL
N
Legend
5.03436 - 5.03543
5.03544 - 5.03654
5.03655 - 5.03712
5.03713 - 5.03760
5.03761 - 5.03811
5.03812 - 5.03889
5.03890 - 5.04040
0 5 10 20 Kilometers

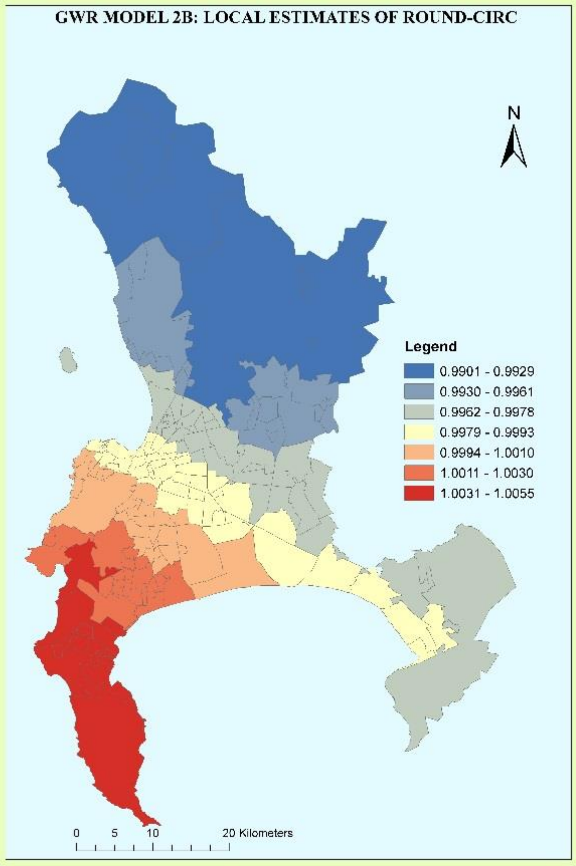
APPENDIX I2: Local estimates for predictors in GWR Model 2B

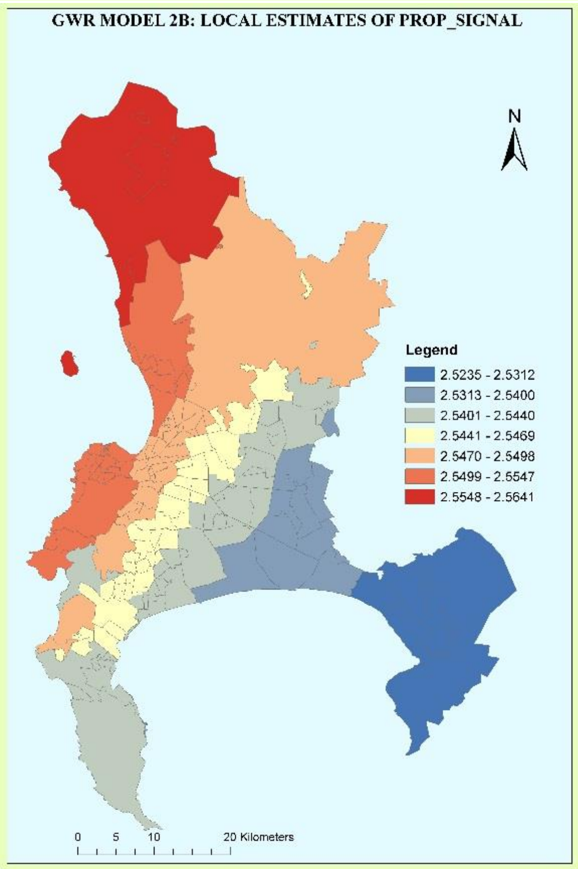



GWR MODEL 2B: LOCAL ESTIMATES FOR PROP_NOTWORK
N
Legend
-0.2121 - -0.2119
-0.2118 - -0.2117
-0.2116 - -0.2115
-0.2114 - -0.2113
-0.2112 - -0.2111
-0.2110 - -0.2108
-0.2107 - -0.2104
0 5 10 20 Kilometers
GWR MODEL 2B: LOCAL ESTIMATES FOR PROP_UPPERINC
N
Legend
-0.6069 - -0.6062
-0.6061 - -0.6056
-0.6055 - -0.6051
-0.6050 - -0.6048
-0.6047 - -0.6045
-0.6044 - -0.6043
-0.6042 - -0.6039
0 5 10 20 Kilometers
GWR MODEL 2B: LOCAL ESTIMATES OF ENT_9CAT
N
Legend
5.1105 - 5.1819
5.1820 - 5.2288
5.2289 - 5.2704
5.2705 - 5.3139
5.3140 - 5.3649
5.3650 - 5.4464
5.4465 - 5.5533
0 5 10 20 Kilometers
GWR MODEL 2B: LOCAL ESTIMATES OF PROP_GI9CAT
N
Legend
0.1167 - 0.1182
0.1183 - 0.1194
0.1195 - 0.1199
0.1200 - 0.1205
0.1206 - 0.1211
0.1212 - 0.1221
0.1222 - 0.1235
0 5 10 20 Kilometers

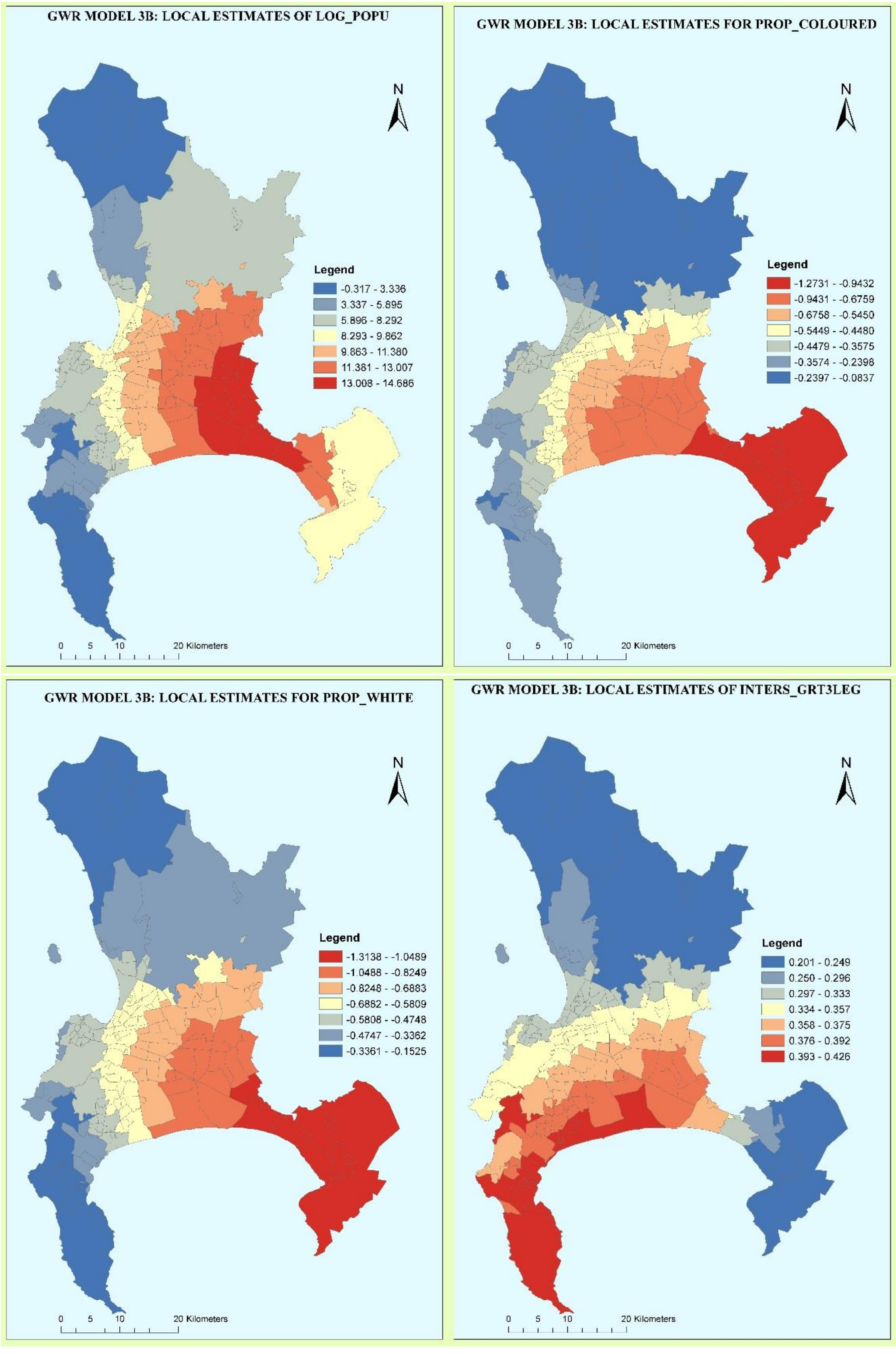

GWR MODEL 2B: LOCAL ESTIMATES OF RATIO_INTERS-CDS
N
Legend
0.614 - 0.617
0.618 - 0.621
0.622 - 0.624
0.625 - 0.626
0.627 - 0.629
0.630 - 0.632
0.633 - 0.635
0 5 10 20 Kilometers
GWR MODEL 2B: LOCAL ESTIMATES OF PROP_FREEWAYS
N
Legend
0.1766 - 0.1773
0.1774 - 0.1778
0.1779 - 0.1782
0.1783 - 0.1787
0.1788 - 0.1792
0.1793 - 0.1798
0.1799 - 0.1806
0 5 10 20 Kilometers
GWR MODEL 2B: LOCAL ESTIMATES OF PROP_EXPRESSWAYS
N
Legend
0.3474 - 0.3488
0.3489 - 0.3500
0.3501 - 0.3507
0.3508 - 0.3512
0.3513 - 0.3519
0.3520 - 0.3532
0.3533 - 0.3544
0 5 10 20 Kilometers
GWR MODEL 2B: LOCAL ESTIMATES OF PROP_PRIMARYARTER
N
Legend
0.3128 - 0.3133
0.3134 - 0.3149
0.3150 - 0.3155
0.3156 - 0.3160
0.3161 - 0.3164
0.3165 - 0.3170
0.3171 - 0.3178
0 5 10 20 Kilometers

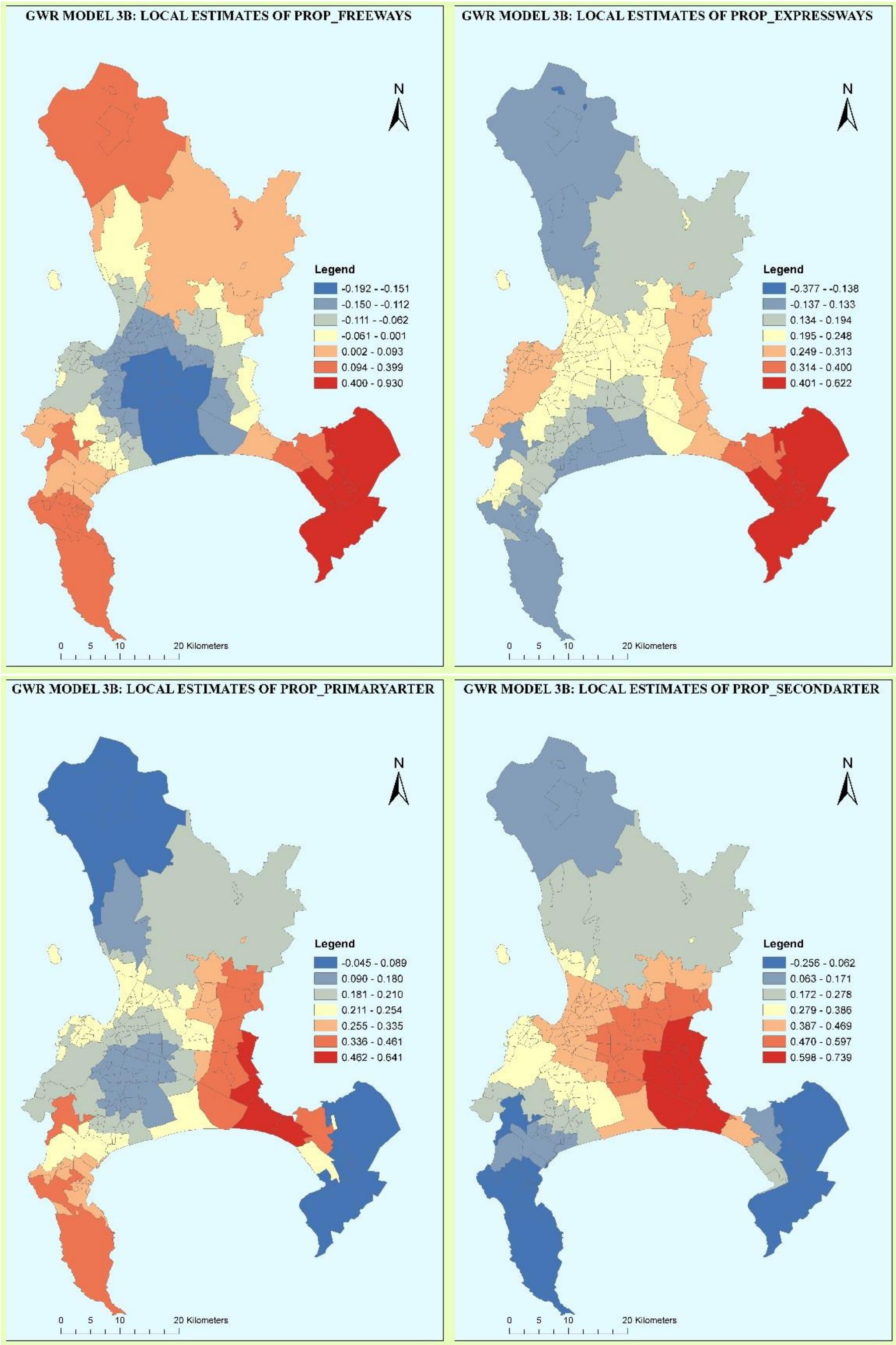

GWR MODEL 2B: LOCAL ESTIMATES OF PROP_SECONDARTER
N
Legend
0.0996 - 0.1004
0.1005 - 0.1013
0.1014 - 0.1018
0.1019 - 0.1024
0.1025 - 0.1028
0.1029 - 0.1034
0.1035 - 0.1041
0 5 10 20 Kilometers


GWR MODEL 2B: LOCAL ESTIMATES OF ROUND-CIRC
N
Legend
0.9901 - 0.9929
0.9930 - 0.9961
0.9962 - 0.9978
0.9979 - 0.9993
0.9994 - 1.0010
1.0011 - 1.0030
1.0031 - 1.0055
0 5 10 20 Kilometers


GWR MODEL 2B: LOCAL ESTIMATES OF PROP_SIGNAL
N
Legend
2.5235 - 2.5312
2.5313 - 2.5400
2.5401 - 2.5440
2.5441 - 2.5469
2.5470 - 2.5498
2.5499 - 2.5547
2.5548 - 2.5641
0 5 10 20 Kilometers

APPENDIX I3: Local estimates of predictors in GWR Model 3B




GWR MODEL 3B: LOCAL ESTIMATES OF PROP_FREEWAYS
N
Legend
-0.192 - -0.151
-0.150 - -0.112
-0.111 - -0.062
-0.061 - 0.001
0.002 - 0.093
0.094 - 0.399
0.400 - 0.930
0 5 10 20 Kilometers
GWR MODEL 3B: LOCAL ESTIMATES OF PROP_EXPRESSWAYS
N
Legend
-0.377 - -0.138
-0.137 - 0.133
0.134 - 0.194
0.195 - 0.248
0.249 - 0.313
0.314 - 0.400
0.401 - 0.622
0 5 10 20 Kilometers
GWR MODEL 3B: LOCAL ESTIMATES OF PROP_PRIMARYARTER
N
Legend
-0.045 - 0.089
0.090 - 0.180
0.181 - 0.210
0.211 - 0.254
0.255 - 0.335
0.336 - 0.461
0.462 - 0.641
0 5 10 20 Kilometers
GWR MODEL 3B: LOCAL ESTIMATES OF PROP_SECONDARTER
N
Legend
-0.256 - 0.062
0.063 - 0.171
0.172 - 0.278
0.279 - 0.386
0.387 - 0.469
0.470 - 0.597
0.598 - 0.739
0 5 10 20 Kilometers