DEVELOPMENT OF A ROBBERY PREDICTION MODEL FOR THE CITY OF TSHWANE METROPOLITAN MUNICIPALITY

By

Nicolas James Kemp
25128095

A dissertation submitted in fulfilment of the requirements for the degree

MSc Geoinformatics

in the Department of Geography, Geoinformatics and Meteorology
at the

UNIVERSITY OF PRETORIA
FACULTY OF NATURAL AND AGRICULTURAL SCIENCES
SUPERVISOR: PROF GREGORY BREETZKE
CO-SUPERVISOR: DR ANTONY COOPER

January 2020

DECLARATION

I, Nicolas James Kemp, declare that the dissertation, which I hereby submit for the degree Master of Geoinformatics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: …………………………..

DATE: …………………………..

Summary

Crime is not spread evenly over space or time. This suggests that offenders favour certain areas and/or certain times. People base their daily activities on this notion and make decisions to avoid certain areas or feel the need to be more alert in some places rather than others. Even when making choices of where to stay, shop, and go to school, people take into account how safe they feel in those places. Crime in relation to space and time has been studied over several centuries; however, the era of the computer has brought new insight to this field.

Indeed, computing technology and in particular geographic information systems (GIS) and crime mapping software, has increased the interest in explaining criminal activities. It is the ability to combine the type, time and spatial occurrences of crime events that makes the use of these computing technologies attractive to crime analysts.

This current study predicts robbery crime events in the City of Tshwane Metropolitan Municipality. By combining GIS and statistical models, a proposed method was developed to predict future robbery hotspots. More specifically, a robbery probability model was developed for the City of Tshwane Metropolitan Municipality based on robbery events that occurred during 2006 and this model is evaluated using actual robbery events that occurred in the 2007. This novel model was based on the social disorganisation, routine activity, crime pattern and temporal constraint crime theories. The efficacy of the model was tested by comparing it to a traditional hotspot model.

The robbery prediction model was developed using both built and social environmental features. Features in the built environment were divided into two main groups: facilities and commuter nodes. The facilities used in the current study included cadastre parks, clothing stores, convenience stores, education facilities, fast food outlets, filling stations, office parks and blocks, general stores, restaurants, shopping centres and supermarkets. The key commuter nodes consisted of highway nodes, main road nodes and railway stations. The social environment was built using demographics obtained from the 2001 census data. The selection of these features that may impact the occurrence of robbery was guided by spatial crime theories housed within the school of environmental criminology. Theories in this discipline

argue that neighbourhoods experiencing social disorganisation are more prone to crime, while different facilities act as crime attractors or generators. Some theories also include a time element suggesting that criminals are constrained by time, leaving little time to explore areas far from commuting nodes. The current study combines these theories using GIS and statistics.

A programmatic approach in R was used to create kernel density estimations (hotspots), select relevant features, compute regression models with the use of the caret and mlr packages and predict crime hotspots. R was further used for the majority of spatial queries and analyses. The outcome consisted of various hotspot raster layers predicting future robbery occurrences. The accuracy of the model was tested using 2007 robbery events. Therefore, this current study not only provides a novel statistical predictive model but also showcases R's spatial capabilities.

The current study found strong supporting evidence for the routine activity and crime pattern theory in that robberies tended to cluster around facilities within the city of Tshwane, South Africa. The findings also show a strong spatial association between robberies and neighbourhoods that experience high social disorganisation. Support was also found for the time constraint theory in that a large portion of robberies occur in the immediate vicinity of highway nodes, main road nodes and railway stations. When tested against the traditional hotspot model the robbery probability model was found slightly less effective in predicting future events. However, the current study showcases the effectiveness of the robbery probability model which can be improved upon and used in future studies to determine the effect that future urban development will have on crime.

Key words: Robbery, crime, kernel density estimation, hotspots, South Africa (S.A), R, social disorganisation, crime pattern, routine activity, time constraint, spatial, geographic information systems (GIS), risk terrain modelling (RTM).

Table of contents

List of Figures

List of Tables

Acknowledgements

I wish to thank and acknowledge the following people for their insights, love and support during my studies.

Thank you to my wife Anette for all your encouragement, patience and help.

To my son Jamian for brightening the mood and enforcing the little breaks that was needed.

My supervisors Prof Gregory Breetzke and Dr Antony Cooper, for your guidance, advice and patience.

To AfriGIS, particularly Magnus Rademeyer, Charl Fouche and Lourens Snyman. Thank you for the opportunity to further my studies and the assistance provided.

Abbreviations and Acronyms

caret   Classification and Regression Training

CSV   Comma-separated Values

EA   Enumerated Area

GIS   Geographic Information Systems

MAUP   Modifiable Areal Unit Problem

mlr   Machine Learning in R

RTM   Risk Terrain Modelling

SA   South Africa

SAPS   The South African Police Service

SP   Sub-place

Stats SA   Statistics South Africa

UK   United Kingdom

USA   United States of America

CHAPTER 1: BACKGROUND

"Though no common definition of the term hotspot of crime exists, the common understanding is that a hotspot is an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization." (Eck et al., 2005:2)

Whether causing physical harm, psychological trauma or economic setbacks, crime affects everyone. In South Africa crime has become part of our everyday life. Higher fences, alarm systems and armed response services have become the norm and those who cannot afford these safety measures simply have to live with more fear. Robbery in particular is a type of crime with an element of fear because the victim is normally confronted by the criminal with threats of physical harm.

This study introduces a novel model with the aim to predict robberies using the physical and sociological environment, mapped as kernel density hotspots. The aim and objectives are explained in more detail during this chapter along with the significance of this study. Furthermore, a brief overview of the crime situation in South Africa is given to inform the reader why there is a need for more crime prevention strategies in South Africa.

1.1   Crime in South Africa

The latest Victims of Crime Survey (2018) made a distinction between crime impacting a household as a whole and crime committed against an individual. According to the survey, both household crime and crime against the individual has increased in South Africa by 5 percent from 2017 to 2018, estimated at 1.5 and 1.6 million incidents respectively). Table 1 shows the number of household crimes and crimes against the individual per province from 2013 to 2018.

Table 1: Incidents of household crime by province (Stats SA 2018)

	2013/14	2014/15	2015/16	2016/17	2017/18
Western Cape	345 600	300 164	307 572	217 428	213 697
Eastern Cape	376 178	315 564	190 397	196 187	183 007
Northern Cape	74 918	86 826	28 534	48 009	39 287
Free State	52 231	57 300	80 291	77 061	85 467
KwaZulu-Natal	157 579	156 134	239 091	282 805	304 626
North West	133 905	138 831	113 425	72 496	103 276
Gauteng	445 768	402 500	498 474	377 834	401 139
Mpumalanga	252 783	248 081	138 609	102 802	128 953
Limpopo	275 909	171 871	103 341	93 658	86 250
South Africa	**2 114 871**	**1 877 271**	**1 699 734**	**1 468 279**	**1 545 701**

Table 2: Incidents of individual crime by province (Stats SA 2018)

	2013/14	2014/15	2015/16	2016/17	2017/18
Western Cape	420 193	377 725	366 138	346 048	261 758
Eastern Cape	219 256	226 347	247 602	262 161	233 166
Northern Cape	50 949	59 455	43 489	40 113	34 710
Free State	137 340	110 452	76 943	81 916	124 333
KwaZulu-Natal	219 485	306 090	226 997	184 980	162 943
North West	142 612	127 208	105 765	90 982	164 383
Gauteng	544 484	558 552	432 340	375 643	486 270
Mpumalanga	162 356	123 688	142 868	116 591	112 276
Limpopo	138 180	74 904	111 903	104 206	102 786
South Africa	**2 034 854**	**1 964 421**	**1 754 044**	**1 602 640**	**1 682 624**

Table 1 and Table 2 show that an increase in household crime occurred in five provinces from 2013 to 2018 while three provinces, including Gauteng, experienced an increase in crimes against the individual. The current study focuses on the City of Tshwane Metropolitan Municipality, which is located in the province of Gauteng. Table 1 and Table 2 further show that Gauteng experienced the most crimes in both household crime and crime against the individual. Gauteng however also has the

highest population, calculated at 6.8 percent households and 4.2 percent of individuals being affected by crime. Of greater concern, Gauteng also experienced a steep increase of six percent in household crime and an increase of three percent crimes against the individual from 2016/17 to 2017/18.

Murder is most often seen as a key indicator of violence in a country (Breetzke, 2010a). Accordingly, South Africa experiences roughly 50 murders a day which is comparable to the United States and China. However, South Africa has a population of six and thirty times less than the United States and China respectively (Breetzke, 2010a). It is also worrisome that the extremely high figure of violent crimes in South Africa has stabilised at such a high level compared to other countries.

According to the Victims of Crime Survey (2017/2018), the public are more sceptical about crime now than during the 2016/2017 period. Approximately 46 percent of the public believe that violent crime increased during the last three years, which suggests that many people do not feel safe or protected against crime. Of greater concern, only 32 percent of the population felt safe to walk alone in their neighbourhoods during the night. These statistics indicate that interventions are needed in the South African context to reduce crime and/or the fear of crime in general.

1.2   Robbery in South Africa

Robbery is seen as the act of taking valuables from a person by the use of force or threat of force, making it a crime that directly impacts the victim as a person. This is different from theft for example, where valuables are taken without the knowledge of the victim at that time. This section also highlights robbery statistics most relevant to the current study which was conducted using robbery data for the years 2006 and 2007.

The 2003 Victims of Crime Survey reported that 14 percent of respondents believed that robbery was the most common crime in their area, which was the second highest crime type following housebreaking with 38 percent (du Plessis and Louw, 2005). Furthermore, it was reported that robbery was among the four highest crimes people fear the most; this is not surprising as robberies as a crime type are perceived as

violent. Du Plessis and Louw (2005) compared the Victims of Crime Surveys for 1998 and 2003 and found that there was a slight decline in the overall crime, as well as robberies which showed a decline from 2.4 to two percent. There was however a steep decline from 41 to 29 percent in robbery reporting from the years 1998 to 2003.

Table 3 contains the figures of reported robberies categorised under robbery with aggravating circumstances and common robbery. These figures were obtained from the South African Police Service's 2007/08 annual crime report.

Table 3: Robberies from 2004 to 2008 (SAPS, 2007/08)

Robbery	Incidence of crime per 100 000 of the population					Raw figures/frequencies				
	2004/ 2005	2005/ 2006	2006/ 2007	2007/ 2008	06/07 vs 07/08	2004/ 2005	2005/ 2006	2006/ 2007	2007/ 2008	06/07 vs 07/08
Aggravating	272.2	255.3	267.1	247.3	-7,4%	126 789	119 726	126 558	118 3	-6,5%
Common	195	159.4	150.1	135.8	-9,5%	90 825	74 723	71 156	64 9	-8,7%

Table 3 shows a decrease in reported robberies from 2006/07 to 2007/08, however there appears to be an increase every second year. The current study also uses the 2006/07 reported robbery cases which had an increase of 6 832 robberies with aggravating circumstances and a slight decrease of 3 567 common robbery occurrences reported. The reported cases for 2006/07 were calculated to 267.1 and 150.1 per 100 000 of the population being victims of robbery with aggravating circumstances and common robbery respectively. SAPS analysed four sub-categories under aggravated robberies and found that 50 percent of street robbery, house robbery, carjacking and business robbery occurs in Gauteng.

1.3 The spatial study of crime

Investigating the link between crime and the spatial location where it took place on the earth's surface is not new. It is however the more recent development of new

technologies such as online mapping, navigation and tracking capabilities which has opened up the field and offered new insights on how geography can aid in the study of crime and place.

Broadly speaking, GIS is a computerised tool created for displaying and manipulating spatial data. With the use of GIS, it is relatively easy to retrieve and analyse spatial data which is subsequently displayed in a visually pleasing manner, making it more understandable to the reader (Burrough, 2001). Spatial data are displayed in a GIS through either vector (points, lines or polygons) or raster representations. Vector data normally displays features on a map as how they occur in the real world, for an example, bus stops as points, roads as lines and dams as polygons. Raster on the other hand is made out of grid cells containing values as a representation of the terrain of interest.

Although displaying maps are normally the expected outcome of a GIS, the computing power, along with the ability to incorporate different data sets like statistics and demographics, with their spatial dimension, is another major strength (Murray et al., 2001). These capabilities and strengths of GIS have generated many spatial analysis methods and techniques. Unwin (1996) noted that what sets a GIS apart from other information systems is the ability to manipulate spatial data through the use of spatial queries, buffers, spatial overlays and detecting changes on the earth's surface. Exploring any kind of data set is essentially the act of statistical detection of patterns, relationships and anomalies within data, the same as with spatial data. The incorporation of spatial statistical methods into the GIS environment has greatly improved exploring spatial data including crime data (Dall'erba, 2009).

Crime analyses are often complex exercises which require the examination of different criteria and variables, while being able to display them all in a map. As mentioned above, GIS software was developed to do precisely that, which is why GIS has been increasingly used by police departments worldwide to develop mapping solutions to various crime problems (Johnson, 2000). Anselin et al. (2000) highlighted two GIS technologies as crime analysis tools. First, the capability of spatial aggregation which makes it possible to measure place-based crime, and second, the ease of which GIS

are able to display relationships between neighbouring matrices across a variety of areal units.

GIS has been used in a variety of studies to address different aspects of criminal activities. For example, Andresen (2005) measured population at risk by investigating the locality of different criminal occurrences in Vancouver, Canada while Gill et al. (2017) examined street segment crime trends in Brooklyn Park, Minnesota, to first test whether the "law of crime concentration" can be applied on a suburban area, and to test the suburban area on stability, variability and concentration of crime trends. Many of these and other spatial crime studies are based on a number of spatial crime theories such as the social disorganisation, routine activity, crime pattern and temporal constraint theory. In this study, these theories inform the development of a robbery prediction model to forecast future robbery events in the Tshwane.

1.4   Research questions

The main aim of this research is to develop a robbery prediction model and to compare the output of the model with a traditional hotspot model

In order to achieve this aim, the following research objectives have been identified.

1) To outline various spatial and temporal crime theories that have been used to explain the crime.

2) To identify a number of geospatial techniques that are used to analyse crime.

3) To identify daytime and night-time robbery hotspots using 2006 robberies occurrences.

4) To develop a daytime and night-time robbery prediction model using variables and features informed by various spatio-temporal theories of crime.

5) To compare the output of the robbery probability model with the output of a traditional crime hotspot model in order to propose a best-fit model for the City of Tshwane Metropolitan Municipality during the day and during the night.

1.5 Significance of the study

The study is significant for a number of reasons. First, very little research has been done analysing spatio-temporal patterns of crime in South Africa. The current study proposes a model which can be used to predict future robbery events in Tshwane. However, the methods proposed can easily be changed to allow for different crime categories and be applied to the rest of the country. Second, robbery is an increasing problem in South Africa generally and Tshwane in particular, and the proposed model can be used as a platform to identify areas, if prioritised by the South African Police Service (SAPS), and which can be targeted for police intervention. Third, geospatial predictive techniques are becoming increasingly used in the spatial analysis of crime. The current study investigates the effectiveness of hotspot analysis combined with regression analysis and a combination of multiple international spatial crime theories as a predictive tool. Fourth, little is known about the applicability of international spatial theories of crime in a South African context. The current dissertation links spatial and time crime theories with the spatio-temporal analysis of robberies within Tshwane. Spatial and time crime theories provide insight on why crime occurs in certain locations more than others, and the theories also depict crime timeslot favoured by criminals. Fifth, the analysis was split into two time periods (day and night). This was done to determine whether the impact of different sets of facilities on crime vary during the day

1.6 Ethics approval

Ethical approval for this research was obtained from the Faculty of Natural and Agricultural Sciences at the University of Pretoria (see Appendix 1). As such, the study was undertaken adhering to all the ethical procedures and processes as outlined by the Faculty. This study was done using publicly available as well as commercially owned datasets. The publicly available data is the Census 2001 Community Profile Database (StatisticsSA) dataset which is freely available for use as long as the user acknowledge Statistics South Africa as the source of the base data. Furthermore, the analysis done in this study was a result of my own processing of the census base

data. The full copyright and disclaimer conditions can be found at the following url: http://www.statssa.gov.za/?page_id=425 (StatisticsSA, 2019b).

The following permissions were granted to use these privately owned datasets:
- Points of interest, municipality boundaries and street centre lines were provided by AfriGIS (See Appendix 2)
- Some maps were also produced using an AfriGIS web map service (WMS) as background imagery. (See Appendix 2)

1.7   Structure of the thesis

Chapter 2 introduces the reader to three spatial crime theories: the social disorganisation, routine activity and crime pattern theories. This chapter also informs the reader about the link between crime and time and explains the basic principles of the time constraint theory as the fourth international spatio-temporal crime theory used to inform the methods employed in the current study.

Chapter 3 reviews a number of geospatial techniques traditionally used in analysing the spatial distribution of crime. These include the basic spatial techniques available in GIS software

Chapter 4 outlines the site of the research, namely the City of Tshwane Metropolitan Municipality. Details of the city are provided in terms of history, demographics and crime specific to the municipality.

Chapter 5 guides the reader through the steps used in the methodology. This chapter includes the techniques and methods used to create the traditional robbery hotspot and robbery probability models. The steps followed to compare the two models in terms of their prediction accuracy are also outlined.

Chapter 6 presents the results and findings of the current study. First, each model presented in the methodology is analysed separately and then compared against each other

Chapter 7 summarises the results and findings of the current dissertation. The various hypotheses proposed are also discussed in light of evidence gathered in the study. Last, the limitations and scope for further studies are also presented.

CHAPTER 2: SPATIAL AND TEMPORAL THEORIES OF CRIME

2.1 Introduction

In this chapter a brief overview of a number of spatial crime theories are presented. This includes an overview of how these spatial and temporal theories of crime originated and evolved over time, as well as some examples of how they have been implemented in various case studies by both international and local scholars. The intention of this chapter is to inform the reader why space and time are important factors in criminal analysis and provide theoretical evidence for decisions undertaken in the current study's methodology.

2.2 Crime and space

Crimes do not uniformly exist across space and time. The notion that geography has a role to play in criminal behaviour has been studied over many decades. In the 19th century, Guerry (1833) mapped personal and property crime occurrences in France. A few years previously Guerry and Balbi (1829) used shaded maps as a display for crime rate intensity. The researchers were also the first to combine multiple variables using statistical graphics and compared personal and property crimes. Their results indicated that both personal and property crime were high in urban areas. Comparing the results to the level of education in areas, they found that areas where residents were more highly educated were highly targeted by property crimes. The researchers argued that wealthier provinces provided more opportunity for housebreakings which was contrary to their expectations. Current studies exploring the relationship between crime and the environment are largely informed by work done in the University of Chicago by Shaw and McKay early in the 20th century.

The term 'environmental criminology' was coined by C. Ray Jeffrey who published '*Crime Prevention Through Environmental Design'* in 1971 (Wortley and Mazerolle, 2008). Jeffrey examined the relationship between crime and the immediate environment and suggested implementing immediate environmental design strategies to reduce crime. Wortley and Mazerolle (2008) describe environmental criminology as

a group of theories structured around criminal events and how their occurrence relates to their immediate surroundings. According to Brantingham and Brantingham (1991), environmental criminology seeks to explain offenders, targets and laws existing at place and time and how these factors work in unison with each other.

The current study seeks to explain the spatial distribution of robberies using a number a spatial theories of crime. Some of the more seminal theories are expanded upon below.

2.2.1   Social disorganisation theory

The seminal theory housed within the school of environmental criminology is the social disorganisation theory of Shaw and McKay (1942). These two theorists developed their theory after discovering that neighbourhood ecological conditions could be used to explain crime and delinquency rather than the individual characteristics of the individual offender. Shaw and McKay (1942) examined delinquency distributions across Chicago neighbourhoods and found associations between high delinquency rates and the underlying socio-demographics of a neighbourhood. In their research they also found three main neighbourhood characteristics contributed to high delinquency rates, namely low economic status, high ethnic heterogeneity and high residential mobility. They argued that neighbourhoods with a low economic status do not have the necessary opportunities for better social ties, nor to enforce social control. According to Breetzke (2010b, p. 1) "The concept that a breakdown of informal social control in families and communities can lead to crime is most directly associated with the ecological perspectives at the macro-level." Furthermore, both ethnic heterogeneity and residential mobility disrupts community cohesion, leading to social instability. Therefore, communities able to reach consensus on what values are important to them and how to address mutual problems are more likely to maintain social order than communities that do not.

Sampson (1986) expanded on the work of Shaw and McKay (1942) by adding social control as a community characteristic. According to Sampson (1986), one of the most fundamental limitations to the social disorganisation theory was the negligence of social control within a city. Sampson (1987) described social control as a manner of

regulating an individual's behaviour at a macro-level; social control refers to a community's ability to reach consensus on mutual principles and values. Sampson further argues that although social stratification such as economic deprivation is important it is also necessary to include family structure and criminal justice sanctions in understanding urban crime. Sampson introduced family disruption as a structural variable while exploring the effect male joblessness and disrupted family values have on violent crimes in black urban communities. He argued that the way families function within a society impacts how informal and formal control are applied on a macro-level. Formal control refers to the enforcement of social norms by authorities such as government laws or policing, while informal control refers to community interventions implemented to create a neighbourhood with desirable living conditions. Examples of informal control includes neighbours keeping a watch out for each other and their property, residents taking interest in the youth and their activities and intervening when someone is not acting in the social norm of the society.

Testing the social disorganisation theory:

Sampson and Groves (1989) were the first to formally test the original model of Shaw and McKay (1942) and added family disruption and urbanisation to the original tenets of the theory. Local friendship networks, control of street-corner teenage peer groups and prevalence of social organisation participation were included as measures of community social organisation. The model was tested and replicated in the United Kingdom which provided additional evidence that structural characteristics have a strong association with criminal victimisation and criminal offending. Since then numerous studies have been undertaken testing this theory in different locations outside of the United States. In Britain, Markowitz et al. (2001) examined whether crime can suddenly change the characteristics of a neighbourhood . Using the British Crime Survey (BCS) conducted in 1984, 1988 and 1992, 151 neighbourhoods in Britain were evaluated based on the effect disorder has on burglaries. The researchers then used non-recursive models to test the effect cohesion and disorder have on each other. The researchers found that there was indeed a feedback loop where neighbourhoods with low cohesion tend to have higher crime rates, which generate more fear and in turn leads to further deterioration of unity within a community.

In Australia, Jobes et al. (2004) investigated the effects of social disorganisation within different kinds of rural communities in New South Wales. The researchers used demographic, economic and social structural measures to group rural communities into different clusters. Structural characteristics in each cluster were then statistically associated with assault, breaking and entering, malicious damage to property and motor vehicle theft. The researchers found that cohesive communities tend to have less crime while one particular small community type with highly disorganised characteristics experienced an extremely high crime rate.

In China, Liu et al. (2016) also tested the social disorganisation theory by examining the relationship between violent crime rates and the neighbourhood characteristics of Jiedaos in Changchun and found that socioeconomic, demographic and land use characteristics has a strong relation with the spatial distribution of violent crime within the city.

In South Africa, Breetzke (2010, p. 448) used the social disorganisation theory to explain the country's high rates of crime. and noted that : "…. a history of repressive racial policing, criminal gangsterism, and vigilantism combined with racist apartheid-era policies had not only contributed to the social disorganization of South African society but established a climate of distrust and fear between and within the diverse races of the country" (Breetzke, 2010a p. 448). The researcher applied the social disorganisation theory to examine the distribution of violent crimes in Tshwane, South Africa, for the years 2001 to 2003 and found that the percentage of unemployed in the population, as well as a 'deprivation index' that was built using dwelling type, water source, toilet facilities, refuse and rubbish removal and type of energy used for lightning, heating or cooking had a positive relationship with violent crimes in Tshwane.

Breetzke (2010) also found that the percentage estranged or deceased fathers and percentage of female-headed households, as measures of family disruption, were not significant predictors of violence. In contrast, residential mobility, measured as the percentage of people that moved into the area in the past five years, was found to be a significant predictor

In another study conducted in South Africa, Swart et al. (2016) investigated whether there is a relationship between adolescent homicides and various neighbourhood characteristics informed by the social disorganisation theory. These researchers defined adolescents as victims between the age of 15 and 19 years, while sub-places provided by Statistics South Africa were delineated as neighbourhoods. Sub-places were specifically demarcated for the national census and therefore the researchers used 2001 census information related to economic status, ethnic heterogeneity, residential mobility and family disruption and found that both female and male adolescent homicides had a significant association with disorganised societies. In another recent study, Lancaster and Kamman (2016) used the 2011 census data to examine whether there are associations between the murder rate and socioeconomic characteristics of police precincts within the city of Johannesburg. The researchers found that renting, informal dwellings, relative poverty, orphans and urbanisation were significant predictors of murder during 2014/2015, while for the 10 year murder rate, significant predictors were population density, unemployment and relative poverty.

The current study uses the social disorganisation theory to inform the data and methods employed. The theoretical framework presented in the above section is used to select social disorganised variables proven to have significant value in explaining crime.

2.2.2 Routine activity theory

The routine activity theory of Cohen and Felson (1979) states that there are three main factors that combine to create a favourable opportunity for crime to occur. First, there needs to be a motivated offender who makes the decision to commit a crime; second, a suitable target and; third, the absence of a capable guardian. A target can also be a range of objects which the offender finds of value whereas a suitable guardian can describe a person or entity, with a CCTV surveillance system being an example of an entity.

The routine activity theory is based on the notion that crime opportunities arise with the movement of people as they attend to their daily activities (Chainey and Ratcliffe, 2013). These researchers explain crime opportunity as a measure of available targets and the absence of capable guardians, while they consider the offender as a person permanently on the lookout for crime opportunities. In terms of external controllers,

the influence an individual may have on the behaviour of a motivated offender has led to some enhancements to the routine activity theory. For example, Eck et al. (2010) introduced the notion of a handler, a guardian and a place manager as controllers. A handler is someone the offender respects enough not to commit a crime in their presence. Examples of handlers include parents, coaches and religious leaders. A guardian is seen as someone feeling obliged to protect the target such as security guards, police or residents looking out for each other. Both handlers and guardians are seen as controllers over the offender but differ in the sense that guardians reduce the opportunity of crime, while handlers reduce the chance of an offender acting on the opportunity that exists. Finally, a controller is a place manager who discourages crime at a target, usually because of an investment or obligation towards that particular place. Examples of place managers include shop owners, landlords, janitors, and floor workers. According to Eck and Weisburd (2015), routine activity theory suggests that an offender and a target need to be at the same place at the same time in order for the crime event to take place.

Testing the routine activity theory:

A number of studies have been forthcoming examining the applicability of the routine activity theory in a variety of contexts. Andresen (2006) examined the influence of social disorganisation and routine activity theory on crime in Vancouver, British Columbia, and found that unemployment was the greatest predictor of crime rates in the city. Motivated offenders, measured by the percentage of young population, were also found to be a significant predictor.

Mustaine (1999) investigated woman's stalking victimisation involving 1 513 college or university students from eight different states in the US and found that employment status was the only significant predictor. The researcher argued that a plausible reason could be that women who are employed spend more time in public, while women who are unemployed spend most of their time during the day at home and are thus more likely to be targeted.

Drawve et al. (2014) examined aggravated assault cases between 2007 to 2009 in the United States and found that the chance of being arrested was higher when the victim was between the age of 15 and 55, the time of the aggravated assault happened

between 6pm and 7am, and when the offender was suspected of being intoxicated. It can be assumed that targets between the age of 15 and 55 are more capable of defending themselves than people either younger or older and that people are more crime alert during the evenings, making them less suitable targets, leading to more arrest. In terms of the intoxication, generally people under the influence of drugs or alcohol do not make rational choices and act more on impulse, rather than planned decisions, leading to lack of judgement when selecting a target, place and time. Furthermore, although the results failed to prove that bias-motivated crimes are more likely to be cleared by an arrest, the motivation factor of a non-gang related and intoxicated assault still aligns with the routine activity theory.

The routine activity theory has also been tested by investigating the linkage between crime and places. Sherman et al. (1989) investigated predatory crime hotspots in Minneapolis, Minnesota, USA, to find out whether certain places are 'criminogenic'. The hotspots were identified by linking 323 979 police calls, received over a period of 1 year, to 115 000 addresses and intersections. The researchers found that robberies are highly concentrated and can effectively be avoided by staying clear of certain places. Even if there is no direct location association between the place investigated and a specific type of crime, the proximity effect of that place might stretch outside its borders, meaning that the place creates crime opportunities in the near vicinity, but not necessarily at the place itself. Freisthler et al. (2016) also found a positive association between the area immediately adjacent to medical marijuana dispensaries and violent and property crimes committed in California, USA Relatedly, Drawve et al. (2014) investigated the risk involved for an offender to commit a crime by measuring the relationship between place committed and arrests made. The researchers found that offenders are at a greater risk of being arrested when committing a crime at a school or college, where an arrest is less likely when the crime occurs at the victim's residence. The researchers also found that the lowest risk for offenders were at parking areas, roads, bars and other places expected to have low guardianship and therefore are less like to be arrested.

The routine activity theory has also been investigated in a South African context where Breetzke and Cohn (2013) investigated the effect gated communities have on residential burglaries in Tshwane. In contrast to the expectation that burglary offences

will be lower in gated neighbourhoods, the researchers found that burglaries may in fact be higher in these locations. The researchers however noted some limitations in their research, including the fact that they did not differentiate between security villages and enclosed neighbourhoods.

Warchol and Harrington (2016) investigated illegal trading in abalone in Table Mountain National Park and Cape Agulhas National Park using the routine activities theory as a guiding framework. The researchers suggest that the economic turnover from illegal harvesting and trading of abalone outweighs the effort and risks involved, making abalone a suitable target. As for motivated offenders, the researchers found that there are mainly three types of offenders, local poachers from small villages near the parks, small scale commercial poachers who gain access to abalone within the park and commercial poachers who use vessels to gain access from the ocean. In terms of guardianship, the researchers found that capable rangers had a greater positive influence than natural or man-made barriers.

The routine activity theory (RA) provides insight on how the population's routine activities can increase their risk of criminal victimisation. Osgood et al. (1996) noted that it is not as much the person but rather the situation that motivates delinquency. Physical characteristics can influence how and when the daily activities are being performed. According to the RA theory, facilities increase the foot traffic and the influx and outflow of people in an area. The notion is that with the increase of people the likelihood of available targets increases and this creates motivation for crime.

2.2.3 Crime pattern theory

Crime pattern theory seeks to explain crime and place by combining rational choices made by offenders while participating in everyday activities (Eck and Weisburd, 2015). The theory was developed by Brantingham and Brantingham (1975, 1981, 1993) who argue that urban planning spatially distributes the daily activities of people. The theory motivates that offenders, just like non-offenders, traveling in a routine manner, build a cognitive map of their surroundings, making them aware of spaces they feel comfortable in. It is in these areas that they are most likely to commit a crime. Crime pattern theory can therefore be explained through the concepts of nodes, paths and

awareness spaces. Brantingham and Brantingham (1995) describe nodes as places people visit during their lives which include central nodes such as a person's home, work-place, school, grocery stores and other places one is expected to visit almost on a daily basis. Paths refer to the networks connecting the latter activity nodes. According to Brantingham and Brantingham (1993) property crime not only cluster around high activity nodes but are also spread out between them, following the paths as connection lines. This is especially true when investigating main roads which are used by many people on a daily basis. The notion here is that when a large number of people travel from node to node, the higher the probability is that a potential offender is among them. While people, including potential offenders, travel from node to node using the pathways, they become aware of their surroundings, referred to as awareness spaces by Brantingham and Brantingham (1993). According to these researchers, the proximity in which crime occurs in relation to paths and nodes broadens when an offender is familiar with an area and clusters around the main paths when the area is less well known. This is because of the cognitive awareness of spaces that offenders build up and feel safe to operate in. The physical characteristics of a neighbourhood usually determine why the area becomes an awareness space in the first place. In the current study, facilities are used to explain the physical characteristics of an area.

Crime generators, attractors and risky facilities:

Anselin et al. (2000) explain that certain facilities that bring people together on a routine basis can be used to explain the spatial location of crime. These facilities can broadly be classified into either crime generators or crime attractors. Brantingham and Brantingham (1995) describe crime generators as facilities where a group of people, with no initial intention of committing a crime, are attracted to. Examples of crime generators are entertainment districts, education facilities, and grocery stores. Crime attractors on the other hand are places known to attract people seeking crime opportunities such as bar districts, drug markets, and shopping centres.

Crime generators are generally easily accessible to the public and attract people in large groups, which in return create opportunities for crime and may become crime hotspots (Bernasco and Block, 2011). Crime attractors can also create crime hotspots but mainly as a result of the type of activities available at those areas.

Different facilities generate or attract crime because of the different dynamics each facility offers. Some studies have shown the importance of only one type of facility. For example Groff and McCord (2012) found that a neighbourhood park acts as a crime generator. The researchers examined the role of parks as crime generators in Philadelphia, Pennsylvania, USA, and found that higher levels of crime were found in the close proximity of neighbourhood parks. Parks were noted as crime generators because they are areas open to the public and do not belong to any one citizen making guardianship problematic. They found that parks with low guardianship create increased opportunities for criminal behaviour.

One particular type of facility may explain a large number of crimes, but the number of crimes may differ for each of those facilities depending on a combination of other type of facilities found in proximity. For example, Hart and Miethe (2014) investigated interpersonal violence in proximity to bus stops and other activity nodes in Henderson, Nevada and found that there is a higher chance of being robbed around a bus stop than any other activity node, but the risk varies widely with the presence of different activity node combinations. For example, there was always a bus stop present when a robbery occurred in an area with an ATM, bar, fast food restaurant, gas station, shopping plaza and smoke shop, but only 10 percent of the time when a robbery incident occurred close to a bar only. So although crime may be more likely at one type of facility, any other facility causing an influx of people or attracting criminals has the potential of attracting criminal activity nearby.

Bernasco and Block (2011) investigated the influence of crime generators and attractors on robberies using a block level analysis in Chicago, Illinois, USA. 75,065 incidents of street robberies were recorded during 1996 to 1998 at one of 24,594 street blocks. In their study, main roads and public transportation stations were identified as crime generators, as they cause an influx of people to the same place at the same time. Numerous shops and businesses thought to have large cash transactions and which had only ten or fewer employees were seen as general crime attractors whereas drug, gambling and prostitution related incidents were seen as more specific crime attractors. The researchers also used offender anchor points such as gang activity and addresses of known robbers, which could be geocoded to a street block. Street blocks containing either a crime generator, an attractor or an anchor point, had higher

robbery activity than adjacent street blocks, whereas street blocks without any of the former points and that are not adjacent to street blocks containing such points had the lowest robbery counts.

In a more recent study, Demeau and Parent (2018) studied the effect crime generators and attractors have on crime in Montreal, Canada and found that schools, convenience stores, grocery stores and subway exits have a positive effect on assaults, theft, robberies and motor vehicle thefts The authors however also found that laundries and pawnshops did not have any effect on the four latter mentioned crime types which is in contrast with many North American studies. This implies that it is not always possible to generalise certain types of facilities as crime generators or attractors. The researchers provide possible reasons for the difference in criminogenic facilities including the fact that Montreal is surrounded by a natural barrier which may have an effect on crime distribution as well as the fact that different cultures exist in different cities which impact the utilisation of the same type of facility across different areas.

Of course not all facilities of the same type experience the same amount of crime. Eck et al. (2007) describe facilities with higher crime rates as "risky facilities". The researchers suggest that only a few facilities within a facility group will contribute to the majority of crime events, resulting in a J-curve crime distribution. Five possible reasons why the majority of crime occurs only at a selected few facilities include random variation, crime reporting, targets, offenders, and place managers. Random variations and crime reporting limit the true effect of a facility to be known and refers to higher crime levels found at a specific facility simply by chance and is only relevant to that point in time. Reporting of crime refers to the fact that some facilities may simply report higher crime than others, creating the illusion that crime levels differ at facilities. Legitimate reasons for higher crime rates at certain facilities revolve around the crime pattern theory suggesting that facilities create targets by attracting people which also generates offenders. In terms of the place manager, decisions made at a facility involving security, employment, opening hours will also have an impact on the amount of crime, resulting in different crime levels found at each facility.

In many cases risky facilities are used to explain crime patterns. Groff and La Vigne (2001) created a burglary opportunity surface which identified areas and facilities as a crime risk grid. A uniform grid was created over the Grier Heights neighbourhood in Charlotte, North Carolina, USA. For each facility, the underlying grid cells were added a value of one, resulting in summed cell values for all variables. The results were tested on known burglary events and found that 73 percent of all burglaries fell between one below and one above the standard deviation of the mean grid cell values, while 21 percent of burglaries were found from one to three standard deviations above the mean. The study concludes that 94 percent of all burglaries fell in the medium, high or very high potential areas, showing the benefits of using risky facilities as crime predictors.

Risky facilities are often nodes a large group of people visit on a daily basis, causing an influx of movement patterns inside the facility as well as in neighbouring areas. It is therefore not uncommon to see the crime impact extend beyond the borders of the risky facility. Bowers (2014) investigated the relationship between crimes happening inside facilities and crime occurrences in the surrounding area. The author investigated 30,144 thefts in a large metropolitan area in the UK dated from 1st January 2005 to 31st August 2009 and found that 21 969 thefts occurred between facilities that experienced one or two internal thefts. The results have therefore shown a positive relationship between facilities experiencing crime and crime occurring in close proximity to those facilities.

The awareness space:
Eck and Weisburd (2015) explain that whilst the routine activity theory only focuses on the type of people and their locational presence related to the target, crime pattern theory is also concerned about the offender's activities in terms of access to a place and how that place got their attention in the first place. Brantingham and Brantingham (1993) suggest that criminals act in areas where they are most comfortable and therefore commit crime close to their central nodes on interest and activity. Characteristics of places also determine the frequency and type of available activities, presenting the argument that some nodes have a stronger temporal tendency towards crime than others. Offenders also become more aware and familiar with routes while commuting between familiar areas, broadening their cognitive map and in doing so,

create new awareness spaces. However, not all awareness spaces have something of value to offenders and therefore crime opportunity areas exist where potential targets are located, but only become at risk when an offender becomes spatially aware of these spaces. According to Chainey and Ratcliffe (2013) not all awareness spaces have crime opportunities and not all crime opportunity areas are at the same risk, because of different types of facilities within the area. A previously known high crime area might force offenders to explore other areas, or new facilities can attract people, making the offender aware of new crime opportunity spaces.

The researchers also present the idea that crime opportunity areas exist outside the offender's cognitive map because of insufficient funds. Daily activities cost money and offenders often resort to crime because of their economic circumstances in the first place, so the lack of money further restricts their movement, limiting their awareness spaces. Another reason why offenders are less likely to commit crime outside their awareness spaces is the fear of standing out. Offenders tend to commit crimes where they themselves feel safe, often keeping the poor from committing crimes in more affluent areas or offenders acting in a different ethnicity neighbourhoods as their own. One might argue that this is different in South Africa because of the racial segregation that happened under the apartheid government, forcing many black African families to reside on the urban periphery of cities. It's therefore not uncommon for the poor to travel a far distance into the more affluent neighbourhoods in search of work or "piece jobs", sometimes to such an extent that they cannot afford to go home or pay for alternative housing, forcing them to a life on the streets.

Testing crime pattern theory:
Numerous international studies have been conducted testing the central tenets of crime pattern theory. For example, Groff et al. (2014) investigated the physical environment and crime changes in 355 municipalities in the metropolitan area of Philadelphia, Pennsylvania, USA and found that land use, transportation networks, street and highway patterns and rail barriers as measures of permeability across Philadelphia were positively associated with violent and property criminal events. Caplan et al. (2011) investigated crimes committed in Newark, New Jersey, USA, by reoffending parolees and found weak but positive spatial relationships between the local crime hotspots and reoffending parolee criminal activities. As expected, the

researchers found no associations between crime and the parolees' homes, with the fear of being recognised thought to be the reason. Summers and Johnson (2017) on the other hand examined the notion that crime would be more frequent on busy street networks, as offenders are more likely to travel on these street segments and so become aware of potential crime spaces while on route. The researchers conducted their study in London, UK, and found that serious violent crimes that occurred outdoors are concentrated on high-choice street segments and therefore provide evidence in favour of the crime pattern theory.

In a South African study, Hiropoulos and Porter (2014) examined the influence of crime attractors and crime generators on theft from motor vehicles in the province of Gauteng. The study found that theft from motor vehicles did not occur randomly over space and found clusters occurred in specific precincts. To examine the role of crime pattern theory, the researchers studied the location of shopping centres, major roads and major nodes in relation to crime and found that most shopping centres fell in areas with a high rate of theft out of motor vehicles as was the case with major nodes, which consisted out of retail and industrial nodes as well as central business districts. Major roads provide access points well known to offenders and frequently used by potential targets, making major road locations attractive to crime. This was indeed the case as the researchers found that locations of major road hubs were associated with precincts considered to have a high theft out of vehicles rate.

The crime pattern theory provides important insight on which facilities might act as major nodes, creating awareness spaces in close proximity to them. The current study seeks to explain the spatial relationship between facilities and robberies occurring in Tshwane by identifying awareness spaces around facilities as hotspots. The crime pattern theory therefore helps in the decision making of which facilities are known to create awareness spaces for a large number of people. The crime pattern theory in combination with the routine activity theory provides the framework of the facility risk model, used in the current dissertation to measure the effect of facilities on robberies.

2.3 Crime and time

2.3.1 Introduction

The proliferation of Geographic Information Science (GISc) and geographic information systems (GIS) over the past few decades has resulted in a strong focus on understanding the spatial dimension of crime. The important role that time plays in the commission of a crime has as a result been sorely neglected. Time is however intertwined with spatial-based crime theories, as there is always a time attached to a criminal event. Ratcliffe (2006) divided the temporal framework for time geography into three categories: capability constraints, coupling constraints, and authority constraints. The former can be described as the physical and biological limits the human body adheres to such as the need for rest and the maximum ability humans possesses in using their senses. Coupling constraints refer to human activity in which individuals need to interact in society to sustain their standard of living, while authority constraints dictate the admission rights of individuals determining where and when they are allowed to interact in society. This time-geography framework dictates when people are more likely to participate in their daily activities, linking time to victim availability, guardianship and access to awareness spaces, which according to location-based theories are necessary for a crime to take place.

Despite time in itself being a continuous variable, it is most often measured using discrete intervals known as "snapshots" (Ratcliffe, 2006). Time "snapshots" can include events happening at various levels of temporal resolution including yearly, monthly, weekly, daily, and hourly. Gorr and Lee (2015) raised their concerns about using these fixed time slots as it presents a similar "boundary" issue as caused by the modifiable areal unit problem (MAUP) when using fixed boundary sizes. In other words, selecting a fixed time slot can cause the misrepresentation of crime, particularly if crime occurs across different time intervals. Nevertheless, an increasing number of studies use these discrete time intervals to examine crime and its temporal distribution (see Breetzke (2016); Conrow et al. (2015); Ratcliffe (2010)).

2.3.2 Time in crime studies

The role of time in crime has been investigated in a number of studies covering a variety of aspects involved with time such as time of day, season, and periodicity.

Furthermore, the physical characteristics that changes with time such as traffic and daylight have also been investigated. However, it is not always possible to determine the exact time of the crime occurrence. Ashby and Bowers (2013) raise the concern of indeterminate or so-called aoristic crimes, which make it difficult to determine the exact time of the criminal event. Burglary is a typical example of an aoristic crime as owners are usually not present and can only specify a time range in which the burglary could have taken place. The researchers examined whether the exact time can be estimated from the time range by conducting a study of cycle thefts from railway stations in London, UK. After victims reported the time they left their pedal cycles at the train station and the time they noticed that it has been stolen, CCTV systems were used to determine the exact time of theft. The results of the different estimation methods showed that it is not advisable to use the start, midpoint or end point as the exact time. Rather an aoristic and random method were developed and found to be most suitable to estimate pedal cycle theft peak times. Their study concluded that half of the pedal cycle thefts happened between 13:04 and 18:52, which can be used to determine optimal deployment of police resources. Time and place are however not always a constant in crime analysis and certain places can sometimes become heightened for only a distinct time period

Johnson et al. (2007) examined the pattern of space-time burglary events in ten areas situated in five different countries and found that in all ten areas, houses within 200m of a burglarised house experienced a higher risk of burglary for a time period of at least two weeks post the initial burglary. The magnitude of burglary victimisation across space and time did however differ between countries. In Australia and the Netherlands the probability of a house being burglarised had a wider range in distance than in the USA, which experienced a more localised effect. In terms of time, the probability of a house being burglarised that is in close vicinity to a previously burglarised home was greater for a longer period of time in Canberra, Australia, The Hague, Netherlands, and Philadelphia, USA than in the other study areas. Studies interested in the probability of a crime happening at the same place or in close proximity are called repeat and near-repeat victimisation studies. These studies specify time periods in which to expect heightened crime in and close to the initial incident.

Temporal constraints on places:

Seeking to explain the spatio-temporal relationship between bars and crime, Conrow et al. (2015) conducted a study in Buffalo, USA, to find out whether crime around alcohol outlets became more clustered after receiving their license. Space-time analysis was done on a 100 ft. incremental spatial range starting from 20 ft. to 2640 ft. and a 15-day temporal incremental range starting from 30 to 180 days after outlet's alcohol license was issued. The results showed a relationship between alcohol outlets and crime and found that for more than half of the 432 individual alcohol outlets, the maximum risk occurred between 30 and 45 days after the license was issued. Furthermore, crime in general seems to decrease in Buffalo between 12pm and 5am, except for areas within increased spatial proximity to bars where the opposite effect was observed. Haberman and Ratcliffe (2015) also inspected places thought to have an influence on crime by examining whether those places show different criminogenic time periods during the day. The authors conducted their study on street robberies between 2009 and 2011 in Philadelphia, Pennsylvania, USA, and found that some places contributed to higher activities of street robbery than others but only during certain times of the day. They also found more occurrences of street robberies at the same place but at different times during the day. The results provide evidence that in some cases a facility's operational dynamics and business hours can dictate street robbery occurrences i.e. fast food outlets receive customers on a regular basis and therefore create street robbery opportunities during multiple time slots whereas check-cashing stores are busier during non-working hours, generating more crime opportunities during those particular hour slots. Pawn shops on the other hand have specific business hours and only contribute to high levels of street robbery during open hours.

Guardianship at specific facilities also varies across time slots during the day, changing the likelihood of street robberies. Large groups of people can be seen as guardians (routine activities theory), resulting in lower levels of street robbery at train stations during the peak morning hours but higher levels later the same day when there are less guardians present. Criminogenic places (crime pattern theory) such as high schools are more complex as robbery counts can be high in the immediate vicinity

during and after closing hours. Sometimes the spatially lagged effect is seen well after closing hours as a result of extracurricular activities held on school premises.

Natural time constraints:

Natural factors such as weather and daylight vary temporally and can also impact on crime patterns. Indeed a large and growing number of studies have examined the impact of various meteorological parameters and crime (Linning, 2015, Pereira et al., 2016, Cohn, 1990, Ranson, 2014, Schutte and Breetzke, 2018). Uittenbogaard and Ceccato (2012) examined the influence of time of day, weekdays and weekends as well as seasons on space-time crime clusters. The results showed that property crime tends to take place during the afternoon while violent crimes seem to peak during night-time. Generally weekends tend to be more socially inclined with the assumption that people have more free time available than during the week. Therefore, the peak in violent crimes during weekends can be explained by routine activity theory which suggests that a greater potential of crime exists where a larger number of people interact with each other. The peak of property crime during the evening can also be explained by people generally having more free time during the evenings. Uittenbogaard and Ceccato (2012) also investigated the crime seasonality trends and found no fluctuations in property crime, but violent crime levels were noticeably higher in the warmer summer months than during winter. Linning (2015) conducted a seasonality study on property crime in Vancouver, BC and Ottawa, ON Canada and found that there were no extreme variations in micro-spatial patterns of property crime during the year. However, the researcher found that the more humid Ottawa displayed temporal peaks whereas the temperate Vancouver did not. In the more tropical city of Recife, Brazil, Pereira et al. (2016) also found that there were only slight increases of homicides during warmer and drier months, but were not statistically significant. There was, however, evidence that homicides peak during weekends and evenings.

Crime happening at specific times is more often than not caused by the routine activities people follow. However, in some cases, restrictions during certain times of the day, whether implemented by humans or nature, can cause crime to temporally spike. Carrillo et al. (2018) studied the implications of driving restrictions in Quito, Ecuador on criminal activity. The driving restriction was implemented during peak hours for both pollution and traffic congestion reasons. Although these restrictions

were implemented for good reason, its impact causes displacement of police resources, resulting in higher crime rates in both the restricted areas and during peak hours.

In a South African context, Breetzke (2016) examined the periodicity of crime in Tshwane, South Africa, and found that violent crime peaks during the warmer summer months. This suggests that crime opportunities increase where people are drawn together during their daily activities; in summer people tend to spend more time outdoors than in the cooler winter months. The researcher also noted that violent crimes peak over the weekends and also during the night from about 8pm to 7am. These are typical time slots that people are not at work and most likely have time to spend on events where time is less of a constraint. It is however interesting to note that Suffla and Seedat (2016) found the opposite results in homicidal strangulation in Johannesburg, South Africa where the researchers found that most of the recorded strangulation incidents happened over weekdays and during the day.

2.3.3 Temporal constraint theory

One of only a few theories that specifically focus on time as their central element includes the temporal constraint theory of Ratcliffe (2006). The time constraint theory motivates that the likelihood of crime occurring is related to time spent by individuals along activity pathways and at nodes. Indeed, time dictates when individuals are busy with their daily activities and when they relocate between activity nodes. Ratcliffe (2006) explains the temporal constraint theory using events happening in a day to day journey and the time involved to complete the particular journey. Employees usually have to be at work at a certain time or risk losing their jobs; therefore they need to plan on when to leave home in order to arrive at work on time. Not wanting to be late, people tend to include transportation method, route, travel time, traffic and extra times for unexpected events into their trip planning. Accordingly, an employee will estimate the time spent at each traveling node, calculating whether they are on time. While on the journey to work, the employee may become aware of a crime opportunity at a travelling pathway, while the likelihood of becoming aware or acting on a crime opportunity increases with the time available at that node.

Ratcliffe (2006) demonstrated this theory using an artificial street network to explain spatial patterns in opportunity-based crime. The researcher hypothesized that participating in daily life leaves little time for offenders to explore areas outside their least-distance path. Ratcliffe emphasizes the fact that the temporal constraint theory does not explain all crime, for instant, career criminals will be less concerned about time constraints. Nevertheless, implementing time constraints has long being recognized by communities as a valuable crime restrictor and by knowing the temporal constraints of a potential offender, then areas can be identified as a potential crime risk. Furthermore, Ratcliffe (2006) notes that property crime in particular can be determined by implementing the temporal time constraint theory in relation to the offenders traveling nodes.

The time constraint theory has not yet been applied to South Africa. The current study aims to fill this gap in part by incorporating commuter nodes into the analysis. The commuter nodes acts as visiting points between a potential offender's home, work, school and other activity nodes. According to the time constraint theory, while an offender is at one of these nodes an offender has limited time to explore other areas. The current study therefore examines the relationship between robberies and proximity to commuter nodes, as a time constraint model, within the City of Tshwane Metropolitan Municipality.

2.4   Summary

The three spatial crime theories discussed explain in different ways how locations influence crime. The social disorganisation theory is based on the notion that neighbourhoods experience different crime levels based on the existence of social disorganisation within those neighbourhoods. The social disorganisation theory is therefore focussed on the demographic characteristics of a neighbourhood and how that impacts the risk of crime. The theory is different to the other two theories as the routine activity and crime pattern theory are more interested in how population movement and facilities that the population visit influence lifestyles and behaviours that can lead increase the risk of criminal behaviour.

In this study I use aspects of all three theories in the construction of robbery risk prediction model. The social disorganisation theory is used to motivate for the inclusion of the neighbourhood as the unit of analysis. The routine activity and crime pattern theory aids in the selection of criminogenic facilities. Of course, facilities also fall within a neighbourhood and therefore the selected facilities are weighted in the analysis according to the level of social disorganisation in the neighbourhood.

Crime is not evenly distributed over time, stressing the importance of understanding time as a component in crime analysis. The temporal constraint theory implies that crime has a direct link to time spent at a traveling node. The general assumption is that there is a vast difference between day and night-time activities and therefore crime patterns will differ between the two time zones. In this study I used time as a variable to explore the difference between day and night-time robberies as well as incorporating proximity around commuter nodes as a measure of time constraint while traveling between activity nodes.

CHAPTER 3: THE SPATIO-TEMPORAL ANALYSIS OF CRIME

3.1 Introduction

Place-based crime analysis has made some fundamental advances over the past few decades due to new computer capabilities. This has been driven mainly by Geographic information systems (GIS) which have the capability to not only represent crime visually but to explain the relationship between criminal activities and place. Computers have also made it easier to manage data such as police records and store information gathered from computer aided dispatch (CAD) systems in such a manner that criminal activities at a certain place can be systematically analysed.

The popularity of GIS has resulted in this technology finding itself into crime-specific programming and software platforms. Indeed, the integration of spatial capabilities into programming platforms has also made it possible to develop predictive models using spatial referenced data. In this way spatial data can be analysed in a manner not always easily available in standard GIS packages, but can be exported in a format that can be utilised by these GIS packages.

In the section below I outline two main geospatial methods/techniques commonly used in crime analysis. The purpose of this chapter is to inform the reader about these two techniques and describe their relevance to this study.

3.2 Hotspot analysis

Hotspot analysis is a popular technique used to visualise crime. Eck et al. (2005,p. 2) explain that "…the common understanding is that a hotspot is an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization". Crime hotspots can be visualised on many spatial hierarchical levels including point, street, neighbourhood and other larger political or statistical boundaries.

Eck et al. (2005) lists a number of different hotspot crime mapping techniques including spatial ellipses, thematic mapping, grid thematic mapping and kernel density

estimation (KDE). Spatial ellipses search for the areas with the most points concentrated close to each other and create a standard deviational ellipse around those areas. The intensity and spread of point cluster areas are shown by the size and orientation of the ellipses. A thematic map is an easy to produce, easy to read method of displaying geographic distribution. Aggregated crime events are thematically mapped to identify and display crime patterns that may exist in the study area. A limitations of thematic mapping is that thematic maps rely on a user's ability to determine the type of distribution and number of classes, which can lead to different maps using the same data. In other words, the user can determine whether a neighbourhod is a hotspot or not.

Grid thematic mapping is another mapping technique. This involves aggregating the points of interest to cell blocks of a user specified size and then thematically classifying those cells. Creating uniform grids over the study area not only eliminates the different boundary size issue, but can identify hotspots within the bigger boundary. Of course, the granularity of the resultant hotspot map in dependent on the resolution being employed. A coarse resolution will result in a coarse hotspot map, while a finer resolution will result in a smoother appearing hotspot. The kernel density estimation (KDE) is similar to grid thematic mapping, but uses a propability method to smooth the points over the grid cells instead of just aggregating points to a single cell. KDE is explained in greater detail in the section below.

Numerous studies have shown the effectiveness of hotspot analysis in a variety of contexts (Liu and Brown, 2003, Weisburd and Eck, 2004, Braga and Bond, 2008, Chainey et al., 2008, Ratcliffe et al., 2011, Wain et al., 2017). For example, Braga et al. (2014) examined a multitude of 25 crime prevention tests using hotspot analysis. The researchers found that 20 out of the 25 hotspot tests proved to be successful in preventing crime. However, these 25 hotspot tests were presented in 19 case studies of which the vast majority were conducted in the United States of America. Nevertheless, the results of their meta-study proves that identifying crime hotspots, no matter where or what the cause, is beneficial to law enforcement agencies.

Chainey et al. (2008) examined the accuracy of hotspot maps for predicting spatial patterns of crime in London, UK. A comparison between various crime analysis

techniques such as spatial ellipses, thematic mapping, grid thematic mapping and KDE were undertaken in order to find the best hotspot mapping technique. The hotspot techniques were tested for different time periods and crime types. The researchers found that out of all the techniques used, the kernel density estimation (KDE) technique outperformed the rest in regard to predicting future crime events throughout all of the variations explored.

3.2.1 Kernel Density Estimation (KDE)

The kernel density estimation (KDE) is a common method used to create hotspots. The method involves the placement of a uniform grid over the points in question. Next a specified bandwidth is used to search for points within a distance from an identified cell, and finally, an estimation of point density for each cell is calculated. Mathematically, kernel density estimation is a probability function which smooths point distributions over grid cells as a density value. The density value for each cell is calculated by taking the distance of each point in relation to other points within a specific radius into account and distributing the density of these points among the underlying cell. A cell block receives more of the distributed value the closer it is to points. The mathematical equation of a KDE is as follows:

$$f(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} k\left(\frac{d_i}{h}\right)$$

The density value $f(x, y)$ is at the location (x, y). In the equation n represents the total number of observations with h as the bandwidth selected, d_i represents the distance between observations, i and the location at (x, y), and finally k as the kernel. Gorr and Lee (2015) describe a kernel as a three-dimensional density function which uses each event containing cell as a centre. The kernel distributes weights to each grid cell by searching for events within the specified radius from the centre, the closer to the centre the higher the weight which is used to calculate a final score for each grid cell by adding all the weights for that specific cell together.

There are two main factors to consider when using KDE to construct hotspot maps. These include 1) selecting the appropriate cell size as well as 2) selecting the

appropriate bandwidth. By cell size, I refer to the length and width of each individual cell. By bandwidth I refer to the search radius in which each cell searches for points to use in the density estimation.

Regarding the cell size, Chainey and Ratcliffe (2005) suggest a method of dividing the shortest side of the study area's bounding box by 150. A bounding box is the smallest rectangle that can be drawn to cover the whole area of interest. The method requires that the sides of a rectangle drawn over the study area should be measured and the length of the shortest side, divided by 150, should be used as the side length of the square block resembling a grid cell. Chainey (2011) notes that this method of choosing a cell size can also be used to calculate an optimal bandwidth by multiplying the value by five. Chainey (2013) notes that this method should only be used as a starting point and needs further intervention by the user, as the size of the bandwidth is directly linked to how the densities of points are smoothed out over the study area. A large bandwidth size can therefore decrease the prediction accuracy when points far away have an influence on a cell value, which is unlikely to be the case in a real world scenario. However, a smaller bandwidth size can lead to cellblocks not incorporating the value of near neighbours leading to fuzzy maps with lots of small hotspots.

Chainey (2013) tested the prediction accuracy on a 30 meter incremental range of cell sizes, starting from 30 meters up until 240 meters. They found little evidence that cell size has an influence on the prediction accuracy for both residential burglary and assaults with injury. However, different bandwidth sizes starting from 100 meters to 800 meters in 100 meter increments have shown that the smaller bandwidths predict crime more accurately. The researchers however note that small bandwidths cause too many hotspot areas and are not practical for identifying only a few areas that need strategic attention. The proposed solution is to start with a large bandwidth to identify only a few areas and use smaller bandwidths inside those areas if more detail is needed.

The kernel density estimation function spatially distributes crime density over an area and assigns each grid cell a value that can be represented as a colour on a map. Out of the commonly known hotspot mapping techniques, the kernel density method offers a visually aesthetic representation of crime distribution making it a popular choice

among researchers (Chainey, 2013). One concern with KDE however is choosing which cell values are considered as high and should therefore be represented as hotspots. High value cells are normally given shades of deep red, drawing attention to those areas which typically are surrounded with less intense colours scaling down from orange to green to blue. However, deciding which cells represent red and which should represent orange is important because the colouring is what draws the attention of a reader to a specific area and therefore ultimately indicates what is considered to be a hotspot. Some of the commonly known methods of dividing the cell values into ranges are by dividing values into classes of equal number of features, equal intervals, standard deviation or nested averages. Eck et al. (2005) however suggests the statistical mean method which is similar to the nested averages and breaks the cell values into a set number of classes by recalculating the mean after each new class was created. In other words, all values lower than the initial mean will be placed into a class while the values larger than the mean will be seen as a new dataset on which the process is repeated. The researchers found it best to assign colours by using six classes in the statistical mean method, ranging from 0 to mean 1, mean 1 to mean 2 and so on. The suggested method is also best applied to values greater than 0 and only to grid cells that fall in the study area.

3.3 Risk Terrain Modelling (RTM)

Another commonly employed GIS technique used to generate a surface of crime risk (hotspot) is risk terrain modelling (RTM). "Risk terrain modelling (RTM) is an approach to spatial risk assessment that utilizes a geographic information system (GIS) to attribute qualities of the real world to places on a digitized map" (Caplan and Kennedy, 2011, p7). The technique makes use of separate coverage maps over the same geographical area each representing some aspect of the socio-demographic and built environment, overlaid together to form a composite risk map. Risk factors can be anything from facilities to population information. A composite map is then produced by adding all the terrain maps together, resulting in a risk value for each raster cell where the higher the risk value the higher the probability of crime occurring in that cell. Therefore, RTM maps are not considered as crime hotspots per se, but the factors included in the model rather indicate crime vulnerability of the area.

Kennedy et al. (2011) explain that the crime association each risk factor has on the landscape is a key aspect of risk terrain modelling. Crime associations are measured by observing the environment in which previous crime events occurred, identifying factors and their relationship towards crime. It is therefore essential to not only understand the influence crime attracting and crime generating facilities have on their immediate location but on their neighbourhoods as well. The authors use the example of bars and the location of shootings. The notion is that although shootings do not necessarily take place inside a bar, but rather in its surroundings, the bar still has a link to the shootings such as the kind of people it attracts and the association with alcohol. Of course, crime incidents are needed to understand the crime risk associated with different facilities but as these associations become more familiar, the risk value of facilities can be applied to find potential crime hotspots or used on areas where crime data are not available.

RTM maps are usually modelled over a continuous surface, rather than using political or statistical boundaries, hence the common practise of using raster cells as representation of the terrain. According to Caplan et al. (2011) vector points, lines and polygons are not sufficient representation of crime risk at a specific point as it has no relationship over the terrain it is part of. RTM maps can therefore forecast future crime events based on existing environmental conditions and do not rely on traditional past crime occurrences to predict future crime occurrences.

Caplan et al. (2011) estimated the risks of future shootings in Irvington, New Jersey, USA by fitting a risk terrain model (RTM) over the study area. The researchers were interested in testing the predictive significance of the map by comparing the RTM to traditional past crime hotspot maps (an aim mimicked in my research). The risk terrain models were created in a GIS by combining multiple spatial layers thought to have an impact on shootings in Irvington into a composite risk terrain map. The three main variables used as predictors were the dwellings of known gang members, the locations of retail business infrastructures and the locations of drug arrests. These variables were geocoded before being individually converted to a density map per variable. The researchers went on to classify the three density maps into four categories ranging from low to high risk per variable per map. The shooting data itself was divided into three time periods of six months each. The reason for doing this was

twofold. First, to test the predictive power of the risk terrain model by comparing the predicted shootings from period one and period two with their succeeding periods and second, to create two traditional maps to test the hypothesis that risk terrain models have a stronger predictive ability than traditional hotspot maps. Comparing the top 10 percent of high-risk cells, it was found that risk terrain models out-performed traditional hotspot maps by correctly predicting 42 percent of shootings, while traditional hotspot maps only predicted 21 percent. The researchers further found that overall risk terrain models predicted almost 21 percent more shootings than traditional hotspot maps. Although the current study does not use RTM, the notion of using facilities to create a prediction map is similar to the methods I will follow to generate a robbery prediction model. Similar to Caplan et al. (2011) I will also be comparing my robbery model to a traditional hotspot map.

3.4 Software

A number of GIS software options were considered for use in this study. The advantage of using open source GIS software such as QGIS is that the users are usually guided through a few steps using a user interface. In most cases, the output is displayed on the canvas of the GIS making it easy to view, evaluate and apply changes as required. However, in the current study it was important to overlay multiple hotspot maps and in order to do so it was important to ensure the same grid size and extent was used throughout. The programming language R generally provides more control over the dimensions used in hotspot creation and eliminates the tedious process of running the same process for multiple layers. Furthermore, R has extensive spatial and statistical libraries, easing the integration process allowing the user to run R scripts, without switching back and forth between different software packages. Therefore, the language R was used to build up scripts which run a series of statistical and spatial functions needed in this study.

3.5 Summary

In this chapter two different crime analysis techniques were discussed namely hotspot analysis and more specific, kernel density estimation hotspots and RTM.

In this study kernel density estimation is used to create traditional hotspot maps using historical 2006 robbery incidences. The KDE technique is also employed to generate my robbery probability model informed by previous spatio-temporal crime theories. The robbery probability model builds on the idea of the RTM model which seeks to explain crime based on the physical environment of an area, rather than solely on historical crime incidents. In the current study, the physical environment is examined using facilities and commuter nodes. Using the kernel density estimation and rasterising techniques, the current study builds a similar model to that of the RTM by combining features found in the physical environment into a risk map, called the robbery probability model.

Implementing the spatial techniques discussed above generally requires licensed software. The current study however uses R which is freely available. R was also chosen based on its statistical and spatial capabilities. The current study also presents a novel approach in which density values are predicted for each grid cell in a raster using regression analyses, resulting in a raster containing predicted robbery density values which can be displayed as a hotspot map. This is different from the RTM approach which stacks raster layers on top of each other to form a composite risk map.

CHAPTER 4: THE STUDY SETTING

The study site selected for this research is the City of Tshwane Metropolitan Municipality (CTMM). The CTMM is 6 368 km^2 in size and is part of the northern part of the central Gauteng province of South Africa.

According to StatisticsSA (2013) the City of Tshwane is the largest municipality in South-Africa and has seven regions consisting of 105 wards. Tshwane's active economy, which contributes 9.4 percent of South Africa's GDP and 26.8 percent of Gauteng's, makes it an attractive municipality to live in. The municipality houses roughly 2.9 million residents making it the 5th largest metropolitan municipality in South Africa by population size. The vast majority of residents in Tshwane are Black African 75.4 percent, while the White population is estimated at 21 percent, Coloured population at 2 percent and Asian residents are 1.8 percent of the population. Tshwane is classified as a young city as the youth (aged 10 to 30) form 37 percent of the total population.

There are roughly 46 percent of households in Tshwane that earns less than R76 401 per annum, with the average household income estimated at R60 642 per annum. (StatisticsSA, 2013).

Figure 1: Locality map of the City of Tshwane Metropolitan Municipality, South Africa.
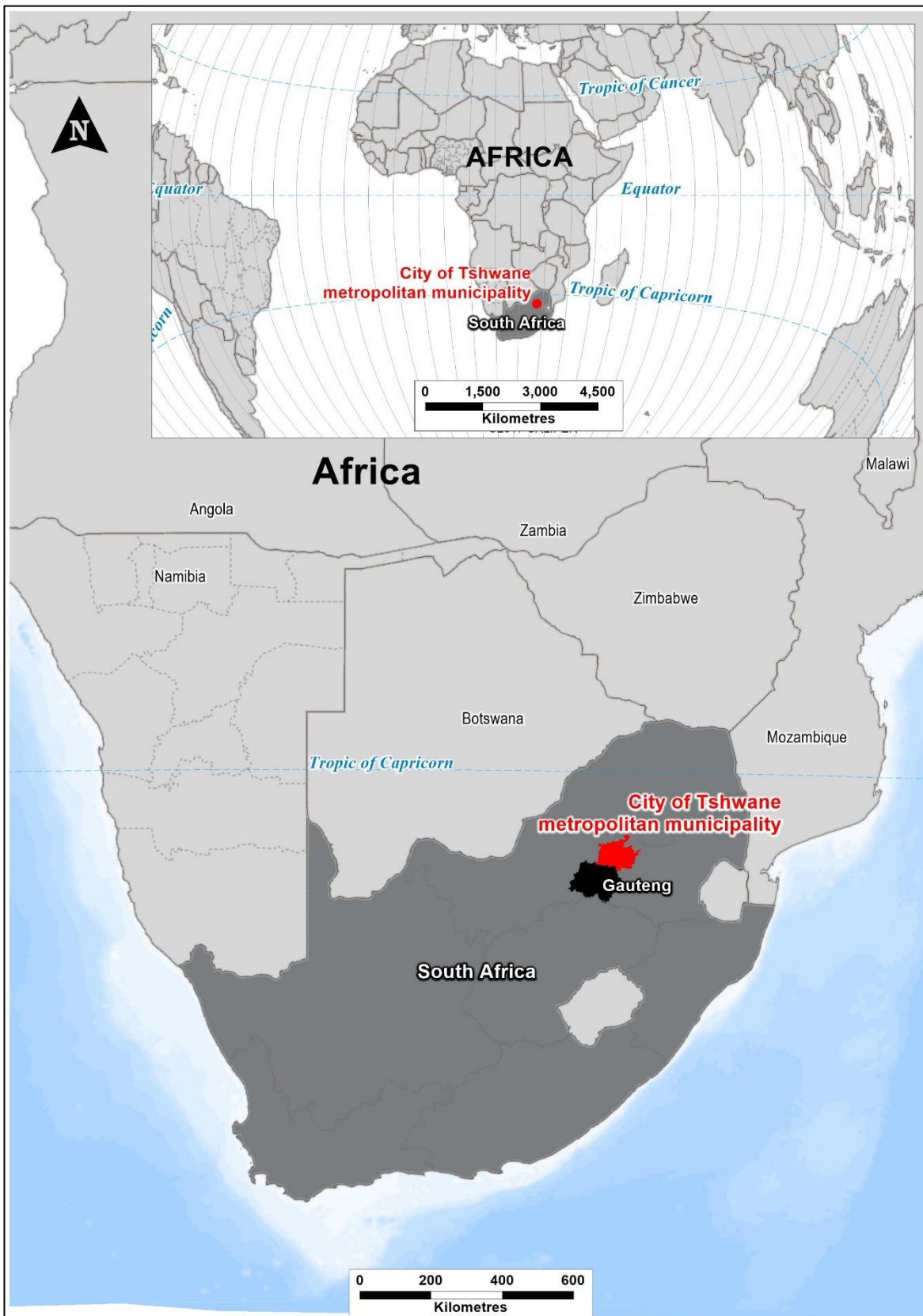
There are roughly 455 census sub-places in Tshwane. A sub-place is the smallest spatial boundary level released by Statistics South Africa that contains census information. Figure 2 shows the spatial boundary of Tshwane as well as the boundaries of the 455 sub-places contained within the city.

Figure 2: Tshwane and the 2001 sub-place boundaries

Comparing Borders: Tshwane and 2001 Subplace

KwaMhlanga NU

Bronkhorstspruit

Ekangala

City of Tshwane Rural

Cullinan

Refilwe

Rayton

Roodeplaat

Pretoria

Temba Hammanskraal

Soshanguve

Akasia

Winterveldt

Mabopane

Centurion

Ga-Rankuwa

Tshwane border
Subplace border

10   0   10   20   30   40 km
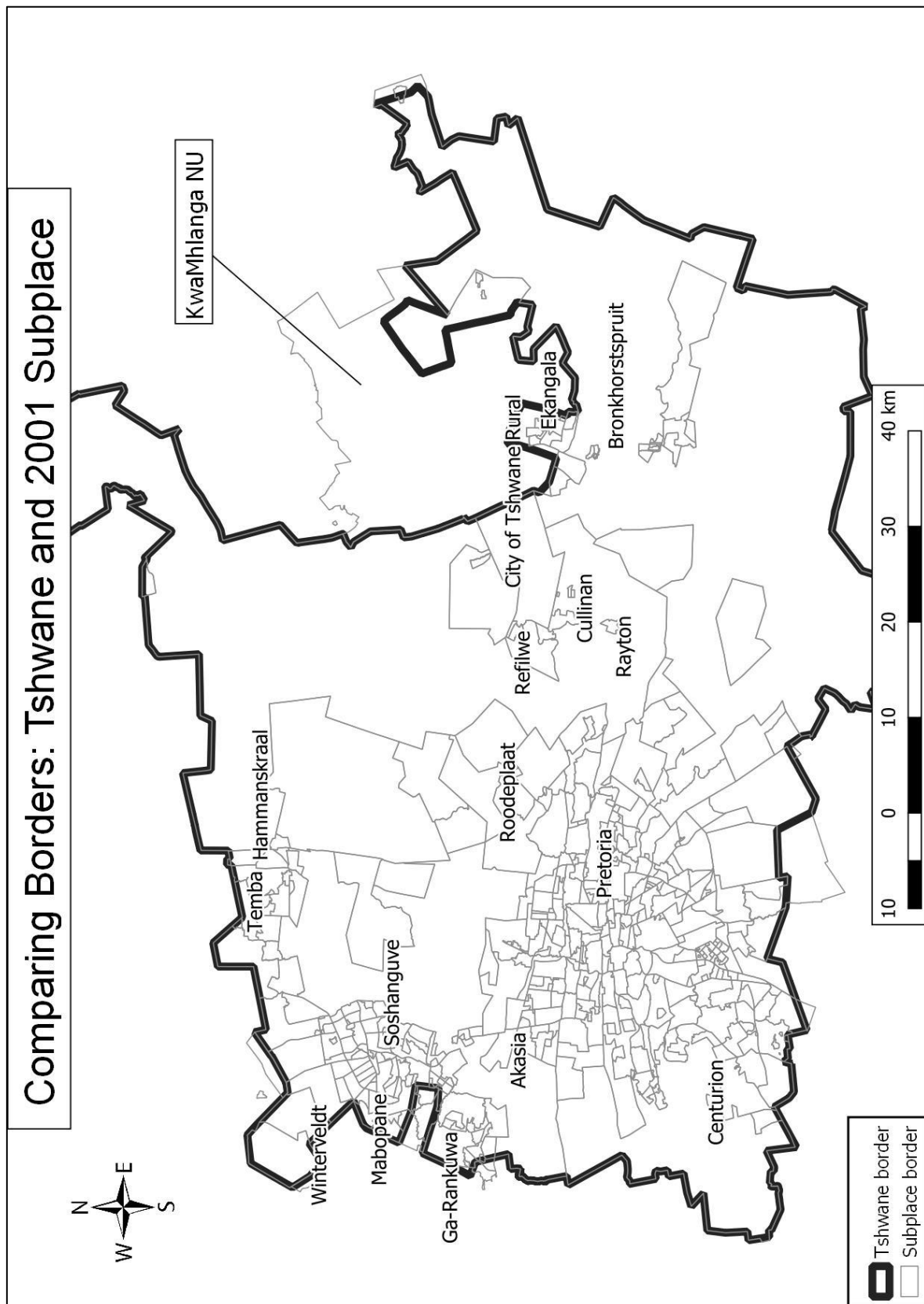
4.1 Crime in Tshwane

Table 4 displays the crime counts reported to the SAPS for 2001 to 2006 for the City of Tshwane Metropolitan Municipality. These years were selected as they roughly correspond to the crime data that was available for the study. Robbery was used in this study and therefore the two main subcategories of robbery, namely common robbery and robbery aggravating, are included in the table. The counts include attempted robberies in both categories while robbery with fire-arms and robbery with weapon other than firearm was included under robbery aggravating. Other crime includes all other crime committed in Tshwane aggregated.

Table 4: Tshwane crime counts for 2001 to 2006

Robbery	2001	2002	2003	2004	2005	2006	2005 vs 2006	Total
Common	5 029	8 902	9 021	7 989	6 832	7 131	299	**44 904**
Aggravating	3 872	7 783	10 483	9 705	8 459	8 581	122	**48 883**
Other Crime	122 102	166 322	176 493	164 832	155 614	149 629	-5 985	**934 992**
Grand Total	**131 003**	**183 007**	**195 997**	**182 526**	**170 905**	**165 341**	**-5 564**	**1 028 779**

Two items in Table 4 warrant attention. First, all the crime categories presented peeked during 2003, recording the highest crime occurrences after which crime occurrences steadily decreased during the next three years. Second, 2003 was also the year in which robbery aggravating surpassed common robbery incidents and remained the highest robbery category for the years presented. Robbery aggravating incidents are usually more violent than common robbery, which is of concern, as the counts indicate that crime in Tshwane became increasingly more violent. Furthermore, although crime overall decreased from 2005 to 2006, both robbery categories had a slight increase of 299 common robberies and 122 robbery aggravating reported incidents in Tshwane.

Figure 3 shows the reported robberies per police station in Tshwane, with the names of the police stations provided in the x-axis and accompanied with robbery counts in y-axis per year (z-axis). The SAPS (2005, p. 36) stated that "The geographical

demarcation of a police station area into manageable sectors, taking into account Crime Administration Blocks, the geographical size of areas, topographical features, community resources, crime types and patterns." is one of the pillars needed by sector policing to normalise crime. These geographical demarcated areas are known as the police station areas.

Figure 3: Robberies reported per South African police station in Tshwane
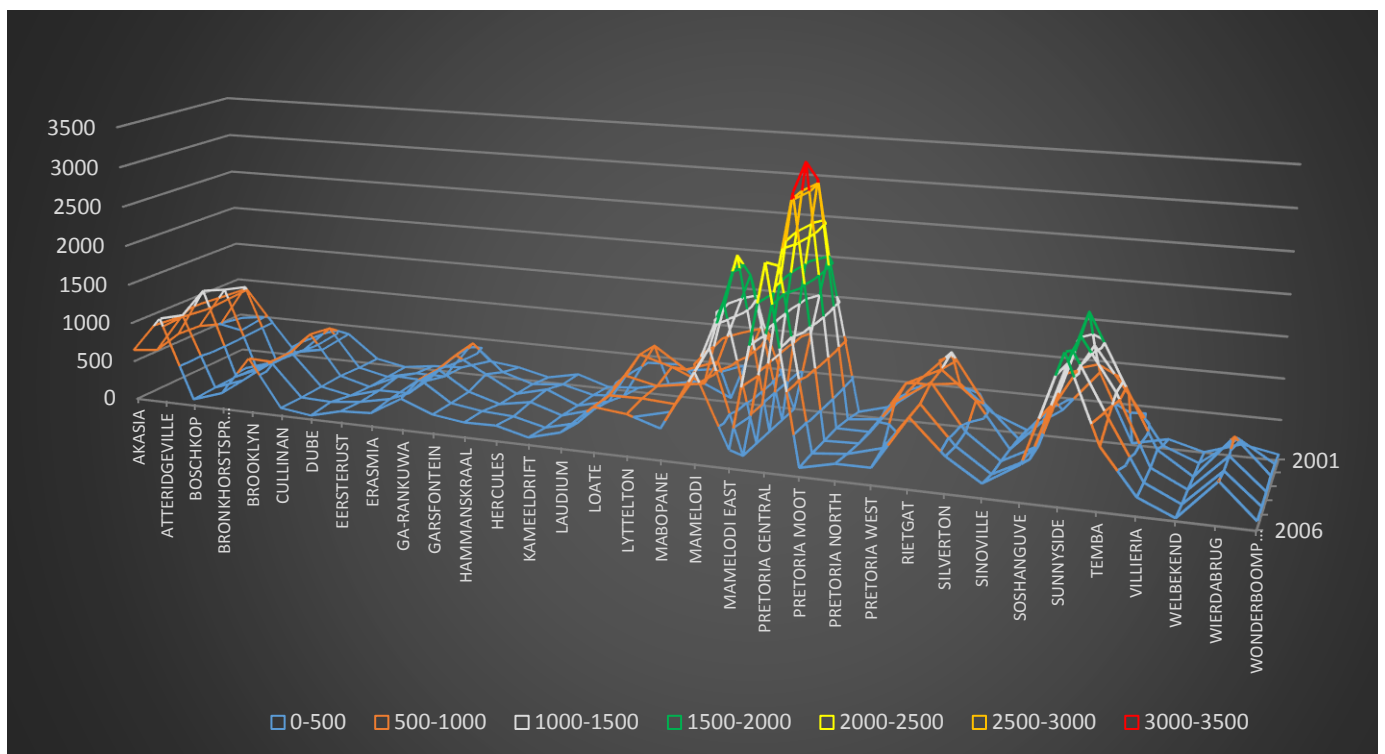
From Figure 3 it is evident that robberies are not uniformly spread out over Tshwane between 2001 and 2006. Pretoria Central is shown to be the most affected by robberies followed by Mamelodi and Sunnyside. Furthermore, Akasia, Atteridgeville, Loate and Rietgat police stations also stand out from other police stations in terms of reported robberies. There were also noticeably more robberies at Brooklyn, Ga-Rankuwa, Lytelton, Mabopane, Silverton, Soshanguve, Temba and Wierdabrug. The current study aims to investigate robberies on a micro-level that is much smaller than a police precinct; even so, the latter identified areas are expected to be prominent places in which robbery hotspots will be identified.

CHAPTER 5: DATA AND METHODS

5.1 Introduction

In this chapter the reader will be guided through a series of steps that were followed to produce a model to predict robbery hotspots in the City of Tshwane Metropolitan Municipality. In the first step the data sources, and selection, manipulation and transformation methods used are discussed. The second step focuses on the creation of a traditional hotspot model using historical robbery locations from 2006. The traditional hotspot model is then validated using robbery locations of 2007. The third step explains how a robbery probability hotspot model was created using census (demographic) data, facilities and commuter nodes. In the final step, the traditional hotspot model (step 2) was compared against the developed robbery probability hotspot model (step 3) to determine which method best predicts robberies for the City of Tshwane Metropolitan Municipality. Importantly, the analysis described above was done for both day- and night-time, Daytime refers to (07:00am to 19:00pm) and night-time to (19:00pm to 07:00am). The methodology followed for day and night-time were exactly the same and therefore only explained as a singular model in the steps below.

5.2 Data

In this section the data sources that were used in the study are listed along with the selection, manipulation and transformation methods used. The data was separated into four categories namely crime, census, facility and commuter nodes.
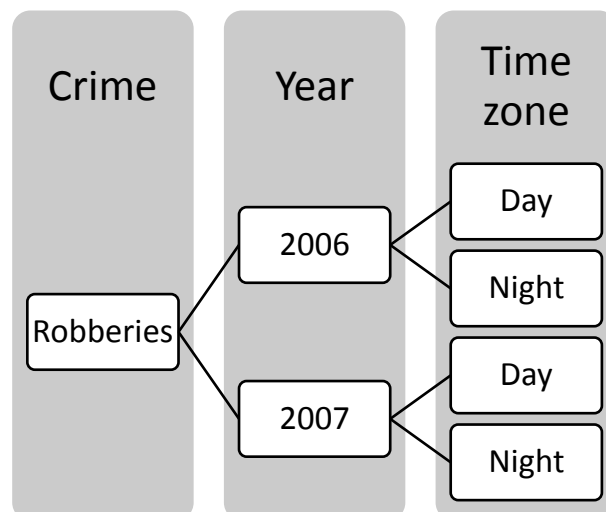
5.2.1 Crime data

The crime data used for the study was obtained from the South African Police Services. The crime data contained GPS (global positioning system) located crime events in Tshwane, South Africa from 2001 to 2007. The crime locations were accompanied with attributes including the longitude, latitude, committed time and the offence classification.

Regarding the offence classification, there were a total of 47 different types of crime. The current study focussed on five subcategories of robberies namely robbery aggravating, attempted robbery aggravating with firearm, common robbery, attempted common robbery and robbery with weapon other than firearm.

Robberies committed in 2006 were used to build the traditional hotspot model and 2007 robberies were used to validate the traditional hotspot model. Furthermore, the robbery data were then split into daytime (7:00am to 19:00pm) and night time (19:00pm to 7:00am) as there could be different predictors and spatial patterns for robbery depending on the time of day and night. Rummens et al. (2017) found that splitting crime into daytime (7:00am to 19:00pm) and night time (19:00pm to 7:00am) had a considerable increase in crime prediction accuracy while Breetzke (2016) found that violent crime peaks between 8pm and 7am. Figure 4 displays the robbery selection flow.

Figure 4: Crime selection and split process

5.2.2 Census data

Statistics South Africa is responsible for conducting the census in South Africa. The last three population censuses were held in 1996, 2001 and 2011. Since the crime data was collected from 2001 to 2007, it was decided to use the data collected during the 2001 census. The smallest geographical boundaries created by Statistics South Africa is called the Enumerator Areas (EA's) which vary in size and are considered small enough to administer a questionnaire for each household in each EA. The EAs, however, were deemed to be a privacy risk and so Statistics South Africa decided to use sub-places (SP) as the smallest level to convey census data to the public. Sub-

places are roughly the size equivalent of suburbs and which is still small enough to approximate a 'neighbourhood'.

A vast amount of census data is collected for each SP relating to person and household demographics. This included the count of people, households and their associated data such as age, citizenship, education, employment, gender, marital status, migration and means of transport to work.

According to the social disorganisation theory there are four main factors that increase the risk of crime in communities; these include deprivation, ethnic heterogeneity, residential mobility and family disruption. In the current study the decision was made to select the following census demographic variables that most closely align with the central tenets of the social disorganisation theory: unemployed population (social deprivation), people moved in five years (residential mobility), divorcees (family disruption), male population between the age of 15 and 34 (young population), African foreign born population (ethnic heterogeneity) and people living in deprivation. The latter variable was used based on Noble and Wright (2013) who itemised the deprivation variables as households without a pit latrine with ventilation or flush toilet, households without use of electricity for lighting, households without piped water within at least 200 meters and shack as dwelling type.

The selection of these variables to represent social disorganisation theory also mimics the results of previous work in Tshwane that has tested the theory. For example, Breetzke (2010a) found unemployment, deprivation and residential mobility as significant predictors of crime in the city, while female headed households, father dead or estranged, black population and people renting houses were found to be insignificant predictors. Furthermore, in a socio-structural analysis of crime Tshwane, Breetzke (2010b) noted that disaffected youth measured by population of males older than 15, divorce or separated and African foreign born had a positive relationship with contact crime in Tshwane.

The current study uses these identified social disorganisation variables to highlight neighbourhoods which according to the literature are more prone to crime than others. Therefore, the social structure of neighbourhoods aids the robbery probability models on a high level by identifying boundaries in which robberies are more likely to occur.

5.2.3 Facility data

Facilities may increase or decrease the risk of crime in their vicinity (Eck and Weisburd, 2015). In the current study a number of facilities were chosen based on crime pattern theory and informed by the RTM work of Caplan and Kennedy (2011). According to the latter proximity to public transport, pubs, bars exotic clubs, schools, banks/cash points, post offices and leisure and fast food outlets increase the risk of crime. Previous research using crime pattern theory by Hiropoulos and Porter (2014) found positive associations between shopping centres, major roads, retail and industrial nodes and car theft in Gauteng. Of course, car theft and robberies may have different underlying spatial dynamics; however, the latter facilities can also be seen as major nodes and crime generators or attractors. The current study uses similar facilities to investigate their relationship with robberies. The final chosen eleven types of facilities used in the study were obtained from AfriGIS and are shown in Table 5. These facilities are used in the robbery probability model, where their relationships with robberies are evaluated and mapped as predictive robbery hotspots.

Table 5: List of facilities used in the study per sub-place (n = 455)

Facility	Count	Min	Max	Mean	St. dev
Parks	1 748	0	193	5.81	14.21
Clothing store	143	0	23	0.48	1.91
Convenience store	84	0	4	0.28	0.71
Education	365	0	28	1.21	2.52
Fast food	123	0	19	0.41	1.39
Filling station	260	0	18	0.86	1.70
Office park/block	85	0	9	0.28	1.03
Other stores	138	0	21	0.46	1.66
Restaurant	244	0	33	0.81	2.92
Shopping centre	163	0	7	0.54	1.14
Supermarket	147	0	9	0.49	1.03

Figure 5: Facility locations used in the current study
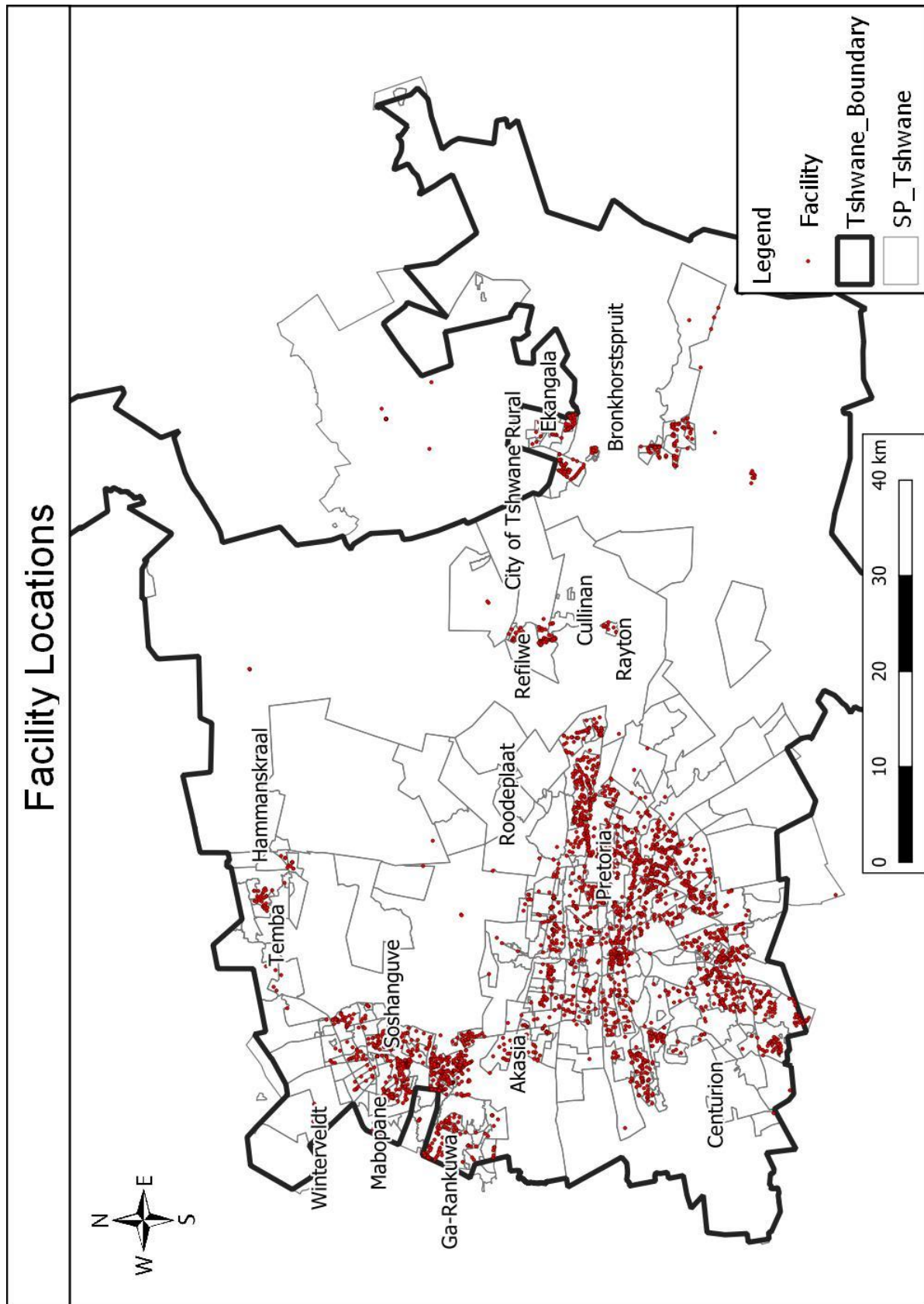
A total of 3 296 facilities were used in this study (see Figure 5). They capture a range of land-use types from educational, recreational, entertainment and stores. All the facilities selected were sourced from the historical database of AfriGIS (Pty) Limited. The database contains spatial information which mainly includes the formats shapefile (.shp) and Maptitude's (.dbd) spatial layers for each facility. The facility data were extracted for use in this study using queries in Maptitude. A typical attribute query takes a column name, mathematical operator and value i.e. the query to select shopping centres form the points of interest were entered as TYPE_2 = "SHOPPING CENTRE", where TYPE_2 is the column name containing shopping centres as a subcategory.

A number of improvements on facility data held by AfriGIS have been made since 2006. In 2006 the facility data was accompanied with an accuracy level which was later renamed to a confidence level. The confidence level indicated how accurate the point on the map is relative to its location in the real world. The confidence levels are as follow:

Table 6: AfriGIS confidence levels technical description

Confidence level	Technical Description
1	Accurate Erf-Portion Level
2	Erf Level
3	Street Corner
4	Within 5 of Street Number
5	Street Name and Suburb
6	SG town
7	Suburb
8	Town

Each facility also has a unique AfriGIS ID which is kept the same throughout the years, and therefore it was possible to link facilities in 2006 to the latest data release (May 2017) and update the confidence level, longitude and latitude accordingly. Furthermore, the current study aims to predict crime at a fine level of aggregation and therefore only points with a confidence level lower than five were selected for inclusion

in the study, which means the facility should fall at least within five street numbers of the actual street number.

Figure 6 shows an example of improvements made on the points of interest accuracy from 2006 to 2017. In the example the blue pins indicate the location of two filling stations as captured in 2005. From the satellite imagery it is clear that the two filling stations were wrongfully captured but were relocated during 2006 to 2017 to the precise location, indicated with the red arrows. Improvements were done, where necessary on all the 3 296 facilities used in the study.

**Figure 6:** Improving the confidence level of a facility

5.2.4 Commuter node data

Traveling nodes such as bus stops are nodes people visit while traveling between their daily activities. The reason for including these commuter data in the study was threefold. First, according to the crime pattern theory, offenders become aware of potential targets while commuting between points. Therefore, the commuting points indicate the potential of that area being visited, increasing the risk of a potential offender noticing crime opportunities. Second, public transport nodes in particular have been found to be key street robbery risk factors (see (Newton, 2014, Hart and Miethe, 2014, Block and Davis, 1996, Clarke, 1996)). The third reason was derived from the temporal constraint theory which suggests that potential criminals have little time to explore crime opportunity areas outside their traveling network. Therefore, the commuter nodes, as shown in Figure 7, were included in the analysis based on the notion that while on journey, potential offenders spend limited time outside these nodes, while time spend at these nodes increase the potential awareness of robbery opportunities.

The commuter node data was sourced from the same AfriGIS spatial database as the facility data. The main road and street nodes were extracted from the street central lines. The latter nodes indicate where the street central lines meet. This means that a point is placed wherever certain roads cross each other. The decision was made to only include the road categories 'highway' and 'main road' nodes, excluding streets, arterial and secondary roads, based on the notion that not much time is spent on the latter nodes, nor is it as congested as main roads or highways. The last set of nodes, railway stations, was extracted from the same points of interest layer as where the facilities were sourced. There were however limitations on the data available and therefore could not source taxi ranks and bus stops as commuter nodes which would have made the selected data more complete.

Figure 7: Commuter nodes
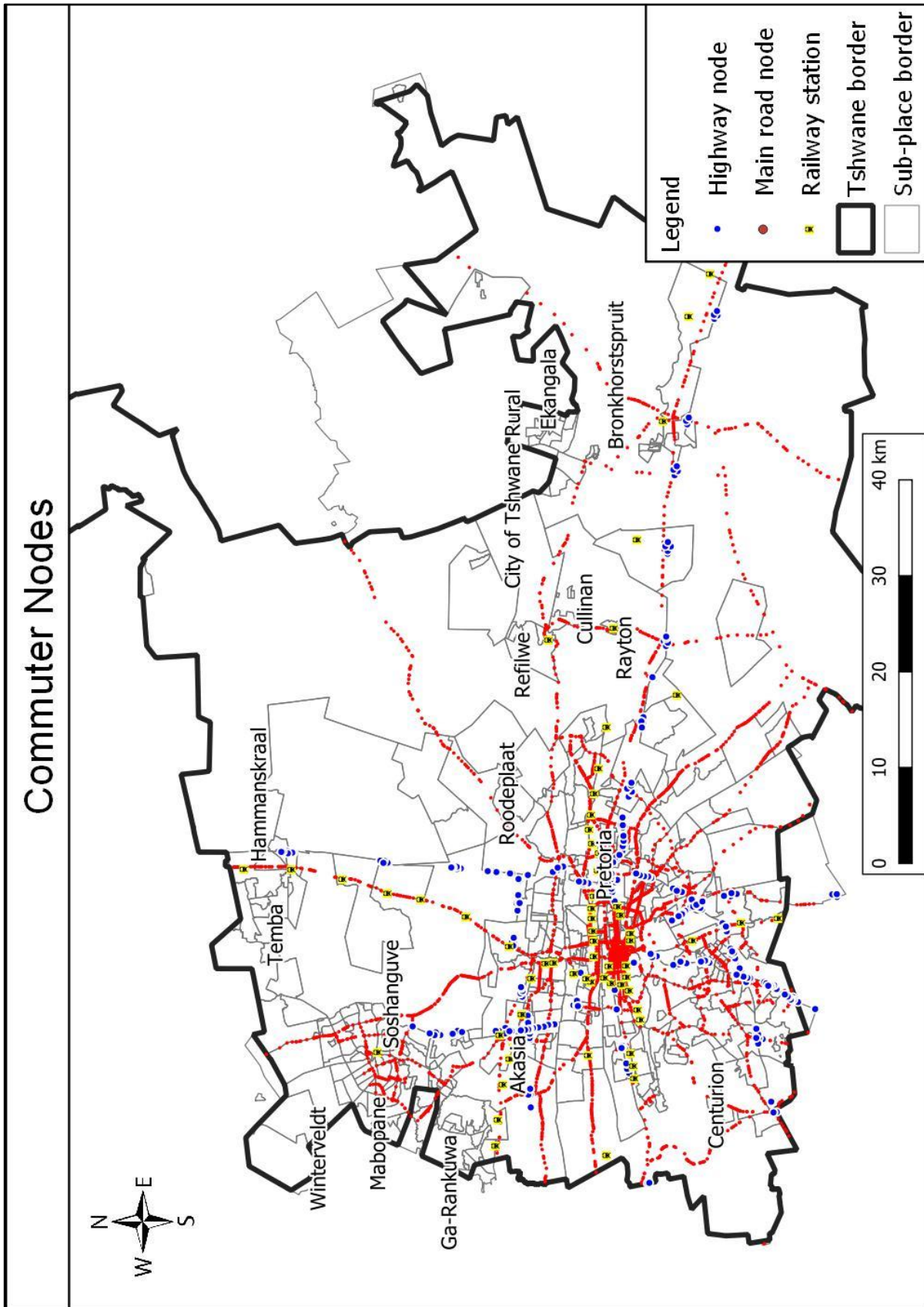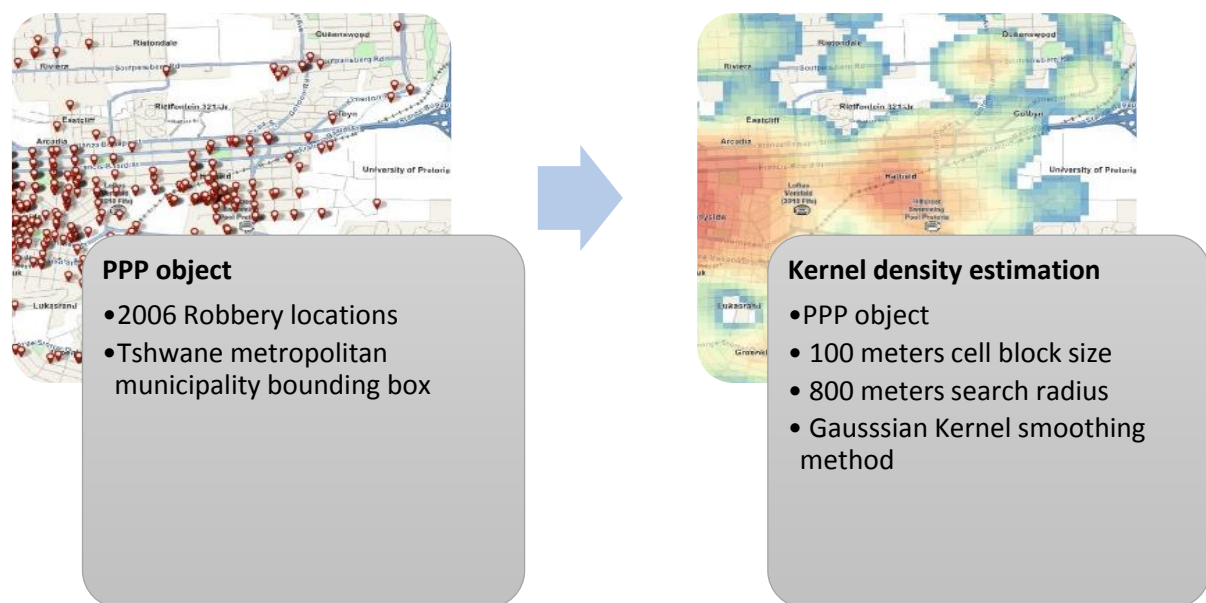
5.3   Analysis

5.3.1   Traditional Hotspot Model

The first step in the analysis was to create a traditional hotspot model for robbery in Tshwane for 2006. This sub-section therefore guides the reader through steps taken to create the 2006 robbery hotspot map accompanied with R code snippets.

The advantage of the traditional hotspot model lies in its design simplicity. The only data that is required in the model is robbery locations, transformed into kernel density estimation maps. In order to create the hotspot maps, robbery locations were extracted from the crime data for the year 2006. Thereafter, the data was further filtered to create two separate robbery time periods: 7am to 19pm represents day-time and 19pm to 7am night-time.

Figure 8 explains the process that was followed to create the crime kernel density estimation map for robbery hotspots for 2006. The same method was used for both the day and night analysis.

Figure 8: Crime kernel density estimation process

PPP object
- 2006 Robbery locations
- Tshwane metropolitan municipality bounding box

Kernel density estimation
- PPP object
- 100 meters cell block size
- 800 meters search radius
- Gausssian Kernel smoothing method

Coordinates to ppp.object:

The first step undertaken to create the 2006 robbery hotspot maps was to convert the 2006 robbery locations into a ppp.object. In the R code the 2006 robbery table is represented by xy where xy$x and xy$y are the longitude and latitudes respectively. Tshwanebbox represents the City of Tshwane Metropolitan Municipality's bounding box, stored in a matrix. The bounding box were included to extract 2006 robberies falling within the latter bounding box, before running the kernel density estimation on the extracted points.

```
###--------------------- Coordinates to ppp.object ---------------------###

Tshwane_ppp.object <- ppp(xy$y,xy$x, Tshwanebbox[c(1,3)],Tshwanebbox[c(2,4)])
```

Density function:

The ppp.object along with a grid size of 100 meters, sigma value of 200. Notably, the sigma value is not in meters but rather a numeric value or function assigned as the

 standard deviation used in the smoothing kernel. Simply put, the sigma value does not represent metres; therefore in the current study the sigma value of 200 represents approximately 800 meters.

```
###--------------------- Kernel density estimation (Spatstat) ---------------------###

li_density.raster = raster::raster(density(gauteng_points_projected,eps=100,sigma=200,kernel="gaussian")
```

The raster layer created from the steps above is illustrated in Figure 9. The kernel density values were divided into 32 classifications using the nested average method. The code snippet below shows how the data were classified and assigned to a class. The latter 32 classes were then assigned to a colour ranging from blue to red as displayed in the map.

```
################# Create classes based on the nested average method ###################

breaks <- cartography::getBreaks(v = li_density.raster[], method = "em", nclass = 32)

robbery_prob_raster[] <- cut(li_density.raster[], breaks=breaks, labels = c(1:32))
```

Figure 9: Traditional model – 2006 traditional hotspots
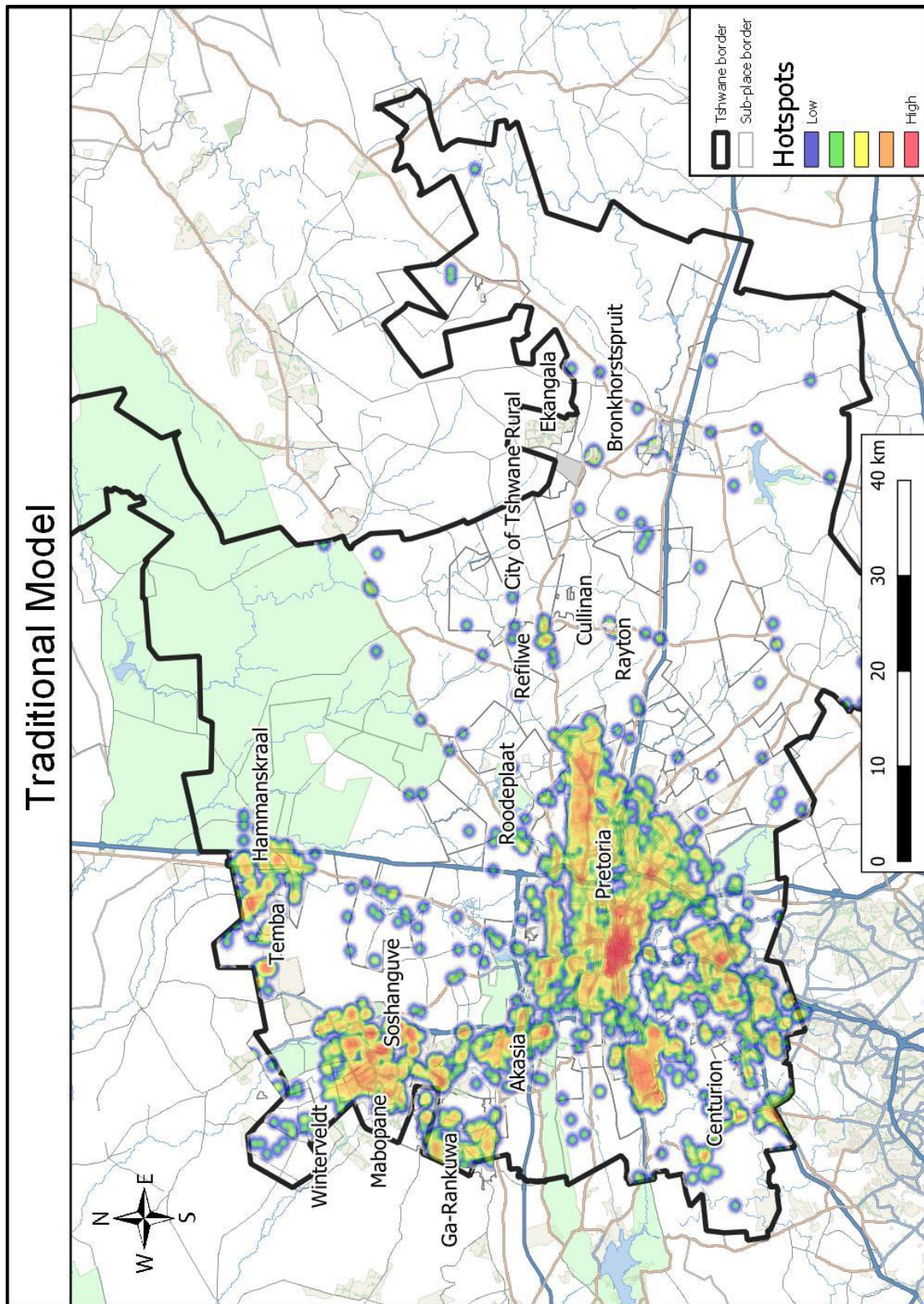
5.3.2 Robbery probability model process

The second step in the analysis was the creation of a robbery prediction model. The model was constructed using the following steps: First, a facility model was constructed by examining the relationship between 2006 robbery incidents and facilities in Tshwane. The process involved a series of data modelling techniques used to predict robbery probability around facilities. Second, the social disorganisation index was constructed per sub-place in Tshwane using 2001 census data. Third, commuter data was used to determine the probability of robberies around commuter nodes. Finally, the latter three models were combined into a robbery probability model.

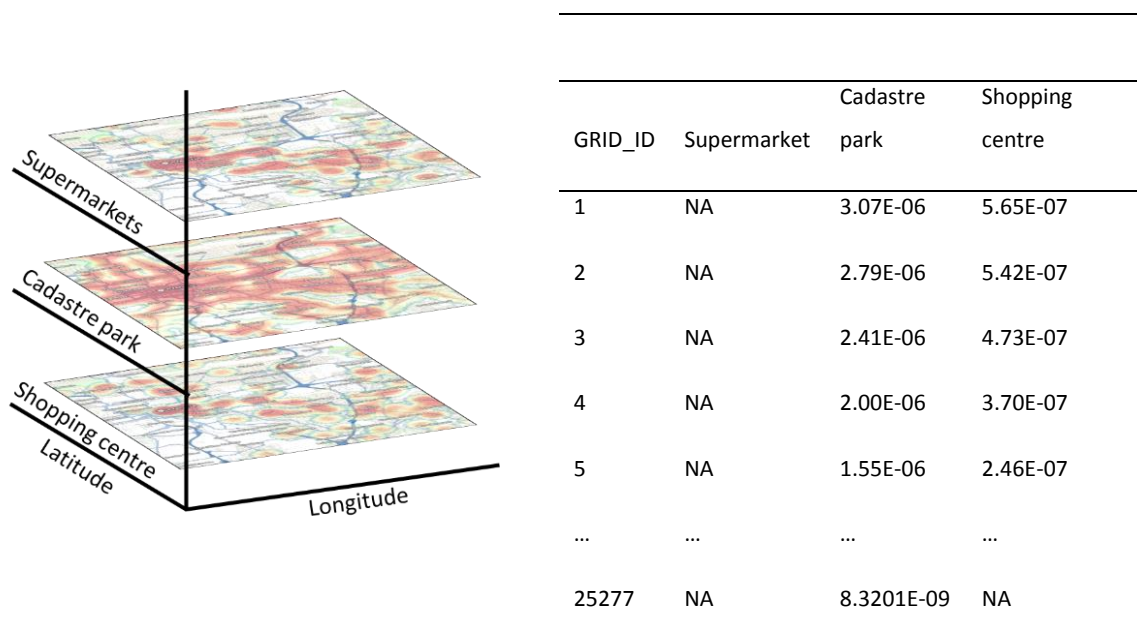**Figure 10:** Robbery probability model methodology
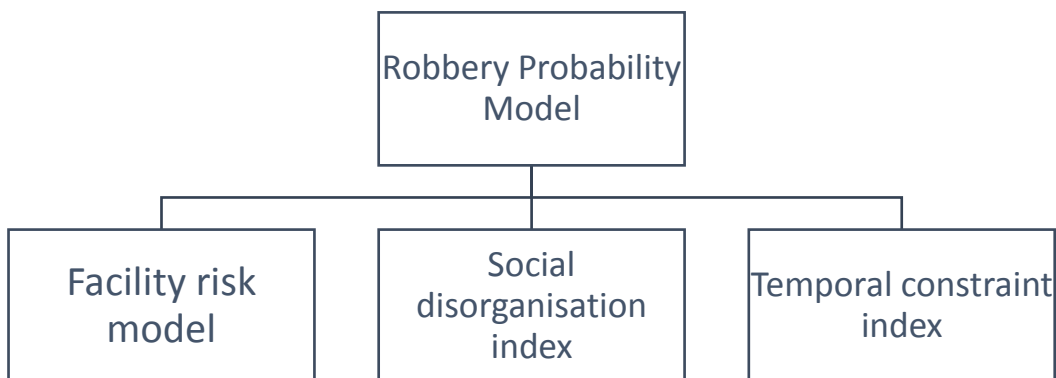
5.3.3 Facility risk model

The facility risk model builds on the notion that some facilities have a spatio-temporal influence on crime patterns. The facility risk model aims to find relationships between crime and facilities. Once the relationship has been established the model can then be applied in areas where crime locations are unknown but have similar facilities.

The first step in the creation of the facility model was to create a hotspot map of facilities. This was done by following the exact same method used to create the 2006 robbery hotspot map. A hotspot map was created for each of the eleven types of

facilities used in the study (see Table 5) and then stacked together to form a composite map for all the facilities together. Figure 11 shows an example of the raster stacking process undertaken to get the stacked hotspot map. In this example supermarkets, cadastre parks and shopping centre hotspot maps were stacked together. The example shows the individual tables of each raster layer and the result of being stacked. The "GRID ID" field indicates the cell block number, therefore each of the tables show the first five cell block's density values as well as the last one. By looking at the GRID ID numbers from these tables, it is evident that the raster layers have the same amount of rows, (n = 25 277). The composite grid can be seen as a table containing columns of each raster layer involved in the stacking process. Note that "NA" values were given to raster cells outside the bandwidth distance used to calculate the kernel density values for each facility.

Figure 11: Facility raster stacking process

GRID_ID	Supermarket	Cadastre park	Shopping centre
1	NA	3.07E-06	5.65E-07
2	NA	2.79E-06	5.42E-07
3	NA	2.41E-06	4.73E-07
4	NA	2.00E-06	3.70E-07
5	NA	1.55E-06	2.46E-07
...	...	...	...
25277	NA	8.3201E-09	NA

Stacking is a tedious and difficult process. However, if the raster layers have the same dimensions, stacking is relatively simple to execute using R. The code snippet below reads all the .tif files in the working directory and stack them together.

```
###--------------------- Import raster files ---------------------###
hab = list.files(getwd(), pattern="tif$" ,full.names=FALSE)
###--------------------- Create a raster stack ---------------------###
r <- stack(hab)
```

After the stacking process was undertaken a regression analysis was undertaken in raster to determine how well the various facilities (independent variables) predict crime (dependent variable). Regression analysis models seek to explain the dependent variable in a regression model by observing the influence a set of independent, also known as predictor variables. The purpose of the regression model is therefore to train predictor variables to best fit the dependent variable so that those predictor variables can serve as the input in a prediction model.

The eleven stacked raster layers – each representing a facility hotspot map - were the independent variables in the regression model and the 2006 robbery hotspots created in section 5.3 was the dependent variable. The result of this step was a table with grid numbers as row names, 2006 robbery density as the dependent variable column and each facility as separate independent variable columns.

During the running of the model it became apparent that there were a number of missing values in the raster grids both for the independent and the dependent variables. In the current study, these missing values represents grid blocks which had no density values for each facility (i.e., there were no crime located in the grid cell locations, and/or no public facilities in those grid cell locations). In the analysis these grid cells were excluded in the analysis. Indeed, it was found that removing grid blocks where there were no robbery and facility density values had no significant impact on the regression model results.

In preparation for regression modelling, rows were split into a training and a testing dataset. The reason for doing this is to determine how the independent variables respond to known dependent variable values in the training dataset. The response is then applied to the unseen data of the testing dataset. The testing data also contains robbery density values that are used to validate the predicted outcome. In the current study a random split was applied, assigning 75 percent of cells to the training data

and 25 percent of cells to the testing dataset. The method is shown in the code snippet below where 'Prediction_Table' is the table as created above by stacking the raster files and 'dependent_v' the 2006 robbery density values.

```
###---------------------- Training and testing datasets ---------------------###
inTrain <- createDataPartition(y=Prediction_Table[,dependent_v],p=0.75,list=FALSE)
training <- Prediction_Table[inTrain,]
testing <- Prediction_Table[-inTrain,]
```

It is good practice to normalise the independent variables prior to inserting them in a regression as normalised variables are an assumption of regression. In the current study, the independent variables were normalised using range as method. The range method simply scales the values between zero and one. The following code snippet applies and sets the range method using the preProcess function from the caret library in R.

```
###---------------------- Normalise the independent variables ---------------------###
###----- Normalise independent variables using ranges in the training dataset
preObj<-preProcess(training[,independent_v],method="range")
NPP<-predict(preObj,training[,independent_v])
###----- Combine the dependent variable with the normalised independent variables
training<-cbind(training[,dependent_v],NPP)
###----- Normalise testing and full table's independent variables using the training data ranges
NPP_test<-predict(preObj,testing[,independent_v])
testing<-cbind(testing[,dependent_v],NPP_test)
NPP_full<-predict(preObj,Full_Table[,independent_v])
Full_Table_norm<-cbind(Full_Table[,dependent_v],NPP_full)
```

The statistical method of attempting to omit irrelevant predictors and only selecting the most significant predictor variables is known as feature selection. The current study examined two types of feature selection methods namely, filters and wrappers.

Filters and wrapper methods are used to calculate the weight each feature (variable) brings to the prediction model by analysing the influence each predictor variable has

on the dependent variable. The major difference is that filter methods analyse a predictor variable as a singular entity where wrapper methods validate the usefulness of a variable within a subset of variables. Guyon and Elisseeff (2003) outline these two methods by explaining that filters are best used as a pre-processing step by calculating the importance of each variable separately. Variables are then filtered by keeping only the variables found to be important and adding to the performance of the model. I.e. variable ranking is a filtering method which involves scoring variables from most to least important. Variables are then selected based on an importance level or a selection from the top ranking variables. Variable ranking is however criticised for the selection of redundant variables which result in similar variables being used in the prediction process without adding any value. A possible solution to this is to correlate each variable with one another, in doing so, identify redundant variables and remove the variable that rank the lowest between the two variables in question (i.e., avoid multicollinearity). Wrapper methods on the other hand assess subsets of variables to the prediction performance. Wrappers therefore differ from filters by incorporating interactions between variables.

The process of wrappers involves the use of search strategies to loop through possible variable subsets and therefore can be computational strenuous and in some cases can lead to overfitting. Overfitting is when the model tries to find a solution by fitting all or almost all predictor variables for the training data which in return leads to unreliable predictions when the model is introduced to new data. Overfitting can however be avoided by the so-called greedy strategies in the form of forward selection and backwards elimination. As the names suggests, the forward selection strategy starts with an empty variable set and sequentially adds one variable to the evaluated variable subset. The best performing predictor in the subset is added to the list of previous best performing variables. The list of selected variables grows until either a predefine number of features are selected or the model does not improve further. Backwards elimination on the other hand starts with all variables, removing the least important variables. Backwards elimination continues until either a predefined number of variables are left or the model decrease in accuracy.

In the current study, feature selection was used to eliminate facilities (variables) not adequately predicting the 2006 robberies. After testing numerous combinations, a

wrapper feature selection method with a back elimination search strategy was ultimately used to predict 2006 robberies (dependent variable) using the set of facility hotspot maps as predictors (independent variables). The code snippet below shows how the feature selection method was defined in R code, using the mlr package.

```
#--- Create a search control for the feature selection method
ctrl = makeFeatSelControlSequential(method = "sbs", beta = -0.001)
#--- resample the training data set using 5-fold cross validation
rdesc = makeResampleDesc("CV", iters = 5)
#--- Create a feature selection wrapper
Feature_selection_method = makeFeatSelWrapper("regr.lm", resampling = rdesc, control = ctrl, show.info = FALSE)
```

There are a number of steps to create the feature selection wrapper. First, the wrapper needs to know which type of statistical model will be used in the analysis. A linear regression model was chosen based on the structure of the underlying data. The second decision to be made was the method of resampling to be employed. Resampling is the method of splitting the training dataset into training and validating subsets. Molinaro et al. (2005) compared a few resampling methods by estimating the prediction error and found that the leave-one-out cross-validation (LOOCV), 5 and 10-fold cross-validation and the .632+ bootstrap resampling methods produced the smallest mean square errors on small samples.

As a result the current study used a 5-fold cross-validation, which means that the training dataset was randomly split into five sets of approximately the same size. Each set is the validated against the four remaining sets. The last step is to define the search strategy. The search strategy is created in the feature selection control. In the current study, feature selection control was created using a sequential backwards selection (sbs) search strategy with a beta of -0.001. Therefore, the selection control started with all the predictor variables and progressively deleted variables that performed the worst. This was done until the improvement in the model was less than -0.001. A negative value was used to allow the model to remove a variable even if there was a slight decrease in the model's improvement, allowing leniency with the aim to improve the model as it keeps on deleting variables. The beta value was however small enough

64

not to have a significant impact on the overall prediction accuracy. The wrapper then combines the feature selection control, resample method and linear regression model to create the feature selection learner.

The final step was to run the prediction model. Specifying a regression task in R is outlined in the code snippet below. In the code snippet, the regression task is created on the training dataset using the target as dependent variable. In the current study, the target was the 2006 robberies. The prediction model is then trained using the feature selection learner on the regression task.

```
#--- Create the regression task
crime.task <- makeRegrTask(data = training, target = target_name)
#--- Create a prediction model
mod <- train(lrn, task = crime.task)
```

The trained prediction model created above consists of a list of the selected predictor variables and because the model uses linear regression, a formula can be derived using the coefficients from each independent variable, as seen in Table 7.

Table 7: Multiple regression coefficients

Coefficients:	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	1054.99	39.12	26.97	***
Cadastre Park	8900.99	413.79	21.51	***
Clothing Store	128391.66	3091.43	41.53	***
Convenience Store	-69109.71	1012.13	-68.29	***
Educational Facility	30684.35	874.36	35.09	***
Fast Food Outlet	246696.62	2373.34	103.94	***
Filling Station	37131.86	796.72	46.61	***
Office Parks/Blocks	12248.05	1080.22	11.34	***
Other Stores	63499.4	3377.8	18.8	***
Restaurants	238303	1505.21	158.32	***
Shopping Centres	-63665.06	102.19	-62.41	***
Supermarkets	16015.69	1162.25	13.78	***

Signif. codes: High: 0 '***' Moderate: 0.001 '**' Low: 0.01 '*'

No significance : 0.05 '.' 0.1 ' ' 1

Residual standard error: 12890 on 143146 degrees of freedom
Multiple R-squared: 0.5905,   Adjusted R-squared: 0.5904
F-statistic: 1.876e+04 on 11 and 143146 DF, p-value: < 2.2e-16

The multiple linear regression formula is defined as:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i,p-1} + E_i$$

Where **β** is the coefficient variable, **x** the estimated value for the coefficient and **E_i** the random error component. Using this formula, a prediction value can be calculated for each grid cell. The predicted value describes the robbery risk that exists in a grid cell based on the presence of facilities in the near vicinity. Therefore in the current study, the predicted value is referred to as facility risk.

The performance of the prediction model created above was tested by applying the same model on the unseen testing dataset. In this step, robbery density values are

being predicted using the selected predictor variables without taking the dependent 2006 robbery variable into account. Tests were then performed by comparing the predicted density value to the actual 2006 robbery density value with the purpose to estimate how well the model performs on new data. The final step was to apply the same prediction model to the whole dataset containing grid cells for the whole of Tshwane. The following code snippet shows how the explained steps were implemented in R.

```
#--- Test the prediction model
test_predt<-predict(mod,newdata=testing)
#--- Test the performance of the prediction model
test_predict_performance <- performance(test_predt, measures = list(mse))
#--- Add the predicted values to the testing table
test_pred <- as.data.frame(test_predt)
testing$test_Output <- test_pred[,2]
#--- Predict robbery density estimation values for the whole study area
predt<-predict(mod,newdata=FUll_Table_norm)
pred <- as.data.frame(predt)
#--- Add the predicted values to the original table
FUll_Table_norm$Output <- pred[,2]
#--- Set all cells not containing any facilty data to NA
FUll_Table_norm$Output[rowSums(FUll_Table_norm[2:ncol(FUll_Table_norm)-1])==0] <- NA
```

Figure 12 shows an example of how the predicted values (facilities) predict robberies in 2006 (crime). The facility risk value was calculated using the multiple regression formula and are displayed from blue to red, signifying robbery risk intensity based on the presence of facilities in those areas.

Figure 12: Facility risk hotspot map
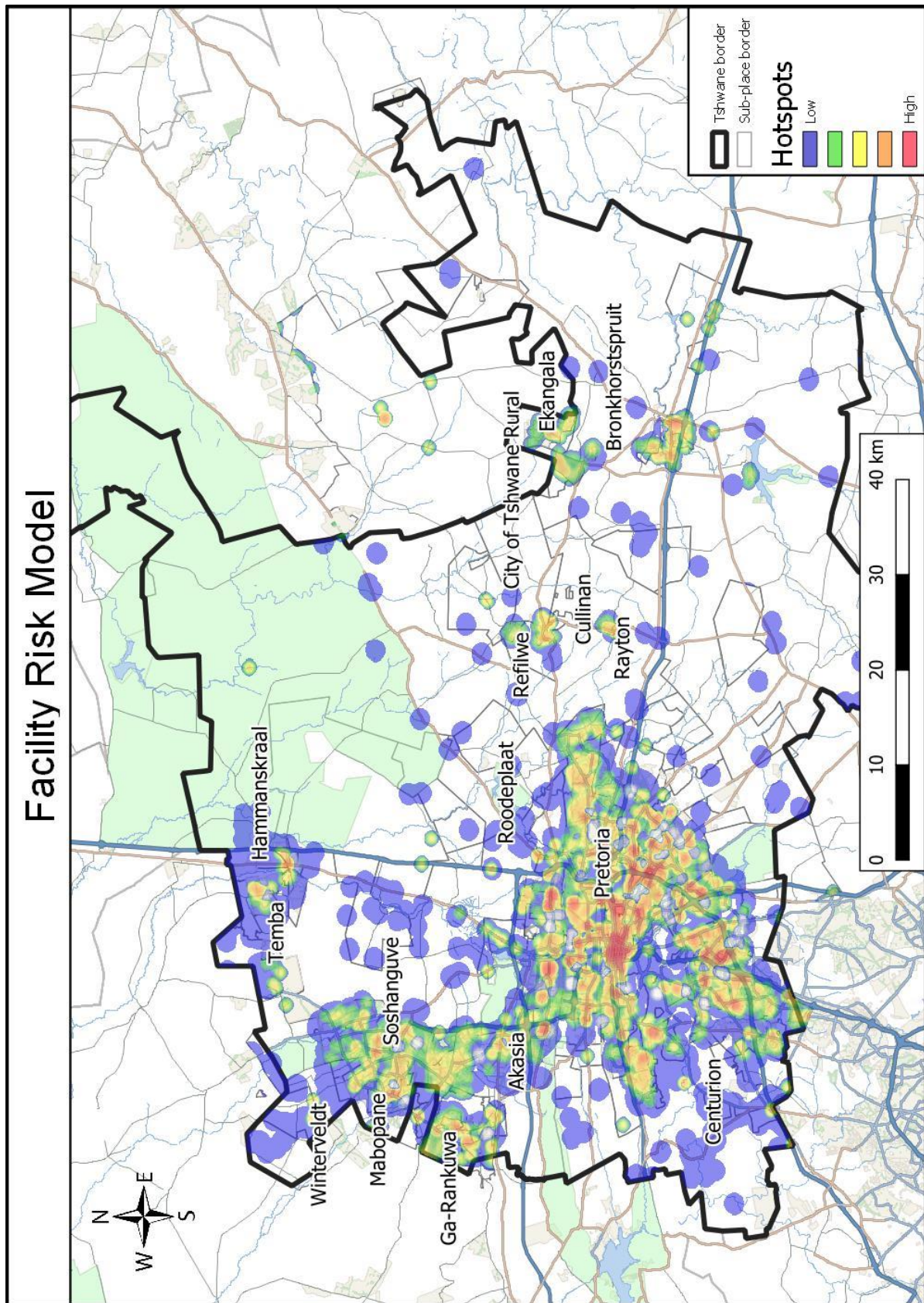
5.3.4 Social disorganisation model

According to the social disorganisation theory, crime is more prominent in neighbourhoods with low-economic status, high ethnic heterogeneity, high residential migration and high family disruption. In this study a social disorganisation index was built at the sub-place level of spatial aggregation. This was done to supplement the risk facility model developed above because it could be that certain facilities act as crime generators in more socially disorganised neighbourhoods than in more affluent neighbourhoods. The reason for including this analysis in the overall model was to incorporate the importance of neighbourhood level social disorganisation in the prediction of robbery in Tshwane.

Different methods for standardising the data were considered including using percentages. However, it was decided to use density-based measures instead for two reasons. First, most of the analyses are density driven. Second, it was found that using other measures such as percentages skewed the index, highlighting areas with small population counts in large sub-place areas as more deprived. Of course, these areas are more deprived than the others based on percentages, however considering that this study aims to find places with the highest density of robberies; the deprivation per square kilometre within a sub-place area gives a more accurate portrayal of the 2006 robbery footprint.

The first step involved in the creation of the social disorganisation model was to select the variables used to represent social disorganisation. The variables selected include the number of male population between the age of 15 and 34 per square kilometre, the number of households without electricity for lighting per square kilometre, the number of dwellings without water within 200 meters per square kilometre, the number of households without pit latrine with ventilation or flush toilet per square kilometre, the number of unemployed population per square kilometre, the number of shacks as dwelling type per square kilometre, the number of divorcees per square kilometre, the number of people migrated to sub-place in last 5 years per square kilometre and the number of people who were born in other African countries per square kilometre.

Table 8: Social disorganisation demographics per sub-place measured in squared kilometres ($n = 455$)

	Min	Max	Mean	St. dev
Male (age 15-34)	0.00	6 768.17	470.87	700.34
No Electricity	0.00	3 829.46	157.87	472.77
No access to water	0.00	2 566.67	98.85	301.43
Inadequate Access to toilets	0.00	5 839.43	205.38	583.15
Unemployment	0.00	8 910.03	859.92	1 224.15
Shack	0.00	5 898.37	197.54	608.64
Divorced	0.00	525.02	53.69	64.13
Moved (Last 5 years)	0.00	8 132.18	591.16	949.34
African foreign born	0.00	556.41	17.10	47.57

In order to compare variables of different sub-places with each other the data was normalised. This was done by converting the data into z-scores. This forces each variable to a value between zero and one (see Table 9). Finally the social disorganisation index was calculated by adding the z-scores together. Therefore, the social disorganisation index is a value between 0 and 9 (the amount of deprivation variables) with 9 being the highest deprivation a sub-place can score. Alternative multivariate statistical techniques, such as exploratory factor analysis and principal component analysis, were considered in the construction of the index; but in the interest of simplicity a summation of the selected variables, with equal weighting, was applied.

**Table 9:** Social disorganisation demographics z-scores

	Min	Max	Mean	St. dev
Male (age 15-34)	0.00	1.00	0.07	0.10
No Electricity	0.00	1.00	0.04	0.12
No access to water	0.00	1.00	0.04	0.12
Inadequate Access to toilets	0.00	1.00	0.04	0.10
Unemployment	0.00	1.00	0.10	0.14
Shack	0.00	1.00	0.03	0.10
Divorced	0.00	1.00	0.10	0.12
Moved (Last 5 years)	0.00	1.00	0.07	0.12
African foreign born	0.00	1.00	0.04	0.09

The decision to use equal weighting of the social disorganisation variables was twofold. First, there is a lack of local research on the different influences each of these variables has specifically on robberies to help inform on the weightings between these variables. Second, the model that is being employed is a global model, which does not take local variations into account. For instance, unemployment may have a larger influence on robberies in Pretoria CBD than in the rural areas. Therefore, further research is necessary to inform on weightings in order to improve upon this model.

Figure 13 shows the social disorganisation index, created from the steps above, for each sub-place on a map. For display purposes, the social disorganisation index values were divided into 32 classifications using the nested average method, displayed from blue (lowest deprived) to red (highest deprived).

Figure 13: Social Disorganisation Index - Vector
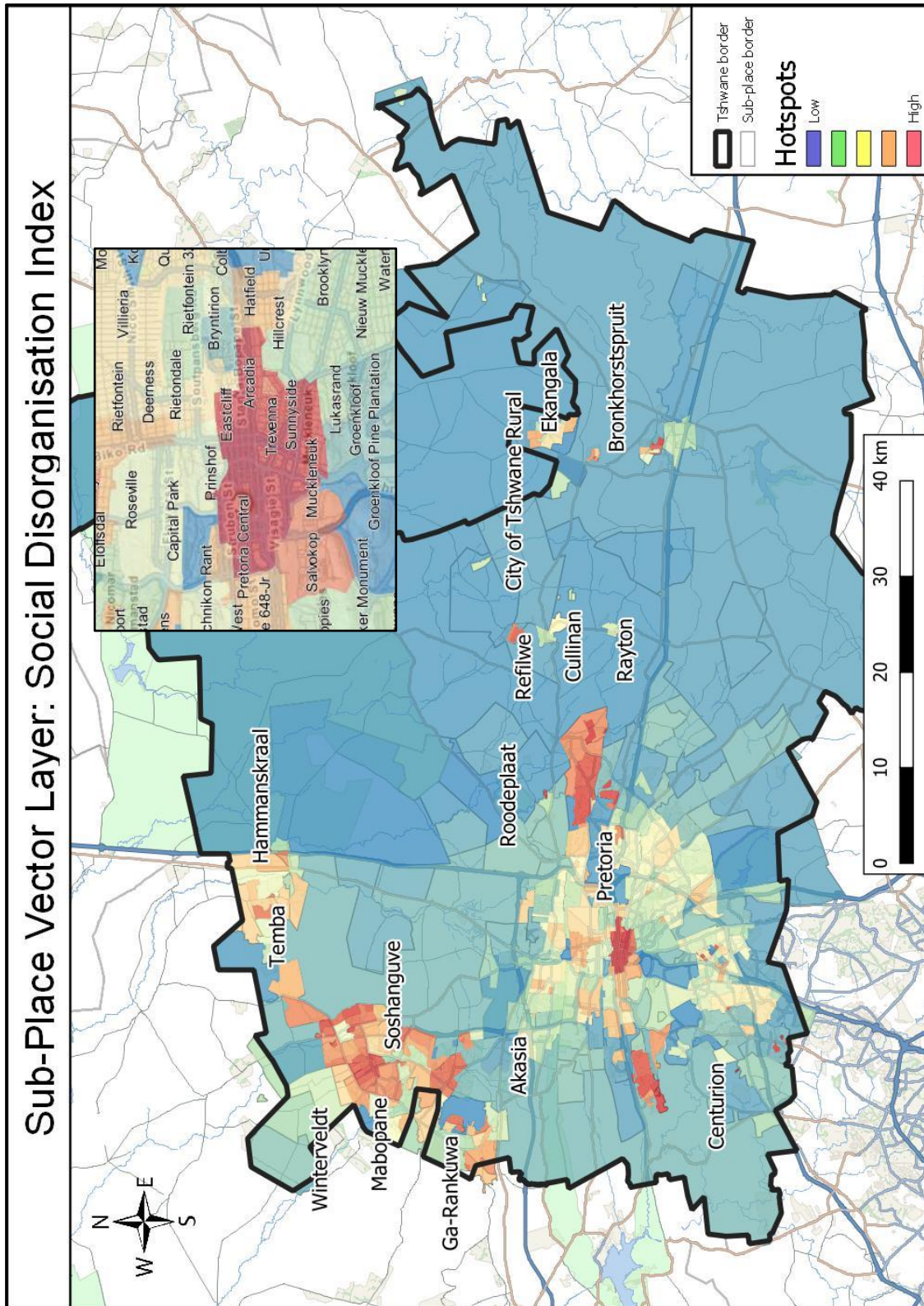
The next step in the process was to convert the vector data, at the sub-place unit of analysis, to raster. This was done in order to combine the output of the social disorganisation index with the previously created facility risk model. The vector to raster conversion was done using the following code in R.

```
#--- Create a empty raster with same specifications as the robbery probability raster
raster_specs <- raster('C:/1_CrimeAnalysis/Inputs/Raster_Specs.tif')
sd_raster <- raster()
extent(sd_raster) <- extent(Tshwane_border)
ncol(sd_raster) <- ncol(raster_specs)
nrow(sd_raster) <- nrow(raster_specs)
proj4string(sd_raster) <- "+proj=tmerc +lat_0=0 +lon_0=29 +k=1 +x_0=0 +y_0=0 +axis=wsu +ellps=WGS84
+towgs84=0,0,0,0,0,0,0 +units=m +no_defs"
###-------Convert SP boundary vector to raster layer-------###
sd_raster <- rasterize(SD_Tshwane_master, sd_raster, 'SD_Index')
```

The rasterised sub-place layer is shown in Figure 14. For display purposes, the social disorganisation index values in the raster layer were also divided into 32 classifications using the nested average method, similar to vector sub-place layer. However, the colour scheme has changed from white to black, in order to see the pixels (grid blocks) better, with black being the highest deprived areas.

73

Figure 14: Social disorganisation index - Raster

5.3.5 Temporal constraint model

The final step in the analysis was to create a temporal constraint model. According to the temporal constraint theory, crime probability increases with the time potential criminals spend at commuting nodes. Furthermore, the temporal constraint theory suggests that potential criminals have limited time to explore areas beyond their commuting nodes. Therefore, the temporal constraint model uses commuting nodes to create a temporal constraint raster with locations (grid cells) closer to the commuter nodes at higher risk values than locations (grid cells) further away from commuter nodes. Theses commuting nodes consist out of the intersections of highways, main roads and train stations. Similar to the facility model, the commuting nodes were transformed into a raster grid with proximity values approximating the likelihood of robberies happening closer to the commuting nodes.

Combining the various commuter raster layers (i.e., highways stacked over main roads stacked over train stations) into a composite commuter layer was once again done through the process of stacking the commuter raster layers on top of each other. The final temporal constraint raster grid was then created by adding each commuter kernel density value together for each cell. The process of creating the temporal constraint raster is displayed in Figure 15.

Figure 15: Commuter raster stacking process
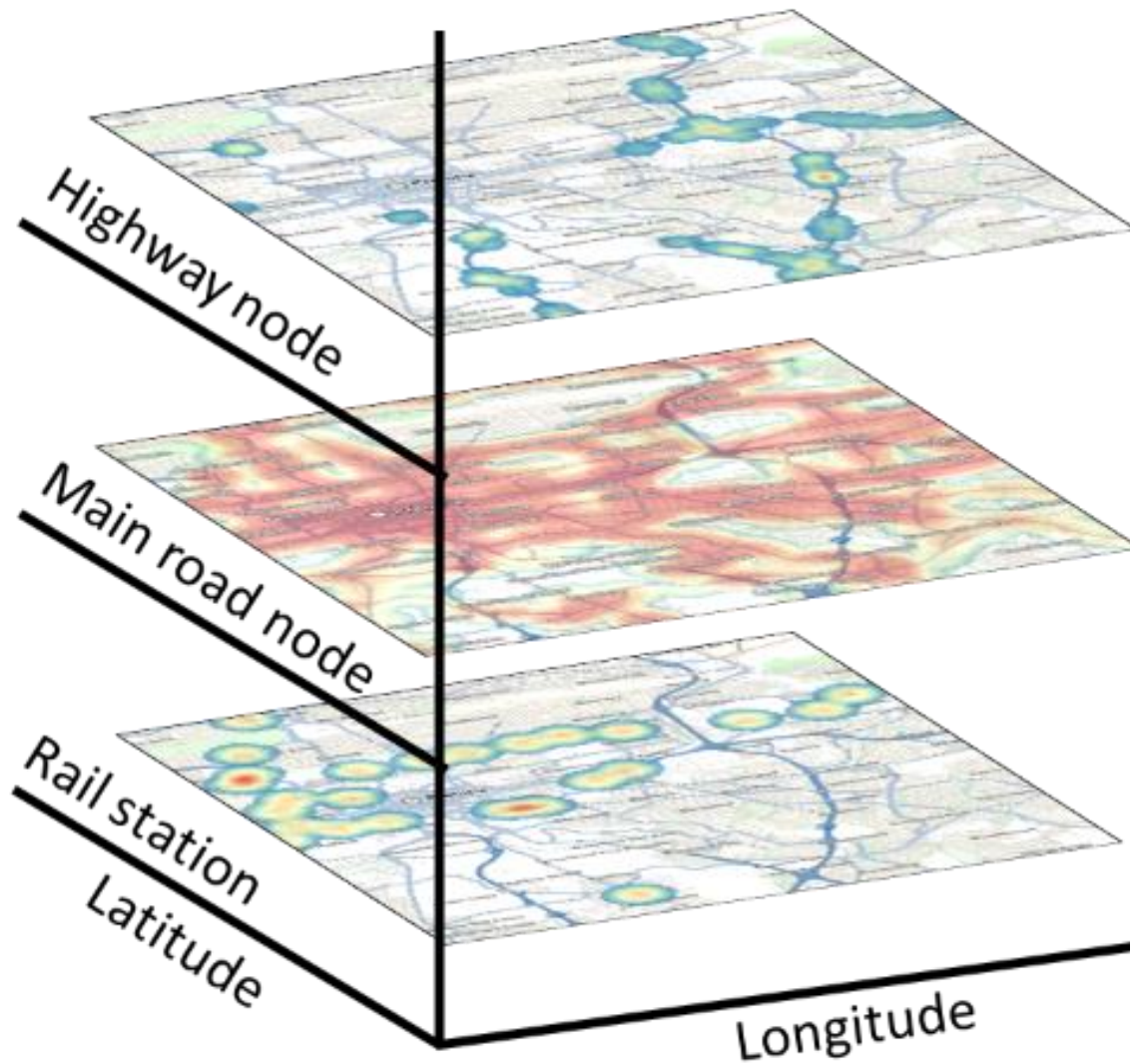
The temporal constraint value was calculated by adding the grid cell values for each commuter node type (highways, main roads and train stations) and is shown in Figure 16. For display purposes, the temporal constraint index values were divided into 32 classifications using the nested average method, displayed from blue (low) to red (high).

**Figure 16:** Temporal constraint index raster
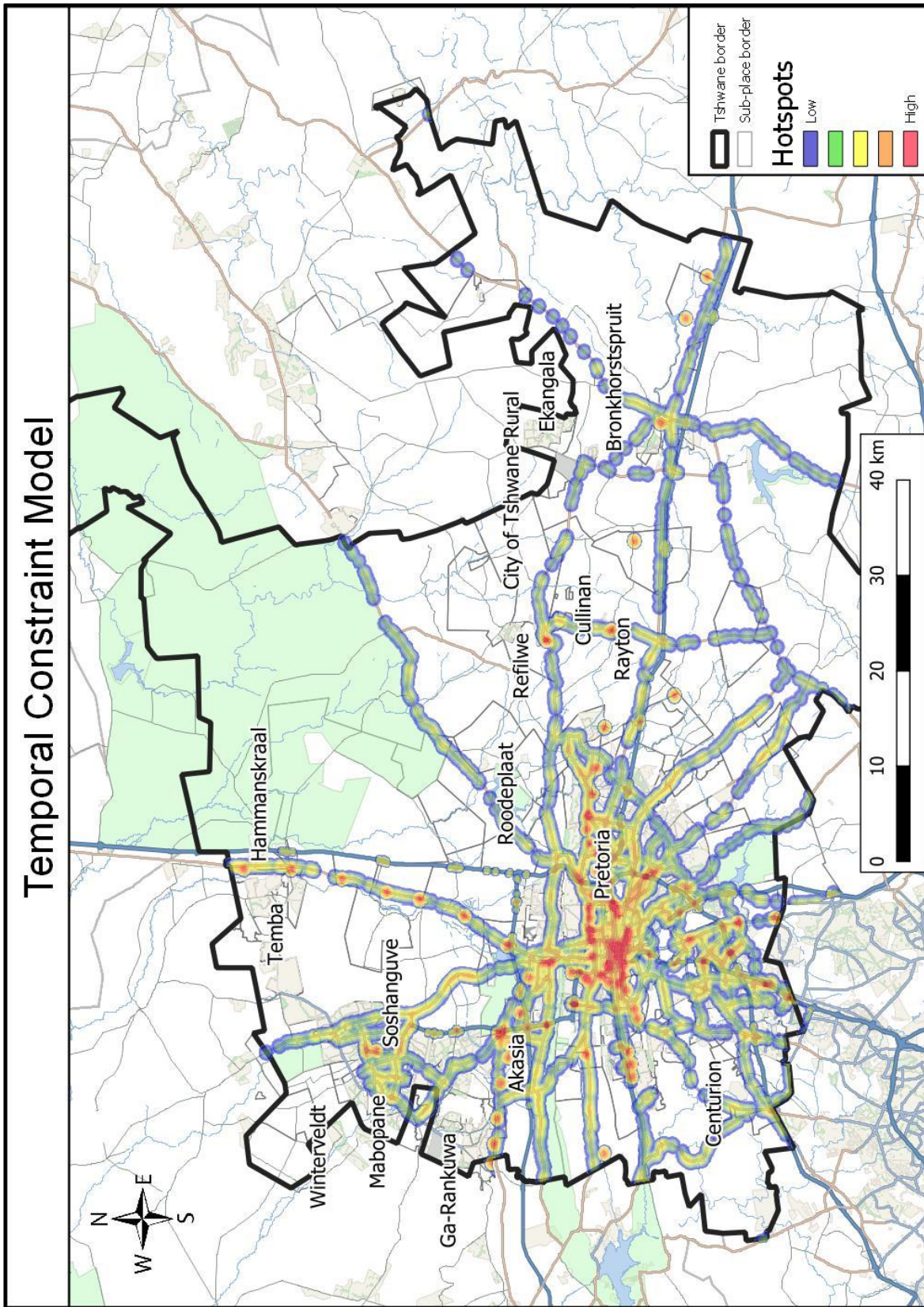
5.3.6 Robbery probability model

The final stage in the development of the robbery prediction model was to combine the previous three models (i.e., risk facility mode, social disorganisation model and the commuter node model) into a robbery probability model. Inspired by spatio-temporal crime theories namely routine activity, crime pattern, social disorganisation and temporal constraint theory, the above sections created three different robbery probability risk layers. In this section, the facility probability, social disorganisation index and temporal constraint models are combined to create the final robbery probability model. Shown in Figure 17, the process of creating the final robbery probability model involved stacking the three together.

Figure 17: Robbery probability model process
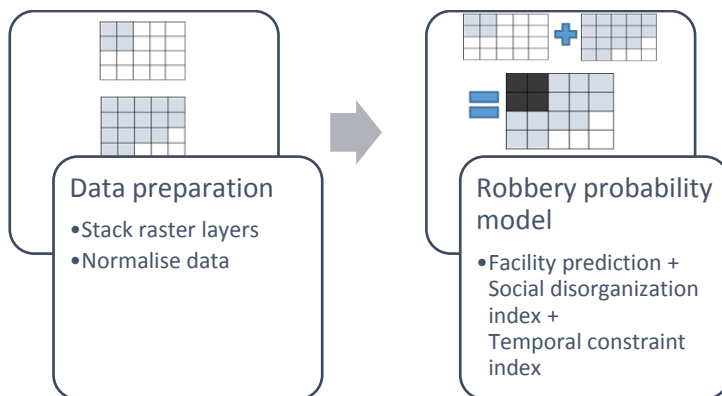
The facility risk, social disorganisation and temporal constraint raster layers were built on the same raster template, meaning that their grid blocks overlap each other exactly. This enabled the raster layers to be stacked on top of each other and summed to produce the final robbery probability raster.

Figure 18: Robbery probability raster stacking process
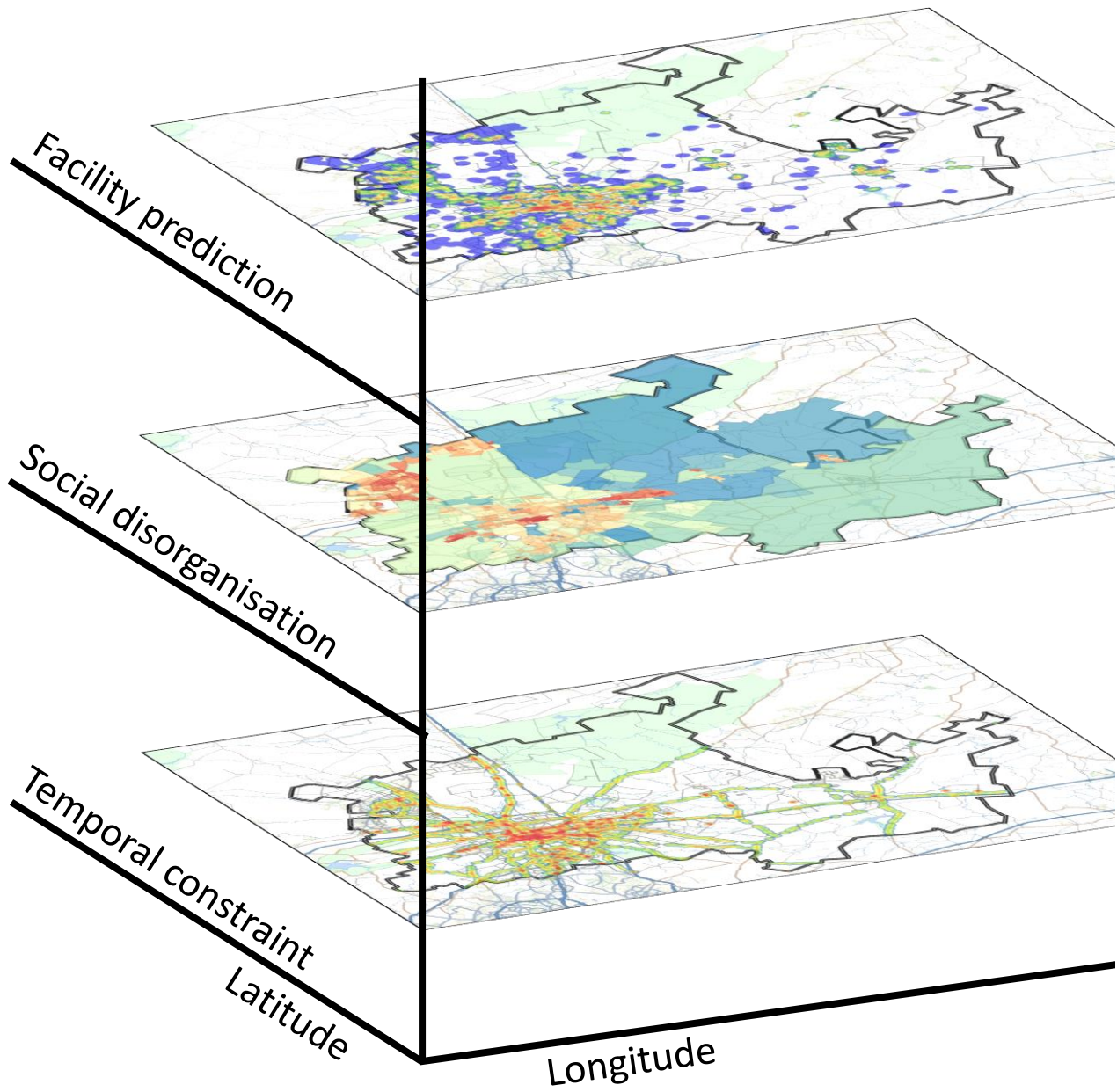
Before stacking the three models together, the grid cell variables needed to be normalised. In the analysis no weightings were applied to any model as it was assumed that each model plays an equal role in the commission of robbery in the city. It could be that facilities play a bigger role than social disorganisation in the prediction of robbery but that was not the assumption followed. As a result all grid cells for each model were normalised to allow them to be stacked and combined to form the final robbery risk model. The final robbery prediction model is shown in Figure 19. For display purposes, the robbery probability values were divided into 32 classifications using the nested average method, displayed from blue (lowest deprived) to red (highest deprived).

Figure 19: Robbery probability raster map
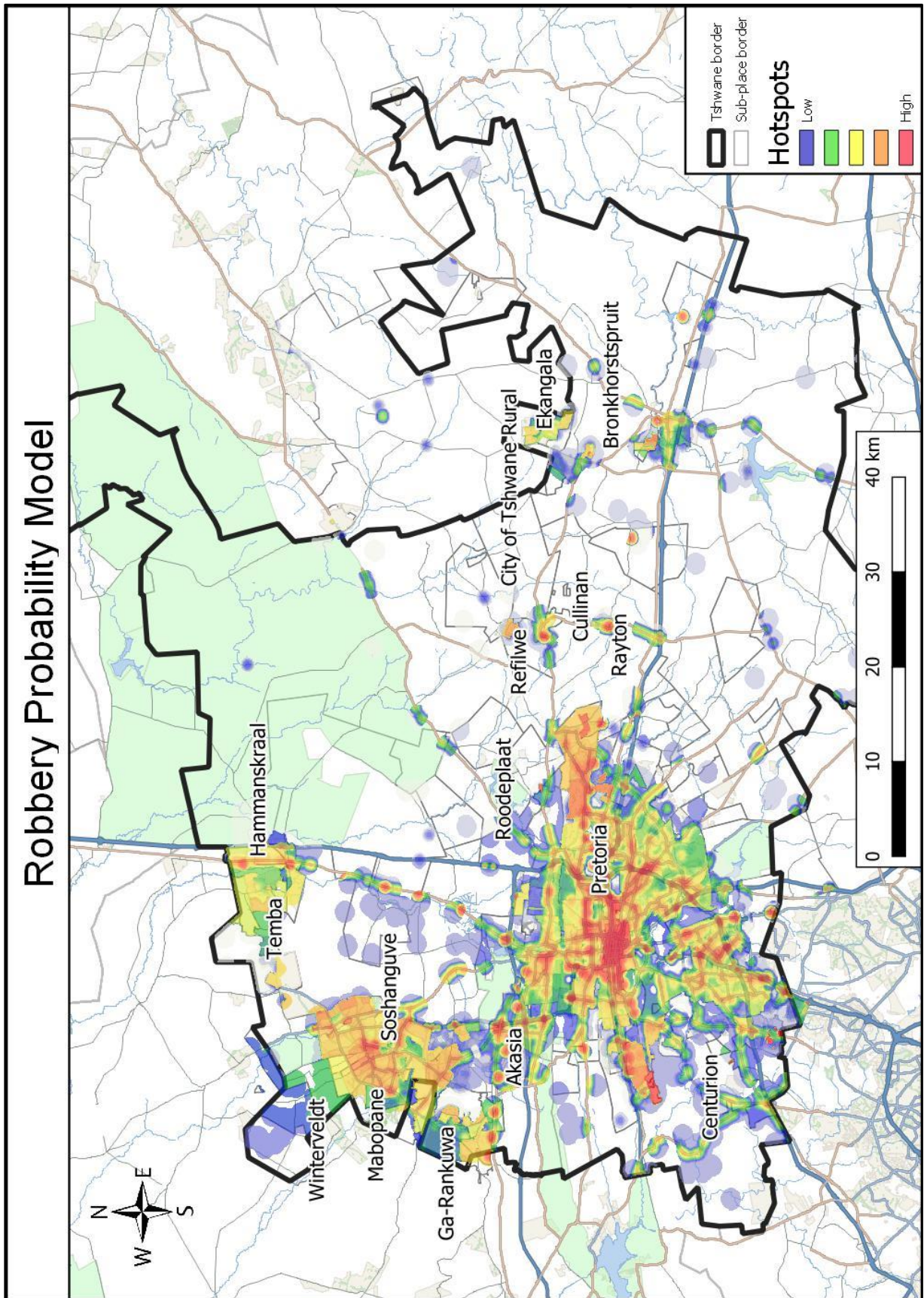
5.4 Traditional hotspots compared to robbery probability model

The main aim of this study was to develop a robbery prediction model for robbery in Tshwane. A secondary aim was to compare the robbery probability hotspot model with a 'traditional' hotspot model to determine which model best predicts robbery hotspots. This remainder of this chapter outlines how the two models were compared. Furthermore, the exact same methodology was used to prepare separate models for day and night-time for both the robbery probability and the traditional hotspots model.

5.4.1 Prediction Accuracy Index (PAI)

In order to compare the traditional hotspot model (generated in section 5.3) with the final robbery prediction model (generated in section 5.3.6) a prediction accuracy index (PAI) was used. Chainey et al. (2008) developed the prediction accuracy index (PAI) as a measure to test the prediction accuracy of hotspot techniques in relation with the study area size. The PAI equation is as follows.

Figure 20: Prediction Accuracy Index Equation

$$\frac{\left(\frac{n}{N}\right)*100}{\left(\frac{a}{A}\right)*100} = \frac{HitRate}{AreaPercentage} = \text{Prediction Accuracy Index}$$

In this equation, **n** represents the number of crimes which fall in the hotspot areas, **N** is the total number of crime events in the study area, **a** represents the area size of the hotspots and **A** the area size of the study area. The higher the PAI value the better the hotspot model explains future crime events.

Unlike the hit rate method - which only calculates the percentage of new crime events within the predicted hotspot areas, the PAI method gives a more accurate description on the performance of a prediction model on a micro-level analysis by including area size. The inclusion of the area size is what makes the PAI value attractive as it is more relevant to a real life scenario, i.e. police are often under resourced and therefore physically incapable of being at all the hotspot areas all of the time. Therefore, the

smaller the area police has to cover in which most of the potential crime occurs, the better.

5.4.2 Standardising the prediction models

In order to compare PAI values with each other the size of the hotspots needs to be controlled. Chainey et al. (2008) specify that three percent of the whole study area needs to be considered as hot. Therefore, the current study also selected three percent of the traditional and robbery probability raster grid blocks falling within the whole City of Tshwane Metropolitan Municipality predicted to contain the most robbery occurrences.

Figure 21 shows the comparison between the traditional and the robbery probability model with three percent of the grid cells selected as hotspots for the City of Tshwane Metropolitan Municipality.

Figure 21: Selecting hotspots from a kernel density estimation raster
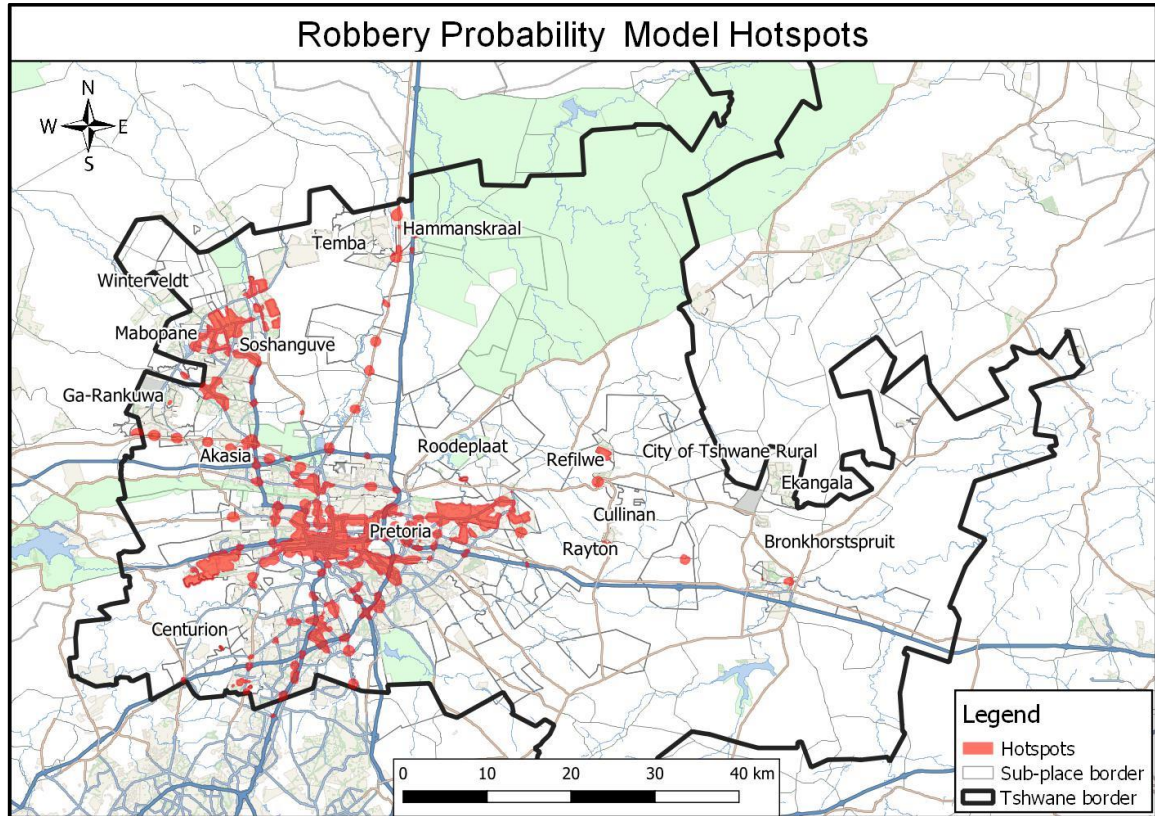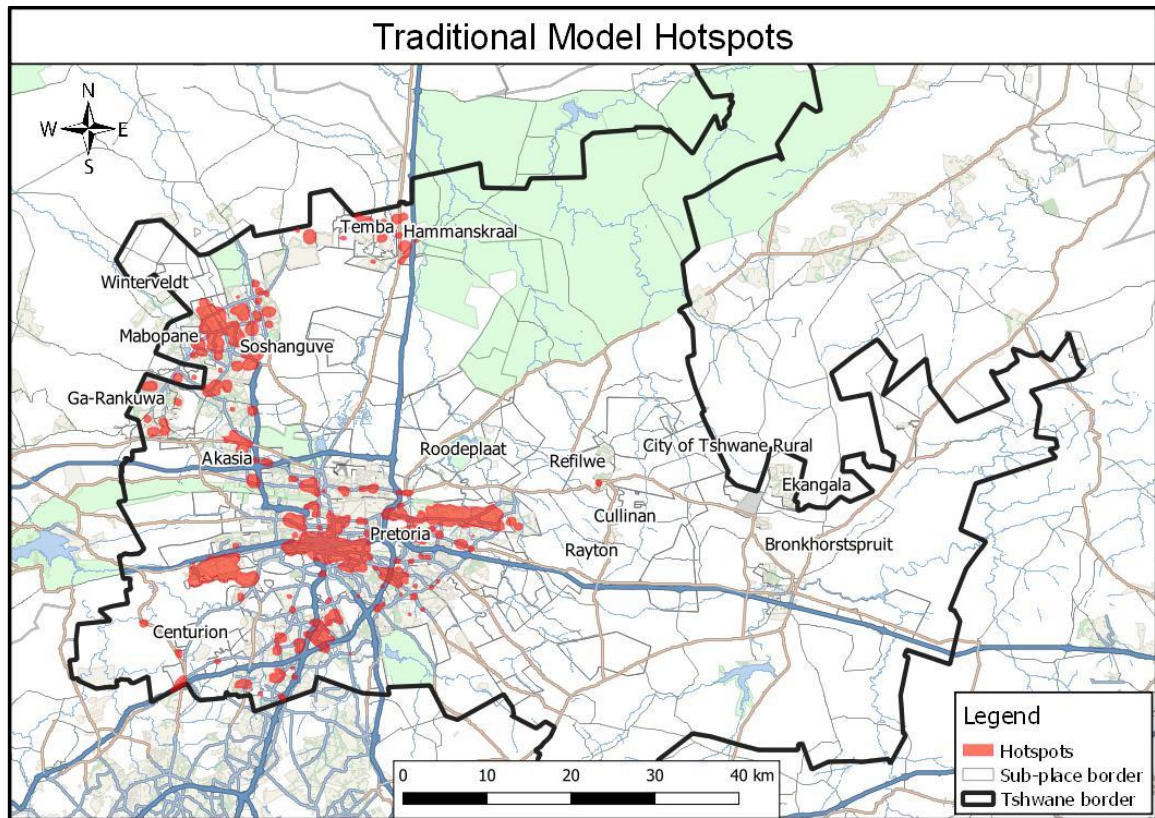
5.4.3 PAI implementation

In order to calculate the PAI values for the traditional and robbery probability models a number of steps are first required.

The first step was to extract the areas deemed to be "hot" from the raster layer. The current study decided to extract the top three percent robbery hotspot areas falling within the City of Tshwane Metropolitan Municipality. The code snippet below crops the raster layer and retrieves the number of grid blocks falling within the Tshwane border. The code then continues to extract the grid blocks with the highest robbery density values, accumulating to three percent of the study area.

```
###--------------------- Calculate amount of gid blocks accumulating to 3% of study area  ---------------------###
Top_perc_value <- 3 #The percent of study area needed = 3%
Study_Area <- area(Tshwane_border_size)/1000000
grid_cell_sizeKm <- 0.01 #Grids size = 100m*100m (0.01km2)
perc_area_grids <- ((Top_perc_value/100)*Study_Area)/grid_cell_sizeKm
###--------------------- Crop raster to study area  ---------------------###
raster <- projectRaster(raster, crs=crs(Tshwane_border))
raster <- crop(raster, Tshwane_border)
raster <- mask(raster, Tshwane_border)
###--------------------- Select top grids from cropped raste accumulating to 3% of the study area size  ---------------------###
total_n <- ncell(raster) #Total number of grid cells in study area
raster[order(-raster[])[(perc_area_grids+1):total_n]] <- NA #Set 97% of area to NA
raster[!is.na(raster[])]<- 1 #Set top 3% area to 1
```

The second step in the PAI process was to convert each model (the 'top three percent' traditional and the 'top three percent' robbery probability models) into vector. The vector layers were then exported to a geopackage for display purposes. The code snippet below creates the polygon layer from the raster and re-projects the polygon layer to meters.

```
###-------------------- Raster to polygon conversion (Remove NA values and dissolve similar values ---------------------###
  pol <- rasterToPolygons(raster, na.rm=TRUE, dissolve = TRUE)
###--------------------- Re-project polygon to meters ---------------------###
  pol <- spTransform(pol, c("+proj=tmerc +lat_0=0 +lon_0=29 +k=1 +x_0=0 +y_0=0 +axis=wsu +ellps=WGS84
+towgs84=0,0,0,0,0,0,0 +units=m +no_defs"))
```

The reason for re-projecting the polygon layer to meters was to calculate the hotspot area coverage in square kilometres. The code snippet below aggregates the count of 2007 robberies found within the hotspot areas and calculates the area size in square kilometres.

```
###--------------------- Count of 2007 robberies ---------------------###
  total_robbery <- nrow(2007_robbery_locations)
###--------------------- Count of 2007 robberies inside hotspots ---------------------###
  RobberyCount <- as.integer(poly.counts(robbery_locations, pol))
###--------------------- Study area calculation (km2) ---------------------###
  Study_Area <- area(Tshwane_border_size)/1000000
###--------------------- hotspot area calculation (km2) ---------------------###
  Hotspot_AreaKm <- area(pol)/1000000
```

The last step in the methodology was to calculate the PAI value for the traditional and robbery probability model. This was done by comparing the 2007 robbery occurrences in Tshwane with the traditional and robbery probability models in order to see which of the two models best predicts 2007 robberies in the city. It should be noted that both the traditional and the robbery probability models were built using 2006 robbery data. By validating both models using 2007 data, allows the user to see which model best predicts future robbery events. The code snippet below implements the PAI equation by inserting the above calculated variables. First, the hit rate percentage for each model was calculated. This was done by dividing the 2007 robbery count (individual points) within the hotspot area by the total robberies that occurred during 2007. Second, the percentage area of Tshwane covered by the hotspot area was calculated. This was done by dividing the hotspot area by the total area of the City of Tshwane Metropolitan Municipality. The last step taken was to divide the hit rate percentage

(first step) with the area percentage (second step). The code below shows how the latter steps were implemented in R.

```
###--------------------- Hit rate ---------------------###
  HitRate <- (RobberyCount/total_robbery)*100
###--------------------- Area percentage ---------------------###
  AreaPercentage <- (Hotspot_AreaKm/Study_Area)*100
###--------------------- PAI value ---------------------###
  PAI <- HitRate/AreaPercentage
```

5.5 Limitations

The models proposed in the current study are built on the assumption that the datasets used are complete and accurate. In reality this is almost never the case and therefore it is important to understand the shortcomings of the current study to improve on and or to apply the models in a different scenario. The two main limitations experienced in the current study were, first, the availability of data and second, the computer processing capabilities.

5.5.1 Data availability

Data is the single most important variable in prediction models as the term, garbage in, garbage out, describes that the prediction output can only be as accurate as the data that fed the model. In the current study the two main data sources were the crime data sourced from the South African Police Service and the facility data sourced from AfriGIS (Pty) Ltd.

Robbery data limitations:

For the current study to predict robbery locations as accurate as possible, all robberies should be reported and spatially accurately captured. According to Mistry (2004) the 2003 national victims of crime survey reported that only 29 percent of robbery cases were reported to the police. Of course, the percentage may have changed during 2006; nevertheless, it can be assumed that the robbery counts used in the current study does not represent the actual amount of robberies that have occurred. This is

of concern as not only does this impact the traditional hotspot model but also the facility risk model which uses robbery occurrences to determine the spatial relationship between robberies and facilities. Furthermore, the robbery counts are also being used to test the prediction strength of each model, which according to the 2003 national victims of crime survey does not account for 71 percent of robberies.

The current study assumed that all locations of the SAPS robbery data are accurate. There may however be some anomalies especially in areas such as informal settlements or open fields where the exact locations are unknown. It is also common practice to place incidents at a specific facility when the exact locations are unknown. Furthermore, certain facilities such as hospitals are often used to take victim statements and therefore also used as a georeferenced incident point. This can result in facilities faulty showing as crime hotspot.

Robberies that occurred in the year 2006 were used to develop the models created in the current study. The reason for this was because the facility and commuter nodes were sourced from the same year. Of course, conducting the study on more recent data will provide a more accurate portrayal of the current robbery events in Tshwane. The dataset that was available is however the only official and spatially replete dataset available in the country by which to conduct spatial analysis. Additional ways to conduct crime analysis using other data sources – recently outlined by Faull (2019) – could be used in the future when apply, or improving upon, the model developed here.

Facility data limitations:
The facility data obtained from AfriGIS (Pty) Ltd contained a confidence level indicating how accurate the facility's location on the map resembles its true location. The locational accuracy of the current study requires the facilities to be plotted on the street and within the suburb of the facility's address. There were however a few facilities that could only be located at the suburb level and therefore excluded from the analysis. Other than known facilities that could not be located accurately enough to be used in the current dissertation, there is also a high probability that unknown 'informal' facilities exist in Tshwane – particularly in township settings - but has not formally been documented yet. This is especially the case in the northern, more rural

parts of the City of Tshwane Metropolitan Municipality where neighbourhoods consist of informal dwellings. Informal dwellings often contain points of interest known to the community but are not officially listed as businesses and therefore not recorded as facilities. The last concern was the limited available facilities, which forced the current study to leave out other facilities such as alcohol outlets, clubs, parking lots, ATM locations, which according to the crime pattern and routine activity theories are nodes often seen as crime attractors or generators.

Census data limitations:

The 2001 census data was included in the analysis. The first limitation was that the last three national censuses were conducted in the years 1996, 2001 and 2011, making 2001 the logical choice for the current study. The current study was however conducted using crime data the years 2006 and 2007, and therefore not taking into account five years of demographic changes that might have occurred in the City of Tshwane Metropolitan Municipality during this time. Furthermore, density of African foreign born were used as a heterogeneity variable, this however does not include undocumented foreigners which may cause the total number of foreign born residence of an area to be misleading. Further studies may also benefit in expanding on the heterogeneity variable by including population and language groups in South Africa.

5.6 Methodology summary

In the analysis two main robbery prediction models were created; a traditional hotspot and a robbery probability model. The aim was then to compare the latter two models in terms of how accurately they predicted future robbery events (i.e., in this study, this refers to the 2007 robbery incidents).

The main difference between the traditional and robbery probability model is that the traditional hotspot model relies solely on previous crime events to build a hotspot model whereas the robbery probability model identifies at-risk areas based on the underlying built characteristics of the environment as well as the underlying socio-demographics. Figure 22 shows how a flow diagram for the methods followed.

Figure 22: Daytime and night-time robbery prediction model methodology
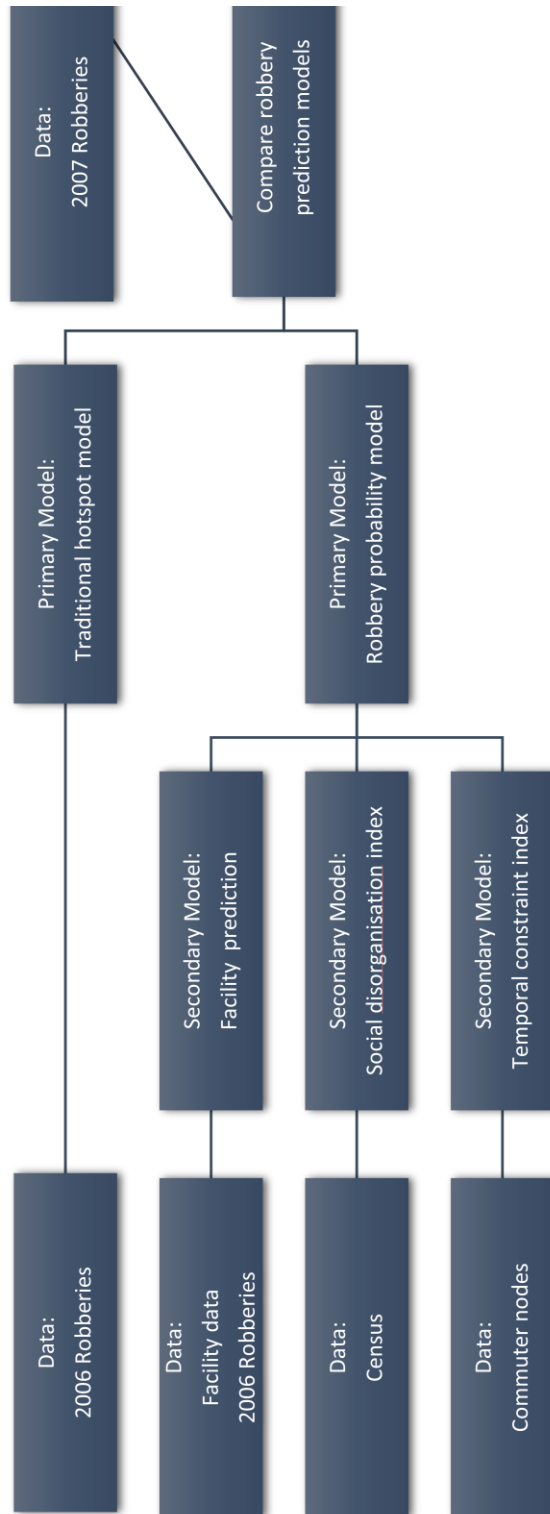
Accordingly, the 2006 robbery locations were used to create a traditional hotspot model. Three datasets were combined to create the robbery probability model namely, facilities, commuter nodes and socio-demographic data.

1) With the facility data as independent variables and the 2006 robbery locations as the dependent variables, a regression model was run which identified facilities most closely associated with robbery and a predicted robbery density grid was calculated.

2) Socio-demographics were extracted from the census data using the social disorganisation theory as a guide

3) A social disorganisation index was created per SAL

4) The commuter nodes were selected based on the temporal constraint theory and were collected as vector point layers. The commuter nodes were transformed into raster grid and combined into the temporal constraint index.

5) All three datasets were made to conform to the same dimensions as the traditional hotspot model raster layer. Therefore, the raster layers of the three secondary models could be combined into one robbery probability model. Both the traditional hotspot and robbery probability model were tested on 2007 robbery locations and compared against each other using the prediction accuracy index (PAI) value.

The final outcome was two raster files: one representing traditional hotspots and the second, a map representing robbery probability areas. The two raster files were used to visually show the predicted crime hotspots in Tshwane while the tables were used to compare the two latter models with each other.

CHAPTER 6: RESULTS

6.1 Introduction

This chapter describes the results obtained. It outlines the accuracy of both prediction models (traditional hotspots versus the robbery probability model) and how they compare against each other. The findings of this study are related back to the existing literature as well as the various spatial crime theories used to inform the methodology. The current study divided the analysis into daytime (07:00am to 19:00pm) and night-time (19:00pm to 07:00am). Therefore the results will showcase both timeslots, while comparing the traditional hotspot to the robbery probability models.

6.2 Traditional hotspot results

The traditional hotspot model was built using robberies that occurred during 2006 within the City of Tshwane Metropolitan Municipality. The crime data were filtered based on type of crime, year and day of time. Robberies that occurred outside Tshwane's sub-place border were also discarded.

The traditional hotspot maps were built on a grid with a cell size of approximately 100 by 100 meters. Each of the grid cells contained a 2006 robbery density value as created with the kernel density method. Figure 23 and Figure 24 below shows the result of the traditional hotspots created for both day and night-time respectively. The red circles highlight areas with high robbery density.

Figure 23: Traditional hotspot map – Daytime
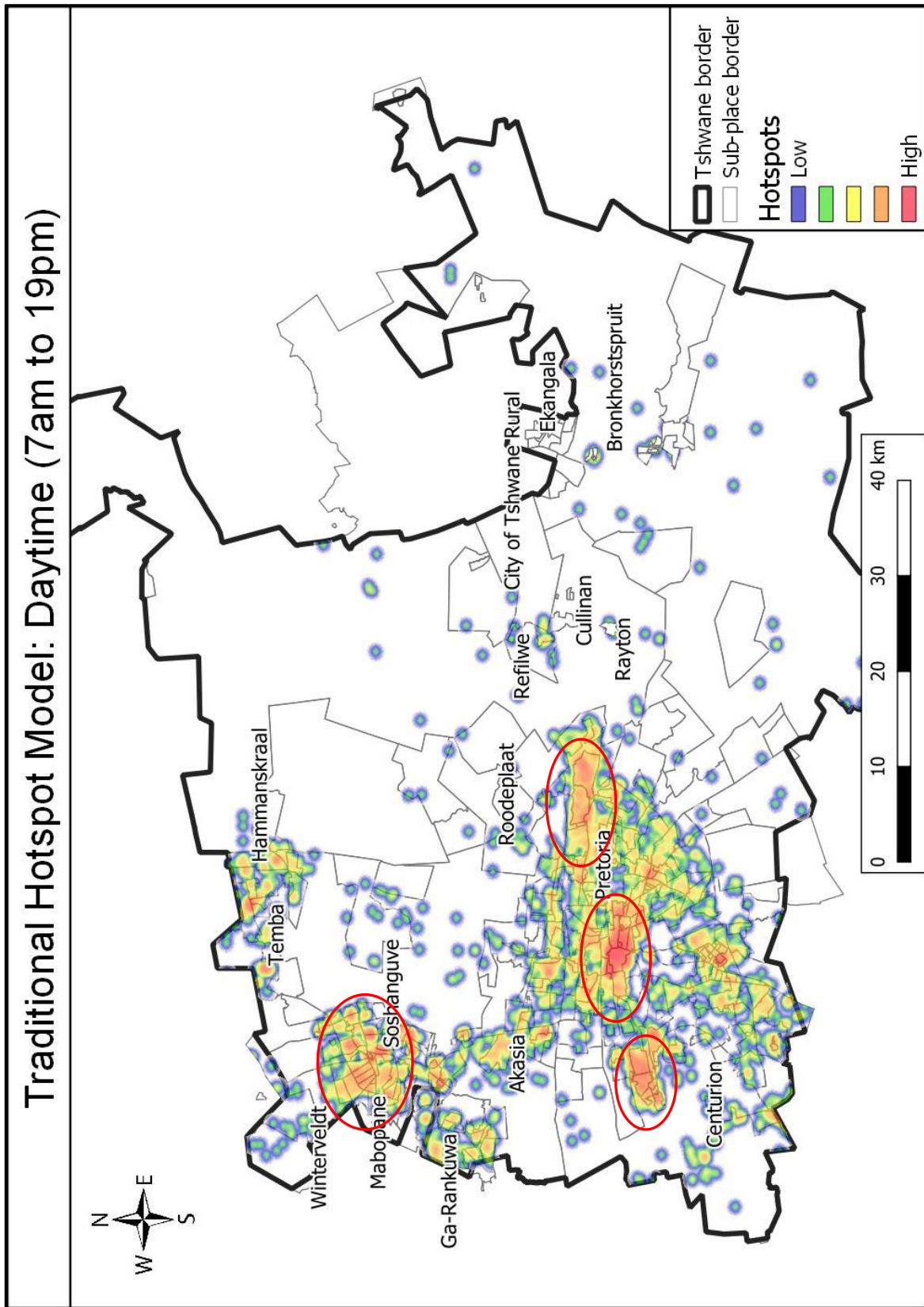
Figure 24: Traditional hotspot map – Night-time
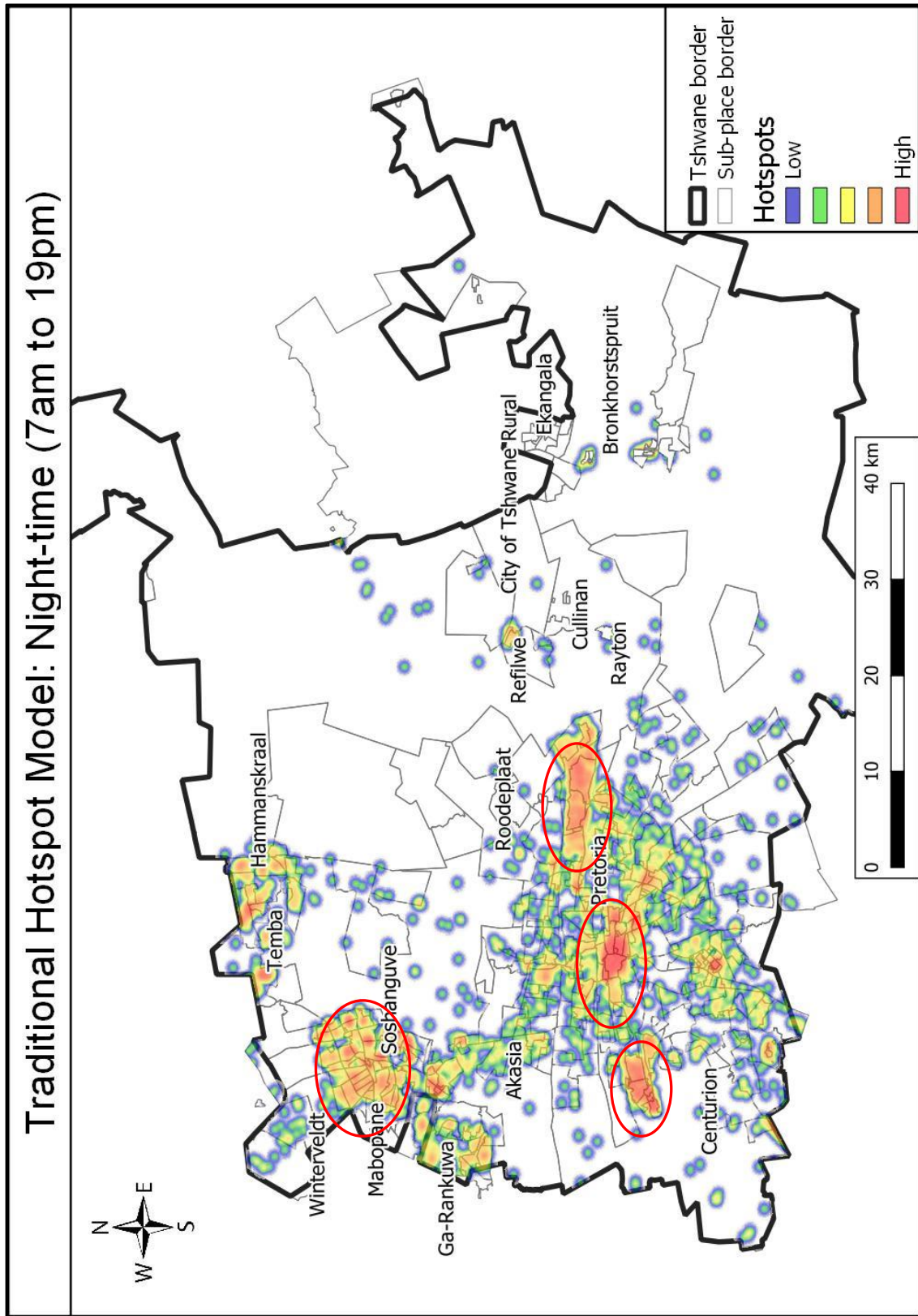
From Figure 23 and Figure 24 it is evident that similar spatial robbery patterns are found during day and night, meaning that the same areas seem to be targeted regardless of time of day. Furthermore, a visual observation shows that robberies mostly occur in the urban areas and noticeably less in the rural areas. This was expected as not only do more people reside in the urban areas but are also the place where most conduct their daily activities and according to the crime pattern and routine activity theory, these are the areas prone to criminal activity. Most robberies occur in the Pretoria central business district, the northern township of Soshanguwe, Mamelodi in the east of Tshwane and Atteridgeville in the west of Tshwane. The latter areas are highlighted by the red circles. The results support the social disorganisation theory, as Soshanguve, Mamelodi and Atteridgeville are generally more socially disorganised than other suburbs in the City of Tshwane Metropolitan Municipality (see Breetzke, 2010).

From Figure 25 it is noticeable that during the night, the towns of Hammanskraal and Temba experience a high increase in robbery activity. According to the crime pattern theory, one explanation might be that these are typical mobile areas, meaning that people living in these areas travel to work on a daily basis, spending most of their time during the day in another area. The same can be said for the town of Refilwe (Figure 26), where robberies recorded in 2006 increases from only two robberies during the day to 31 during the night.

Figure 25: Traditional hotspots in the towns of Hammanskraal and Temba
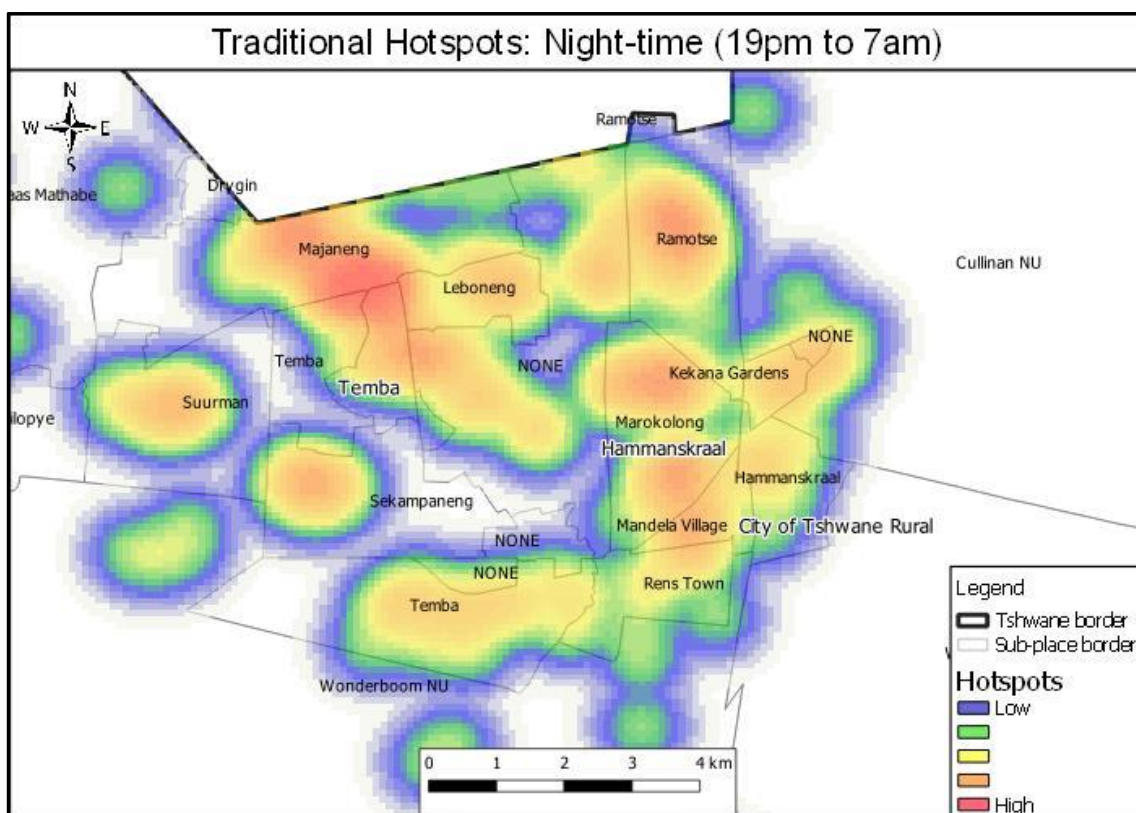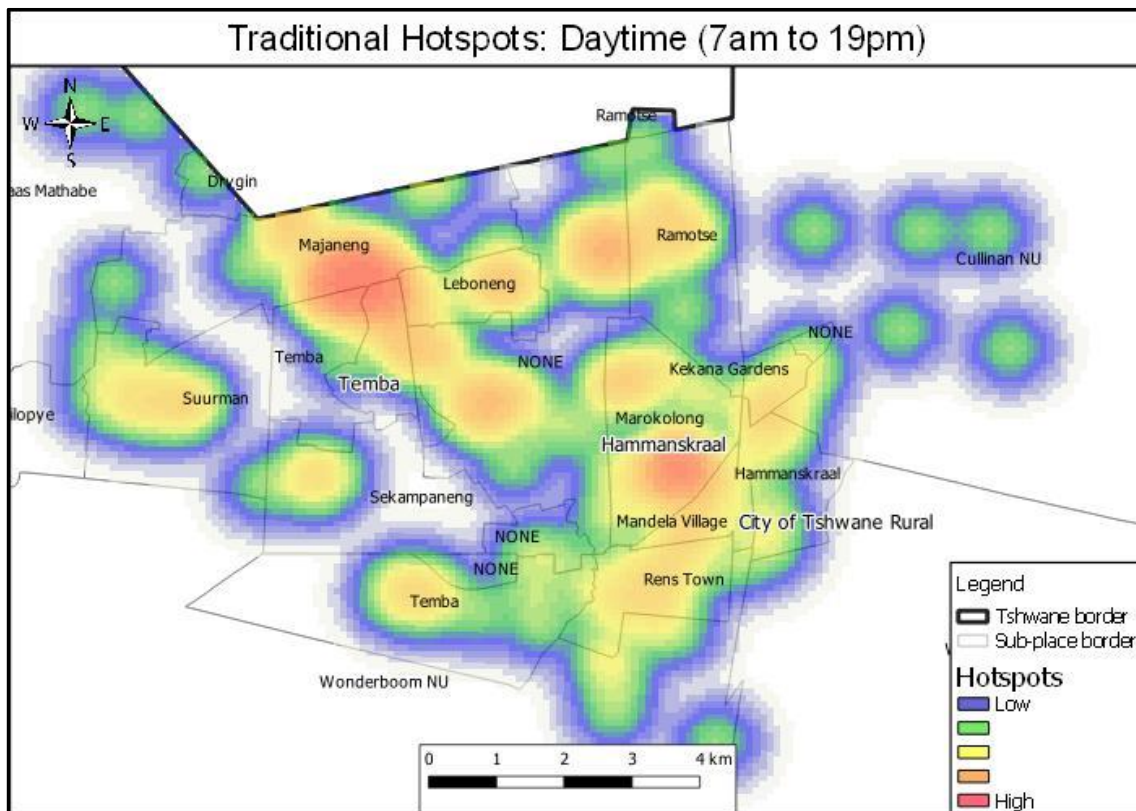
Figure 26: Traditional hotspots in the town of Refilwe
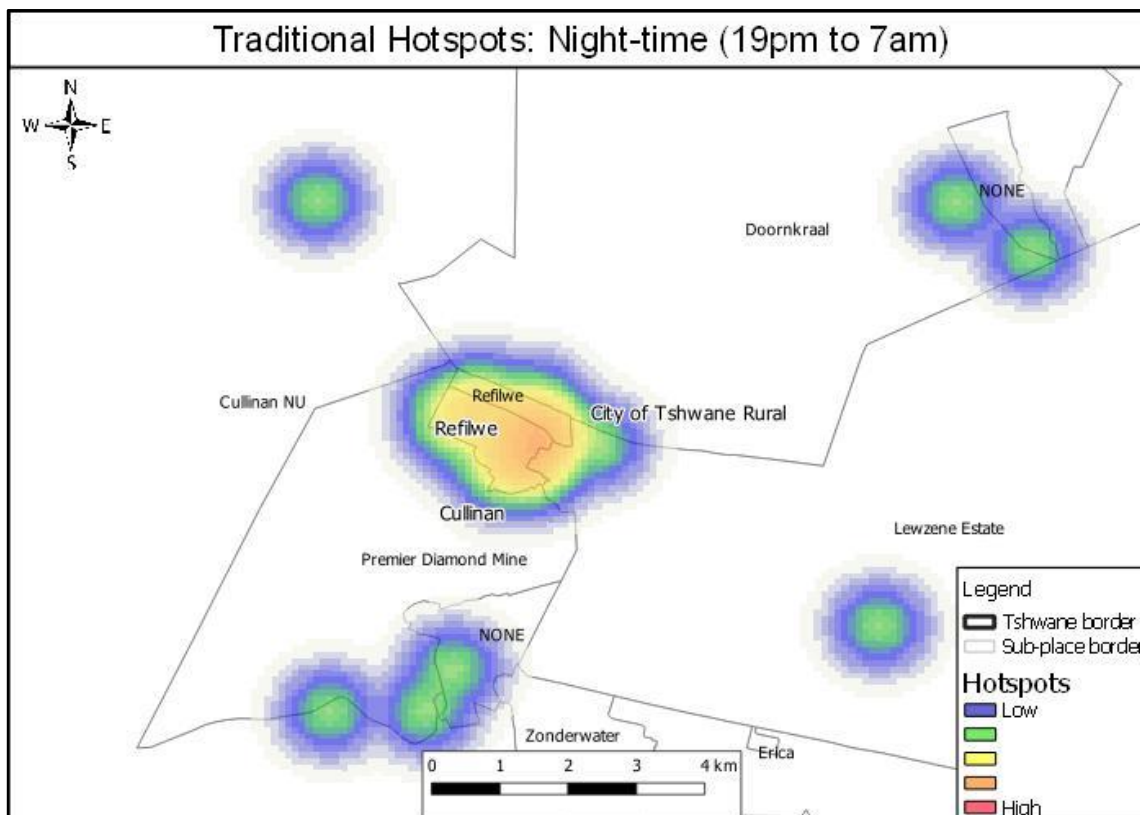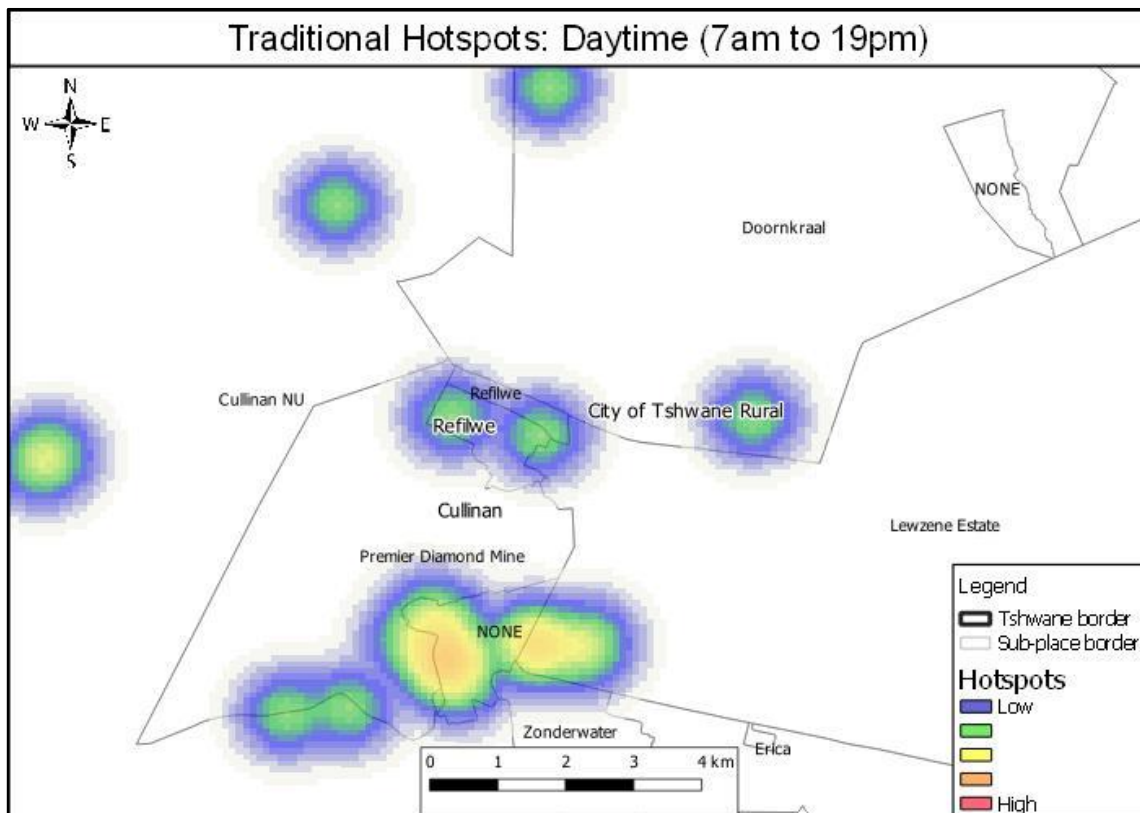
Traditional hotspots PAI value

The PAI value was used to measure how well the traditional model predicts future robberies. The PAI value measures the percentage of robberies that are accurately predicted while taking into account the percentage of area. However, the PAI value does not say much on its own but is rather a measure to compare the different hotspots with each other, where a higher PAI value means that the predicted hotspot area contains a higher concentration of robberies. A higher PAI value is therefore better because it indicates areas where police can target a large percentage of robberies using limited resources. Table 10 contains the values used to calculate the PAI value for both the day and night-time.

Table 10: Traditional hotspots PAI value for day and night-time (2007 robberies)

Time off Day	Study Area (km2)	Hotspot Area (Km2)	Hit Rate	PAI
Day	6947.92	208.08	72.05	24.06
Night	6947.92	208.14	73.09	24.40

Table 10 shows the values necessary to calculate the PAI value. "Study Area" represents the surface area of Tshwane's sub-places in square kilometres. The "Hotspot Area" field is the size of the extracted hotspots (as calculated in section 5.3) in square kilometres. The hit rate value was derived from the percentage of total robberies that fell in the hotspot area. Finally the PAI value was calculated by dividing the hit rate with the percentage of the total area being covered by the hotspot area. The higher the PAI value, the greater the accuracy of the model.

In the current study I limited the hotspot area to be three percent of the total study area. This was done to make the PAI values calculated for each model comparable with each other. Table 10 above shows that 72.05% and 73.09% of 2007 robberies fell within the day and night-time traditional hotspots respectively. Dividing the latter predicted 2007 robbery percentages by approximately 3 percent indicates that the

night-time model, with a PAI value of 24.4, performed slightly better than the daytime model with a PAI value of 24.06.

6.3 Robbery probability model results

The robbery probability model was inspired by various spatial crime theories namely, social disorganisation, routine activity, crime pattern and temporal constraint and uses a similar technique as RTM where raster layers are stacked upon each other to create a risk map based on various built and social features.
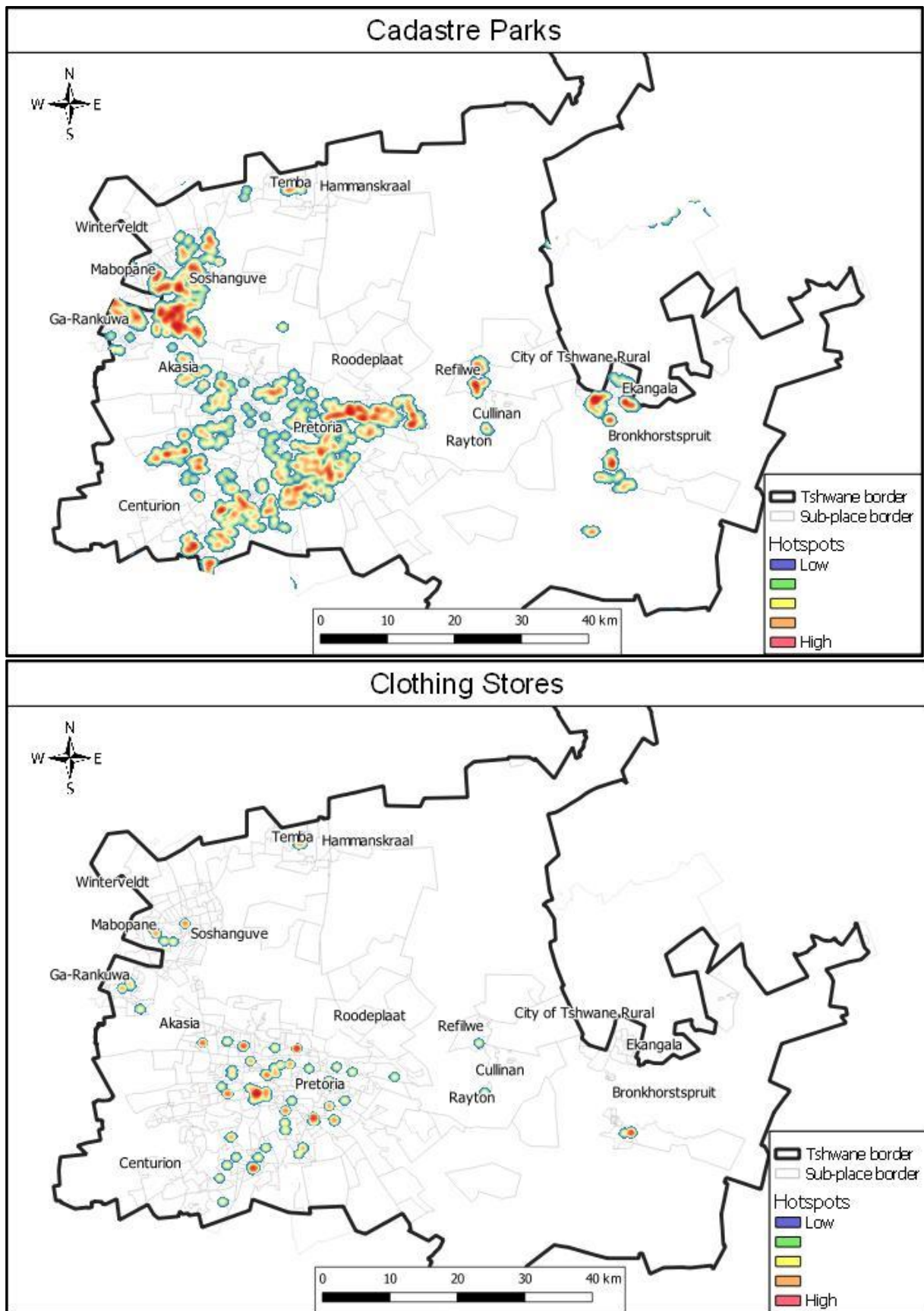
6.3.1 Facility risk model results

Initially regression analysis was used to predict future robbery hotspots using facilities as independent variables and the traditional hotspots as the dependent variable. The facilities used in the current study were chosen based on the routine activity and crime pattern theories.

The facilities included cadastre parks, clothing stores, convenience stores, educational facilities, fast food outlets, filling stations, office parks/blocks, other stores, restaurants, shopping centres and supermarkets. The grids for each facility are shown below in Figure 27.
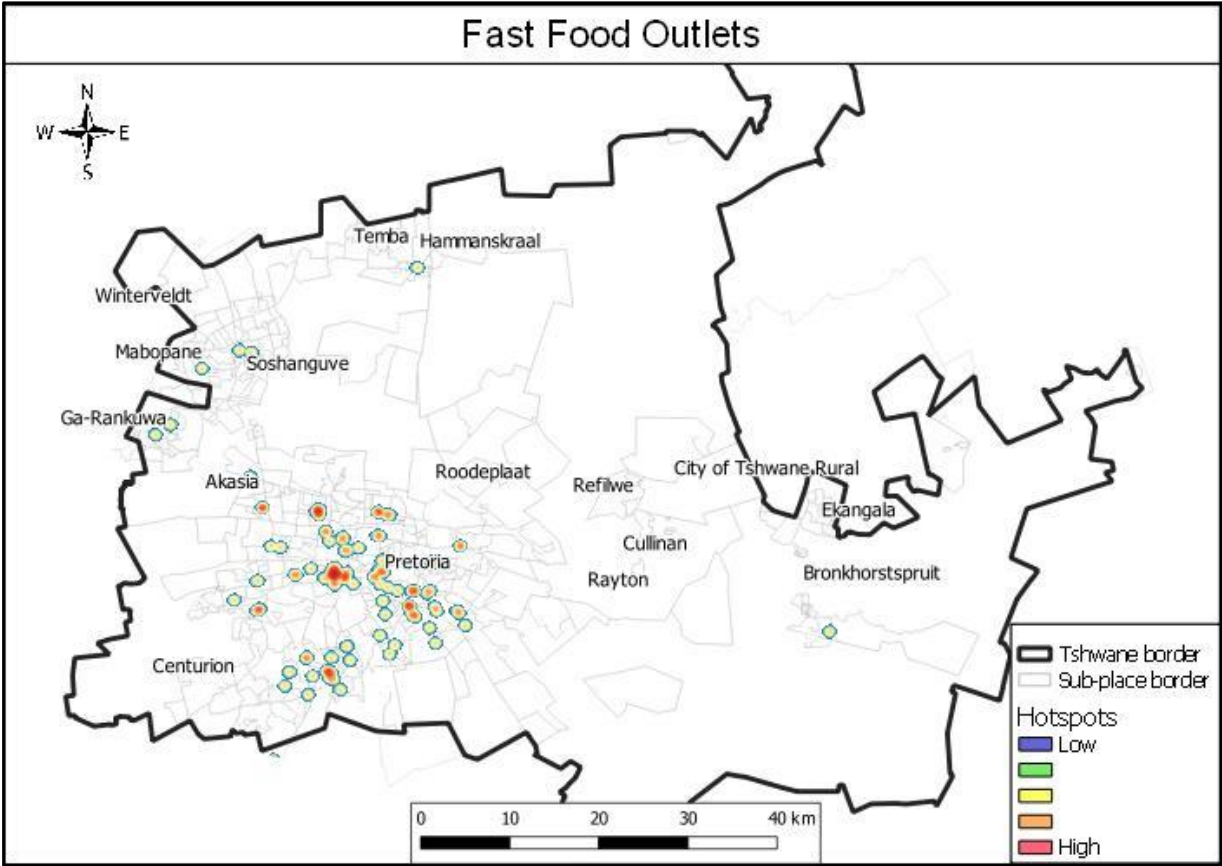
The facility hotspots were divided into 32 classifications using the nested average method, displayed from blue (low facility density) to red (high facility density).
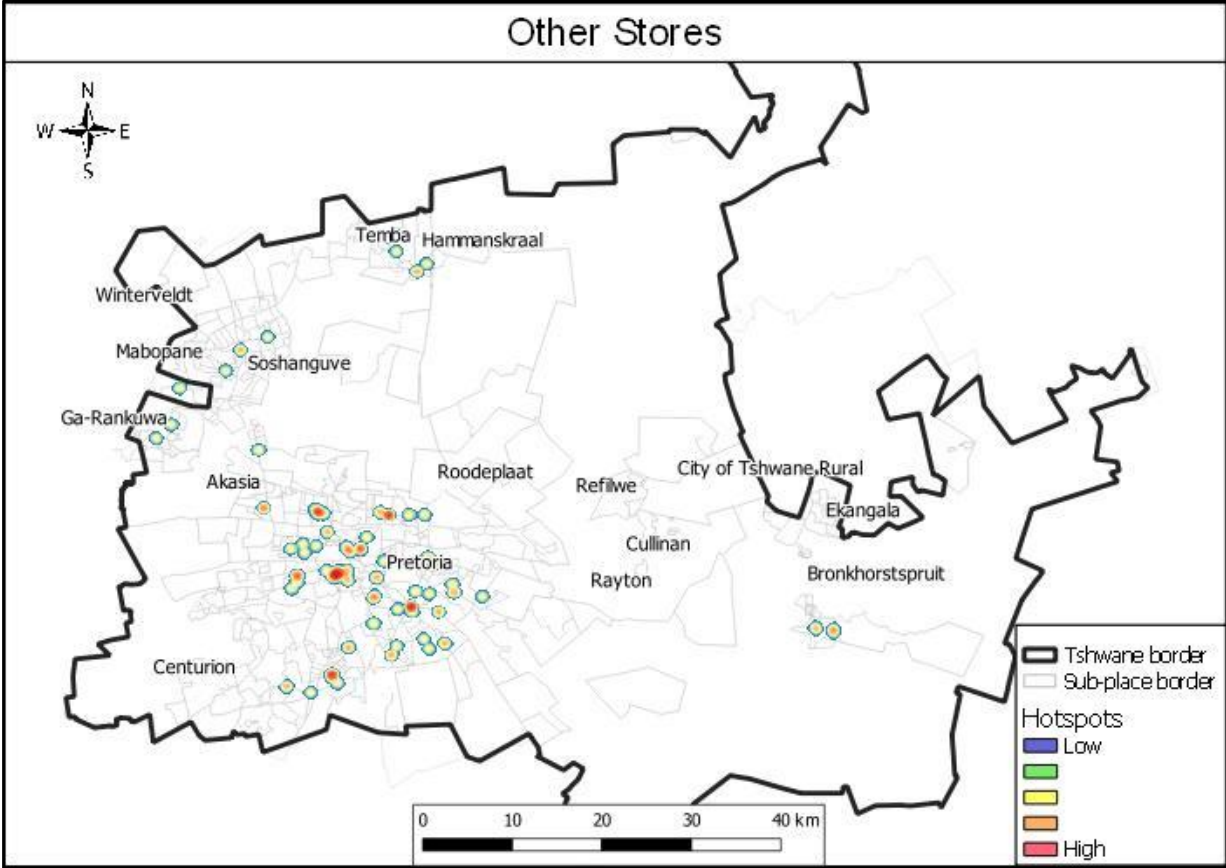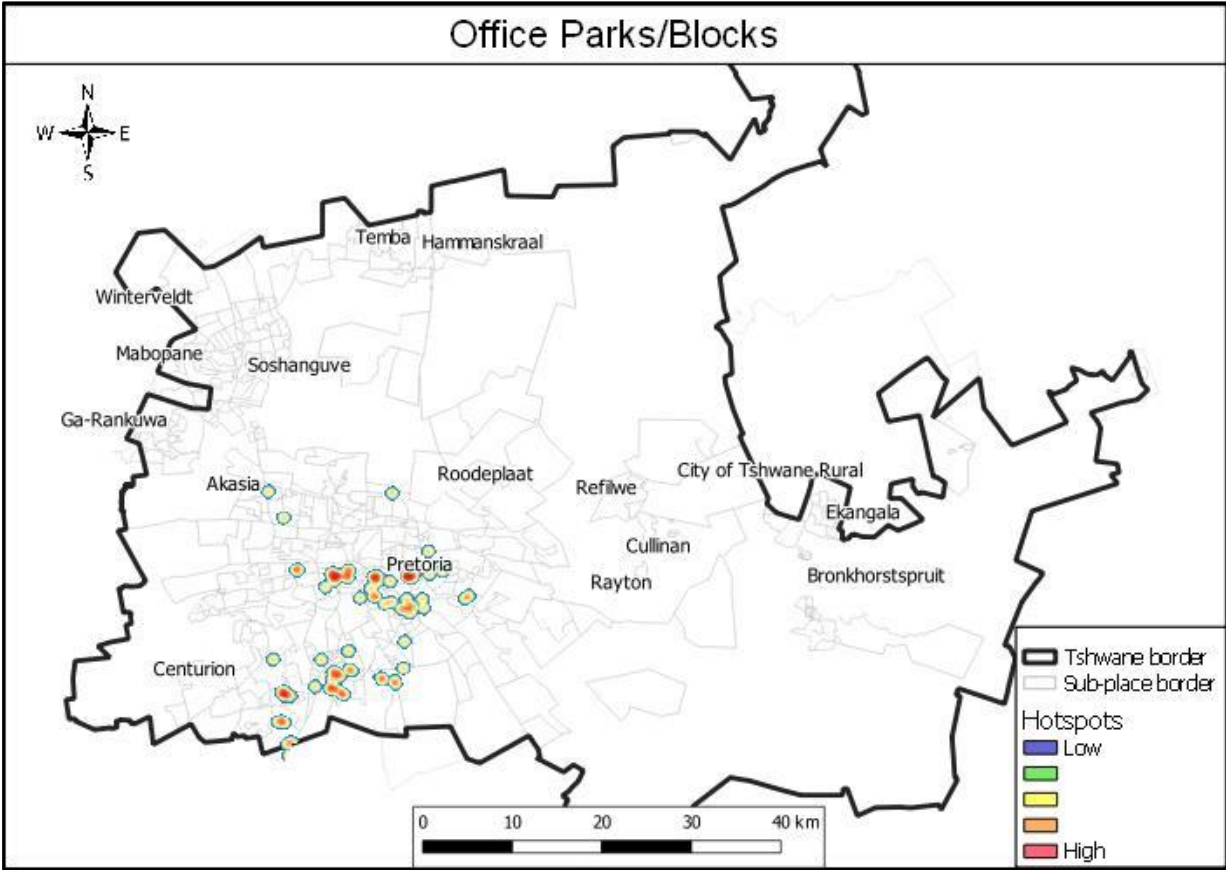
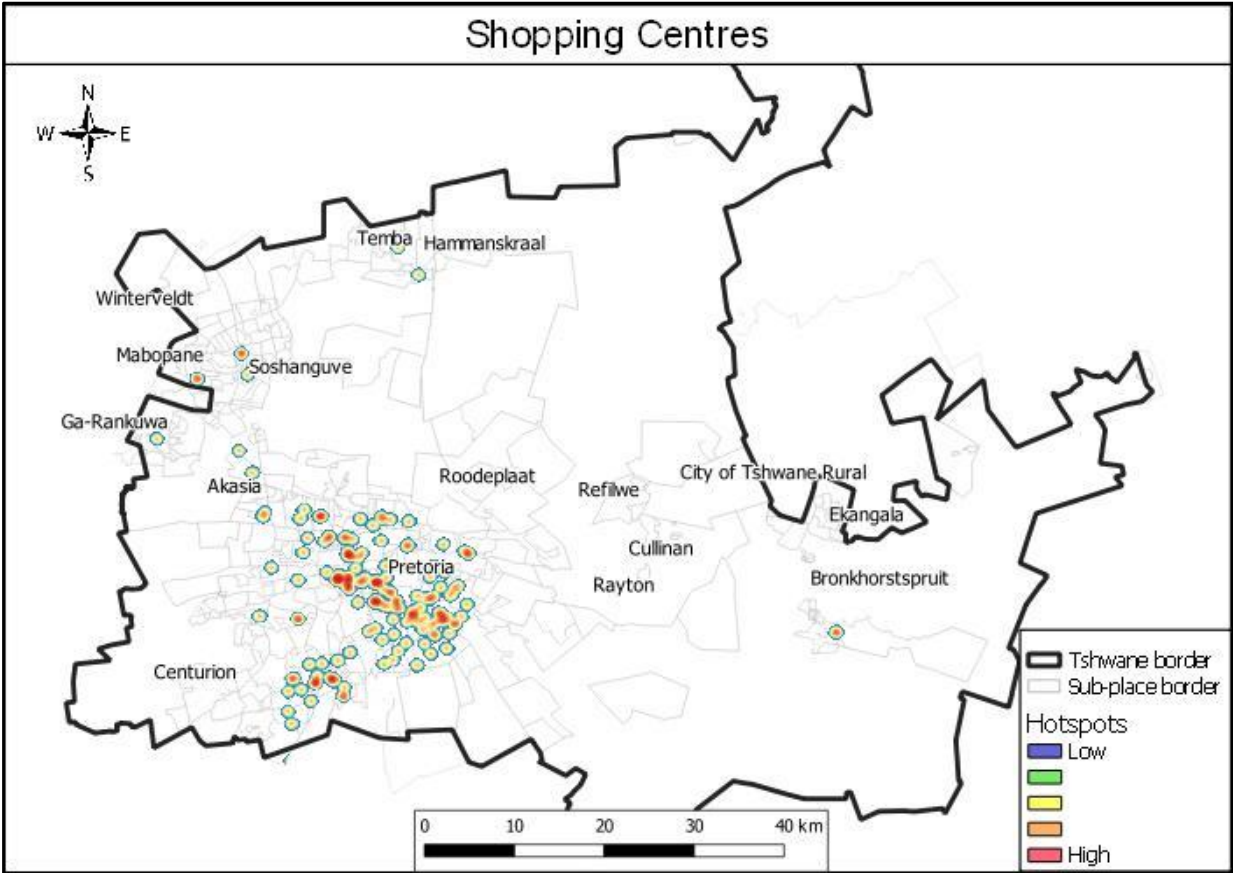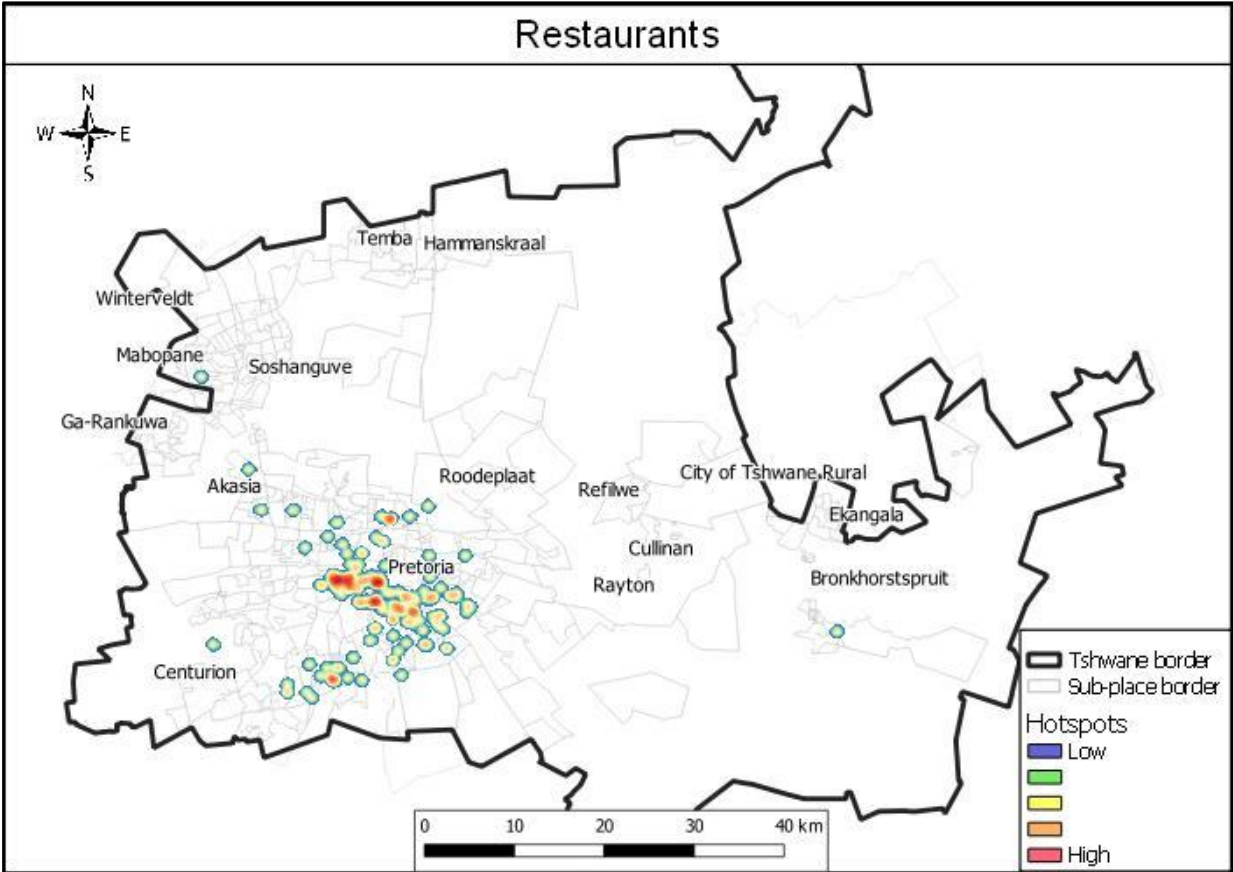**Figure 27:** Facility kernel density estimation heatmaps

Convinience Stores
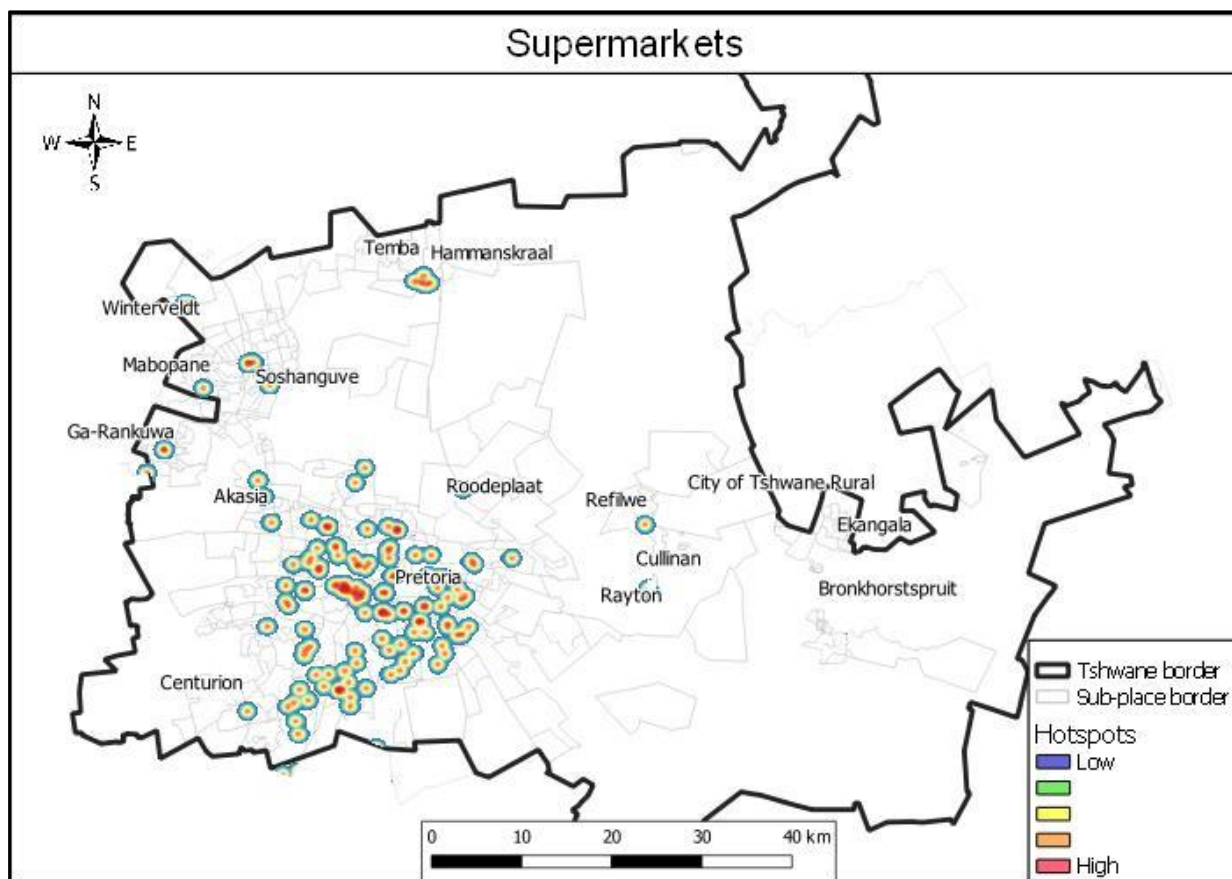
Temba
Hammanskraal
Winterveldt
Mabopane
Soshanguve
Ga-Rankuwa
Akasia
Roodeplaat
Refilwe
City of Tshwane Rural
Ekangala
Pretoria
Cullinan
Rayton
Bronkhorstspruit
Centurion

0   10   20   30   40 km

Tshwane border
Sub-place border
Hotspots
Low
High

Educational Facilities

Temba
Hammanskraal
Winterveldt
Mabopane
Soshanguve
Ga-Rankuwa
Akasia
Roodeplaat
Refilwe
City of Tshwane Rural
Ekangala
Pretoria
Cullinan
Rayton
Bronkhorstspruit
Centurion

0   10   20   30   40 km

Tshwane border
Sub-place border
Hotspots
Low
High

Fast Food Outlets

Filling Stations

Office Parks/Blocks

Other Stores

Restaurants

Shopping Centres

104

Most facilities have a presence in Pretoria central where also most robberies occurred during 2001 to 2005. Of course, Pretoria central was also found to be at a high risk because of its social structure and the volume of people residing in that area. It is therefore difficult to speculate if the robberies and facilities have a direct link, and if so, which facilities are the main contributors. By visual inspection, Pretoria central is expected to be a high robbery risk zone in the facility risk model based on the number and closeness of facilities in that area. Although cadastre parks are scarce in Pretoria central, hotspots created by this facility resembles the same pattern as seen in the traditional hotspots especially in the towns of Soshanguve, Mamelodi and Atteridgeville. Therefore, cadastre parks appear to be one of the main contributors to robberies outside the Pretoria central business district. Groff and McCord (2012) did a study highlighting parks as a crime generator. In their research they note that neighbourhood parks have a significant association with crime in Philadelphia, Pennsylvania, USA which through visual inspection seems to be a similar case for the City of Tshwane Metropolitan Municipality.

One concern with the facility risk model is that, except for parks and educational facilities, there are very limited numbers of facilities in the western, north-western and northern parts of Tshwane. This is direct consequence of apartheid-era spatial planning with restricted and limited the urban development of resources and amenities in township locations. Township area are however high in robberies as indicated by the traditional hotspots, which implies that the facilities may fail to be associated with robberies in these areas. Township areas are generally low income neighbourhoods with numerous informal dwellings, resulting in unregistered informal shops, food outlets, and transport hubs complicating the relationship between facilities and crime.

The raster layers were then stacked on top of each other to form individual independent variables and a linear regression model was then built using the traditional hotspots as the dependent variable and the raster grids as the independent variables. The outcome was a predicted robbery density value for each grid block in the City of Tshwane Metropolitan Municipality.

The results of the day and night time linear regression are shown in Table 11 and Table 12 respectively.

Table 11: Daytime spatial regression results for facility risk

Coefficients:	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	1054.99	39.12	26.97	***
Cadastre Park	8900.99	413.79	21.51	***
Clothing Store	128391.66	3091.43	41.53	***
Convenience Store	-69109.71	1012.13	-68.29	***
Educational Facility	30684.35	874.36	35.09	***
Fast Food Outlet	246696.62	2373.34	103.94	***
Filling Station	37131.86	796.72	46.61	***
Office Parks/Blocks	12248.05	1080.22	11.34	***
Other Stores	63499.4	3377.8	18.8	***
Restaurants	238303	1505.21	158.32	***
Shopping Centres	-63665.06	102.19	-62.41	***
Supermarkets	16015.69	1162.25	13.78	***

Signif. codes: High: 0 '***' Moderate: 0.001 '**' Low: 0.01 '*'

No significance : 0.05 '.' 0.1 ' ' 1

Residual standard error: 12890 on 143146 degrees of freedom

Multiple R-squared: 0.5905,   Adjusted R-squared: 0.5904

F-statistic: 1.876e+04 on 11 and 143146 DF, p-value: < 2.2e-16

Table 12: Night-time spatial regression results for facility risk

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1 243.58	35.53	35.00	***
Cadastre Park	16 025.62	380.59	42.11	***
Clothing Store	80 758.89	2 797.70	28.87	***
Convenience Store	-53 354.81	936.18	-56.99	***
Educational Facility	29 131.93	804.44	36.21	***
Fast Food Outlet	160 637.97	2 191.53	73.30	***
Filling Station	62 804.70	737.29	85.18	***
Office Parks/Blocks	-14 829.58	1 016.19	-14.59	***
Other Stores	-159 594.42	3 110.72	-51.30	***
Restaurants	253 362.38	1 366.51	185.41	***
Shopping Centres	-48 743.96	933.20	-52.23	***
Supermarkets	34 237.12	1 057.38	32.38	***

Signif. codes: High: 0 '***' Moderate: 0.001 '**' Low: 0.01 '*'

No significance : 0.05 '.' 0.1 ' ' 1

Residual standard error: 11820 on 145176 degrees of freedom

Multiple R-squared: 0.4442,   Adjusted R-squared: 0.4441

F-statistic: 1.055e+04 on 11 and 145176 DF, p-value: < 2.2e-16

Both convenience stores and shopping centres were found to have a negative influence on robberies irrespective of time. Relating back to the crime pattern theory, both the latter facilities can be seen as crime attractors, however, convenience stores and shopping centres generally have security guards, surveillance cameras and high foot traffic acting as guardians, which might explain the negative association found here. Office parks/blocks were found to have a positive association with robberies during the day but negative during night. Of course, office areas are busier during the day which not only lead to more possible targets but also strengthens the routine and crime pattern theory stating criminals become aware of crime opportunities while busy with their daily routines. Other stores are the only other variable that has a positive influence on robberies during the day but negative during the night. The types of stores

included in the other store variable are not typically large stores attracting many people at a time and even less so during the night. The rest of the facilities namely cadastre parks, clothing stores, educational facilities, fast food outlets, filling stations, restaurants and supermarkets were all found to have a positive association on robberies irrespective of time of day. From the latter facilities, cadastre parks fast food outlets, filling stations and restaurants are generally known to be open during the day and night-time, attracting people in both time periods. It was however notable that clothing stores, educational facilities and supermarkets had a positive association with robberies during the night, as these facilities normally only trade during the day.

The r-squared values for the day and night-time models were 0.59 and 0.44 respectively which shows that the facility risk model during the day are more likely to predict robbery hotspots than the night-time facility risk model.

The daytime and night-time regression models produced the facility risk values displayed in Figure 28. The facility risk hotspots were divided into 32 classifications using the nested average method, displayed from blue (low facility risk) to red (high facility risk).

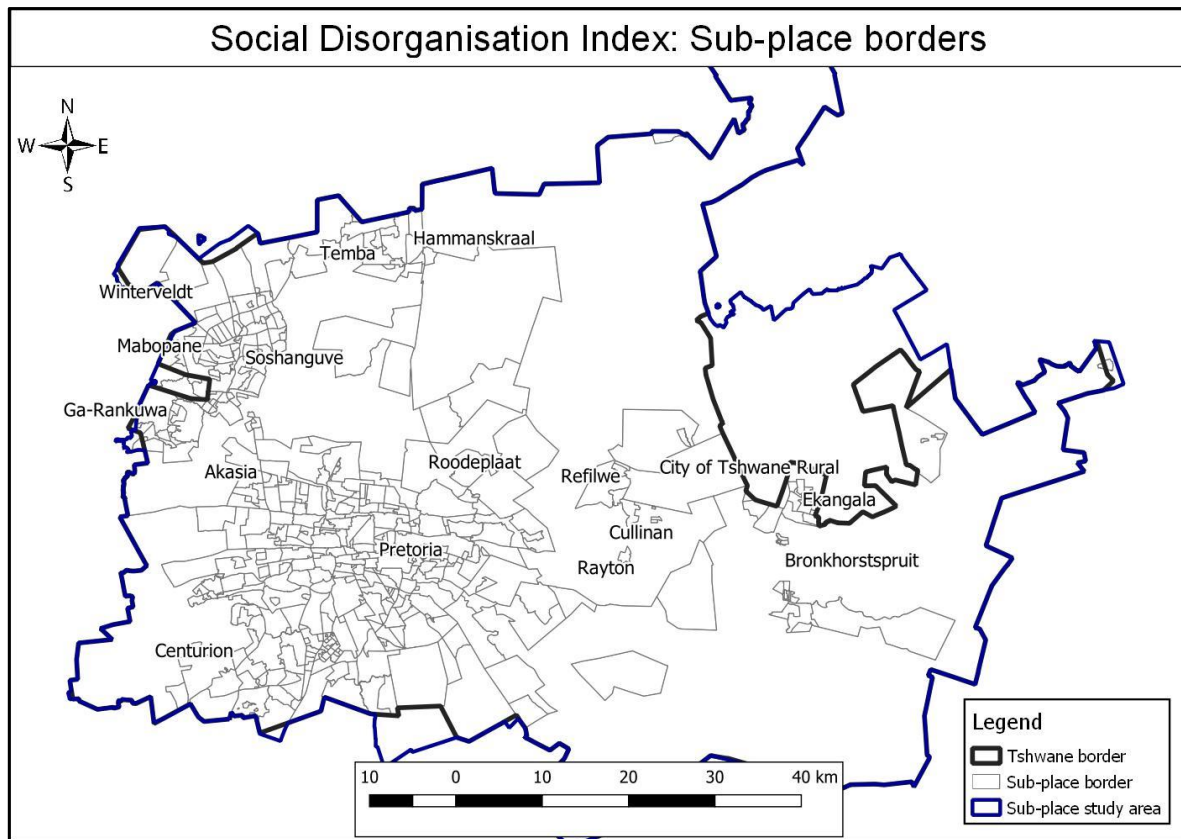**Figure 28:** Daytime and Night-time facility risk
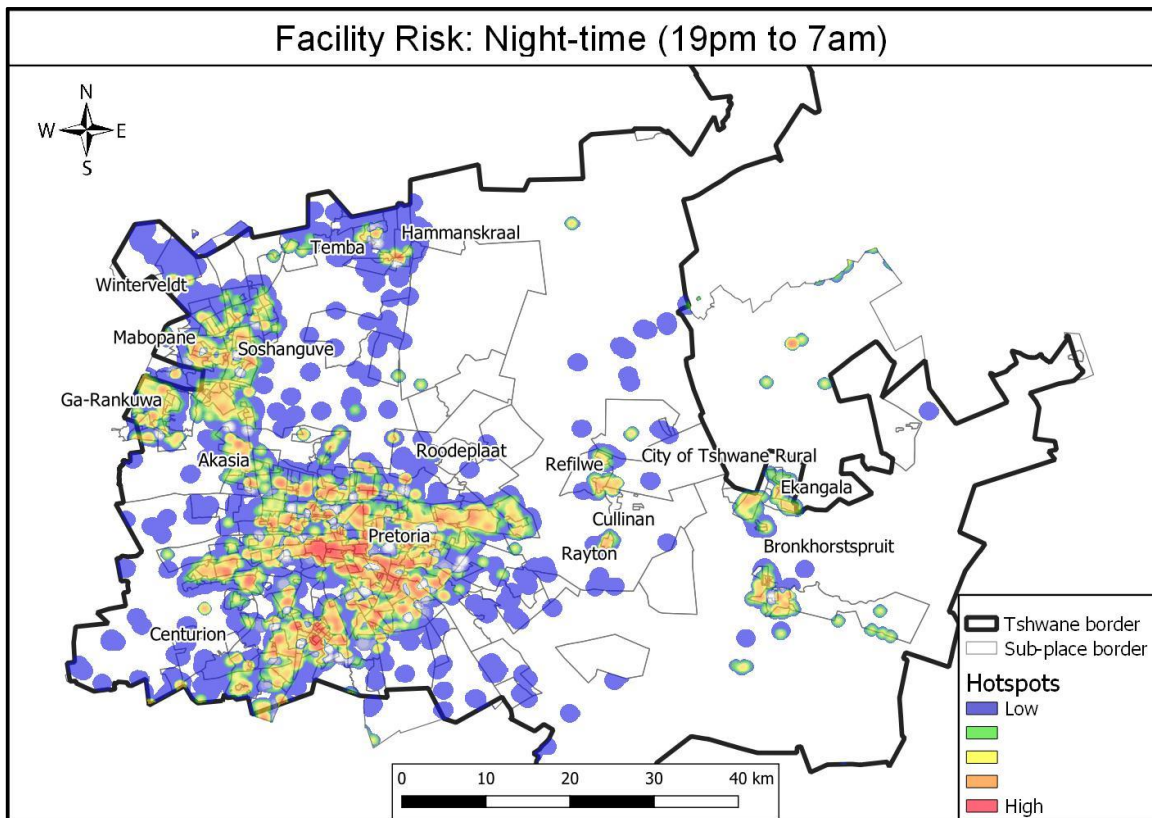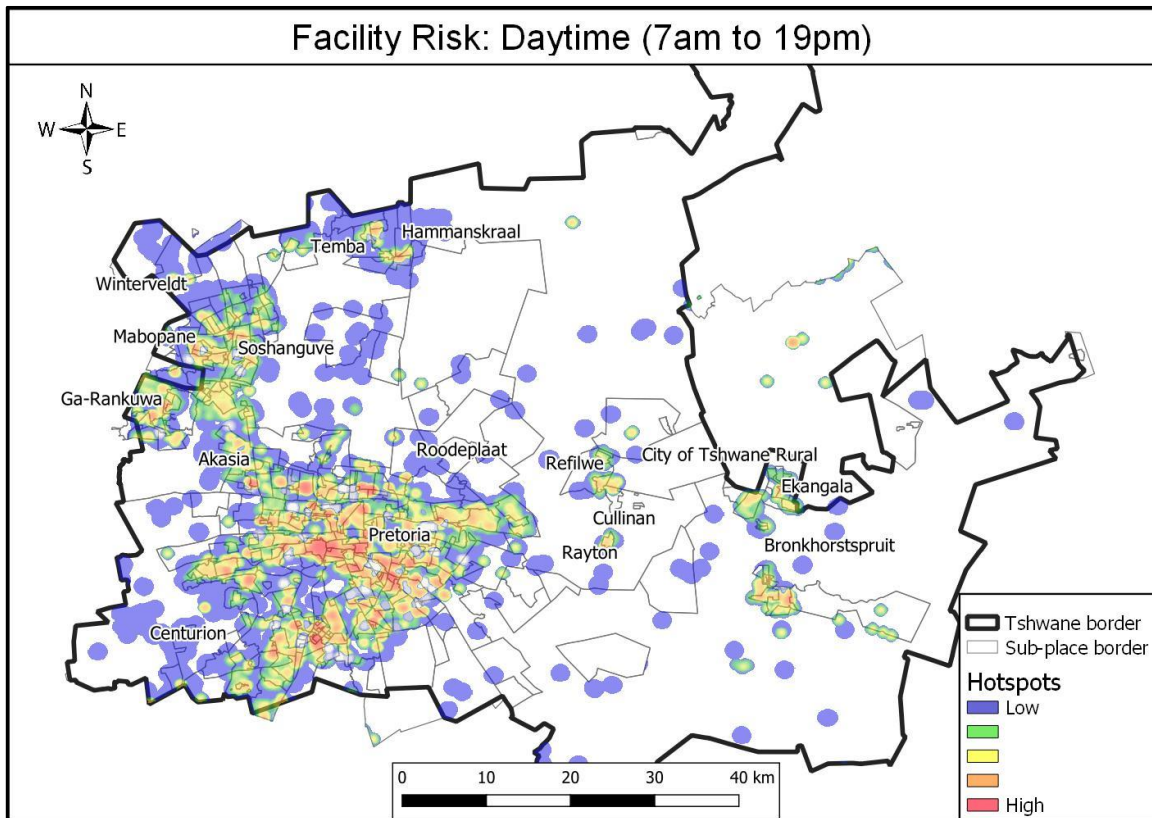
6.3.2 Social disorganisation index results

The social disorganisation index was built using the 2001 Statistics South Africa census data using sub-places as the spatial unit of analysis. Figure 29 below display the sub-place boundaries used in this study.

Figure 29: Study area sub-place border

The sub-place demarcated census variables were used to create a socially disorganised index for each sub-place (see 5.3.4). According to the literature, the more socially disorganised a neighbourhood function in the higher the neighbourhood is acceptable to crime. A map of the index is provided in Figure 30. For display purposes, the social disorganisation index values were divided into 32 classifications using the nested average method, displayed from blue (lowest deprived) to red (highest deprived). The red circles indicate the areas with the highest social disorganisation in the City of Tshwane Metropolitan Municipality. These areas include

Soshanguve, Atteridgeville, Pretoria central business district and Mamelodi, which were also the towns experiencing the most robberies as shown in the traditional hotspot model. Therefore, there seems to be a strong visual association between social disorganisation and robberies.

Figure 31 to Figure 34 provides a zoomed in view of the latter areas depicting the similarities between the areas identified as highly social disorganised and robbery hotspots as created from the 2006 reported robbery locations. In order to compare the maps, the robbery hotspots were also divided into 32 classifications using the nested average method, displayed from blue (low robbery hotspots) to red (high robbery hotspots).
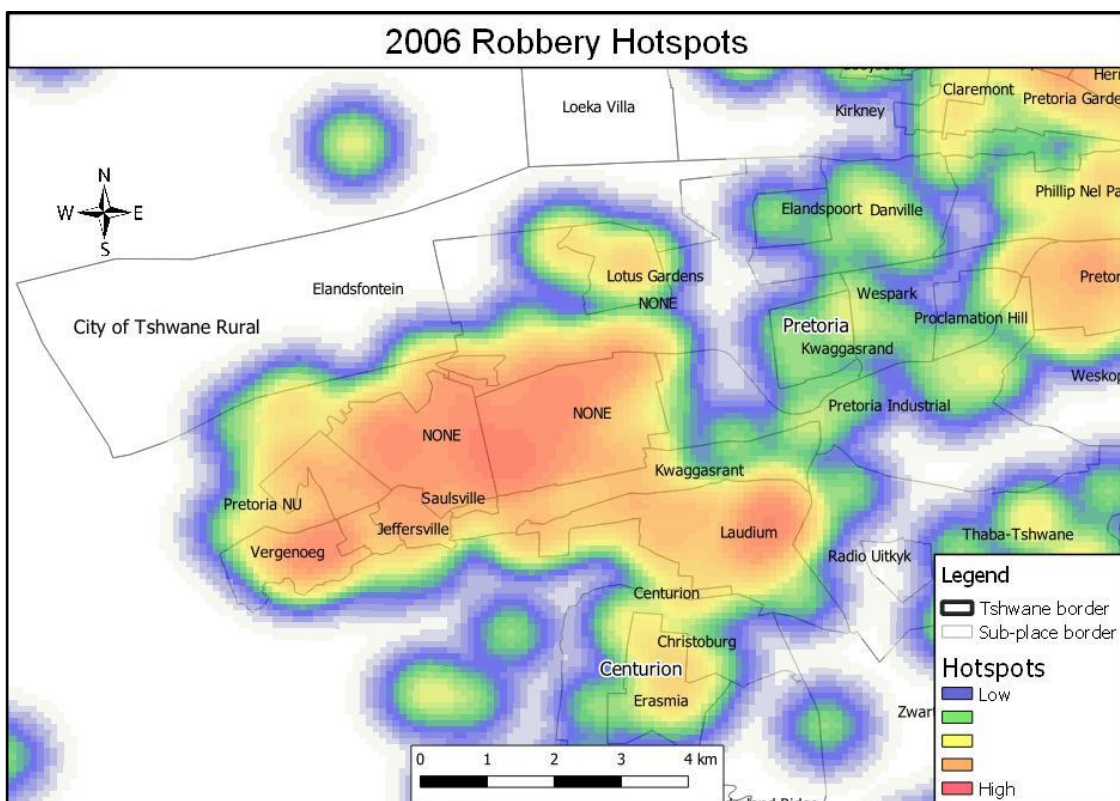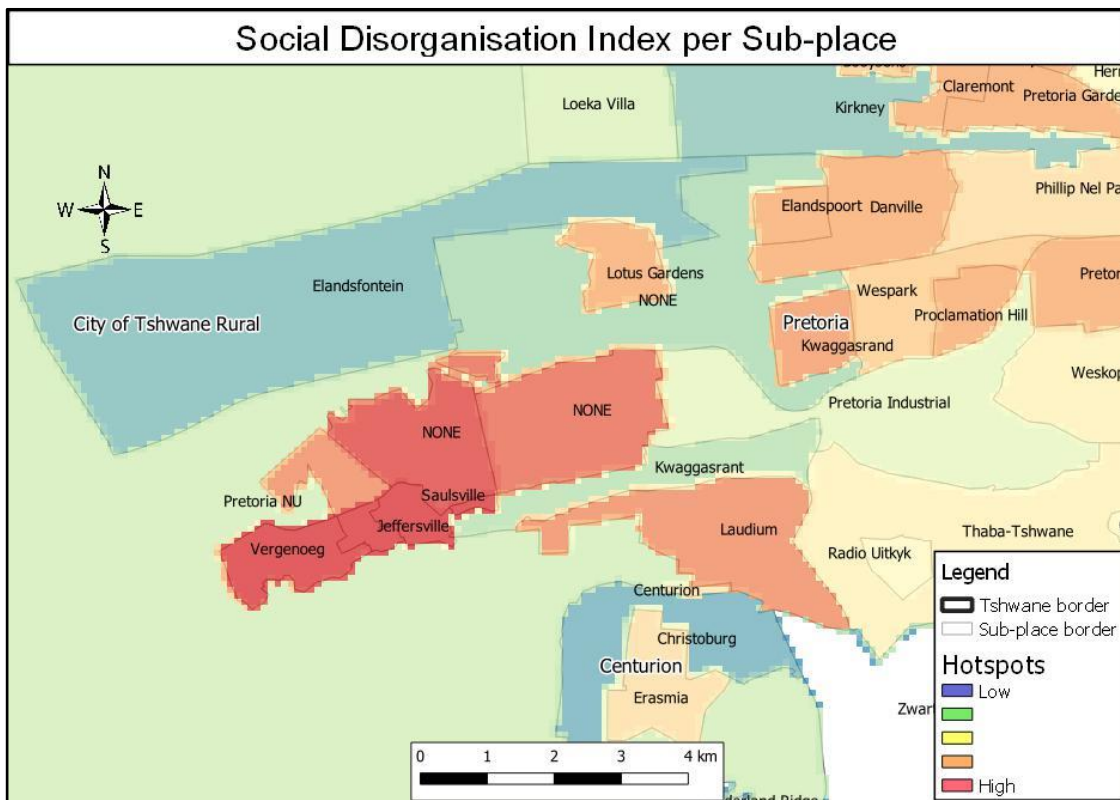
Figure 30: Social disorganisation index per sub-place
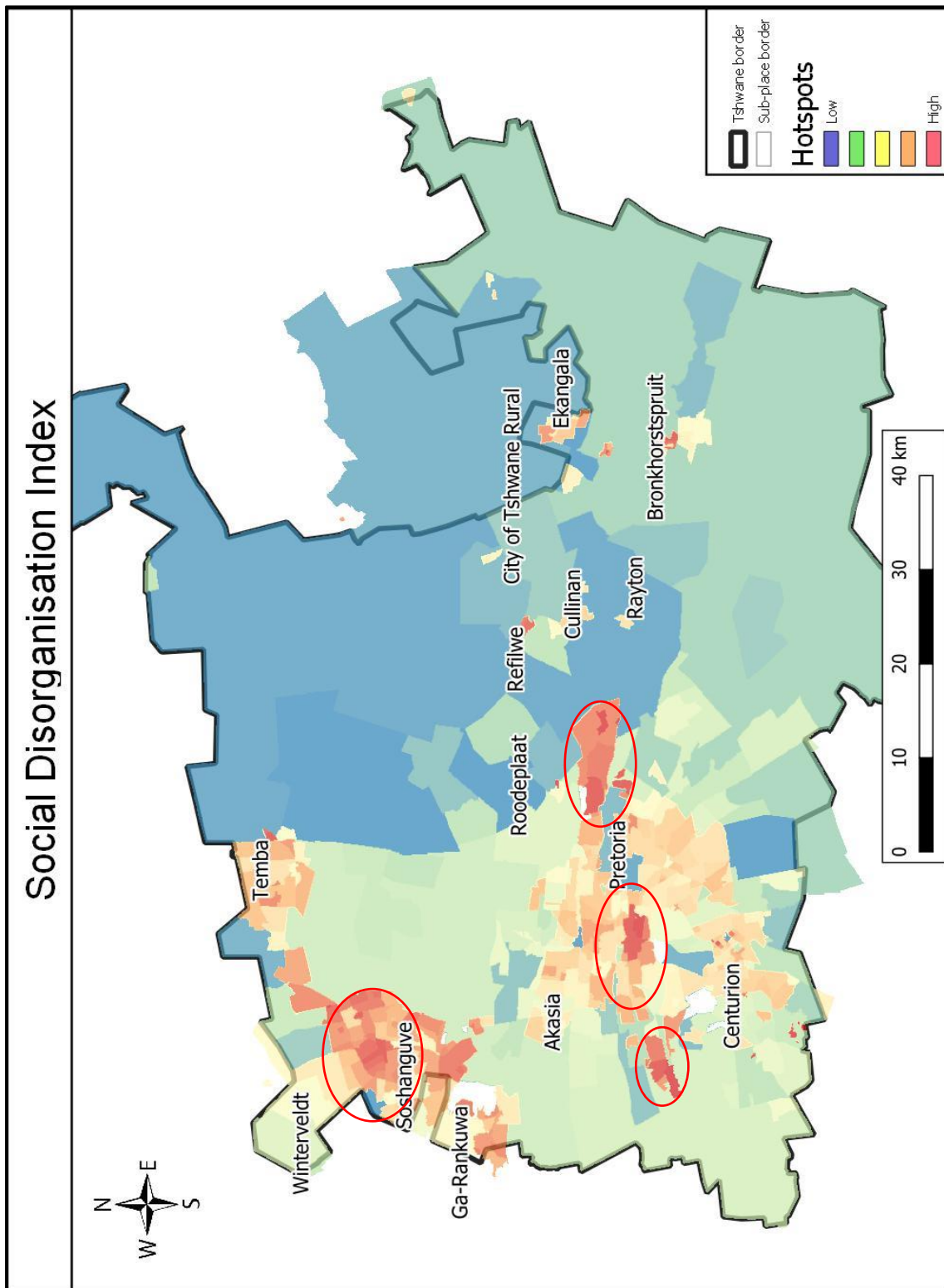
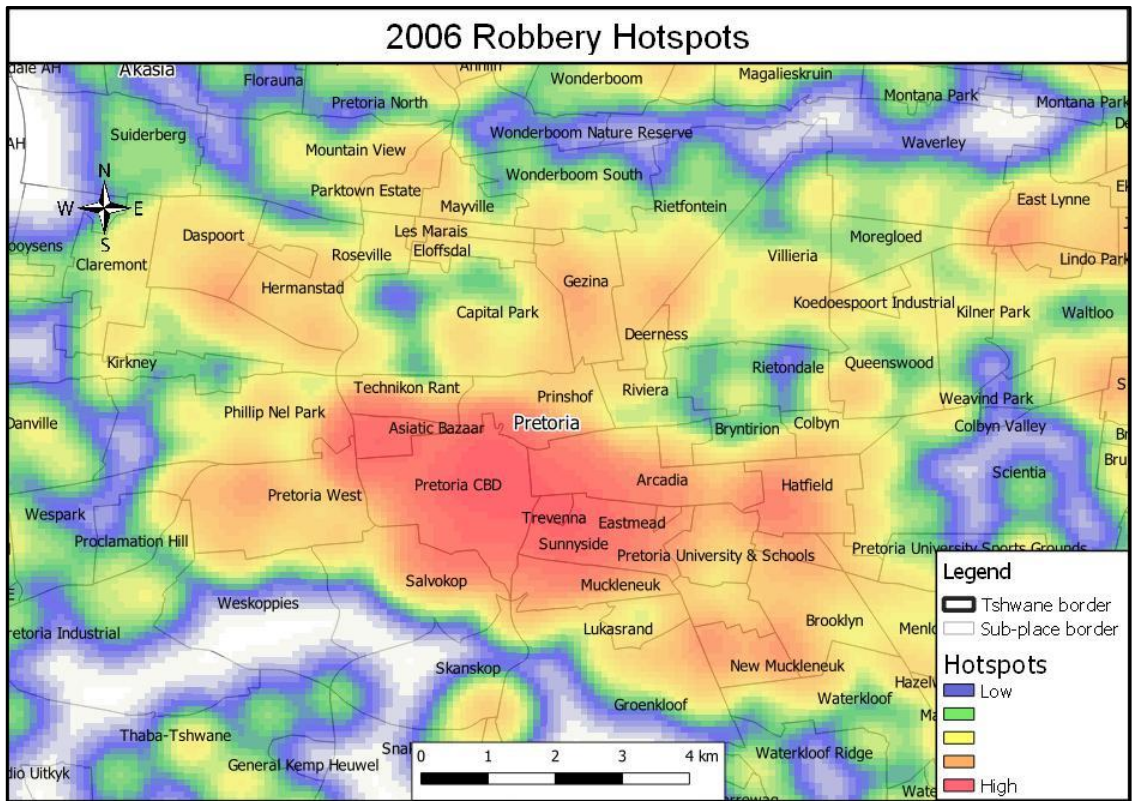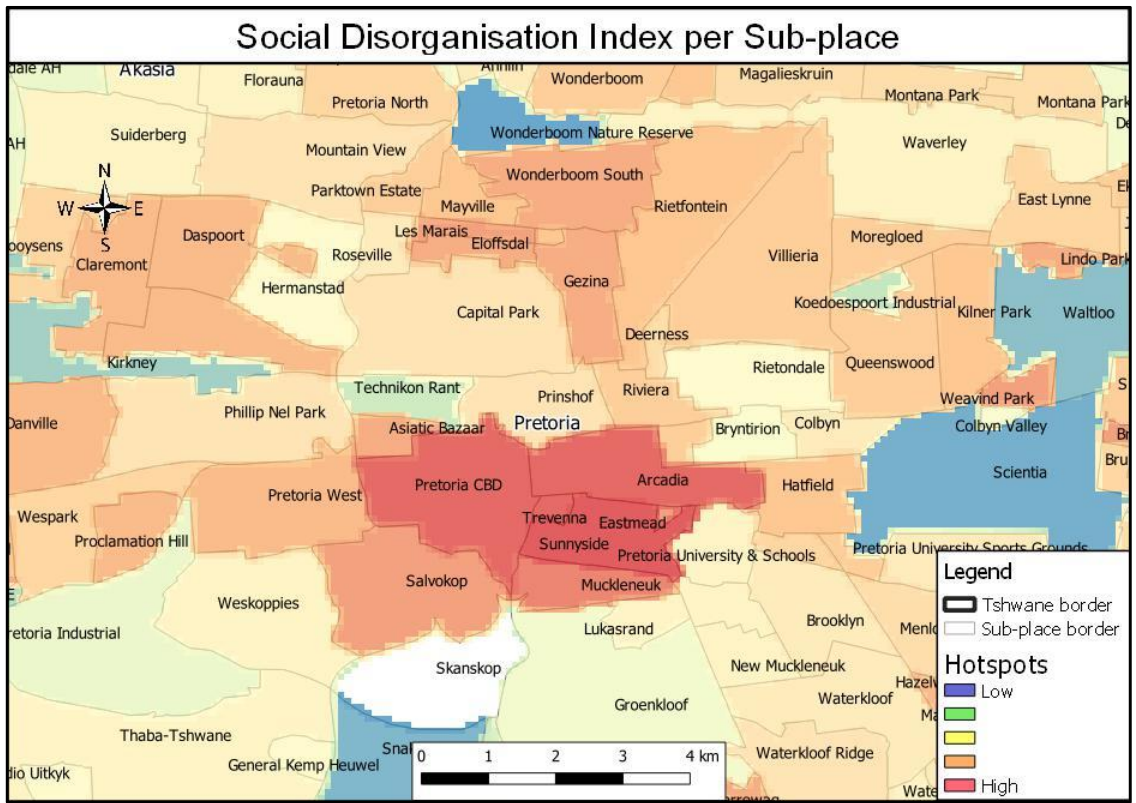**Figure 31:** Soshanguve: social disorganisation index and 2006 robbery
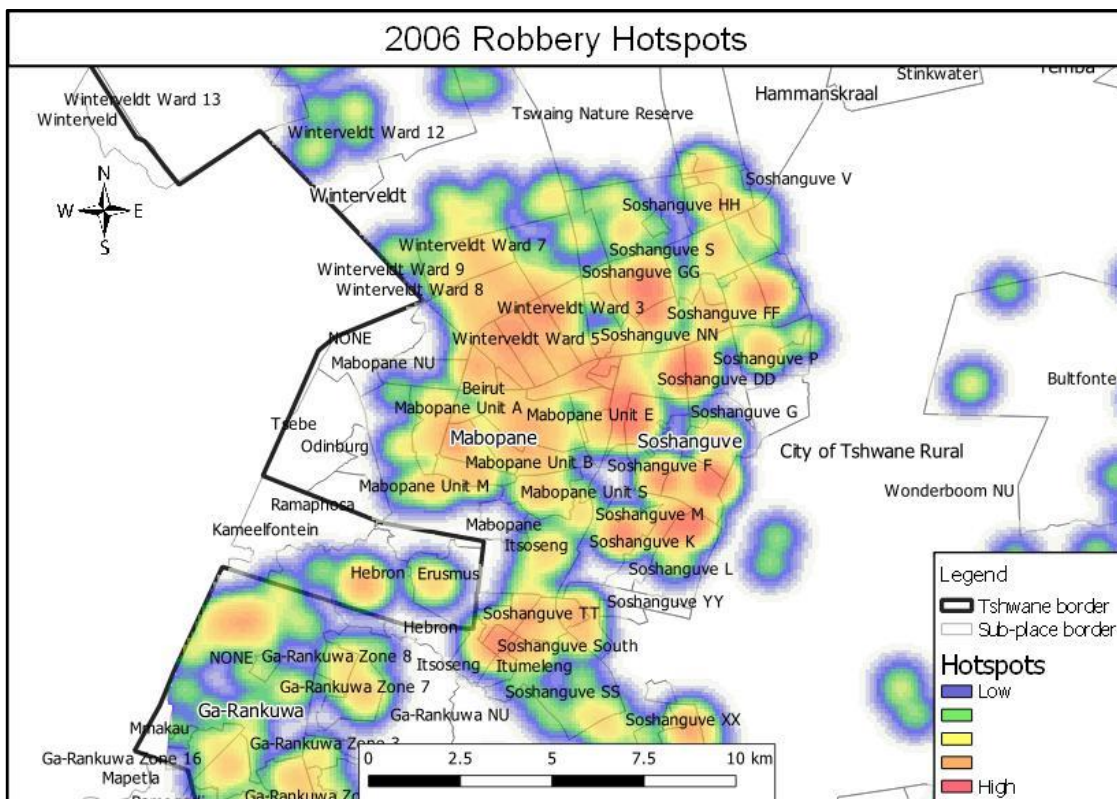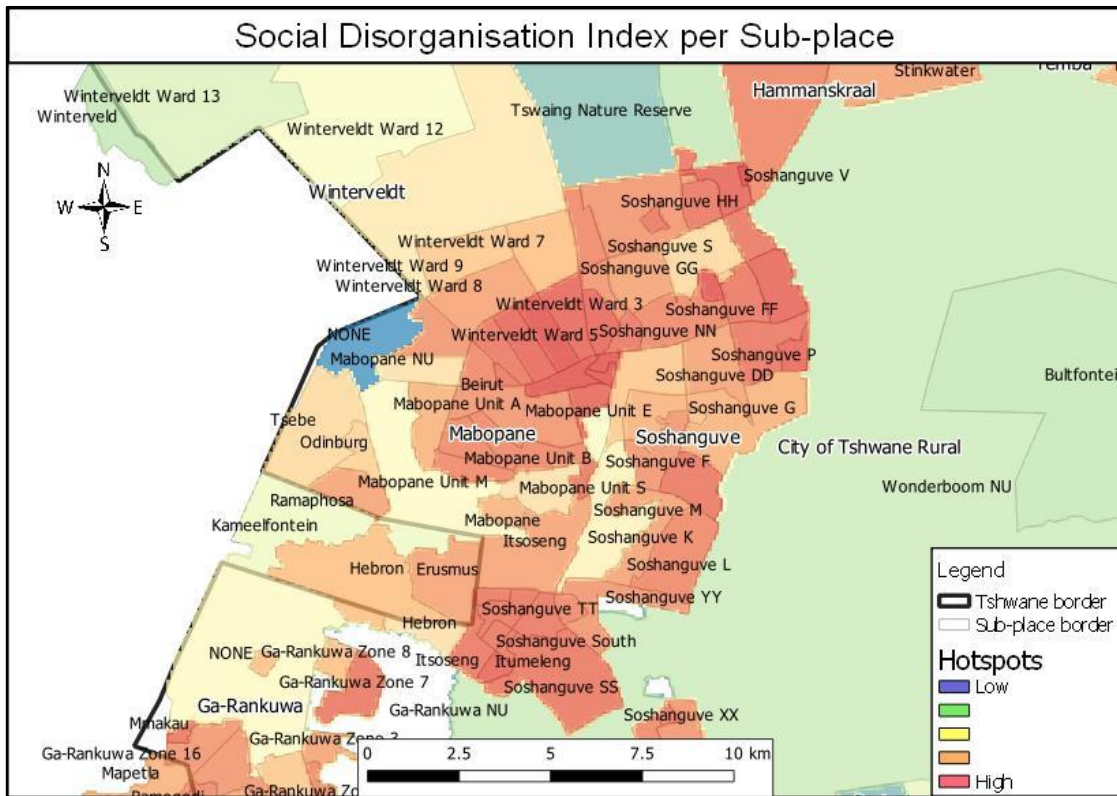
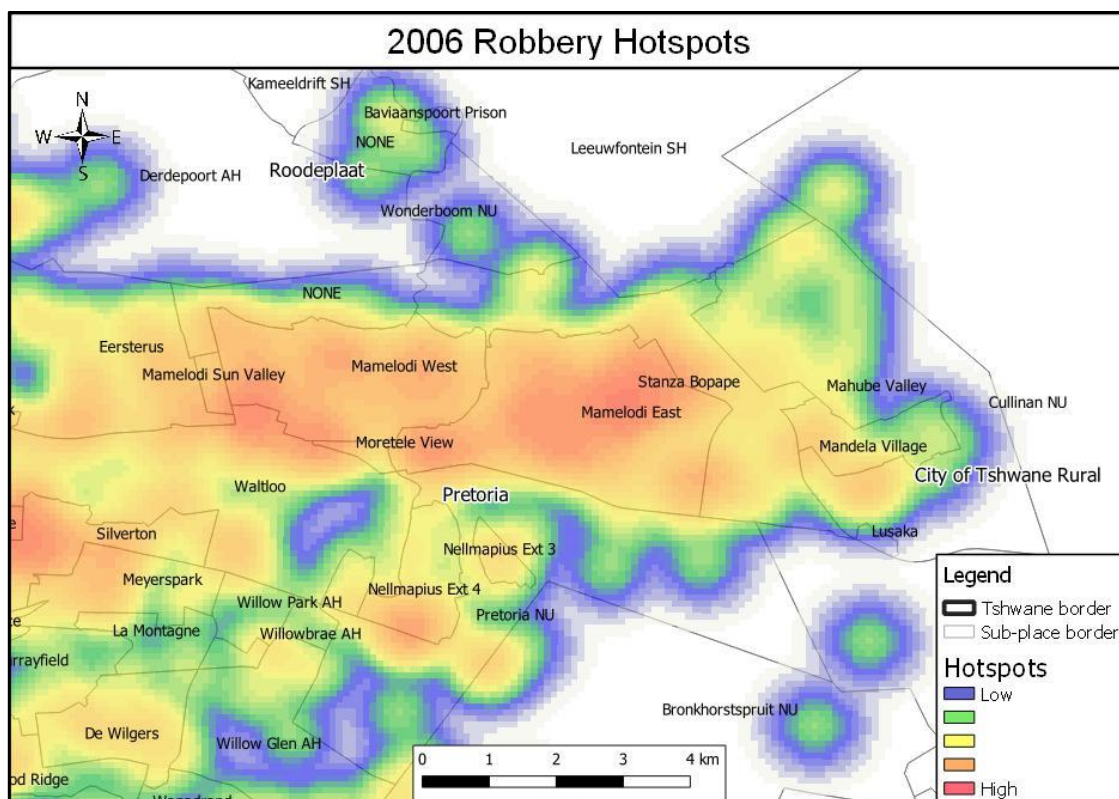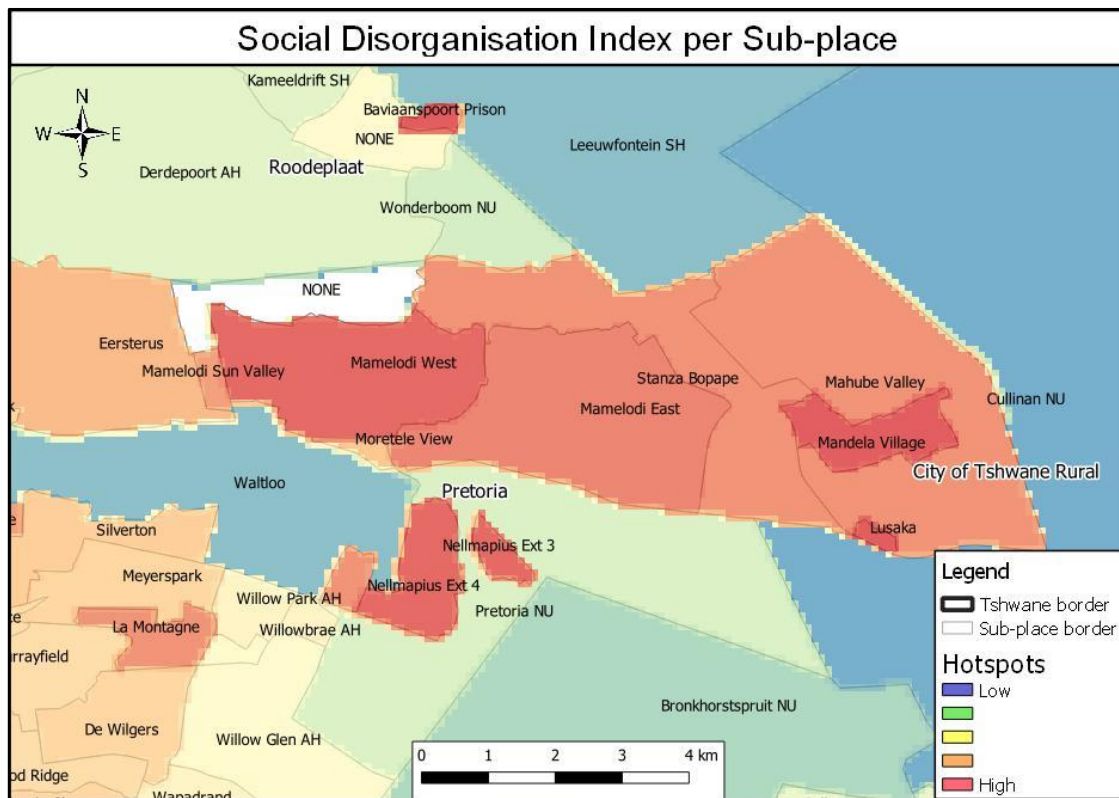**Figure 32:** Atteridgeville: social disorganisation index and 2006 robbery

Figure 33: Pretoria CBD: social disorganisation index and 2006 robbery

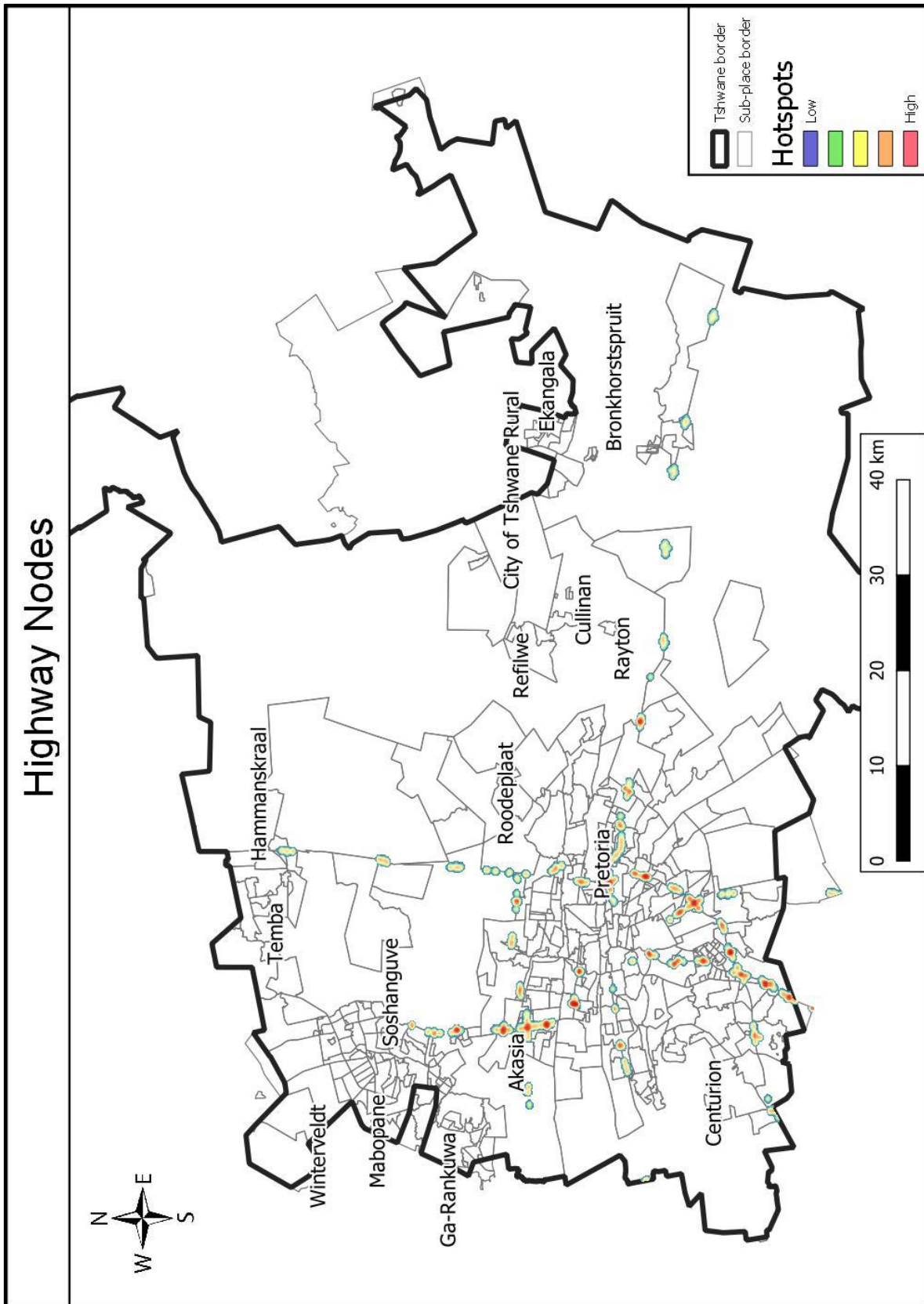**Figure 34:** Mamelodi: social disorganisation index and 2006 robbery

6.3.3 Temporal constraint index

The temporal constraint theory suggest that most crimes happen near traveling nodes because criminals are also constraint by time, leaving them little opportunity to explore areas outside their least distance traveling path. Inspired by the temporal constraint theory, the decision was made to include highways, main roads and train stations in the current study as commuter nodes. Following the same pattern as the facility risk model, the commuter nodes were transformed into hotspots as displayed from Figure 35 to Figure 37 and summed together to form the temporal constraint hotspots shown in Figure 38.

The commuter hotspots shown in the latter maps were created by dividing the commuter density values into 32 classifications using the nested average method, displayed from blue (low commuter density) to red (high commuter density).

Figure 35: Density map of highway nodes

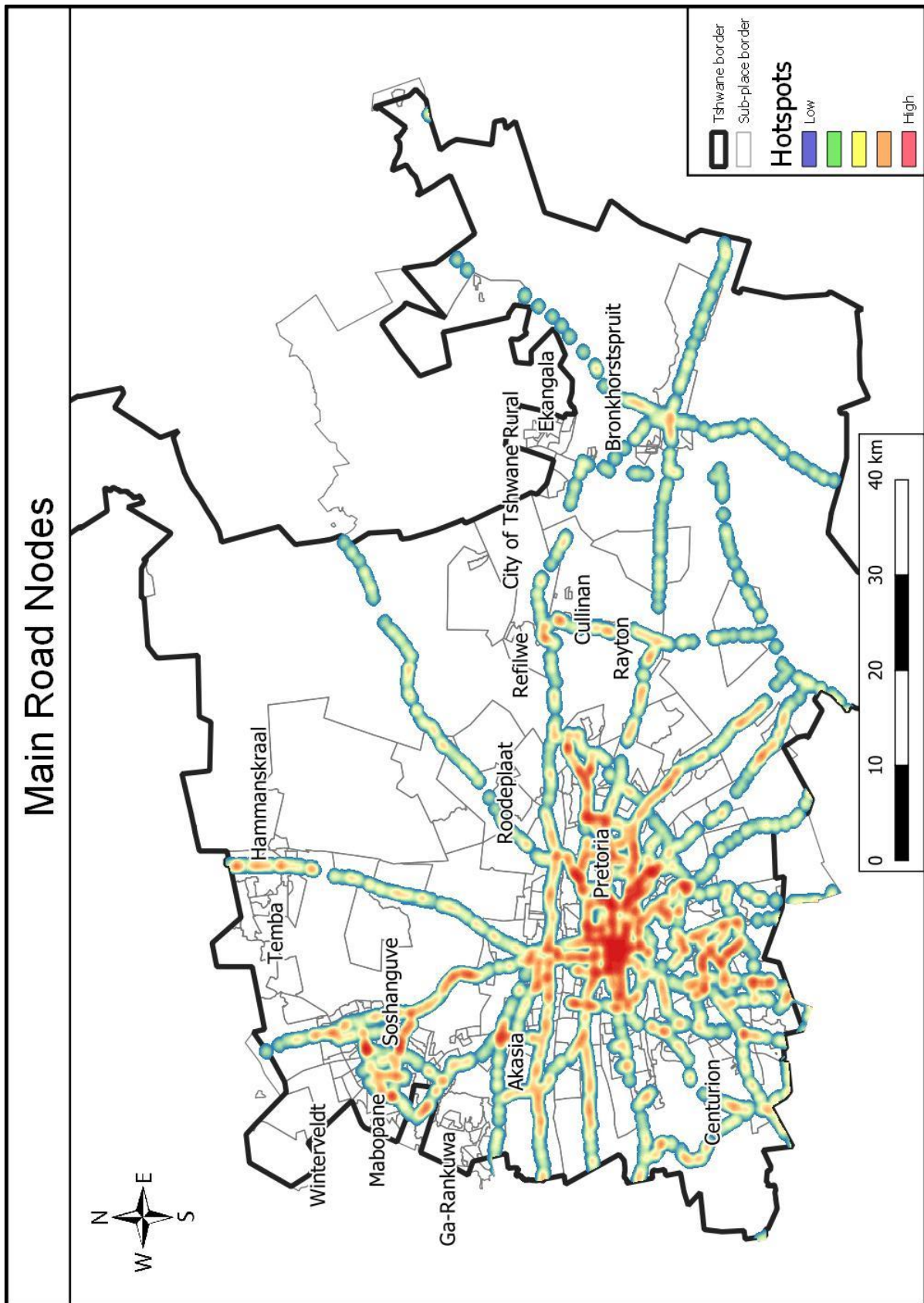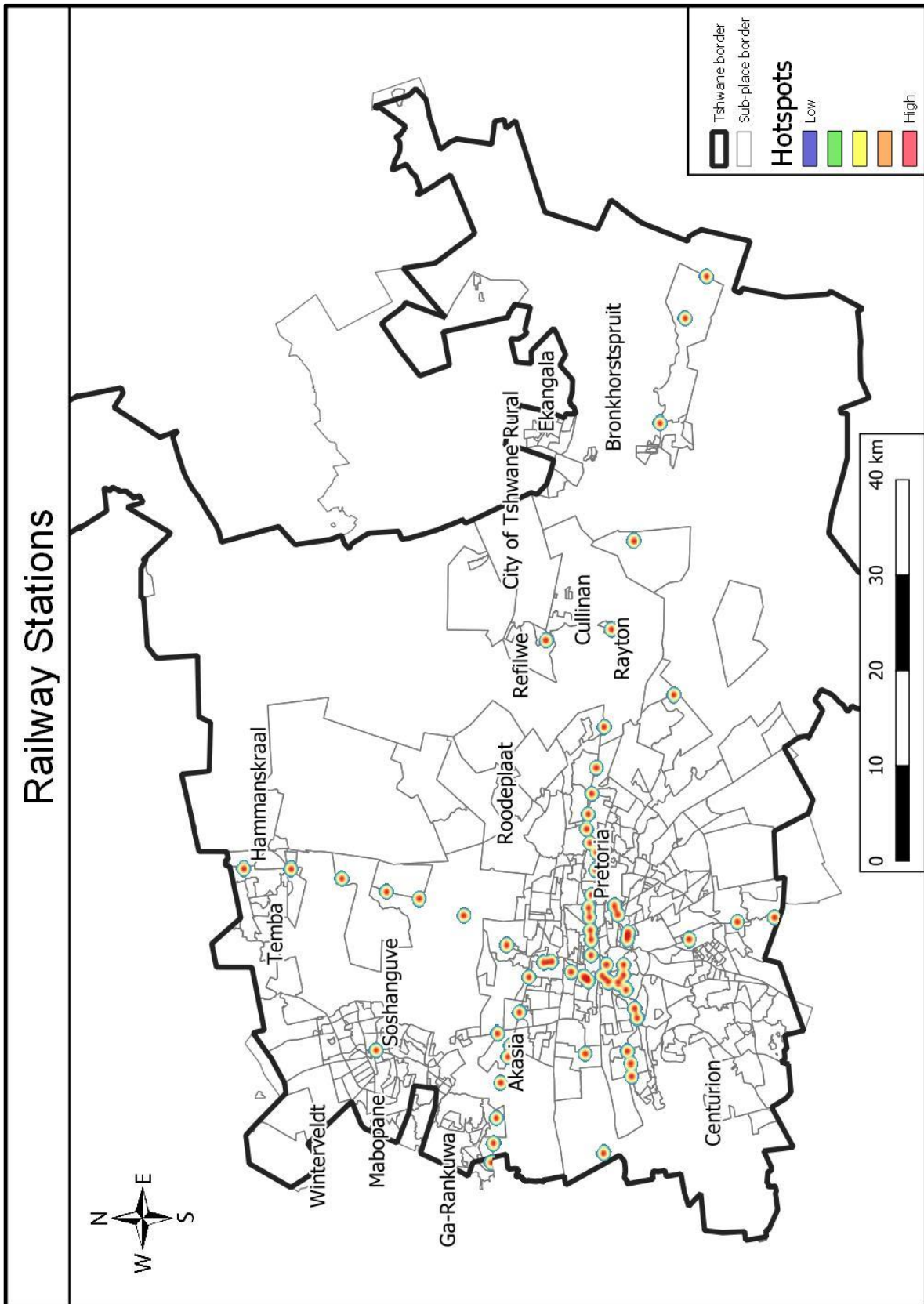**Figure 36:** Density map of main road nodes

Figure 37: Density map of rail stations

Figure 38: Temporal constraint hotspots
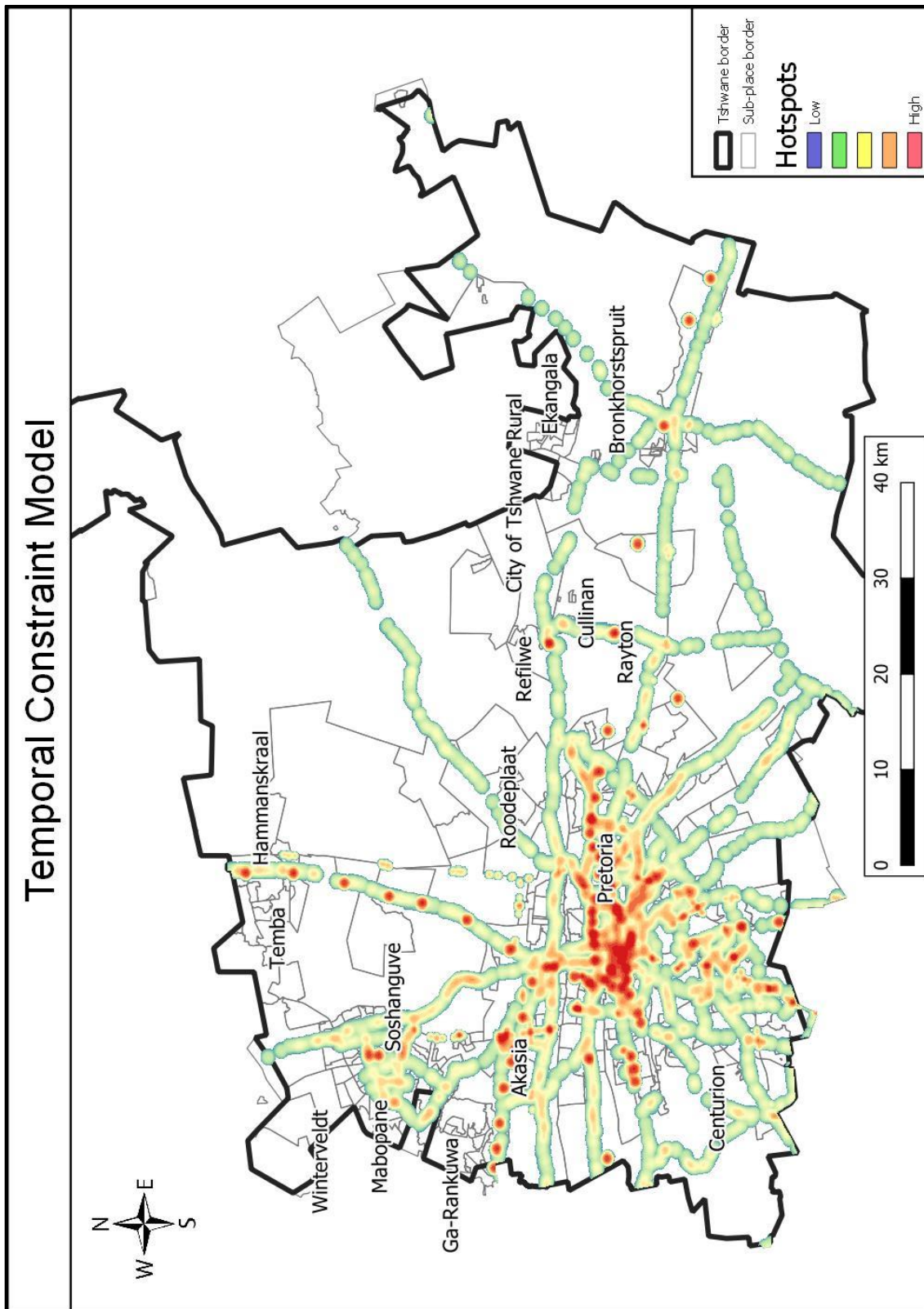
6.3.4 Final robbery probability model

The facility risk, social disorganisation and temporal constraint models were combined to create the final robbery probability model. Combining the latter raster layers created the robbery probability hotspots as displayed for day and night-time in Figure 39 and Figure 40 respectively.

The robbery probability density values were divided into 32 classifications, for both the day and night-time maps, using the nested average method, displayed from blue (low robbery probability) to red (high robbery probability).

Figure 39: Robbery probability hotspot map for daytime
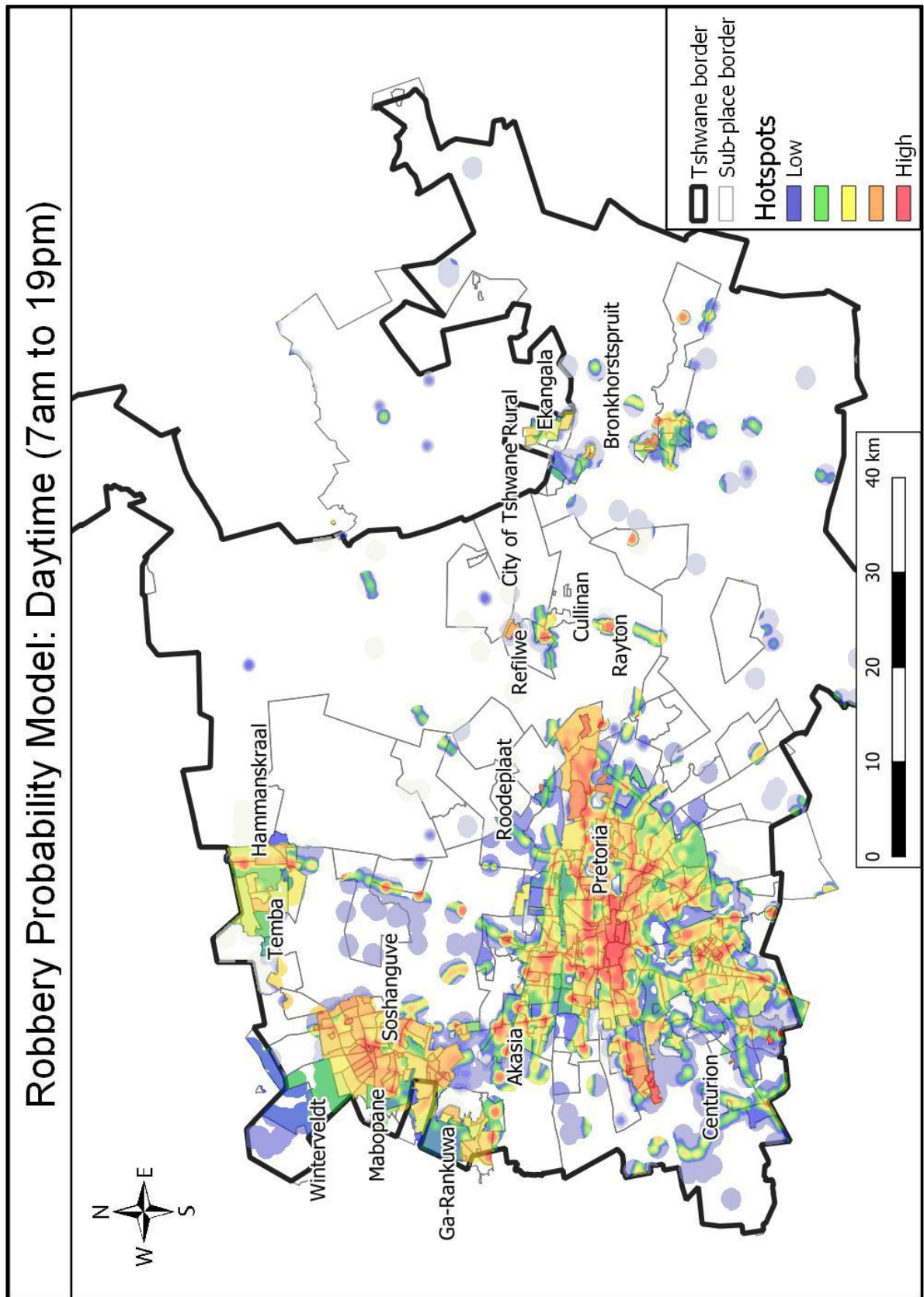
Figure 40: Robbery probability hotspot map for night-time
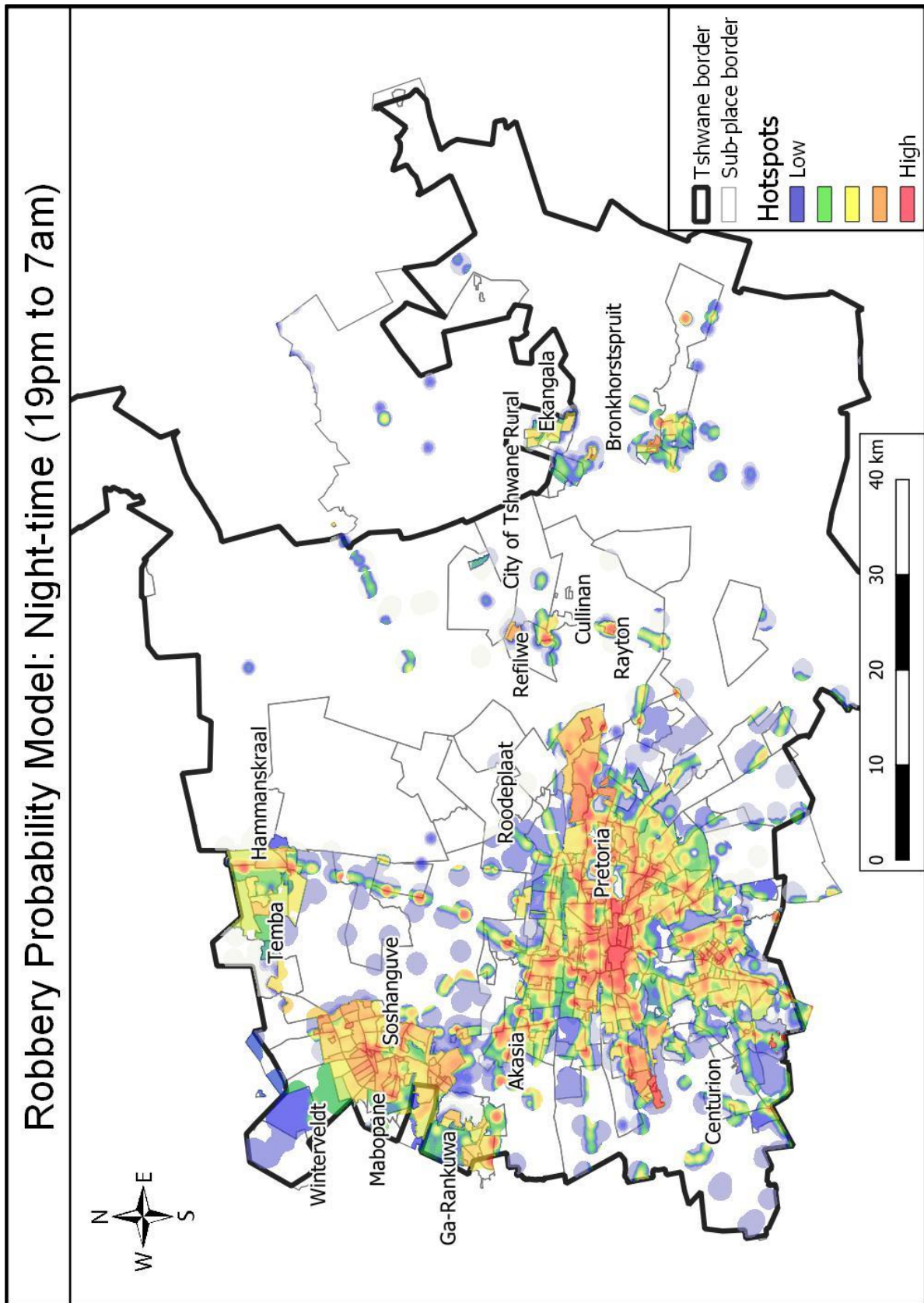
The highest density grid blocks were extracted from the above created robbery prediction model and were used to create and compare the PAI values for day and night-time as well as the PAI values from the traditional hotspots. The extracted three percent areas along with the 2007 robberies falling within and outside these areas are shown below in Figure 41 to Figure 44 followed by the PAI values shown in Table 13.

Figure 41: Daytime robbery probability model hotspot areas

Figure 42: 2007 robberies within the daytime robbery probability model

Robbery Probability Model: Daytime (7am to 19pm)

Figure 43: Night-time robbery probability hotspot areas

Figure 44: 2007 robberies within the night-time robbery probability model

Robbery Probability Model: Night-time (19pm to 7am)

Table 13: Robbery probability daytime and night-time PAI

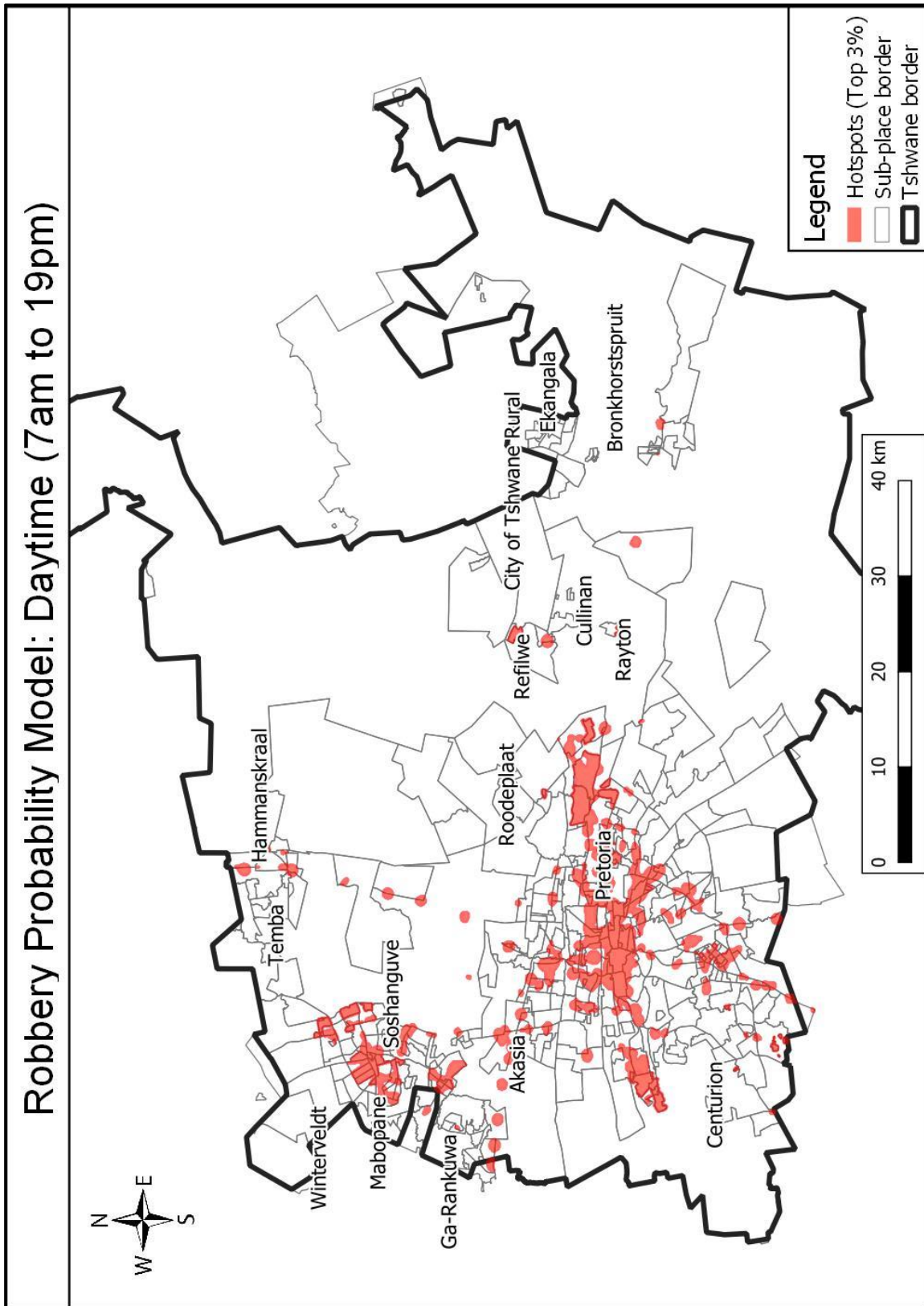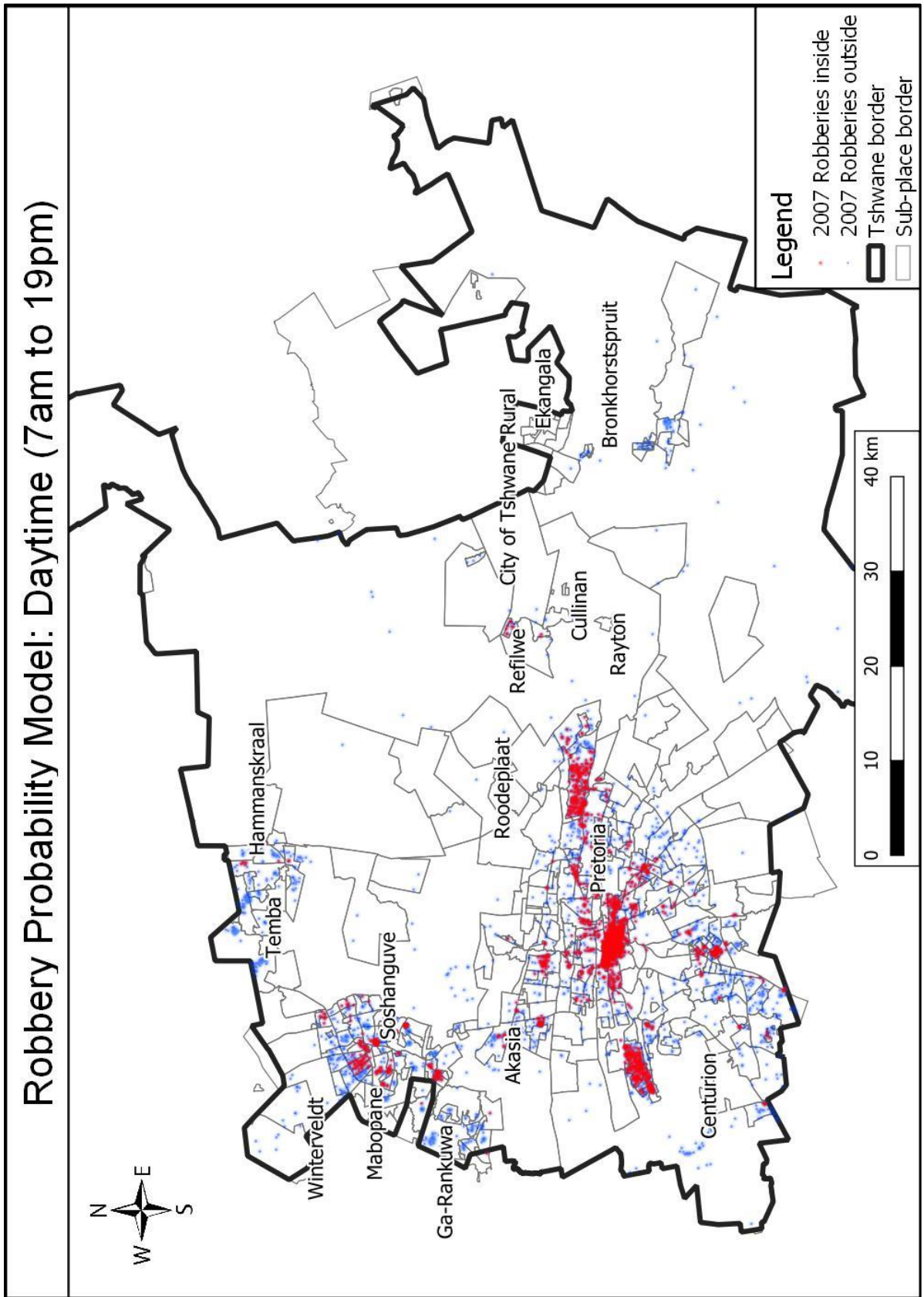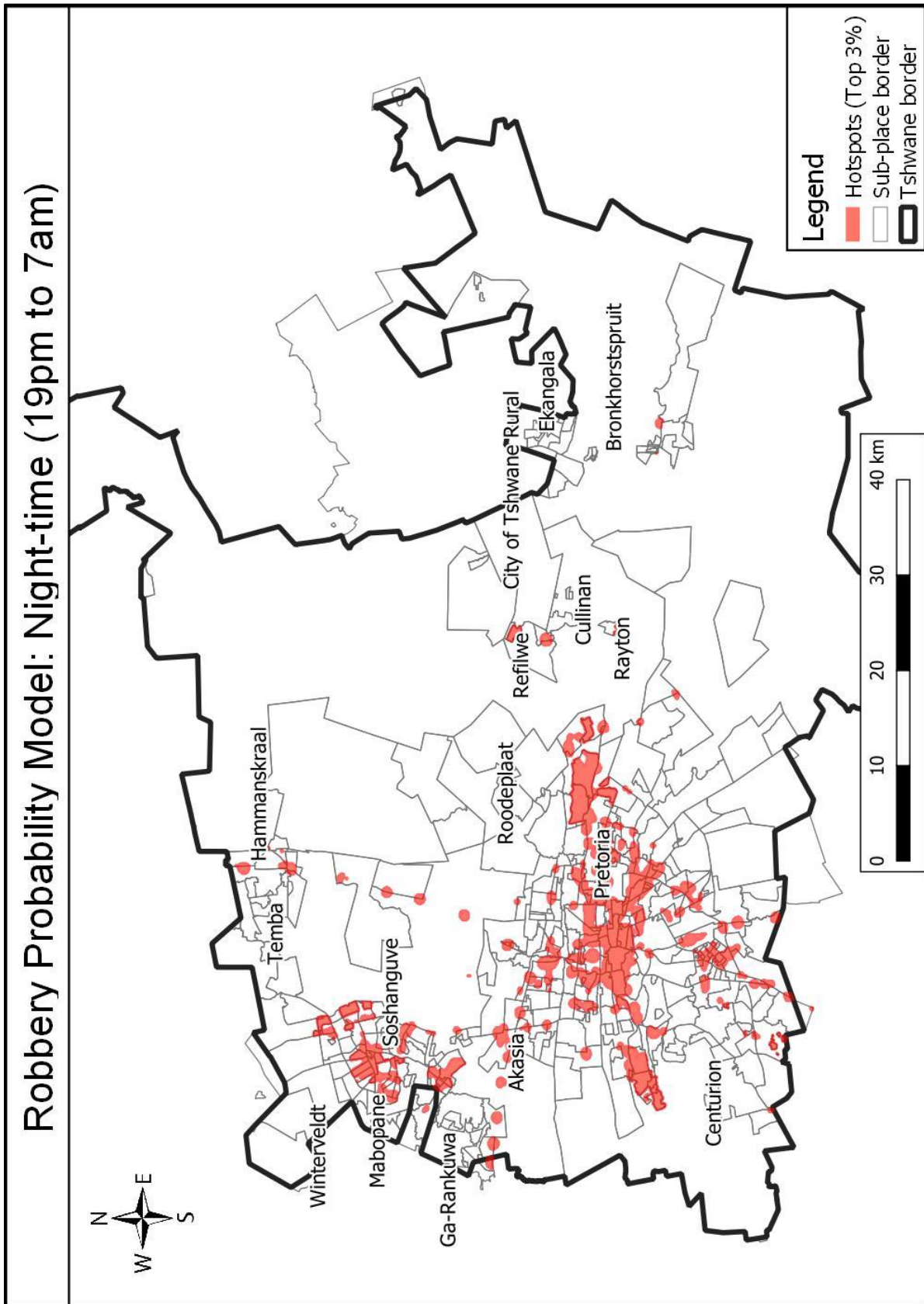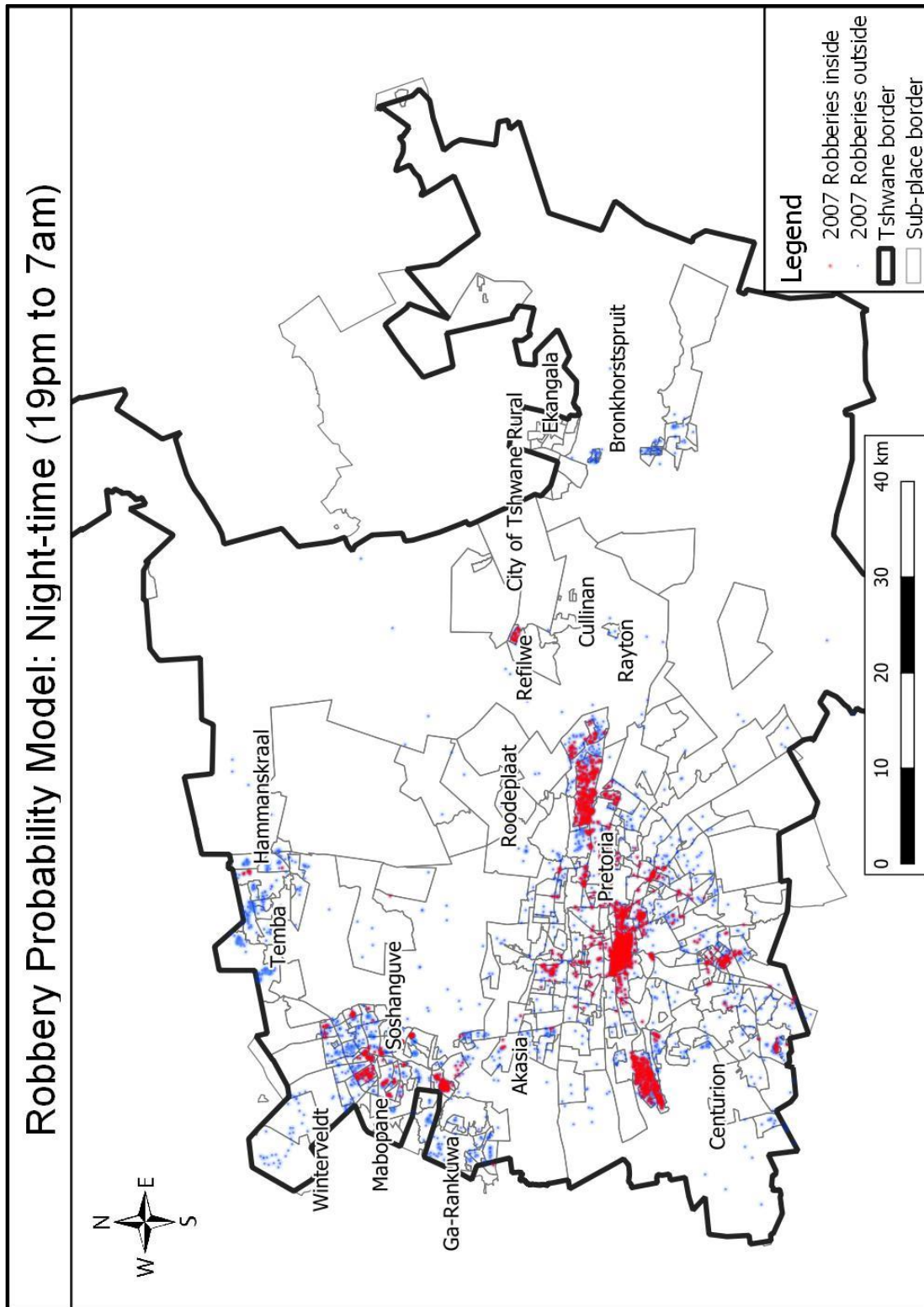| Time off Day	Study Area (km2)	Hotspot Area (Km2)	Hit Rate	PAI
Day	6947.92	208.08	60.51	20.21
Night	6947.92	208.14	57.19	19.10

The hotspot areas for both day and night-time were limited to three percent of the total study area size. This was done to enable the comparison between the traditional and robbery probability hotspots. Table 13 shows that 60.51% and 57.19% of 2007 robberies fell within the top three percent day and night-time robbery probability hotspots respectively, which is indicated by the hit rate field. Dividing the latter predicted 2007 robbery percentages by the three percent of the total study area indicates that the daytime model, with a PAI value of 20.21, performed slightly better than the night-time model with a PAI value of 19.10. The reasons why the daytime performed better than the night-time were thought to be twofold. First, according to the literature, people are more time constraint during working hours, which causes crime to occur closer to commuter nodes during the day than during night. Second, the facility regression model performed better during the day, indicating that more robberies were accurately predicted by facilities during the day than during the night.

6.4 Traditional hotspots compared to robbery probability

The main aim of the current study was to develop and compare a traditional hotspot model against a new robbery probability model used to predict robberies in the City of Tshwane Metropolitan Municipality. This section therefore answers the main aim by comparing the results shown for the traditional and robbery probability models. This section focusses on which model performs the best for daytime and which model for night-time. Table 14 and Table 15 below show results of the PAI analysis for the day and night-time results of the models respectively.

Table 14: Robbery prediction model PAI comparison for daytime

Model name	Study Area (km2)	Hotspot Area (Km2)	Hit Rate	PAI
Robbery probability	6947.92	208.08	60.51	20.21
Traditional hotspots	6947.92	208.14	72.05	24.06

Table 15: Robbery prediction model PAI comparison for night-time

Model name	Study Area (km2)	Hotspot Area (Km2)	Hit Rate	PAI
Robbery probability	6947.92	208.08	57.19	19.10
Traditional hotspots	6947.92	208.14	73.09	24.40

It is evident from the PAI values in Table 14 and Table 15 that the traditional model outperforms the robbery probability model both in the day and night time. Using the traditional hotspot model in the City of Tshwane Metropolitan Municipality will have a 10 percent higher prediction rate than the robbery probability model in both the day and night-time. Therefore the conclusion can be made that even though there is evidence in favour of the spatio-temporal crime theories, a higher percent of robberies are predicted by the traditional hotspots meaning that robberies tend to cluster in the same areas year on year

6.5 Results summary

Both the traditional and facility risk model had much higher prediction accuracy during the day than during the night.

During the process of building the traditional hotspot model there were a few interesting findings. First, the results showed that over 70 percent of robberies

accumulate in three percent of the City of Tshwane Metropolitan Municipality's area, showing the strength of hotpots created from historical crime events to predict future robbery occurrences. Further results indicated that not only was the Pretoria central business district, the northern townships of Soshanguwe, Mamelodi in the east and Atteridgeville in the west historically the location of robbery hotspots but these areas will continue to be hotspots in the future.

In conclusion, it was found that for the City of Tshwane Metropolitan Municipality it is best to use the traditional hotspot model to predict future robbery occurrences.

CHAPTER 7: DISCUSSION AND CONCLUSION

7.1 Overview

This study introduced a novel model which used the physical and sociological environment to predict robberies, mapped as kernel density hotspots. This model was compared to the traditional hotspot model which used previously known robbery hotspots as robbery prediction areas.

The following objectives were identified in this study:

1) To outline various spatial and temporal crime theories that have been used to explain the crime.

2) To identify a number of geospatial techniques that is used to analyse crime.

3) To identify daytime and night-time robbery hotspots using 2006 robberies occurrences.

4) To develop a daytime and night-time robbery prediction model using variables and features informed by various spatial theories of crime.

5) To compare the output of the robbery probability model with the output of a traditional crime hotspot model in order to propose a best-fit model for the City of Tshwane Metropolitan Municipality during the day and during the night.

In this section the outcome of these five objectives are summarised using the results and findings of the analyses undertaken and ends with further research and further studies needed to build upon this study.

7.2 Spatial and temporal crime theories.

The methods followed in this study were loosely based on four main spatial-temporal crime theories: social disorganisation, routine activity, crime pattern and temporal constraint theory. The social disorganisation theory focuses on the social composition

of neighbourhoods in order to identify characteristics that increase the risk of crime. Where the social disorganisation theory concentrates on the inner workings of communities, the routine activity and crime pattern theories are more interested in the movement patterns of individuals. The temporal constraint theory is built on the notion that an individual's movements are restricted by the time available while commuting.

Aspects of these spatial-temporal theories were combined in the current study and formed the basis of the facility risk, social disorganisation and commuter models. The facility risk model was built mainly on facilities chosen on the assumption that these facilities may contribute to robberies according to the central tenets of the crime pattern theory. The facilities are also routinely visited by residents and act as nodes. The social disorganisation theory proved useful with the selection of demographic data to include in the analysis. The impacts of temporal crime theories in the current study were twofold. First, the time constraint theory inspired the inclusion of commuter nodes as a confined space in which robberies occur. Second, it provided evidence from previous studies (Breetzke (2016); Suffla and Seedat (2016)) that splitting crime data into day and night-time proved to be vital in a South African context.

The spatio-temporal theories discussed in this study inspired the creation of the robbery probability model. The theories proved to be effective in the City of Tshwane Metropolitan Municipality as the robbery probability model predicted over 55 percent of robberies for both the day and night-time timeslots in only three percent of the study area. There is however more research needed regarding the weighting of social disorganised variables, facilities such as alcohol outlets, bars and commuter nodes such as bus stops and taxi ranks which were not available for the use in this study. The daytime model also outperformed the night-time model; the reason for this is thought to be because people are generally not as time constraint during the day spreading crime over a farther distance from commuter nodes. Also, some of the facilities chosen based on the routine activity and crime pattern theory are more active during the day leading to a higher concentration of crime around those facilities during the day than during the night.

7.3 Geospatial crime techniques

Place-based crime has been studied over several decades. However, new computer capabilities, and more specific GIS, have brought new insight and spurred numerous geospatial-based crime studies. GIS software is powerful in storing, managing and handling spatial data. GIS makes it possible to detect geospatial crime patterns in the physical environment. A combination of visually displaying and the capability of detecting patterns had brought forth some exiting spatial analysis techniques useful in crime analysis.

This study discussed two spatial techniques namely, kernel density hotspots and RTM. Kernel density estimation methods were used to create the traditional hotspot model by generating hotspots from actual robbery locations as well as to generate hotspot maps for each feature utilised in the robbery probability model. The latter hotspots were stacked together using a similar technique used in RTM.

The current study used the programming language R as the platform in which the spatial techniques were executed. R is freely available and was mainly developed for statistical purposes. However, numerous spatial techniques have been incorporated into the R framework, making it possible to run the spatial and statistical processes discussed in the current study programmatically.

7.4 Daytime and night-time 2006 robbery hotspots

The current study made use of KDE to predict future crime occurrences (see (Liu and Brown (2003); Weisburd and Eck (2004); Braga and Bond (2008); Chainey et al. (2008); Ratcliffe et al. (2011); Wain et al. (2017)). The traditional hotspots model was created using 2006 robbery occurrences and then validated based on the PAI value of 2007 robbery occurrences. The current study found that for both the day- and night-time traditional hotspots, the hit rate was above 70 percent. The hit rate was calculated as the percentage of 2007 robberies that fell within the traditional hotspots created from the 2006 robberies. Furthermore, the latter hotspots were limited to cover only three percent of the total study area. The results of the study suggested that the police

service can target 70 percent of robberies occurring in the City of Tshwane Metropolitan Municipality by focussing on only the percent of the study area.

7.5   The daytime and night-time robbery prediction model.

Research on routine activity and crime pattern theory has most often found positive associations between facilities and different types of crime. In the current study it was necessary to find facilities that have a spatial association with robberies and to investigate how well those facilities predict future robbery hotspots.

The results of the regression analysis found that during any time of the day facilities such as cadastre parks, clothing stores, educational facilities, fast food outlets, filling stations, restaurants and supermarkets are positively associated with robbery while convenience stores and shopping centres show a negative association. Office parks and other stores had both a positive association with robberies during daytime but negative during the night. With regards to performance, the linear regression model had a fairly good adjusted r-squared value of 0.59 during the day but a moderate to weak 0.44 for the night-time model. This implies that facilities are estimated to explain 59 and 44 percent of the variation of robbery density values during the day and night-time respectively. This is supported by the crime pattern and routine activity theories which suggests that people become aware of crime opportunities while conducting their daily routines at facilities. The current study uses facilities seen as crime attractors or generators by the crime pattern theory. The general assumption is that most of the latter facilities are only operational during the day and therefore attracting less people at night-time.

It was subsequently found that the facility model (as created in section 5.3.3) explained 51.92 percent and 44.2 percent of the variation of daytime and night-time robberies respectively. As expected by the regression model, the daytime proved to be more accurate than the night-time, but even so, facilities still predicted a fairly large percentage of robberies during the night. Therefore, facilities play a definite role in robbery locations within the City of Tshwane Metropolitan Municipality accompanied by evidence to support the routine activity and crime pattern theories.

In terms of the social disorganisation theory, previous studies have found positive associations between numerous social disorganisation measures and violent crime (see Breetzke, 2010a). The current study builds on Breetzke's findings by utilising similar measures into the development of a social disorganisation index. The social disorganisation measures included unemployment population, people migrated in the past five years, people with marital status as divorced, male population between the age of 15 and 34 and people born in other African countries. The study further included living environment deprivation variables itemised by Noble and Wright (2013), which included households without a pit latrine with ventilation or flush toilet, households without use of electricity for lighting, households without piped water within at least 200 meters and shack as dwelling type.

To test the temporal constraint theory, the current study included highway nodes, main road nodes and railway stations. The latter variables were chosen on the assumption that these nodes are visited daily by a variety of people and according to the temporal constraint theory, criminals have limited time to explore areas outside their least distance traveling path which most likely are connected by the latter nodes. Transforming the commuter nodes into kernel density estimations created hotspots which indicate the likelihood of a robbery taking place based on the closeness to a traveling node. The closer to a traveling node, the higher the risk of robbery.

It was found that almost 50 percent of robberies are being predicted during the day but only 38.69 percent during the night using the commuter model. The drop in more than 10 percent predicted robberies during the night is further evidence of the temporal constraint theory when considering that people generally have less to do during the night and therefore are not as time constrained as during the day.

The current study combined the spatio-temporal crime theories into a single robbery probability model. This was done by ensuring the outputs of all the models conformed to the exact same raster dimensions, making it possible to add the values together. The output of the robbery probability model was a new raster grid layer with the combined cell values, of which the highest values were extracted accumulating to three percent of the study area, a requirement before the implementation of the PAI process.

The robbery probability model predicted 60.53 percent and 57.26 percent of day and night-time respective robberies

7.6   Comparison between the robbery probability model and a traditional crime hotspot model.

A secondary aim of the current study was to find out which of the models between the traditional hotspots and robbery probability models better predict robberies. This was done by comparing the PAI values between the two latter models.

The results were in favour of the traditional hotspots model for both day- and night-time. The traditional model predicted almost 10 percent more of the robberies occurred in 2007 than the robbery probability model for both during the day- and night-time. However, the facility risk model was built on robberies as the dependent variable, after the model has been created it can be used in areas where robberies are unknown or in a study to calculate the impact new facilities, road networks and demographic changes will have on robberies. The latter is not possible with the traditional model which is solely dependent on previous robbery locations.

7.7   Further research and scope for further studies

The current study introduced a new prediction method whereby robbery was predicted per grid block. Further research is required to better understand the impact and reliability of the results generated here and to improve of the methods and data used. Further research is also needed to compare the prediction results to a more traditional dependent variable such as a simple count of robberies per grid block. Using a count of robberies per grid block will also reduce the number of grid blocks acting as the dependent variable, which makes it possible to use other prediction models such as the random forest model. Therefore further research can also be done on choosing the right prediction model to predict a variable per grid block.

The current dissertation used data from 2001 to 2007 which is more than ten years old. Further studies are required whereby more recent data are used to create both a

traditional hotspot and a robbery probability models before making the models a viable option for the City of Tshwane Metropolitan Municipality to implement as a real life scenario. Additional data such as alcohol outlets, ATMs, foot traffic also opens up the possibility to improve and expand on the models presented in this study. The time constraint model in particular could be vastly improved on by finding routes most travelled, especially from and to neighbourhoods classified as socially disorganised.

Even though the robbery probability model introduced in the current study was specifically built for the City of Tshwane Metropolitan Municipality, the model can be applied on any geographical location with the same type of facilities. Further studies can therefore test the same model on different geographical locations to find out whether the results are context-specific. It is also possible to interchange the facilities included in the current study and in doing so, find the best possible set of variables for any given location.

CHAPTER 8: REFERENCES

ANDRESEN, M. A. 2005. Crime measures and the spatial analysis of criminal activity. *British Journal of Criminology,* 46**,** 258-285.

ANDRESEN, M. A. 2006. A spatial analysis of crime in Vancouver, British Columbia: a synthesis of social disorganization and routine activity theory. *Canadian Geographer,* 50**,** 487-502.

ANSELIN, L., COHEN, J., COOK, D., GORR, W. & TITA, G. 2000. Spatial analyses of crime. *Criminal Justice,* 4**,** 213-262.

ASHBY, M. P. & BOWERS, K. J. 2013. A comparison of methods for temporal analysis of aoristic crime. *Crime Science,* 2**,** 1.

BERNASCO, W. & BLOCK, R. 2011. Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency,* 48**,** 33-57.

BLOCK, R. & DAVIS, S. 1996. The environs of rapid transit stations: A focus for street crime or just another risky place. *Crime prevention studies,* 6**,** 237-57.

BOWERS, K. 2014. Risky Facilities: Crime Radiators or Crime Absorbers? A Comparison of Internal and External Levels of Theft. *Journal of Quantitative Criminology,* 30**,** 389-414.

BRAGA, A. A. & BOND, B. J. 2008. Policing crime and disorder hot spots: A randomized controlled trial. *Criminology,* 46**,** 577-607.

BRAGA, A. A., PAPACHRISTOS, A. V. & HUREAU, D. M. 2014. The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly,* 31**,** 633-663.

BRANTINGHAM, P. & BRANTINGHAM, P. 1995. Criminality of place. *European journal on criminal policy and research,* 3**,** 5-26.

BRANTINGHAM, P. L. & BRANTINGHAM, P. J. 1993. Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology,* 13**,** 3-28.

BREETZKE, G. D. 2010a. Modeling violent crime rates: A test of social disorganization in the city of Tshwane, South Africa. *Journal of Criminal Justice,* 38**,** 446-452.

BREETZKE, G. D. 2010b. A SOCIO-STRUCTURAL ANALYSIS OF CRIME IN THE CITY OF TSHWANE, SOUTH AFRICA. *South African Journal of Science,* 106**,** 41-47.

BREETZKE, G. D. 2016. Examining the spatial periodicity of crime in South Africa using Fourier analysis. *South African Geographical Journal,* 98**,** 275-288.

BREETZKE, G. D. & COHN, E. G. 2013. Burglary in Gated Communities: An Empirical Analysis Using Routine Activities Theory. *International Criminal Justice Review (Sage Publications),* 23**,** 56-74.

BURROUGH, P. 2001. GIS and geostatistics: Essential partners for spatial analysis. *Environmental and ecological statistics,* 8**,** 361-377.

CAPLAN, J. M. & KENNEDY, L. W. 2011. Risk terrain modeling compendium. *Rutgers Center on Public Security, Newark*.

CAPLAN, J. M., KENNEDY, L. W. & MILLER, J. 2011. Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. *Justice Quarterly,* 28**,** 360-381.

CARRILLO, P. E., LOPEZ-LUZURIAGA, A. & MALIK, A. S. 2018. Pollution or crime: The effect of driving restrictions on criminal activity. *Journal of Public Economics,* 164**,** 50-69.

CHAINEY, S. 2013. Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime. *Bulletin of the Geographical Society of Liege,* 60**,** 7-19.

CHAINEY, S. & RATCLIFFE, J. 2013. *GIS and crime mapping*, John Wiley & Sons.

CHAINEY, S., TOMPSON, L. & UHLIG, S. 2008. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal,* 21**,** 4-28.

CLARKE, R. V. G. 1996. *Preventing mass transit crime*, Criminal Justice Press Monsey, NY.

COHEN, L. E. & FELSON, M. 1979. Social change and crime rate trends: A routine activity approach. *American sociological review***,** 588-608.

COHN, E. G. 1990. Weather and crime. *The British Journal of Criminology,* 30**,** 51-64.

CONROW, L., ALDSTADT, J. & MENDOZA, N. S. 2015. A spatio-temporal analysis of on-premises alcohol outlets and violent crime events in Buffalo, NY. *Applied Geography,* 58**,** 198-205.

DALL'ERBA, S. 2009. Exploratory Spatial Data Analysis. *In:* KITCHIN, R. & THRIFT, N. (eds.) *International Encyclopedia of Human Geography.* Oxford: Elsevier.

DEMEAU, E. & PARENT, G. 2018. Impacts of Crime Attractors and Generators on Criminality in Montreal. *Canadian Journal of Criminology & Criminal Justice,* 60**,** 387-412.

DRAWVE, G., THOMAS, S. & WALKER, J. 2014. The Likelihood of Arrest: A Routine Activity Theory Approach. *American Journal of Criminal Justice,* 39**,** 450-470.

DU PLESSIS, A. & LOUW, A. 2005. THE TIDE IS TURNING: The 2003/04 SAPS crime statistics. *SA Crime Quarterly***,** 1-8.

ECK, J., CHAINEY, S., CAMERON, J. & WILSON, R. 2005. Mapping crime: Understanding hotspots.

ECK, J. E., CLARKE, R. V. & GUERETTE, R. T. 2007. Risky facilities: Crime concentration in homogeneous sets of establishments and facilities. *Crime prevention studies,* 21**,** 225.

ECK, J. E. & WEISBURD, D. L. 2015. Crime places in crime theory.

FREISTHLER, B., PONICKI, W. R., GAIDUS, A. & GRUENEWALD, P. J. 2016. A micro-temporal geospatial analysis of medical marijuana dispensaries and crime in Long Beach, California. *Addiction,* 111**,** 1027-1035.

GILL, C., WOODITCH, A. & WEISBURD, D. 2017. Testing the "Law of Crime Concentration at Place" in a Suburban Setting: Implications for Research and Practice. *Journal of Quantitative Criminology,* 33**,** 519-545.

GORR, W. L. & LEE, Y. 2015. Early warning system for temporary crime hot spots. *Journal of Quantitative Criminology,* 31**,** 25-47.

GROFF, E. & MCCORD, E. S. 2012. The role of neighborhood parks as crime generators. *Security Journal,* 25**,** 1-24.

GROFF, E. R. & LA VIGNE, N. G. 2001. Mapping an opportunity surface of residential burglary. *Journal of Research in Crime and Delinquency,* 38**,** 257-278.

GUERRY, A.-M. 1833. *Essai sur la statistique morale de la France*, Clearwater.

GUYON, I. & ELISSEEFF, A. 2003. An introduction to variable and feature selection. *Journal of machine learning research,* 3**,** 1157-1182.

HABERMAN, C. P. & RATCLIFFE, J. H. 2015. Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology,* 53**,** 457-483.

HART, T. C. & MIETHE, T. D. 2014. Street robbery and public bus stops: A case study of activity nodes and situational risk. *Security Journal,* 27**,** 180-193.

HIROPOULOS, A. & PORTER, J. 2014. Visualising property crime in Gauteng : applying GIS to crime pattern theory. *SA Crime Quarterly,* 2014**,** 17-28.

JOBES, P. C., BARCLAY, E., WEINAND, H. & DONNERNIEYER, J. F. 2004. A Structural Analysis of Social Disorganisation and Crime in Rural Communities in Australia. *Australian & New Zealand Journal of Criminology (Australian Academic Press),* 37**,** 114-140.

JOHNSON, C. Crime mapping and analysis using GIS. Geomatics Conference on Geomatics in Electronic Governance, January, 2000.

KENNEDY, L. W., CAPLAN, J. M. & PIZA, E. 2011. Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology,* 27**,** 339-362.

LANCASTER, L. & KAMMAN, E. 2016. Risky localities Measuring socioeconomic characteristics of high murder areas. *South African crime quarterly,* 56**,** 27-35.

LINNING, S. J. 2015. Crime seasonality and the micro-spatial patterns of property crime in Vancouver, BC and Ottawa, ON. *Journal of Criminal Justice,* 43**,** 544-555.

LIU, D., SONG, W. W. S. L. E. & XIU, C. 2016. Spatial patterns of violent crimes and neighborhood characteristics in Changchun, China. *Australian & New Zealand Journal of Criminology (Sage Publications Ltd.),* 49**,** 53-72.

LIU, H. & BROWN, D. E. 2003. Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting,* 19**,** 603-622.

MARKOWITZ, F. E., BELLAIR, P. E., LISKA, A. E. & LIU, J. 2001. Extending social disorganization theory: Modeling the relationships between cohesion, disorder, and fear. *Criminology,* 39**,** 293-319.

MISTRY, D. 2004. Falling crime, rising fear: 2003 national victims of crime survey. *South African Crime Quarterly*.

MOLINARO, A. M., SIMON, R. & PFEIFFER, R. M. 2005. Prediction error estimation: a comparison of resampling methods. *Bioinformatics,* 21**,** 3301-3307.

MURRAY, A. T., MCGUFFOG, I., WESTERN, J. S. & MULLINS, P. 2001. Exploratory spatial data analysis techniques for examining urban crime: Implications for evaluating treatment. *British Journal of Criminology,* 41**,** 309-329.

NEWTON, A. 2014. Crime on Public Transport.

NOBLE, M. & WRIGHT, G. 2013. Using indicators of multiple deprivation to demonstrate the spatial legacy of apartheid in South Africa. *Social Indicators Research,* 112**,** 187-201.

OSGOOD, D. W., WILSON, J. K., O'MALLEY, P. M., BACHMAN, J. G. & JOHNSTON, L. D. 1996. Routine activities and individual deviant behavior. *American Sociological Review***,** 635-655.

PEREIRA, D. V. S., ANDRESEN, M. A. & MOTA, C. M. M. 2016. A temporal and spatial analysis of homicides. *Journal of Environmental Psychology,* 46**,** 116-124.

RANSON, M. 2014. Crime, weather, and climate change. *Journal of environmental economics and management,* 67**,** 274-302.

RATCLIFFE, J. 2010. Crime mapping: spatial and temporal challenges. *Handbook of quantitative criminology.* Springer.

RATCLIFFE, J. H. 2006. A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquency,* 43**,** 261-291.

RATCLIFFE, J. H., TANIGUCHI, T., GROFF, E. R. & WOOD, J. D. 2011. The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology,* 49**,** 795-831.

RUMMENS, A., HARDYNS, W. & PAUWELS, L. 2017. The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied Geography,* 86**,** 255-261.

SAMPSON, R. J. 1986. Crime in cities: The effects of formal and informal social control. *Crime and justice,* 8**,** 271-311.

SAMPSON, R. J. 1987. Urban black violence: The effect of male joblessness and family disruption. *American journal of Sociology,* 93**,** 348-382.

SAMPSON, R. J. & GROVES, W. B. 1989. Community structure and crime: Testing social-disorganization theory. *American journal of sociology,* 94**,** 774-802.

SAPS. 2005. *Strategic Plan for the SA Police Service (2005 -2010)* [Online]. Available: https://www.saps.gov.za/about/stratframework/strategic_plan/2005_2010/strategic_plan_2005_2010.pdf [Accessed 01/09/2019 2019].

SAPS. 2007/08. *South African Police Service: Annual Report 2007/2008* [Online]. Available: https://www.gov.za/sites/default/files/gcis_document/201409/saps-anrep-complete.pdf [Accessed 02/06/2019 2019].

SCHUTTE, F. H. & BREETZKE, G. D. 2018. The influence of extreme weather conditions on the magnitude and spatial distribution of crime in Tshwane (2001–2006). *South African Geographical Journal,* 100**,** 364-377.

SHAW, C. R. & MACKAY, H. D. 1969. *Juvenile deliquency and urban areas a study of rates of delinquency in relation to differential characteristics of local communities in American cities,* Chicago [u.a.], Univ. of Chicago Press.

SHERMAN, L. W., GARTIN, P. R. & BUERGER, M. E. 1989. Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology,* 27**,** 27-56.

STATISTICSSA. 2013. *City of Tshwane* [Online]. Available: http://www.statssa.gov.za/?page_id=993&id=city-of-tshwane-municipality [Accessed 06/08/2019 2019].

STATISTICSSA. 2018. *Victims of Crime Survey, 2018* [Online]. Available: http://www.statssa.gov.za/?page_id=1854&PPN=P0341&SCH=7373 [Accessed].

STATISTICSSA. 2019a. *Community Profile Databases* [Online]. Available: http://www.statssa.gov.za/?page_id=4089 [Accessed 20/06/2019 2019].

STATISTICSSA. 2019b. *Copyright and disclaimer* [Online]. Available: http://www.statssa.gov.za/?page_id=425 [Accessed 03/10/2019 2019].

SUFFLA, S. & SEEDAT, M. 2016. The epidemiology of homicidal strangulation in the City of Johannesburg, South Africa. *Journal of Forensic and Legal Medicine,* 37**,** 97-107.

SUMMERS, L. & JOHNSON, S. 2017. Does the Configuration of the Street Network Influence Where Outdoor Serious Violence Takes Place? Using Space Syntax to Test Crime Pattern Theory. *Journal of Quantitative Criminology,* 33**,** 397-420.

SWART, L.-A., SEEDAT, M. & NEL, J. 2016. Neighborhood Sociostructure and Adolescent Homicide Victimization in Johannesburg, South Africa. *Homicide Studies,* 20**,** 220-238.

UITTENBOGAARD, A. & CECCATO, V. 2012. Space-time clusters of crime in Stockholm, Sweden. *Review of European Studies,* 4**,** 148.

UNWIN, D. J. 1996. GIS, spatial analysis and spatial statistics. *Progress in Human Geography,* 20**,** 540-551.

WAIN, N., ARIEL, B. & TANKEBE, J. 2017. The collateral consequences of GPS-LED supervision in hot spots policing. *Police Practice and Research,* 18**,** 376-390.

WARCHOL, G. & HARRINGTON, M. 2016. Exploring the dynamics of South Africa's illegal abalone trade via routine activities theory. *Trends in Organized Crime,* 19**,** 21-41.

WEISBURD, D. & ECK, J. E. 2004. What can police do to reduce crime, disorder, and fear? *The Annals of the American Academy of Political and Social Science,* 593**,** 42-65.

WORTLEY, R. & MAZEROLLE, L. 2008. Environmental criminology and crime analysis: Situating the theory, analytic approach and application. *Environmental criminology and crime analysis,* 6**,** 1-18.

CHAPTER 9: APPENDICES

9.1 Appendix 1 – Crime data letter of approval from University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Natural and Agricultural Sciences
Ethics Committee

E-mail: ethics.nas@up.ac.za

Date: 26 June 2017

ETHICS SUBMISSION: LETTER OF APPROVAL

Name of Applicant	Prof G Breetzke
Department	Geography, Geoinformatics and Meteorology
Reference number	EC170518-119
Title	Crime analysis in Tshwane, South Africa: Understanding spatio-temporal crime

Dear Prof Breetzke

The submission conforms to the requirements of the NAS EC. Any amendments must be submitted to the NAS EC on a relevant application form as used for the original application quoting the reference number and detailing the required amendment. An amendment would be for example differentiating within the research target population.

You are required to submit a progress report no later than two months after the anniversary of this application as indicated by the reference number. The progress report document is accessible of the NAS faculty's website: Research/Ethics Committee.

You are required to notify the NAS EC upon the completion or ending of the project using the form Project Completed. Completion will be when the data has been analysed and documented in a postgraduate student's thesis or dissertation, or in a paper or a report for publication.

The digital archiving of data is a requirement of the University of Pretoria. The data should be accessible in the event of an enquiry or further analysis of the data.

The NAS EC wishes you well with your research project.

Yours sincerely,

Chairperson
NAS Ethics Committee

9.2 Appendix 2 – Letter of approval from AfriGIS

A F R I G I S
everything about everywhere.

26 June 2019

To: Magnus Rademeyer
        Managing Director: AfriGIS

From: Nicolas Kemp

RE: LETTER OF REQUEST TO USE AFRIGIS DATASETS FOR MY STUDIES AT THE UNIVERSITY OF PRETORIA

I am currently enrolled at the University of Pretoria for my MSc - GIS. My studies focus on robbery hotspots in the City of Tshwane Metropolitan Municipality, specifically for the years 2006 and 2007. I am kindly requesting to use the following AfriGIS datasets for my research:

- AfriGIS municipality boundaries to select data within Tshwane
- Town boundaries for labelling purposes
- AfriGIS street centrelines
- AfriGIS points of interest
- AfriGIS WMS services as background imagery for maps

The data was sourced from the year 2006, however, the points of interests were updated with 2018 coordinates and confidence levels. Furthermore, the above mentioned datasets will only be used for the purpose of my studies and will remain the property of AfriGIS. I will be the only person with direct access to the datasets and they will not be distributed to other parties.

Regards

Nicolas Kemp

Approved by:

Magnus Rademeyer   27/06/2019
                                Date