NELSON MANDELA UNIVERSITY

A Model for Crime Management in Smart Cities

Lindsay Westraadt

Submitted in fulfilment of the requirements for the degree of Philosophiae Doctor in the Faculty of Science at the Nelson Mandela University

Promoter: Professor André P. Calitz

December 2019

Declaration

I, Lindsay Westraadt (student number: 202304736), hereby declare that this thesis for the degree of Philosophiae Doctor is my own work and that it has not previously been submitted for assessment or completion of any postgraduate qualification to another University or for another qualification.

L. Westraadt

Lindsay Westraadt

Dedicated to my family

"[T]he justice system alone cannot solve many of the underlying conditions that give rise to crime. It will be through partnerships across sectors and at every level of government that we will find the effective and legitimate long-term solutions to ensuring public safety."

President's Task Force on 21st Century Policing. (2015)

Acknowledgements

I would like to thank all those who have supported me throughout this study. In particular, I would like to thank my promoter, Professor André Calitz, for his guidance, availability and support. Your dynamic approach to research is truly refreshing, and I hope that this is only the start of exciting things to come!

I would also like to acknowledge Professors Chris James and Anne Fitchette from the School of Civil and Environmental Engineering at the University of the Witwatersrand for introducing me to the wonderful worlds of complexity theory and urban design.

To my colleagues at the Centre for High Resolution Transmission Electron Microscopy, I would like to thank you for your friendship and moral support. Particular thanks goes to my line-manger, Professor Jan Neethling, for his unyielding patience towards the end of this study. Special thanks also goes to Prof Mike Lee, Mr William Goosen, Ms Marisa Kolver and Ms Genevéve Marx for always putting a smile on my face.

To the broader Nelson Mandela University, I would like to thank Dr Derek Du Preez from the Department of Botany for granting me access to Esri® ArcMap™, Dr MC du Plessis from the Department of Computing Sciences for his guidance regarding neural networks, and Professor Gary Sharp from the Department of Statistics for his guidance regarding exploratory factor analysis. I would also like to thank Nelson Mandela University for their financial support.

Special thanks goes to the staff at the Nelson Mandela Bay Municipality for their open and collaborative spirit. In particular, I would like to acknowledge Mr Chris van Lelyveld and Mr Rajan Naidoo from the ICT Department; Ms Melanie Smit, Mr Franco Cilliers, and Mr Gideon Louw from the Corporate GIS Department; Mr Gerald du Toit from the Traffic Department; and Mr Whiteboy from the Fire Department. Special thanks also goes to Mandy Dicker from the Humewood police station.

Thanks goes to Colonel Kenqu from the National Crime Registrar and Mr Garth Sampson from the South African Weather Service for providing me with data.

Most importantly, I would like to thank my family. Without the unyielding support and sacrifice of my husband, my parents, and my two little boys, none of this would have been possible. I am truly blessed to have such a supportive family. I would also like to thank my spiritual family for carrying me through a very trying three years.

Abstract

Observation of global smart city trends shows a shift in focus from sector-based interventions towards integrated decision-making informed by Big Data. This move towards integration is evident in the emergence of Integrated City Management Platforms, which aim to orchestrate smart city infrastructure at a system-of-systems level. Despite the fact that smart city solutions in South Africa are in their infancy, South Africa already has the policy framework in place to support integrated decision-making at the planning level through the implementation of Integrated Development Plans (IDPs). In this study, it was shown how emerging trends in smart city integration can complement existing IDP practices in South Africa, with the potential of transforming the smart city status of South African cities.

The main research problem addressed in this study is that South African cities are not effectively integrating and utilising available, and rapidly emerging smart city data sources for planning and management. To this end, it was proposed that a predictive model, that assimilates data from traditionally isolated management silos, could be developed for prediction and simulation at the system-of-systems level. As proof of concept, the study focused on only one aspect of smart cities, namely crime management. Subsequently, the main objective of this study was to develop and evaluate a predictive model for crime management in smart cities that effectively integrated data from traditionally isolated management silos. The Design Science Research process was followed to develop and evaluate a prototype model.

It was proposed that emerging smart city KPI frameworks could be leveraged to represent smart cities as a set of sectoral KPIs. Taking the latter as input, it was further proposed that a combined approach employing Bayesian Neural Networks and sensitivity analysis could be used as a tool for systems-level scenario analysis. Based on these design specifications, two prototype models were developed and evaluated, one for street larceny and one for street robbery. Due to the limited accessibility of South African data at the time of this study, readily accessible open data for New York City was used to develop and demonstrate the prototype models.

It was found that the prototype modelling approach successfully integrated data from traditionally isolated management silos, and provided an effective tool for systems-level scenario analysis and synergistic cross-sector collaboration. The models also proved effective in identifying the key government agencies at play in the fight against crime. During the initial stage of model exploration, it was found that there was a high degree of correlation among input features. Further investigation showed that the input data tended to cluster together to represent different system “states”. Due to the clustering in input space, care needed to be taken when fixing variables for sensitivity analysis. Subsequently, it was shown that the proposed modelling approach could most effectively be implemented when latent system states were first identified with the use of Exploratory Factor Analysis, followed by contextualised sensitivity analyses within the respective city states.

The prototype model was evaluated *ex ante* within the South African context by way of a mixed-method case study of Nelson Mandela Bay Municipality. It was shown that comparable KPI frameworks to the one used in this study were already implemented in South African IDPs. It was anticipated therefore, that the demonstrated modelling approach could supplement existing IDP processes with relatively little effort and disruption of existing management processes. Guidelines for the implementation of the developed modelling approach within the South African context were developed based on the implementation and evaluation of the prototype model.

The practical contributions of this study was the development of a prototype model for integrated decision-making in smart cities, and the associated guidelines for the implementation of the developed modelling approach within the South African IDP context. Theoretically, this work contributed towards the development of a modelling paradigm for effective integrated decision-making in smart cities. This work also contributed towards developing strategic-level predictive policing tools aimed at proactively meeting community needs, and contributed to the body of knowledge regarding complex systems modelling.

Keywords: Smart cities, Integrated City Management Platforms, Integrated Development Plans, Predictive Policing, Bayesian Neural Networks.

Table of Contents

SECTION 2: SUPPORTING LITERATURE 31

List of Abbreviations

Abbreviation	Term in full
ACS	American Community Survey
ANN	Artificial Neural Network
CCTV	Closed-Circuit Television
CI	Core Indicator
CAD	Computer-Aided Dispatch
CompStat	Compare Statistics
CPA	City Protocol Agreement
DEA	Department of Environmental Affairs
DEAT	Department of Environmental Affairs and Tourism
DSR	Design Science Research
EAP	Environmental Assessment Practitioner
EFA	Exploratory Factor Analysis
EIA	Environmental Impact Assessment
EIAMS	Environmental Impact Assessment and Management Strategy
EMF	Environmental Management Framework
EMP	Environmental Management Plan
FBI	Federal Bureau of Investigation
GDP	Gross Domestic Product
GHG	Greenhouse Gas
GIS	Geographic Information System
GPS	Global Positioning System
IBM	International Business Machines
ICT	Information and Communications Technology
ICMP	Integrated City Management Platform
IDP	Integrated Development Plan
IEC	International Electrotechnical Commission
IEM	Integrated Environmental Management
IoT	Internet of Things
ISO	International Standards Organisation
IT	Information Technology
ITU	International Telecommunication Union
KPA	Key Performance Area
KPI	Key Performance Indicator
LAPD	Los Angeles Police Department
LEED	Leadership in Energy and Environmental Design
MAP	Maximum a Posteriori
MMR	NYC Mayor’s Management Report
MOO	NYC Mayor’s Office of Operations
MPAC	Municipal Public Accounts Committee
NEMA	National Environmental Management Act
NGO	Non-governmental Organisation
NIJ	National Institute of Justice
NMBM	Nelson Mandela Bay Municipality
NTA	Neighbourhood Tabulation Area

NYC	New York City
NYPD	New York Police Department
OECD	Organisation for Economic Co-operation and Development
PLUTO	Primary Land Use Tax Lot Output
PMS	Performance Management System
PUMA	Public Use Microdata Area
R&D	Research & Development
RACR	Real-Time Analysis Critical Response
RAND	The RAND Corporation
RMS	Records Management System
RTIC	Real-Time Intelligence Centre
RTM	Risk Terrain Modelling
RSA	Republic of South Africa
RO	Research Objective
RQ	Research Question
SALGA	South African Local Government Association
SANRAL	South African National Roads Agency
SDBIP	Service Delivery and Budget Implementation Plan
SI	Supporting Indicator
SDF	Spatial Development Framework
SDO	Standards Developing Organisation
SDG	Sustainable Development Goal
SSL	Strategic Subjects List
UCR	Uniform Crime Reporting
UK	United Kingdom
US	United States
USA	United States of America
WCED	World Commission on Environment and Development

List of Figures

List of Tables

SECTION 1: RESEARCH CONTEXT AND METHODOLOGY

Chapter 1. Research Context

1.1 Background

The drive for smart cities is rooted in the United Nations Sustainable Development Goals (United Nations., 2015), which calls all member countries to focus on development that meets basic human needs, promotes economic growth, reduces inequality, and protects the natural environment (The Economist, 2009). Cities have a major impact on human quality of life and the natural environment. The plethora of challenges faced by city managers worldwide include the increasing pressures of service delivery, housing, unemployment, food supply, natural disasters, disease, traffic congestion, pollution, resource consumption and waste, inequality, and crime. The latter are all intensified by rapid urbanisation and the increasing effects of climate change (Kourtit *et al.*, 2014).

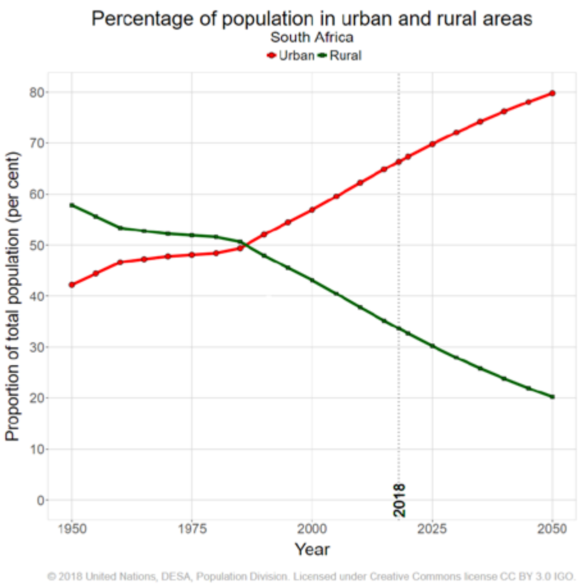


Figure 1-1: Percentage of South African population living in urban areas. Source: United Nations (2018).

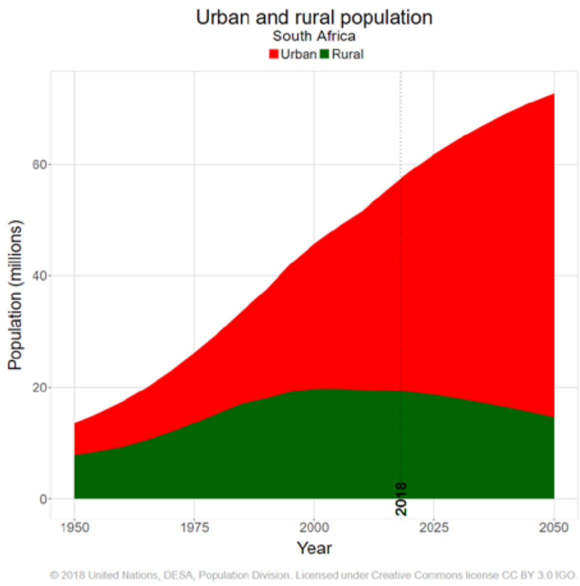


Figure 1-2: Total South African population living in urban areas. Source: United Nations (2018).

In South Africa, the total population living in urban areas is expected to increase from just over 65% in 2018 to 80% in 2050 (Figure 1-1; United Nations, 2018). This translates into an urban population increase of over 10 million people (Figure 1-2; United Nations, 2018). Now, more than ever, city officials are being called to manage increasingly stressed resources with unprecedented efficiency (IBM., 2010; IEC., 2015). Moreover, the increasing effects of climate change requires a high degree of resilience to be imbedded into the design and management of cities (The Rockefeller Foundation | Arup., 2015).

Globally, the Information Technology (IT) industry has stepped up to this challenge in a major way, and over the last decade there has been an explosion in smart city solutions (IEC., 2015). While there is some ambiguity with the definition of smart cities (Albino *et al.*, 2015; SALGA., 2015; Ahvenniemi *et al.*, 2017; Mattoni *et al.*, 2017), it is evident that most authors define smart cities as those that employ IT to achieve the goals of sustainability and resilience.

Connectivity is fundamental to smart cities (ITU., 2016b). Smart people are connected through smart phones, and infrastructure and the urban environment are connected through the Internet of Things (IoT). The IoT is the name given to the growing trend in which large numbers of networking sensors are embedded into various devices, enabling information-gathering and control functions (Chen *et al.*, 2014). This ubiquitous connectivity allows for real-time monitoring and the management of infrastructure and citizens. By combining real-time monitoring, data analytics and advanced citizen engagement, cognitive technologies are leveraging the IoT and Big Data to radically reduce inefficiencies in all city sectors such as healthcare, public safety, water, transportation and energy (IBM., 2018).

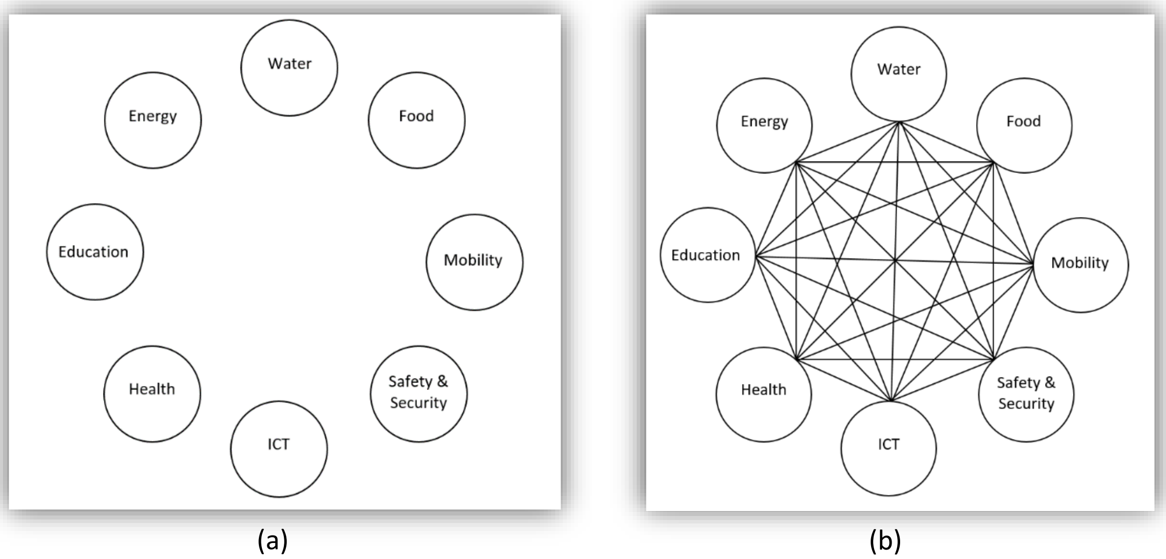


Figure 1-3: (a) Optimisation within historical verticals (intra-sector optimisation) is the core of today's smart city solutions (IBM., 2010; IEC., 2015). (b) In recent years, however, there has been a call for horizontal (inter-sector) integration aimed at eliminating systemic inefficiencies (IEC., 2015). Source: Author's own construction.

Despite the interdependence between city sectors, optimisation within historical verticals (intra-sector optimisation) is the core of today's smart city solutions (Figure 1-3a; IBM, 2010; IEC, 2015). In recent years, however, it has become apparent that this silo approach to smart cities development is reaching its limits, and research is being conducted on developing appropriate frameworks for horizontal (inter-sector) integration aimed at eliminating systemic inefficiencies and fostering integrated decision-making, interoperability and cross-sector collaboration (Figure 1-3b; IEC, 2015).

This move towards integration is evident in the emergence of Integrated City Management Platforms (ICMPs) (City Protocol Society., 2015b), such as the Intelligent Operations Centre solutions of IBM (Figure 1-4; IBM, 2013) and Huawei (Huawei, 2018), which aim to orchestrate smart city infrastructure at a system-of-systems level (IEC., 2015). Yan Lida, president of Huawei Enterprise Business Group, likened these information platforms to smart city central nervous systems, which comprise "a 'brain' and 'peripheral nerves', gathering real-time information about the status of the city, transmitting the data, enabling the "brain" to analyze and make informed decisions, delivering feedback commands, and ultimately carrying out intelligent actions" (fin24tech, 2017).

Figure 1-4: IBM Intelligent Operations Center. Source: IBM. (2013).

Much work is currently being done by international standards organisations to develop the data, technical and management standards required to effectively support integrated decision-making and collaboration within ICMPs (IEC., 2015). Thus far, development of management standards has focused on the design of conceptual models aimed at creating a common visual understanding of core smart city components and their interactions (City Protocol Society., 2015b); and on developing globally comparable Key Performance Indicators (KPIs) aimed at setting clear development targets required for city operation, evaluation and transformation (City Protocol Society., 2015a).

1.2 Problem statement

In South Africa, initial smart city efforts have focused on the provision of free Wi-Fi access at strategic locations, the adoption of smart meters and e-government applications for municipal utilities management, and the implementation of Intelligent Traffic Management Systems (SALGA., 2015). Smart policing (Happimo, 2016; Head, 2017) and open data (Open Government Partnership, 2018) are also emerging as areas of interest.

Although a number of South African city departments have implemented smart city solutions, these initiatives have been fragmented with limited integration across departments (SALGA., 2013; SALGA., 2015). There is however, indications that this will change in the near future, as smart city solutions mature. The City of Johannesburg, for example, aims to become the leading smart city in South Africa with the implementation of an Intelligent Operation Centre aimed at providing an integrated view of the city's strategic and operational issues. The latter aims to focus on public safety in the initial phase of implementation (City of Johannesburg., 2011; City of Johannesburg., 2015; SALGA., 2015).

Despite the fact that smart city solutions in South Africa are in their infancy, South Africa already has the policy framework in place to support integrated decision-making at the planning level. In South Africa, integrated city planning is enforced through the implementation of Integrated Development Plans (IDPs) and their supporting Spatial Development Frameworks required by the Local Government:

Municipal Systems Act, 32 of 2000 (RSA., 2000). The main aims of an IDP are to accelerate service delivery in municipalities and to deliver the spatial, social, ecological and economic urban patterns that are in line with the country's democratic and sustainability visions (Ngamlana and Eglin, 2015; Mnguni, 2016). IDPs are carried out by identifying the development needs and concerns in a municipality and then formulating and prioritising possible intervention programmes and projects in an inclusive and integrated manner (RSA., n.d.).

While the introduction of the IDP policy has been positively received, there exists a chronic disconnect between planning and implementation (DEA., 2014; Ngamlana and Eglin, 2015). The number of civic protests in South Africa is on the increase, with over 100 protest incidents every year (Mnguni, 2016). The main reason cited for these often violent protests has been dissatisfaction with municipal service delivery (Mnguni, 2016). While there are a number of factors contributing to the non-performance of the IDP process, two of the major criticisms include a lack of true cross-sector integration and collaboration (Ngamlana and Eglin, 2015), and a lack of effective information systems (DEA., 2014).

Reliable, current and readily accessible data is an essential requirement for effective decision-making and participation by all role-players in the planning process. Coupled to this, is the need for this data to be assimilated in such a way as to enable effective cross-sector integration and collaboration. While a wide variety of data and information is being generated by various government departments in South African cities, these data are not being integrated and utilised in a way that optimally supports integrated planning and management (DEA., 2014).

The main research problem addressed in this study is that: **South African cities are not effectively integrating and utilising available data sources for smart city planning and management.**

1.3 Thesis statement

This study will investigate how existing, and rapidly emerging smart city data sources can be integrated and utilised to more effectively support planning in South African cities. In order to limit the scope of the investigation, the study will focus on only one aspect of smart cities, namely crime management. In this study, it is proposed that emerging trends in smart city integration can complement existing planning practices in South Africa. Specifically, it is proposed that a predictive model, that incorporates data from traditionally isolated management silos, can be developed for whole-system scenario analysis applications.

The thesis statement is as follows: **A predictive model, that effectively integrates and utilises data from traditionally isolated management silos, can support an integrated approach to crime management in smart South African cities.**

1.4 Research contribution

In order to create the common strategic vision necessary for cooperative decision-making, effective integrated decision-making requires clear development objectives, indicators and targets. Furthermore, effective integrated decision-making requires knowledge of the status quo, and reliable tools for scenario analysis (DEA., 2014). To this end, emerging smart city conceptual models and KPI frameworks (see Section 1.1) are effective in delineating key smart city components and their desired states, and provide a framework for harnessing data generated by ICMPs to create objective automated performance dashboards for status quo analysis (City Protocol Society., 2015b; City Protocol Society., 2015a).

Despite the deluge of data generated by ICMPs and the accompanying growth in computing power, limited research has been done on exploring the use of this data to develop objective quantitative tools for project prioritisation and scenario analysis at the system-of-systems level (Lombardi *et al.*, 2012; Mattoni *et al.*, 2015; Schleicher *et al.*, 2016; Mattoni *et al.*, 2017). In this study, it is proposed that a predictive model for whole-system scenario analysis can be developed, by building upon emerging smart city management solutions.

It is anticipated that this can be achieved by leveraging smart city KPI frameworks (City Protocol Society., 2015a) as a tool for assimilating and representing data from traditionally isolated management silos as a set of sectoral KPIs (Figure 1-5a). It is envisaged that the inter-dependencies between sectoral KPIs can be encapsulated in an artificial neural network (Figure 1-5b), which can be used for prediction and simulation applications at the system-of-systems level (Figure 1-5c).

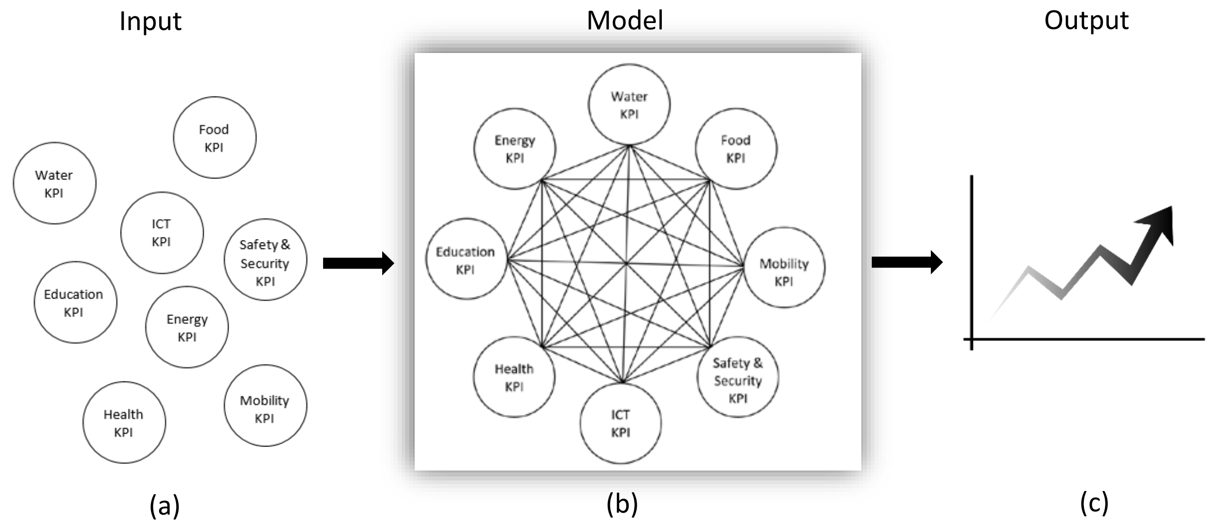


Figure 1-5: In this study, it is proposed that a predictive model for whole-system scenario analysis can be developed by building upon emerging smart city management solutions. (a) It is anticipated that this can be achieved by using emerging smart city KPI frameworks to represent data from traditionally isolated management silos as a set of sectoral KPIs. (b) It is envisaged that the inter-dependencies between sectoral KPIs can be encapsulated in an artificial neural network, which (c) can be used for prediction and simulation at the system-of-systems level. Source: Author's own construction.

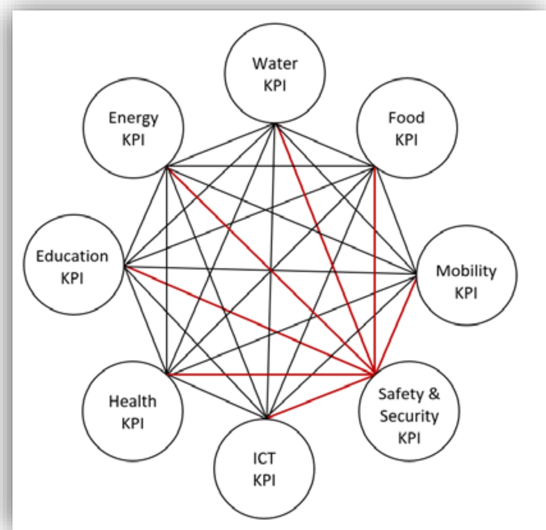


Figure 1-6: In order to limit the scope of the investigation, this study will focus on only one aspect of smart cities, namely crime management. Source: Author's own construction.

The aim of this study is to test the feasibility of this supposition. The Design Science Research approach will be followed to design, develop, demonstrate and evaluate a prototype model. In order to limit the scope of the investigation, the study will focus on only one aspect of smart cities, namely crime management (Figure 1-6).

It is anticipated that the research contributions of this study will be in the form of a prototype model demonstrating the proposed integrated approach to crime management in smart cities, accompanied by a set of model implementation guidelines for South African cities. It is anticipated that the theoretical contribution of this study will be in the form of design science knowledge regarding the efficacy of the proposed approach.

1.5 Research questions and objectives

The main research question of this study is:

RQ_m: How can a predictive model for crime management in smart South African cities be developed that effectively integrates and utilises data from traditionally isolated management silos?

The sub-questions following from the main research question are:

RQ_1: What are the requirements of an efficacious smart city model?

RQ_2: Which input and output parameters are relevant to the model?

RQ_3: Which data sources are available in smart cities?

RQ_4: Which modelling approach could be used?

RQ_5: How could the developed model be used in practice?

RQ_6: What is the efficacy of the prototype model?

RQ_7: What implementation guidelines for the South African context can be derived from the development and evaluation of the prototype model?

Associated with the above research questions are the following research objectives. The primary research objective of this study is:

RO_m: To develop and evaluate a predictive model for crime management in smart South African cities that effectively integrates data from traditionally isolated management silos.

The secondary objectives are:

RO_1: Identify the functional, construction and environmental requirements of an effective model.

RO_2: Identify relevant input and output parameters.

RO_3: Identify and characterise available data sources.

RO_4: Identify the modelling technique to be used to develop the model.

RO_5: Develop the model.

RO_6: Demonstrate the application of the model.

RO_7: Evaluate the efficacy of the model.

RO_8: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

1.6 Research scope and limitations

Due to the limited accessibility of South African data at the time of this study, readily accessible open data for New York City will be used to develop and demonstrate the prototype model. The efficacy of the model will be evaluated for the South African context, however, based on a case study of the Nelson Mandela Bay Municipality.

As already mentioned, in order to limit the scope of the investigation, the study will focus on only one aspect of smart cities, namely crime management (Figure 1-6). Chapter 5 summarises the state-of-the-art in crime forecasting, and distinguishes between the two broad focus areas of the discipline, namely place-based and person-based prediction. In order to limit the scope even further, this study will focus only on location-based crime prediction, with particular focus given to robbery and larceny.

Big Data (Section 4.2.2) is the name given to the unprecedented amount of data being generated resulting from the datafication of many aspects of our lives. Smart city connectivity is at the forefront of Big Data generation, and globally, much deliberation is being given to the development of effective hardware and software architectures necessary to store, process and analyse Big Data (Chen *et al.*, 2014). Although Big Data architecture is critical to the success of ICMPs, it will not be a focus of this study. The main focus of this study is the development of an effective modelling paradigm for integrated decision-making in smart cities.

Lastly, this study will not address important issues such as data security (Chen *et al.*, 2014), privacy and civil rights considerations inherent in crime forecasting (Perry *et al.*, 2013); and feedback loops that occur between the application environment and data products (O'Neil and Schutt, 2014). O'Neil and Schutt (2014:5) point out that "we're witnessing the beginning of a massive, culturally saturated feedback loop where our behaviour changes the product and the product changes our behaviour. Considering the impact of this feedback loop, we should start thinking seriously about how it's being conducted, along with the ethical and technical responsibilities for the people responsible for the process."

1.7 Structure of thesis

This thesis is divided into four main sections as shown in Figure 1-7. The content of each thesis chapter is discussed below. Table 1-1 provides an overview of the research questions and objectives which are addressed in each chapter.

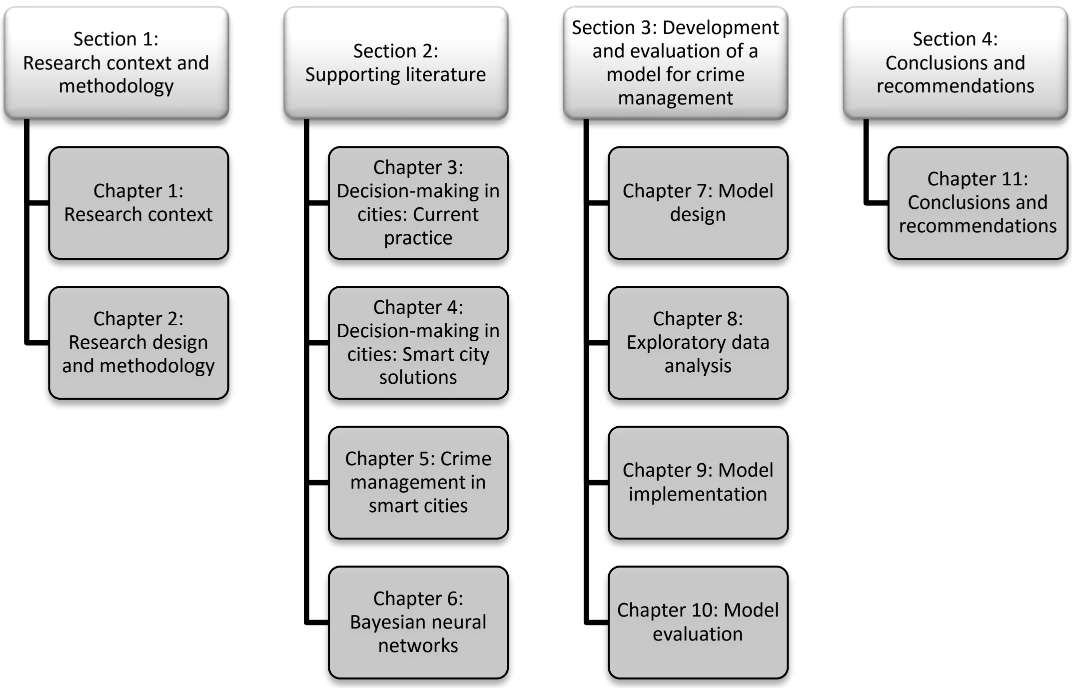


Figure 1-7: Thesis structure.

Table 1-1: Research questions, objectives and chapter deliverables.

Research questions	Research objectives	Supporting Chapters	Reporting chapters
RQ_1	RO_1	Chapters 3 and 5	Chapter 7
RQ_2	RO_2	Chapters 4 and 5	Chapter 7
RQ_3	RO_3	Chapter 4	Chapter 7
RQ_4	RO_4	Chapters 4 - 6	Chapter 7
	RO_5		Chapter 9
RQ_5	RO_6		Chapter 9
RQ_6	RO_7		Chapter 10
RQ_7	RO_8		Chapter 10

1.7.1 Section 1: Research context and methodology

Section 1 provides research context and specifies the research approach and strategies adopted in this study. In **Chapter 1 (Research context),** a brief background to the state of smart city management solutions is given from both a global and South African perspective. The research problem is introduced, and the associated research questions and objectives of this study are stated. The research contribution and scope of the study are defined, and the layout of the thesis is outlined.

In this study, the Design Science Research (DSR) process will be followed to develop, demonstrate and evaluate a model for crime management in smart South African cities. In **Chapter 2 (Research design and methodology),** the DSR paradigm will be explicated within the framework of commonly adopted research paradigms, and the research strategies and data collection methods to be employed in this study will be introduced. The DSR methodology and the concept of design science knowledge will be explained, and the application of the DSR process as applied in this study will be outlined. Lastly, ethical considerations will be noted.

1.7.2 Section 2: Supporting literature

Section 2 of this thesis provides the supporting literature necessary for problem formulation and model design. In **Chapter 3 (Decision-making in cities: Current practice)**, the requirements of a potential solution to the main research problem (Section 1.2) will be identified. This will be achieved by performing a gap analysis of current city management practices in South Africa, and by reviewing proposed information system solutions to the identified challenges.

In **Chapter 4 (Decision-making in cities: Smart city solutions)**, current smart city activities will be reviewed, and smart city design solutions aimed at meeting the model requirements identified in Chapter 3 will be proposed. An overview of the state-of-the-art in crime forecasting will then be given in **Chapter 5 (Crime management in smart cities)**. Knowledge gained in this chapter will be used to further refine the proposed solution requirements (Chapters 3) and potential smart city design interventions (Chapter 4) within the context of crime management.

Bayesian neural networks will be introduced as the modelling approach of choice in this study in **Chapter 6 (Bayesian neural networks)**. The theory explaining Bayesian neural networks will be explained, and available tools for Bayesian learning will be summarised. The implementation of Bayesian neural networks followed in this study will also be summarised.

1.7.3 Section 3: Development and evaluation of a model for crime management

Section 3 covers the practical design, implementation and evaluation of a prototype model for crime management in smart cities. In **Chapter 7 (Model design)**, the solution requirements identified in Chapters 3 and 5 will be consolidated, together with the potential smart city design solutions identified in Chapters 4 to 6. The final design of the prototype model will then be developed.

New York City open data will be used to develop and demonstrate the prototype model. Available data will be explored in **Chapter 8 (Exploratory data analysis)**.

The purpose of this exploratory data analysis is to provide context for model design in Chapter 7 and model interpretation in Chapter 9.

The prototype model will be implemented and demonstrated in **Chapter 9 (Model implementation)** according to the design specifications laid out in Chapter 7. The model will then be evaluated in **Chapter 10 (Model evaluation)**.

1.7.4 Section 4: Conclusions and recommendations

Chapter 11 (Conclusions and recommendations) concludes this thesis, and will summarise the study, discuss the achievement of the research objectives, highlight research contributions (both practical and theoretical), discuss the limitations and challenges of the study, and provide recommendations for future research.

1.8 Summary

In this chapter, a brief background to the state of smart city management solutions was given from both a global and South African perspective. The research problem was introduced (Section 1.2), and the associated research questions and objectives of this study were stated (Section 1.5). The research contribution (Section 1.4) and scope (Section 1.6) of the study were defined, and the layout of the thesis was outlined (Section 1.7). In the following chapter, the research design and methodology adopted in this study will be discussed.

Chapter 2. Research Design and Methodology

2.1 Introduction

In the previous chapter, the research problem addressed in this study was introduced, and the associated research questions and objectives were stated. The research contribution and scope of the study were defined, and the structure of the thesis was outlined. In this chapter, the research design and methodology adopted in this study will be discussed.

Research design involves the often implicit selection of a research paradigm, and the explicit selection of research strategies and data collection methods that best suit the research problem, environment and resources at hand. In this study, the Design Science Research (DSR) process will be followed to develop, demonstrate and evaluate a model for crime management in smart South African cities. Section 2.2 of this chapter explicates the DSR paradigm within the framework of commonly adopted research paradigms in the natural and social sciences, and introduces the research strategies and data collection methods employed in this study.

The DSR methodology and the concept of design science knowledge are described and explained in Section 2.3 of this chapter. The application of the DSR process as applied in this study is then outlined in Section 2.3.5. Specifically, the research objectives (Section 1.5), thesis chapters (Section 1.7), research strategies and data collection methods, and knowledge contributions are all contextualised within the DSR process. Lastly, ethical considerations are noted in Section 2.4.

2.2 Research Design

A common delineation of research design choices is the research onion (Figure 2-1) of Saunders *et al.* (2009). The outer two layers of the research onion constitute the research paradigm of a scientific community (Collis and Hussey, 2014). These layers embody the often implicit assumptions that a particular scientific community adopt with regards to reality and the attainment of knowledge (Saunders *et al.*, 2009; Collis and Hussey, 2014). In contrast, the inner layers relating to research strategies and choices are more practical in nature, and focus on how the research will be carried out. In this section, commonly adopted research paradigms will be introduced, and the DSR paradigm adopted in this study will be explained with reference to these paradigms. The research strategies and data collection methods employed in this study will also be introduced.

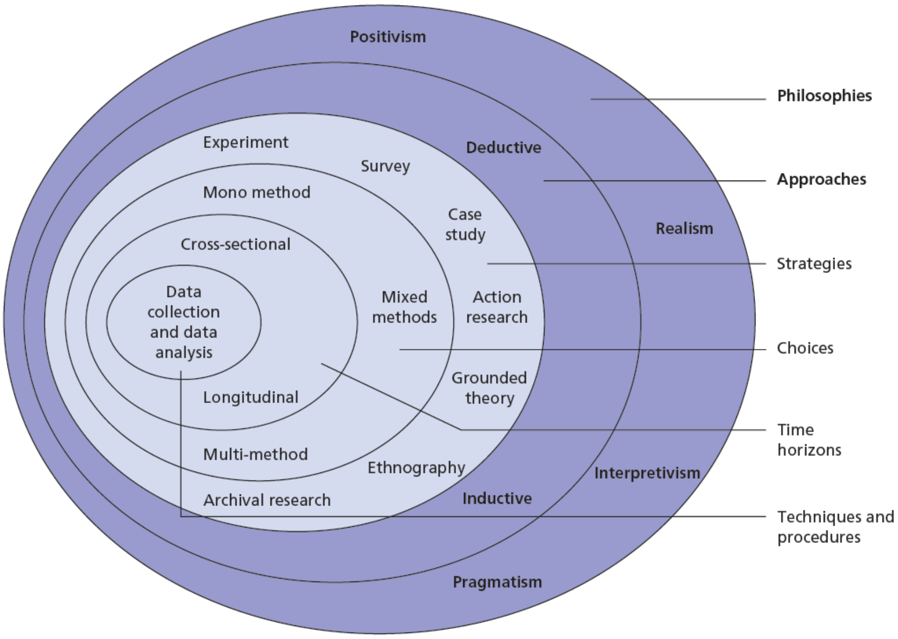


Figure 2-1: The research onion. Source: Saunders *et al.* (2009).

2.2.1 Research paradigm

The research paradigm of a particular research community encompasses the set of commonly held beliefs and assumptions held by that community about how the world works, and what is considered acceptable knowledge (Saunders *et al.*, 2009). It is a mental model comprising a set of beliefs that the research community adopts with respect to research ontology, epistemology, axiology and methodology (Vaishnavi and Kuechler, 2004; Saunders *et al.*, 2009; Johannesson and Perjons, 2012; Collis and Hussey, 2014).

Research **ontology** refers to the researcher's beliefs about reality, which entities exist and how they interact (Johannesson and Perjons, 2012), and is reflected in the researcher's perceived interaction with the system under investigation (Saunders *et al.*, 2009). Objectivism assumes that the system under investigation exists and behaves independent of the researcher concerned with its existence; while subjectivism assumes that reality is socially constructed, and that the presence of the observer influences the behaviour of the observed system (Saunders *et al.*, 2009).

Research **epistemology** refers to what is accepted as valid knowledge (Saunders *et al.*, 2009); while research **axiology** is concerned with the role of values (Saunders *et al.*, 2009) and research **methodology** is concerned with the process of research (Collis and Hussey, 2014).

Two commonly employed research paradigms are positivism and interpretivism (Saunders *et al.*, 2009; Collis and Hussey, 2014). **Positivism** is the traditional philosophy of the natural scientist (Johannesson and Perjons, 2012). Positivists assume an objective ontological philosophy (Saunders *et al.*, 2009; Collis and Hussey, 2014). They believe that there is only one reality, and everyone perceives reality in the same way (Collis and Hussey, 2014).

The positivistic philosophy developed as a reaction to theological and metaphysical world views that valued authority, divine revelation and tradition as legitimate sources of knowledge (Johannesson and Perjons, 2012). From an epistemological stand point, positivists believe that only measurable phenomena can be regarded as valid knowledge (Collis and Hussey, 2014). Positivists favour quantifiable observations that lend themselves to statistical analysis (Saunders *et al.*, 2009).

In terms of axiological assumptions, positivists believe that the process of research is value-free (Collis and Hussey, 2014). Consequently, positivists maintain an independent and objective stance (Collis and Hussey, 2014), and aim to minimise their interaction with the natural or social system under investigation (Johannesson and Perjons, 2012). Methodologically, positivists most often take a deductive approach to research (Collis and Hussey, 2014). Deductive research involves the observation of a phenomena, the development of a hypothesis that could potentially explain the phenomenon, and then the subsequent design and implementation of a research strategy to test the hypothesis (Saunders *et al.*, 2009).

Contrary to positivists, **interpretivists** assume a subjective ontological philosophy (Saunders *et al.*, 2009; Collis and Hussey, 2014). They believe that social reality is socially constructed (Saunders *et al.*, 2009; Collis and Hussey, 2014). Each person has their personal interpretation of reality and there are multiple realities (Collis and Hussey, 2014). Interpretivism emerged as a reaction to positivism (Johannesson and Perjons, 2012). Interpretivists argue that the social world can only be understood by exploring the subjective meanings and purposes that people attach to their actions (Johannesson and Perjons, 2012).

From an epistemological stand point, interpretivists attempt to minimise the distance between the researcher and the system under investigation (Collis and Hussey, 2014). Interpretivists believe that deep knowledge of a social phenomenon can be gained by actively participating in the phenomenon together with the people who actually created it (Johannesson and Perjons, 2012).

In terms of axiological assumptions, interpretivists recognise that researchers have values, and that these values determine what is recognised as facts and how these facts are determined (Collis and Hussey, 2014). Methodologically, interpretivists most often take an inductive approach to research (Collis and Hussey, 2014). In contrast to deductive research, the purpose of inductive research is not to test an hypothesis, but rather to develop a hypothesis based on deep analysis of only a few instances of the phenomena of interest (Saunders *et al.*, 2009).

Both positivism and interpretivism have associated strengths and weaknesses. A positivist approach to research is effective in identifying regularities among phenomena and deducing a generalised understanding of the cause and effect relationships between phenomena (Johannesson and Perjons, 2012). An interpretivist approach, however, will argue that while generic rules are reliable, they only provide superficial knowledge (Johannesson and Perjons, 2012). Interpretivist research activities provide deep understating of social phenomena derived from only a few instances of the phenomenon of interest. While interpretivists provide deep knowledge, the knowledge is not readily generalised (Johannesson and Perjons, 2012).

In contrast to pure positivist or interpretivist philosophical stances, **pragmatists** are flexible in their philosophy, and adopt either stance depending on the research problem (Saunders *et al.*, 2009). Johannesson and Perjons (2012) suggest that in order to overcome the weaknesses of the respective research approaches, they can be combined. An interpretivist approach can be followed to generate a hypothesis, while a positivist approach can be used to verify the hypothesis.

The **DSR** paradigm is not bound to fixed ontological, epistemological and axiological assumptions, and can accommodate either a positivist or interpretivist research paradigm (Johannesson and Perjons, 2012). The DSR approach differs from conventional deductive and inductive research approaches, however, in the sense that the latter focus on understanding reality, while design science research focuses on utility, and aims to create and evaluate artefacts intended to solve a practical problem (Vaishnavi and Kuechler, 2004; Peffers *et al.*, 2008; Johannesson and Perjons, 2012).

2.2.2 Research approach, strategies and methods

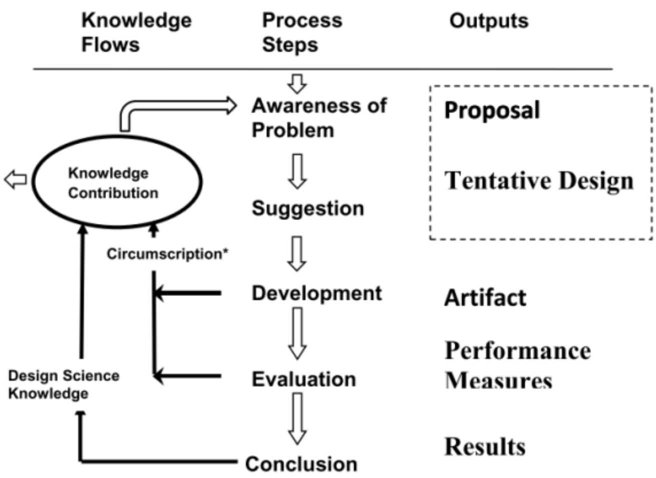
In this study, the DSR paradigm will be adopted to develop and evaluate a model for crime management in smart cities using a deductive research approach. The application of the DSR process as applied in this study is outlined in Section 2.3.5. A proposed design will be determined by way of a literature review through which design requirements and available modelling approaches will be identified. The proposed design will then be developed and evaluated.

As indicated in the research onion (Figure 2-1), commonly employed research strategies include experiments, surveys, case studies, action research, grounded theory, ethnography and archival research (Saunders *et al.*, 2009; Collis and Hussey, 2014). In this study, archival research, simulation experiment and case study strategies will be employed. The contextual motivation for these choices will be elucidated in Section 2.3.5.

2.3 Design Science Research

2.3.1 Definition and process overview

This study will adopt a learning-through-doing research strategy, formally known as design science research (DSR) (Vaishnavi and Kuechler, 2004; Peffers *et al.*, 2008; Hevner and Chatterjee, 2010; Johannesson and Perjons, 2012; Vaishnavi and Kuechler, 2015). To identify the missing knowledge in a new area of design, it is useful to carry out the design using existing knowledge. This generates knowledge about the extent of missing knowledge and the challenges faced in filling the knowledge gaps (Vaishnavi and Kuechler, 2004). Hevner and Chatterjee (2010:5) provide a formal definition of DSR: "Design science research is a research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artifacts, thereby contributing new knowledge to the body of scientific evidence. The designed artifacts are both useful and fundamental in understanding that problem."



Circumscription is the discovery of constraint knowledge about theories gained through detection and analysis of contradictions when things do not work according to theory.

Figure 2-2: DSR process model. Source: Vaishnavi and Kuechler (2004).

The DSR process (Figure 2-2) is described as a cycle in which existing knowledge is used to design and construct artefacts, and the artefacts are evaluated to build knowledge (Vaishnavi and Kuechler, 2004). A DSR artefact can be in the form of a construct, model, method or instantiation. The different types of DSR artefacts are further explained in Section 2.3.4.

In the DSR process (Figure 2-2) all design begins with an awareness of a problem, and the first step in the process is to define the problem (Hevner and Chatterjee, 2010). The next stage is a preliminary suggestion for a problem solution. Once a tentative design is settled on, the next stage is the development of the design (Hevner and Chatterjee, 2010). In this stage, the design is further refined and an artefact is produced (Hevner and Chatterjee, 2010). Once the artefact is ready, it is evaluated according to implicit or explicit functional specification in the suggestion (Hevner and Chatterjee, 2010). The output of a DSR project is a novel artefact and design science knowledge (Vaishnavi and Kuechler, 2004; Johannesson and Perjons, 2012).

2.3.2 Design science knowledge

Critical to DSR is design science knowledge which distinguishes DSR from routine design (Vaishnavi and Kuechler, 2004; Peffers *et al.*, 2008; Hevner and Chatterjee, 2010; Johannesson and Perjons, 2012; Vaishnavi and Kuechler, 2015). Design science research not only produces a novel artefact, but it also produces design science knowledge about the artefact and its interaction with its environment (Johannesson and Perjons, 2012). Design science knowledge can be embedded in the developed artefact, or it can be in the form of design theories which have been derived from the DSR process (Vaishnavi and Kuechler, 2004).

As highlighted in Section 2.3.4, design science knowledge is embedded in DSR artefacts. Constructs hold definitional knowledge; models hold descriptive, prescriptive or predictive knowledge; and methods hold prescriptive knowledge. In addition to this embedded knowledge, the DSR process also aids in the development of design theories.

During the design process, existing design knowledge regarding model requirements and intended practice, environmental conditions and the artefact's anticipated interaction with its environment, and artefact construction (Figure 2-3) are made explicit (Johannesson and Perjons, 2012). These assumptions regarding design best practice are then implemented and tested, and refined throughout the DSR process. The purpose of the DSR process therefore is to develop or refine existing design theories. Hevner and Chatterjee (2010) emphasise this aspect of DSR by highlighting the role of circumscription (Figure 2-2) in the DSR process.

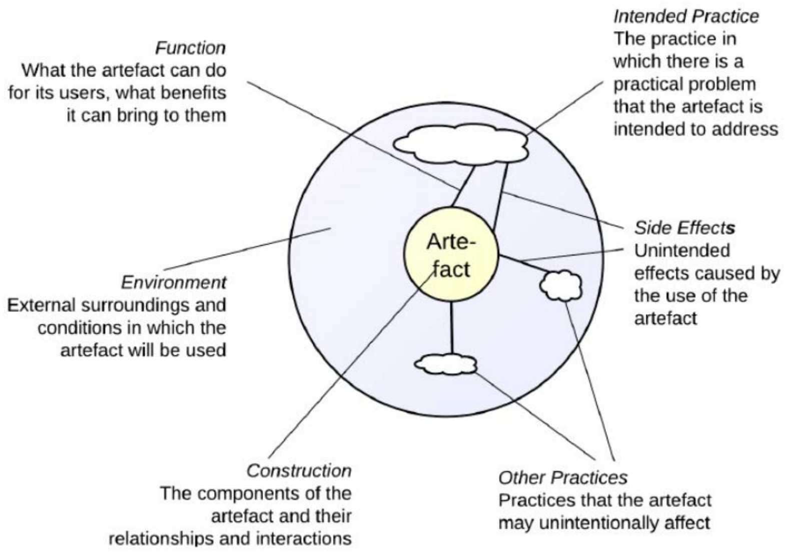


Figure 2-3: DSR artefact construction, function and environment. Source: Johannesson and Perjons (2012).

2.3.3 Distinction between DSR and general design

Design science knowledge distinguishes DSR from general design practice. DSR and design both focus on the development of novel solutions to problems, however they differ in purpose with respect to generalisability and knowledge contribution (Johannesson and Perjons, 2012). While general design work may only be relevant to a single client, DSR aims to produce and communicate design knowledge that is of general interest (Johannesson and Perjons, 2012). DSR therefore, makes use of rigorous research methods, has an emphasis on communicating results to both practitioners and researchers, and builds upon existing knowledge bases to ensure that proposed results are well founded and original (Johannesson and Perjons, 2012).

Vaishnavi and Kuechler (2004) distinguish routine design from DSR by the number of unknowns in the proposed design or the amount of missing knowledge. Figure 2-4 is a useful aid for distinguishing between DSR and routine design. This study falls within the adaptation domain of Figure 2-4, and therefore classifies as a DSR project.

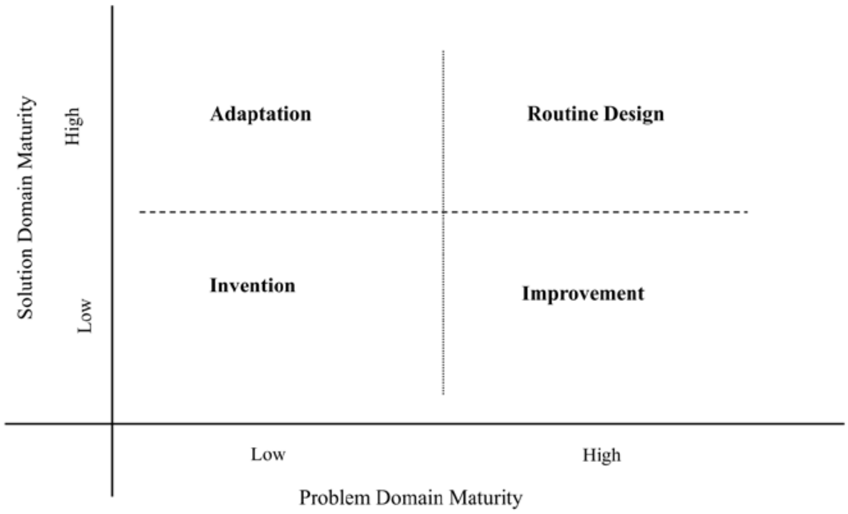


Figure 2-4: DSR knowledge contribution framework. Source: Vaishnavi and Kuechler (2004).

2.3.4 DSR artefacts

A DSR artefact can be any designed object with an embedded solution to a research problem (Peffers *et al.*, 2008). The four main types of DSR artefacts are constructs, models, methods and instantiations (Peffers *et al.*, 2008; Johannesson and Perjons, 2012).

Constructs are a form of definitional knowledge comprising terms, notations, definitions, and concepts that are needed for formulating problems and their possible solutions (Johannesson and Perjons, 2012). Constructs do not make any statements about the world, but they provide the conceptual vocabulary (Vaishnavi and Kuechler, 2004) that make it possible to formulate problems and communicate solutions (Johannesson and Perjons, 2012). Examples include the concepts of methods in Java and functional dependencies in relational database theory (Johannesson and Perjons, 2012).

A **model** is an artefact that expresses relationships among constructs (Vaishnavi and Kuechler, 2004). A model can be descriptive (representing an existing situation which can be used for describing and analysing problem situations), prescriptive (describing potential solutions to practical problems), or predictive (can be used to forecast the behaviour of objects and systems) (Johannesson and Perjons, 2012). In design science the focus is often on prescriptive models e.g. business process models and systems architecture (Johannesson and Perjons, 2012).

A **method** is a form of prescriptive knowledge and defines an algorithm, process or set of guidelines that can be used to solve a problem (Vaishnavi and Kuechler, 2004; Johannesson and Perjons, 2012). Examples include methods for database design or change management initiatives (Johannesson and Perjons, 2012).

Instantiations are working systems that can be used in practice (Vaishnavi and Kuechler, 2004; Johannesson and Perjons, 2012). An instantiation can be the realisation of a DSR artefact (construct, model or method) in an environment (Vaishnavi and Kuechler, 2004), or it can be an existing working solution that can be investigated to gain design science knowledge.

2.3.5 Application

A commonly used DSR methodology for information systems research is that of Peffers *et al.* (2008). In this section, each activity in the methodology and its application in this research project is explained. The DSR process model of Peffers *et al.* (2008) is shown in Figure 2-5. The process model closely resembles that presented by Johannesson and Perjons (2012). Both Peffers *et al.* (2008) and Johannesson and Perjons (2012) stress that, although the method is presented in a sequential manner, a DSR project is an iterative process (Johannesson and Perjons, 2012); and the method is not prescriptive with respect to work order (Peffers *et al.*, 2008; Johannesson and Perjons, 2012).

Table 2-1 summarises the DSR process as applied in this study. Specifically, Table 2-1 specifies the research strategies and methods employed in each of the six DSR activities, together with the relevant research objectives (Section 1.5) and thesis chapter(s) (Section 1.7).

Table 2-1: Design science research process as applied in this study. Based on the methodologies prescribed by Peffers *et al.* (2008) and Johannesson and Perjons (2012).

Activity	Identify problem and motivate	Outline artefact and define requirements	Design and Development	Demonstrate	Evaluate	Communicate
Research Strategies and methods	Literature survey	Literature survey	Literature survey; Archival Research (NYC open data)	Simulation experiments	Mixed-methods case study: interview and document study	Articles; Conferences; Thesis; NMBM collaboration; Future studies
Research Objectives		RO_1	RO_{2-5}	RO_6	RO_{7-8}	
Reporting Thesis Chapters	1	1, 7	7, 9	9	10	
Supporting Thesis Chapters	3	3, 5	3-6, 8	7, 8	7, 9	

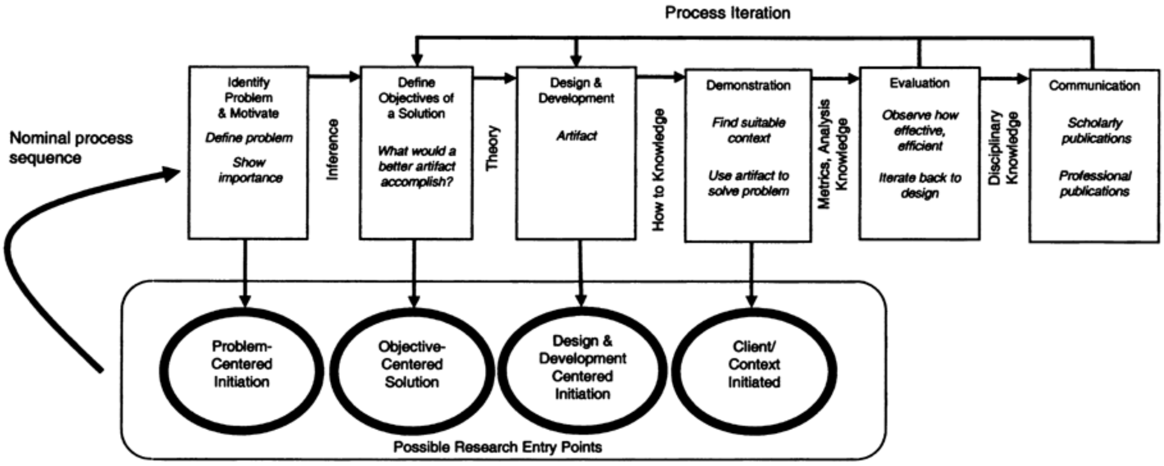


Figure 2-5: DSR methodology process model of Peffers *et al.* (2008).

2.3.5.1 Activity 1: Problem identification and motivation

The first step in the DSR process is to define the practical problem to be solved, and to justify the value of solving the problem (Peffers *et al.*, 2008; Johannesson and Perjons, 2012). In this study, the research problem was formulated and justified by carrying out a literature survey of current decision-making practices and challenges in South African cities (Chapter 3). The research problem was introduced and justified in Chapter 1.

2.3.5.2 Activity 2: Define the objectives of a solution

The next step in the DSR process is to define the objectives of a solution (Peffers *et al.*, 2008). Specifically, a solution to the problem identified in Activity 1 should be outlined in the form of an artefact, and the requirements of the artefact should be clearly defined (Johannesson and Perjons, 2012).

Requirements can be seen as a transformation of the problem into demands on the proposed artefact (Johannesson and Perjons, 2012). Artefact requirements defined in Activity 2 will be used in Activity 5 to measure the efficacy of the developed artefact. When defining artefact requirements, the designer should consider functional requirements as well as construction and environmental requirements (Johannesson and Perjons, 2012) (see Figure 2-3).

The choice of artefact and artefact requirements requires knowledge of the problem, the state of current solutions and their efficacy, and what is feasible (Peffers *et al.*, 2008). In this study, artefact requirements were formulated in Chapter 3, based on a gap analysis of current city management practices in South Africa, and a review of proposed information system solutions to the identified challenges. Solution requirements were further refined in Chapter 5, within the context of crime management.

The proposed artefact was outlined in Chapter 1 (Sections 1.3-1.5). It was proposed that a predictive model for integrated crime management can be developed that effectively utilises data from traditionally isolated management silos. It was anticipated that such a model could be used for prediction and simulation at the system-of-systems level. Model requirements identified in Chapters 3 and 5 are consolidated in Chapter 7.

2.3.5.3 Activity 3: Design and development

In Activity 3, the artefact is designed and developed. Based on the chosen artefact and its requirements as defined in Activity 2, this activity focuses on determining the artefact's architecture and then creating the actual artefact (Peffers *et al.*, 2008; Johannesson and Perjons, 2012). Each component of the artefact is to be clearly defined, and the purpose of each component is to be explained with regards to which requirement it addresses (Peffers *et al.*, 2008).

Design elements aimed at meeting each of the artefact requirements prescribed in Activity 2 were identified in Chapter 4, based on a literature review of the state-of-the-art in smart city management solutions. The design elements identified in Chapter 4 were further refined in Chapter 5 within the context of crime management. Design elements identified in Chapters 4 and 5 were then consolidated in Chapter 7. The model design was developed in Chapter 7, and the prototype model was implemented in Chapter 9 according to the design specifications presented in Chapter 7.

In Chapter 1 (Sections 1.3-1.5), it was proposed that a predictive model for whole-system scenario analysis can be developed by building upon emerging smart city management solutions. It was proposed that this can be achieved by using smart city KPI frameworks to represent data from traditionally isolated management silos as a set of sectoral KPIs (Figure 1-5a). It was anticipated that the inter-dependencies between these KPIs can be encapsulated in an artificial neural network (Figure 1-5b), which can be used for prediction and simulation at the system-of-systems level (Figure 1-5c). The aim of this study was to test the feasibility of this supposition by developing and evaluating a prototype model.

Due to the limited accessibility of South African data at the time of this study, readily accessible open data archives for New York City were used to develop and demonstrate the prototype model. Datasets used in this study are summarised in Section 7.6.

2.3.5.4 Activity 4: Demonstration

The goal of Activity 4 is to demonstrate the use of the developed model (Peffers *et al.*, 2008), with the intention of proving the feasibility of the model to solve an instance of the identified problem (Johannesson and Perjons, 2012). Appropriate activities include experimentation, simulation, case study, proof etc. (Peffers *et al.*, 2008). In this study, the application of the prototype model was demonstrated by way of simulation experiments in Chapter 9. Specifically, it was shown how the neural network developed in Activity 3, coupled with sensitivity analysis, could be used as a decision-making tool at the system-of-systems level.

2.3.5.5 Activity 5: Evaluation

The purpose of Activity 5 is to evaluate the developed model based on the model requirements specified in Activity 2 (Peffers *et al.*, 2008; Johannesson and Perjons, 2012). In Chapter 10, the prototype model was evaluated *ex ante* (Peffers *et al.*, 2008) within the South African context by way of a mixed-method case study. Nelson Mandela Bay Municipality (NMBM) was selected as the most feasible metro to investigate as research agreements were already in place between the Nelson

Mandela University and the NMBM. The efficacy of the model was evaluated by way of an interview with NMBM employees, and a document study of their IDP and performance management systems.

2.3.5.6 Activity 6: Communication of DSR knowledge

The final activity in the DSR process is to communicate design science knowledge. It is anticipated that the research contributions of this study will be in the form of a prototype model demonstrating the proposed integrated approach to crime management in smart cities, accompanied by a set of model implementation guidelines for South African cities. It is anticipated that the theoretical contribution of this study will be in the form of design science knowledge regarding the efficacy of the proposed approach. Design science knowledge generated in this study has and will be communicated through the following platforms:

- Research and conference papers (a copy of the published conference paper originating from this study is provided in Appendix 2);
- This doctoral thesis;
- Collaborative meetings with local municipal officials; and
- Future graduate studies.

2.4 Ethical considerations

Ethical approval was not required for this study, as no vulnerable groups participated in the study and data was largely sourced from open data sources.

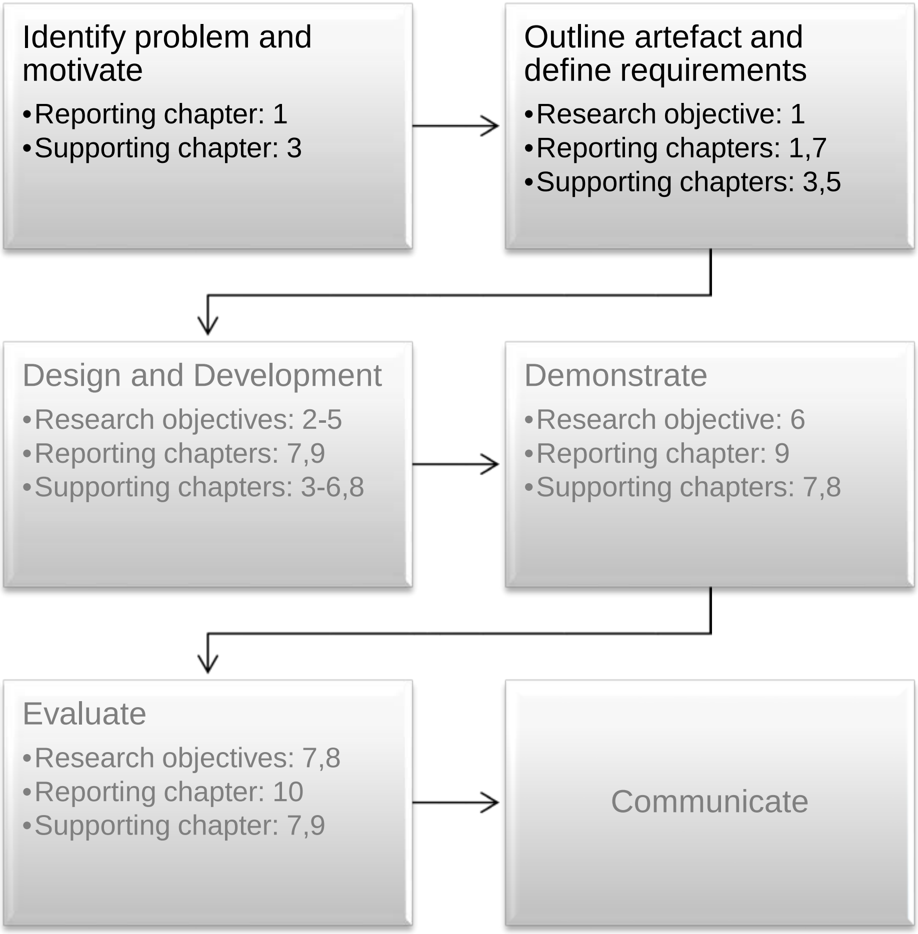
2.5 Summary

In this study, the DSR process will be followed to develop, demonstrate and evaluate a prototype model for crime management in smart South African cities. In this chapter, the DSR paradigm was explicated within the framework of commonly adopted research paradigms, and the research strategies and data collection methods to be employed in this study were introduced (Section 2.2). The DSR methodology and the concept of design science knowledge were explained, and the application of the DSR process as applied in this study was outlined (Section 2.3). Lastly, ethical considerations were noted (Section 2.4).

The following Section of this thesis will provide the supporting literature necessary to formulate the research problem, identify solution requirements and propose smart city design interventions. Reviews of current decision-making practices in South African cities, and emerging smart city management solutions will be presented in Chapters 3 and 4, respectively. An overview of crime management in smart cities will be given in Chapter 5, while Bayesian neural networks will be introduced in Chapter 6 as the modelling approach of choice in this study.

SECTION 2: SUPPORTING LITERATURE

Chapter 3. Decision-making in Cities: Current Practice



Research objective addressed in this chapter:

RO$_1$: Identify the functional, construction and environmental requirements of an effective model.

RO$_2$: Identify relevant input and output parameters.

RO$_3$: Identify and characterise available data sources.

RO$_4$: Identify the modelling technique to be used to develop the model.

RO$_5$: Develop the model.

RO$_6$: Demonstrate the application of the model.

RO$_7$: Evaluate the efficacy of the model.

RO$_8$: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

Figure 3-1: Research objective and design science research activities addressed in this chapter.

3.1 Introduction

In this study, the Design Science Research (DSR) process will be followed to develop, demonstrate and evaluate a prototype model for crime management in smart South African cities. In the previous chapter, the DSR methodology and the concept of design science knowledge were explained, and the application of the DSR process as applied in this study was outlined. The research strategies and data collection methods to be employed in this study were also introduced.

The purpose of this chapter (Figure 3-1) is to provide the supporting literature necessary to formulate the research problem, and to identify the requirements of a potential solution. The research problem addressed in this study was introduced in Section 1.2 of this thesis. This chapter addresses RO_1 (Figure 3-1) by defining the functional, construction and environmental requirements of a potential solution to the stated research problem.

Solution requirements will be identified by performing a gap analysis of current city management practices in South Africa. This will be achieved by first defining the goals of city planning and management (Section 3.2). The complex nature of cities will then be explained (Section 3.3); and the implications of complexity for management will be explicated in the form of principles for management best practice (Section 3.4).

Current city management practices in South Africa will then be introduced (Section 3.5); and will be critically evaluated against the management principles identified in Section 3.4. The results of this gap analysis are reported in Section 3.6. The challenges identified in Section 3.6 formed the basis for formulating the research problem stated in Section 1.2.

Proposed information system solutions to these challenges will be discussed in Section 3.7. Section 3.8 concludes this chapter with a list of proposed solution requirements based on these recommendations.

3.2 The goal: Sustainable and resilient cities

The drive for smart cities is rooted in the Sustainable Development Goals (SDGs) (Figure 3-2) which call for development that synergistically meets basic human needs, promotes job creation and economic growth, reduces inequality, and respects the natural environment (The Economist, 2009). The SDGs, as defined in *Transforming Our World - the 2030 Agenda for Sustainable Development* (United Nations., 2015), is the global development agenda that replaced the Millennium Development Goals upon their expiration in 2015.

Figure 3-2: Sustainable Development Goals. Source: Project Everyone (2016).

The term “sustainable development” was first used in 1987 by the World Commission on Environment and Development (WCED) in their Brundtland report. The Brundtland report (formally titled “Our Common Future”) was the result of an urgent call by the General Assembly of the United Nations for the WCED to formulate "a global agenda for change” (World Commission on Environment and Development., 1987). The report was commissioned amidst global concerns relating to resource depletion, pollution and environmental degradation resulting from the unbridled growth of developed countries. These challenges were intensified by the strong need for development and growth in developing countries.

In their report, the WCED defined sustainable development as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (World Commission on Environment and Development, 1987: 41). Sustainable development realises the need for development and economic growth in order to meet the needs of the world's poor, but at the same time realises that there is limits to growth, and that growth needs to take place within the carrying capacity of the earth.

Cities have a major impact on human quality of life and the natural environment; and, although different cities face different challenges (IEC., 2015), the driving force behind all city planning and management is to create sustainable and resilient cites (United Nations., 2015). City managers are tasked with addressing a plethora of socio-economic challenges, such as service delivery, transportation, education, job creation, food security, public safety, healthcare services and quality of life. This must be achieved against the pressures of aging infrastructure, fiscal and socio-political constraints, and limitations in the sink and source capacity of the environment. The latter are all intensified by rapid urbanisation and the increasing effects of climate change (Kourtit *et al.*, 2014).

In South Africa, for example, the total population living in urban areas is expected to increase from just over 65% in 2018 to 80% in 2050 (United Nations, 2018). This will translate into an urban population increase of over 10 million people. Now, more than ever, city officials are being called to manage increasingly stressed resources with unprecedented efficiency (IBM., 2010; IEC., 2015). Moreover, the increasing effects of climate change requires a high degree of resilience to be imbedded into the design and management of cities (The Rockefeller Foundation | Arup., 2015).

While sustainability indicators focus on the social, economic and environmental performance of cities (McCarney, 2015), urban resilience focuses on the capacity of individuals and systems within a city to survive, adapt and thrive in the face of chronic stresses and acute shocks (Rockefeller Foundation, 2016). Over and above meeting basic human needs, fostering economic prosperity, and protecting assets, resilient cities focus on emergency response, reliable communication and mobility, and inclusive and integrated planning and management informed by data

(McCarney, 2015; The Rockefeller Foundation | Arup., 2015; United Nations., 2015; The Rockefeller Foundation | Arup., 2018).

3.3 The city as a complex system

In light of the different agendas at play within a city, cities are often seen as a multifaceted system of functional sub-systems (Fernández-Güell *et al.*, 2016). Each sub-system is composed of an integrated web of stakeholders from both the public and private sectors; all trying to, independently, optimise their respective interactions within a particular domain (IBM., 2010; IEC., 2015; Schleicher *et al.*, 2016).

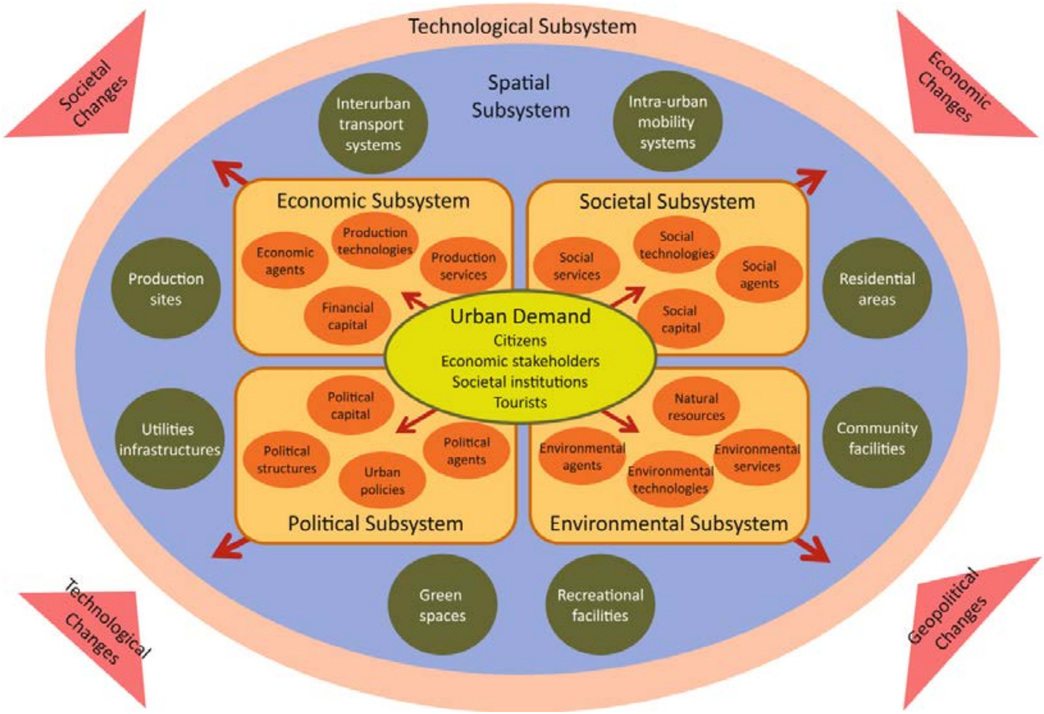


Figure 3-3: The city as a system of functional sub-systems. Source: Fernández-Güell *et al.* (2016).

Despite the fact that the large majority of decisions are made independently within functional sub-systems (IBM., 2010; IEC., 2015; SALGA., 2015), city systems are highly interdependent (Schleicher et al., 2016). In fact, economists estimated in 2010 that on average almost 50 percent of economic outputs generated by a particular sub-system relies on input from other systems (IBM., 2010). The model of functional systems developed by Fernández-Güell *et al.* (2016) (Figure 3-3) is just one of many such models aimed at encapsulating this diversity and interdependence in cities (Chourabi *et al.*, 2012; City Protocol Society., 2015b; IEC., 2015; Kourtit *et al.*, 2017).

Complex systems are defined to be systems that consist of many interdependent components and have distinct properties arising from these inter-dependencies, such as adaptability, emergence, self-organisation, attractors, chaos and non-linearity (OECD., 2009; Haken, 2012; Fernández-Güell *et al.*, 2016; Wikipedia, 2018d). Cities, by their very nature, are complex systems (Allen, 1997a; Allen, 1997b; OECD., 2009; Batty and Marshall, 2012; Haken, 2012; Kourtit *et al.*, 2014; Angelidou, 2015; Fernández-Güell *et al.*, 2016).

As explained by Fernández-Güell et al. (2016: 87), "[A] city is built from multiple singular initiatives taken through time by a great number of players who are tightly interconnected among themselves. In this ecosystem, any spatial or structural alteration in one of its elements can modify the other parts of the system. Cities, understood as complex systems, are adaptive as they evolve and are not readily predictable because they do not necessarily act in a deterministic fashion."

3.4 Implications of complexity for management

Managing complex city systems independently from within sectoral silos can result in poor performance, especially when inter-dependencies and feedback loops result in unexpected externalities and outcomes. This concern becomes more pronounced as cities continue to grow (Hardin, 1968), and become more connected (Schleicher *et al., 2015*). In recent decades, there has been a ubiquitous call for an integrated approach to managing cities (IBM., 2010; Chourabi *et al.*, 2012; City Protocol Society., 2015b; IEC., 2015; Mattoni *et al.*, 2015; Fernández-Güell *et al.*, 2016; Schleicher *et al.*, 2016; Kourtit *et al.*, 2017; Mattoni *et al.*, 2017).

While the call for integrated city planning and management is relatively new, the tools for managing complex socio-ecological systems have been under development for decades. Since the start of the environmental movement in the 1960's, scientists and resource managers have been exploring ways of successfully managing complex socio-ecological systems (DEAT., 2004); and a number of management paradigms and principles have been developed to take into account the inherent uncertainty in complex systems (DEAT., 1998; OECD., 2009; Biggs *et al.*, 2015).

In order to limit externalities, a key principle in managing complex systems is to promote integrated decision-making. Integration in this context implies a whole-systems approach to decision-making (IBM., 2010); and practically refers to the vertical (within sectoral silos) and horizontal (inter-silo) integration of all stakeholders and sustainability concerns (social, economic and ecological) into all decision-making processes across the full life-cycle of activities (DEAT., 2004).

Here, the life-cycle of activities refers to each stage of development, from plans and programmes, to project design, to project implementation and monitoring, to project decommissioning (DEAT., 2004). The desired outcomes of integrated decision-making is sustainability-led optimisation of financial and environmental resources, and the minimisation of socio-economic and ecological externalities.

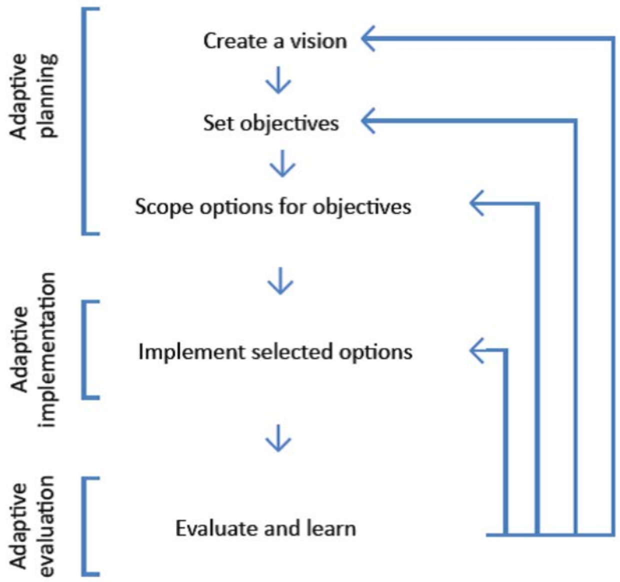


Figure 3-4: Schematic summary of the steps in the strategic adaptive management process. Source: Biggs *et al.* (2015).

Another key principle in managing complex socio-ecological systems is adaptive management (Biggs *et al.*, 2015). In order to manage for unexpected outcomes resulting from feedback loops and other properties of complex systems, adaptive management methods aim to identify a set of system performance indicators, and then monitor the impact of interventions to ensure the desired outcome of decisions. The steps involved in the strategic adaptive management process are summarised by Biggs *et al.* (2015) in Figure 3-4.

At the operations level, the need for systems-level optimisation is also apparent (Hardin, 1968; IBM., 2010; IEC., 2015). In 2010, IBM economists estimated that inefficiencies were costing the world US$15 trillion annually in waste and lost resources (IBM., 2010). IBM. (2010) noted that more than 50% of food supplied was not eaten, nearly 35% of water designated for agricultural use was being wasted, 25% of electricity generated was not being used, and inefficiencies of over 35% in the Healthcare, Government and Safety, and Education sectors were being observed.

Over and above the direct costs, the indirect costs associated with these inefficiencies had secondary effects in other areas such as consumer spending, pollution and quality of life indicators (IBM., 2010). These huge levels of inefficiency were attributed to managing city systems from within sectoral silos; and, while some level of inefficiency is inescapable, economists estimated an annual global savings potential of US$4 trillion should a systems approach be adopted (IBM., 2010).

Based on the above analysis, three key management principles, namely *effective stakeholder engagement*, *adaptive management* and *sustainability-led systems-level optimisation,* across the full activity life-cycle*,* have been identified as vital for effective and efficient integrated city planning and management. These principles will be used as a benchmark when assessing the effectiveness of existing management tools in Section 3.6.

3.5 Current management practices in South Africa

In South Africa, integrated city planning and management is supported mainly through the implementation of Integrated Development Plans (RSA., n.d.) and their supporting Spatial Development Frameworks (RSA., 2011) required by the Local Government: Municipal Systems Act, 32 of 2000 (RSA., 2000). In addition, integrated decision-making is supported through the application of Integrated Environmental Management principles and tools, promoted and regulated by the National Environmental Management Act, 107 of 1998 (DEAT., 1998).

3.5.1 Integrated development plans

The main aim of an Integrated Development Plan (IDP) is to accelerate service delivery in municipalities and to deliver the spatial, social, ecological and economic urban patterns that are in line with the country's democratic and sustainability visions (Ngamlana and Eglin, 2015; Mnguni, 2016). An IDP is a principal strategic development plan prepared by local government for a five year period, which guides and informs all planning, budgeting, management and decision-making in a municipality (RSA., n.d.). The purpose of an IDP is to identify the development needs and concerns in a municipality and then formulate and prioritise possible intervention programmes and projects in an inclusive and integrated manner, taking into account Key Performance Indicators (KPIs) and targets, stakeholder concerns, spatial development considerations, budgeting constraints and legislative and sectoral considerations (RSA., n.d.).

A Spatial Development Framework (SDF) is a core component of an IDP and consists of an in-depth mapping of the bio-physical, socio-economic and built environment status quo, patterns and trends in a municipality (RSA., 2011). It quantifies the spatial development needs (e.g. housing and mobility) and infrastructure capacities within the municipality and qualitatively assesses the performance of the municipality against desired spatial form and principles (RSA., 2011). The purpose of an SDF is to identify the location of spatial tools (e.g. nodes, corridors, infill and densification, containment, protection and growth areas) that are required to meet a municipality's priorities (RSA., 2011). It is imperative that both the IDP and the SDF are guided by comprehensive stakeholder engagement programmes (RSA., 2011; RSA., n.d.).

3.5.2 Environmental assessment

Integrated Environmental Management (IEM) is a South African term that is equivalent to the globally applied term Environmental Assessment and Management (DEAT., 2004). IEM aims to support sustainable development through the use of a wide range of environmental assessment and management tools – such as Environmental Impact Assessments and Environmental Management Systems –

throughout the full activity life-cycle and by all sectors of society (DEAT., 2004). An overview of commonly used IEM tools at each stage in the activity life-cycle is shown in Figure 3-5 (DEAT., 2004).

IEM tools foster stakeholder engagement and the incorporation of sustainablity considerations into the development, evalution and monitoring of strategic level plans, programmes and policies. IEM tools also provide an effective framwork for incorporating sustainability considerations into project-level planning and design processes, as well as into the establishment, operations and closure of projects.

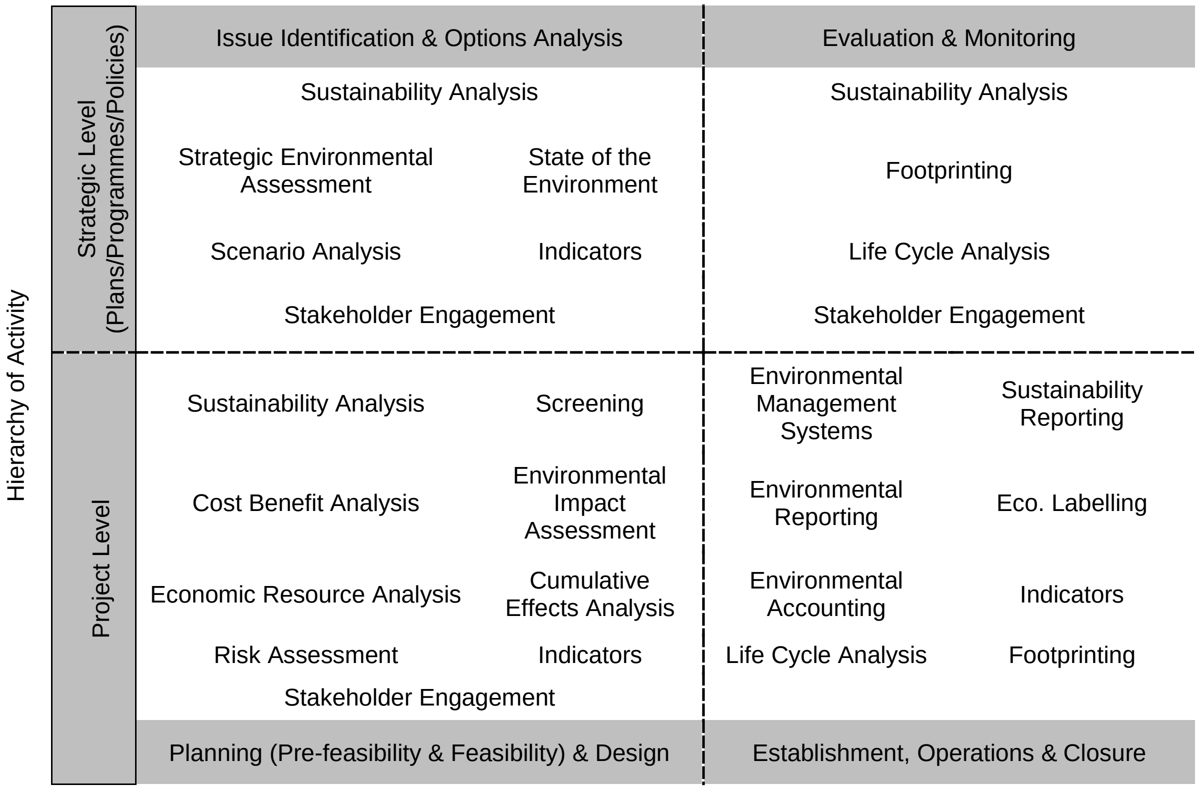


Figure 3-5: An overview of commonly used IEM tools at each stage in the activity life-cycle. Source: Adapted from DEAT. (2004).

Although the National Environmental Management Act (NEMA) (107 of 1998) (DEAT., 1998) promotes the application of a wide range of IEM tools (DEAT., 1998; DEAT., 2004), project-level Environmental Impact Assessment (EIA) (RSA., 2012a; RSA., 2017) and more recently, strategic level Environmental Management Frameworks (EMF) (RSA., 2012b) are the only regulated tools for IEM in South Africa (DEA., 2014). Project-level EIA requires that the social, biophysical and other impacts associated with certain listed activities be identified, predicted, evaluated and mitigated prior to major decisions being made (RSA., 2017; Wikipedia, 2018e). While strategic-level EMFs are aimed at providing context for project-level environmental authorisations (DEA., 2014; Cilliers and Retief, 2016).

EIAs are prepared by independent consultants, namely Environmental Assessment Practitioners (EAPs), who typically compile a report based on a number of specialist impact studies identified in the scoping stage of the EIA (RSA., 2017). Listed activities may only proceed once the relevant government Competent Authority approves the EIA and its accompanying Environmental Management Plan (EMP) aimed at avoiding and mitigating potential impacts.

The purpose of the EMF is to function as a support mechanism in the EIA process and to inform decision-making regarding land-use planning applications in IDPs and SDFs (RSA., 2012b). By performing a status quo assessment (including sensitivity analysis, environmental opportunities and constraints), the EMF makes significant and detailed spatial information available about a specific geographical area. In addition to the status quo, EMFs identify the desired state of the environment and propose the way forward by identifying specific management zones and management guidelines. Management guidelines include limits and cumulative impacts, the identification of existing impacts to be addressed, the identification of the scope of potential impacts and information needs of EIAs, and the identification of activities requiring EIAs in delineated geographical areas.

3.6 Gap analysis

In Section 3.4, three key management principles, namely *effective stakeholder engagement*, *adaptive management* and *sustainability-led systems-level optimisation,* across the full activity life-cycle, were identified as vital for effective and efficient integrated city planning and management. A literature review of the effectiveness of the existing management tools summarised in Section 3.5 showed that they fail to effectively achieve these goals. The results of this literature review are summarised below.

Despite the positivity embodied in the introduction of the key policy tools aimed at achieving sustainable socio-economic reform in South Africa, there exists a chronic disconnect between planning and implementation (DEA., 2014; Ngamlana and Eglin, 2015). The current IEM system is often criticised as being inefficient and ineffective (DEA., 2014) and IDPs have consistently failed to achieve transformation (Ngamlana and Eglin, 2015). The number of civic protests in South Africa is on the increase, with over 100 protest incidents every year (Mnguni, 2016). The level of violence associated with these protests is also rising (Mnguni, 2016). The main reason cited for violent protests has been dissatisfaction with municipal service delivery (Mnguni, 2016).

Effective service delivery has been adversely affected by a crippling lack of capacity within local government (Mnguni, 2016), staggering estimates of corruption and fruitless expenditure (Mnguni, 2016), political agendas, and factional battles within governing parties (Mnguni, 2016). Although to a lesser degree, similar concerns regarding quality and ethics have plagued the implementation of EIA and EMF (DEA., 2014; Marais *et al.*, 2014).

Across all activity levels there is a general lack of understanding and internalisation of sustainability principles (DEA., 2014; Ruwanza and Shackleton, 2015). *Sustainability-led development* aims to shift the focus of environmental management away from impact mitigation and instead, aims to focus on synergistic solutions that optimise positive socio-economic and ecological outcomes. However, it was found that the main focus of environmental intervention in South Africa is on impact identification, mitigation and compliance; and sustainability-led development has not been embraced (DEA., 2014; Ruwanza and Shackleton, 2015). This is made evident by the limited scope of KPIs used in IDPs (Marais *et al.*, 2008; Mautjana and Mtapuri, 2014) and the associated limited focus on environmental issues in IDPs (Ruwanza and Shackleton, 2015).

Another concern is the lack of *integration* and *effective cooperative governance.* Sector plans are still largely drawn up independently when preparing IDPs and “the level of integration is determined by the degree to which municipal departments talk to one another” (Ngamlana and Eglin, 2015: 4). In addition, there is a lack of coordination across the spheres of government, resulting in fragmented, and sometimes conflicting, planning and implementation (Ngamlana and Eglin, 2015). Regarding project-level EIA, fragmentation and duplication of authorisation processes often lead to frustration (DEA., 2014). Even in EIA reporting, there is a lack of integration, with numerous specialists' reports being prepared independent of each other (Morrison-Saunders *et al.*, 2014).

A major flaw in IEM in South Africa is the absence of *adaptive management* and *monitoring* (DEA., 2014). Likewise, the IDP process has been criticised for its failure to acknowledge limited foresight and to deliver a framework that fosters adaptive planning approaches (Ngamlana and Eglin, 2015). Such a framework would need to facilitate responsive scenario analysis and ongoing monitoring, learning and adaptation (Hummelbrunner and Jones, 2013).

Meaningful public engagement is one of the cornerstones of sustainable development (Everatt *et al.*, 2010), fundamental in achieving bottom-up sustained transformation. However, a large body of literature records the failure of IDP and IEM processes to achieve meaningful stakeholder engagement (Everatt *et al.*, 2010; Aklilu *et al.*, 2014; DEA., 2014; Ngamlana and Eglin, 2015). Failure has been attributed to the non-functionality of ward committees aimed at facilitating engagement (Aklilu *et al.*, 2014; Ngamlana and Eglin, 2015), coupled with a low sense of ownership of development initiatives and capacity of citizens to participate (Aklilu *et al.*, 2014).

Although the IDP only facilitates citizen participation in the planning phase of development, there has been a call for participatory monitoring and evaluation aimed at involving citizens and other stakeholders in the implementation of IDPs (Ngamlana and Eglin, 2015). It is envisaged that such an initiative will facilitate accountability, co-creation and networked solutions (Ngamlana and Eglin, 2015).

3.7 Proposed information systems solutions

Analysis of the existing framework for environmental assessment and integrated decision-making in South Africa (Section 3.6) showed a significant lack of true sustainability-led decision-making and cross-sector collaboration within the respective IDP and IEM processes, which are the primary tools for sustainable socio-economic development in the country. Despite the fact that IDPs and IEM provide the policy frameworks necessary to enforce the management principles identified in Section 3.4, they have largely failed to do so.

The socio-political challenges identified in Section 3.6 – such as chronic lack of capacity within local government, corruption and poor stakeholder engagement processes – are major contributing factors to this non-performance. However, another major contributing factor is the lack of effective supporting information systems (DEA., 2014).

Current and readily accessible data is an essential requirement for effective decision-making and participation by all role-players in the planning process. Coupled to this, is the need for this data to be assimilated in such a way as to enable effective cross-sector integration and collaboration. While a wide variety of data and information is being generated by various government departments in South African cities, these data are not being integrated and utilised in way that optimally supports integrated planning and management (DEA., 2014).

Recently, the Environmental Impact Assessment and Management Strategy (EIAMS) was developed, delineating the pillars of effective IEM (DEA., 2014) and proposing the way forward with regards to improving IEM practices. It was suggested in the EIAMS that the generic guiding principles of sustainability described in NEMA be elucidated through the development of *clear sustainability objectives, indicators and targets* at national and local government levels (DEA., 2014). It was hoped that clear sustainability targets would provide the necessary strategic vision to promote sustainability-led development, and the required framework for cooperative and efficient decision-making, performance monitoring and adaptive management (DEA., 2014).

In addition, the EIAMS (DEA., 2014) also called for *effective environmental information management.* Reliable, current, publicly available and understandable environmental information is an essential requirement for effective decision-making and participation by all role-players in IEM processes. Effective environmental information management is therefore critically important to the success of IEM. While a wide variety of data and information is being generated by various government departments, industry, non-governmental organisations and consultants, there are no information systems fully implemented to collate the data and make it publicly available (DEA., 2014). The EIAMS (DEA., 2014) called for a centrally maintained catalogue of available information, and the development of appropriate data standards.

Table 3-1: Current policy tools and proposed information systems solutions aimed at achieving key management principles across the development cycle.

Stage in the Development Cycle	Key Management Principle	Current Policy Tool	Proposed Information Systems Solutions**
Strategic- & Project-level Planning	Stakeholder Engagement	Strategic-level: IDP, SDF, EMF *Project-level:* EIA supported by EMF	**Spatial data*:** • current state and trends of the environment • desired state of the environment, including sustainability objectives, indicators and targets * Data to be continuously updated with feedback from monitoring and new information acquired during impact assessment studies. **Tools for scenario analysis**
	Adaptive Management		
	Sustainability-led Systems-level Optimisation		
Implementation, Monitoring, Auditing, Enforcement & Feedback	Stakeholder Engagement	-	**Community feedback:** • Community and other environmental monitoring forums • Registering of complaints and whistleblowing
	Adaptive Management	EIA & accompanying Env. Management Plan indicating mitigation requirements Optional implementation of Env. Management Systems (ISO 14001)	**Documents*:** • EIA and EMP reports • Environmental Authorisation and associated conditions of approval • Compliance notices • Norms, standards and legislative requirements • Site plans and layout plans • Baseline monitoring data *Information generated is often in the form of difficult-to-access reports. Standardising the requirements for digital formats will facilitate accessibility to documents. **Routine monitoring data** **Reporting:** • trend identification • non-compliance and incident reporting • Need for adaptive management intervention
	Sustainability-led Systems-level Optimisation	-	-

**Guided by the EIAMS for South Africa. EIAMS is the Environmental Impact Assessment and Management Strategy developed for South Africa in 2014 (DEA., 2014).

Table 3-1 indicates how current policy tools are supporting the three key management principles aimed at effective and efficient integrated city planning and management (identified in Section 3.4) at each stage of the activity life cycle; and indicates how information systems interventions can meet current shortfalls in the implementation of these principles.

Guided by the EIAMS (Table 3-1; DEA, 2014), it is proposed that, at the strategic level, effective decision-making requires knowledge of the status quo, clear development objectives and targets, and tools for scenario analysis. Clear development targets are essential to creating the common vision necessary for collaborative decision-making (DEA., 2014); while, one of the major challenges limiting true cross-sector collaboration and sustainability-led decision-making is a lack of appropriate planning tools at the cross-sector level (Hummelbrunner and Jones, 2013).

Furthermore, it is proposed that the above interventions be implemented within an adaptive management framework, informed by routine monitoring activities. Such a framework could provide decision-makers with knowledge of the status quo, knowledge of the desired state of city systems, tools for scenario analysis at the cross-sector level, and feedback on the effectiveness of past interventions. In response to the research problem identified in Section 1.2, therefore, it is proposed that available data can be effectively integrated and utilised for city planning and management by assimilation into an adaptive integrated decision-making framework such as that described above.

Two possible criticisms of the EIAMS specifications (Table 3-1; DEA, 2014), however, is the limited emphasis on stakeholder engagement tools, and the absence of decision-making tools aimed at fostering operations-level systems optimisation. The latter is the backbone of modern smart city solutions and a likely fixture in future city operations, yet IEM remains limited in its adoption of operations-level environmental management tools.

A major weakness in the IDP and IEM processes is the inability to achieve meaningful citizen engagement. As mentioned in Section 3.6, failure has been attributed to the non-functionality of ward committees aimed at facilitating engagement (Aklilu *et al.*, 2014; Ngamlana and Eglin, 2015), coupled with a low sense of ownership of development initiatives and capacity of citizens to participate (Aklilu *et al.*, 2014).

Ward committees, chaired by the ward councillors, are made up of members representing various ward interests (Bathembu, 2016) and are the major legislative vehicle through which citizens can participate in the development of IDPs (Ngamlana and Eglin, 2015). Ward committees, however, are highly politicised and fraught with power struggles (Ngamlana and Eglin, 2015; Mnguni, 2016) and in their current state, are not effective platforms for engagement (Ngamlana and Eglin, 2015; Mnguni, 2016).

In order to achieve effective public engagement, citizens need to be appropriately capacitated with meaningful information regarding their rights, and sustainability and development issues (Everatt *et al.*, 2010; DEA., 2014). Emerging smart citizen engagement tools (Xenos *et al.*, 2014; Kleinhans *et al.*, 2015; Making Sense *et al.*, 2018) could be very effective in transforming the state of stakeholder engagement in South Africa, particularly in light of substantial efforts by local government to provide free Wi-Fi access at strategic locations across most of the country's major cities (SALGA., 2015).

3.8 Conclusions regarding model requirements

The research problem addressed in this study is that South African cities are not effectively integrating and utilising available data sources for smart city planning and management (Section 1.2). Proposed information systems solutions to the challenges faced in city planning and management in South Africa were discussed in Section 3.7 and summarised in Table 3-1. These recommendations will be used in this section to delineate the functional, construction and environmental requirements of a solution to the research problem stated above.

3.8.1 Functional requirements

Functional requirements of a DSR artefact refer to the intended use of the artefact (Figure 2-3; Johannesson and Perjons, 2012). Section 3.4 concluded that effective decision-making (and therefore effective use of data in decision-making) incorporates all stakeholder and sustainability considerations. In addition, effective decision-making does not occur in isolation, but is carried out within an adaptive management framework.

Furthermore, as indicated in Table 3-1, effective city planning requires clear development targets, knowledge of the status quo, and reliable tools for scenario analysis. Subsequently, an artefact that effectively utilises available data sources for city planning will need to incorporate all stakeholder and sustainability considerations, provide clear development targets, provide knowledge of the status quo, provide reliable tools for scenario analysis, and form part of an adaptive management framework.

3.8.2 Construction requirements

The construction requirements of a DSR artefact refer to the components of the artefact and their relationships and interactions (Johannesson and Perjons, 2012) (Figure 2-3). Based on the functional requirements identified in Section 3.8.1, a prototype model will be developed that incorporates all stakeholder and sustainability considerations, and serves as a reliable tool for systems-level scenario analysis. The following construction requirements of a solution to the research problem are proposed:

- *Requirement 1: KPI framework* – The solution should incorporate a set of Key Performance Indicators (KPIs) and their associated targets. KPIs should be selected in such a way as to quantitatively represent all stakeholder and sustainability considerations. By so doing, the KPI framework will explicate the goals of all involved, and will simultaneously provide knowledge of the status quo;

- *Requirement 2: Predictive model* – The solution should incorporate a predictive model that can be used as a reliable tool for systems-level scenario analysis:

- *Requirement 2.1:* The model should take as input the KPIs identified in Requirement 1, and should predict the relative influence of selected KPIs and combinations thereof on target KPIs. By so doing, the model will predict the collective impact of stakeholder decisions (represented as a set of changing KPIs) on target KPIs. The aim, therefore, is to represent a city as a set of state variables (Requirement 1), and then model the relationship between state variables (Requirement 2). In order to integrate stakeholder and sustainability concerns, it is proposed that city state variables be expressed as a set of KPIs;

- *Requirement 2.2:* The model should be able to incorporate known or unknown complexities and inter-dependencies between variables; and

- *Requirement 2.3:* The predictive model should be reliable. Predictions should be accurate and precise. A measure of prediction uncertainty should be included with the results.

3.8.3 Environmental requirements

The environmental requirements of a DSR artefact relate to the external surroundings and conditions in which the artefact will be used (Johannesson and Perjons, 2012) (Figure 2-3). The main environmental requirement considered in this study is described below:

- *Requirement 3: Data availability and accessibility:* Reliable and available data is fundamental to the success of any model.

    - *Requirement 3.1:* It is intended that the developed model will complement the existing IDP process in South Africa. As such, relevant data needs to be available and accessible for South African cities.

    - *Requirement 3.2:* The model will be implemented at the strategic planning level. Available data therefore needs to be applicable at corresponding spatial and temporal scales.

3.9 Summary

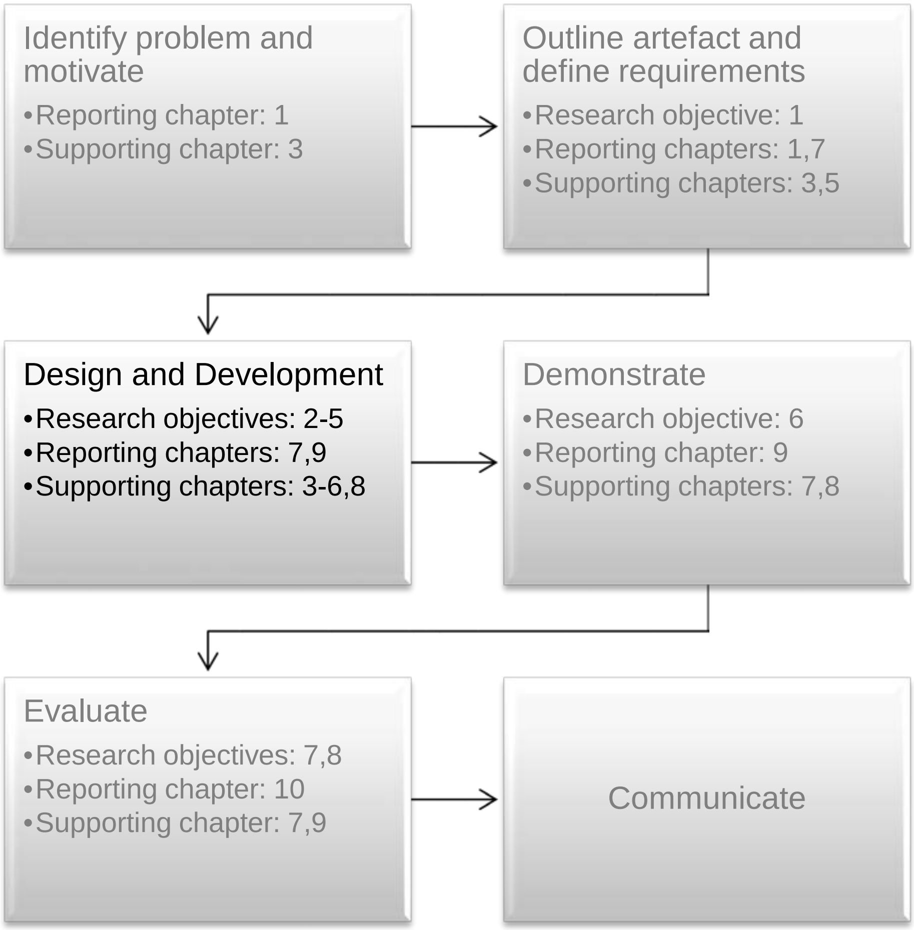
The purpose of this chapter (Figure 3-1) was to provide the supporting literature necessary to formulate the research problem stated in Section 1.2, and to identify the requirements of a potential solution. This was achieved by performing a gap analysis of current city management practices in South Africa (Section 3.6), and by reviewing proposed information system solutions to the identified challenges (Section 3.7).

It was concluded that an artefact that effectively utilises available data sources for city planning will need to incorporate all stakeholder and sustainability considerations, provide clear development targets, provide knowledge of the status quo, provide reliable tools for scenario analysis, and form part of an adaptive management framework (Section 3.8). To this end, it was proposed that a prototype model be developed that incorporates all stakeholder and sustainability considerations, and serves as a reliable tool for scenario analysis (Section 3.8).

The outcome of this chapter was to address RO_1 (Figure 3-1) by defining the functional, construction and environmental requirements of a potential solution to the stated research problem. Section 3.8 listed the proposed solution requirements based on the recommendations highlighted in Section 3.7.

The objective of this study is to develop and evaluate a prototype model for crime management in smart South African cities. To this end, the solution requirements identified in Section 3.8 will be further developed in Chapter 5 within the context of crime management. Model requirements identified in Chapters 3 and 5 will be consolidated in Chapter 7 (Section 7.2). The following chapter of this thesis aims to identify smart city design solutions to the identified model requirements.

Chapter 4. Decision-making in Cities: Smart City Solutions



Research objectives addressed in this chapter:

RO_1: Identify the functional, construction and environmental requirements of an effective model.

RO_2: Identify relevant input and output parameters.

RO_3: Identify and characterise available data sources.

RO_4: Identify the modelling technique to be used to develop the model.

RO_5: Develop the model.

RO_6: Demonstrate the application of the model.

RO_7: Evaluate the efficacy of the model.

RO_8: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

Figure 4-1: Research objectives and design science research activity addressed in this chapter.

4.1 Introduction

The previous chapter highlighted shortfalls in current city management practices in South Africa, and proposed requirements of a potential solution to these challenges. The goal of this chapter (Figure 4-1) is to review current smart city activities, and to propose how these tools can be used to complement existing city management practices.

As a basis for further discussion, key smart city concepts are defined in Section 4.2 of this chapter. Smart cities are defined; and the concepts of the Internet of Things, Big Data, Data Science and data mining algorithms are all explained within the context of smart cities. The trend towards systems thinking and integrated decision-making in smart cities is then discussed in Section 4.3, and the increasingly important role of Integrated City Management Platforms is highlighted.

Emerging data and management standards aimed at promoting smart city best practice are discussed in Section 4.4. These standards form the basis of the smart city design solutions proposed in this study. In order to gain an understating of locally available technology, the state of smart city maturity in South Africa is discussed in Section 4.5. Section 4.6 concludes this chapter by proposing smart city design solutions aimed at meeting the model requirements identified in Section 3.8 of the previous chapter.

The design solutions proposed in Section 4.6 address $RO_{2\text{-}4}$ (Figure 4-1). It will be shown how emerging smart city standards can be used to identify the relevant input and output parameters of the prototype model described in Chapter 1. Readily available open data will be highlighted as a key source of data required to develop the prototype model. Lastly, the machine learning and data mining tools widely adopted in smart cities will be identified as a feasible source of modelling techniques to be explored to develop the model.

4.2 Smart city concepts

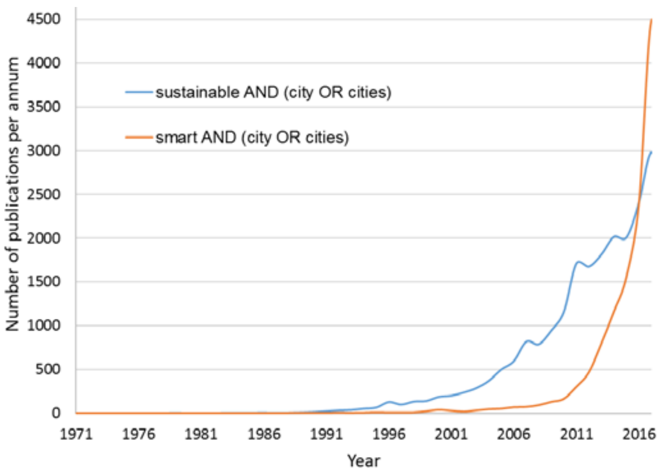
4.2.1 Smart city definition

There is some ambiguity in the definition of smart cities (Neirotti *et al.*, 2014; Albino *et al.*, 2015; Ahvenniemi *et al.*, 2017; Mattoni *et al.*, 2017). SALGA (2015: 7) quoted Joe Bignan from the Economist in saying that "different industries approach the subject from their comfort zones. IT companies define a smart city through a technology lens; developers concentrate on physical infrastructure; utilities insist it is about sustainable energy; and the green lobby champions the environment. Smart Cities are all of the above."

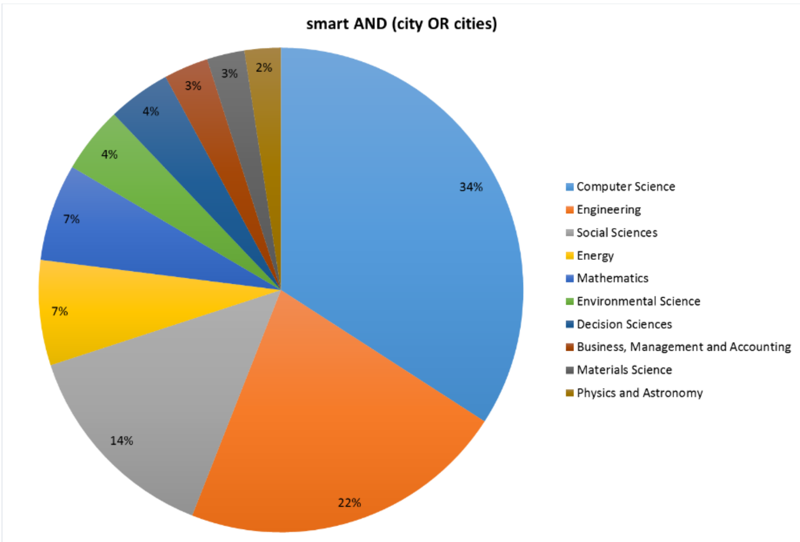
Albino *et al.* (2015) cited 23 definitions of smart cities in order to clarify the definition of a smart city. The chosen definitions came from a range of fields, including architecture, urban planning, engineering, business, and information management. On reviewing the definitions, it is evident that most define smart cities as those whose goals are to achieve sustainability (see Section 3.2), and who make use of technology to achieve these goals.

The smart city definitions varied, however, in their emphasis on technology. The difference in emphasis is clear when considering smart city frameworks developed by different sectors (Neirotti *et al.*, 2014; Albino *et al.*, 2015; Ahvenniemi *et al.*, 2017; Mattoni *et al.*, 2017). IBM (2013), for example, focuses strongly on ICT; while urban strategist, Boyd Cohen (Cohen, 2012) takes a more holistic view. In the latter, ICT is only one aspect of a smart city, together with other focus areas such as reinventing a city's physical, social and business infrastructures (Harrison *et al.*, 2010; Neirotti *et al.*, 2014).

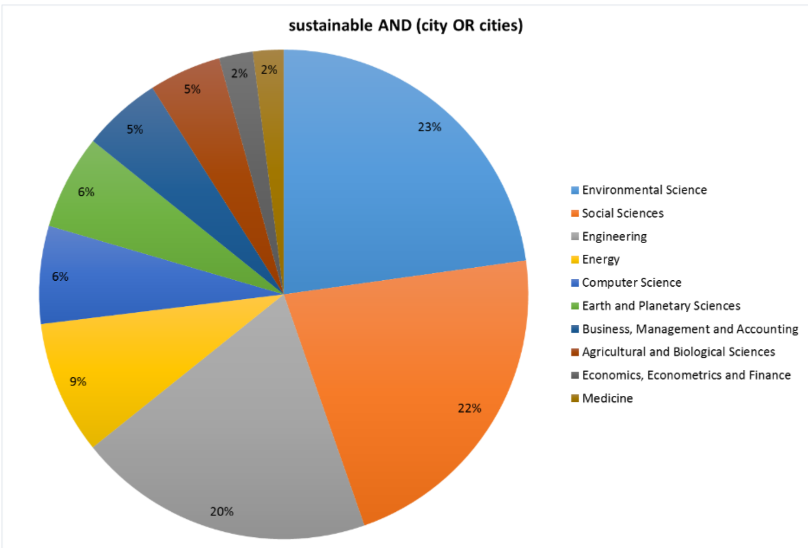
The ambiguity in the definition of smart cities, therefore, is in part due to the different disciplines involved in city planning and management. Ahvenniemi et al. (2017) distinguishes between smart city frameworks and sustainable city frameworks, and promotes the consolidation of these frameworks into a unified focus on smart sustainable cities. Figure 4-2 illustrates this distinction by showing the results of Scopus analytics that compare the publication trends and subject focus areas of



(a)



(b)



(c)

Figure 4-2: Scopus analytics comparing (a) publication trends, and (b, c) subject focus areas between smart and sustainable cities research. Source: Author's own construction.

smart and sustainable cities research. Sustainable cities research preceded smart cities research by close to two decades (Figure 4-2a); and is dominated by the environmental and social sciences domains, followed by engineering (Figure 4-2c). Smart cities, on the other hand, is a more recent phenomenon, and is dominated by the computer science and engineering domains, followed by the social sciences (Figure 4-2b).

The difference in focus between the different domains is further illustrated by the work of Kuyper (2016) that compared the smart city strategies of Barcelona and Amsterdam. Both cities are leaders in smart city solutions (Gasco-Hernandez, 2018), yet their approaches differ greatly. Barcelona has taken a top-down digital city strategy by implementing comprehensive technical infrastructure aimed at improving citizen services. For example, through the renewal of the transit system, the use of smart lighting, and the placement of sensors for water use optimisation. Amsterdam, in contrast, has taken a participatory bottom-up approach, resulting in an extensive amount of smart city pilot projects generally aimed at environmental sustainability.

A number of authors, from various disciplines, have summarised the components of a smart city (Frost & Sullivan., 2013; IBM., 2013; Albino *et al.*, 2015; SALGA., 2015), and various city development frameworks and indictors have been developed to benchmark smart cities (Neirotti et al., 2014; Albino et al., 2015; Ahvenniemi et al., 2017; Mattoni et al., 2017). A well referenced delineation of smart cities is Cohen's Smart Cities Wheel (Cohen, 2012), which classifies smart cities into six dimensions: smart governance, smart living, smart mobility, smart people, smart economy and smart environment. There are 18 total sub-dimensions (or working areas) to his model, with a total of 62 indicators (Cohen, 2014).

Table 4-1 is a summary of Cohen's smart cities indicators, highlighting the key working areas within each of the six dimensions. The principles of sustainability and resilience (Section 3.2) are clearly reflected in the indicators, which focus on reducing resource consumption; promoting clean, multimodal and integrated energy and transport systems; promoting connectivity and inclusivity; and fostering integrated planning and management informed by data.

Table 4-1: Summary of Cohen's Smart Cities Wheel indicators. Source: Adapted from Cohen (2014).

Dimension	Working Area	Indicators
Smart Environment	Smart buildings	LEED certified, smart meters, building automation systems
	Resource management	Energy: Renewable energy, smart grid (self-healing, net metering, real-time info for customers)
		Pollution levels: GHG, air quality
		Solid waste: Reduce and recycle
		Water consumption: Reduce consumption per capita, smart water meters
	Sustainable urban planning	Climate resilience strategy
		Density
		Green space per capita
Smart mobility	Efficient transport	Clean energy transport: Bicycle paths, shared bicycles and shared vehicles, electric vehicle charging stations
	Multi-modal access	Public transport: Integrated fare system, high percentage of public transport use
	Technology infrastructure	Smart cards
		Access to real-time information: Demand-based pricing, real-time traffic management system, multi-modal transit app, real-time information to the public re transit services e.g. bus, train, bike sharing
Smart government	On-line services	On-line procedures and electronic benefit payments
	Infrastructure	Wi-Fi and Broadband coverage
		Sensor coverage: Infrastructure components with installed sensors (traffic, public transit demand, parking, air quality, waste, H_2O, public lighting)
		Integrated health and safety operations: Services integrated in a singular operations center leveraging real-time data (ambulance, emergency/disaster response, fire, police, weather, transit, air quality)
	Open government	Open data and apps
		Privacy policy
Smart economy	Entrepreneurship and innovation	Start-ups, R&D, innovation cities index
	Productivity	GDP per capita
	Local and global connexion	Technology exports
		International congresses and fairs
Smart people	Inclusion	Internet connected house holds
		Smart phone penetration
		Civic engagement
	Education	Secondary and tertiary education
	Creativity	Foreign-born immigrants
		Urban living lab
		Creative industry jobs
Smart living	Culture and wellbeing	Living conditions: Housing, electricity, potable water, sanitation, overcrowding
		Gini coefficient of inequality
		Quality of life ranking
		Investment in culture
	Safety	Smart crime prevention: Live streaming video cameras, predictive crime software technologies, apps
	Health	Single health history
		Life expectancy

4.2.2 The IoT and Big Data

Connectivity is fundamental to smart cities. Smart people are connected through smart phones and infrastructure and the urban environment are connected through the Internet of Things (IoT). The IoT is the name given to the growing trend in which large numbers of networking sensors are embedded into various devices, enabling information-gathering and control functions (Chen et al., 2014). This ubiquitous connectivity allows for real-time monitoring and management of infrastructure and citizens.

By combining real-time monitoring, event management, data analytics and advanced citizen engagement, cognitive technologies are leveraging the IoT and Big Data to radically reduce inefficiencies in all government sectors including smart buildings, healthcare, education, emergency management, public safety, city planning and operations, government administration, water, transportation and energy (Al Nuaimi *et al.*, 2015; IBM., 2018).

Smart mobility, for example, combines sustainable urban design concepts (Ritchie and Thomas, 2009) such as vehicle sharing and multi-modal transportation solutions, with ICT solutions such as smart parking (Rao, 2017) and real-time traffic management (Wang *et al.*, 2018), to improve the resilience and efficiency of smart city transportation systems. Similarly, smart buildings combine green architecture and design with ICT solutions that help occupants monitor and efficiently use energy and water resources.

As datafication resulting from the IoT and smart phone usage intensifies, an unprecedented amount of data is now available and increasingly accessible to city officials that potentially holds great value for city planning and management. Real-time traffic data, data from mobile devices (e.g. social media feeds, geographical locations, pictures, etc.), weather data, pollution levels, data from infrastructure sensors, CCTV footage, medical records, etc., all hold great value (Chen *et al.*, 2014; Steenbruggen *et al.*, 2015; Ferguson, 2017b). Collectively, these data sources have the potential to catapult city planning and management from off-line activities informed by heuristics to real-time decision-making informed by Big Data.

The world is undergoing a paradigm shift in its view and use of data (O'Neil and Schutt, 2014). For the first time, as a result of on-line and off-line datafication, massive amounts of data is available about many aspects of citizen's lives and the computing power to analyse this data has become readily accessible through advances in cloud computing. As a result, there is a growing influence of data in most sectors and in most industries (O'Neil and Schutt, 2014).

Big Data is the name given to the unprecedented amount of data being generated from the datafication of many aspects of our lives and is defined in terms of the challenges that these new data streams pose. Big Data is frequently defined with reference to the "3Vs" of volume, variety and velocity (Chen *et al.*, 2014; Gandomi and Haider, 2015; Kacfah Emani *et al.*, 2015). Despite the challenges involved in analysing Big Data, it holds great value, and the "3Vs" are often expanded upon to define Big Data with reference to the "4Vs": volume, variety, velocity and value (Chen et al., 2014). In addition to the "4Vs", other dimensions of Big Data include veracity and variability (Al Nuaimi et al., 2015; Gandomi and Haider, 2015). The "4Vs" are described below:

- *Volume (great volume):* Big Data is "big" in terms of volume (Chen et al., 2014; Gandomi and Haider, 2015). This is a relative term referring to when the size of the data outstrips state-of-the-art computational solutions, and a host of new tools and methods need to be developed and employed to manage and process the data (O'Neil and Schutt, 2014; Chen et al. 2014). According to Chen et al. (2014) Big Data ranges from several terabytes (TB) to several petabytes (PB). For example, Google processes data of hundreds of PB, while Facebook generates log data of over 10 PB per month (Chen et al., 2014). Taobao generates data of tens of TB for on-line trading per day (Chen et al., 2014).

- *Variety (various modalities):* Variety indicates the various types of data, which include semi-structured and unstructured data such as audio, video, webpage, and text, as well as traditional structured data (Chen et al., 2014). Gandomi and Haider (2015) point out that a high level of variety is not new. Organisations have been storing unstructured data for years. However, the emergence of new data management technologies and analytics, which enable organisations to leverage this data in their business processes, *is* new. For example, facial recognition technologies empower retailers to acquire intelligence about store traffic, the age or gender composition of their customers, and their in-store movement patterns. This information can be leveraged in decisions related to product promotions and placement (Gandomi and Haider, 2015). Variety also refers to the complexity of connecting, matching, cleaning and transforming data received from different sources (Gandomi and Haider, 2015).

- *Velocity (rapid generation):* Velocity means that data collection and analysis must be rapidly and timeously conducted, so as to maximise the value of Big Data (Chen et al., 2014; Gandomi and Haider, 2015). The proliferation of digital devices such as smartphones and sensors has led to an unprecedented rate of data creation and is driving a growing need for real-time analytics and evidence-based planning (Gandomi and Haider, 2015). For example, the data emanating from mobile devices provides information about customers, such as their geospatial location, demographics, and past buying patterns; this, sometimes 'perishable' data, can be used to generate real-time, personalised offers for customers (Gandomi and Haider, 2015).

- *Value (huge value but very low density):* Big Data received in the original form usually has a low value relative to its volume (Chen et al., 2014; Gandomi and Haider, 2015). However, a high value can be obtained by analysing large volumes of such data (Gandomi and Haider, 2015).

4.2.3 Data Science

Data science is a "new"[1] profession aimed at extracting value from Big Data. Data Science combines aspects of computer science, statistics and domain knowledge (Conway, 2010; Shan et al., 2015) to tackle the scientific, analytic, and engineering

[1] The term "data scientist" was first coined in 2008, and only gained entry into Wikipedia in 2012 (O'Neil and Schutt, 2014).

challenges associated with extracting value from massive disparate sources of data, sometimes in fields not accustomed to rich data sources.

Despite the hype surrounding Big Data and Data Science, the term “Data Science” is often criticised as being nothing more than a rebranding of statistics or machine learning (Provost and Fawcett, 2013; O’Neil and Schutt, 2014). In order to determine what a data scientist is, O’Neil and Schutt (2014) looked at the origin of the term, as well as the skill sets typically required of data scientists in job descriptions. Much of the current data explosion is coming from the high-tech world, and the term “data scientist” originated in companies such as Facebook, Google and LinkedIn, where people were working in teams on problems that required a hybrid skill set of statistics and computer science.

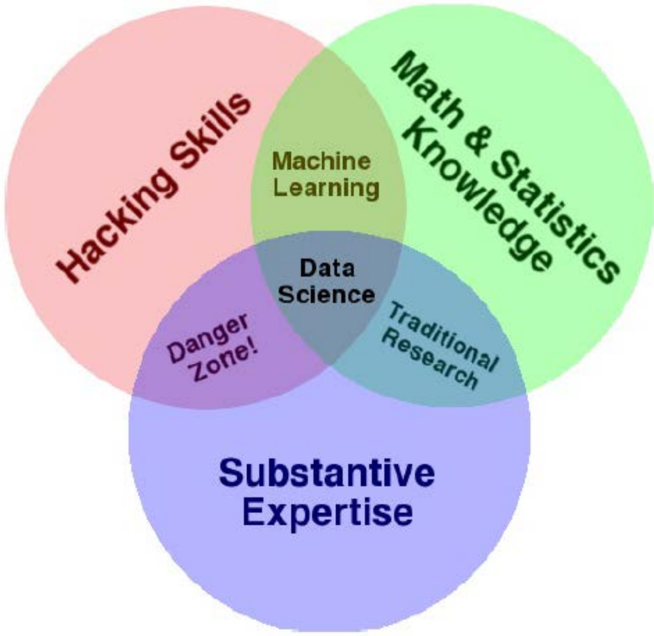


Figure 4-3: Drew Conway's Venn diagram of Data Science. Source: Conway (2010).

Most job descriptions for data scientists require data scientists to be experts in computer science, statistics, communication, and data visualisation, and to have extensive domain expertise. In general, individuals are not simultaneously experts in all these fields, and O’Neil and Schutt (2014) argue that it makes better sense to have Data Science teams of people covering all the required skills set. Never-the-less, their findings are in alignment with Drew Conway’s Venn diagram of Data Science (Figure 4-3; Conway, 2010) which shows Data Science as the intersection of hacking skills, mathematics and statistics knowledge, and domain knowledge. It distinguishes Data Science from traditional research by limiting traditional research to the intersection of domain knowledge and knowledge in mathematics and statistics.

It follows, that Data Science is arguably the application of statistics and machine learning to unprecedented amounts of data across a wide range of fields, even fields that are traditionally not data-intensive. To process massive data sets requires new algorithms and computer architectures, and hence hacking skills that are typically not held by traditional data miners. With access to such a high quantity of data from such a broad range of sources, scientists in fields not accustomed to rich data sources need to become acquainted with statistics and machine learning algorithms as well as the hacking skills necessary to handle the data.

Data Science, therefore, is possibly not a new "science", but rather a set of skills required by the latest generation of scientists from all disciplines in the age of Big Data. O'Neil and Schutt (2014:15) offer a definition for "Data Science" in academia: "an academic data scientist is a scientist, trained in anything from social science to biology, who works with large amounts of data, and must grapple with computational problems posed by the structure, size, messiness, and the complexity and nature of the data, while simultaneously solving a real-world problem."

As the smart city concept matures, attention is shifting from domain-specific ICT solutions, towards an interconnected and synergistic approach to solving urban challenges (Fernández-Güell et al., 2016) (see Section 4.3). However, little progress has been observed in this regard (Fernández-Güell et al., 2016). While frameworks for the management and use of Big Data in smart cities have been proposed (Batty, 2013; Hashem *et al.*, 2016; Pan *et al.*, 2016; Rathore *et al.*, 2016; Silva *et al.*, 2017; Thakuriah *et al.*, 2017), these have focused on architecture considerations. Limited attention has been given to the use of data for integrated decision-making at the system-of-systems level (Lombardi *et al.*, 2012; Mattoni *et al.*, 2015; Schleicher *et al.*, 2016; Mattoni *et al.*, 2017).

As the call for integration intensifies, city scientists and managers will need to develop their capacity to process and analyse large amounts of data from disparate sources. Conversely, while ICT vendors have the required hacking skills for managing and analysing Big Data, they may not have the domain knowledge held by urban planners and city managers. Data Science, therefore, holds great value in bridging the gap between disciplines, and may lead to effective solutions in the new era of integrated smart cities.

4.2.4 Data mining algorithms

Fundamental to Data Science is machine learning and data mining. Data mining is a relatively new field which developed during the 1990's (Nisbet et al., 2009), and represents a confluence of the established fields of traditional statistical analysis, artificial intelligence, machine learning and database technology (Nisbet et al., 2009). Nisbet et al. (2009:17) defines data mining as "the use of machine learning algorithms to find faint patterns of relationship between data elements in large, noisy, and messy data sets, which can lead to actions to increase benefit in some form (diagnosis, profit, detection, etc.)".

There are a number of texts (Hastie et al., 2009; Nisbet et al., 2009; Witten et al., 2011; Han et al., 2012) explaining frequently used data mining techniques. The main styles of learning in data mining applications include classification, association, clustering, numeric prediction, and link mining (Witten et al., 2011; Chen et al., 2014). These are discussed below as a brief introduction to Big Data analytics.

Numerical prediction can draw from a number of techniques such as regression and neural networks to determine a relationship between predictor variables and response variables (Nisbet et al., 2009). Numerical prediction provides a numerical value for a response variable under set conditions; and, depending on the technique employed, may provide a probability associated with that value.

Classification is the operation of separating various entities into two groups (binary classification) or into several classes (multiple classification) (Nisbet et al., 2009). An example of binary classification is a spam filter (O'Neil and Schutt, 2014) which classifies incoming mail as either spam or ham. An example of multiple classification is a diagnostic model that may have several possible outcomes (e.g. influenza, strep throat, chicken pox, etc.) (Nisbet et al., 2009).

Cluster analysis divides a heterogeneous group of records into several more homogeneous classes, or clusters. These clusters contain records that are similar in their values for particular variables. Unlike classification, clustering can be performed on unlabelled classes, where the output value or class is unknown in the training set. Clustering algorithms, such as k-means, employ distance metrics to group records (Nisbet et al., 2009). Clustering is useful in that it can lead to the discovery of previously unknown groups within data (Han et al., 2012). An example of clustering analysis is in business intelligence, where clustering can be used to enhanced customer relationship management by organising a large number of customers into groups, where customers within a group share strong similar characteristics (Han et al., 2012). Another example is in predictive policing, where cluster analysis can be used to make predictions by stating that a future situation will likely be similar to a previous cluster of situations (e.g., "This neighbourhood is showing attributes similar to those of other neighbourhoods labelled as high-crime") (Perry et al., 2013:35).

Another area of data mining is the identification of frequent patterns such as itemsets, sub-sequences, or sub-structures, from which association rules and correlations can be deduced (Han et al., 2012). For example, a set of items, such as milk and bread that appear frequently together in a transaction data set, is a frequent itemset. A sub-sequence, such as buying first a PC, then a digital camera, and then a memory card, if it occurs frequently in a shopping history database, is a frequent sequential pattern. A popular application of pattern identification is in retail, where the discovery of frequent itemsets and sequences can help retailers develop more effective marketing strategies such as strategic item placement in stores, or recommendation algorithms for on-line shopping (Han et al., 2012).

Social network analytics is concerned with synthesising the structural attributes of a social network and extracting intelligence from the relationships among the participating entities (Gandomi and Haider, 2015). The structure of a social network is modelled through a set of nodes and edges, representing participants and relationships, respectively. Various techniques have recently emerged to extract information from the structure of social networks; including community detection, social influence analysis, and link prediction (Gandomi and Haider, 2015). Link prediction aims to predict future linkages between existing nodes in an underlying network. In security, link prediction helps to uncover potential collaborations in terrorist or criminal networks, for example. In the context of on-line social media, the primary application of link prediction is in the development of recommendation systems, such as Facebook's "People You May Know" (Gandomi and Haider, 2015).

4.3 Integrated City Management Platforms

Despite the explosion in smart city solutions over the last decade, the transformation of cities is not following as expected (IEC., 2015). Optimisation and integration in historical verticals (sectoral silos) is the core of today's smart cities projects and it is being proposed that this silo approach to smart cities development is reaching its limits (IEC., 2015; Fernández-Güell *et al.*, 2016) (see Sections 3.3 and 3.4).

Practitioners are calling for a coordinated approach (IBM., 2010; Chourabi *et al.*, 2012; City Protocol Society., 2015b; IEC., 2015; Mattoni *et al.*, 2015; Fernández-Güell *et al.*, 2016; Schleicher *et al.*, 2016; Kourtit *et al.*, 2017; Mattoni *et al.*, 2017). In recent years, much work is being done on developing appropriate frameworks aimed at eliminating systemic inefficiencies and fostering integrated decision-making, interoperability and collaboration (IBM., 2010; IEC., 2015).

The call for integration has led to the emergence of informational or systems platforms aimed at facilitating knowledge acquisition and information transfer between trans-disciplinary systems (Cohen, 2014; City Protocol Society., 2015b). These systems platforms, also known as Integrated City Management Platforms (ICMPs) (IEC., 2015), support integrated decision-making through the automated monitoring of city performance indicators for situational analysis; and tools and applications for system-level data analysis and representation, decision support and management actions (City Protocol Society., 2015b; City Protocol Society., 2016).

Example integrated city management platforms are the Intelligent Operations Centre solutions of IBM (IBM., 2013; Zhuhadar *et al.*, 2017) and Huawei (fin24tech, 2017; Huawei, 2018). IBM's Intelligent Operations Center for Smarter Cities (IBM., 2013), depicted in Figure 1-4, integrates information from a wide range of sources (e.g. citizen apps, weather, traffic, crime reports, social media analytics, infrastructure sensors and news feeds) to provide near-real time situation awareness (to optimise infrastructure, events and resources) and near-real time key performance indicators (to monitor the effectiveness and implementation of programmes).

Figure 4-4: Schematic of the IBM Intelligent Operations Center architecture. Source: IBM. (2013).

The system presents a unified view of operations across many agencies, thereby enabling city officials to improve efficiency and optimise services in departments such as emergency response, transportation, energy, water, and public safety in an integrated and synergistic way. This is in-line with city development frameworks that call for integrated planning and management informed by real-time data (The Rockefeller Foundation | Arup., 2015; United Nations., 2015).

A schematic of IBM's Intelligent Operations Center's system architecture is shown in Figure 4-4. Data from the IoT, smart phones, and other sources are integrated into the system and assimilated by the different sector modules. Situational awareness is achieved through geospatial representation of variables and real-time monitoring of data against event rules for alerts and KPIs which are reflected on a dashboard. The system also boasts tools for modelling and simulation.

4.4 Standards

As the concepts and practices of smart cities emerge and mature, standards developing organisations (SDOs) are developing standards to promote best practice. Initially, standards have focused on sectoral best practice such as building, energy, wastewater, smart grids and intelligent transportation systems standards (American National Standards Institute, 2018). Emerging standardisation activities, however, are focusing on the ‘bigger picture’ (IEC., 2015; American National Standards Institute, 2018), and a number of SDOs are developing specifications aimed at fostering a system-of-systems approach to management (The British Standards Institution., 2014; City Protocol Society., 2015c; IEC., 2015; ISO., 2016; ITU., 2016a).

The International Electrotechnical Commission (IEC) (IEC., 2015) identified three key standard requirements to orchestrate smart city infrastructure: namely, data standards focusing on data format and security (The British Standards Institution., 2017); technical standards (City Protocol Society., 2016; ITU., 2016b) focusing on the integration of interoperable infrastructures and services; and management standards.

Management standards are aimed at fostering integrated decision-making and collaboration between stakeholders by providing mutual communication tools necessary for creating a common vision, benchmarking, knowledge transfer, quality assurances, project assessments and collaboration between different operators and service providers (IEC., 2015). Thus far, management standards have focused on developing conceptual models of the city as a system-of-systems (The British Standards Institution., 2014; City Protocol Society., 2015b; ISO/IEC., 2017) and performance metrics (ISO., 2014; City Protocol Society., 2015a; McCarney, 2015; U4SSC., 2017a; ITU., 2018).

Standards relating to smart city conceptual models, performance metrics and open data will be discussed in the following sections.

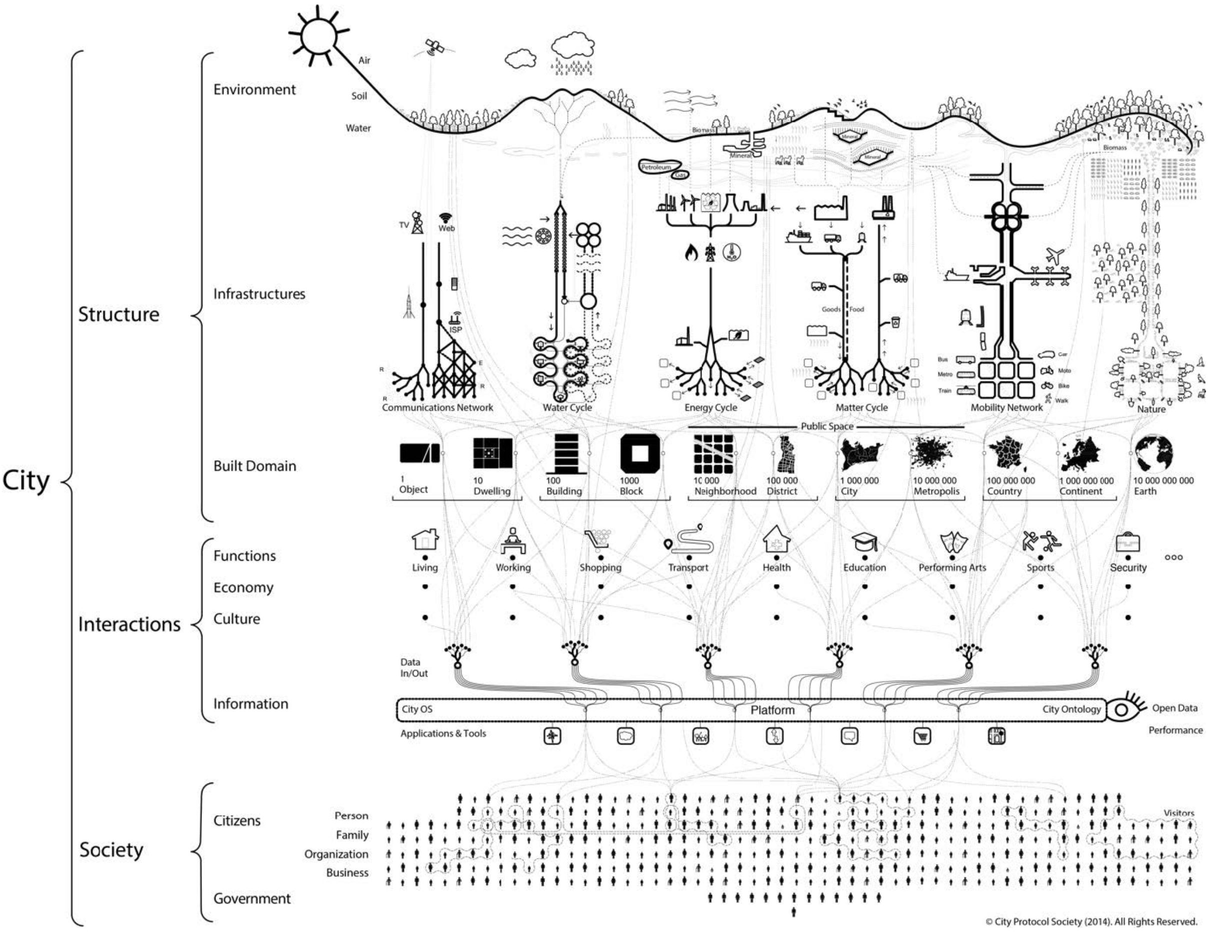


Figure 4-5: Bird's-eye view of the City Protocol Society's *City Anatomy*. Source: City Protocol Society. (2015b).

4.4.1 Conceptual Models

The purpose of conceptual models is to create a common visual understanding of the core components of a city, and their interactions (The British Standards Institution., 2014; City Protocol Society., 2015b; ISO/IEC., 2017). Figure 4-5 is an illustration of a conceptual model developed by the City Protocol Society (City Protocol Society., 2015b) termed the *City Anatomy*. The *City Anatomy* offers a common language describing the city ecosystem as a set of physical structures coupled with the living entities that make up a city's society, and the flow of interactions between them (City Protocol Society., 2015b).

The *City Anatomy* is just one example of a conceptual model aimed at creating a common visual understanding of the core components of a city and their interactions. Other examples include those developed by the British Standards Institution (2014) and ISO/IEC. (2017).

4.4.2 Performance Metrics

Now, more than ever, sustainable urban planning and management is dependent upon evidence-based decision-making (McCarney, 2015; Open Data Charter., 2015). Big Data generated through sensing platforms contains massive amounts of city information (City Protocol Society., 2015a). However, the effective use of this data in decision-making has been hampered by a lack of integration and a clear vision (City Protocol Society., 2015a; IEC., 2015).

Recently, attention has been given to developing KPIs aimed at steering city planning and management activities towards meeting sustainable development goals (United Nations., 2015; U4SSC., 2017b) and measuring and monitoring smart city performance with respect to these goals (ISO., 2014; City Protocol Society., 2015a; McCarney, 2015; U4SSC., 2017a; ITU., 2018). In addition to setting clear development targets and creating a framework for prioritising city challenges, globally comparable KPIs are essential for comparative learning across cities and evaluating the impact of interventions (City Protocol Society., 2015a; McCarney, 2015).

Typically, standards focus on developing a comprehensive set of KPIs to measure a city's social, economic, environmental and governance performance and resilience (ISO., 2014; City Protocol Society., 2015a; McCarney, 2015; U4SSC., 2017a; ITU., 2018) (see Sections 3.2 and 4.2.1). An example of a set of KPIs is the City Protocol Society's *City Anatomy Indicators* (City Protocol Society., 2015a) shown in Figure 4-6. The *City Anatomy Indicators* are an expansion of the indictors proposed by the International Organisation for Standardisation (ISO., 2014) and are aimed at assessing the various sub-systems in the *City Anatomy* framework (City Protocol Society., 2015b) (see Section 4.4.1). The *City Anatomy Indicators* are often referred to as the City Protocol Agreement (CPA) indicators in this report.

KPIs allow the linking of short-term and long-term goals through metrics (City Protocol Society., 2015a; IEC., 2015). The *City Anatomy Indicators* are calculated using real-time data generated from city management platforms (City Protocol Society., 2016) (see Section 4.3), and allow the evaluation of city metrics either from a short-term operations perspective (e.g., in emergency management situations) or from a long-term strategic perspective. Figure 4-6 depicts the top-level evaluation framework of the *City Anatomy Indicators* as a dashboard view where city functioning and status quo is visualised using green, yellow and red indicators for the various systems and subsystems that form part of a city (City Protocol Society., 2015a).

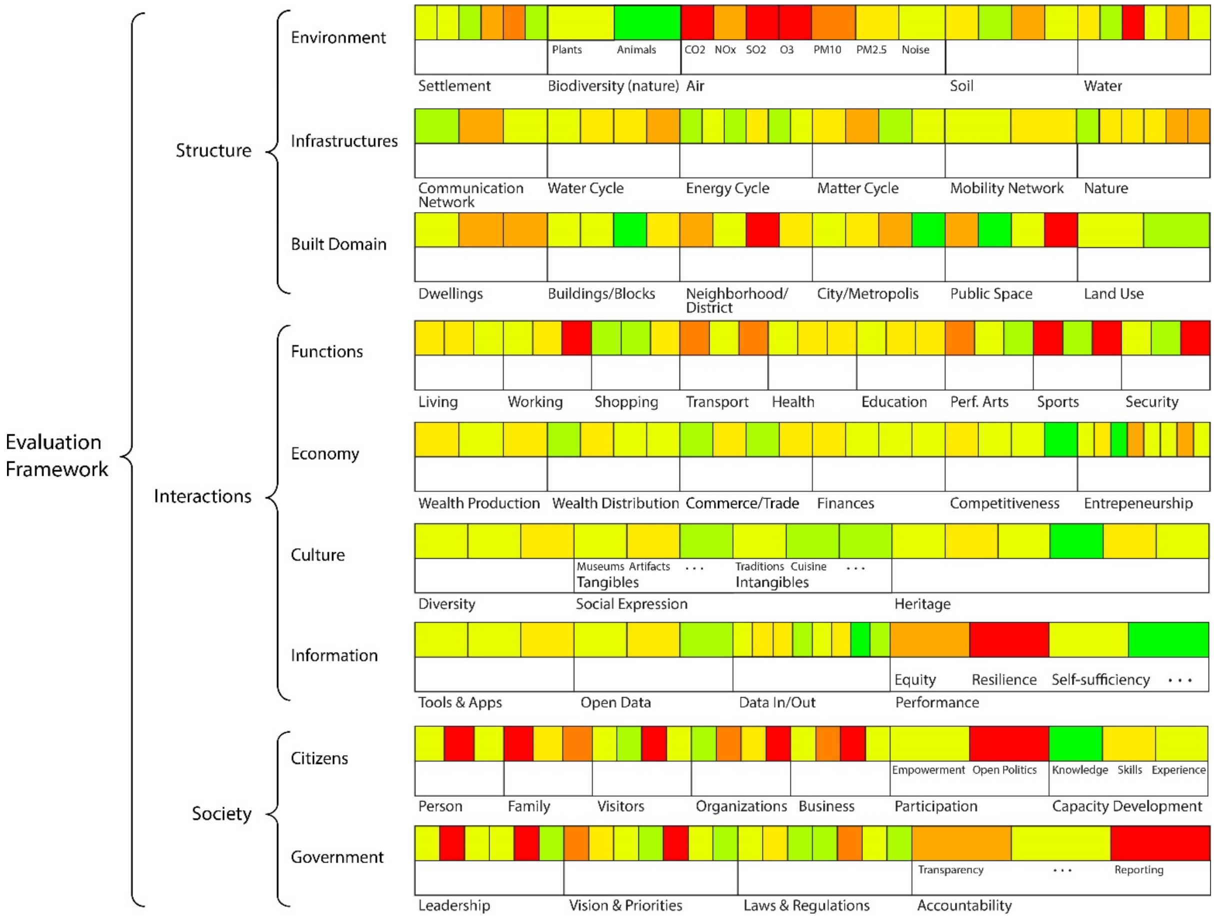


Figure 4-6: Dashboard view of the City Protocol Society's *City Anatomy Indicators*. Source: City Protocol Society. (2015a).

4.4.3 Open Data

According to the International Open Data Charter (2015: 1), "Open data is digital data that is made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime, anywhere." As seen in Section 4.2.2, the world is witnessing a significant global transformation, fuelled by data and information. This transformation has enormous potential to foster more effective governance, and to support the development of sustainable development solutions (Open Data Charter., 2015). Open data is at the centre of this revolution.

Open data supports evidence-based policy making, and enables new insights through unprecedented cross-sector collaboration. Furthermore, open data supports transparency in government decision-making and the spending of public money, thereby promoting accountability and good governance. Open data also promotes public debate and monitoring of the impact and relevance of government decisions and programmes. Open data contributes to the generation of inclusive economic growth through the creation and strengthening of markets, enterprises, and jobs (Manyika *et al.*, 2013; Open Data Charter., 2015; Berends *et al.*, 2017).

The international Open Data Charter. (2015) delineates six principles of effective open data strategies, and provides guidelines on the achievement of these principles in practice. Signatories of the Open Data Charter. (2015) commit to developing action plans aimed at implementing these principles. The principles mandate that data should be open by default, timely and comprehensive, accessible and usable, comparable and interoperable, be used for improved governance and citizen engagement, and be used for inclusive development and innovation (Open Data Charter., 2015).

The Open Data Barometer aims to uncover the true prevalence and impact of open data initiatives around the world (World Wide Web Foundation., 2017). It provides comparative data on governments and regions (Figure 4-7), and ranks governments on their readiness for, implementation of, and impact of open data programmes (Figure 4-8). In 2017, South Africa ranked 46th globally, with an overall score of 34% (Figure 4-7 and Figure 4-8).

The Open Data Barometer's fourth edition report indicates that while some governments are showing progress (Carrara *et al.*, 2016), most governments are not meeting the basic Open Data Charter principles (World Wide Web Foundation., 2017). The main reasons cited are a lack of effective policies, and the insufficient breadth and quality of the datasets released (World Wide Web Foundation., 2017). This is despite the fact that a number of open data standards have been developed to address challenges relating to the quality of released datasets (Bird, 2015; Center for Government Excellence, 2017; Open Data Institute, 2019). Similar observations were made by Open Knowledge International in their latest State Of Open Government Data in 2017 report (Lämmerhirt et al., 2017).

South Africa has only just started gaining momentum in global open data trends. South Africa endorses the Open Government Declaration as a founding member of the Open Government Partnership, which was formally launched in September 2011 (Open Government Partnership, 2018). As a signatory to the Open Government Declaration, South Africa has committed to the development and implementation of action plans aimed at making their government more inclusive, responsive and accountable (Open Government Partnership, 2018). Since South Africa endorsed the Open Government Declaration in 2011, however, implementation with respect to open data has been slow (Humby, 2018).

The open data movement in South Africa is still in its infancy (Humby, 2018). In 2015, the Deputy Minister of Public Service and Administration, Minister Ayanda Dlodlo, commissioned the development of a pilot national open data portal (as part of an Open Government Partnership commitment) that consolidated 409 datasets from national and provincial government. The portal is accessible at www.data.gov.za. However, no new datasets have been added to the site since 2015 (Humby, 2018) due to lack of public funding, reticence on the part of some government departments to make datasets available, and the lack of demand for open data.

Locally, many metropolitan municipalities, including Cape Town, Ekurhuleni, Johannesburg, and eThekwini, are in the initial stages of establishing open data portals (Humby, 2018). Other useful open data websites at the time of writing included www.scoda.co.za, www.municipalbarometer.co.za, sacities.net, stepsa.org, southafrica.opendataforafrica.org, www.statssa.gov.za and data.code4sa.org. At the time of this study, however, these websites did not contain sufficient data to be of benefit to this study.

Regional Rank	East Asia & Pacific			Europe & Central Asia			Latin America & Caribbean			Middle East & North Africa			North America			Sub-Saharan Africa		
		Global Rank	Score (/100)		Global Rank	Score (/100)		Global Rank	Score (/100)		Global Rank	Score (/100)		Global Rank	Score (/100)		Global Rank	Score (/100)
1	Korea	5th	81	UK	1st	100	Mexico	11th	73	Israel	28th	46	Canada	2nd	90	Kenya	35th	40
2	Australia	5th	81	France	3rd	85	Uruguay	17th	61	Tunisia	50th	32	USA	4th	82	South Africa	46th	34
3	New Zealand	7th	79	Netherlands	8th	75	Brazil	18th	59	UAE	60th	26				Mauritius	59th	26
4	Japan	8th	75	Norway	3rd	74	Colombia	24th	52	Kazakhstan	59th	26				Ghana	59th	26
5	Philippines	22nd	55	Spain	11th	73	Chile	26th	47	Qatar	74th	19				Tanzania	67th	22

Figure 4-7: Open Data Barometer's fourth edition regional champions with their respective overall rankings and scores. Source: World Wide Web Foundation. (2017).

ODB 4th Edition Ranking

Rank	Score	Country	Readiness	Implementation	Impact
1	100	United Kingdom	99	100	94
2	90	Canada	96	87	82
3	85	France	100	71	88
4	82	United States of America	96	71	80
5	81	Korea	95	59	100
5	81	Australia	85	78	78
7	79	New Zealand	92	58	99
8	75	Japan	84	60	89
44	36	Ukraine	55	35	19
46	34	South Africa	51	28	29
46	34	Poland	61	24	23

Figure 4-8: Open Data Barometer's fourth edition rankings and scores. Selection chosen to indicate top performers as well as South Africa. Source: World Wide Web Foundation. (2017).

4.5 The state of smart city maturity in South Africa

To date, South Africa has been in the early stages of smart city development (Fernández-Güell *et al.*, 2016), with e-governance, connectivity and ad-hoc sectoral based solutions dominating the smart city landscape (Misuraca, 2007; SALGA., 2015). In recent years, there have been substantial efforts to provide free Wi-Fi access at strategic locations across most of the country's major cities (SALGA., 2015). Major progress has also been made in implementing e-governance at the national and local levels (Misuraca, 2007; SALGA., 2015).

In terms of transport, much attention has been focused on Intelligent Traffic Management Systems, an example of which is SANRAL's i-TRAFFIC website (SANRAL., 2018) which provides real-time information on incident alerts, traffic speeds, and construction updates for regions in Gauteng, the Western Cape and KwaZulu-Natal. There have also been pockets of initiative in other sectors; examples of which are predictive policing efforts in Cape Town (Head, 2017; Venktess, 2017), and the crime fighting application, Namola, in the City of Tshwane (Happimo, 2016) (see Section 5.2). South Africa has also started gaining momentum in global open data trends (see Section 4.4.3).

Although a number of departments are implementing smart city solutions, these initiatives are not integrated (SALGA., 2015). Utilities concentrate solely on their own performance, and there is little cross-functionality in the service provision of energy, waste, water, and transport, for example. There is also a general lack of adequate data at the local and regional levels to support effective programme planning (SALGA., 2015); and data collection is often ad hoc and fragmented (SALGA., 2013). There is however indications that this will change in the near future, as cities start adopting smart city and open data solutions. The City of Johannesburg, for example, aims to become the leading smart city in South Africa with the implementation of an Intelligent Operation Centre aimed at providing an integrated view of the city's strategic and operational issues. The latter aims to focus on public safety in the initial phase of implementation (City of Johannesburg., 2011; City of Johannesburg., 2015; SALGA., 2015).

4.6 Conclusions regarding model design

The functional, construction and environmental requirements of a solution to the research problem stated in Section 1.2 were delineated in Section 3.8. Based on the state-of-the-art in smart city solutions reviewed in this Chapter, the following design elements are proposed to meet the solution requirements:

- *Design element 1: KPI framework – Requirement 1* requires that the solution incorporate a set of Key Performance Indicators (KPIs) that quantitatively represent all stakeholder and sustainability considerations. The emerging smart city conceptual models and KPI frameworks discussed in Section 4.4 delineate key smart city components and their desired states, and therefore provide a potential design solution to *Requirement 1*. Here, it is proposed that a subset of the City Protocol Society's *City Anatomy Indicators* (Section 4.4.2) be used as input to the prototype model developed in this study.

- *Design element 2: Predictive model – Requirement 2* requires that the solution incorporate a predictive model that can be used as a reliable tool for systems-level scenario analysis. Specifically, the model should take as input the KPIs identified in *Requirement 1*, and should predict the relative influence of selected KPIs and combinations thereof on any given KPI. The model should also be able to incorporate known or unknown complexities and inter-dependencies between variables. In addition, predictions should be accurate and precise, and a measure of prediction uncertainty should be provided. Here, it is proposed that a modelling approach employing neural networks (Section 4.2.4) will be effective in meeting *Requirement 2*. Bayesian Neural Networks, which incorporate measures of prediction uncertainty, will be described in Chapter 6.

- *Design element 3: Data – Requirement 3* states that reliable and available data is fundamental to the success of any model. Due to the limited accessibility of South African data at the time of this study, readily accessible open data (Section 4.4.3) will be used to develop and demonstrate the prototype model.

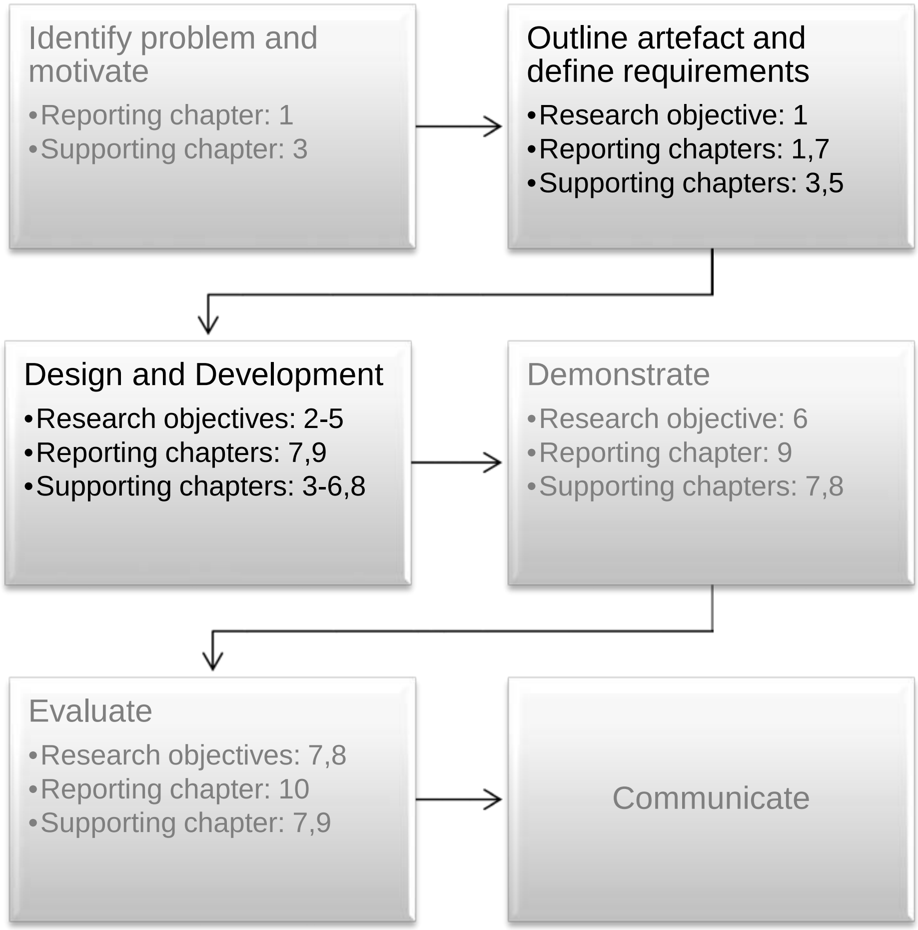
4.7 Summary

The purpose of this chapter was to review current smart city activities, and to propose how associated tools can be used to complement existing city management practices in South Africa. Section 4.6 concluded this chapter by proposing smart city design solutions aimed at meeting the model requirements identified in Section 3.8. The design solutions proposed in Section 4.6 address $RO_{2\text{-}4}$ (Figure 4-1) by indicating smart city KPI standards, open data and machine learning as potential solutions to parameter identification, data requirements and choice of modelling approach, respectively.

Despite the deluge of data generated in smart cities and the accompanying growth in computing power, little research has been done on exploring the use of this data to develop objective quantitative tools for project prioritisation and scenario analysis at the system-of-systems level (Lombardi *et al.*, 2012; Mattoni *et al.*, 2015; Schleicher *et al.*, 2016; Mattoni *et al.*, 2017) (see Section 4.2.3). In this study, it is proposed that a predictive model for whole-system scenario analysis can be developed by building upon emerging smart city management solutions. It is hypothesised that this can be achieved by using smart city KPI frameworks to represent data from traditionally isolated management silos as a set of sectoral KPIs (Figure 1-5a). It is envisaged that the inter-dependencies between sectoral KPIs can be encapsulated in an artificial neural network (Figure 1-5b), which can be used for prediction and simulation at the system-of-systems level (Figure 1-5c).

The aim of this study is to test the feasibility of this supposition. In order to limit the scope of the investigation, the study will focus on only one aspect of smart cities, namely crime management (Figure 1-6). The purpose of the following chapter is to further develop the proposed model requirements (Section 3.8) and smart city design solutions (Section 4.6) within the context of crime management.

Chapter 5. Crime Management in Smart Cities



Research objectives addressed in this chapter:

RO$_1$: Identify the functional, construction and environmental requirements of an effective model.

RO$_2$: Identify relevant input and output parameters.

RO$_3$: Identify and characterise available data sources.

RO$_4$: Identify the modelling technique to be used to develop the model.

RO$_5$: Develop the model.

RO$_6$: Demonstrate the application of the model.

RO$_7$: Evaluate the efficacy of the model.

RO$_8$: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

Figure 5-1: Research objectives and design science research activities addressed in this chapter.

5.1 Introduction

The research problem identified in Section 1.2 is that South African cities are not effectively integrating and utilising available data sources for smart city planning and management. The main supposition (Section 1.3) of this study is that a predictive model, that effectively integrates and utilises data from traditionally isolated management silos, can be developed to support an integrated approach to decision-making. In order to limit the scope of the investigation, the study focuses on only one aspect of smart cities, namely crime management (Section 1.4). The main objective of this study, therefore, is to test the feasibility of the thesis statement by developing and evaluating a predictive model for crime management in smart cities (Section 1.5).

Chapter 3 of this thesis highlighted shortfalls in current city management practices in South Africa, and identified the requirements of a potential solution to these challenges. Chapter 4 reviewed current smart city activities, and proposed smart city design solutions aimed at meeting the requirements of the prototype model identified in Chapter 3. Both Chapters 3 and 4 have focused on city management practice in general. The purpose of this chapter is to further develop the proposed model requirements and smart city design solutions within the context of crime management.

This chapter will further develop $RO_{1,2,4}$ (Figure 5-1) by reviewing the state-of-the-art in crime management in smart cities. The rise of Big Data (see Section 4.2.2) and its applications in crime management will be discussed in Section 5.2, which provides an overview of emerging real-time intelligence centres. Specific attention will then be given to crime forecasting in Section 5.3. The concept of crime forecasting, or "predictive policing" as it is colloquially known, will be defined. The recent emergence and prevalence of predictive policing will be discussed; together with its applications, associated modelling techniques, and commonly employed predictors of crime. The challenges of predictive policing in practice will also be explored.

The requirements and design of the prototype model will then be refined in the context of crime management in Sections 5.4 and 5.5, respectively. This chapter will therefore further develop $RO_{1,2}$ (Figure 5-1) by identifying model requirements specific to crime management, and by identifying model parameters which are relevant predictors of crime. This chapter will further develop RO_4 by identifying modelling techniques commonly employed in crime forecasting.

5.2 Real-time intelligence centres

Ferguson (2017b) paints a picture of the rise of Big Data in policing. The Los Angeles Police Department's (LAPD) Real-Time Analysis Critical Response (RACR) Division (Figure 5-2), for example, is a high-tech command centre that analyses citywide crime on a real-time basis to identify patterns and make deployment recommendations.

Figure 5-2: The Los Angeles Police Department's Real-Time Analysis and Critical Response Division. Source: Pike and Schulz (2014).

The command centre boasts a digital map of alerts to 911 calls, breaking news stories on television screens, surveillance of cameras monitoring the streets, on-line law enforcement intelligence, real-time crime data analytics, and Computer-Aided Dispatch (CAD) (Ferguson, 2017b). The LAPD RACR is just one example of emerging real-time intelligence centres (RTICs) in the United States (US) that are leveraging predictive analytics and data-driven surveillance to revolutionise policing (Ferguson, 2017b; Brooks, 2018).

RTICs integrate major operational and investigative systems and databases including CAD, automated Records Management Systems (RMS), as well as new and emerging data sources such as IoT and sensor data, social media feeds, and biometrics (Brooks, 2018). RTICs then leverage Big Data analytics (see Sections 4.2.4 and 5.3) to automatically search through these traditionally fragmented and unstructured data sources with unprecedented efficiently to make crime predictions and find otherwise-hidden clues (Ferguson, 2017b).

RTICs are at the forefront of data visualisation and sharing (Ferguson, 2017b; Brooks, 2018). Centres typically use geographic information systems (GIS) to display real-time information onto video walls within intelligence centres; and can instantaneously share background information such as maps, photos and suspect histories onto responding officers' mobile phones or patrol car mobile display terminals (Brooks, 2018). Predictive analytics are also used to produce daily crime forecasts which are digitally sent to patrol officers to guide them during their shift (Brooks, 2018).

The age of Big Data policing is also emerging in South Africa. Known examples include predictive policing efforts in Cape Town (Head, 2017; Venktess, 2017), and the launch of Namola in the City of Tshwane (Happimo, 2016). Namola, which is an evolution of the StellieSafe application piloted in Stellenbosch (Alfreds, 2016), is a crime fighting application used by the Tshwane Metro Police Department (Happimo, 2016), and has been referred to as an “Uber for police” (Mabuza, 2016).

The application, which was officially launched in the Tshwane Metropolitan Municipality in 2016 (Mabuza, 2016), aims to reduce the average police response time (Chutel, 2016). When activated, Namola locates and dispatches the nearest available officer to the scene using the caller’s GPS coordinates, essentially circumventing the traditional call centre (Figure 5-4, Chutel, 2016).

The caller is shown a photo and name of the responding officer (Memeburn., 2015), and police vehicles are fitted with a dashboard-mounted smartphone, allowing officers to receive alerts and messages directly from citizens and the control room. Although the call centre is circumvented, the system is still monitored by a control room which keeps sight of real-time tracking maps showing the location of all response vehicles and active citizen alerts (Figure 5-3).

Since the implementation of the system, an 11 times improvement in response time has been observed (Mabuza, 2016). By circumventing the traditional call centre, Namola removes the need for callers to first communicate their needs and whereabouts to an operator, and instead automatically picks up on their GPS location. This has added benefits, as ambiguities resulting from language barriers are reduced (Alfreds, 2015). The app can also improve accountability among police officers by monitoring the whereabouts of patrol cars (Chutel, 2016).

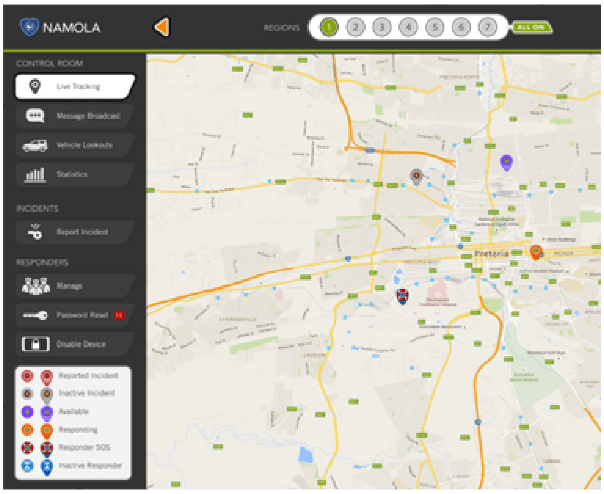


Figure 5-3: The Namola control room. Source: Happimo (2016).

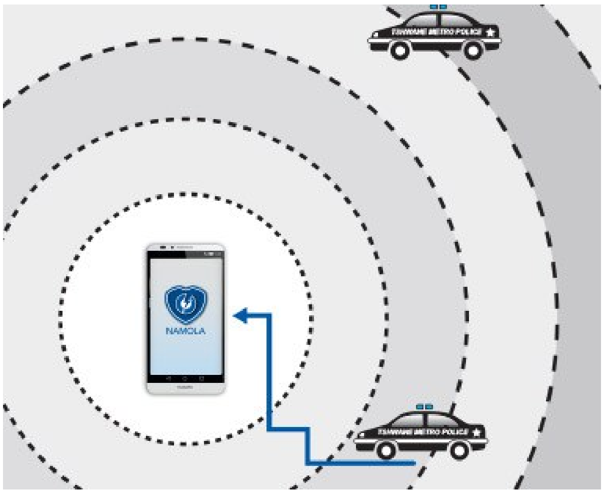


Figure 5-4: The Namola app connects callers with their nearest police vehicle. Source: Happimo (2016).

5.3 Predictive analytics

An important aspect of real-time intelligent centres is predictive analytics. The main objective of this study is to develop a predictive model for crime management in smart cities. In order to aid the design of the prototype model, this section explores the state-of-the-art in predictive analytics as applied to crime forecasting.

5.3.1 Definition

Predictive analytics as applied to crime management is commonly referred to as “predictive policing” (Perry et al., 2013; Robinson and Koepke, 2016; Babuta, 2017).

Predictive policing refers to the use of Big Data and advanced analytical techniques to make statistical predictions about potential criminal activity, in order to prevent crime by identifying likely targets for police intervention, or to solve past crimes (Ferguson, 2012; Bachner, 2013; Perry et al., 2013; Brayne et al., 2015).

Although the use of statistical and geospatial analyses to forecast crime levels has been around for decades, the past decade has seen a surge of interest in data intensive approaches to crime forecasting and analytics (Bachner, 2013; Perry *et al.*, 2013; Ratcliffe, 2016; Ferguson, 2017b; Smith *et al.*, 2017). Predictive policing is based on the notion that the hugely successful data analytics tools used by the private-sector to predict consumer behaviour can be adapted and applied to policing to predict criminal behaviour (Bachner, 2013; Perry *et al.*, 2013; Robinson and Koepke, 2016; Joh, 2017). Predictive policing, therefore, is an application of Data Science to policing (see Section 4.2.3).

5.3.2 Emergence and global prevalence

Predictive policing originated in the USA as a natural progression of the highly successful CompStat programme introduced in the 1990's (Gorr and Harries, 2003; Bachner, 2013; Perry et al., 2013; Ratcliffe, 2016; Babuta, 2017). The New York Police Department was at the forefront of the intelligence-led policing revolution, with the introduction of the CompStat (Compare Statistics) system in 1994 (Babuta, 2017). By mapping crime hotspots, the system allowed the performance of each precinct to be quantitatively measured and tracked (Babuta, 2017). These spatial statistics formed the basis of weekly meetings aimed at reducing crime numbers (Babuta, 2017); and provided an unprecedented level of accountability in law enforcement (Babuta, 2017).

Dramatic reductions in crime rates were observed in precincts that had implemented CompStat (Corman and Mocan, 2002; Roeder *et al.*, 2015), and soon almost every law enforcement agency in the USA had adopted the practice of crime mapping and statistical analysis (Babuta, 2017). After the major successes of crime mapping in the 1990's, the US National Institute of Justice (NIJ) started exploring crime forecasting as a proactive extension of crime mapping in 1998 (Gorr and Harries, 2003).

However, it was not until November 2009, that widespread interest in predictive policing was sparked at the first predictive policing symposium hosted by the NIJ in Los Angeles (Perry *et al.*, 2013). The symposium – which was a gathering of expert researchers, practitioners, government officials, and law enforcement leaders – generated a great deal of interest in predictive policing, and consultants and private companies soon began providing software solutions for predictive policing (Perry *et al.*, 2013).

By 2016, predictive policing software was employed (or had been employed) by at least 40% of the largest police departments in the USA, while a further 20% were actively procuring software (Robinson and Koepke, 2016). Predictive policing is reportedly being used today by law enforcement agencies in several USA states, the United Kingdom (UK), Germany, the Netherlands, Switzerland, and China (Anže Žitnik, 2019). While the adoption of predictive policing techniques is prevalent in the United States, the uptake of predictive policing elsewhere has been slower. Recent reports (Babuta, 2017; HMIC., 2017), for example, state that a limited number of UK police forces have adopted crime prediction tools as part of their digital strategies, and not until very recently has there been talk of predictive policing in South Africa (Head, 2017; Venktess, 2017).

5.3.3 Applications and modelling techniques

Since the inception of predictive policing, the development of software solutions has been spearheaded by academic and private-sector vendors. Leading academic packages include PredPol (PredPol Inc., 2018), Risk Terrain Modelling (Rutgers Center on Public Security, 2019), Chicago's Strategic Subject List (City of Chicago, 2017), and HunchLab (Azavea., 2015; Azavea., 2017). A review by Robinson and Koepke (2016) showed that by far, the most adopted package by USA police departments is PredPol, followed by Azavea's HunchLab. Other commonly employed packages include BAIR Analytics ATACRAIDS offered by LexisNexis, IBM's Crime Prediction and Prevention Package, solutions offered by Information Builders, and Risk Terrain Modelling developed by Rutgers University.

By definition (Section 5.3.1), predictive policing makes use of data mining and machine learning algorithms (Section 4.2.4) to predict crime and criminal behaviour. What follows is a general overview of predictive policing techniques with the aim of guiding the design of the prototype model in this study. The overview is based on the state-of-the-art in the USA as summarised by Bachner (2013) and Perry *et al.* (2013). While authors such as Bachner (2013), Perry *et al.* (2013), Hassani et al. (2016) and Pramanik *et al.* (2017) provide comprehensive summaries of predictive policing techniques, it is important to note that the techniques employed by private-sector proprietary software are ultimately unknown (Robinson and Koepke, 2016; The Leadership Conference on Civil and Human Rights *et al.*, 2016).

Predictive policing techniques can be divided into place-based and person-based solutions (Perry et al., 2013; Robinson and Koepke, 2016). Place-based solutions aim to identify places and times that correspond to an increased risk of crime, while person-based solutions focus on the "who" of future (or past) crimes (Perry et al., 2013; Robinson and Koepke, 2016). The majority of predictive policing software focuses on place-based prediction (Robinson and Koepke, 2016). Place-based prediction will also form the focus of this study. However, both place-based and person-based techniques will be discussed in the sections that follow in order to develop a full understanding of the state-of-the-art in predictive policing.

5.3.3.1 Place-based techniques

The fundamental assumption of predictive policing is that crime is predictable; and there is a strong body of evidence that supports this assumption (Gorr and Harries, 2003; Bachner, 2013; Perry et al., 2013). Criminal decision-making is based on some amount of rationality. Offenders seek to achieve a purpose, and the decision to act is influenced by situational and environmental factors (Bachner, 2013).

Criminality of places is based on theories such as routine activity theory (Cohen, 1979), rational choice theory and crime pattern theory (Brantingham and

Brantingham, 1984; Sherman, 1989). Perry et al. (2013) notes that such theories are best applied to stranger offenses such as robberies, burglaries and thefts. Placed-based techniques are consequently less applicable to relationship violence for example, which involves human connections and leads to decisions that do not fit into traditional “criminal rational choice” frameworks (Perry et al., 2013).

There are two main types of placed-based predictors: those based purely on previous instances of crime, and those that take into account environmental (or structural) considerations of a community (Ferguson, 2012; Caplan et al., 2013; Taylor et al., 2015). Incident-based predictors (such as hot spot analysis and near-repeat methods) assume that the best way to predict future crime is to use past incidents as indicators of future behaviour. On the other hand, the alternative approach (for example Risk Terrain Modelling) considers the environment in which crimes occur and identifies features of the landscape that would be conducive to crime (Caplan et al., 2013). Both types of predictors will be described in the sections that follow, together with combined spatiotemporal techniques.

5.3.3.1.1 Hot spot and near-repeat methods

The most basic predictive models rely on past crime data (Bachner, 2013). This has roots in the well-established theory of repeat victimisation (Ferguson, 2012; Bachner, 2013; Perry *et al.*, 2013), where victims who are victimised once are likely to be the targets of crime again, and that offenders often return to the place of their first crime to use their knowledge about an area (Anže Žitnik, 2019).

A rudimentary example of this is hot spot analysis. Hot spot analysis maps historical crime data with the aim of visualising where crimes are concentrated (the hot spots). Conventional approaches use grid mapping, for example, and human judgment to determine the location of hotspots, while more advanced methods make use of various clustering techniques such as covering ellipses and kernel density estimation (Perry et al., 2013).

The application of hotspot analysis within CompStat in the 1990’s (Section 5.3.2) generated a great deal of interest in the spatial analysis of crime, which lead to the emergence of predictive policing as an extension of hot spot analysis. PredPol, one of the first truly predictive tools for policing, was based on the examination of “near repeats” (Caplan et al., 2013).

Near-repeat theory is based on the phenomena that, for certain crimes (such as residential burglary, automobile theft, and theft from automobiles), once a particular location has been subject to a crime, that location and the nearby area are statistically more likely to be subject to additional, similar crime events in the near future (Ferguson, 2012). The timeframe is usually within one week of the initial crime, and the risk of repeat victimisation decays over time (Ferguson, 2012).

The developers of PredPol (Mohler *et al.*, 2016; PredPol Inc., 2018) were able to predict future crimes, at locations other than previous crimes locations, by likening near-repeat events to seismic aftershocks (D'Orsogna and Perc, 2015). PredPol was therefore developed based on models similar to those used to describe seismic activity; and was based on the assumption that additional crimes often closely follow the initial event in time and space, comparable to a seismic aftershock (Anže Žitnik, 2019). PredPol uses only incident variables, such as the type of offense, the location, the time of day, and the day of the week to generate predictive maps indicating the locations and timing of future crimes of the same type (Bachner, 2013).

5.3.3.1.2 Risk Terrain Modelling

While hotspot mapping and near-repeat analysis allow police to more efficiently allocate resources, proactive crime prevention that addresses the underlying causes of crime cannot proceed without an understanding of the social and environmental factors contributing towards crime in an area (Brantingham and Brantingham, 1984; Weisburd *et al.*, 2009; Caplan *et al.*, 2013).

Risk terrain analysis comprises a family of techniques that attempt to identify geographic features that contribute to crime risk, and then make predictions about crime risk based on how close given locations are to these risk-inducing features (Kennedy et al., 2011; Perry et al., 2013). Geospatial features include the presence of individuals on probation and parole, recent crime incidents, recent calls for disorderly conduct or acts of vandalism, and buildings or infrastructure known to be “at risk” such as bars, liquor stores, and certain types of major roads (Bachner, 2013; Perry et al., 2013).

Risk terrain analysis offers an alternative approach to crime prediction; and assumes that future crimes are less determined by previous events and more a function of the dynamic interaction between social, physical and behavioural factors at a given location (Ferguson, 2012). Risk terrain analysis can therefore predict new hot spots that are similar to other hot spots, without necessarily the occurrence of recent crime in the newly predicted areas (Perry et al., 2013). There are both heuristic and statistical approaches to risk terrain analysis (Perry et al., 2013). These are briefly discussed below.

Risk terrain modelling (RTM) is a heuristic approach to assessing how geospatial factors contribute to crime risk (Kennedy *et al.*, 2011). In RTM, separate map layers representing the spatial influence and intensity of a crime risk factor are created in a geographic information system. All map layers are combined to produce a composite risk terrain map with values that account for all risk factors at every place throughout the landscape (Perry et al., 2013). RTM requires a substantial amount of time and analytical skill to develop risk layers and generate risk-composite maps (Bachner, 2013). Consequently, police departments are reluctant to embrace this method of crime prediction (Bachner, 2013).

The statistical approach to risk terrain analysis involves two major phases (Perry et al., 2013). In the first phase, the algorithm tracks the distances between crimes and the nearest geospatial feature of each type. In the second phase, the algorithm assesses how “similar” each point on a grid is to locations that have seen crimes with respect to distances to the geospatial features. Points at which distances to geospatial features resemble those of crime locations are judged to be at higher risk (Perry et al., 2013).

Unlike near-repeat methods, risk terrain analysis has successfully been applied to violent crimes, and has therefore broadened the application landscape of predictive policing techniques (Ferguson, 2012).

5.3.3.1.3 Spatiotemporal methods

Near-repeat and risk-terrain modelling do not take advantage of temporal patterns in crime, and do not illustrate how the incidence of crime changes over time (Bachner, 2013). Most crimes are strongly influenced by cyclical patterns, such as the day of the week, the time of day, or the season (Bachner, 2013; Perry *et al.*, 2013). For example, during summer, when children are not in school, there may be a spike in petty crimes and burglary (Perry *et al.*, 2013).

Sometimes it is necessary to identify both the spatial and temporal components of crime patterns (Bachner, 2013). Crime analysts commonly conduct spatiotemporal analyses by creating heat maps. A heat map is a table (possibly created in Microsoft Excel) that shows, through colour intensity, the relative frequencies of crimes with different dates, times, and conditions (Perry *et al.*, 2013).

A more complex heuristic method involves the use of spatiotemporal additive models (Perry *et al.*, 2013). These models are extensions of regression models on grids. Input data include probabilities that each grid cell has a particular spatiotemporal feature at a particular time. The models combine the spatiotemporal features of the crime area with crime incident data to predict the location and time of future crimes. The models produce a probability that a crime will be committed at a certain place and time conditioned on the spatiotemporal features of the area in which past crimes were committed (Perry *et al.*, 2013).

5.3.3.1.4 A combined approach

An example software package that uses Big Data analytics to combine all the place-based techniques discussed above into a single predictor of crime, is HunchLab by Azavea. (2015). As mentioned in the introduction to Section 5.3.3, HunchLab is the second most employed predictive policing software package in the USA. HunchLab's forecasting engine uses an ensemble decision tree learning approach to incorporate a wide range of crime theories into a single prediction of criminal risk (Azavea., 2015).

HunchLab includes each crime theory by deriving individual sets of variables that represent the concepts within each theory. Specifically, the learning algorithm incorporates the following crime theories and associated data sources (Azavea., 2015):

- Traditional hotspot maps are represented by baseline crime levels;
- Near-repeat patterns are represented by measures of event recency;
- Risk Terrain Modelling is represented by measures of the proximity and density of geographic features such as bars, schools, and transit stops;
- Routine activity theory is incorporated by measures of the proximity and concentration of known offenders, police guardianship, and exposure of targets (such as population, parcels, or automobiles);
- Collective efficacy is represented by socioeconomic indicators, measures of heterogeneity, etc.;
- Temporal cycles are represented by incorporating seasonality, time of month, day of week, time of day, etc.;
- Recurring temporal events are included e.g. Holidays, sporting events, etc.; and
- Weather is included e.g. Temperature, precipitation, etc.

5.3.3.2 Person-based techniques

Compared to place-based predictive techniques, person-based methods are much less mature; they are also laden with privacy and civil rights concerns (Ferguson, 2012; Perry *et al.*, 2013; Robinson and Koepke, 2016; The Leadership Conference on Civil and Human Rights *et al.*, 2016; Ferguson, 2017a). Person-based techniques focus on predicting future offenders and likely victims of crime. They also focus on identifying perpetrators by creating perpetrator profiles that accurately match likely offenders with specific past crimes (Perry et al., 2013).

5.3.3.2.1 Predicting future offenders

By predicting future offenders, it is anticipated that person-based predictions can assist with decisions regarding reasonable suspicion, bail, sentencing and parole

(Simmons, 2016; Joh, 2017; Kehl *et al.*, 2017). Specifically, person-based predictors can assist:

- patrol officers in focusing their investigative efforts by identifying likely suspects;
- magistrates in granting search warrants by predicting the likelihood of an infraction based on the facts presented in a warrant application;
- judges in granting bail by predicting the chances that a defendant will return to court for trial; and
- sentencing judges by predicting the risk that a convicted defendant will likely reoffend (Simmons, 2016; Joh, 2017; Kehl *et al.*, 2017).

Predictive analysis focused at predicting offenders, focuses on identifying individuals who may become offenders, and on predicting the dynamics of organised crime (Perry et al., 2013). Examples of possible offenders are probationers and parolees at greatest risk of reoffending, domestic violence cases with a high risk of injury or death, and mental health patients at greatest risk of future criminal behaviour or violence (Perry et al., 2013). Here, conventional methods rely on clinical techniques that add up the number of risk factors to create an overall risk score. The corresponding predictive analytics methods use regression and classification models to associate the presence of risk factors with a percentage chance that a person will offend (Perry et al., 2013).

The city of Chicago, for example, uses the hotly contested computer-generated Strategic Subjects List (SSL, or “heat list”) (City of Chicago, 2017). The SSL is a list of people that are predicted most likely to become involved in a shooting, either as a perpetrator or as a victim (Robinson and Koepke, 2016). The SSL algorithm uses a social network analysis method (see Section 4.2.4), where each person’s risk score depends on their past behaviour, as well as the past offenses, recorded gang affiliations, and criminal justice records of their co-accused (Robinson and Koepke, 2016).

A similar system is Intrado’s “Beware” (Weller, 2016). The system uses information collected by commercial data brokers to assign a “threat score” to each member of the community (Robinson and Koepke, 2016). Data sources include public records (such as criminal history, vehicle registrations, address databases, property records, etc.) and social media data such as Twitter and Facebook feeds; and are used to assess the likelihood of an individual committing a crime (Weller, 2016).

Both Chicago's SSL and Intrado's "Beware" are shrouded by civil rights concerns. These issues, challenging the future of predictive policing, will be further discussed in Section 5.3.5. Also of relevance to predicting offenders, are methods that identify criminal groups such as gangs that are likely to carry out violent assaults on each other in the near future. One way of assessing the latter is near-repeat modelling on recent intergroup violence (Perry et al., 2013). These methods can also be used to assess the risk that an individual will become a victim of crime (Perry et al., 2013).

5.3.3.2.2 Predicting perpetrator identities

Predictive analysis focused at predicting perpetrator identities use available information from crime scenes to link suspects to crimes, both directly and by processes of elimination (Perry et al., 2013). Conventional approaches largely trace links manually, with assistance from simple database queries (usually, the names, criminal records, and other information known about the suspects), and manual requests to review license plates and sensor data such as CCTV footage (Perry et al., 2013). Predictive analytics automate the linking, matching available clues to potential suspects across very large data sets. The latter includes intelligence and master name databases and third-party databases such as motor vehicle registries, and pawn data (Perry et al., 2013).

Predictive analytics also makes use of GPS tracking. Analysts can perform queries of GPS tracking databases to see whether any tracked offenders were in the immediate vicinity of a crime (Perry et al., 2013). As text and image mining techniques mature, predictive analytics will increasingly leverage sensor data such as CCTV footage, and various sources of social media data (Perry et al., 2013) such as Facebook or Twitter. Analysts could, for example, determine whether the same license plate was spotted repeatedly near multiple burglaries or robberies that appear to be part of the same crime series; or facial recognition software could identify perpetrators who commit crimes in front of surveillance cameras (Perry et al., 2013).

Other areas of interest include anchor point analysis (or geographic profiling) and crime linking (Perry et al., 2013). Crime linking aims to identify which crimes are part of a series (Perry et al., 2013). This sort of analysis can be done manually or with computer assistance. In the manual version, the analyst sets up a table that compares key attributes of crimes committed by a known offender to other crimes that have not yet been matched. A crime that matches a large fraction of the attributes has a high probability of being part of the same crime series (Perry et al., 2013). In the automated version, computer software computes a probability that a recent crime is part of a crime series (Perry et al., 2013).

Geographic profiling determines the most probable area of a serial offender's search base (residence or other familiar location) through an analysis of his or her crime locations (Perry et al., 2013). Software programs employ distance-decay functions to model the geographic space associated with a spree of crimes, in order to produce a probability grid over an area estimating the offender's home base based on each grid cell's spatial relation to the crimes (Perry et al., 2013).

5.3.4 Predictors of crime

Emerging predictive policing methods incorporate data from a wide range of sources to generate the place-based or person-based predictions discussed in Section 5.3.3. Depending on their underlying theoretical basis, some predictive systems, such as PredPol (Mohler *et al.*, 2016; PredPol Inc., 2018), input crime incident variables (and event recency) only. Other predictive systems, such as HunchLab (Azavea., 2015), take various factors (e.g. community events and weather) into consideration.

Predictors that rely solely on crime incident variables (sourced from 911 calls or police reports) are often based on near-repeat theory. These predictors and the associated near-repeat theory were explained in Section 5.3.3.1.1. In addition to crime incident variables, supplementary variables can be divided into three categories, namely spatial, temporal, and behavioural variables (Bachner, 2013;

Perry et al., 2013). Table 5-1 shows examples of supplementary variables commonly used in predictive policing listed by Bachner (2013).

Spatial variables include indicators of potential targets such as shopping malls, property values, area demographics, population density and residential instability. Spatial variables also include indicators of potential escape routes such as highways and dense foliage; and indicators of criminal residences such as bars, adult retail stores and public health information (Bachner, 2013).

Table 5-1: Examples of variables used in predictive policing. Source: Bachner (2013).

Spatial Variables	Temporal Variables	Social Network Variables
Indicators of Areas with Potential Victims/Targets • Shopping malls • Property values • Hotels • Area demographics • Population density • Residential instability **Indicators of Escape Routes** • Highways • Bridges • Tunnels • Public transportation • Railways • Dense foliage **Indicators of Criminal Residences** • Bars and liquor stores • Adult retail stores • Fast food restaurants • Bus stops • Public health information • Areas with physical decay	• Payday schedules • Time of day • Weekend vs. weekday • Seasonal weather (e.g., hot versus cold weather) • Weather disasters • Moon phases • Traffic patterns • Sporting and entertainment events	• Kinship • Friendship • Affiliation with an organization • Financial transaction • Offender/victim

Crime is commonly observed to concentrate in certain neighbourhoods, particularly those characterised by poverty, racial segregation of minority groups, and high concentrations of single parent families (Sampson, 2006). Consequently, socioeconomic and demographic indicators such as poverty and race are often used as risk-factors for crime (Sampson, 2006; Taylor et al., 2015). However, the theory of collective efficacy (Bandura, 2000; Browning, 2002; Sampson, 2006) advocates a shift away from community-level correlations, and sheds light on the underlying social mechanisms at work within high-crime neighbourhoods.

Collective efficacy refers to a community's ability to self-regulate crime; and it is highly dependent on social cohesion (Browning, 2002). Factors such as poverty, racial segregation, single parent families and inequality (Hagan and Peterson, 1995; Kelly, 2000) degrade the social cohesion within a community, and consequently weaken the community's resilience to crime.

Another theory guiding the choice of spatial indicators is the broken windows theory (Sampson and Raudenbush, 2004). According to the "broken windows" theory of urban decline, minor forms of public disorder lead to serious crime and a downward spiral of urban decay (Sampson and Raudenbush, 2004). Two alternate explanations are given for this phenomenon.

The first explanation is that visual cues such as graffiti, public intoxication, garbage, and abandoned cars are thought to attract criminal offenders, who assume that residents are indifferent to the state of the neighbourhood (Sampson and Raudenbush, 2004). The alternative reasoning is the notion that perceived disorder leads to perceived disinterest of authorities in the community, and the subsequent hopelessness of residents (Sampson and Raudenbush, 2004). Following from the broken windows theory is the leading indicators theory (Groff and Vigne, 2002; Perry et al., 2013) in which minor crimes such as graffiti precede more serious crimes.

Temporal cycles have a strong effect on most types of crime (Perry *et al.*, 2013; Towers *et al.*, 2018). Temporal variables, therefore, incorporate temporal cycles by including features such as seasonality, time of the month, day of the week, time of the day, etc. (Bachner, 2013). Temporal variables also often include recurring temporal events (e.g. payday, holidays, social and sporting events, school schedules, etc.), traffic patterns, and daily weather patterns (Bachner, 2013).

While less pronounced than seasonality, it has been found that weather has an influence on certain types of crime (Perry et al., 2013; Towers et al., 2018). For example, Bushman et al. (2005) showed that higher temperatures correlate with higher levels of aggression, even when controlling for such factors as season and time of day. An environmental factor not commonly employed in predictive policing models is air pollution. However, new studies have shown a positive correlation between crime and this previously unexplored environmental indicator (Mapou et al., 2017; Bondy et al., 2018). In this study, the relationship between crime and air pollution will be further explored.

Social network variables include family and friendship networks, affiliation to organisations, financial transactions, offender or victim status (Bachner, 2013), and, more recently, social media activity such as Twitter messages (MEDIA4SEC., 2016; Anže Žitnik, 2019).

5.3.5 Challenges in practice

As adoption of predictive policing instruments becomes more prevalent, police departments and developers are being met with opposition from civil rights advocates expressing major concerns regarding privacy, undue suspicion and discrimination (Robinson and Koepke, 2016; The Leadership Conference on Civil and Human Rights *et al.*, 2016; Ferguson, 2017a). These concerns are exasperated by the limited amount of evidence supporting the accuracy and effectiveness of predictive policing tools; and the shroud of secrecy surrounding proprietary software (Robinson and Koepke, 2016; Simmons, 2016; The Leadership Conference on Civil and Human Rights *et al.*, 2016; Ferguson, 2017a; Joh, 2017). In the following sections, four major factors challenging the future of predictive policing will be discussed.

5.3.5.1 Limited evidence of effectiveness

Despite widespread vendor claims of effectiveness, there is limited evidence that predictive policing works (Groff and Vigne, 2002; Bennett Moses and Chan, 2016; Robinson and Koepke, 2016; Weller, 2016; Anže Žitnik, 2019).

Regarding place-based predictions, it is generally accepted that hot spot policing generates noteworthy crime reductions (Braga, 2001; Bowers *et al.*, 2011; Braga *et* *al.*, 2014); and studies by the developers of PredPol showed that its predictive capacity incrementally improves on these earlier methods of predicting crime (Mohler *et al.*, 2016). However, improvements are arguably insufficient, as three police departments in the US (two users of PredPol, and one user of IBM software) have reportedly rejected predictive policing due to minimal benefit (Robinson and Koepke, 2016). Furthermore, a RAND evaluation (Hunt *et al.*, 2014) of a predictive policing tool developed in-house by the Shreveport police department in Louisiana, found no statistical evidence that the program reduced crime.

Similar results were found for person-based predictors. A RAND evaluation of an early version of Chicago's Strategic Subject List found that the effort was not successful in reducing gun violence (Saunders *et al.*, 2016). Chicago's police department claimed that the report's findings are no longer relevant, however, as the prediction model has since been updated and improved (Chicago Police Department., 2016).

5.3.5.2 Sampling bias and discrimination

It is well known that police crime reports primarily document law enforcement's response to the reports they receive and situations they encounter, and are not an accurate record of the crime that happens in a community (Robinson and Koepke, 2016; Simmons, 2016; The Leadership Conference on Civil and Human Rights et al., 2016; Joh, 2017). Sampling bias, therefore, is inherent in crime data.

The number of crime reports are greatly influenced by what crimes citizens choose to report, the places police are sent on patrol, and how police decide to respond to the situations they encounter (Robinson and Koepke, 2016). A US National Crime Victimisation Survey found that around 40 to 50 percent of violent crime victimisations are not reported to police, while around 60 percent of property crime victimisations go unreported (Langton *et al.*, 2012).

Furthermore, for certain types of crimes (such as drug offenses and illegal gambling) police statistics are not reflective of the level of crime, but rather reflect the level of resources dedicated to its detection (Robinson and Koepke, 2016). Enforcement practices therefore, which vary widely from one neighbourhood to another, have a much larger impact on statistics for some crimes than for others (Robinson and Koepke, 2016).

A major concern is that seemly objective predictive policing algorithms based on this biased data will further intensify unwarranted discrepancies in enforcement (Robinson and Koepke, 2016; Simmons, 2016; The Leadership Conference on Civil and Human Rights et al., 2016; Joh, 2017). Marijuana possession arrests, for example, are notoriously biased towards black Americans. As a consequence, predictive systems that incorporate these biased historical crime statistics may lead to a cycle of self-fulfilling prophecies and reinforced discrimination (Robinson and Koepke, 2016). The same concern can be extended to the use of personal information like posts from social media (Anže Žitnik, 2019); where monitoring social media can lead to discrimination, over-criminalisation and unwarranted monitoring of youth or minority groups.

5.3.5.3 Limited focus on community needs

A major concern of civil rights advocates is that predictive policing algorithms are primarily used to further police already over-policed communities, rather than to meet community needs (Robinson and Koepke, 2016; The Leadership Conference on Civil and Human Rights *et al.*, 2016). Despite the challenges inherent in predictive policing, The Leadership Conference on Civil and Human Rights *et al.* (2016) advocate that predictive policing can be used for good; for example, by exploring its role in more effectively allocating social services resources.

Social services interventions such as educational opportunities, job training, and health services can help to address problems for at-risk individuals and communities before crimes occur (The Leadership Conference on Civil and Human Rights *et al.*, 2016). As stated in the final report of the President's Task Force on 21st Century Policing (2015: 8) , "[T]he justice system alone cannot solve many of the underlying conditions that give rise to crime. It will be through partnerships across sectors and at every level of government that we will find the effective and legitimate long-term solutions to ensuring public safety."

However, despite the opportunities, the evidence points to the contrary (Robinson and Koepke, 2016; The Leadership Conference on Civil and Human Rights et al., 2016). A comprehensive review carried out by Robinson and Koepke (2016) showed that applications of place-based predictions focused narrowly on reducing crime rates through intensified enforcement; with little focus on incorporating other measures of community need and police performance such as building community trust, and reducing coercive tactics (Hunt *et al.*, 2014).

Similar outcomes were observed for person-based predictors. Chicago's Strategic Subjects List was originally intended to be a carrot-and-stick approach, where individuals on the list would be warned against further criminal activity, while at the same time be offered assistance in obtaining a job or social services (Robinson and Koepke, 2016). However, it was found that offers of assistance did not materialise, and the program only led to increased contact with a group of people already in relatively frequent contact with police (Robinson and Koepke, 2016; Saunders *et al.*, 2016).

A major factor contributing to the lack of efficacy of predictive policing systems is a lack of guiding standards on how to use the predictions (Willis, 2011; Robinson and Koepke, 2016). Rather than changing their tactics, police tend to focus on generating more citations and arrests (Willis, 2011; Robinson and Koepke, 2016). What is lacking from current policing systems, therefore, is a holistic framework for community safety that proactively and systematically identifies community problems, prioritises tactics and document results (Robinson and Koepke, 2016).

Such a framework would be similar to the adaptive management frameworks proposed for complex systems in Section 3.4 of this thesis. To the author's knowledge, only HunchLab (Azavea., 2015) incorporates such a framework for tracking the effectiveness of police tactics (Robinson and Koepke, 2016).

5.3.5.4 Lack of transparency

With privacy and freedom at stake, coupled with issues surrounding sampling bias and model effectiveness, civil rights advocates have called for transparency regarding predictive policing algorithms (Joh, 2017). Greater transparency is needed to foster well-informed public debate about system aims and implementation, the selection of input features and data sources, and algorithm design (The Leadership Conference on Civil and Human Rights et al., 2016).

5.4 Conclusions regarding model requirements

After reviewing the content of this chapter, the model requirements proposed in Section 3.8 can be refined within the context of crime management as follows:

- *Requirement 2: Predictive model* – In Section 3.8, it was prescribed that the prototype model be a reliable tool for systems-level scenario analysis; capable of predicting the relative influence of selected KPIs and combinations thereof on target KPIs. In terms of crime management, "target KPIs" are further specified to represent indicators relating to crime. In addition, Section 5.3.5.3 provides further support for this requirement. Predictive policing activities were criticised for their strong focus on numbers, and limited attention to community needs. A predictive model capable of identifying leading cross-sector causes of crime will aid in addressing the call for cross-sector partnerships aimed at developing effective long-term solutions to ensuring public safety (President's Task Force on 21st Century Policing., 2015).

5.5 Conclusions regarding model design

After reviewing the content of this chapter, the smart city design solutions presented in Section 4.6 can be refined within the context of crime management as follows:

- *Design element 1: KPI framework* – In Section 4.6, it was proposed that a subset of the City Protocol Society's *City Anatomy Indicators* be used as input to the prototype model. Here, it is proposed that the commonly employed predictors of crime summarised in Section 5.3.4 be used as a guide when selecting KPIs from the set of CPA indicators.

- *Design element 2: Predictive model* – In Section 4.6, it was proposed that a modelling approach employing Bayesian Neural Networks would be effective in handling the anticipated complexity of city interactions, while at the same time providing a measure of model reliability. In this chapter, Section 5.3.3 describes commonly employed place-based and person-based predictive policing techniques. Due to the nature of the required model (see model requirement 2), it is evident that a place-based neural network including location features will be the most effective modelling approach.

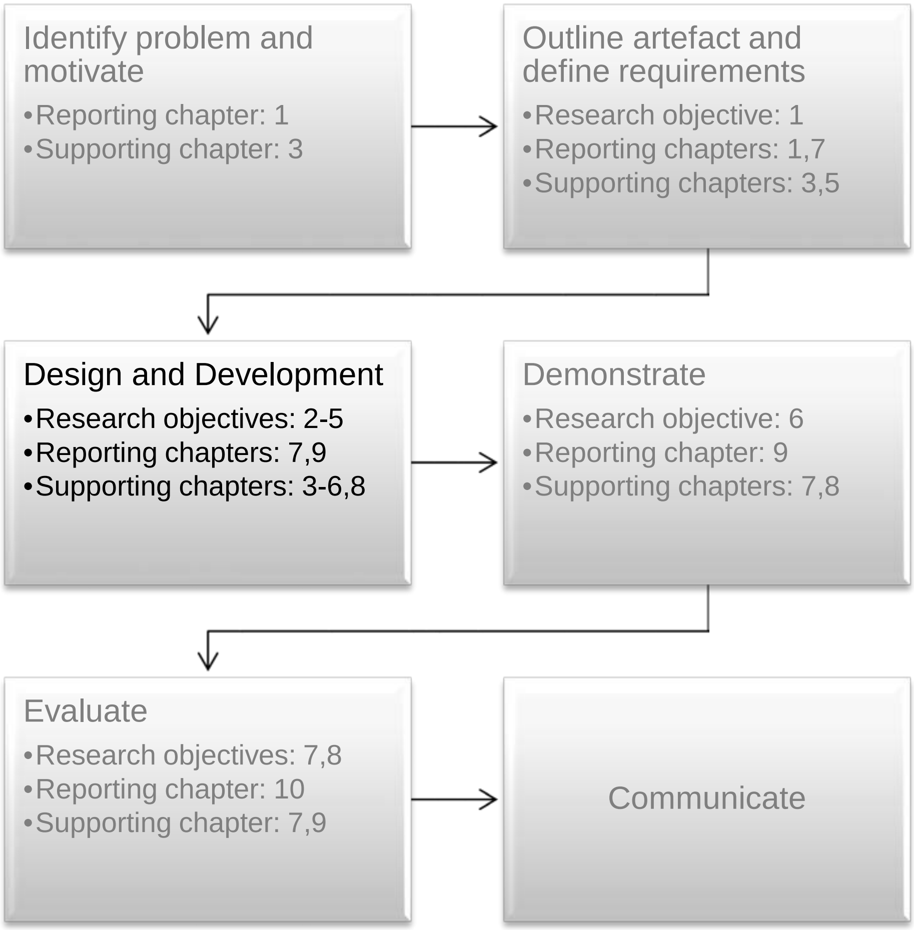
5.6 Summary

The purpose of this chapter was to further develop the requirements of the prototype model to be developed in this study (Section 3.8), and the associated smart city design solutions (Section 4.6), within the context of crime management. This was achieved (Sections 5.4 and 5.5) by reviewing the state-of-the-art in real-time intelligence centres (Section 5.2) and predictive policing (Section 5.3).

Specifically, this chapter further developed $RO_{1,2,4}$ (Figure 5-1) by identifying model requirements specific to crime management; identifying model parameters which are relevant predictors of crime; and by identifying modelling techniques commonly employed in crime forecasting.

In Section 4.6, it was a proposed that a modelling approach employing Bayesian Neural Networks would be effective in handling the anticipated complexity of city interactions, while at the same time providing a measure of model reliability. The following chapter explores Bayesian Neural Networks in detail.

Chapter 6. Bayesian Neural Networks



Research objective addressed in this chapter:

RO$_1$: Identify the functional, construction and environmental requirements of an effective model.

RO$_2$: Identify relevant input and output parameters.

RO$_3$: Identify and characterise available data sources.

RO$_4$: Identify the modelling technique to be used to develop the model.

RO$_5$: Develop the model.

RO$_6$: Demonstrate the application of the model.

RO$_7$: Evaluate the efficacy of the model.

RO$_8$: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

Figure 6-1: Research objective and design science research activity addressed in this chapter.

6.1 Introduction

The main objective of this study is to develop a prototype model for crime management in smart cities (Section 1.5). Chapters 3 to 5 of this thesis address $RO_{1\text{-}3}$ (Figure 6-1) by providing the supporting literature necessary to refine the requirements of the proposed model, identify relevant input and output parameters, and identify sources of available data. The purpose of this chapter (Figure 6-1) is to address RO_4, by identifying the modelling technique to be used to develop the prototype model.

In order to choose a modelling technique, knowledge of the model requirements is required. Model requirements derived in Chapters 3 and 5 are consolidated in Section 7.2 of this thesis. Based on the requirements identified in Section 7.2, an effective model would need to:

1. Be able to incorporate known or unknown complexities and inter-dependencies between variables;
2. Indicate the uncertainty in model predictions; and
3. Be a reliable tool for interpreting the relative influence of input parameters and combinations thereof on crime.

Artificial neural networks are well known for their ability to automatically approximate any function (Tan *et al.*, 2006; Han *et al.*, 2012). For this reason, they are often used when the relationship between variables is unknown or complex. In order to meet the first requirement identified above, artificial neural networks are the modelling approach of choice in this study. An overview of artificial neural networks is given in Section 6.2. The fundamental concepts of a network architecture and learning algorithms are introduced; followed by discussions relating to optimising the predictive accuracy of a model. Specifically, choices relating to network topology, regularisation, feature selection and ensemble methods will be discussed.

In Section 6.3, Bayesian neural networks are introduced as a means of capturing and expressing modelling uncertainty. The theory explaining Bayesian neural networks is discussed, and available tools for Bayesian learning are summarised.

The implementation of Bayesian neural networks followed in this study is summarised in Section 6.4.

Lastly, Section 6.5 discusses matters relating to the interpretation of the developed model. In this study, sensitivity analysis will be used to determine the relative influence of input parameters and combinations thereof on crime. However, issues surrounding the interpretation of causality in observational studies are noted.

6.2 Artificial neural networks

6.2.1 Architecture and application

An artificial neural network (ANN) is a data mining technique that mimics the learning process of biological neural systems (Tan *et al.*, 2006; Han *et al.*, 2012). The human brain consists of nerve cells called neurons, linked together with other neurons via strands of fibre called axons (Tan *et al.*, 2006). Axons are used to transmit nerve impulses between neurons whenever the neurons are stimulated (Tan *et al.*, 2006). A neuron is connected to the axon of another neuron via dendrites. The contact point between a dendrite of a receiving neuron and the axon of a transmitting neuron is called a synapse (Tan *et al.*, 2006). The structure of the biological neural system is illustrated in Figure 6-2. It is understood that the human brain learns by changing the strength of the synaptic connection between neurons upon repeated stimulation by the same impulse (Tan *et al.*, 2006).

There are different kinds of artificial neural networks; however, feed-forward neural networks are the most common (Han *et al.*, 2012). A multilayer feed-forward neural network consists of an input layer, one or more hidden layers, and an output layer (Han *et al.*, 2012). An example of a multilayer feed-forward network is shown in Figure 6-3. The number of hidden layers is arbitrary, although in practice, usually only one is used (Han *et al.*, 2012).

As the name implies, in a feed-forward network, the nodes in one layer are connected only to the nodes in the next layer (Tan *et al.*, 2006). This is in contrast to recurrent neural networks, in which there may be connections between nodes in the same layer, or between a layer and a previous layer (Tan *et al.*, 2006).

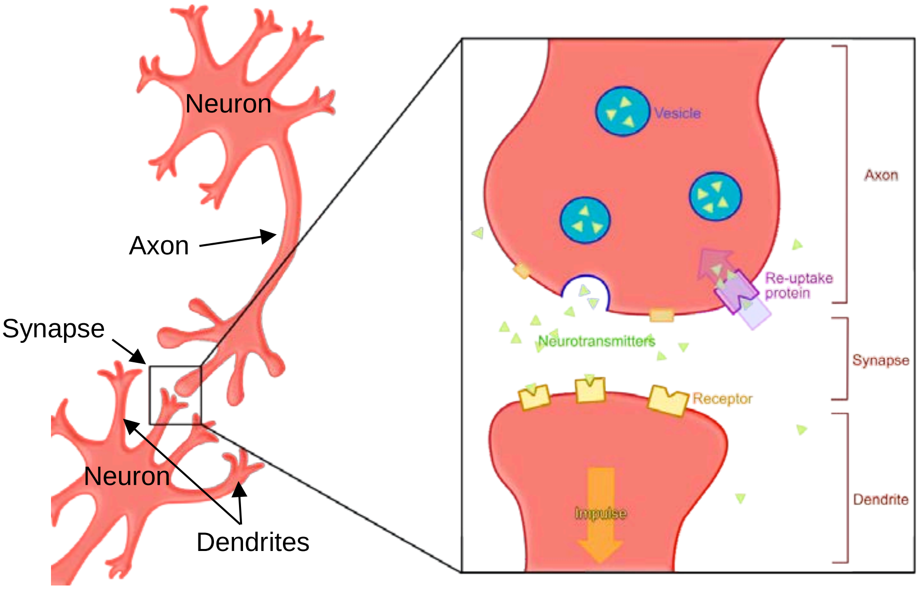


Figure 6-2: The structure of the biological neural system. Source: adapted from Wikimedia Commons.

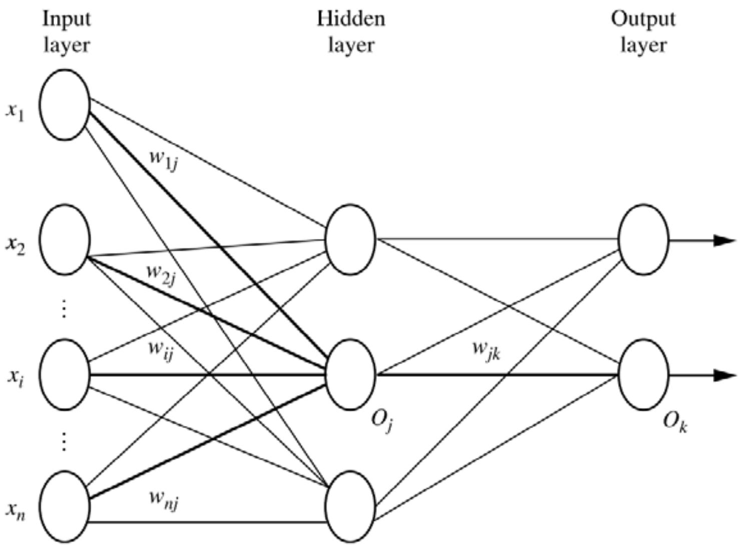


Figure 6-3: Example of a feed-forward neural network. Source: Han *et al.* (2012).

The output of each unit in the hidden and output layers of a feed-forward ANN is calculated as a function of the weighted sum of the outputs from the units in the previous layer (Han *et al.*, 2012). For example, the output O_j of unit j in the hidden layer in Figure 6-3 is given by:

$$O_j = f\left(\sum_i w_{ij}\, O_i + \theta_j\right) \qquad \text{Equation 6-1}$$

where w_{ij} is the weight of the connection from unit i in the previous layer to unit j; O_i is the output of unit i from the previous layer; and θ_j is the bias of the unit (Han *et al.*, 2012). The function f is called the activation function of the unit, and symbolises the activation of the neuron represented by the unit (Han *et al.*, 2012). Commonly used activation functions include the linear, sigmoid, hyperbolic tangent, and sign functions (Tan *et al.*, 2006).

The output(s) of the output later gives the networks predictions (Han *et al.*, 2012). ANNs can be used for classification or numeric prediction (Tan *et al.*, 2006; Han *et al.*, 2012). For classification problems, two or more output units are used, representing the possible prediction classes. The feed-forward ANN shown in Figure 6-3 is an example of a classification neural network. For numeric prediction applications, typically only one output unit is used. In this study, neural networks will be used to make numeric predictions.

Given enough hidden units and training samples, multilayer feed-forward networks can closely approximate any function (Bhadeshia, 1999; Han *et al.*, 2012). From a statistical point of view, multilayer feed-forward neural networks perform nonlinear regression; and are of great value when little is known about the relationship between model attributes and targets (Han *et al.*, 2012).

6.2.2 Learning and model evaluation

The weights and biases introduced in Equation 6-1 are initially chosen at random. The output of the neural network, therefore, does not initially match with experimental data. Consequently, neural networks are “trained” to determine the set of weights and biases that best fit the input data.

Typically, available data are split into a training set and a testing set. For a given training set of input-target pairs, $D = \{x^m, t^m\}$ (where m is a label running over the pairs), the goal of an ANN learning algorithm is to determine a set of weights w that minimise the total sum of squared distances between the network's predictions and the known target values in the training set (MacKay, 1992; Foresee and Hagan, 1997; Tan *et al.*, 2006; Han *et al.*, 2012). This error term, $E_D(w)$, is expressed as follows:

$$E_D(w) = \frac{1}{2}\sum_{m}[y^m(x^m, w) - t^m]^2 \qquad \text{Equation 6-2}$$

where y^m is the prediction of the neural network for inputs x^m and weights w.

The learning algorithm most commonly employed to train feed-forward neural networks is back-propagation (Tan *et al.*, 2006; Han *et al.*, 2012). Back-propagation randomly initialises the set of weights w, and then learns by using the gradient descent method to iteratively determine the optimal weights of the output and hidden units of a neural network (MacKay, 1992; Tan *et al.*, 2006; Han *et al.*, 2012). For each iteration, the weights are modified so as to minimise the mean-squared error between predictions and target values. The algorithm is called back-propagation because weight modifications are propagated in the backwards direction (starting from the output layer, and propagating through to the first hidden layer) (Han *et al.*, 2012).

Once the model has been trained, it is evaluated against the testing set. Similar to Equation 6-2, the test error, or generalisation error, of the trained model is simply a measure of the total squared distances between the network's predictions and the known target values in the testing set. While the goal of the back-propagation learning algorithm is to minimise E_D, the goal of network design is to develop a model that generalises well to new data (MacKay, 1992; Foresee and Hagan, 1997).

Key design parameters that influence how well a neural network model generalises to new data include the choices of network typology, initial weights and biases, regularisation constant and input parameters. These are briefly discussed in the following sections.

6.2.3 Network topology and the initialisation of weights

Before training begins, the user decides on the network topology by specifying the number of units in the input layer, the number of hidden layers, the number of units in each hidden layer, and the number of units in the output layer (Han *et al.*, 2012). There are no clear rules as to the "best" number of hidden layer units; and, in general, network design is a trial-and-error process (Han *et al.*, 2012).

By increasing model complexity, multilayer feed-forward neural networks can closely approximate any function (Bhadeshia, 1999; Han *et al.*, 2012). However, it is generally accepted that overly complex models, while fitting the training set well, may not generalise well to new data (Foresee and Hagan, 1997). This is termed over-fitting, and is illustrated by Bhadeshia (1999) in Figure 6-4.

Figure 6-4 shows three attempts at modelling a given set of data. A linear model (Figure 6-4a) is too simple and does not capture the real complexity in the data. This model is said to underfit the data. An overly complex function such as that illustrated in Figure 6-4c accurately models the training data but generalises poorly to the test set. This model is said to overfit the training data. The optimum model is illustrated in Figure 6-4b.

The training and test errors are shown schematically as a function of model complexity in Figure 6-4d. The training error tends to decrease continuously as the model complexity increases. However, once the model complexity surpasses what is necessary to model the system under investigation, it is said to overfit the data, and the test error will increase. In practice, the choice of network topology is chosen by first training a few network topologies, and then choosing the one that minimises the test error. The model with the lowest test error generalises best to unseen data.

Another factor influencing the accuracy of the trained model is the initialisation of the weights. The set of weights w that minimises E_D (Equation 6-2) is not unique. This is because, depending on the complexity of the model, the error surface $E_D(w)$ in weight space w may not have a single global minimum, but rather a number of local minima (Tan *et al.*, 2006). The initial values of the weights may therefore affect the resulting accuracy. It is therefore common to repeat the training process for a given network topology with a different set of initial weights (Han *et al.*, 2012).

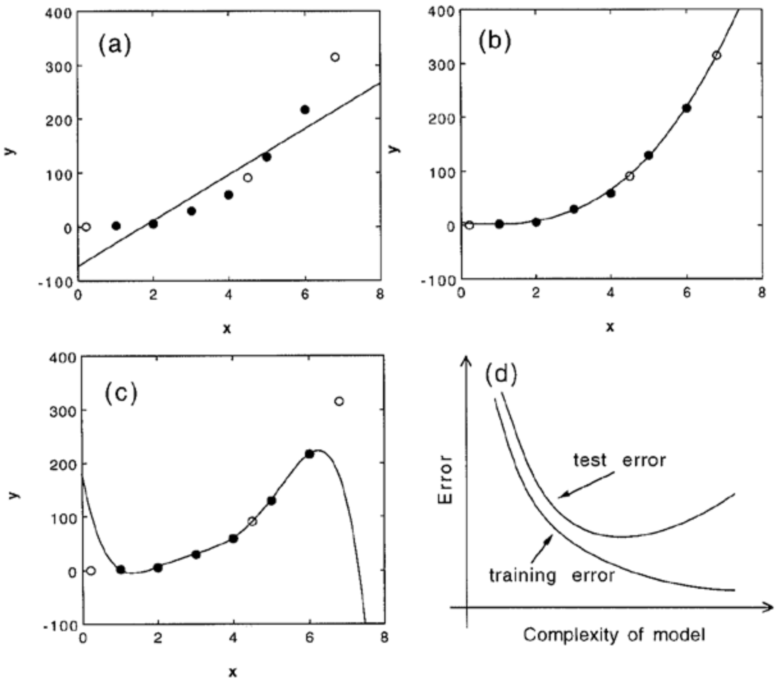


Figure 6-4: Variations in the test and training errors as a function of model complexity in a case where y should vary with x³. The filled points represent the training data, and the circles represent the test data. (a) A linear function is too simple. (b) A cubic polynomial fits the training data well and generalises well to the test data. (c) A fifth order polynomial fits the training data well but generalises poorly to the test data. (d) Schematic illustration of the variation in the test and training errors as a function of model complexity. Source: Bhadeshia (1999).

6.2.4 Regularisation

Another method for improving generalisation constrains the size of the network weights and is referred to as regularisation (Foresee and Hagan, 1997) or weight decay (MacKay, 1992). In weight decay, regularisation terms that penalise large weights are added to the training cost function in the hope of achieving smoother or simpler models (MacKay, 1992; Foresee and Hagan, 1997). The intention of regularisation is to prevent the overfitting of noise. Building on Equation 6-2, the cost function of the learning optimisation problem becomes:

$$M = \underbrace{\alpha E_w(w)}_{\text{Regularisation term}} + \beta E_D(w) \qquad \text{Equation 6-3}$$

where α and β are black box parameters; and $E_w(w)$ is an expression of the sum of squares of the network weights given by:

$$E_W(w) = \frac{1}{2}\sum_i w_i{}^2 \qquad \text{Equation 6-4}$$

The relative size of the objective function parameters dictates the emphasis for training. If α << β, then the training algorithm will drive the errors smaller. If α >> β, training will emphasise weight size reduction at the expense of network errors, thus producing a smoother network response. The main problem with implementing regularisation is setting the correct values for the objective function parameters (Foresee and Hagan, 1997). As with topology optimisation (Section 6.2.3), this is also a trial and error process.

6.2.5 Dimensionality reduction and feature subset selection

Careful selection of input variables is essential in the construction of a good model (Sourmail, 2002; Tan *et al.*, 2006; O'Neil and Schutt, 2014). As datafication intensifies, the number of potential model inputs increase. Unlike model complexity (Section 6.2.3), however, including more and more input features ultimately leads to poorer performance (Neal, 2012). Accordingly, redundant and irrelevant features need to be reduced. Two key ways of reducing the number of input parameters are dimensionality reduction and feature subset selection (Tan *et al.*, 2006; Han *et al.*, 2012).

Dimensionality reduction refers to the reduction of the dimensionality of the data set by creating new attributes that are a combination of old attributes; while feature selection reduces the dimensionality of the dataset by simply selecting a subset of the original data set (Tan *et al.*, 2006). Typically, dimensionality reduction techniques use linear algebra to project the data from a high-dimensional space into a lower-dimensional space. Principal Components Analysis (PCA), for example, is a linear algebra technique that finds new attributes (principal components) that are linear combinations of the original attributes, and that are orthogonal to each other. The principal components capture the maximum amount of variation in the data (Tan *et al.*, 2006).

While some irrelevant and redundant attributes can be eliminated by using domain knowledge, selecting the best subset of features frequently requires a systematic approach (Tan *et al.*, 2006). The ideal approach to feature selection is to try all possible subsets of features as input to the target data mining algorithm. This is, however, not feasible; as for n attributes, there are 2^n subsets of features (Tan *et al.*, 2006; Han *et al.*, 2012).

The three standard approaches to feature subset selection are embedded, filter and wrapper approaches (Tan *et al.*, 2006). Embedded feature selection approaches occur naturally as part of the data mining algorithm. Filter approaches use an approach independent of the data mining task to perform feature selection before the data mining algorithm is run. Wrapper approaches, use the target data mining algorithm to find the best subset of features in a way similar to the ideal approach discussed above, but without enumerating all possible subsets.

6.2.6 Ensemble methods

Ensemble methods are often used to improve the accuracy of a learned model. Ensemble methods use a combination of learned models with the aim of creating an improved composite model (Han *et al.*, 2012). A commonly employed ensemble method is bagging. For a given combination of learned models, bagging simply takes the average value of each predictor for a given test tuple. It has been theoretically proven that a bagged predictor will always have an improved accuracy over a single learned predictor (Han *et al.*, 2012). In this study, a form of bagging is used to improve the accuracy of the learned model. This is described in Section 6.4.1.4.

6.3 Bayesian neural networks

6.3.1 Noise and uncertainty

While E_D (Equation 6-2) gives an overall perceived level of noise in the output parameter, it is an insufficient description of the uncertainties of prediction (Bhadeshia, 1999). This problem is illustrated in Figure 6-5. The best-fit function (i.e. the most probable values of the weights w) does not adequately describe the uncertainties in regions of the input space where data are sparse (B), or where the data are noisy (A) (Bhadeshia, 1999). There are many functions which can be fitted or extrapolated into uncertain regions of the input space, without compromising the fit in regions which are rich in accurate data (Bhadeshia, 1999). MacKay (1992) developed a Bayesian framework for neural networks which allows for the calculation of error bars representing this uncertainty. This Bayesian framework for neural networks will be explained in the following sections.

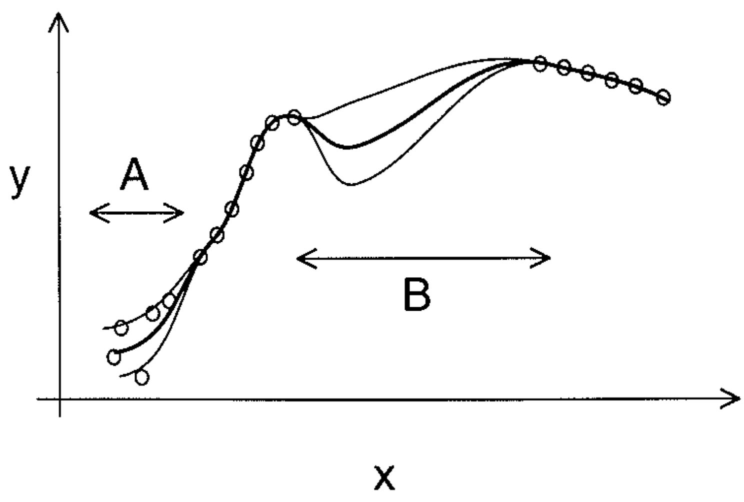


Figure 6-5: Schematic illustration of the uncertainty in defining a fitting function in regions where data are sparse (B) or where they are noisy (A). The thick line represents the best-fit function, while the thinner lines represent error bounds due to uncertainties in determining the weights of the neural network. Source: Bhadeshia (1999).

6.3.2 Bayes' theorem

A Bayesian approach to learning uses probability to express uncertainty in model predictions (MacKay, 1992; Neal, 1995; Neal, 2012). While the conventional Back-propagation algorithm (Section 6.2.2) determines the single best set of weights w that minimises the training error, a Bayesian approach to learning aims to determine a probability distribution of likely weights. Similarly, instead of determining the single best prediction based on the optimum weights, a Bayesian approach to prediction aims to determine a probability distribution of likely outputs.

Bayes' Theorem provides a way to incorporate uncertainty into the conventional Back-propagation algorithm (MacKay, 1992; Neal, 1995; Neal, 2012). Consider again the training set of input-target pairs, $D = \{x^m, t^m\}$ and the set of weights w introduced in Section 6.2.2. Bayes' Theorem states that:

$$P(w|D) = \frac{P(D|w)P(w)}{P(D)} \qquad \text{Equation 6-5}$$

The term $P(w|D)$ is the posterior probability of w conditioned on D. It is the probability that w correctly fits the model given the "evidence" D. In contrast, $P(w)$ is the prior probability of w. It is the unconditioned probability of w, and represents the prior assumptions regarding the distribution of w. $P(D|w)$ is the posterior probability of D conditioned on w. It is the probability of predicting D given the set of weights w. Lastly, $P(D)$ is the prior probability of D. It is the unconditioned probability of D, and represents the prior assumptions regarding the distribution of D.

6.3.3 Bayesian regularisation

Hinton (2013) provides a simplified explanation of a reduced form of Bayesian learning called Maximum a Posteriori (MAP) learning (MacKay, 1992; Neal, 1995; Neal, 2012). In MAP learning, a Gaussian prior is assumed for the weights, and Gaussian noise is added to the output of the neural network.

From a probability perspective, the goal of learning is to maximise the probabilities of producing the target values, t^m, in the training set i.e. the goal of learning is to maximise $P(D|w)$. Alternatively, the goal of learning is to find the most probable set of weights w given the training set D i.e. the goal of learning is to maximise $P(w|D)$. The cost function of the neural network learning algorithm can be expressed in terms of negative log probabilities (Hinton, 2013). The cost function therefore becomes:

$$Cost = -\log p(w|D) \qquad \text{Equation 6-6}$$

Equation 6-6 can then be expanded using Bayes' Theorem (Equation 6-5) as follows:

$$Cost = -\log p(D|w) - \log p(w) + \log p(D) \qquad \text{Equation 6-7}$$

If a Gaussian prior is assumed for the weights, then $p(w)$ can be expressed as:

$$p(w) = \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{w^2}{2\sigma_w^2}} \qquad \text{Equation 6-8}$$

and the negative log probability of $p(w)$ becomes:

$$\begin{aligned} -\log p(w) &= \frac{w^2}{2\sigma_w^2} + constant \\ &= \frac{1}{2\sigma_w^2} \sum_i w_i^2 + constant \end{aligned} \qquad \text{Equation 6-9}$$

If Gaussian noise is added to the output of the neural network, then the probability that y^m correctly predicts the target value t^m for a particular training tuple, can be given by:

$$p(t^m|y^m) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t^m - y^m)^2}{2\sigma^2}} \qquad \text{Equation 6-10}$$

where y^m is the prediction of the neural network for inputs x^m and weights w. Assuming that the output errors on the different training tuples are independent, $p(D|w)$ becomes:

$$p(D|w) = \prod_m p(t^m|y^m) \qquad \text{Equation 6-11}$$

Since the log function is monotonic, the negative log probability of $p(D|w)$ is approximated as a sum of the negative log probabilities of $p(t^m|y^m)$:

$$-\log p(D|w) \approx \frac{1}{2\sigma_D^2} \sum_m (t^m - y^m)^2 + constant \qquad \text{Equation 6-12}$$

Substituting Equation 6-9 and Equation 6-12 into Equation 6-7, the cost function becomes:

$$Cost = \frac{1}{2\sigma_D^2}\sum_m (t^m - y^m)^2 + \frac{1}{2\sigma_w^2}\sum_i w_i^2 + \log p(D) + constant \qquad \text{Equation 6-13}$$

Furthermore, by removing the terms that are independent of w and multiplying through by $2\sigma_D^2$, the cost function becomes:

$$Cost = \sum_m (t^m - y^m)^2 + \frac{\sigma_D^2}{\sigma_w^2}\sum_i w_i^2 \qquad \text{Equation 6-14}$$

The Bayesian approach to learning therefore provides an objective interpretation of the regularisation constant introduced in Section 6.2.4. Based on this interpretation, MacKay (1992) showed that the optimum regularisation constant can be inferred from the data; removing the need for computationally expensive trial and error optimisation. Furthermore, MacKay (1992) proposed how Bayesian learning could be implemented using this finding for a simple feed-forward neural network with one hidden layer (MacKay, 1992). This implementation framework is explained in Section 6.4. For a more detailed introduction to Bayesian neural networks the reader is referred to the original papers of MacKay (1992) and Neal (1995).

6.3.4 Tools for Bayesian learning

Pre-packaged tools for Bayesian learning include Model Manager developed by Sourmail (2004), software developed by Neal (2004), the *trainbr* function in MATLAB (MathWorks., 2018), Edward (Tran *et al.*, 2016) for Python, and the *brnn* package in R (Rodriguez and Gianola, 2018).

Sourmail (2004) developed software for training Bayesian neural networks called Model Manager as part of his PhD at Cambridge University (Sourmail, 2002). The software package creates a neural network model based on the optimisation algorithm proposed by MacKay (1992). The software is freely available for download, and is a collection of *tcl* scripts with a *tk* interface (Sourmail, 2004). The software is only applicable to simple feed-forward neural networks with one hidden layer.

Both the *trainbr* function in MATLAB (MathWorks., 2018) and the *brnn* package in R (Rodriguez and Gianola, 2018) implement Bayesian regularisation as proposed by MacKay (1992), and further developed by Foresee and Hagan (1997). The functions therefore, can be used to find the optimum regularisation parameters for two layer feed-forward neural networks. The functions do not however, return the standard deviation of the weights, and cannot be used for uncertainty applications.

Edward is a Python library for probabilistic modeling, inference, and criticism (Tran *et al.*, 2016). Edward fuses the three fields of Bayesian statistics and machine learning, deep learning, and probabilistic programming (Tran *et al.*, 2016). Both Edward and the software of Neal (2004) are based on the practical implementation of Bayesian neural network learning proposed by Neal (1995). The implementation is an expansion of that proposed by MacKay (1992) to more complex deep learning scenarios; and uses Markov chain Monte Carlo methods.

6.4 Model implementation

In this study, Bayesian Neural Networks will be implemented using the Model Manager software package developed by Sourmail (2004). In this section, the training process is explained, and the means of expressing uncertainty in model predictions is described.

6.4.1.1 Model architecture

A simple feed-forward neural network with one hidden layer forms the basis of the model (Sourmail, 2002). Since the model will be used to make numeric predictions, only one output unit will be used. The choice of input units and the number of units in the hidden layer are described next.

6.4.1.2 Choice of variables

As noted in Section 6.2.5, careful selection of input variables is essential in the construction of a good model. In this study, the dimensionality of the input data will be reduced using a combination of domain knowledge, and embedded and filter approaches. Inherent in the Bayesian framework for neural networks is the capacity for automatic relevance determination (MacKay, 1995; Neal, 2012). As such, the learning algorithm implemented in this study incorporates automatic relevance detection (Sourmail, 2002), where variables that are either redundant or irrelevant are affected a zero weight. The algorithm therefore has built-in resilience against an ‘incorrect’ choice of variables (Sourmail, 2002).

6.4.1.3 Data processing

Before training begins, the data is randomised and equally divided into a training and a testing set (Sourmail, 2002). Data is then normalised to bring the range of variations of all the variables to between -0.5 and 0.5 (Sourmail, 2002).

6.4.1.4 Training process

For a given set of input units, the predictive performance of a trained model is dependent on its topology (Section 6.2.3), the choice of initial priors for the weights (Section 6.2.3), and the choice of regularisation parameter (Section 6.2.4). Consequently, finding the optimal model for a given problem is a trial and error process that is often computationally expensive and time consuming.

In addition to providing a means for expressing modelling uncertainty, the learning algorithm employed in this study infers the optimal regularisation constant from the data (Section 6.3.3). This removes the need to manually optimise the regularisation parameters of alternative network topologies. Network design therefore, only focuses on varying the number of hidden units and the choice of initial priors for the weights.

The training process involves the training of a large number of models with different numbers of hidden units (typically 1 to 25), and different priors on the weights (typically 5) (Sourmail, 2002). The models are then used to make predictions on the unseen testing set and are ranked according to predictive performance.

An ensemble of models is then used to further improve the accuracy of the predictor (Section 6.2.6). To determine the optimum number of models to use in the “committee”, the combined prediction error is calculated for an increasingly large number of models (Figure 6-6). The selected models are then retrained on the full database (Sourmail, 2002).

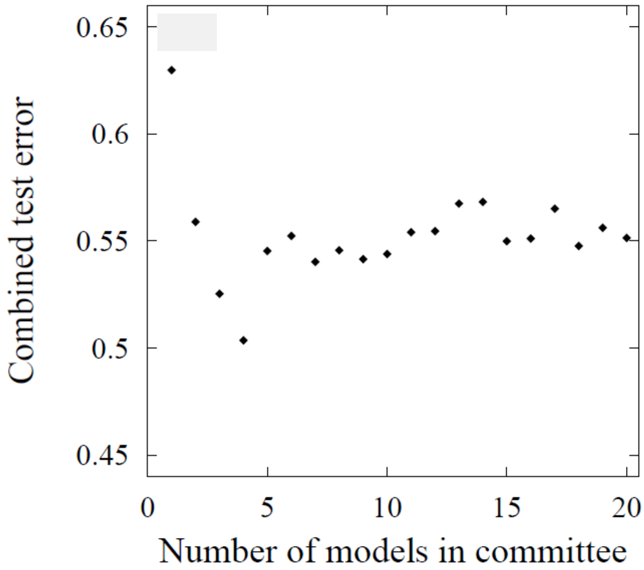


Figure 6-6: An example of the variation of the combined test error when the number of models in the committee is increased. In this case, the optimum number of models is found to be 4. Source: Sourmail (2002).

6.4.1.5 Representing uncertainty

The committee prediction is the average of the predictions of the models constituting the committee, while the error is calculated according to:

$$\bar{y} = \frac{1}{L}\sum_{l} y^{l}$$
$$\sigma^{2} = \frac{1}{L}\sum_{l} {\sigma_{y}}^{l^{2}} + \frac{1}{L}\sum_{l} (y^{l} - \bar{y})^{2} \qquad \text{Equation 6-15}$$

where L is the number of networks in the committee, and σ is the standard deviation. The exponent l refers to the model used to produce the corresponding prediction y^l (Sourmail, 2002).

6.5 Model interpretation

6.5.1 Sensitivity analysis

Artificial neural networks are effective tools for non-linear regression when the relationship between input and target variables is unknown or difficult to model using conventional regression techniques (Section 6.2.1). The nature of the relationship between the predictions of a neural network model and its input parameters is implicit in the architecture of the model and the values of the optimised weights (Bhadeshia, 1999). These weights, however, are not intuitively easy to interpret (Bhadeshia, 1999). For this reason, neural networks are often criticised as being black box predictors (Han *et al.*, 2012).

The purpose of this study is to develop a model for crime management in smart cities that can be used for project prioritisation and scenario analysis (Section 1.4). As such, knowledge of the relationship between input features (crime predictors) and model predictions (crime rates) is essential for identifying leading predictors. A commonly employed method used to assess the impact that a given input variable has on a network output, is sensitivity analysis (Bhadeshia, 1999). The input to the variable is varied while the remaining input variables are fixed at some value. Meanwhile, changes in the network output are monitored. The knowledge gained from this analysis form can be represented in rules such as “IF X decreases 5%THEN Y increases 8%.” (Han *et al.*, 2012).

6.5.2 A note on causality and observational studies

The predominant application of crime forecasting is to predict the crime rate for a particular location and time; or to identify the likelihood of a person offending (Section 5.3.3). For the application of this study, however, the goal is to develop a predictive model that can be used for decision-making and scenario analysis purposes (Section 1.4). As such, the model will be used to make assumptions about causality.

One of the mantras most commonly quoted by statisticians is that “correlation does not imply causation” (O’Neil and Schutt, 2014: 274). Correlation between two variables is often not the result of causality, but rather the result of known or unknown confounding factors causing a spurious association. The gold standard for establishing causality is the randomised experiment (O’Neil and Schutt, 2014). However, randomised experiments are not always possible for ethical or practical reasons.

When randomised experiments are not feasible, researchers revert to observational studies. O’Neil and Schutt (2014: 283) define observational studies as follows: “An observational study is an empirical study in which the objective is to elucidate cause-and-effect relationships in which it is not feasible to use controlled experimentation.” Most data science activity revolves around observational data.

By definition, the lack of controlled conditions in observational studies make them particularly susceptible to incorrect causal inferences. A comprehensive study of causation in observational data falls out of the scope of this study. However, the danger of incorrect causal inference is noted. This is with particular reference to Section 5.3.5 of this study which highlights the public debate surrounding predictive policing and undue racial discrimination resulting from questionable casual inference. The issue of causal inference from observational data is a very current topic, and the interested reader is referred to the statistical literature for further reading (Sloman, 2005; Morgan and Winship, 2007; Pearl, 2009).

6.6 Summary

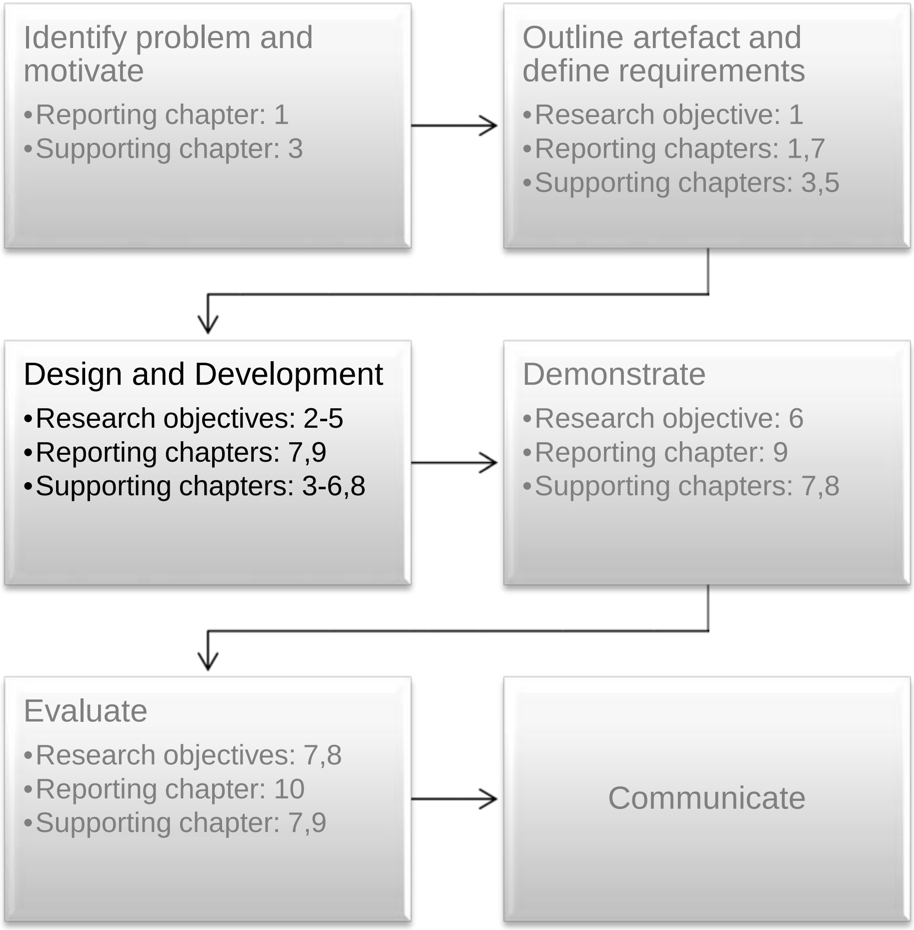
In this study, it is supposed that emerging trends in smart city integration can complement existing planning practices in South Africa. Specifically, it is proposed that a predictive model, that incorporates data from traditionally isolated management silos, can be developed for whole-system scenario analysis applications. In order to limit the scope of the investigation, the study focuses on only one aspect of smart cities, namely crime management.

The purpose of this chapter was to identify the modelling technique to be used to develop the model. Bayesian neural networks were introduced as the modelling approach of choice. The theory explaining Bayesian neural networks was explained, and available tools for Bayesian learning were summarised. The implementation of Bayesian neural networks followed in this study was summarised. The chapter concluded with a discussion on matters relating to the interpretation of the developed model. In this study, sensitivity analyses will be used to determine the relative influence of input parameters and combinations thereof on crime. However, issues surrounding the interpretation of causality in observational studies were noted.

Section 2 of this thesis provided the supporting literature necessary to refine the functional, construction and environmental requirements of the proposed model; identify relevant input and output parameters; identify sources of available data; and identify the modelling technique to be used to develop the model. Knowledge gained from Section 2 will be consolidated in the following chapter of this thesis which will focus on model design.

SECTION 3: DEVELOPMENT AND EVALUATION OF A MODEL FOR CRIME MANAGEMENT

Chapter 7. Model Design



Research objectives addressed in this chapter:

RO_1: Identify the functional, construction and environmental requirements of an effective model.

RO_2: Identify relevant input and output parameters.

RO_3: Identify and characterise available data sources.

RO_4: Identify the modelling technique to be used to develop the model.

RO_5: Develop the model.

RO_6: Demonstrate the application of the model.

RO_7: Evaluate the efficacy of the model.

RO_8: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

Figure 7-1: Research objectives and design science research activity addressed in this chapter.

7.1 Introduction

In this study, the Design Science Research (DSR) process is followed to develop, demonstrate and evaluate a prototype model for crime management in smart South African cities. The DSR methodology and its anticipated application in this study (Figure 7-1) was outlined in Chapter 2. Section 2 of this thesis (Chapters 3 through 6), provided the supporting literature necessary to formulate the research problem, outline and define the requirements of a potential solution, and identify potential design interventions.

The research problem, together with the proposed solution and anticipated design interventions, were briefly introduced and outlined in Chapter 1 (Sections 1.3-1.5). The purpose of this chapter (Figure 7-1), is to refine the model requirements and associated design interventions outlined in Chapter 1, based on knowledge gained in Section 2. The main purpose of this chapter is to describe the design of the prototype model, and thereby address $RO_{1\text{-}4}$ (Figure 7-1).

Section 7.2 of this chapter addresses RO_1 by consolidating the requirements of an effective problem solution identified in Chapters 3 and 5. An overview of key design elements is then described in Section 7.3, based on recommendations made in Chapters 4 to 6. Design elements are further explicated in Sections 7.4 to 7.8. Specifically, RO_2 is addressed by identifying the choice of input and target features used to develop the model (Section 7.5); RO_3 is addressed by summarising and characterising the data sources used (Section 7.6); and RO_4 is addressed by identifying the modelling technique used to develop the model (Section 7.8).

Due to the limited accessibility of South African data at the time of this study, readily accessible open data for New York City was used to develop and demonstrate the prototype model. The rationale behind the choice of city is given in Section 7.4, while an explanation of the choice of spatial and temporal units of analysis is given in Section 7.7.

7.2 Solution requirements

Activity 2 in the DSR process aims to identify the requirements of a solution to the main research problem (Figure 7-1). The purpose of solution requirements is to guide the design process, and to be used as a framework for assessing the efficacy of the final DSR artefact (Section 2.3.5.2) (Peffers et al., 2008).

The research problem addressed in this study is that South African cities are not effectively integrating and utilising available data sources for smart city planning and management (Section 1.2). In Chapter 3, it was shown that smart city data can most effectively be used within an integrated adaptive decision-making framework, supported by a comprehensive set of Key Performance Indicators (KPIs) and tools for quantitative scenario analysis (Section 3.8). To this end, the aim of this study is to develop a prototype model that incorporates all stakeholder and sustainability considerations, and serves as a reliable tool for systems-level scenario analysis. Thereby, improving the efficiency of decisions, and minimising unexpected externalities (Section 3.4).

As proof of concept, the study focuses on only one aspect of smart cities, namely crime management. Subsequently, the main objective of this study is to develop and evaluate a predictive model for crime management in smart cities that effectively integrates data from traditionally isolated management silos (Section 1.5). Model requirements identified in Section 3.8, were refined within the context of crime management in Section 5.4. The requirements identified in Section 3.8 and Section 5.4 are consolidated below:

- *Requirement 1: KPI framework* – The solution should incorporate a set of KPIs, selected in such a way as to quantitatively represent all stakeholder and sustainability considerations. By so doing, the KPI framework will explicate the goals of all involved, and will simultaneously provide knowledge of the status quo.

- *Requirement 2: Predictive model* – The solution should incorporate a predictive model that can be used as a reliable tool for systems-level scenario analysis:

    - *Requirement 2.1:* The model should take as input the KPIs identified in Requirement 1, and should predict the relative influence of selected KPIs and combinations thereof on crime. By so doing, the model will predict the collective impact of stakeholder decisions (represented as a set of changing KPIs) on crime. In effect, in order to truly integrate stakeholder and sustainability concerns, the KPI framework prescribed in Requirement 1 is comparable to a set of state variables used to model a dynamic system;

    - *Requirement 2.2:* The model should be able to incorporate known or unknown complexities and inter-dependencies between variables; and

    - *Requirement 2.3:* The predictive model should be reliable. Predictions should be accurate and precise, and a measure of prediction uncertainty should be included with the results.

- *Requirement 3: Data availability and accessibility:* Reliable and available data is fundamental to the success of any model.

    - *Requirement 3.1:* It is intended that the developed model will complement the existing IDP process in South Africa (Section 3.5.1). As such, relevant data needs to be available and accessible for South African cities; and

    - *Requirement 3.2:* The model will be implemented at the strategic planning level. Available data therefore needs to be applicable at corresponding spatial and temporal scales.

7.3 Overview of key design elements

Based on the state-of-the-art in smart city solutions reviewed in Chapter 4 (Section 4.6), and current predictive policing practices identified in Chapter 5 (Section 5.5), the following design elements were proposed to meet the requirements identified in Section 7.2:

- *Design element 1: KPI framework* – Requirement 1 requires that the solution incorporate a set of KPIs that quantitatively represent all stakeholder and sustainability considerations. The emerging smart city conceptual models and KPI frameworks discussed in Section 4.4 delineate key smart city components and their desired states, and therefore provide a potential design solution to Requirement 1. It is proposed that a subset of the City Protocol Society's *City Anatomy Indicators* (Section 4.4.2) be used as input to the prototype model; and that the commonly employed predictors of crime summarised in Section 5.3.4 be used as a guide when selecting KPIs from the set of CPA indicators.

- *Design element 2: Predictive model* – Requirement 2 requires that the solution incorporate a predictive model that can be used as a reliable tool for systems-level scenario analysis. Specifically, the model should take as input the KPIs identified in Requirement 1, and should predict the relative influence of selected KPIs on crime. The model should also be able to incorporate known or unknown complexities and inter-dependencies between variables. In addition, predictions should be accurate and precise, and a measure of prediction uncertainty should be provided. It is proposed that a combined approach employing Bayesian Neural Networks (Chapter 6) and sensitivity analysis (Section 6.5.1) be used to predict the relative influence of input features on crime. Bayesian Neural Networks will be effective in handling the anticipated complexity of city interactions, and will provide a measure of model reliability.

- *Design element 3: Data* – Requirement 3 states that reliable and available data is fundamental to the success of any model. Due to the limited accessibility of South African data at the time of this study (Section 4.4.3), readily accessible open data will be used to develop and demonstrate the prototype model.

7.4 Choice of city: New York City

Due to the limited implementation of open data portals in South Africa at the time of this study (Section 4.4.3), readily accessible open data for New York City (NYC) (Section 7.6) was used to develop and demonstrate the prototype model. The United States is a global leader in open data (Section 4.4.3); and in 2017, NYC was a leader in the US City Open Data Census (Open Data Census., 2017) rankings.

7.5 Choice of indicators

The indicators used in this study are listed in Table 7-3. As specified in Section 7.3, indicators were selected from the City Protocol Society's *City Anatomy Indicators* (Section 4.4.2), guided by the commonly employed predictors of crime summarised in Section 5.3.4. The availability of data (at the required spatial unit, see Section 7.7) was also a factor, limiting the selection of indicators (Section 7.6).

For the most part, the indicators listed in Table 7-3 are not exact implementations of CPA indicators (City Protocol Society., 2015a), but rather they have been adapted according to closely related crime predictors and available data. Out of the 22 indicators used in this study, only four indicators are not associated with a related CPA indicator. In these instances, new indictors have been created due to the availability of relevant data for which no CPA indicator exists. Specifically, these four indicators relate to the prevalence of single mothers (ID 11), child abuse (ID 12), drug crimes (ID 16) and graffiti (ID 37).

For each indicator, a description of the indicator used in this study, together with the associated CPA indicator, is given in Table 7-3. In addition, it is indicated whether a related indicator currently forms part of the NYC KPI framework, reported annually in the Mayor's Management Report (de Blasio *et al.*, 2018) and the NYC Social Indicators Report (MOO., 2018).

In order to visualise the diversity of domains included in the development of the prototype model, the associated CPA domain and NYC agency are specified in Table 7-1 and Table 7-2 for each indicator, respectively. The anticipated outcome of model development is to include a diverse range of stakeholders in the decision-making process, thereby fostering synergistic solutions that are often overlooked when solutions are solved within sectoral silos. The chosen indicators are briefly explained in the sections that follow.

7.5.1 Indicators related to structure

7.5.1.1 Environment

CPA indicators related to the environment (City Protocol Society., 2015a) measure the amount of permeable surfaces (required to filter out water pollutants and recharge the water table), environmental hazards, pollution, weather related indicators, and species diversity. Indicator ID 38 (Table 7-3) incorporates a measure of air pollution by measuring levels of fine particulate matter in the air (see Section 5.3.4). The effects of temperature (Indicator: ID 9, Table 7-3) were also considered. However, initial data exploration showed insufficient variability in the temperature over the reporting period. For this reason, temperature was not included.

7.5.1.2 Infrastructure

CPA indicators related to infrastructure (City Protocol Society., 2015a) focus on digital connectivity, storm water management, transportation, sanitation, hazardous waste generation and recycling, water security and service interruption, energy security and service interruption, renewable energy sources, and biodiversity. In this study, attention was given to the effect of electricity interruption on crime. It is often theorised that crime increases as the number of street lights out increase (Pease, 1999; Welsh and Farrington, 2008). For this reason, Indicator: ID 22 (Table 7-3) was included, and measures the average number of street lights out per day per unit area using 311 call-centre request data (Data ID 12, Table 7-8).

7.5.1.3 Built domain

CPA indicators related to the built domain (City Protocol Society., 2015a) focus on building age, land use practices, accessibility of public space, densification, urban green, and informal settlements. In this study, measures of land use (Indicator: ID 24, Table 7-3) and pedestrian traffic (Indicator: ID 36, Table 7-3) were included. The land use codes used in this study are listed in Table 7-4. A pedestrian volume index was calculated using biannual pedestrian counts at 114 locations across the city (Data ID 19, Table 7-8) according to Equation 7-1:

$$Pedestrian\ volume\ index = \frac{Average\ count\ in\ PUMA}{Population\ in\ PUMA} \times 100{,}000 \qquad \text{Equation 7-1}$$

Table 7-1: Indicators per CPA domain. *Indicator IDs are specified in Table 7-3.

CPA Domain	Indicator ID(s)*
Built Domain	24; 36
Citizens	11-13; 26
Economy	2-5
Environment	9; 38
Functions	1; 7-8; 15-16; 20-21; 33-34; 37
Government	14
Infrastructure	22

Table 7-2: Indicators per NYC agency. *Indicator IDs are specified in Table 7-3.

Agency	Indicator ID(s)*
Administration for Children's Services (ACS)	12
Civilian Complaint Review Board (CCRB)	14
Department of City Planning (DCP)	24; 26
Department of Education (DOE)	7-8
Department of Environmental Protection (DEP)	38
Department of Finance (DOF)	2
Department of Health and Mental Hygiene (DOHMH)	15-16; 38
Department of Homeless Services (DHS)	20
Department of Transportation (DOT)	22; 36
Department of Youth and Community Development (DYCD)	11; 13
Economic Development Corporation (EDC)	37
Human Resources Administration (HRA)	15
Mayor's Office for Economic Opportunity (MOEO)	3-5
Mayor's Office of Climate Policy and Programs (MOCPP)	9
New York City Housing Authority (NYCHA)	21
New York Police Department (NYPD)	16; 33-34; 37
Department of Sanitation (DSNY)	37
Office of Citywide Event Coordination and Management (CECM)	1

Table 7-3: Indicators used in this study.

ID	CPA Domain	Agency	Indicator(s) used	Related CPA indicators	Current NYC Reporting Scale	Citywide	Borough	CD/PUMA/ Precinct	Coordinates	Variable name
1	Functions	CECM	Events per 100k population	Performing arts shows per 1000 population (Functions SI)	-	14	14	14	14	events
2	Economy	DOF	Assessed (commercial/residential) property values relative to citywide average assessed (commercial/residential) property values	Assessed value of commercial and industrial properties as a percentage of total assessed value of all properties (Economy CI)	-	15	15	15	15	V1 (residential), V5 (commercial)
3	Economy	MOEO	Unemployment rate (%)	City's unemployment rate (Functions CI); Percentage of employed population (Economy CI)	Borough	1;2;3	1;2;3	4;5	-	unemployment
4	Economy	MOEO	Theil's T inequality index (within PUMAs/between PUMAs)	Gini Index (Economy CI)	Citywide	1;2;3	1;2;3	4;5	-	ineqT1r (within), ineqT2r (between)
7	Functions	DOE	Percentage of population without high school diploma	Primary education student/teacher ratio (Functions CI); Students per teacher in mandatory education (Functions SI)	Borough	1;2;3	1;2;3	4;5	-	noHigh
8	Functions	DOE	Percentage of population with higher education degrees	Number of higher education degrees per 100k population (Functions SI)	Citywide	1;2;3	1;2;3	4;5	-	degree
9	Environment	MOCPP	Number of hot days per year	Global solar irradiance yearly average (W/m2) (Environment SI); Average wind speed (km/h) (Environment SI)	-	13	13	-	-	-
11	Citizens	DYCD	Percentage single female householders	-	-	1;2;3	1;2;3	4;5	-	female

Table 7 3: Indicators used in this study. (Continued)

ID	CPA Domain	Agency	Indicator(s) used	Related CPA indicators	Current NYC Reporting Scale	Citywide	Borough	CD/PUMA/ Precinct	Coordinates	Variable name
12	Citizens	ACS	Number of credible abuse/neglect investigations per 100k population	-	Citywide	16	16	16	-	abuse
13	Citizens	DYCD	Fertility rate per 1000 women aged 15-44	Fertility rate: Annual number of live births per 1000 women aged 15-49 years (Citizens CI)	-	36	36	36	-	fertility
14	Government	CCRB	Total civilian complaints against uniformed members of the New York City Police Department per 100k population	Number of convictions for corruption and/or bribery by city officials per 100k population (Government SI)	Citywide	29	29	29	-	integrity
15	Functions	HRA; DOHMH	Adult New Yorkers without health insurance (%)	Public expenditure on health per capita (Functions CI)	Citywide	1;2;3	1;2;3	4;5	-	insurance
16	Functions	DOHMH; NYPD	Drug crimes per 100k population	-	Citywide	10; 12	10; 12	10; 12	10; 12	drugs
20	Functions	DHS	Total number of 311 requests related to homeless encampments and panhandling per 100k population	Number of homeless per 100k population (Functions SI)	Citywide	12; 33; 34; 35; 42; 43	12; 33; 35; 43	12; 33	12	homeless
21	Functions	NYCHA	Percentage social housing	Percentage of social housing (Functions CI)	Citywide	41	41	41	-	socialHousing
22	Infrastructure	DOT	Average number of street lights out per day per unit area	Average length of electrical interruptions (in hours) (Infrastructure SI)	Citywide	12	12	12	12	SL

Table 7 3: Indicators used in this study. (Continued)

ID	CPA Domain	Agency	Indicator(s) used	Related CPA indicators	Current NYC Reporting Scale	Citywide	Borough	CD/PUMA/ Precinct	Coordinates	Variable name
24	Built Domain	DCP	Percentage land use type	Green area (hectares) per 100k population (Built Domain CI); Neighborhood Homogeneity (Built Domain CI); Industrial availability: Space density (Built Domain SI); Percentage parking places off the road (Infrastructure CI); Areal size of mix-use developments as a percentage of city total built area (Built Domain CI)	-	15	15	15	15	P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11
26	Citizens	DCP	Theil's L diversity index	Cultural diversity (Citizens CI)	-	1;2;3	1;2;3	4;5	-	diversity
33	Functions	NYPD	Larceny (street/residential/commercial) per 100k population	Crimes against property per 100k population (Functions SI)	Citywide	10; 11	10; 11	10	10	larStreet, larCommercial, larResidence
34	Functions	NYPD	Robbery (street/residential/commercial) per 100k population; Assault (street/residential) per 100k population	Violent crime rate per 100k population (Functions SI); Number of homicides per 100k population (Functions CI)	Borough	10; 11	10; 11	10	10	robStreet, robCommercial, robResidence, assStreet, assResidence
36	Built Domain	DOT	Pedestrian volume index	Surface of pedestrian priority areas and streets / Total street area (Built Domain CI)	Citywide	19	19	-	19	pedIndex
37	Functions	NYPD; DSNY; EDC	Graffiti reports per 100k population	-	Citywide	10; 12	10; 12	10; 12	10; 12	graffiti
38	Environment	DOHMH; DEP	Fine particulate matter (PM2.5) concentration	Fine particulate matter (PM2.5) concentration (Environment CI)	CD	44	44	44	-	PM

Table 7-4: Land use codes. Source: Adapted from PLUTO data dictionary (Data ID 15, Table 7-8).

Codes	Decodes
P1	One &Two Family Buildings
P2	Multi-Family Walk-Up Buildings
P3	Multi-Family Elevator Buildings
P4	Mixed Residential & Commercial Buildings
P5	Commercial & Office Buildings
P6	Industrial & Manufacturing
P7	Transportation & Utility
P8	Public Facilities & Institutions
P9	Open Space & Outdoor Recreation
P10	Parking Facilities
P11	Vacant Land

7.5.2 Indicators related to interactions

7.5.2.1 Functions

CPA indicators related to functions (City Protocol Society., 2015a) encompass a wide range of indicators. These can be sub-divided into the following categories: built domain, transportation, health care, education, economy, public safety and culture.

Relating to the built domain, attention focused on the percentage social housing (Indicator: ID 21, Table 7-3). In terms of health care, the percentage of adult New Yorkers without health insurance was considered (Indicator: ID 15, Table 7-3). Public safety indicators included measures of robbery and assault (Indicator: ID 34, Table 7-3), larceny (Indicator: ID 33, Table 7-3), drugs (Indicator: ID 16, Table 7-3) and graffiti (Indicator: ID 37, Table 7-3). All crime data was derived from NYPD complaint reports records (Data ID 10, Table 7-8).

In terms of culture, the number of events per 100k population was considered (Indicator: ID 1, Table 7-3). Education related indicators included the percentage of the population with higher education degrees (Indicator: ID 8, Table 7-3), and the percentage of the population without high school diplomas (Indicator: ID 7, Table 7-3). Homelessness (McCarthy and Hagan, 1991; Fischer et al., 2008) (Indicator: ID 20, Table 7-3) was also included, and was derived from 311 requests related to homeless encampments and panhandling (Data ID 12, Table 7-8).

7.5.2.2 Economy

CPA indicators related to the economy (City Protocol Society., 2015a) focus on indicators such as productivity, unemployment, inequality, the number of businesses, property values, debt, revenue, etc. Economy indicators in this study include measures of unemployment (Indicator: ID 3, Table 7-3), income inequality (Indicator: ID 4, Table 7-3), and property values (Indicator: ID 2, Table 7-3).

The most widely used single measure of inequality is the Gini coefficient (World Bank., 2014). The Gini coefficient is the suggested CPA measure of inequality (City Protocol Society., 2015a). However, the Gini coefficient is not decomposable (World Bank., 2014). That is, the Gini coefficient cannot be broken down by partitions – such as regions or population groups – in an additive way. A widely used decomposable metric for inequality is the Theil index. Theil's T index, as explained in the report by the World Bank. (2014), was used to measure regional inequality partitioned into PUMAs. Thereby, developing a measure of poverty within (*ineqT1r*), as well as across PUMAs (*ineqT2r*).

7.5.3 Indicators related to society

7.5.3.1 Citizens

CPA indicators related to citizens (City Protocol Society., 2015a) focus on measuring demographic features such as population density, fertility rate and cultural diversity; as well as measuring levels of entrepreneurship and community engagement in public participation processes. CPA indicators used in this study include cultural diversity (Indicator ID 26, Table 7-3) and fertility rate (Indicator ID 13, Table 7-3). Additional indicators, not forming part of the CPA framework, include measures of child abuse and neglect (Indicator ID 12, Table 7-3), and female-headed households (Indicator ID 11, Table 7-3). Cultural diversity was calculated according to Theil's L diversity index, as explained in the report by the World Bank. (2014).

7.5.3.2 Government

CPA indicators related to government (City Protocol Society., 2015a) focus on measuring the effectiveness of the legal system, corruption, gender equality, and levels of transparency and open governance in a city. Indicator ID 14 (Table 7-3) incorporates a measure of police corruption and misconduct by measuring the total number of civilian complaints against uniformed members of the New York City Police Department per 100k population.

7.6 Summary of data sources

The datasets used in this study are listed in Table 7-8. Datasets were sourced from the URLs listed in Table 7-6 over a six-month timeframe ranging from July to December 2018. For each dataset, the following is specified: a Data ID for reference purposes; the source agency (Table 7-5); a brief description of the data; the data type; the source URL (Table 7-6); the spatial (Table 7-7) and temporal scales at which the data is reported; and the start and end years of reporting. Here, "data type" refers to the accessibility of the data: whether it is presented as a readily accessible data table or shapefile, or if it is presented in a form that requires pre-processing, such as a poorly structured Excel file or pdf report.

Table 7-5: List of source agencies.

ID	Agency Name	Data ID
1	311	12
2	Administration for Children's Services (ACS)	16-17
3	Civilian Complaint Review Board (CCRB)	29
4	Department of City Planning (DCP)	15; 21-25
5	Department of Correction (DOC)	37-38
6	Department of Health and Mental Hygiene (DOHMH)	36; 44
7	Department of Homeless Services (DHS)	33-35; 43
8	Department of Information Technology & Telecommunications (DoITT)	26
9	Department of Transportation (DOT)	19
10	Mayor's Office of Operations (OPS)	40; 42
11	National Centers for Environmental Information: National Oceanic and Atmospheric Administration	13
12	New York City Housing Authority (NYCHA)	41
13	New York State Division of Criminal Justice Services	11
14	NYPD	10
15	Office of Citywide Event Coordination and Management (CECM)	14
16	United States Census Bureau	1-9; 27-28

Table 7-6: List of source URLs.

ID	Website Name / Description	URL
1	American FactFinder	factfinder.census.gov
2	United States Census Bureau data	www.census.gov/geo/maps-data/data/tiger-cart-boundary.html
3	National Oceanic and Atmospheric Administration data	www.ncdc.noaa.gov/cdo-web/datatools
4	New York State open data portal	data.ny.gov
5	NYC Open Data	opendata.cityofnewyork.us
6	NYC Health	www1.nyc.gov/site/doh/index.page
7	NYC Planning	www1.nyc.gov/site/planning/index.page

Table 7-7: List of spatial scales.

ID	Spatial Scales	Reference
1	Citywide	-
2	Borough / County	Section 8.2.1.1
3	Public Use Microdata Area	Section 8.2.2.2
4	Police Precinct	Section 8.2.1.4
5	Community District	Section 8.2.1.2
6	Neighbourhood Tabulation Area	Section 8.2.2.1
7	ZIP Code	Section 8.2.1.3
8	Street Address	-
9	Coordinates	-
10	Other: School District; Council District; Health Area; Tax Lot; Census Tract	-

7.7 Choice of spatial and temporal scales

As mentioned in Section 7.2, it is intended that the prototype model will complement the existing IDP process in South Africa (Section 3.5.1). As such, the model will be implemented at the strategic planning level. Strategic city planning most often makes use of citywide annual trends in KPIs and situational indicators to identify challenges and inform decisions (de Blasio *et al.*, 2018; MOO., 2018; NMBM., 2018b). However, in order to resolve the spatial pattern of crime and its associated predictors across a city, a smaller spatial unit of analysis was sought.

In Section 8.3.2 it was explained that the PUMA was the statistical unit of choice for this study, as it is the smallest statistical geographic unit with sufficiently (temporally) resolved data to meet the needs of this study. Furthermore, the PUMA (or the associated community district) was one of the most commonly used spatial unit of reporting in the NYC open datasets (Table 7-8). For these reasons, the PUMA was chosen as the spatial unit of analysis in this study. Temporally, data was aggregated annually.

Table 7-8: Summary of datasets used in this study.

Data ID	Source Agency	Data Description	Type	URL	Spatial Scale	Temporal Scale	Start Year	End Year
1	16	American Community Survey	Files	7	2	1-Year Estimates	2008	2016
2	16	American Community Survey	Files	7	2	3-Year Estimates	2007	2013
3	16	American Community Survey	Files	7	2	5-Year Estimates	2010	2016
4	16	American Community Survey	Files	7	3	3-Year Estimates	2007	2013
5	16	American Community Survey	Files	7	3	5-Year Estimates	2010	2016
6	16	American Community Survey	Files	7	6	5-Year Estimates	2010	2016
7	16	Decennial Census - Census 2010: Total Population and Persons Per Acre, 2000-2010	File	7	2	Year	2000	2010
8	16	Decennial Census - Census 2010: Total Population and Persons Per Acre, 2000-2010	File	7	6	Year	2000	2010
9	16	PEP: Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2017	Table	1	2	Year	2010	2017
10	14	NYPD Complaint Data Historic	Table	5	2; 4; 9	Date&Time	2006	2017
11	13	Index Crimes by County and Agency: Beginning 1990	Table	4	2	Year	1990	2017
12	1	311 Service Requests	Tables	5	2; 5; 7-9	Date&Time	2004	Present
13	11	Local Climatological Data (LCD)	Tables	3	2	Date&Time	2006	2017
14	15	NYC Permitted Event Information - Historical	Table	5	2; 4-5; 8	Date&Time	2008	Present
15	4	Archived Primary Land Use Tax Lot Output (PLUTO)	Tables	7	2; 4-5; 7-10	Year	2006	Present
16	2	Abuse/Neglect by Community District (CD)	Files	5	2; 5	Year	2010	2017
17	2	Detention and Placement Demographic reports	Files	5	7	Year	2015	2017
19	9	Bi-annual pedestrian counts	Shapefile	5	2; 8-9	Bi-annual	2007	2017
21	4	Borough Boundaries	Shapefile	5	-	As needed	-	-
22	4	Community Districts	Shapefile	5	-	As needed	-	-
23	4	Neighborhood Tabulation Areas	Shapefile	5	-	As needed	-	-
24	4	Police Precincts	Shapefile	5	-	As needed	-	-
25	4	Public Use Microdata Areas (PUMA)	Shapefile	5	-	As needed	-	-
26	8	Zip Code Boundaries	Shapefile	5	-	As needed	-	-
27	16	County	Shapefile	2	-	As needed	-	-
28	16	Public Use Microdata Areas (PUMA)	Shapefile	2	-	As needed	-	-
29	3	Where Incidents That Led To A Complaint Took Place By Precinct	Files	5	4	Year	2005	2009
33	7	Individual Census by Borough, Community District and Facility Type	Table	5	2; 5	Month	2018 (July)	Present

Table 7 9: Summary of datasets used in this study. (Continued)

Data ID	Source Agency	Data Description	Type	URL	Spatial Scale	Temporal Scale	Start Year	End Year
34	7	DHS Daily Report	Table	5	1	Day	2013	Present
35	7	Directory Of Homeless Population By Year	Table	5	2	Year	2009	2012
36	6	Fertility rate per 1000 women aged 15-44	Files	6	5	Year	2006	2016
37	5	Inmate Admissions	Table	5	1	Date&Time	2014	Present
38	5	Daily Inmates In Custody	Table	5	1	Date&Time	2012	Present
40	10	Social Indicator Report Data	Table	5	2	Year	2000	2017
41	12	NYCHA Development Data Book	Table	5	2; 5; 8	As needed	-	-
42	10	MMR Agency Performance Indicators	Tables	5	1	Year	2003	2017
43	7	HOME-STAT Weekly Dashboard	Files	5	1; 2	Week	2018	Present
44	6	New York City Community Air Survey: Neighborhood Air Quality 2008 - 2016	Report	6	5	Year	2009	2016

7.8 Modelling technique

In this study, Bayesian Neural Networks will be used to develop a model for crime management in Smart cities. The Bayesian approach to Neural Networks is explained in Chapter 6 of this thesis. The model will be implemented using the Model Manager software package developed by (Sourmail, 2004). The implementation process is described in Section 6.4 in detail.

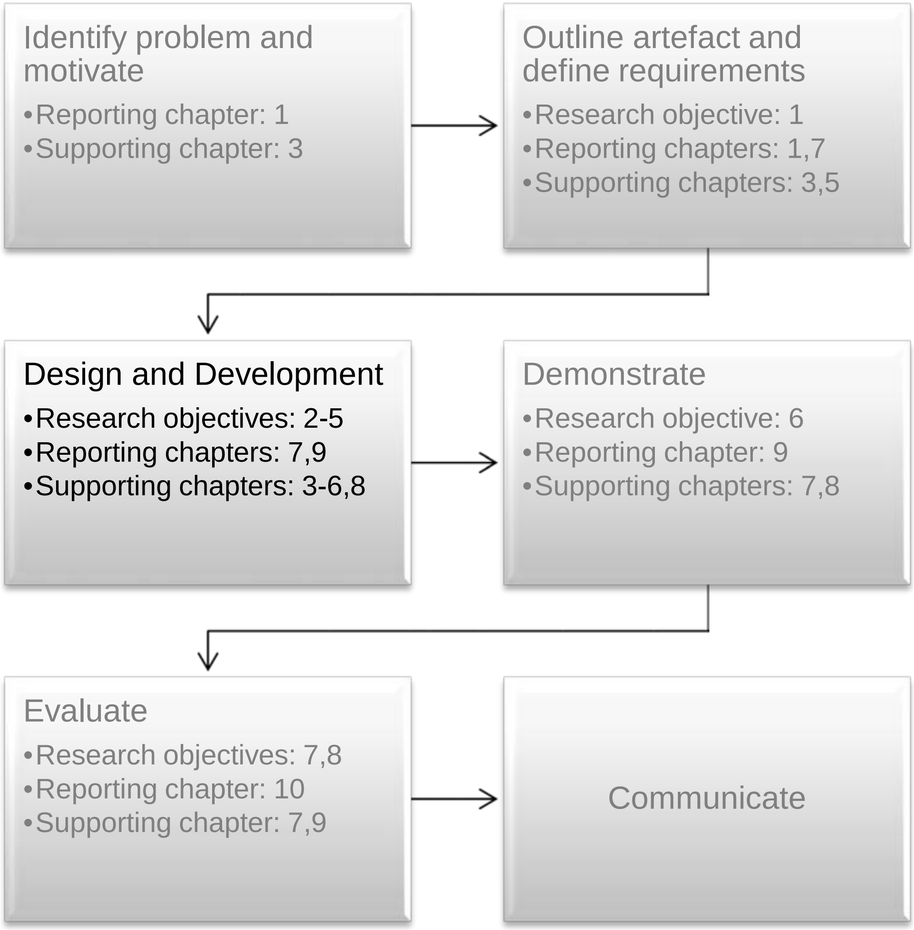
7.9 Summary

In this study, the DSR process is followed to develop, demonstrate and evaluate a prototype model for crime management in smart South African cities (Figure 7-1). The research problem, together with the proposed solution and anticipated design interventions, were briefly introduced and outlined in Chapter 1 (Sections 1.3-1.5). The purpose of this chapter, was to refine the model requirements and associated design interventions outlined in Chapter 1, based on knowledge gained in Section 2. The main purpose of this chapter was to describe the design of the prototype model, and thereby address $RO_{1\text{-}4}$ (Figure 7-1).

Section 7.2 of this chapter addressed RO_1 by consolidating the requirements of an effective problem solution identified in Chapters 3 and 5. An overview of key design elements was then described in Section 7.3, based on recommendations made in Chapters 4 to 6. Design elements were explicated in Sections 7.4 to 7.8.

Specifically, RO_2 was addressed by identifying the choice of input and target features used to develop the model (Section 7.5); RO_3 was addressed by summarising and characterising the data sources used (Section 7.6); and RO_4 was addressed by identifying the modelling technique used to develop the model (Section 7.8). The choice of city was explained in Section 7.4, while the chosen spatial and temporal units of analysis were identified in Section 7.7. In the following chapter, the data used in this study will be explored. This exploratory data analysis informed design decisions made in this chapter, and will provide context for model interpretation in Chapter 9.

Chapter 8. Exploratory Data Analysis



Research objectives addressed in this chapter:

RO$_1$: Identify the functional, construction and environmental requirements of an effective model.

RO$_2$: Identify relevant input and output parameters.

RO$_3$: Identify and characterise available data sources.

RO$_4$: Identify the modelling technique to be used to develop the model.

RO$_5$: Develop the model.

RO$_6$: Demonstrate the application of the model.

RO$_7$: Evaluate the efficacy of the model.

RO$_8$: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

Figure 8-1: Research objectives and design science research activity addressed in this chapter.

8.1 Introduction

In this study, New York City (NYC) open data was used to develop and demonstrate a prototype model for crime management in smart cities (Section 2.3.5.3). In this chapter, the available data will be explored. This exploratory data analysis provided context for the model design in Chapter 7 and the model interpretation in Chapter 9. This chapter, therefore, contributes towards addressing $RO_{5\text{-}6}$ (Figure 8-1).

An understanding of the relationships between commonly used geographic units is essential when combining data from different sources. The geographic units most often used in NYC open data are described and compared in Section 8.2. Section 8.3 provides an overview of available demographic data, and discusses the implication of sample size when selecting a spatial unit of analysis (Section 7.7).

In order to provide context for model interpretation, an overview of crime trends for the period 2006 to 2017 is provided in Section 8.4. This is followed by a description of perpetrator demographics over the same time period in Section 8.5.

8.2 Geographic areas

Depending on their source, the available datasets used in this study are most commonly reported with respect to one of the following administrative (Wikipedia, 2018a) or statistical (United States Census Bureau, 2018a) geographic units: city, borough, community district, Public Use Microdata Area, Neighbourhood Tabulation Area, ZIP code or police precinct. In this section, these units and the relationship between them are delineated.

8.2.1 Administrative geographic areas

The administrative divisions of New York City correspond to the various units of local government (e.g. counties and community boards) as well as to various single purpose units of government (e.g. education and public safety) (Wikipedia, 2018a).

In this section, the following administrative units will be described:

- Local government: Boroughs and community districts; and
- Single purpose units: ZIP codes and police precincts.

8.2.1.1 Boroughs

Data sources: ID 21 & 27 (Table 7-8)

New York City consists of five administrative divisions called boroughs, each corresponding to a county of the state of New York (Wikipedia, 2018a). The five boroughs, shown in Figure 8-2, are: The Bronx (Bronx County), Brooklyn (Kings County), Manhattan (New York County), Queens (Queens County) and Staten Island (Richmond County) (Wikipedia, 2018a).

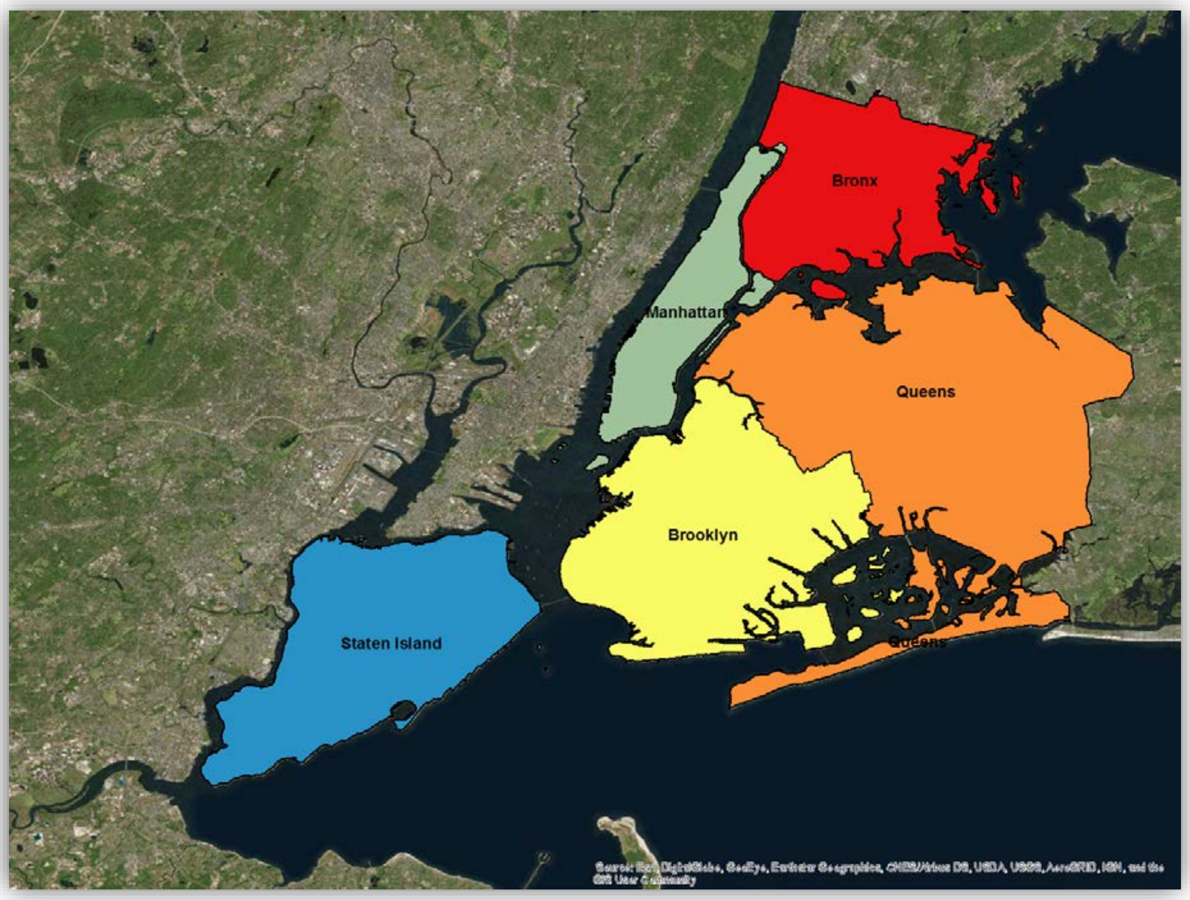


Figure 8-2: Boroughs of New York City. New York City consists of five county-level administrative divisions called boroughs, namely: The Bronx, Brooklyn, Manhattan, Queens and Staten Island. Source: Author's own construction. Software: Esri® ArcMap™.

The county is the primary administrative division of New York state, with the power and fiscal capacity to provide local government services such as health, education, public safety and social services (Wikipedia, 2018a). However, in New York City, borough presidents have minimal executive powers. “Executive functions in New York City are the responsibility of the Mayor of New York City, while legislative functions reside with the New York City Council. The borough presidents primarily act as spokesmen, advocates, and ceremonial leaders for their boroughs, have budgets from which they can allocate relatively modest sums of money to community organisations and projects, and appoint the members of the 59 largely advisory community boards in the city's various neighbourhoods” (Wikipedia, 2018b).

8.2.1.2 Community districts and neighbourhoods

Data source: ID 22 (Table 7-8)

Each borough is sub-divided into community districts (shown in Figure 8-3 and Figure 8-4). There are 59 community districts: twelve in Manhattan, twelve in the Bronx, eighteen in Brooklyn, fourteen in Queens, and three in Staten Island (Wikipedia, 2018c). Each community district is represented by a community board, which is appointed by the New York City government in an advisory capacity. Community boards are appointed to advise on land use and zoning matters, participate in the city budget process, and address service delivery in their district (Wikipedia, 2018c). Most community districts include a number of neighbourhoods. However, neighbourhood names and borders are not officially defined (Wikipedia, 2018f), and the neighbourhood geographic unit is rarely used when reporting data.

8.2.1.3 ZIP codes

Data source: ID 26 (Table 7-8)

Five-digit ZIP codes are used by the United States Postal Service to effectively deliver mail. ZIP is an acronym for Zone Improvement Plan (Wikipedia, 2018h). The first digit divides the United States into 10 large groups of states numbered from 0 to 9. Within these areas, each state is divided into an average of 10 smaller

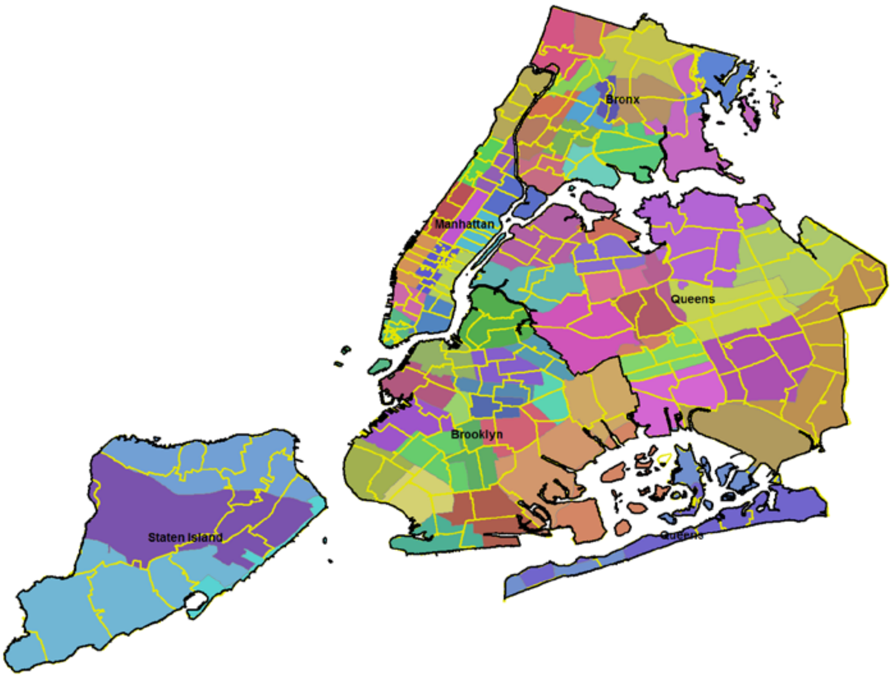


Figure 8-3: ZIP code areas (yellow boundaries) and community districts (shaded polygons) of New York City. Source: Author's own construction. Software: Esri® ArcMap™.

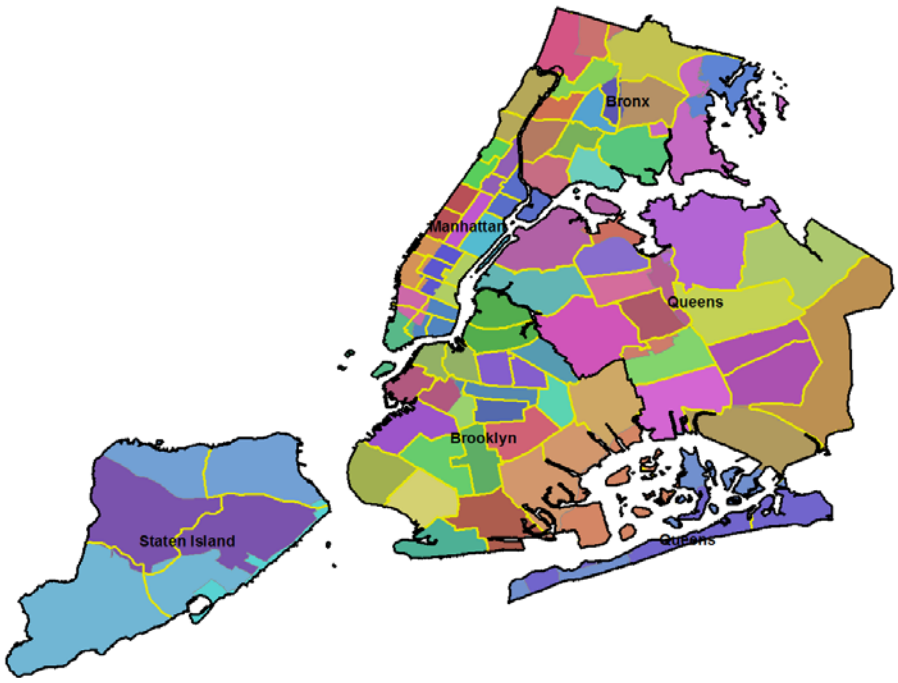


Figure 8-4: Police precincts (yellow boundaries) and community districts (shaded polygons) of New York City. Source: Author's own construction. Software: Esri® ArcMap™.

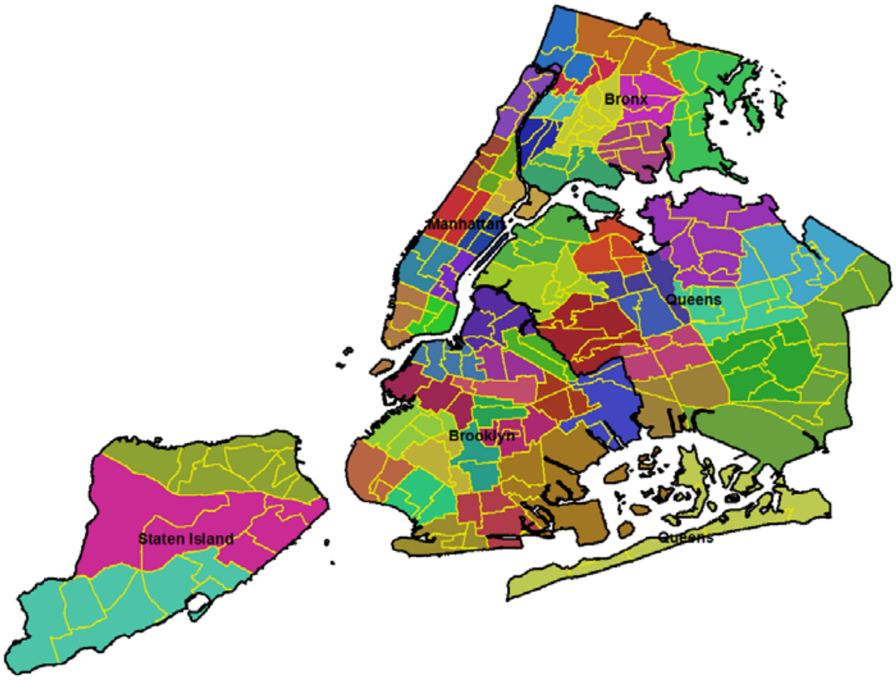


Figure 8-5: Neighbourhood Tabulation Areas (yellow boundaries) and Public Use Microdata Areas (shaded polygons) of New York City. Source: Author's own construction. Software: Esri® ArcMap™.

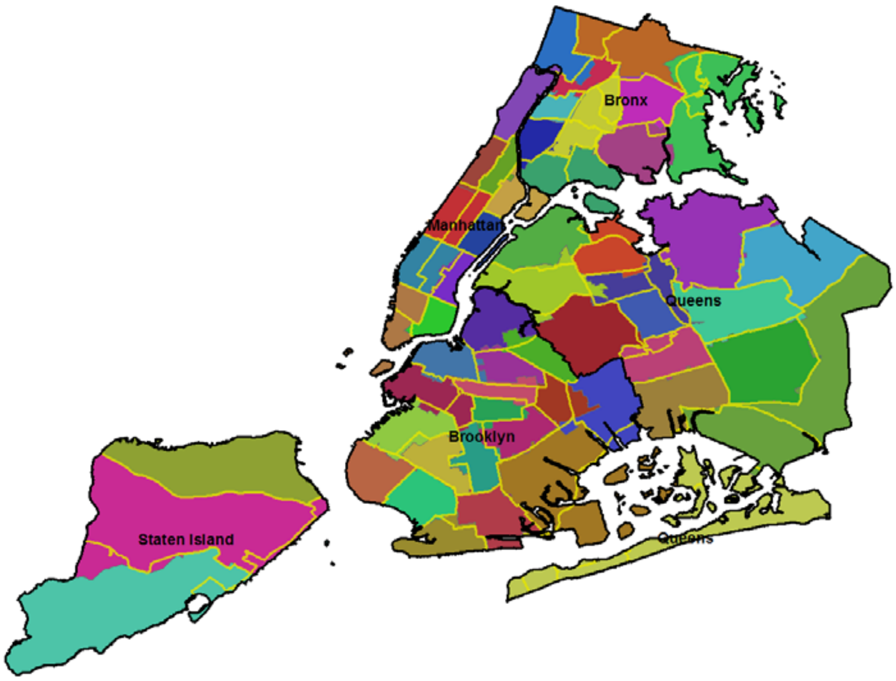


Figure 8-6: Community districts (yellow boundaries) and Public Use Microdata Areas (shaded polygons) of New York City. Source: Author's own construction. Software: Esri® ArcMap™.

geographical areas, identified by the second and third digits. These digits, in conjunction with the first digit, represent a sectional centre facility or a mail processing facility area. The fourth and fifth digits identify a post office, station, branch or local delivery area (ArcGIS, 2018; Wikipedia, 2018h). The areas corresponding to the ZIP codes of New York City are shown in Figure 8-3 with reference to the community districts of New York City. It is evident from Figure 8-3 that ZIP code boundaries do not coincide with community district boundaries.

8.2.1.4 Police Precincts

Data source: ID 24 (Table 7-8)

Police precincts delineate the areas patrolled by police officers. The 77 precincts of the New York Police Department (NYPD) are shown in Figure 8-4 with reference to the community districts of New York City. It is evident from Figure 8-4 that precinct boundaries largely coincide with community district boundaries, with the exception of those falling within Staten Island. In 2013, a new precinct (the 121st precinct) was added to Staten Island, increasing the number of precincts in this borough to four (Wikipedia, 2018g). As far as the author is aware, no historical shape files were available on-line at the time of this study that defined the NYPD precinct boundaries prior to 2018. Caution was therefore taken when interpreting data that were delineated according to precincts.

8.2.2 Statistical geographic areas

The primary purpose of statistical areas is to tabulate and present census data (United States Census Bureau, 2018a). Two commonly used statistical geographic units are Neighbourhood Tabulation Areas (NTAs) and Public Use Microdata Areas (PUMAs).

8.2.2.1 Neighbourhood Tabulation Areas (NTAs)

Data source: ID 23 (Table 7-8)

NTAs are aggregates of census tracts, and are subsets of PUMAs. The aggregation of NTAs into the 55 PUMAs of New York City is shown in Figure 8-5.

8.2.2.2 Public Use Microdata Areas (PUMAs)

Data sources: ID 25 & 28 (Table 7-8)

PUMAs are aggregates of census tracts that have a minimum population of 100,000 people (United States Census Bureau, 2018c). PUMAs approximate community districts, or combinations of community districts. This approximation is shown in Figure 8-6 for New York City. There are 59 community districts and only 55 PUMAs in New York City. Table 8-1 lists the PUMAs of New York City and their associated community districts (New York City Department of City Planning, 2010).

Table 8-1: List of Public Use Microdata Areas (PUMAs) and their associated community districts (New York City Department of City Planning, 2010).

PUMAs	Associated Community Districts
3701	NYC-Bronx Community District 8--Riverdale, Fieldston & Kingsbridge
3702	NYC-Bronx Community District 12--Wakefield, Williamsbridge & Woodlawn
3703	NYC-Bronx Community District 10--Co-op City, Pelham Bay & Schuylerville
3704	NYC-Bronx Community District 11--Pelham Parkway, Morris Park & Laconia
3705	NYC-Bronx Community District 3 & 6--Belmont, Crotona Park East & East Tremont
3706	NYC-Bronx Community District 7--Bedford Park, Fordham North & Norwood
3707	NYC-Bronx Community District 5--Morris Heights, Fordham South & Mount Hope
3708	NYC-Bronx Community District 4--Concourse, Highbridge & Mount Eden
3709	NYC-Bronx Community District 9--Castle Hill, Clason Point & Parkchester
3710	NYC-Bronx Community District 1 & 2--Hunts Point, Longwood & Melrose
3801	NYC-Manhattan Community District 12--Washington Heights, Inwood & Marble Hill
3802	NYC-Manhattan Community District 9--Hamilton Heights, Manhattanville & West Harlem
3803	NYC-Manhattan Community District 10--Central Harlem
3804	NYC-Manhattan Community District 11--East Harlem
3805	NYC-Manhattan Community District 8--Upper East Side
3806	NYC-Manhattan Community District 7--Upper West Side & West Side
3807	NYC-Manhattan Community District 4 & 5--Chelsea, Clinton & Midtown Business District
3808	NYC-Manhattan Community District 6--Murray Hill, Gramercy & Stuyvesant Town
3809	NYC-Manhattan Community District 3--Chinatown & Lower East Side
3810	NYC-Manhattan Community District 1 & 2--Battery Park City, Greenwich Village & Soho

3901	NYC-Staten Island Community District 3--Tottenville, Great Kills & Annadale
3902	NYC-Staten Island Community District 2--New Springville & South Beach
3903	NYC-Staten Island Community District 1--Port Richmond, Stapleton & Mariner's Harbor
4001	NYC-Brooklyn Community District 1--Greenpoint & Williamsburg
4002	NYC-Brooklyn Community District 4--Bushwick
4003	NYC-Brooklyn Community District 3--Bedford-Stuyvesant
4004	NYC-Brooklyn Community District 2--Brooklyn Heights & Fort Greene
4005	NYC-Brooklyn Community District 6--Park Slope, Carroll Gardens & Red Hook
4006	NYC-Brooklyn Community District 8--Crown Heights North & Prospect Heights
4007	NYC-Brooklyn Community District 16--Brownsville & Ocean Hill
4008	NYC-Brooklyn Community District 5--East New York & Starrett City
4009	NYC-Brooklyn Community District 18--Canarsie & Flatlands
4010	NYC-Brooklyn Community District 17--East Flatbush, Farragut & Rugby
4011	NYC-Brooklyn Community District 9--Crown Heights South, Prospect Lefferts & Wingate
4012	NYC-Brooklyn Community District 7--Sunset Park & Windsor Terrace
4013	NYC-Brooklyn Community District 10--Bay Ridge & Dyker Heights
4014	NYC-Brooklyn Community District 12--Borough Park, Kensington & Ocean Parkway
4015	NYC-Brooklyn Community District 14--Flatbush & Midwood
4016	NYC-Brooklyn Community District 15--Sheepshead Bay, Gerritsen Beach & Homecrest
4017	NYC-Brooklyn Community District 11--Bensonhurst & Bath Beach
4018	NYC-Brooklyn Community District 13--Brighton Beach & Coney Island
4101	NYC-Queens Community District 1--Astoria & Long Island City
4102	NYC-Queens Community District 3--Jackson Heights & North Corona
4103	NYC-Queens Community District 7--Flushing, Murray Hill & Whitestone
4104	NYC-Queens Community District 11--Bayside, Douglaston & Little Neck
4105	NYC-Queens Community District 13--Queens Village, Cambria Heights & Rosedale
4106	NYC-Queens Community District 8--Briarwood, Fresh Meadows & Hillcrest
4107	NYC-Queens Community District 4--Elmhurst & South Corona
4108	NYC-Queens Community District 6--Forest Hills & Rego Park
4109	NYC-Queens Community District 2--Sunnyside & Woodside
4110	NYC-Queens Community District 5--Ridgewood, Glendale & Middle Village
4111	NYC-Queens Community District 9--Richmond Hill & Woodhaven
4112	NYC-Queens Community District 12--Jamaica, Hollis & St. Albans
4113	NYC-Queens Community District 10--Howard Beach & Ozone Park
4114	NYC-Queens Community District 14--Far Rockaway, Breezy Point & Broad Channel

8.2.3 Conclusions

With the exception of ZIP codes, there is a fair amount of correlation between the administrative and statistical geographic areas explored in this section. This is illustrated in Figure 8-7, which shows the respective administrative and statistical geographic hierarchies studied. PUMAs closely approximate community districts, as do police precincts (with the exception of Staten Island).

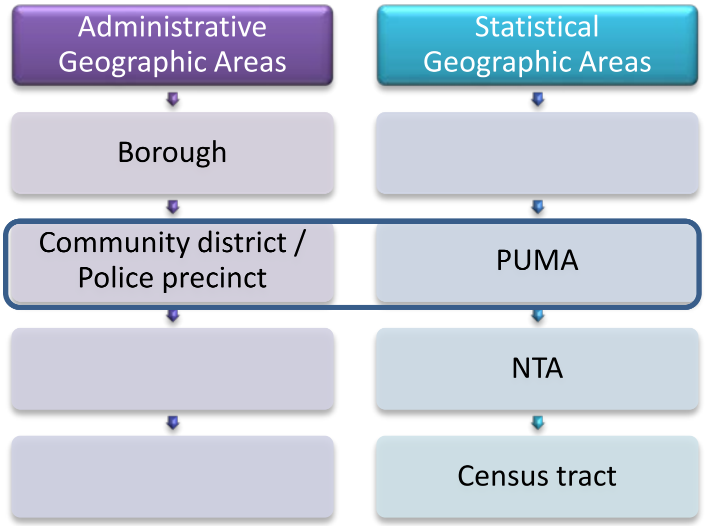


Figure 8-7: Administrative and statistical geographic hierarchies explored in this section. PUMAs closely approximate community districts, as do police precincts (with the exception of Staten Island). Source: Author's own construction.

8.3 Demographic data

8.3.1 Source of demographic data

Data sources: ID 1-8 (Table 7-8)

Demographic data for New York City is available from the United States Census Bureau. Two key sources of demographic data are the United States decennial census and the American Community Survey (ACS) (United States Census Bureau, 2018b). As the name indicates, comprehensive surveys of the US population are carried out every 10 years by means of the decennial census, with the last major census being carried out in 2010.

Since 2005, in order to provide detailed information for local officials, businesses and citizens more frequently, data that were historically collected only once every 10 years by the decennial census have been collected monthly (and released annually) through the ACS (United States Census Bureau, 2018b). The ACS therefore provides an unprecedented ability to annually monitor social, economic, housing and demographic trends in local communities (United States Census Bureau., 2018). Pre-packaged datasets for NYC based on these surveys are available for download on the NYC Planning website (see data sources: ID 1-8, Table 7-8).

8.3.2 Sample size and the significance of spatial scale

The ACS has an annual sample size of about 3.5 million addresses (United States Census Bureau., 2018). Data from these surveys are pooled together for a particular calendar year to produce estimates for that year. As a result, ACS estimates reflect data that have been collected over a period of time rather than for a single point in time as in the decennial census. For geographic areas with smaller populations, the ACS samples too few housing units to provide reliable single-year estimates. The Census Bureau therefore combines 5 consecutive years of ACS data to produce estimates for geographic areas with fewer than 65,000 residents. There are consequently two sets of ACS estimates. ACS 1-year estimates are based on data that have been collected over a 12-month period and are suitable for geographic areas with at least 65,000 people. ACS 5-year estimates represent data collected over a period of 60 months and produce estimates which are suitable for all geographic areas including those with fewer than 65,000 residents (United States Census Bureau., 2018).

Table 8-2: Release schedule for ACS data. Adapted from United States Census Bureau. (2018).

Year of release	1-year estimates (65 000 +)	3-year estimates (20 000 +)	5-year estimates (All areas)
2006	2005	NA	NA
2007	2006	NA	NA
2008	2007	2005-2007	NA
2009	2008	2006-2008	NA
2010	2009	2007-2009	2005-2009
2011	2010	2008-2010	2006-2010
2012	2011	2009-2011	2007-2011
2013	2012	2010-2012	2008-2012
2014	2013	2011-2013	2009-2013
2015	2014	NA	2010-2014
2016	2015	NA	2011-2015
2017	2016	NA	2012-2016

NA: Not available.

The release schedule for ACS data is shown in Table 8-2. Starting with data collected in 2005, 1-year estimates have been published for areas with populations of 65,000 or more since 2006. In 2010, the Census Bureau released the first ACS 5-year estimates. ACS 5-year estimates are updated annually by removing the earliest year and replacing it with the latest one. The Census Bureau also produced ACS 3-

year estimates, starting in 2008, but that series was discontinued in 2015 (United States Census Bureau., 2018).

As stated in Section 8.2.2.2, PUMAs have a minimum population of 100,000 people. ACS 1-year estimates are therefore suitable for these geographic areas. ASC 1-year estimates are not suitable for NTAs (Section 8.2.2.1), which on average do not meet the 65,000 population threshold. For this reason, the PUMA was chosen as the statistical unit of choice for this study. It is the smallest statistical geographic unit with sufficiently (temporally) resolved data to meet the needs of this study.

8.4 Crime trends

8.4.1 Citywide crime trends

Data sources: ID 10 & 11 (Table 7-8)

In this study, crime indicators (Table 7-3) are reported according to the index crimes used in the Federal Bureau of Investigation (FBI) Uniform Crime Reporting (UCR) Program (U.S. Department of Justice Federal Bureau of Investigation, 2018). UCR uses standard offense definitions to count and compare crime in localities across America regardless of local variations in crime laws (Data ID 11, Table 7-8). Reported UCR index crimes include seven crimes, which are sub-divided into either violent crimes or property crimes. Specifically, index crimes include the violent crimes of murder (and non-negligent homicide), rape, robbery and aggravated assault; and the property crimes of burglary, motor vehicle theft and larceny-theft. The counts represent only crimes that are reported to the police, and do not reflect the total crimes that occurred (Data ID 11, Table 7-8).

Total index crimes per 100k population for NYC are shown in Figure 8-8 and Figure 8-10 for the period 1990 to 2017 (Data ID 11, Table 7-8). The total reported crime rate in NYC has been steadily declining since records started in 1990. This decrease in crime is often attributed to the implementation of CompStat in the 1990s (see Section 5.3.2), although the extent of its influence on crime is subject to debate (Corman and Mocan, 2002; Roeder et al., 2015).

Crime offences in NYC (Data ID 10, Table 7-8) are reported according to New York State Penal Law definitions. Since UCR index crimes were used in this study, crime records were first classified according to index crime definitions before use. Figure 8-9 and Figure 8-11 show citywide index crime totals for the period 2006 to 2017, comparing aggregated data provided in Data ID 11 (Table 7-8) to totals derived from the NYPD complaint reports dataset (Data ID 10, Table 7-8). For the most part, the author found that there is good agreement between the two datasets, with the exception of rape and aggravated assault.

While records for aggravated assault followed similar trends for both datasets, it was not possible to exactly match the data reported in Data ID 11 (Table 7-8) using the crime identifiers provided in the NYPD complaint reports dataset (Data ID 10, Table 7-8). Regarding rape, the UCR program revised its definition of rape in December 2011. This change was reflected in UCR data starting in 2013 (U.S. Department of Justice Federal Bureau of Investigation, 2018). In the following sections, crime trends for robbery, larceny and assault will be explored. In order to limit the scope of the study, murder, rape, burglary and motor vehicle theft were not further considered.

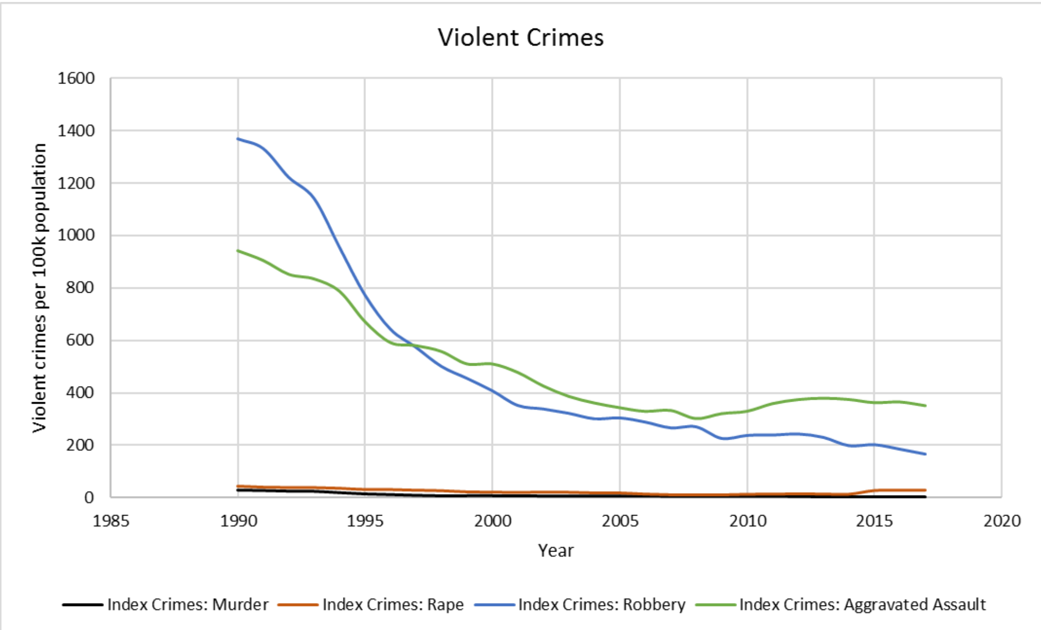


Figure 8-8: Violent crime trends over the period 1990 to 2017 (Data ID 11, Table 7-8). Source: Author's own construction.

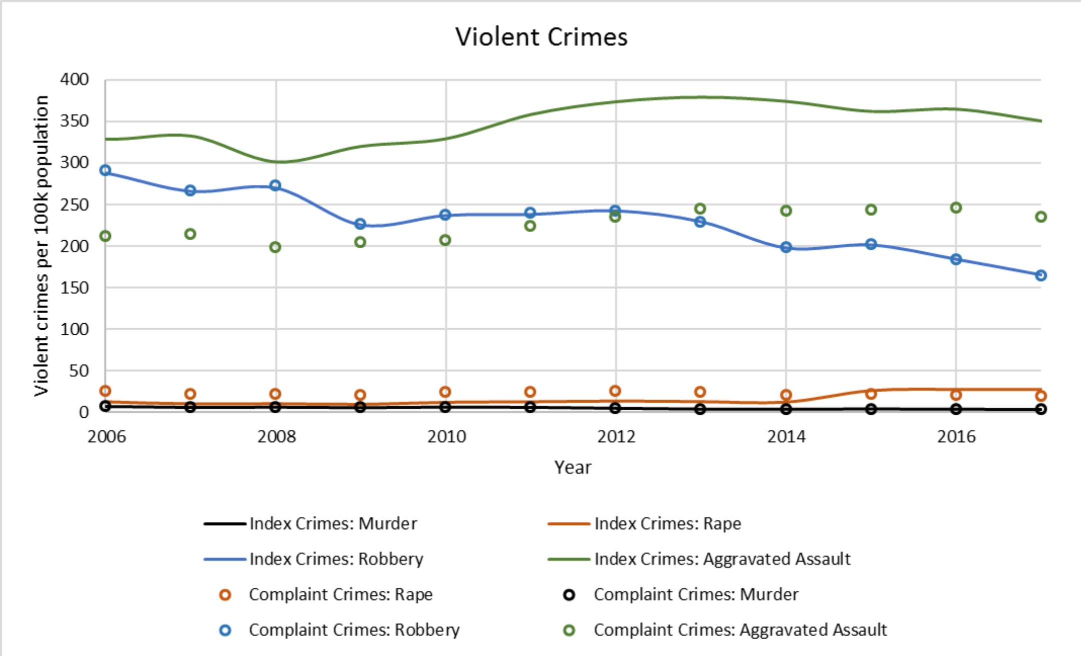


Figure 8-9: Violent crime trends over the period 2006 to 2017 (Data ID 10, 11; Table 7-8). Source: Author's own construction.

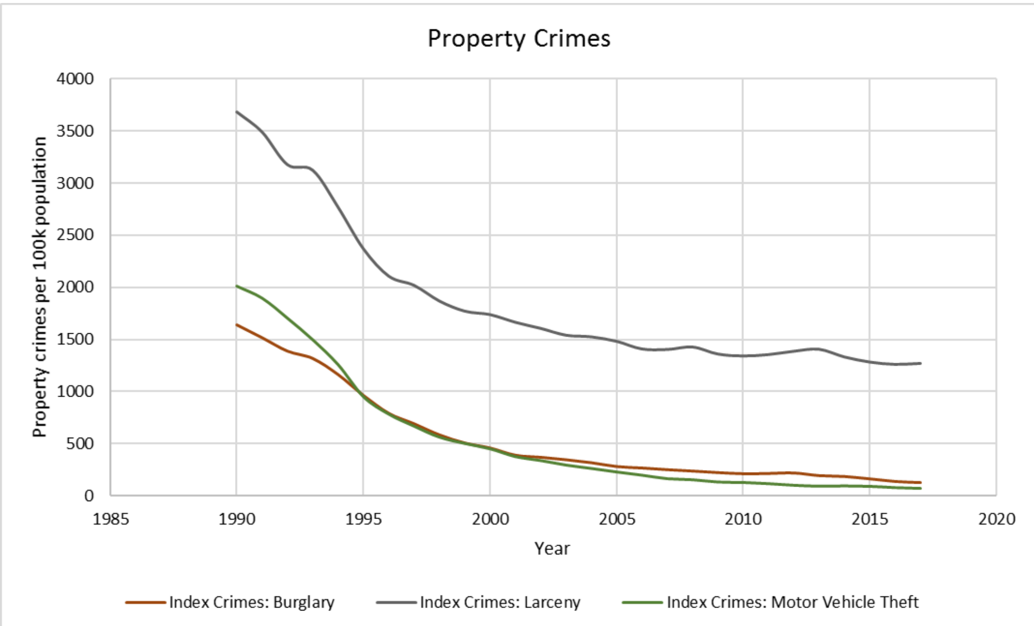


Figure 8-10: Property crime trends over the period 1990 to 2017 (Data ID 11, Table 7-8). Source: Author's own construction.

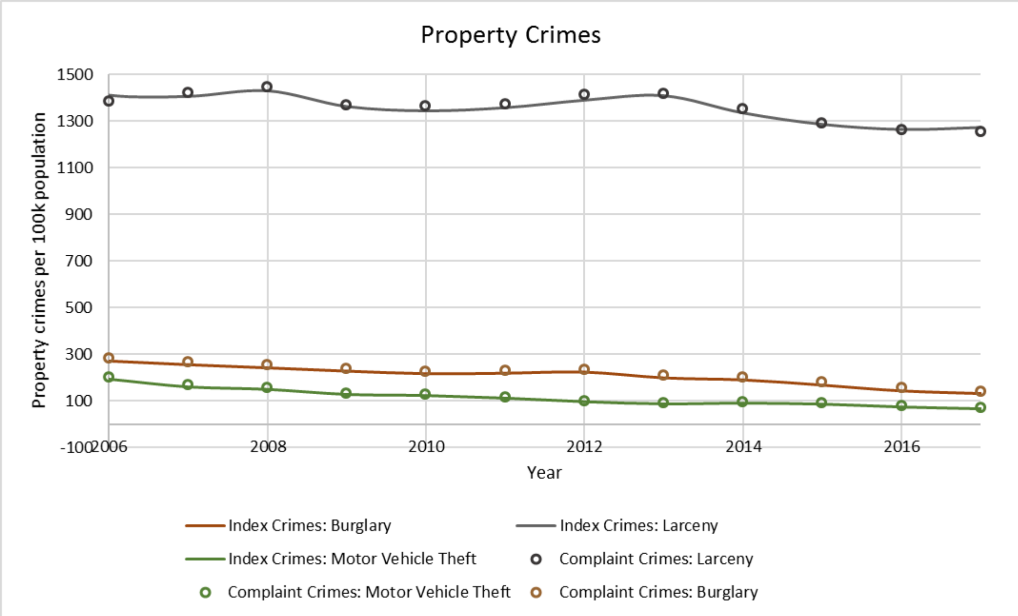


Figure 8-11: Property crime trends over the period 2006 to 2017 (Data ID 10, 11; Table 7-8). Source: Author's own construction.

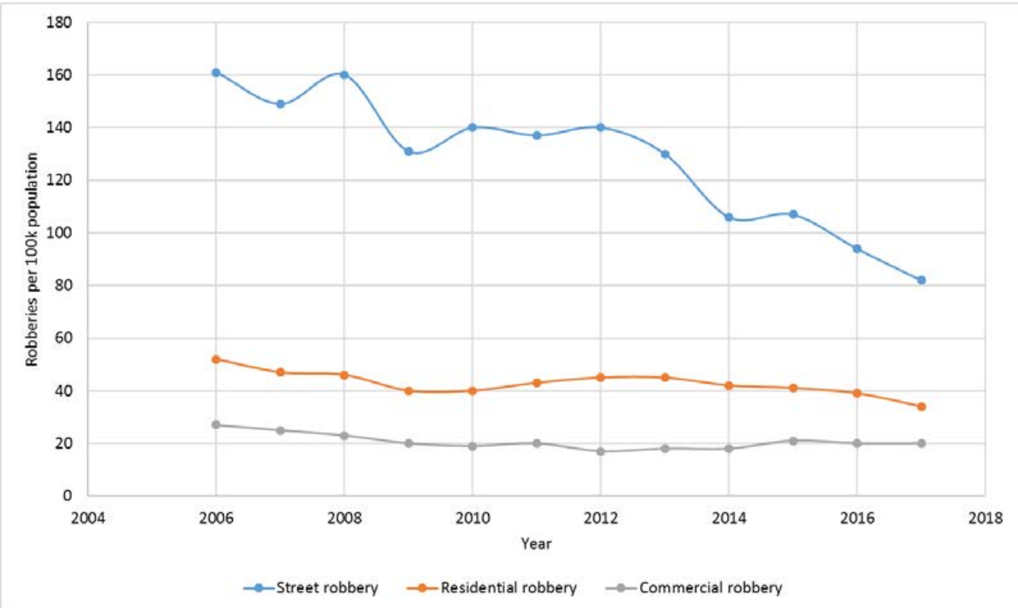


Figure 8-12: Citywide robberies per 100k population per year (Data ID 10, Table 7-8). Source: Author's own construction.

For each of the crimes reviewed in the following sections, the instances of crime per 100k population per PUMA is shown for both 2006 and 2017. Furthermore, the percentage change in crime reports over the reporting period is shown per PUMA. This was done, in order to visualise the hotspot locations of each type of crime, as well as to observe any changes in spatial distribution of the respective crimes over the reporting period.

8.4.2 Robbery

Data source: ID 10 (Table 7-8)

There were 228,078 robberies over the reporting period. Of these robberies, 127,420 (55.9%) occurred on the street, 42,605 (18.7%) occurred in or around residences, and 20,602 (9%) were commercial robberies. At the citywide level, street robberies showed a significant decrease, while residential and commercial robberies remained relatively constant over the reporting period (Figure 8-12).

The number of street robberies per 100k population per PUMA is shown in Figure 8-13 for 2006 and in Figure 8-14 for 2017. The percentage change in street robberies over the reporting period per PUMA in shown in Figure 8-15.

The number of residential robberies per 100k population per PUMA is shown in Figure 8-16 for 2006 and in Figure 8-17 for 2017. The percentage change in residential robberies over the reporting period per PUMA in shown in Figure 8-18.

The number of commercial robberies per 100k population per PUMA is shown in Figure 8-19 for 2006 and in Figure 8-20 for 2017. The percentage change in commercial robberies over the reporting period per PUMA in shown in Figure 8-21.

8.4.3 Larceny

Data source: ID 10 (Table 7-8)

There were 1,357,855 acts of larceny over the reporting period. Of these, 418,266 (30.8%) occurred on the street, 426,069 (31.3%) were thefts from stores or commercial buildings, and 210,068 (15.4%) occurred in or around residences. At the citywide level, street larceny decreased, residential larceny remained constant, and commercial thefts increased over the reporting period (Figure 8-22).

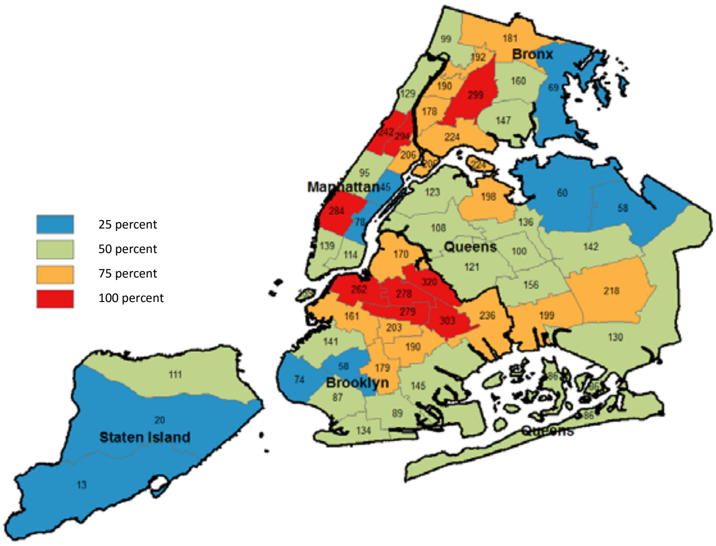


Figure 8-13: Street robberies per 100k population per PUMA in 2006 (Data ID 10, Table 7-8). Source: Author's own construction.

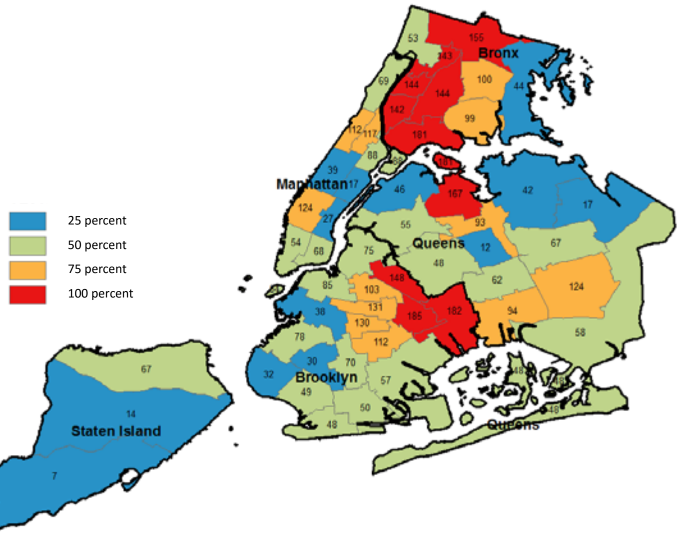


Figure 8-14: Street robberies per 100k population per PUMA in 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

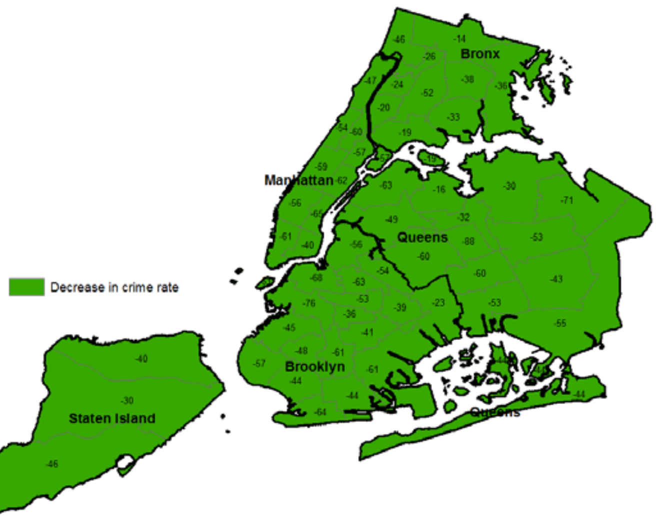


Figure 8-15: Percentage change in street robberies per PUMA over the period 2006 to 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

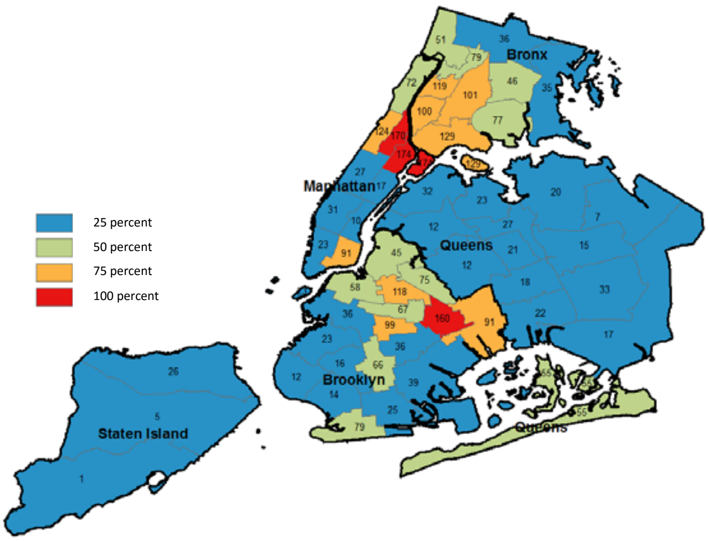


Figure 8-16: Residential robberies per 100k population per PUMA in 2006 (Data ID 10, Table 7-8). Source: Author's own construction.

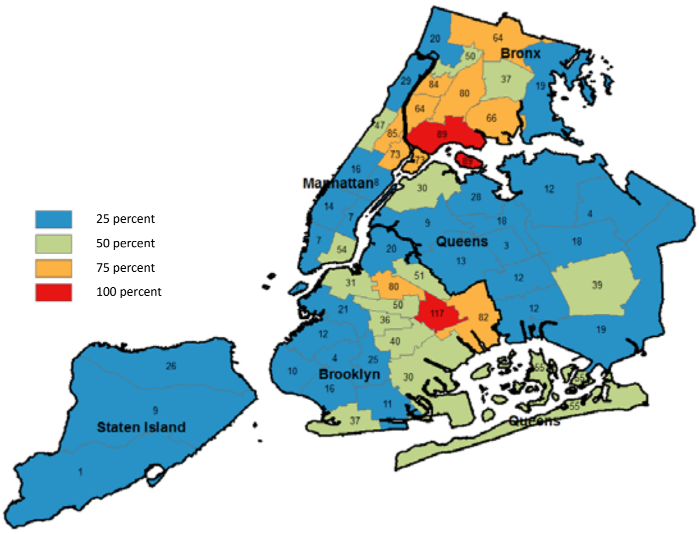


Figure 8-17: Residential robberies per 100k population per PUMA in 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

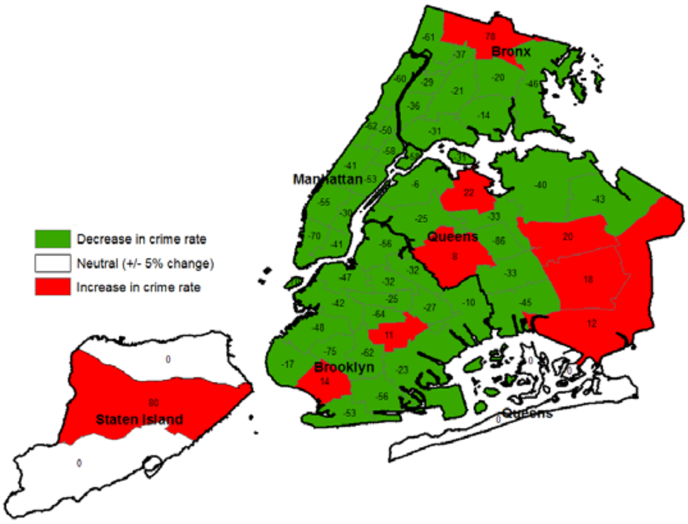


Figure 8-18: Percentage change in residential robberies per PUMA over the period 2006 to 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

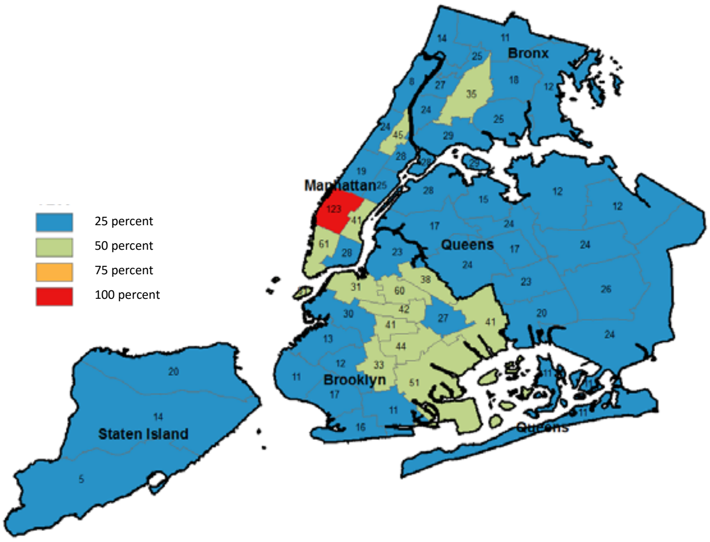


Figure 8-19: Commercial robberies per 100k population per PUMA in 2006 (Data ID 10, Table 7-8). Source: Author's own construction.

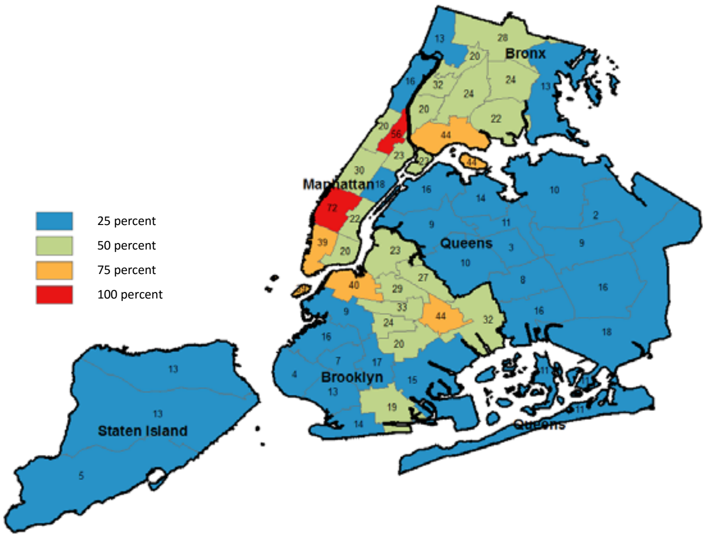


Figure 8-20: Commercial robberies per 100k population per PUMA in 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

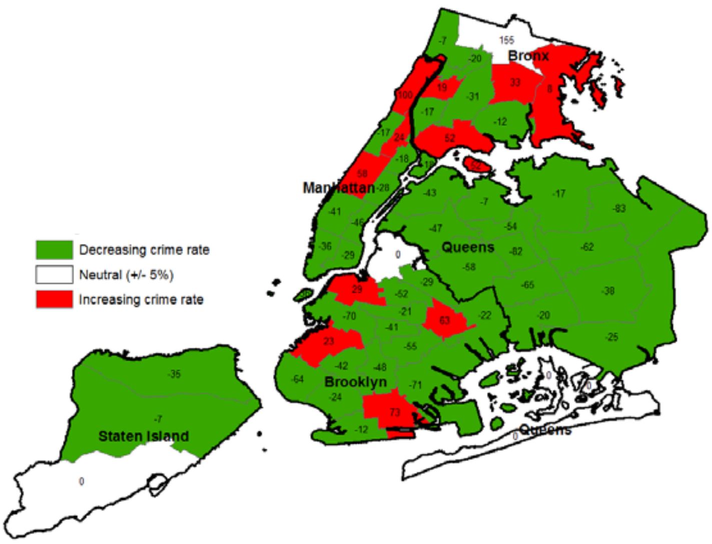


Figure 8-21: Percentage change in commercial robberies per PUMA over the period 2006 to 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

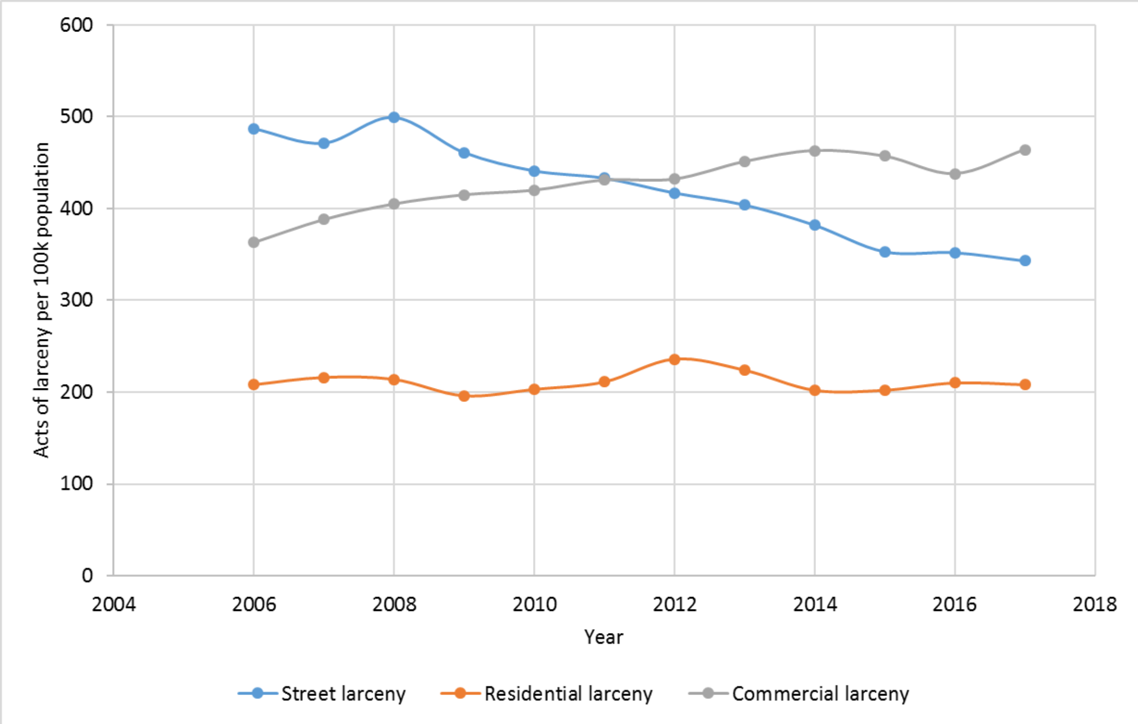


Figure 8-22: Citywide acts of larceny per 100k population per year (Data ID 10, Table 7-8). Source: Author's own construction.

The majority of street larceny is from vehicles. The number of cases of street larceny per 100k population per PUMA is shown in Figure 8-23 for 2006 and in Figure 8-24 for 2017. The percentage change in street larceny over the reporting period per PUMA is shown in Figure 8-25.

The number of cases of residential larceny per 100k population per PUMA is shown in Figure 8-26 for 2006 and in Figure 8-27 for 2017. The percentage change in residential larceny over the reporting period per PUMA in shown in Figure 8-28.

The number of cases of commercial larceny per 100k population per PUMA is shown in Figure 8-29 for 2006 and in Figure 8-30 for 2017. The percentage change in commercial larceny over the reporting period per PUMA in shown in Figure 8-31.

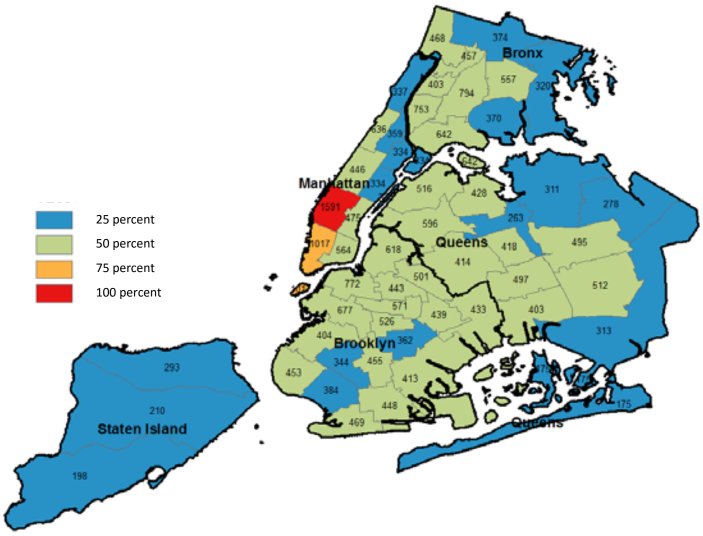


Figure 8-23: Street larceny per 100k population per PUMA in 2006 (Data ID 10, Table 7-8). Source: Author's own construction.

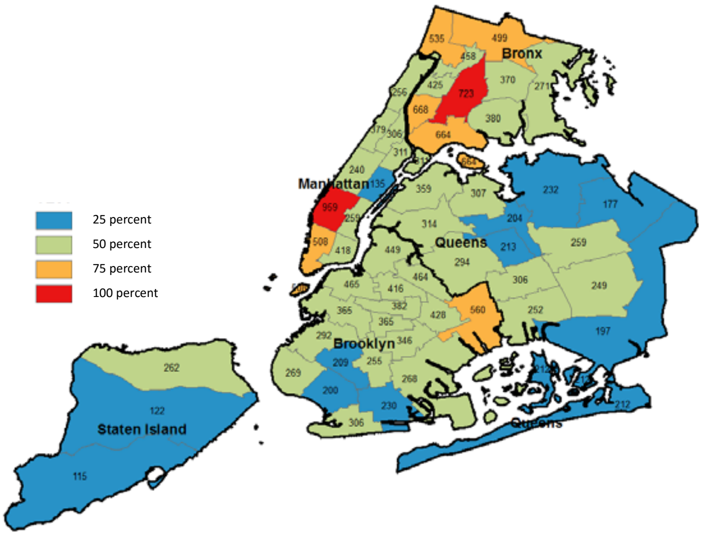


Figure 8-24: Street larceny per 100k population per PUMA in 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

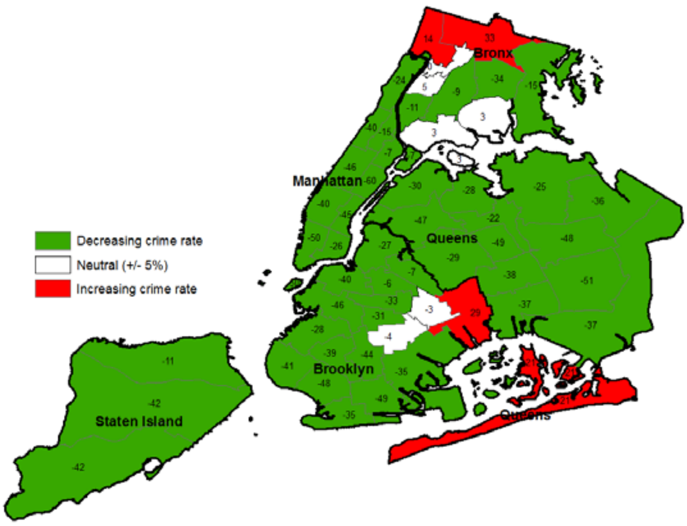


Figure 8-25: Percentage change in street larceny per PUMA over the period 2006 to 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

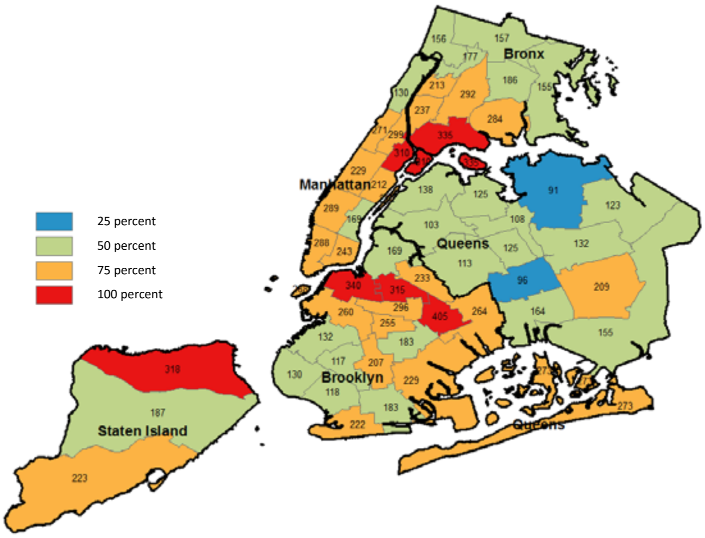


Figure 8-26: Residential larceny per 100k population per PUMA in 2006 (Data ID 10, Table 7-8). Source: Author's own construction.

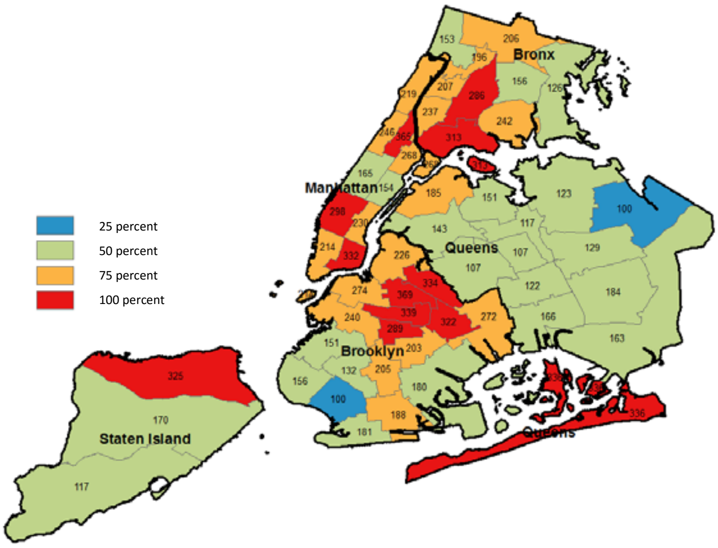


Figure 8-27: Residential larceny per 100k population per PUMA in 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

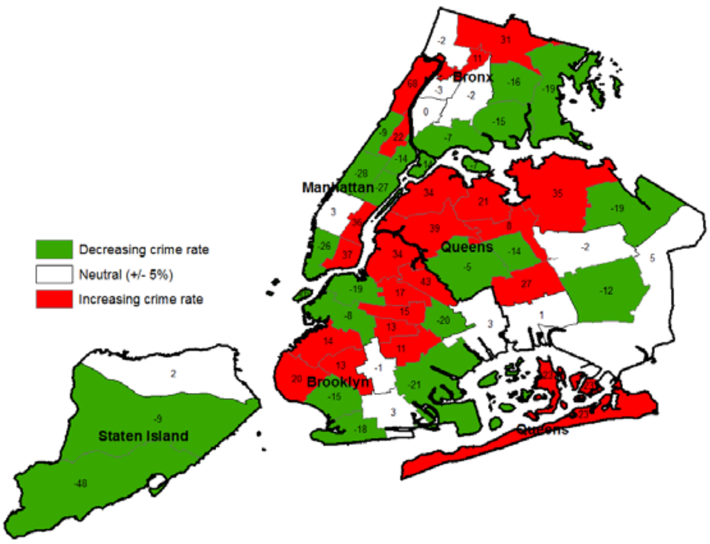


Figure 8-28: Percentage change in residential larceny per PUMA over the period 2006 to 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

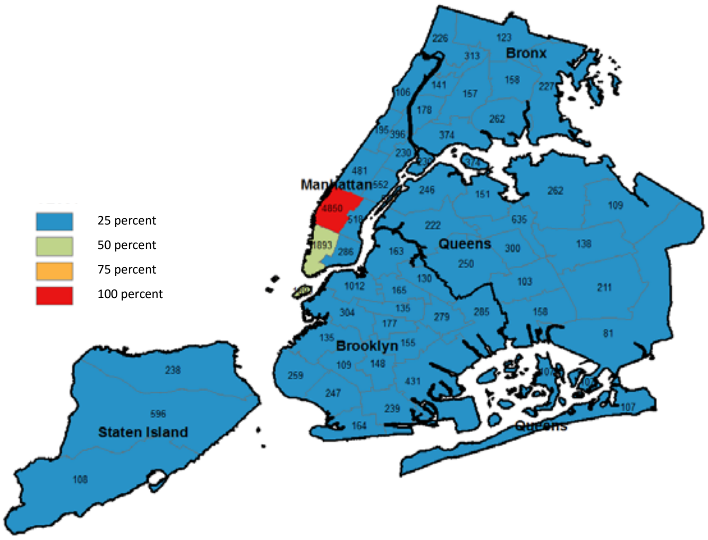


Figure 8-29: Commercial larceny per 100k population per PUMA in 2006 (Data ID 10, Table 7-8). Source: Author's own construction.

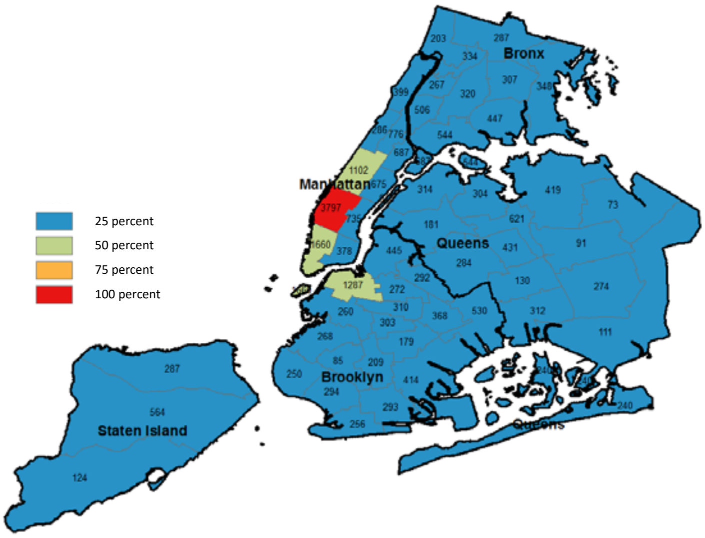


Figure 8-30: Commercial larceny per 100k population per PUMA in 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

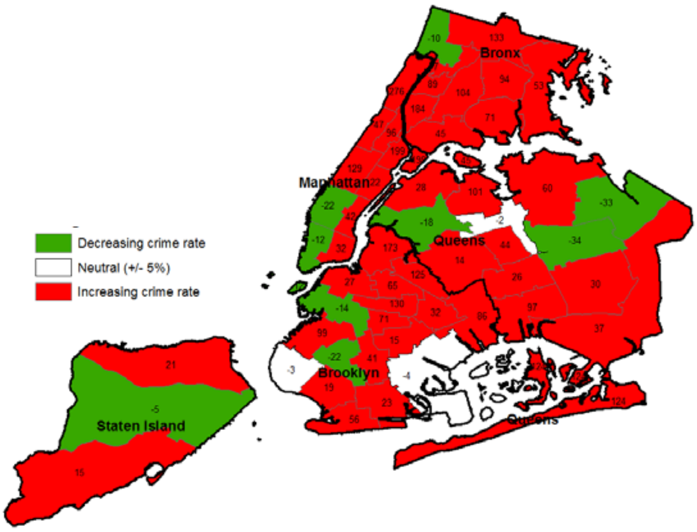


Figure 8-31: Percentage change in commercial larceny per PUMA over the period 2006 to 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

8.4.4 Assault

Data source: ID 10 (Table 7-8)

There were 224,804 acts of felony assault over the reporting period. Of these, 71,272 (31.7%) occurred on the street, while 107,437 (47.8%) occurred in or around residences. At the citywide level, street assault decreased, while residential assault increased over the reporting period (Figure 8-32).

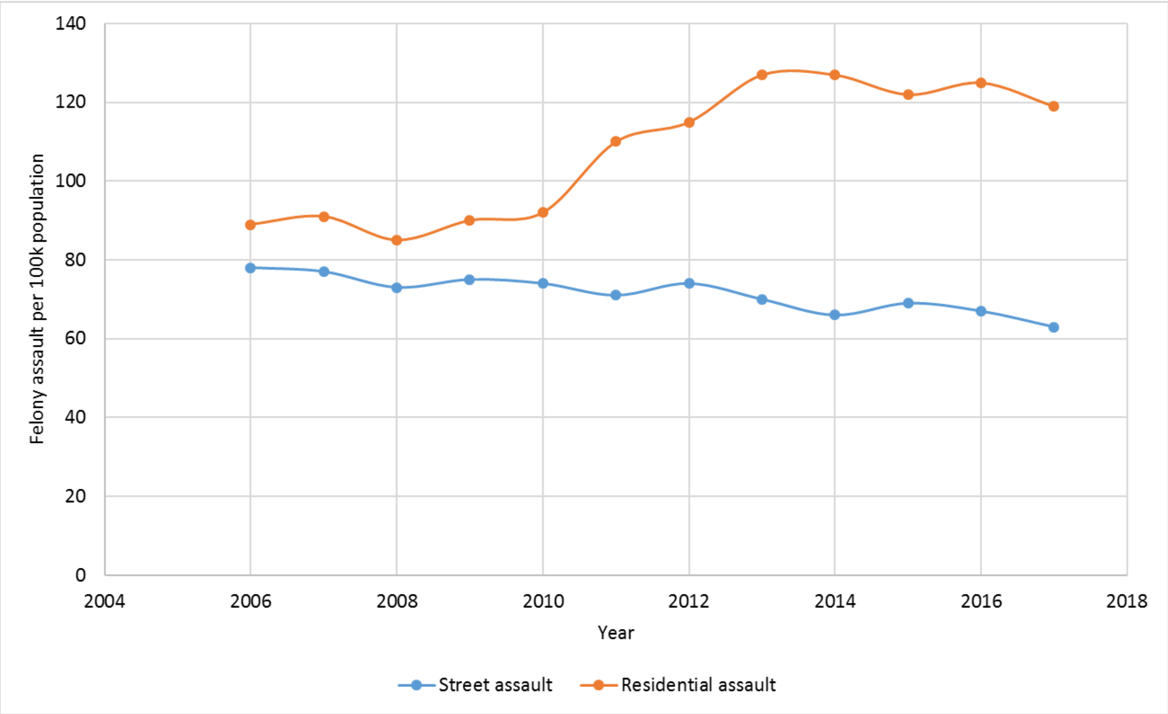


Figure 8-32: Citywide felony assault per 100k population per year (Data ID 10, Table 7-8). Source: Author's own construction.

The number of street felony assault cases per 100k population per PUMA is shown in Figure 8-33 for 2006 and in Figure 8-34 for 2017. The percentage change in street felony assault over the reporting period per PUMA in shown in Figure 8-35.

The number of cases of residential felony assault per 100k population per PUMA is shown in Figure 8-36 for 2006 and in Figure 8-37 for 2017. The percentage change in residential felony assault over the reporting period per PUMA in shown in Figure 8-38.

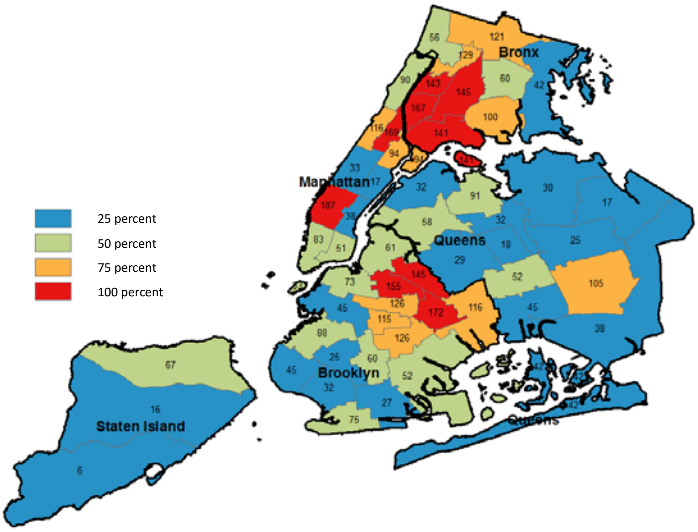


Figure 8-33: Street felony assault per 100k population per PUMA in 2006 (Data ID 10, Table 7-8). Source: Author's own construction.

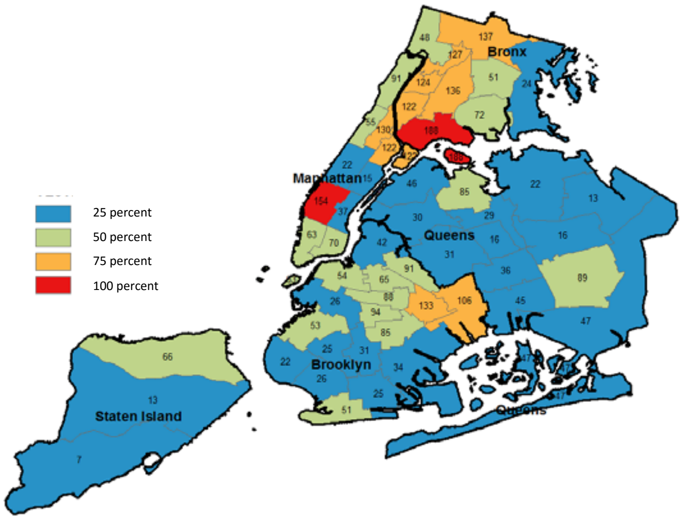


Figure 8-34: Street felony assault per 100k population per PUMA in 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

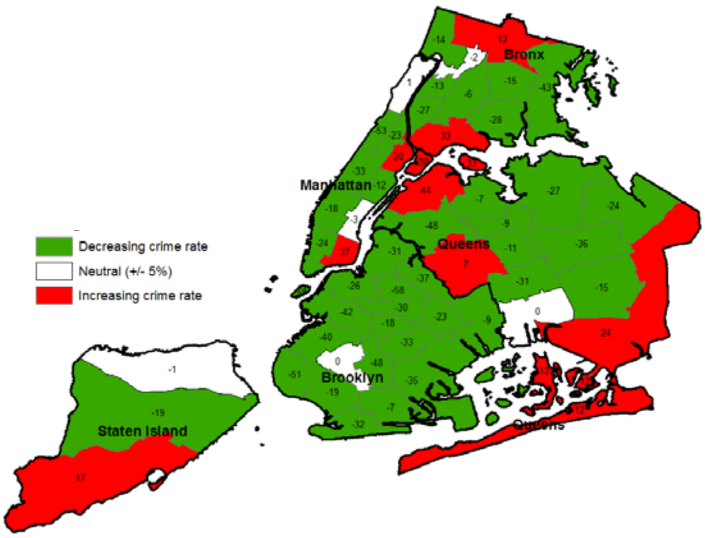


Figure 8-35: Percentage change in street felony assault per PUMA over the period 2006 to 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

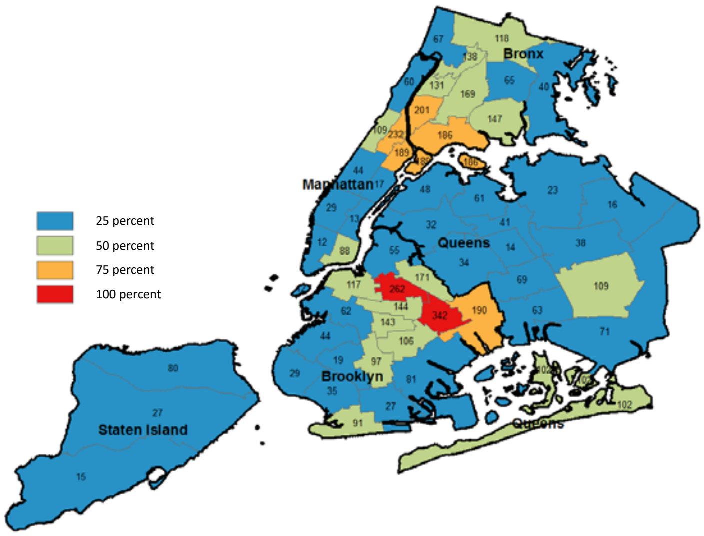


Figure 8-36: Residential felony assault per 100k population per PUMA in 2006 (Data ID 10, Table 7-8). Source: Author's own construction.

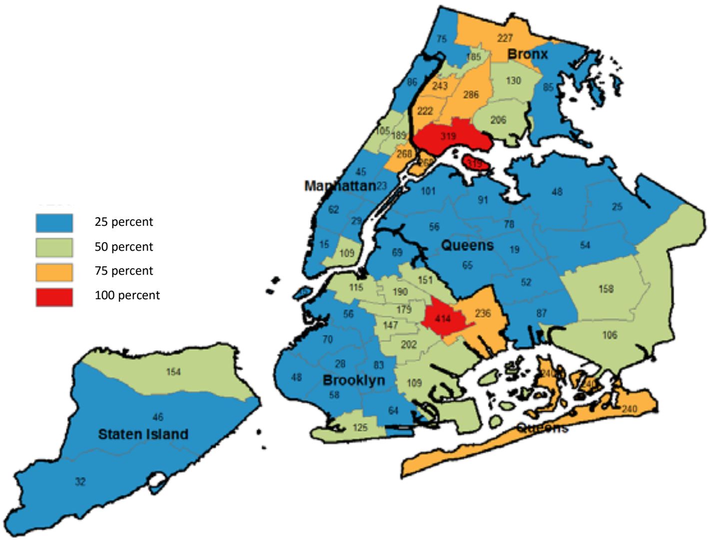


Figure 8-37: Residential felony assault per 100k population per PUMA in 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

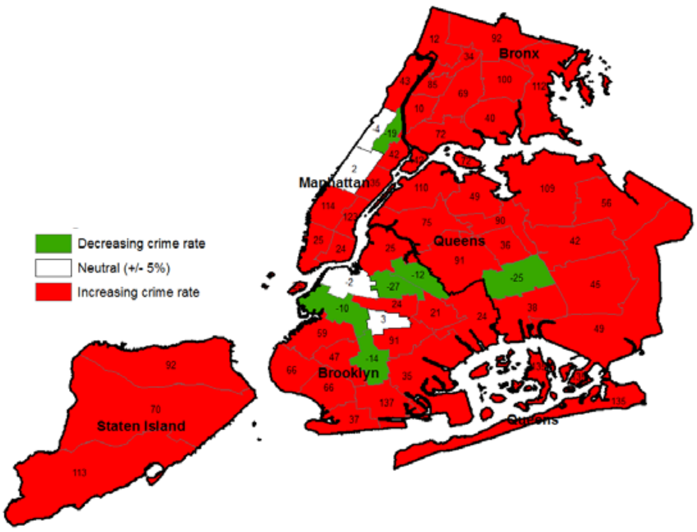


Figure 8-38: Percentage change in residential felony assault per PUMA over the period 2006 to 2017 (Data ID 10, Table 7-8). Source: Author's own construction.

8.5 Demographics of perpetrators

Data sources: ID 17, 37 & 38 (Table 7-8)

Perpetrator demographics are summarised in Figure 8-39 to Figure 8-43. The number of adult inmate admissions and youth detention admissions per year for the period 2015 to 2017 are shown in Figure 8-39. From Figure 8-39, it is apparent that the number of admissions are declining.

The age, race and gender distributions of adult inmates and youths in detention are shown in Figure 8-40, Figure 8-41 and Figure 8-42, respectively. The majority of inmates are Black males in their mid-twenties to thirties. The residential location of youths in detention is shown by ZIP code in Figure 8-43. The majority of youths in detention originate from parts of The Bronx and Brooklyn.

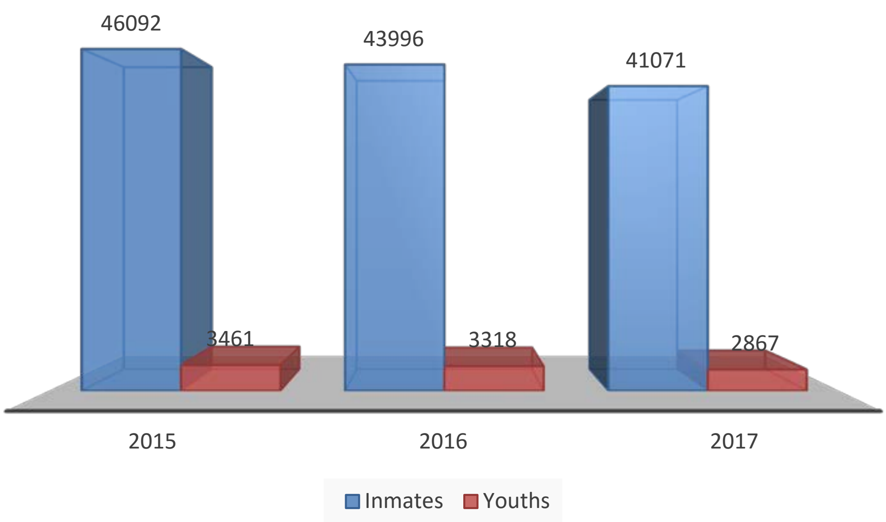


Figure 8-39: Number of adult inmate admissions and youth detention admissions per year for the period 2015 to 2017 (Data ID 17, 37; Table 7-9). Repeat admissions are treated as one for any given year. Source: Author's own construction.

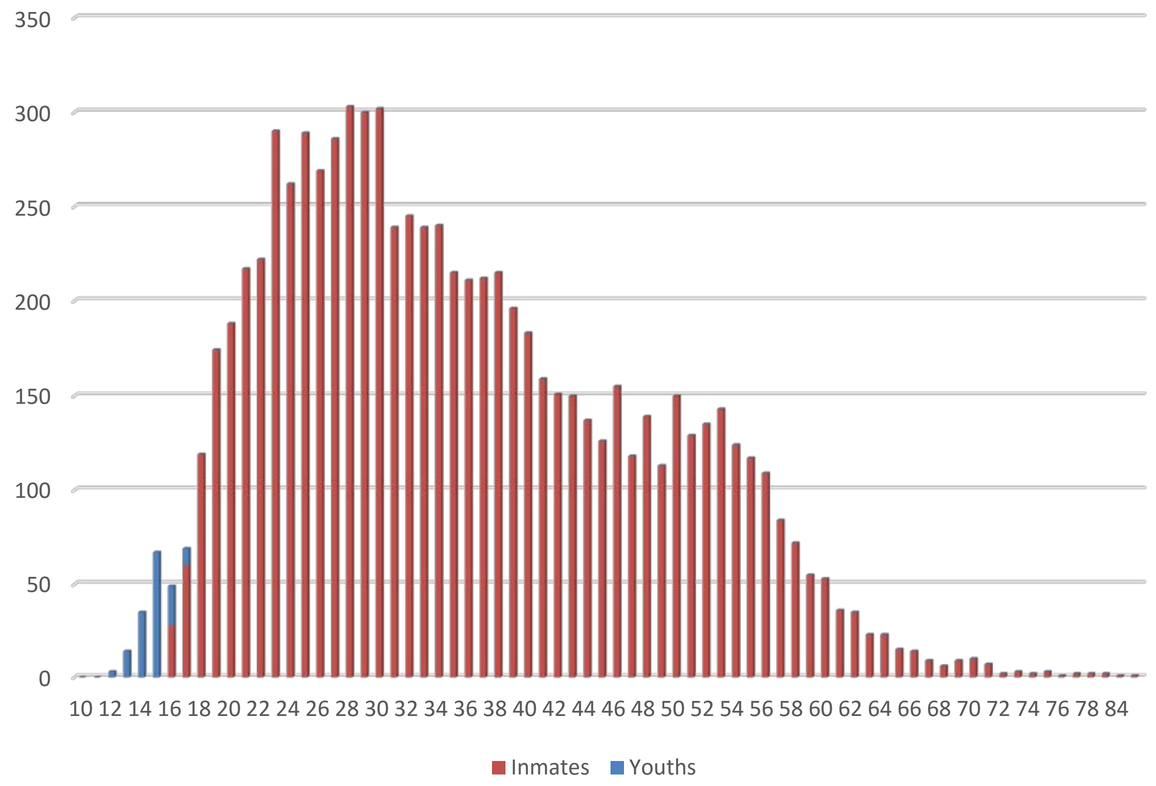


Figure 8-40: Histogram of the age distribution of the adult inmate population (24 August 2018) and youths in detention (averaged over the period 2015 to 2017) (Data ID 17, 38; Table 7-9). Source: Author's own construction.

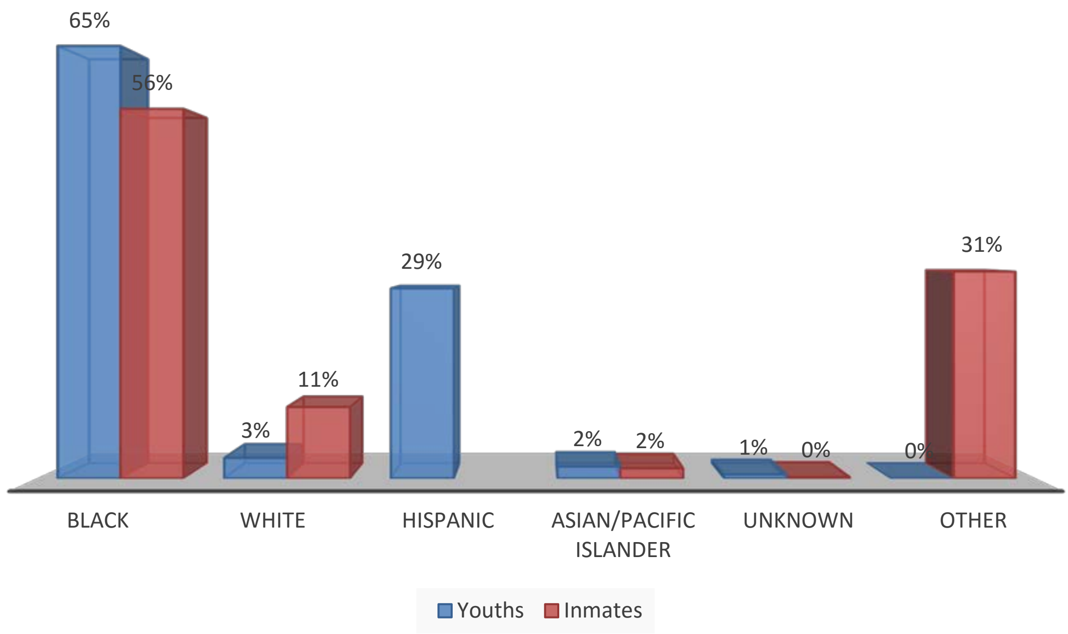


Figure 8-41: Race distribution of adult inmates (24 August 2018) and youths admitted to detention (averaged over the period 2015 to 2017) (Data ID 17, 38; Table 7-9). Source: Author's own construction.

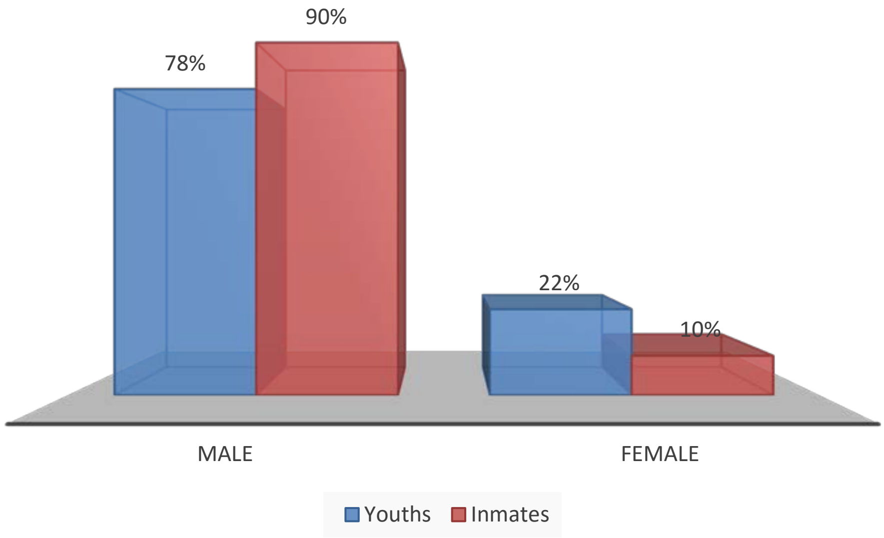


Figure 8-42: Average gender distribution of adult inmates and admitted youth for the period 2015 to 2017 (Data ID 17, 38; Table 7-9). Source: Author's own construction.

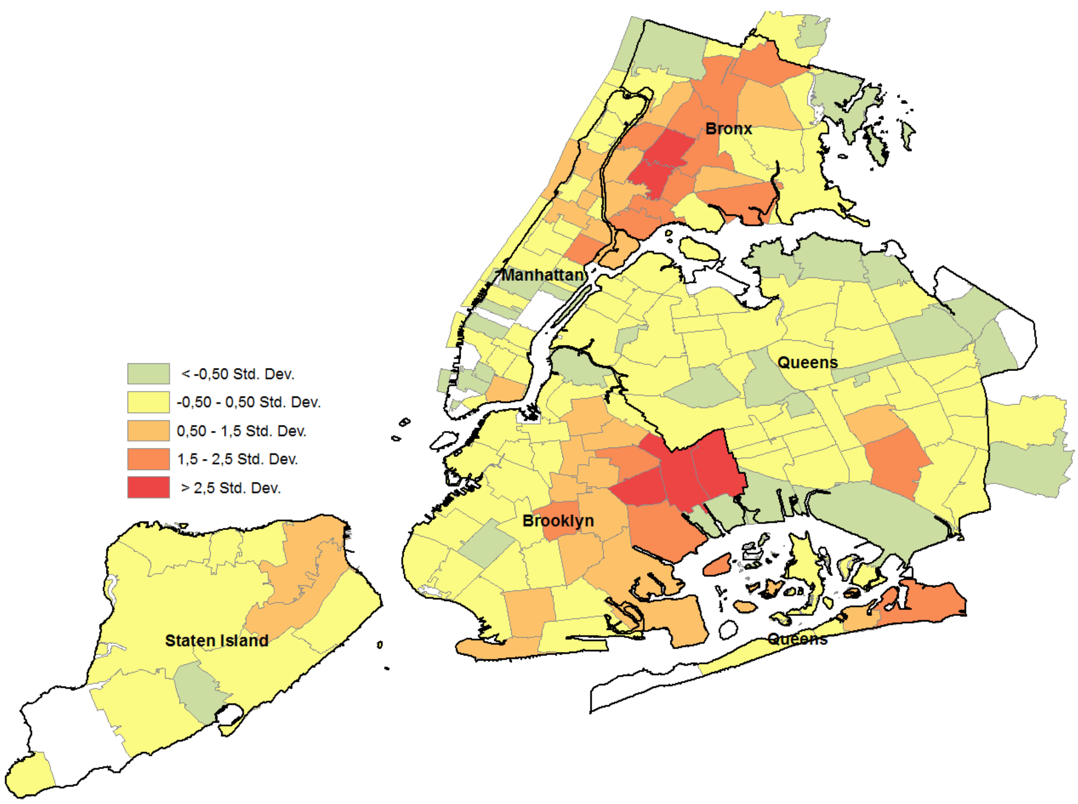


Figure 8-43: Average number of youths in detention by reported ZIP code of youths' primary residence per year for the period 2015 to 2017 (Data ID 17; Table 7-9). Source: Author's own construction.

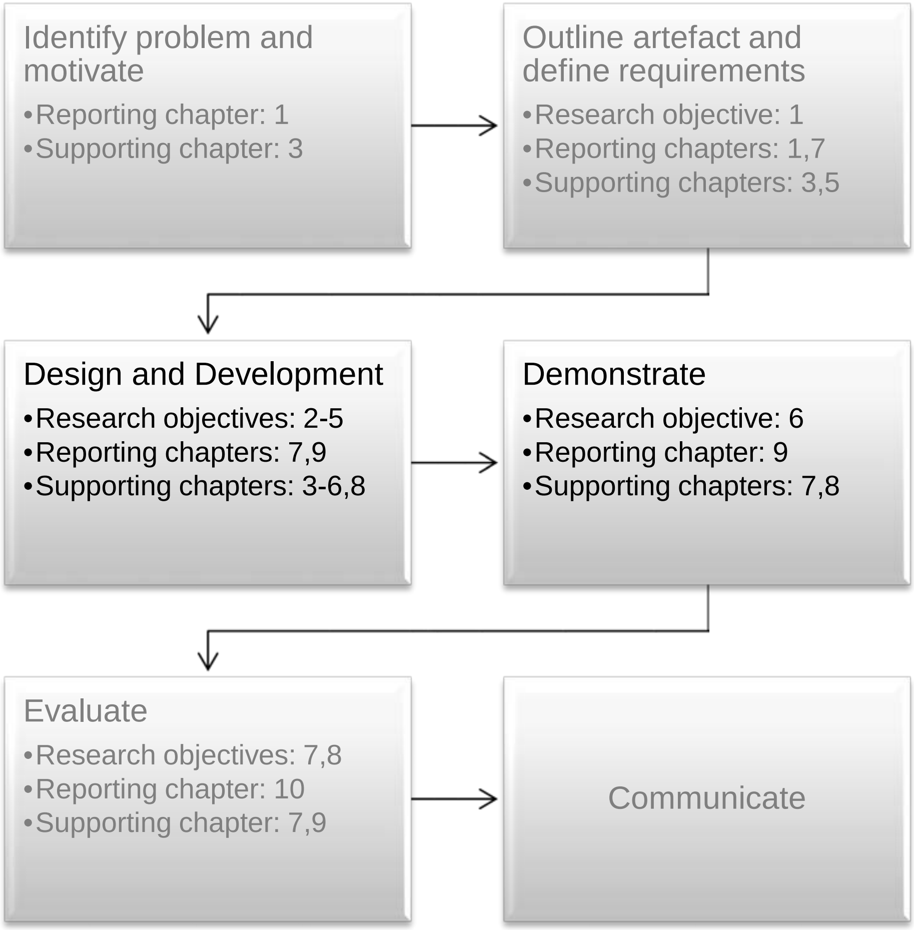
8.6 Summary

In this study, NYC open data was used to develop and demonstrate a prototype model for crime management in smart cities. In this chapter, available data was explored. This exploratory data analysis provided context for model design in Chapter 7 and model interpretation in Chapter 9. This chapter, therefore, contributed towards addressing $RO_{5\text{-}6}$ (Figure 8-1).

An understanding of the relationships between commonly used geographic units is essential when combining data from different sources. The geographic units most often used in NYC open data were described and compared in Section 8.2. Section 8.3 provided an overview of available demographic data, and discussed the implication of sample size when selecting a spatial unit of analysis (see Section 7.7). In order to provide context for model interpretation, an overview of crime trends for the period 2006 to 2017 was provided in Section 8.4. This was followed by a description of perpetrator demographics over the same period in Section 8.5.

The main objective of this study is to develop and evaluate a predictive model for crime management in smart cities (Section 1.5). The model design was developed in Chapter 7, based on guidelines derived from the literature review in Section 2. In the following chapter, the model will be implemented and demonstrated according to the design specifications laid out in Chapter 7.

Chapter 9. Model Implementation



Research objectives addressed in this chapter:

RO_1: Identify the functional, construction and environmental requirements of an effective model.

RO_2: Identify relevant input and output parameters.

RO_3: Identify and characterise available data sources.

RO_4: Identify the modelling technique to be used to develop the model.

RO_5: Develop the model.

RO_6: Demonstrate the application of the model.

RO_7: Evaluate the efficacy of the model.

RO_8: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

Figure 9-1: Research objectives and design science research activities addressed in this chapter.

9.1 Introduction

In this study, the Design Science Research (DSR) process is followed to develop, demonstrate and evaluate a prototype model for crime management in smart South African cities. The DSR methodology and its anticipated application in this study (Figure 9-1) was outlined in Chapter 2. Section 2 of this thesis (Chapters 3 through 6), provided the supporting literature necessary to formulate the research problem, outline and define the requirements of a potential solution, and identify potential design interventions. Section 3 of this thesis (Chapters 7 through 10), focuses on the development and evaluation of the prototype model. The model design was developed and outlined in Chapter 7, followed by the exploration of available data in Chapter 8. In this chapter, the model will be implemented and demonstrated according to the design specifications laid out in Chapter 7. This chapter will therefore address $RO_{5\text{-}6}$ (Figure 9-1).

An overview of the implementation process is given Section 9.2. This is followed by a summary of the data used to develop the model (Section 9.3.1). The data wrangling process will also be discussed, together with a summary of the challenges faced during the data preparation process (Section 9.3.2). The neural network implementation will then be summarised in Section 9.4, followed by a detailed demonstration of the model interpretation process (Section 9.5 and 9.6).

9.2 Process overview

In Section 9.4, it will be shown how Bayesian Neural Networks were used to develop a set of prototype models for crime management in smart cities. Two models were developed for demonstration purposes. Specifically, models were developed to predict street larceny and street robbery in New York City, respectively. The choice of crimes were intentional, with the aim of encapsulating as wide a range of system behaviour as possible. As seen in Section 8.4, the two crimes exhibit different spatial patterns. They also represent different types of crime, namely property crime (larceny) and violent crime (robbery).

An overview of the model development and implementation process, together with the relevant chapter sections, is shown in Figure 9-2. The input data used to develop the models are summarised in Section 9.3, together with a discussion on the data wrangling challenges experienced in the process of preparing the data. The training of the neural networks is described in Section 9.4.

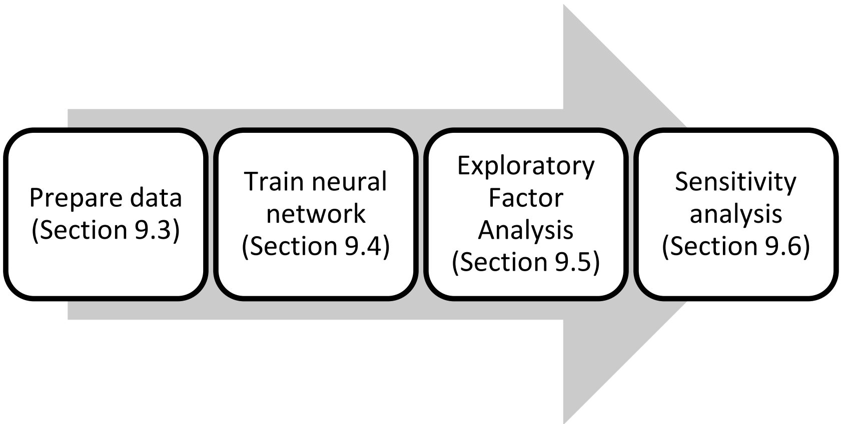


Figure 9-2: Overview of model development and implementation process. Source: Author's own construction.

The anticipated application of the prototype crime predictors is to be a tool for sensitivity analysis aimed at determining the relative influence of decision variables on crime, thereby providing a quantitative means of determining the most effective course of action (Section 7.3). Typically, as part of a sensitivity analysis, an input variable is varied while the remaining input variables are fixed at some value (Section 6.5.1). The changes in the network output are then observed.

During the initial stage of model exploration, it was found that there was a high degree of correlation among input features (Section 9.5). Further investigation showed that the input data tended to cluster together to represent different system "states" (Section 9.5). Due to the clustering in input space, care needed to be taken when fixing variables for sensitivity analysis. The choice of fixed variables is explained in Section 9.6.1.

Exploratory Factor Analysis was used to identify latent "states" within the input data. The process followed, and the observed states, are discussed in Section 9.5. The results of the sensitivity analysis are shown in Section 9.6 for each model, together with a demonstration of its anticipated application in practice. The software packages used in this study included Microsoft Excel for data wrangling and plotting; R (R Core Team., 2018) for data wrangling, plotting and analysis; Esri® ArcMap™ for spatial plotting and performing spatial joins; and Model Manager (Section 6.4) for neural network implementation.

9.3 Data wrangling and summary

9.3.1 Data preparation and summary

The datasets listed in Table 7-8 were cleaned and wrangled to represent the KPIs listed in Table 7-3. The final set of variables used to develop the prototype models are listed in Table 9-1. Variables relating to race (*hispanic, white, black, asian*), and target crimes (*larStreet, larCommercial, larResidence, robStreet, robCommercial, robResidence, assStreet, assResidence*) were not included as input features during neural network development (Section 9.4). However, these variables were included in Exploratory Factor Analysis (Section 9.5).

For each variable listed in Table 9-1, the variable identifier, measurement unit, descriptive statistics, and method(s) used to replace missing values are specified. Missing values were replaced either by linear interpolation (I) or linear extrapolation (E), or by using the nearest non-missing value (N). The associated indicator(s) and data source(s) for each variable are specified in Table 7-3.

As specified in Section 7.7, all KPIs were presented at the PUMA geographical unit. For each variable, a table of annual measures for each PUMA was created for the years 2006 to 2017. A sample input file is shown in Table 9-2. Once input files had been created for each variable, these were combined into a single table of variables with columns representing each variable listed in Table 9-1. The table consisted of 660 data tuples; with each of the 55 PUMAs contributing 12 data tuples, one for each year. PUMA and year identifiers were not used in the development of the neural networks (Section 9.4), and predictions were based purely on location features.

Table 9-1: Variable names and descriptions.

varName	units	Min	Max	Median	Mean	Skewness	Kurtosis	Missing data
larStreet	counts	88	1736	388	417	2.37	10.3	-
larCommercial	counts	73	5522	268	426	5.47	33.01	-
larResidence	counts	66	486	204	212	0.55	-0.42	-
robStreet	counts	6	367	117	127	0.56	-0.22	-
robCommercial	counts	1	123	18	21	2.11	8.7	-
robResidence	counts	1	174	31	44	1.1	0.57	-
assStreet	counts	4	204	59	72	0.65	-0.61	-
assResidence	counts	10	434	84	109	1.22	1.41	-
events	counts	7	428	48	66	3.01	13.21	N
P1	%	0	62	26	26	0.21	-1.16	N
P2	%	1	32	10	12	0.85	-0.05	N
P3	%	0	38	9	10	1.07	1.51	N
P4	%	0	33	5	7	1.93	3.86	N
P5	%	2	36	4	6	3.95	17.9	N
P6	%	0	36	3	4	2.74	7.51	N
P7	%	0	24	3	4	1.92	3.04	N
P8	%	2	24	8	10	0.79	-0.25	N
P9	%	1	51	11	14	1.18	0.7	N
P10	%	0	7	2	2	1.3	1.31	N
P11	%	0	40	3	5	2.17	5.71	N
V1	km^{-2}	21	898	47	100	3.23	10.06	N
V5	km^{-2}	9	983	35	100	3.27	9.97	N
pedIndex	counts	193	11918	1978	2286	1.56	3.24	N
abuse	counts	19	1565	282	401	1.2	0.67	E
unemployment	%	4	20	9	10	0.52	-0.53	IE
insurance	%	2	31	12	12	0.78	1.06	NIE
ineqT1r	index x 100	20	48	32	33	0.31	-0.61	IE
ineqT2r	index x 100	-80	58	-11	-10	-0.03	-0.08	IE
hispanic	%	6	72	21	29	0.84	-0.61	IE
white	%	1	85	28	33	0.4	-1.11	IE
black	%	0	89	13	23	1.07	-0.03	IE
asian	%	0	56	8	13	1.3	0.98	IE
diversity	index x 100	60	357	187	190	0.56	0.77	IE
female	%	1	28	8	10	0.9	0.05	IE
degree	%	12	86	36	40	1.08	0.61	IE
noHigh	%	3	49	20	21	0.37	-0.25	IE
socialHousing	%	0	29	3	6	1.52	1.91	N
drugs	counts	15	2055	247	432	1.44	1.21	-
graffiti	counts	10	983	91	111	3.93	29.77	-
fertility	counts	29	140	58	59	1.79	7.28	N
integrity	counts	0	290	53	76	1.47	1.67	E
PM	$\mu g.m^{-3}$	6	16	10	10	0.43	0.19	E
homeless	counts	0	1071	18	40	7.3	71.55	-
SL	$counts.day^{-1}.km^{-2}$	0.01	113.07	2.99	4.44	8.17	102.13	-

Table 9-2: Sample input file created for each variable. Data for street larceny is shown here.

PUMA	Y2006	Y2007	Y2008	Y2009	Y2010	Y2011	Y2012	Y2013	Y2014	Y2015	Y2016	Y2017
3701	468	469	569	595	486	544	495	382	399	452	488	535
3702	374	308	316	320	345	322	380	353	294	327	450	499
3703	320	322	435	283	392	312	320	335	283	323	268	271
3704	557	469	480	481	436	468	432	457	472	432	443	370
3705	794	858	718	647	614	653	610	576	589	569	595	723
3706	457	410	475	449	398	408	446	389	407	419	472	458
3707	403	400	546	489	420	455	435	487	505	420	561	425
3708	753	622	693	538	508	510	484	476	502	475	505	668
3709	370	425	451	405	360	382	349	408	370	358	398	380
3710	642	803	653	633	573	606	623	598	600	607	728	664
3801	337	371	441	433	401	315	317	286	299	274	305	256
3802	636	514	527	530	432	344	414	348	340	296	347	379
3803	359	305	326	324	308	278	325	361	280	302	296	306
3804	334	397	432	369	374	314	393	408	435	416	306	311
3805	334	355	307	194	279	222	214	192	195	171	155	135
3806	446	471	502	394	451	318	346	357	290	315	289	240
3807	1591	1736	1591	1472	1219	1216	1270	1297	1088	1115	1131	959
3808	475	451	496	489	377	334	322	411	340	325	301	259
3809	564	549	695	660	623	494	517	524	509	464	450	418
3810	1017	1025	1083	850	765	664	696	704	579	589	586	508
3901	198	157	178	156	183	139	119	138	126	107	88	115
3902	210	160	233	197	212	159	147	153	120	104	134	122
3903	293	256	347	366	385	367	412	344	293	278	298	262
4001	618	607	725	633	696	789	645	616	670	585	511	449
4002	501	422	534	496	548	537	494	516	529	427	481	464
4003	443	528	607	489	629	546	471	393	444	405	379	416
4004	772	809	825	628	749	770	694	730	687	515	503	465
4005	677	486	731	655	747	725	646	541	512	373	370	365
4006	571	608	544	516	534	589	569	583	575	411	489	382
4007	439	342	445	487	443	416	514	381	410	391	334	428
4008	433	407	472	475	380	530	586	484	572	498	432	560
4009	413	315	445	424	325	470	397	363	326	333	277	268
4010	362	392	391	385	336	352	330	322	316	323	304	346
4011	526	534	450	596	511	574	465	450	385	381	455	365
4012	404	349	501	421	442	366	403	446	365	277	312	292
4013	453	426	359	429	496	357	405	359	314	298	322	269
4014	344	331	353	386	318	386	302	345	286	274	243	209
4015	455	400	420	449	364	426	458	372	301	340	278	255
4016	448	434	448	439	410	419	418	311	325	272	224	230
4017	384	317	388	374	275	282	220	231	206	202	231	200
4018	469	510	511	456	497	364	426	305	297	329	260	306
4101	516	634	567	421	523	468	402	398	400	384	373	359
4102	428	373	439	450	418	392	285	308	299	273	284	307
4103	311	335	290	344	285	280	239	221	242	270	252	232
4104	278	274	240	334	253	256	240	219	202	193	228	177
4105	313	228	270	335	269	244	214	194	199	219	185	197
4106	495	371	369	417	378	334	301	297	258	299	276	259
4107	263	250	281	210	213	246	223	226	177	219	219	204
4108	418	389	344	258	352	375	248	257	212	192	203	213
4109	596	544	572	464	547	450	376	458	356	301	325	314
4110	414	418	480	400	348	424	367	352	310	286	294	294
4111	497	432	481	417	349	457	387	409	372	347	283	306
4112	512	472	493	411	372	366	336	299	304	279	266	249
4113	403	439	382	362	302	325	408	330	313	295	255	252
4114	175	171	241	216	193	168	241	243	216	188	140	212

In order to eliminate bias introduced by the wide range of scales used among the selected input features evident in Table 9-1, each variable was first standardised before proceeding with model development (Hair *et al.*, 2014). The standard score (or Z-score) for each variable instance was determined by subtracting the mean and dividing by the standard deviation of each variable. The process thereby converted each variable into a standardised set of variable instances, with a mean of 0 and a standard deviation of 1 (Hair *et al.*, 2014).

9.3.2 Data wrangling challenges

A number of challenges were faced in the data preparation process. These are summarised below.

9.3.2.1 Compatibility of spatial units

The use of aggregated data proved challenging when spatial boundaries changed, or when the reporting unit of different data sources were not compatible. As mentioned in Section 7.7, data was aggregated at the PUMA level in this study. Data sources that were aggregated at incompatible spatial units were therefore of little use in this study. For example, while the residential location of youth in detention (Figure 8-43) would have been an effective predictor of crime, it was not included as this data was reported at the ZIP code level. As seen in Section 8.2.1.3, ZIP codes are not compatible with PUMAs.

Another example relates to changing geographic boundaries. As mentioned in Section 8.2.1.4, the geographic boundaries delineating police precincts in Staten Island changed in 2013. As far as the author is aware, no historical shape files were available on-line at the time of this study that defined the NYPD precinct boundaries prior to 2018. Data for Staten Island aggregated according to precincts, therefore, was unreliable prior to 2013.

This had repercussions for the data related to police complaints (Data ID 29, Table 7-8), which listed the number of complaints per precinct for the years 2005 to 2009. The spatial join feature in Esri® ArcMap™ was used to convert the geographic unit of this data from precinct to PUMA. However, because the geographic boundaries of precincts in Staten Island prior to 2013 were unknown, the join was based on the current geographic boundaries. The values used for Staten Island therefore represent regional averages at best.

The outcome was more positive for NYPD complaint data (Data ID 10, Table 7-8). Although the data was reported at the precinct level, the coordinates of crimes were also reported. The dataset was therefore more flexible to changing geographic boundaries, as geocodes are easily joined to any shapefile in Esri® ArcMap™.

9.3.2.2 Missing geocodes

As seen in the preceding section, geocoding data is an effective means of retaining the usable of data in the face of changing or incompatible geographic boundaries. However, geocoding rates across datasets varies widely. On average, only around 3.4% of crime events are not geocoded (Data ID 10, Table 7-8). This includes sex crimes, which are not geocoded to protect victim identities. In contrast, a large proportion of 311 requests related to street lights out were not geocoded (Data ID 12, Table 7-8).

The percentage of 311 requests with unspecified coordinates ranged from 3% in Staten Island to 68% in The Bronx. Although street addresses were provided, major costs, both in terms of time and money, would have been incurred in order to geocode over 1 million 311 requests. Valuable data was therefore inaccessible due to this oversight.

9.3.2.3 Sampling bias

A high degree of sampling bias was inherent in a number of the data sources used in this study. For example, 311 requests relating to homeless encampments (Figure 9-3) were concentrated in Manhattan. It is unclear whether this is solely due to increased levels of homelessness in the area, or to an increased level of intolerance for vagrants in high-income areas.

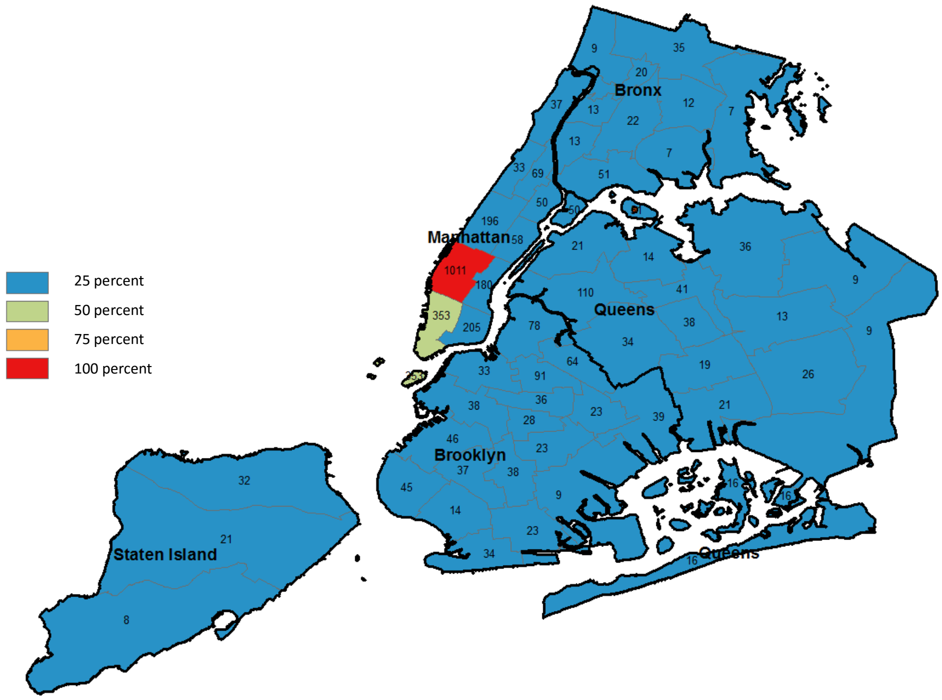


Figure 9-3: Total number of 311 requests related to homeless encampments and panhandling per 100k population per PUMA in 2017 (Data ID 12, Table 7-8).

There was also a high degree of sampling bias inherent in the NYPD complaint reports dataset (Data ID 10, Table 7-8). This dataset includes all valid felony, misdemeanour and violation crimes reported to the NYPD from 2006 to 2017. As noted in the data footnotes downloaded with the dataset, the dataset does not include offenses which do not require a complaint report (e.g. certain drug, trespassing, theft of service, and prostitution offenses). Consequently, these offenses may not be represented accurately in this dataset.

Other datasets containing sampling bias include the inventory of social housing (Data ID 41, Table 7-8) and pedestrian counts (Data ID 19, Table 7-8). The inventory of public housing developments is limited to those under the jurisdiction of the New York City Housing Authority, and exclude a large number of other developments governed by other authorities and NGOs. The bi-annual pedestrian counts provided by the Department of Transport, are only for a small number of select locations, and do not necessarily provided an unbiased measure of pedestrian traffic across the city. Concerns regarding sampling bias were discussed in Section 5.3.5.2.

9.3.2.4 Reporting formats

Data presented as an array of Excel spreadsheets (Data ID 1-6, Table 7-8) was particularly time consuming to clean. This was unnecessary and greatly impacted the time efficiency of research.

9.3.2.5 Reporting standards

Differences in reporting standards across institutions, and across time, proved to be challenging. For example, crime offences in NYPD (Data ID 10, Table 7-8) are reported according to New York State Penal Law definitions. However, many crime indicators, including the ones used in this study (Table 7-3), report crime according to the index crimes which are used in the Federal Bureau of Investigation (FBI) Uniform Crime Reporting (UCR) Program (U.S. Department of Justice Federal Bureau of Investigation, 2018).

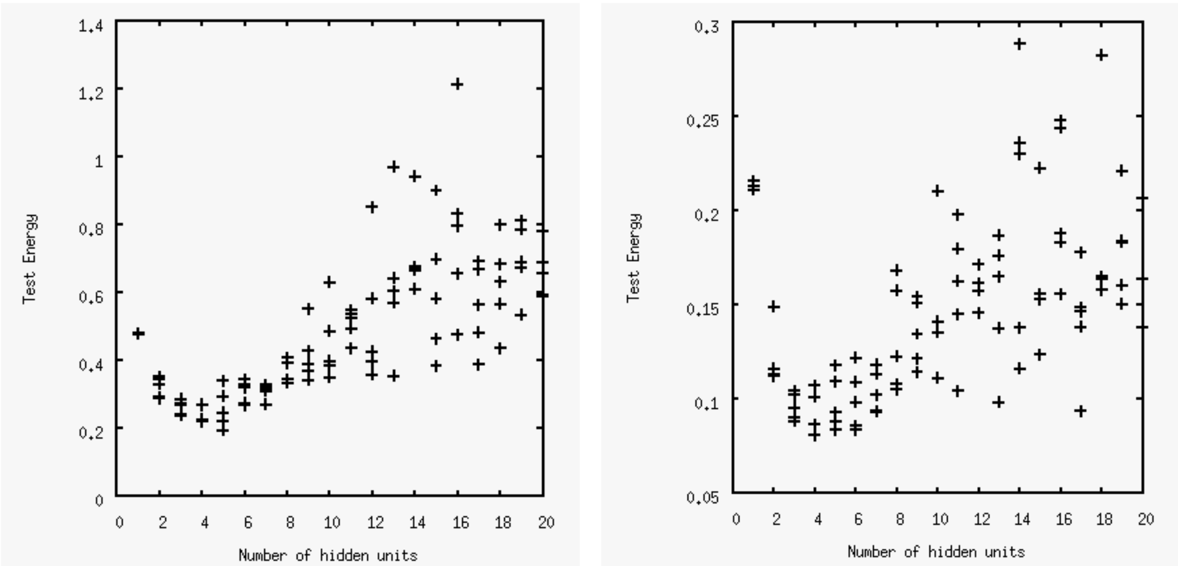
UCR uses standard offense definitions to count and compare crime in localities across America regardless of local variations in crime laws (see Data ID 11, Table 7-8). Reported UCR index crimes include seven crimes which are subdivided into either violent crimes or property crimes. Specifically, index crimes include the violent crimes of murder, rape, robbery and aggravated assault; and the property crimes of burglary, motor vehicle theft and larceny-theft. Since NYPD crime reports were reported according to New York State Penal Law definitions, they needed to be classified according to index crime definitions. This involved manually classifying 560 unique crime identifiers, many of which were difficult to classify based on the information provided.

An example of changing reporting standards over time, was seen in the citywide events dataset (Data ID 14, Table 7-8). The dataset contained a list of information on approved event applications from 2008. Certain events experienced a significant increase in counts in 2013 and again in 2017. It was assumed that this was due to adjustments in reporting methods. For this reason, the data used only represented those events that did not exhibit these spikes in reporting. Again, valuable information was lost.

9.4 Neural network: training and evaluation

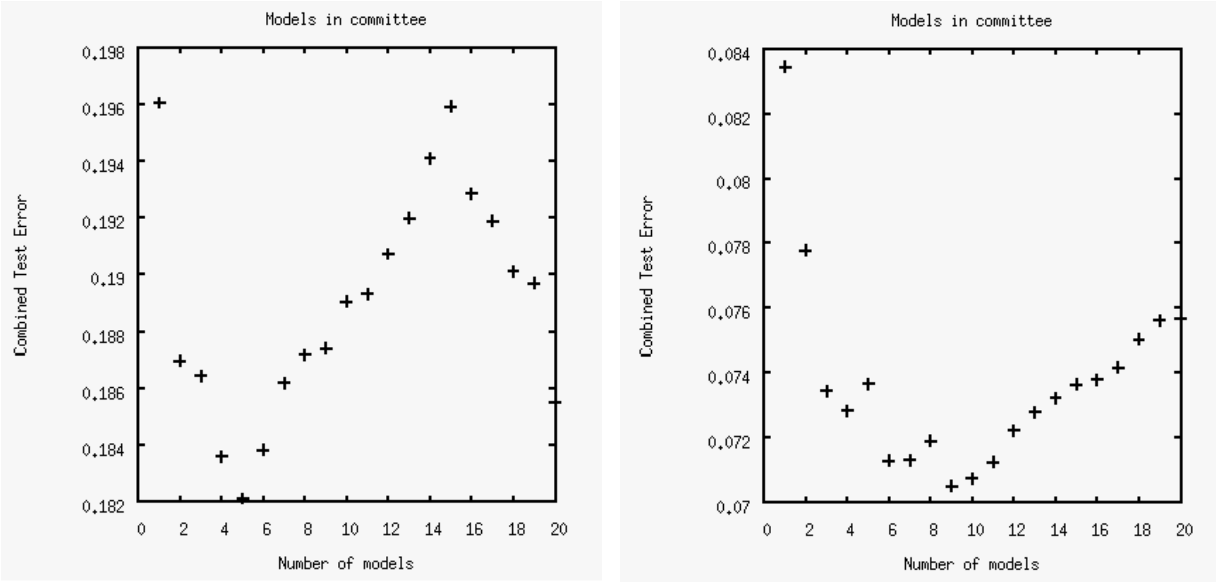
Bayesian Neural Networks were implemented using the Model Manager software package developed by (Sourmail, 2004), and the set of standardised input features listed in Table 9-1. The neural network training and evaluation process was explained in detail in Section 6.4. As noted in Section 9.3.1, variables relating to race (*hispanic, white, black, asian*) and target crimes (*larStreet, larCommercial, larResidence, robStreet, robCommercial, robResidence, assStreet, assResidence*) were not included as input features.

Two models were developed for demonstration purposes; one for predicting street larceny, and another for predicting street robbery. Output from the neural network implementation and evaluation process is shown in Figure 9-4, Figure 9-5 and Figure 9-6. For both predictors, test energy was a minimum for neural networks with around 5 hidden nodes (Figure 9-4). The final model for street robbery consisted of a committee of 5 neural networks, with a combined test error of 0.182 (Figure 7-54a). The final model for street larceny consisted of a committee of 9 neural networks, with a combined test error of just over 0.07 (Figure 7-54b). A visualisation of the accuracy and precision of the final models is shown in Figure 9-6.



(a) Street robbery **(b) Street larceny**

Figure 9-4: Test energy as a function of network architecture. Source: Model Manager standard output.



(a) Street robbery **(b) Street larceny**

Figure 9-5: Combined committee test error as a function of the number of models in the committee. Source: Model Manager standard output.

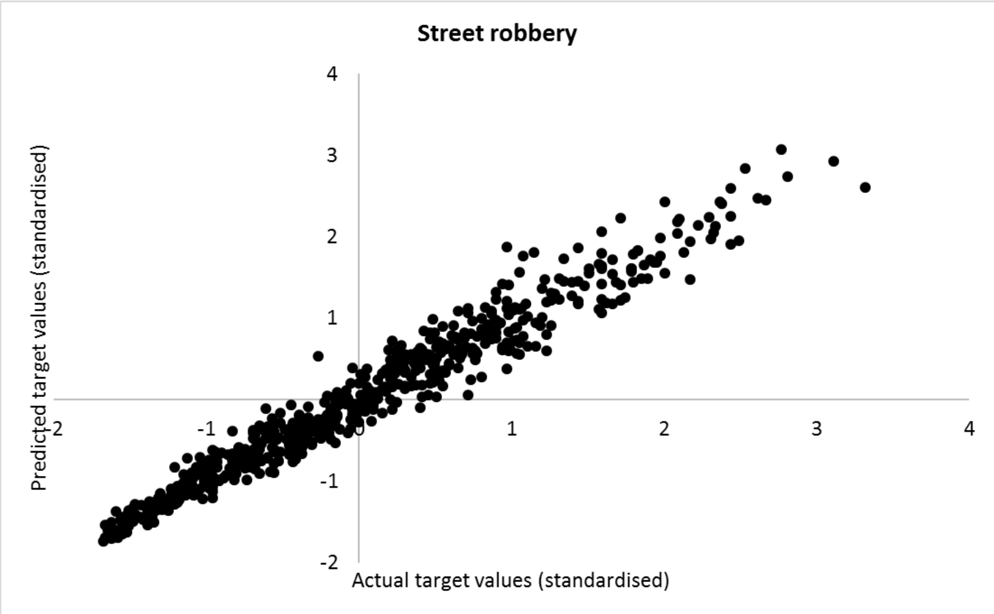
9.5 Exploratory Factor Analysis

During the initial stage of model exploration, it was found that there was a high degree of correlation among input features. The *plot_correlation* function in the *DataExplorer* package in R was used to plot the correlation matrix shown in Figure 9-7. *DataExplorer* creates a correlation matrix by using the *cor* function in base R, which makes use of the Pearson correlation coefficient to determine the linear correlation between two variables.

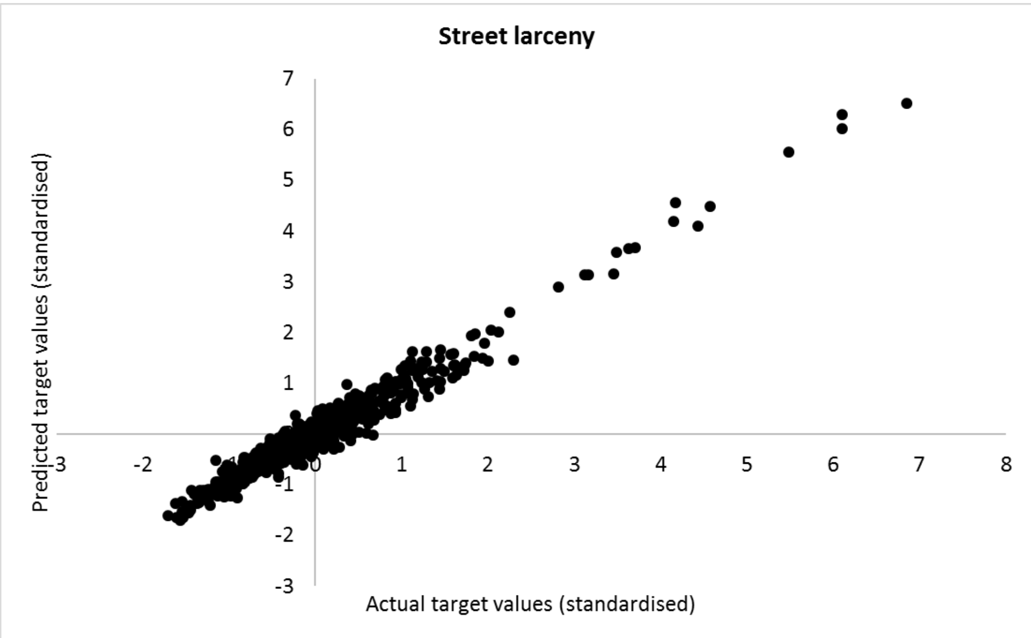
From Figure 9-7, it is apparent that:

A. There is a high degree of correlation between violent crime sub-types, as well as between violent and residential crimes. In addition, these crimes are correlated with socio-economic indictors such as those relating to education, child abuse, drug abuse, inequality and unemployment. Violent and residential crimes also correlate positively with the number of complaints received regarding police integrity.
B. Socio-economic indicators, such as those listed above, are closely correlated with each other. For example, if a region has high unemployment, it is likely to

exhibit higher levels of child abuse, drug abuse, inequality, and lower levels of academic achievement. Socio-economic status is also correlated with the number of complaints regarding police integrity in a given area, and the presence of marginalised communities is associated with an increase in altercations with police.



(a) Street robbery



(b) Street larceny

Figure 9-6: Visualisation of the accuracy and precision of neural network predictions. Source: Author's own construction.

C. Commercial crimes and street larceny are correlated with each other. They also correlate with commercial and transportation land uses, high property values, and high levels of pedestrian traffic, air pollution and homelessness.

From the above observations, it can be deduced that certain types of crime tend to occur alongside each other. For example, the co-occurrences of violent and residential crimes (A), and commercial crime and street larceny (C). Furthermore, crimes tend to be location specific, with particular location features being more prone to specific crime types. It is this strong correlation between crime and location features that has led to the development of the array of place-based predictive policing techniques, discussed in Section 5.3.3.1.

In order to explore the correlations between input features in more detail, Exploratory Factor Analysis (EFA) was used to study the latent structure within the input data. Hair et al. (2014) provides a detailed explanation of the EFA process. Simply put, EFA defines sets of variables, known as factors, that are highly interrelated. These factors are assumed to represent dimensions within the data. That is, it is assumed that the information contained in the original variables can be condensed into a smaller set of new, composite dimensions (factors), with a minimum loss of information (Hair et al., 2014).

EFA was carried out using the *factanal* function in base R. Factors were initially chosen using the latent root criterion (Hair et al., 2014), and further refined using domain knowledge. The VARIMAX orthogonal rotation method was used. Four factors were extracted (Table 9-4): namely A1, A2, A3 and A4. The extracted factors accounted for 59% of the total variance (Table 9-3), which was considered satisfactory for the nature of this study (Hair et al., 2014).

The extracted factor loadings are listed in Table 9-4. Factor loadings of greater than 0.3 are deemed statistically significant for sample sizes greater than 350 (Hair et al., 2014), while factor loadings greater than 0.5 are generally accepted as practically significant (Hair et al., 2014). Significant factor loadings are highlighted in Table 9-4. Positive correlations are highlighted in red, while negative correlations are highlighted in blue.

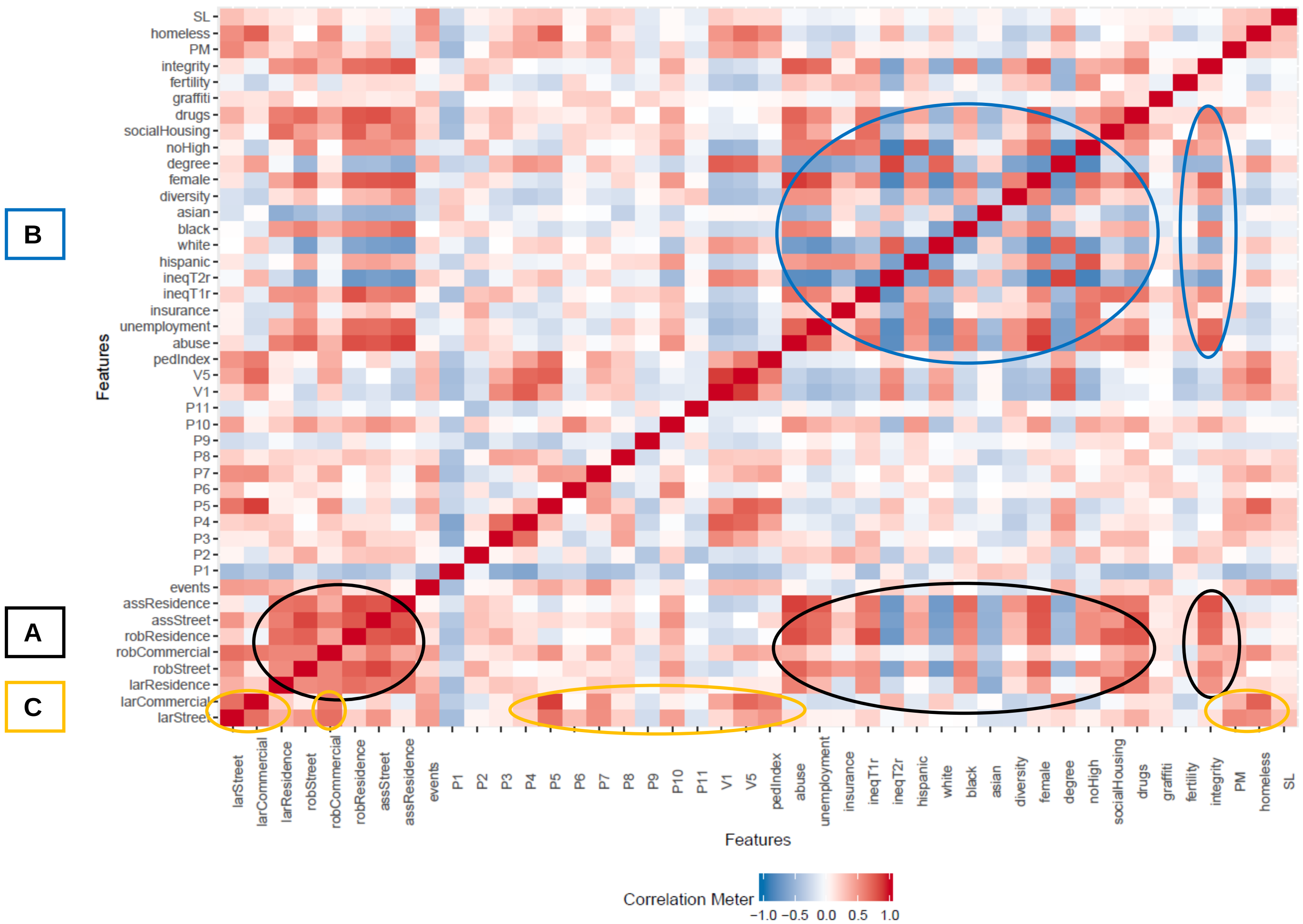


Figure 9-7: Correlation matrix. Variable names and descriptions are listed in Table 9-1.

Table 9-3: Percentage variance explained (n = 660).

	Factors			
	A1	A2	A3	A4
SS loadings	10.531	6.468	5.143	3.727
% of total variance explained	23.9	14.7	11.7	8.5
Cumulative %	23.9	38.6	50.3	58.8

Table 9-4: Extracted factor loadings (n = 660).

	Factors			
Variables	A1	A2	A3	A4
abuse	**0.9**	-0.1	0.3	0.1
asian	**-0.6**	0	0.1	-0.1
assResidence	**0.9**	-0.1	0.1	0.1
assStreet	**0.8**	0.4	0.3	0.1
black	**0.8**	-0.1	-0.2	-0.2
degree	**-0.5**	0.4	**-0.6**	0.4
diversity	**0.5**	-0.3	0.2	-0.3
drugs	**0.7**	0.2	0.3	0.3
events	0.1	**0.5**	-0.2	0.2
female	**0.8**	-0.1	0.4	0
fertility	0.1	-0.2	**0.5**	-0.1
graffiti	0	0.1	0	0.3
hispanic	0.2	0	**0.8**	0.1
homeless	-0.1	**0.7**	-0.2	0.2
ineqT1r	**0.6**	-0.1	0.4	0.4
ineqT2r	**-0.6**	0.2	**-0.7**	0
insurance	0.1	0	**0.7**	-0.2
integrity	**0.8**	0	0.1	0
larCommercial	0	**0.9**	-0.2	0
larResidence	**0.7**	0.2	-0.2	0.3
larStreet	0.2	**0.8**	0.1	0.1
noHigh	0.4	-0.1	**0.9**	0
P1	-0.2	-0.4	-0.1	**-0.8**
P10	0.4	0.2	0.4	0.1
P11	0	-0.2	-0.1	-0.1
P2	0.2	-0.1	0.3	0
P3	0.2	0.1	-0.1	**0.7**
P4	0	0.3	-0.2	**0.7**
P5	-0.1	**0.9**	-0.1	0
P6	0	0.2	0.3	0
P7	0	**0.6**	0.1	0.1
P8	0.2	0.2	0	0.4
P9	0	-0.3	0.1	0.1
pedIndex	-0.1	**0.7**	0	0.2
PM	0	**0.5**	0.1	0.3
robCommercial	**0.5**	**0.7**	-0.1	0
robResidence	**0.8**	0	0.2	0.3
robStreet	**0.8**	0.3	0.3	0
SL	0	0.3	0	0.2
socialHousing	**0.6**	-0.1	0.1	0.4
unemployment	**0.7**	-0.2	0.4	0
V1	-0.3	0.4	-0.4	**0.5**
V5	-0.2	**0.7**	-0.3	0.4
white	**-0.7**	0.1	**-0.5**	0.2

Factor loadings further confirmed the clustering of variables observed in the correlation matrix (Figure 9-7). Similar to what was observed when exploring the correlation matrix, the EFA revealed four highly correlated groups of variables (factors) latent in the input data. These factors were interpreted as city "states", characterised by the types of crime and socio-economic variables that tended to cluster together. The four states identified in this study are illustrated in Figure 9-8, and described below:

- *A1: Collective Efficacy 1:* This state strongly correlated with violent and residential crimes, as well as with socio-economic factors such as drug abuse, single parent households, child abuse, unemployment and inequality. This state also strongly correlated with race and altercations with police. This state, therefore, particularly characterised the challenges faced by many Black communities in NYC. As noted in Section 5.3.4, socioeconomic and demographic indicators such as poverty and race are consequently often used as risk-factors for crime (Sampson, 2006; Taylor et al., 2015). However, the theory of collective efficacy (Bandura, 2000; Browning, 2002; Sampson, 2006) advocates a shift away from community-level indicators such as race, and aims to focus on the underlying social mechanisms at work within high-crime neighbourhoods. The paradigm of collective efficacy was adopted in this study. Consequently, the developed model did not include race as an input feature, but focused on underlying contributors to crime such as abuse or unemployment.

- *A2: Commercial Land Use:* This state correlated with commercial crimes and street larceny; and was characterised by commercial and transportation land uses, high property values, and high levels of pedestrian traffic, air pollution and homelessness.

- *A3: Collective Efficacy 2:* This state represented the challenges faced by a number of Hispanic communities in NYC, and correlated with features such as fertility, no health insurance, and no high school diploma.

- *A4: Mixed Land Use:* This state correlated with multi-family elevator buildings, mixed residential and commercial buildings, and high property values.

Figure 9-8: Illustration of NYC attractor states.

The clustering of state variables observed above is analogous to the concept of basins of attraction in complex dynamic systems. A basin of attraction, as illustrated in Figure 9-9, is a region in state space in which the system tends to remain (Walker *et al.*, 2004; Westley *et al.*, 2011). This equilibrium state represents the current system regime. Westley et al. (2011: 18) define a regime as “the dominant rule-sets supported by incumbent social networks and organisations and embedded in dominant artifacts and prevailing infrastructures, of say, particular industries or social problem arenas.”

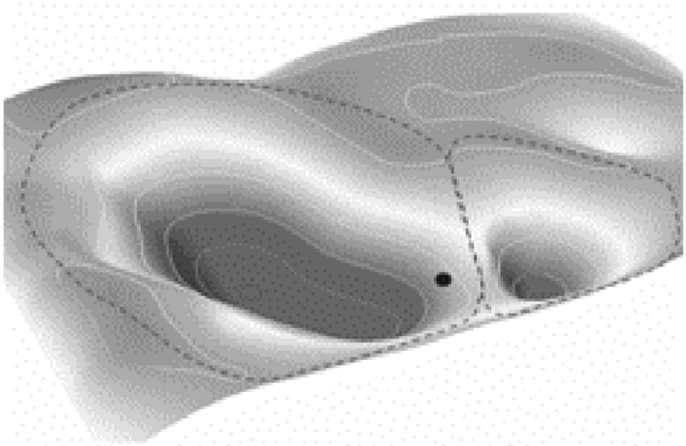


Figure 9-9: Illustration of a basin of attraction. Source: Walker *et al.* (2004).

Change research into complex socio-ecological systems studies the nature and behaviour of attractor basins. Particular attention is focused on understanding the underlying interactions causing the emergence and dissolution of attractor states. Much work has been done on applying the concepts of attractor basins and the associated resilience theory to real-world problems (Walker *et al.*, 2004; Westley *et al.*, 2006; Sendzimir *et al.*, 2007; Gundry *et al.*, 2011; Westley *et al.*, 2011; Westley *et al.*, 2015). Further exploration of these concepts may prove useful in future studies focusing on the modelling of complex city systems.

In order to gain a visualisation of attractor basins across NYC, factor scores for each attractor state was calculated for each PUMA in 2006 and 2017, respectively. Factor scores were calculated by averaging over strongly loading input features as delineated in Table 9-4 (Hair *et al.*, 2014). Specifically, an average of the following standardised input features was used to calculated a factor score for each attractor state:

- *A1: Collective Efficacy 1:* drugs, female, abuse, unemployment
- *A2: Commercial Land Use:* P5, V5
- *A3: Collective Efficacy 2:* insurance, noHigh, fertility
- *A4: Mixed Land Use:* V1, P3, P4

The dominant state in each PUMA was then determined, by selecting the state with the highest factor score. The dominate state per PUMA is shown in Figure 9-10 and Figure 9-11 for 2006 and 2017, respectively. Only factor scores greater than 0.5

were used, as these were considered to be sufficiently deviated from the mean. Figure 9-10 and Figure 9-11 were then compared with the crime trends observed in Section 8.4, which showed good agreement between the crimes associated with the various attractors, and the spatial trends observed in Section 8.4.

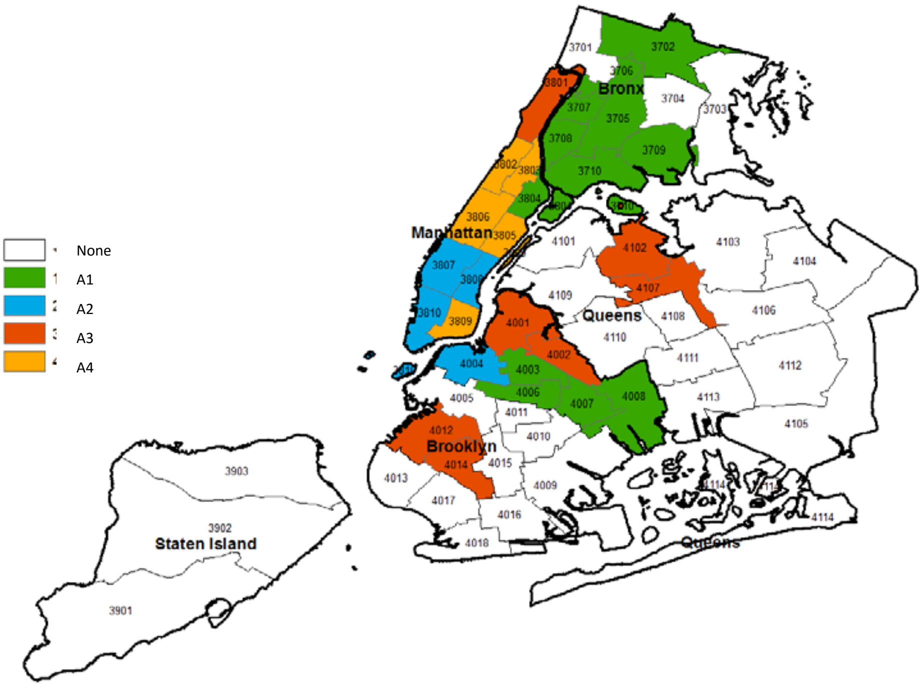


Figure 9-10: Dominant attractor state per PUMA in 2006. Source: Author's own construction.

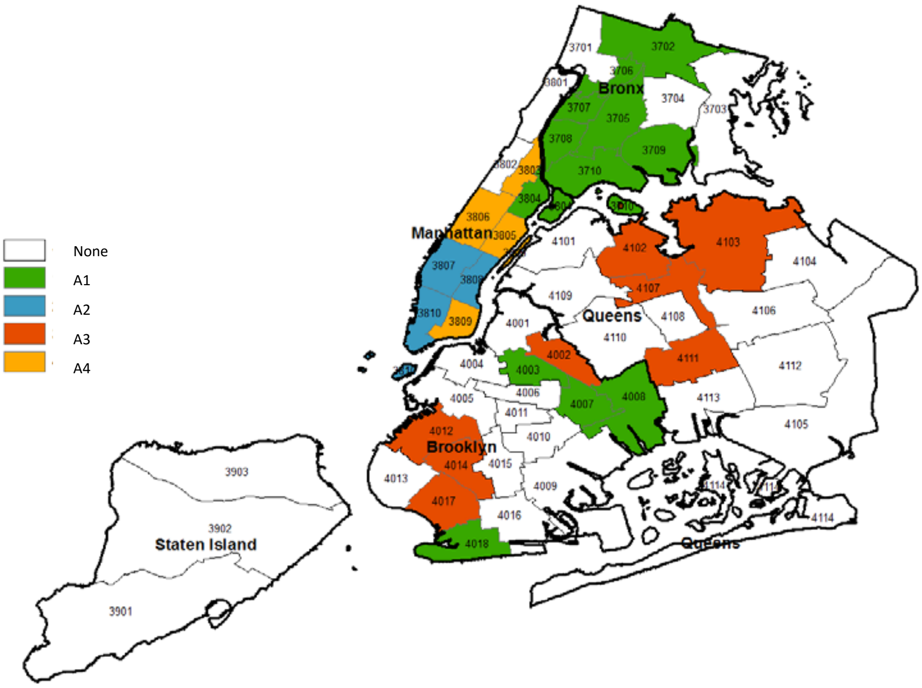


Figure 9-11: Dominant attractor state per PUMA in 2017. Source: Author's own construction.

9.6 Sensitivity analysis

9.6.1 Analysis constraints

The anticipated application of the prototype crime predictors developed in Section 9.4, is to be a tool for sensitivity analysis aimed at determining the relative influence of decision variables on crime (Section 7.3). The ultimate purpose of the model is to be a tool for integrated decision-making, aimed at optimising decisions at the system-of-systems level. Thereby, improving the efficiency of decisions, and minimising unexpected externalities (Section 7.2).

Typically, as part of a sensitivity analysis, an input variable is varied while the remaining input variables are fixed at specific values (Section 6.5.1). The changes in the network output are then observed. Concerning sensitivity analysis, there are two anticipated repercussions from the existence of attractor states (Figure 9-8) observed in Section 9.5:

A. Because different state regimes are governed by different "rules" (Westley *et al.*, 2011), they are likely to respond differently to the same set of input features. In order to meaningfully interpret the results of a sensitivity analysis, the investigator needs to know which state the system under investigation is in (Westley *et al.*, 2015).
B. Because different basins of attraction correspond to specific regions in state space (Walker *et al.*, 2004; Westley *et al.*, 2011), they are likely to be under-represented in regions of input space which fall outside of these domains. Consequently, there will be a higher degree of uncertainty in neural network predictions in these regions, as less data will be available to train the network.

In order to address the anticipated challenges highlighted above, sensitivity analyses were carried out for the two models developed in Section 9.4, separately for each of the four attractor states identified in Section 9.5. This was achieved by fixing each input feature to its average value within a particular state. For any input tuple, the tuple was deemed to be in an attractor state if the factor score for that state was above 1. The calculation of factor scores was explained in Section 9.5.

For a given sensitivity analysis, each input feature was independently varied while the remaining input features were fixed. Each input feature was varied within the numerical range exhibited by that feature within the state under investigation. It was anticipated that varying the input feature outside of this range would result in high levels of model uncertainty, due to limited representation of these ranges in the input space. In contrast, in order to test the validity of the above assumptions, sensitivity analyses were also performed for each model, using the citywide averages and ranges for each input feature. Thereby, testing the feasibility of disregarding attractor states.

9.6.2 Demonstration

The results of the sensitivity analyses for street larceny and street robbery are presented as a set of bar charts in Figure 9-12 and Figure 9-13, respectively. The charts indicate the predicted changes in crime rate per unit change in input features. Bar charts are plotted for each attractor state, as well as for analyses performed using citywide data.

In order to demonstrate the application of sensitivity analysis in decision-making, three scenarios will be explored; namely, street larceny and street robbery in the A2 state, and street robbery in the A1 state. The sensitivity of each crime to the various input features, shown schematically in Figure 9-12 and Figure 9-13, are listed in Table 9-5. The responsible NYC agency and indicator ID (as listed in Table 7-3) are also specified for each feature in Table 9-5. Only immediately actionable variables are listed in Table 9-5. Features relating to land use (P1-P11) and property values (V1 and V5), while valuable indicators of the location of crime, were not included in this analysis.

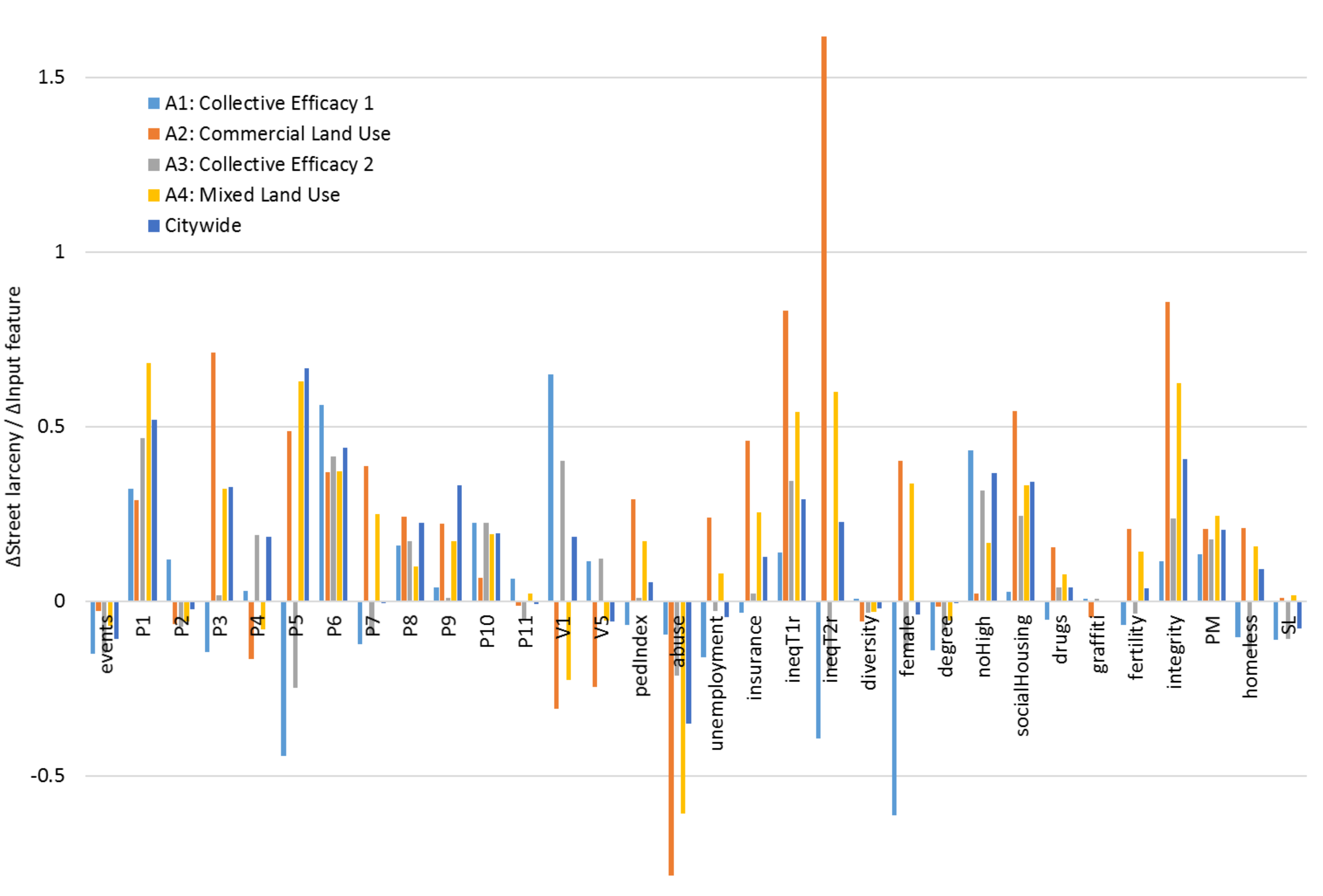


Figure 9-12: Sensitivity plot for street larceny.

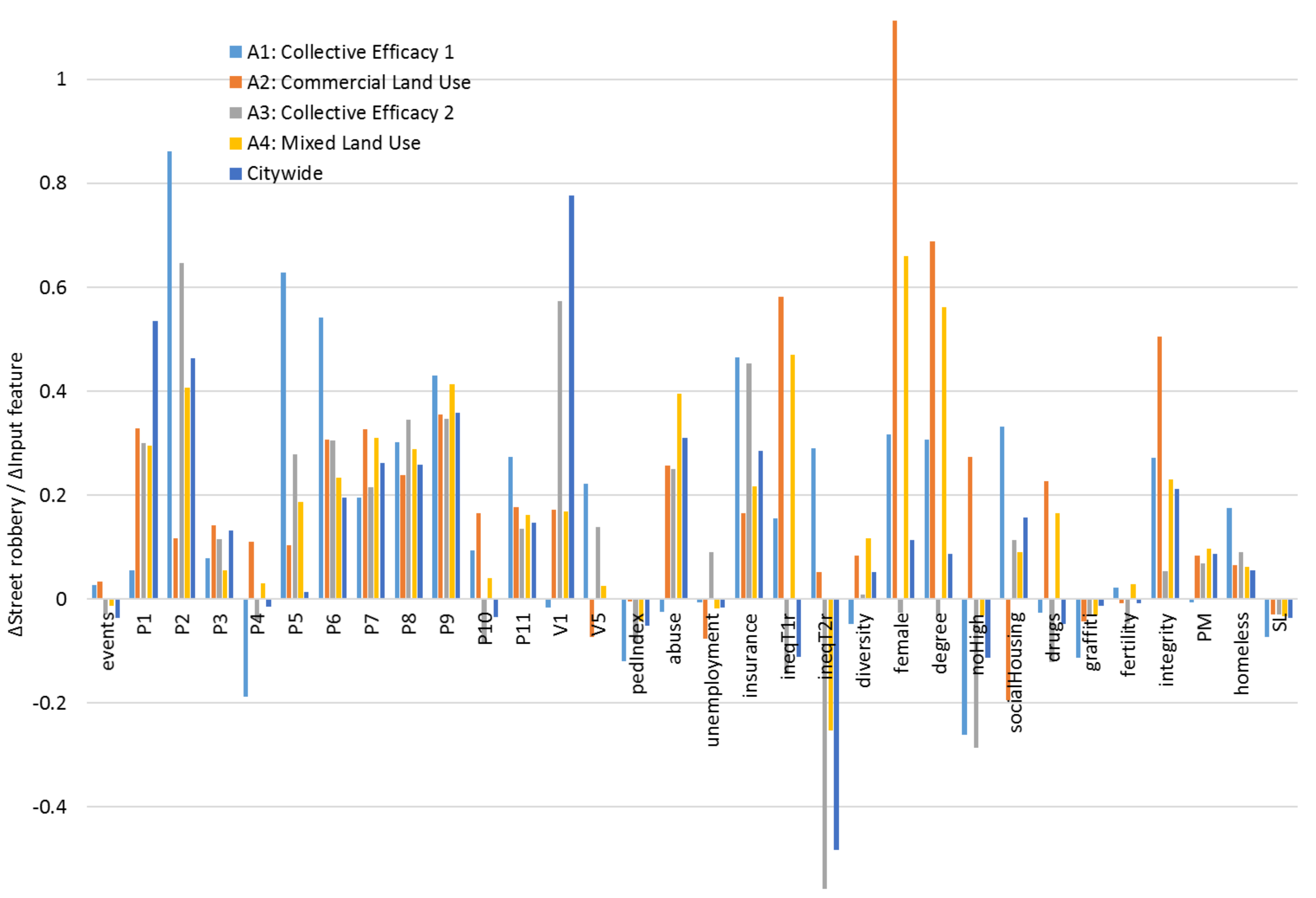


Figure 9-13: Sensitivity plot for street robbery.

Table 9-5: Sensitivity of crime rates to input features for areas in the A1 and A2 attractor states. Identified intervention options are highlighted in blue.

ID	Agency	Variable name	Street larceny (A2)	Street robbery (A2)	Street robbery (A1)
1	CECM	events	-0.03	0.03	0.03
3	MOEO	unemployment	0.24	-0.08	-0.01
4	MOEO	ineqT1r (within) ineqT2r (between)	0.83 1.62	0.58 0.05	0.16 0.29
7	DOE	noHigh	0.02	0.27	-0.26
8	DOE	degree	-0.02	0.69	0.31
11	DYCD	female	0.4	1.11	0.32
12	ACS	abuse	-0.79	0.26	-0.03
13	DYCD	fertility	0.21	-0.01	0.02
14	CCRB	integrity	0.86	0.5	0.27
15	HRA; DOHMH	insurance	0.46	0.17	0.47
16	DOHMH; NYPD	drugs	0.15	0.23	-0.03
20	DHS	homeless	0.21	0.06	0.18
21	NYCHA	socialHousing	0.54	-0.2	0.33
22	DOT	SL	0.01	-0.03	-0.07
26	DCP	diversity	-0.06	0.08	-0.05
36	DOT	pedIndex	0.29	0	-0.12
37	NYPD; DSNY; EDC	graffiti	-0.05	-0.04	-0.11
38	DOHMH; DEP	PM	0.21	0.08	-0.01

For street larceny in the A2 state, the following can be deduced from reviewing the results of the sensitivity analysis presented both in Figure 9-12 and Table 9-5:

- Street larceny appears to be driven by a high concentration of wealthy targets. This is evident in its sensitivity to *ineqT2r*, which indicates that street larceny tends to be elevated in wealthier areas. Furthermore, street larceny is sensitive to indicators relating to high densities of targets, such as pedestrian traffic (*pedIndex*) and air pollution (*PM*).
- Similarly, street larceny appears to be driven by inequality within an area. This is reflected in its sensitivity to *ineqT1r* and *socialHousing*. Street larceny in areas characterised by the A2 state, therefore, can possibly be reduced by addressing inequality in the area. Likely courses of action could address the social efficacy indicators that street larceny is sensitive to in these areas, such as *unemployment*, single-parent households (*female)*, high *fertility* rates, a lack of health *insurance* and abuse of *drugs*. Street larceny is also sensitive to *homeless*ness.
- Features having little effect on street larceny include *events*, education (*noHigh* and *degree*) and street lights (*SL*).
- Increasing levels of *abuse* leads to a reduction in street larceny.

- An increase in *graffiti* and *diversity* tend to decrease street larceny. The reason for this is unclear, however it is possible that areas characterised by these features could represent poorer areas with less wealthy targets.
- The rate of street larceny is sensitive to the number of complaints regarding police officers (*integrity*). This could be an indication that crime is over reported in these areas due to over policing (see Section 5.3.5.2).

Based on the discussion above, and the sensitivities listed in Table 9-5, the following top three interventions can be proposed to decrease street larceny in areas exhibiting the A2 state:

- Increase access to health insurance. This may also have a positive impact on drug-related crimes and homelessness.
- Provide support for women, regarding single parenthood and family planning.
- Create employment opportunities.

For street robbery in the A2 state, the following can be deduced from reviewing the results of the sensitivity analysis presented both in Figure 9-13 and Table 9-5:

- Street robbery appears to be less driven by the absolute wealth of an area (*ineqT2r*), but rather seems to be driven by inequality within an area (*ineqT1r*).
- It appears that street robbers target educated communities (*degree*) with lower levels of unemployment (*unemployment* and *socialHousing*). Street robbers also appear to target busy areas (*PM*) with diverse populations (*diversity*).
- Likely courses of action could address the social efficacy indicators that street robbery is sensitive to in these areas, such as education (*noHigh*), single-parent households (*female)*, *abuse*, a lack of health *insurance* and abuse of *drugs*. Street robbery is also sensitive to *homeless*ness.
- Features having little effect on street robbery include *events*, *fertility,* pedestrian traffic (*pedIndex*) and street lights (*SL*).
- The rate of street robbery is sensitive to the number of complaints regarding police officers (*integrity*). This could be an indication that crime is over reported in these areas due to over policing (see Section 5.3.5.2).

Based on the discussion above, and the sensitivities listed in Table 9-5, the following top interventions can be proposed to decrease street robbery in areas exhibiting the A2 state:

- Provide support to families, regarding single parenthood and abuse.
- Support youth in completing their high school education.

For street robbery in the A1 state, the following can be deduced from reviewing the results of the sensitivity analysis presented both in Figure 9-13 and Table 9-5:
This state seems to be less sensitive to input features, with a larger range of features having little impact on crime. These are namely *events*, *unemployment*, *abuse*, *fertility*, *drugs* and *PM*. It was found that areas characterised by low levels of community efficacy tended to be less sensitive to input features in general. This is shown in

- Table 9-6, which shows the average sensitivity to input features for each state and each crime. This may be an indication that areas with low levels of social efficacy are more difficult to change.
- As with the previous crimes, street robberies tend to be driven by inequality (*ineqT1r, ineqT2r* and *socialHousing*), targeting educated communities (*degree* and *noHigh*).
- Responsive social efficacy measures include single-parent households (*female)*, a lack of health *insurance* and *homeless*ness.
- The rate of street robbery is sensitive to the number of complaints regarding police officers (*integrity*). This could be an indication that crime is over reported in these areas due to over policing (see Section 5.3.5.2).
- Street robberies appear to be sensitive to the outage of street lights. Street robberies tend to decrease as the number of reported outages increase. At first glance, this seems contradictory. However, the number of reports are likely positively correlated with the number of street lights in a given area. Street robberies, therefore, likely decrease as the number of working street lights in an area increase.
- Street robberies appear to be located away from areas characterised by *graffiti*, *diversity* and pedestrian traffic (*pedIndex*).

Based on the discussion above, and the sensitivities listed in Table 9-5, the following high-level interventions can be proposed to decrease street robbery in areas exhibiting the A1 state:

- Provide support to single mothers.
- Increase access to health insurance.
- Ensure that the number of street lights are sufficient and in good working order.

Table 9-6: Average sensitivity to input features by state.

State	Street robbery	Street larceny
A1	0.09	-0.06
A2	0.2	0.27
A3	-0.01	0.01
A4	0.14	0.16
Citywide	0.02	0.08

Based on the analysis above, agencies deemed to have an impact on crime (Figure 9-14), in addition to the NYPD, include:

- The Human Resources Administration (HRA) and the Department of Health and Mental Hygiene (DOHMH). Both these agencies may have an impact on health insurance coverage.
- The Department of Youth and Community Development (DYCD) and the Administration for Children's Services (ACS). These agencies may be able to assist families regarding single parenthood, abuse and family planning.
- The Department of Education (DOE) may be able to assist youth in completing high school.
- The Mayor's Office for Economic Opportunity (MOEO) has an impact on job creation.
- The Department of Transportation (DOT) is in charge of street lights.
- The Civilian Complaint Review Board (CCRB) monitors complaints against the police, and can play a role in combatting discrimination within the policing system.

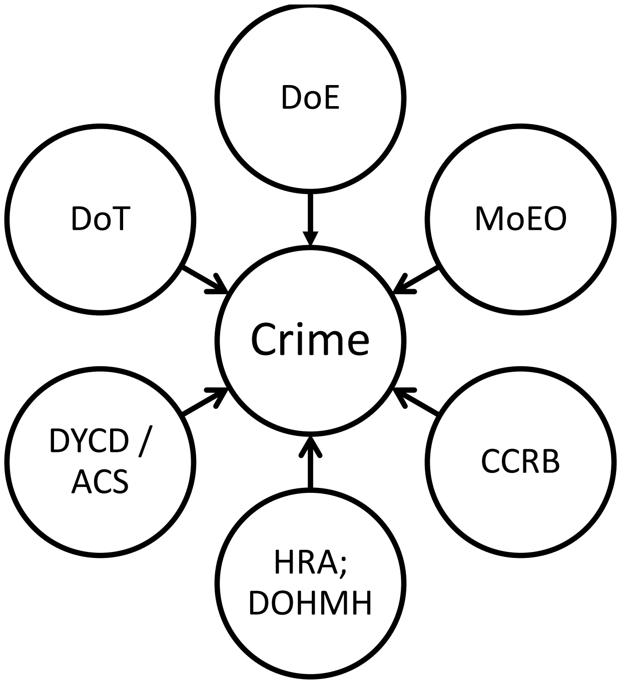


Figure 9-14: NYC agencies identified as key to combatting street robbery and larceny. Acronym descriptions are listed in Table 7-2. Source: Author's own construction.

In order to investigate the effect of disregarding attractor states, the sensitivity of crime to input features was investigated in more detail by observing the change in crime rates as a function of the respective input features. This is illustrated in Figure 9-15 to Figure 9-18 for selected features. In these figures, trends are plotted for each attractor state, including citywide trends. Error bars for each plot are indicated by dashed lines of the same colour. These errors formed part of the Bayesian Neural Network output (Section 6.4.1.5), and indicate sparsity and noise in input data.

When all states exhibit similar behaviour, citywide trends are sufficient to correctly interpret sensitivity analysis results. Example features include health insurance (Figure 9-15) and street lights (Figure 9-16). The relatively large error bars for the A2 and A4 states in Figure 9-16, indicate that limited data was available for such high rates of street light outages.

For street larceny (Figure 9-17), increasing unemployment in states A2 and A4 leads to an increase in crime. However, in areas with low collective efficacy (A1 and A3), an increase in unemployment leads to a decrease in crime. As mentioned earlier, this is likely related to a decrease in inequality in affected areas. Due to the high concentration of unemployment in states A1 and A3, citywide trends shield the effects of increasing crime in more wealthy areas. The low rates of unemployment in wealthy areas are reflected in very large error bars for the A2 and A4 states.

Increasing levels of abuse (Figure 9-18), tend to increase street robberies in all states, except in state A1. A small decrease in street robberies is observed in A1 states as abuse rates reach the upper limit for NYC. This could perhaps indicate that the effects of abuse reach a plateau for high levels of abuse. Citywide trends do not capture this trend. However, it does show a very high error bar for high abuse levels, indicating that the citywide predictions should be interpreted with caution.

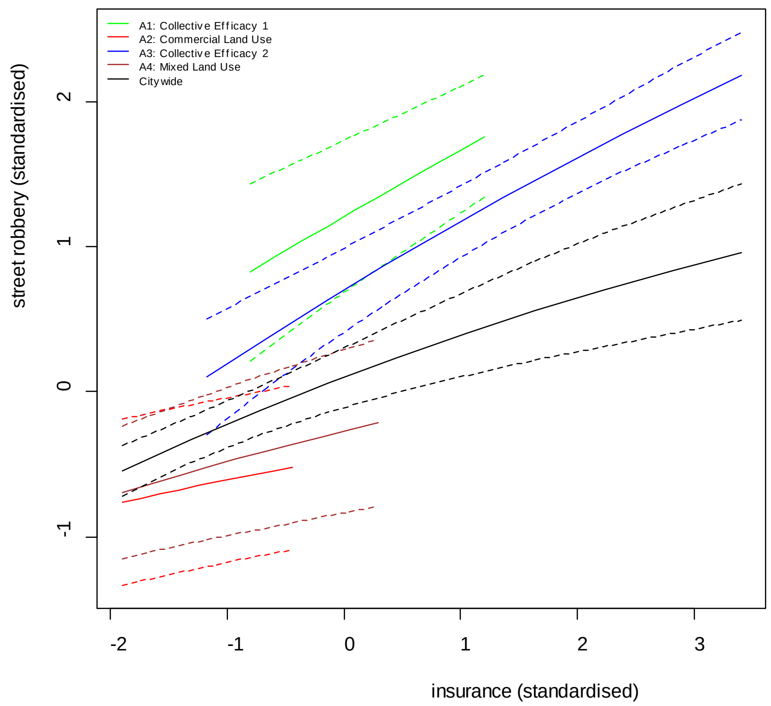


Figure 9-15: Street robbery as a function of the percentage of civilians without health insurance.

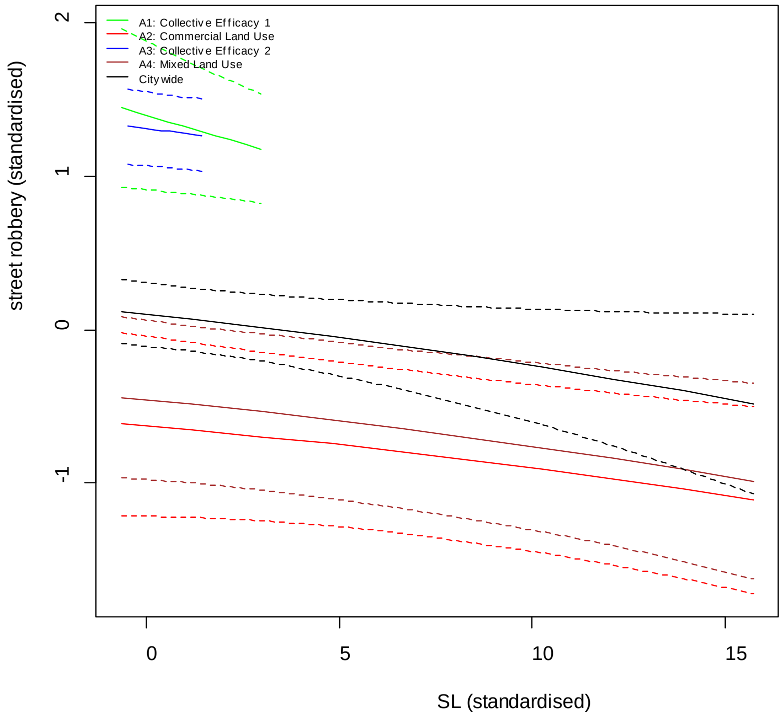


Figure 9-16: Street robbery as a function of the number of street lights out reported.

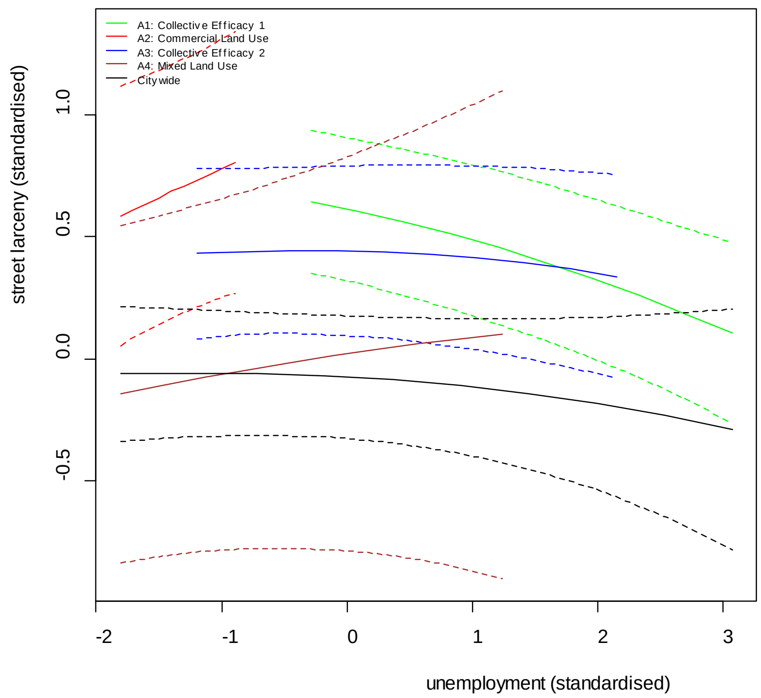


Figure 9-17: Street larceny as a function of the unemployment rate.

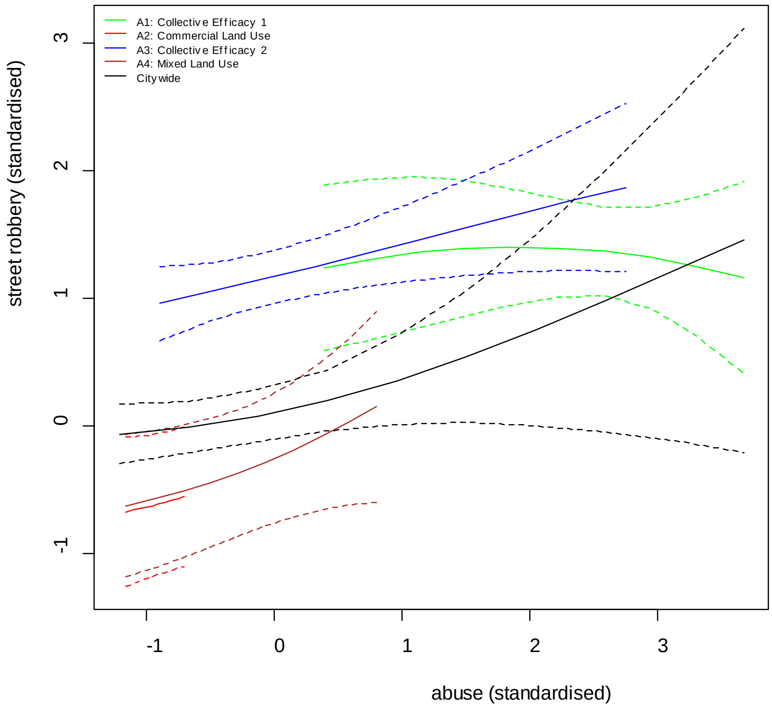


Figure 9-18: Street robbery as a function of abuse.

9.6.3 Highlights

The preceding demonstration showed the following:

- The anticipated application of the prototype model(s) was demonstrated.
- It was shown how the prototype model(s) could effectively be used to identify the government agencies that will have the most impact on crime.
- It was demonstrated that different attractor states respond differently to the same set of input features. Furthermore, it was shown that disregarding the existence of attractor states can lead to misleading results.
- It was consequently shown that the most effective use of the prototype model(s) are achieved when latent system states are first identified with the use of EFA; followed by contextualised sensitivity analyses within respective city states.

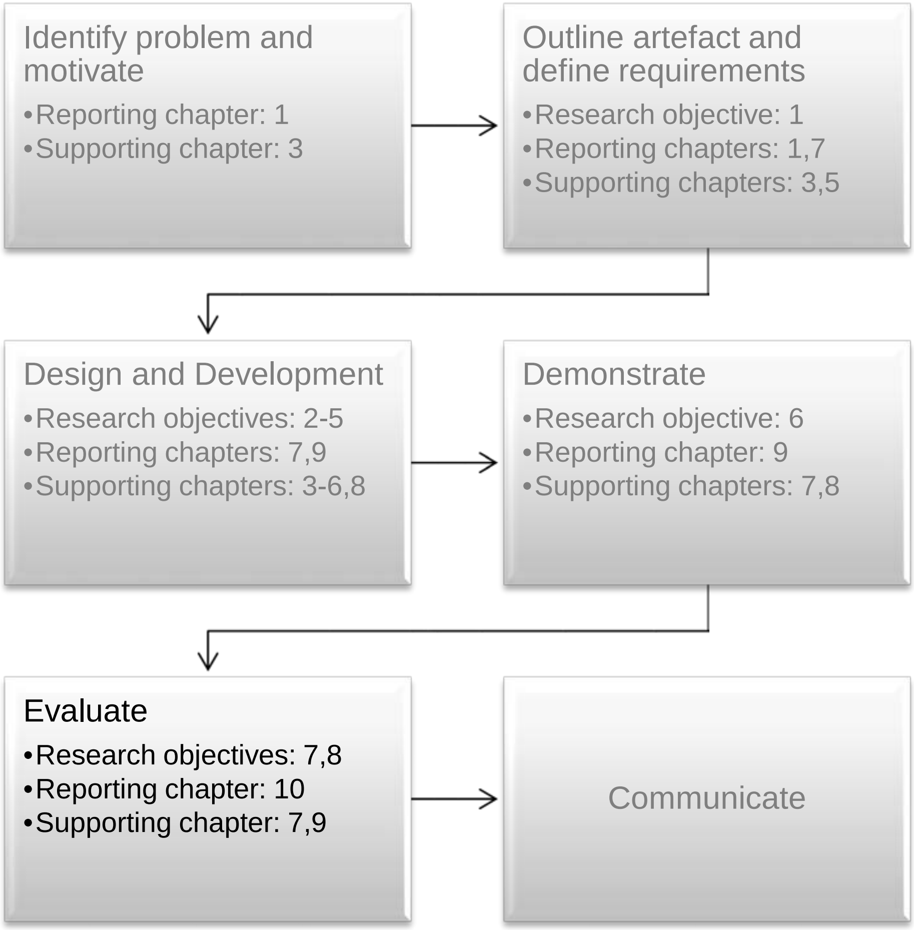
9.7 Summary

The main objective of this chapter was to develop and demonstrate a prototype model for crime management in smart cities according to the design specifications laid out in Chapter 7 of this thesis. This chapter therefore addressed $RO_{5\text{-}6}$ (Figure 9-1).

An overview of the implementation process was given Section 9.2. This was followed by a list of the data used to develop the model (Section 9.3.1), together with a summary of the challenges faced during the data preparation process (Section 9.3.2). Neural network implementation was then summarised in Section 9.4, followed by a detailed demonstration of the model interpretation process (Section 9.5 and 9.6). Two models were developed for demonstration purposes, one for street larceny and one for street robbery.

Due to the limited accessibility of South African data at the time of this study, readily accessible open data for New York City was used to develop and demonstrate the set of prototype models (Section 7.4). In the following chapter, the developed models will be evaluated *ex ante* within the South African context by way of a mixed-method case study.

Chapter 10. Model Evaluation



Research objectives addressed in this chapter:

RO_1: Identify the functional, construction and environmental requirements of an effective model.

RO_2: Identify relevant input and output parameters.

RO_3: Identify and characterise available data sources.

RO_4: Identify the modelling technique to be used to develop the model.

RO_5: Develop the model.

RO_6: Demonstrate the application of the model.

RO_7: Evaluate the efficacy of the model.

RO_8: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model.

Figure 10-1: Research objectives and design science research activity addressed in this chapter.

10.1 Introduction

The main objective of this study was to develop and evaluate a prototype model for crime management in smart South African cities (Section 1.5). In the previous chapter, a set of prototype models were developed and demonstrated. The main objective of this chapter (Figure 10-1) is to evaluate the efficacy of the developed model (RO_7), and to develop a set of implementation guidelines based on knowledge derived from the implementation and evaluation of the prototype model (RO_8). The evaluation approach is described in Section 10.2. This is then followed by model evaluation in Section 10.3, and a list of implementation guidelines in Section 10.4.

10.2 Evaluation approach

The purpose of model evaluation (Figure 10-1) is to evaluate the developed model based on the model requirements specified in Section 7.2 (Peffers *et al.*, 2008; Johannesson and Perjons, 2012). As such, the evaluation process aims to assess the efficacy of the model in meeting each model requirement.

The main objective of this study was to develop and evaluate a prototype model for crime management in smart South African cities (Section 1.5). However, due to the limited accessibility of South African data at the time of this study, readily accessible open data for New York City was used to develop and demonstrate the prototype model (Section 7.4). The developed model was therefore evaluated *ex ante* (Peffers *et al.*, 2008) within the South African context, by way of a mixed-method case study.

Nelson Mandela Bay Municipality (NMBM) was selected as the most feasible metro to investigate as research agreements were already in place between Nelson Mandela University and the NMBM. The efficacy of the developed model was evaluated by way of an interview with NMBM employees, and a document study of key reports such as mid-term (NMBM., 2018a) and annual reports (NMBM., 2017), performance contracts (NMBM., 2018c) and IDP documentation (NMBM., 2018b).

To this end, a meeting was held on 20 November 2018 with members of the NMBM ICT department to discuss the reporting framework in place within the bay. Specific focus was given to understanding the adopted indicator framework, methods of reporting, as well as the supporting data sources. Meeting documentation, including the agenda, supporting documents and attendance list, is given in Appendix 1.

10.3 Evaluation

10.3.1 Requirement 1: KPI framework

Requirement 1 prescribed that the solution incorporate a set of KPIs that quantitatively represent all stakeholder and sustainability considerations. The emerging smart city conceptual models and KPI frameworks discussed in Section 4.4 delineate key smart city components and their desired states, and therefore provide a potential design solution to Requirement 1. It was proposed (Section 7.3) that a subset of the City Protocol Society's *City Anatomy Indicators* (Section 4.4.2) be used as input to the prototype model; and that the commonly employed predictors of crime summarised in Section 5.3.4 be used as a guide when selecting CPA KPIs.

The intended use of the KPIs in the modelling process was explained in Requirement 2. It was explained that, in order to truly integrate stakeholder and sustainability concerns into model predictions, the KPI framework would be used in a manner comparable to the use of state variables in the modelling of dynamic systems. Consequently, in order for the proposed KPI framework to be effective, there needs to be sufficient overlap between predictors of crime (Section 5.3.4) and the CPA indicator framework (Section 4.4.2).

The indicators used in this study are listed in Table 7-3. As specified in Section 7.3, indicators were selected from the City Protocol Society's *City Anatomy Indicators* (Section 4.4.2), guided by the commonly employed predictors of crime summarised in Section 5.3.4. For the most part, the indicators were not exact implementations of CPA indicators, but rather they were adapted according to closely related crime predictors and available data.

Out of the 22 indicators used in this study, only four indicators were not associated with a related CPA indicator. In these instances, new indictors had been created due to the availability of relevant data for which no CPA indicator exists. Specifically, these four indicators related to the prevalence of single mothers (ID 11), child abuse (ID 12), drug crimes (ID 16) and graffiti (ID 37). Overall, however, there was sufficient overlap between crime predictors and CPA indicators to warrant the implementation of the proposed solution.

The good predictive accuracies (Section 9.4) of the prototype models further support the notion that the subset of CPA indicators used as input to the models sufficiently represent key predictors of crime. It is acknowledged, however, that the non-CPA indicators mentioned above had a strong influence on crime (Section 9.6.2), which would have contributed to the accuracy of the models.

Another consideration in evaluating the efficacy of the KPI framework relates to the diversity of NYC agencies included in the framework. The anticipated outcome of model implementation is to include a diverse range of stakeholders in the decision-making process, thereby fostering synergistic solutions that are often overlooked when solutions are solved within sectoral silos.

In order to visualise the diversity of domains included in the development of the prototype model, the associated CPA domain and NYC agency are specified in Table 7-1 and Table 7-2 for each indicator, respectively. KPIs from 18 different NYC agencies were incorporated into the prototype models developed in this study (Table 7-2). Of these, eight agencies, in addition to NYPD, were shown to have a significant impact on the crimes explored (Figure 9-14). The models therefore effectively brought together traditionally overlooked agencies in the fight against crime, such as the Department of Transportation and the Department of Health and Mental Hygiene.

Lastly, the practical efficacy of the KPI framework depends on the readiness of cities to adopt the framework. For New York City, it was found that 73% of the selected indicators (Table 7-3) can be related to similar indicators reported on annually in the Mayor's Management Report (de Blasio *et al.*, 2018) and the NYC Social Indicators Report (MOO., 2018). Similarly, it was found that 59% of the indicators used can be related to similar indicators annually reviewed as part of the NMBM IDP process

(NMBM., 2018b). Comparable KPI frameworks were therefore already actively implemented within the investigated cities.

In conclusion, the KPI framework implemented in this study proved to be an effective solution to Requirement 1. There was sufficient overlap between the predictors of crime (Section 5.3.4) and the CPA indicator framework (Section 4.4.2) to warrant the implementation of the proposed solution. Furthermore, the selected indicators represented a diverse range of stakeholders, supporting the development of synergistic cross-sector solutions. Lastly, the solution was practically ready to implement, as comparable KPI frameworks were already in use within the investigated cities.

10.3.2 Requirement 2: Predictive model

Requirement 2 prescribed that the solution incorporate a predictive model that can be used as a reliable tool for systems-level scenario analysis. Specifically, it was prescribed that the model take as input the KPIs identified in Requirement 1, and be used to predict the relative influence of selected KPIs on crime. The model was to incorporate any known or unknown complexities and inter-dependencies between variables. In addition, predictions were to be accurate and precise, and a measure of prediction uncertainty was to be provided.

It was proposed that a combined approach employing Bayesian Neural Networks (Chapter 6) and sensitivity analysis (Section 6.5.1) be used to develop the prototype model. The proposed approach was implemented and demonstrated in Chapter 9. Two neural networks, one predicting street robbery and the other predicting street larceny, were developed for demonstration purposes (Section 9.4). The intended application of the prototype models as a tool for sensitivity analysis was demonstrated in Section 9.6.

From the model implementation, it was determined that Bayesian Neural Networks were effective in handling the anticipated complexity of city interactions. This is evident in the good predictive accuracies (Section 9.4) of the prototype models. During the initial stage of model exploration, it was found that there was a high degree of correlation among input features (Section 9.5). Further investigation showed that the input data tended to cluster together to represent different system “states” (Section 9.5).

Exploratory Factor Analysis (EFA) was used to identify latent “states” within the input data, and revealed four highly correlated groups of variables (factors) latent in the input data. These factors were interpreted as city “states”, characterised by the types of crime and socio-economic variables that tended to cluster together. The four states identified in this study are illustrated in Figure 9-8. Two of the four states, A1 and A3, corresponded to areas with low collective efficacy (see Section 5.3.4). The remaining two states, A2 and A4, corresponded to commercial and mixed-use land use practices, respectively.

Due to the clustering in input space, care needed to be taken when fixing variables for sensitivity analysis. The associated analysis constraints were explained in Section 9.6.1. In order to address the anticipated challenges, sensitivity analyses were carried out for the two models developed in Section 9.4, separately for each of the four attractor states identified in Section 9.5. It was demonstrated that different attractor states respond differently to the same set of input features (Section 9.6.2). Furthermore, it was shown that disregarding the existence of attractor states can lead to misleading results. It was consequently shown that the most effective use of the prototype model(s) are achieved when latent system states are first identified with the use of EFA, followed by contextualised sensitivity analyses within the respective city states.

The measure of uncertainty inherent in Bayesian Neural Networks proved useful in identifying sparsely populated regions of input space, and thereby identifying potentially unreliable predictions (Section 9.6.2). Large error bands were commonly observed when input data was highly skewed. An example of this can be seen in Figure 9-16, where very high standardised values of street light outages resulted in diverging error bands for street robbery predictions.

Large error bands were also observed in regions of input space that did not correspond to one of the identified attractor states. An example of this is seen in

Figure 9-18, where error bands associated with citywide street robbery rapidly diverged as the number of child abuse cases increased. The reason for this is that child abuse is highly localised within the A1 and A3 attractor states. The number of child abuse cases at the citywide level were therefore comparatively low.

In conclusion, it was shown how a combined approach employing Bayesian Neural Networks and sensitivity analysis can be used as a tool for systems-level scenario analysis, and for identifying the key government agencies at play in the fight against crime (Section 9.6.2). It was also shown that Bayesian Neural Networks proved effective in incorporating the inherent complexities in cities, while at the same time providing a measure of prediction uncertainty. While the reliability of model predictions can only be learned in practice; the model development process itself, proved to be an invaluable communication tool capable of explicating implicit assumptions, and identifying gaps in understanding.

10.3.3 Requirement 3: Data availability and accessibility

Readily available data is fundamental to the success of any model. Furthermore, available data needs to be applicable at the relevant spatial and temporal scales of analysis (Section 7.7). Due to the limited accessibility of South African data at the time of this study (Section 4.4.3), readily accessible open data for New York City was used to develop and demonstrate the prototype model (Section 7.6).

The availability of South African data, however, is critical to the efficacy of the proposed solution in the South African context. From meeting with members of the NMBM ICT department (Appendix 1), it was learned that aggregated data used for performance management purposes were sent to the mayor's office on a quarterly basis by the various government departments. Data informing IDP KPIs are therefore available at the citywide level, and collated by the mayor's office. In addition, demographic data is distributed by Statistics South Africa (StatsSA, 2018).

The extent of data availability within respective government departments, however, was unknown. As only aggregated data were sent to the mayor's office, the spatial and temporal scales of supporting data were unknown. The full extent of additional data collected within departments was also unknown. At the time of this study, there was no integrated data warehouse collating municipal data sources in NMBM.

In addition, throughout the course of this study, officials from a number of other NMBM government departments were approached to discuss the availability of data within respective departments. Departments that were approached included the Corporate GIS Department, the Traffic Department, and the Fire Department. While a large amount of data was being collected by each department, the data was largely underutilised beyond performance reporting. There was also limited cross-department sharing of data.

In conclusion, aggregated data informing IDP and performance management processes in NMBM is available at the citywide level, and is collated quarterly by the mayor's office. However, the full extent of data collected within departments was unknown. At the time of this study, there was no integrated data warehouse collating municipal data sources in NMBM. Data was held within the respective government departments, and there was limited cross-department sharing of data. The full benefit of municipal data, therefore, were not being realised.

10.4 Implementation guidelines

Guidelines for implementing the demonstrated solution within the South African context are proposed below. These guidelines are based on knowledge derived from the implementation and evaluation of the prototype model:

- *Guideline 1: KPI framework* – It was found that the KPI framework already in place within the NMBM, used for planning and performance management purposes, was comparable to the one used in this study. It is therefore proposed that model KPIs be selected in such a way as to supplement the existing KPI framework. The intention of the proposed model is not to promote alternative management activities, but rather to enhance existing practices.

- *Guideline 2: Predictive model:*

  - *Guideline 2.1: Implementation process* – It is proposed that a combined approach employing Bayesian Neural Networks and sensitivity analysis be used as a tool for systems-level scenario analysis in smart cities, and for identifying the key government agencies at play in the fight against crime. The most effective use of such models will be achieved when cognisance is taken of the latent attractor states existing in cities. This can be achieved by first identifying latent states with the use of EFA, followed by contextualised sensitivity analyses within the respective city states. The proposed implementation process is demonstrated in Chapter 9.

  - *Guideline 2.2: Neural network architecture* – In this study, different crimes were predicted separately. This resulted in the development of more than one prototype model. In hindsight, it may be more efficient to incorporate the prediction of a number of different crimes into a single model. This can be achieved by simply altering the number of neural network output nodes trained. The software package used in this study, however, could only accommodate one output node (see Section 6.4).

- *Guideline 3: Data:*

  - *Guideline 3.1: Data inventory* - The full extent of data collected by NMBM departments is unknown. It is therefore proposed that a data inventory be carried out across all NMBM departments, aimed at recording and characterising available datasets. Key characterisation considerations include the time period of available data, as well as the spatial and temporal units of records.

  - *Guideline 3.2: Open data portal* – There is no integrated data warehouse collating municipal data sources in the NMBM, and there is limited cross-department sharing of data. It is consequently proposed that the NMBM subscribe to the South African Open Government Partnership commitments discussed in Section 0. As part of these commitments, South Africa has pledged to implement open data portals (Humby, 2018).

- *Guideline 3.3: Data standards* - A number of challenges were faced during the data preparation process. These were discussed in detail in Section 9.3.2. Key challenges related to the compatibility of geographic units used to report data across agencies and across time, missing geocodes, sampling bias, reporting formats, and differences in reporting standards across agencies and across time. These challenges need to be addressed as the NMBM develops and implements an open data portal. To this end, practitioners should consult the number of emerging data standards aimed at fostering data interoperability and quality (see Section 0).

10.5 Summary

The main objective of this study was to develop and evaluate a prototype model for crime management in smart South African cities (Section 1.5). The main objective of this chapter (Figure 10-1) was to evaluate the efficacy of the developed model (RO_7), and to develop a set of implementation guidelines based on knowledge derived from the implementation and evaluation of the prototype model (RO_8). The evaluation approach was described in Section 10.2. This was then followed by model evaluation in Section 10.3, and a list of implementation guidelines in Section 10.4.

It was found that the prototype model successfully integrated data from traditionally isolated management silos, thereby providing a means for synergistic cross-sector collaboration. By building on emerging smart city KPI frameworks, it was shown how a combined approach employing Bayesian Neural Networks and sensitivity analysis could effectively be used as a tool for systems-level scenario analysis, and for identifying the key government agencies at play in the fight against crime (Section 9.6.2).

Furthermore, it was shown that comparable KPI frameworks to the one used in this study are already implemented in South African IDPs. It is anticipated therefore, that the demonstrated modelling approach could supplement existing IDP processes with relatively little effort and disruption of existing management processes. Guidelines for the implementation of the developed modelling approach within the South African context are listed in Section 10.4.

As seen in Chapter 4, the world is moving towards integrated decision-making informed by Big Data. This move towards integration is evident in the emergence of Integrated City Management Platforms, which aim to orchestrate smart city infrastructure at a system-of-systems level (IEC., 2015). Despite the fact that smart city solutions in South Africa are in their infancy, South Africa already has the policy framework in place to support integrated decision-making at the planning level through the implementation of IDPs. In this study, it was shown how emerging trends in smart city integration can complement existing IDP practices in South Africa, with the potential of transforming the smart city status of South African cities.

SECTION 4: CONCLUSIONS AND RECOMMENDATIONS

Chapter 11. Conclusions and Recommendations

11.1 Introduction

Observation of global smart city trends shows a shift in focus from sector-based interventions towards integrated decision-making informed by Big Data. This move towards integration is evident in the emergence of Integrated City Management Platforms, which aim to orchestrate smart city infrastructure at a system-of-systems level (IEC., 2015). Despite the fact that smart city solutions in South Africa are in their infancy, South Africa already has the policy framework in place to support integrated decision-making at the planning level through the implementation of Integrated Development Plans (IDPs). In this study, it was shown how emerging trends in smart city integration can complement existing IDP practices in South Africa, with the potential of transforming the smart city status of South African cities.

The main research problem addressed in this study is that South African cities are not effectively integrating and utilising available data sources for smart city planning and management (Section 1.2). The goal of this study was to investigate how existing, and rapidly emerging smart city data can be integrated and utilised to more effectively support planning in South African cities.

In this study, it was supposed that emerging trends in smart city integration could complement existing planning practices in South Africa. Specifically, it was proposed that a predictive model, that assimilates data from traditionally isolated management silos, could be developed for prediction and simulation at the system-of-systems level (see Sections 1.3 and 1.4).

As proof of concept, the study focused on only one aspect of smart cities, namely crime management. Subsequently, the main objective of this study was to develop and evaluate a predictive model for crime management in smart cities that effectively integrated data from traditionally isolated management silos (Section 1.5).

The Design Science Research (DSR) process was followed to develop, demonstrate and evaluate a prototype model for crime management in smart South African cities.

The DSR methodology and its anticipated application in this study was outlined in Chapter 2. Figure 11-1 provides an overview of the DSR process followed in this study. Specifically, Figure 11-1 specifies the research objectives addressed during each of the DSR activities, together with the relevant reporting and supporting thesis chapter(s) (Section 1.7).

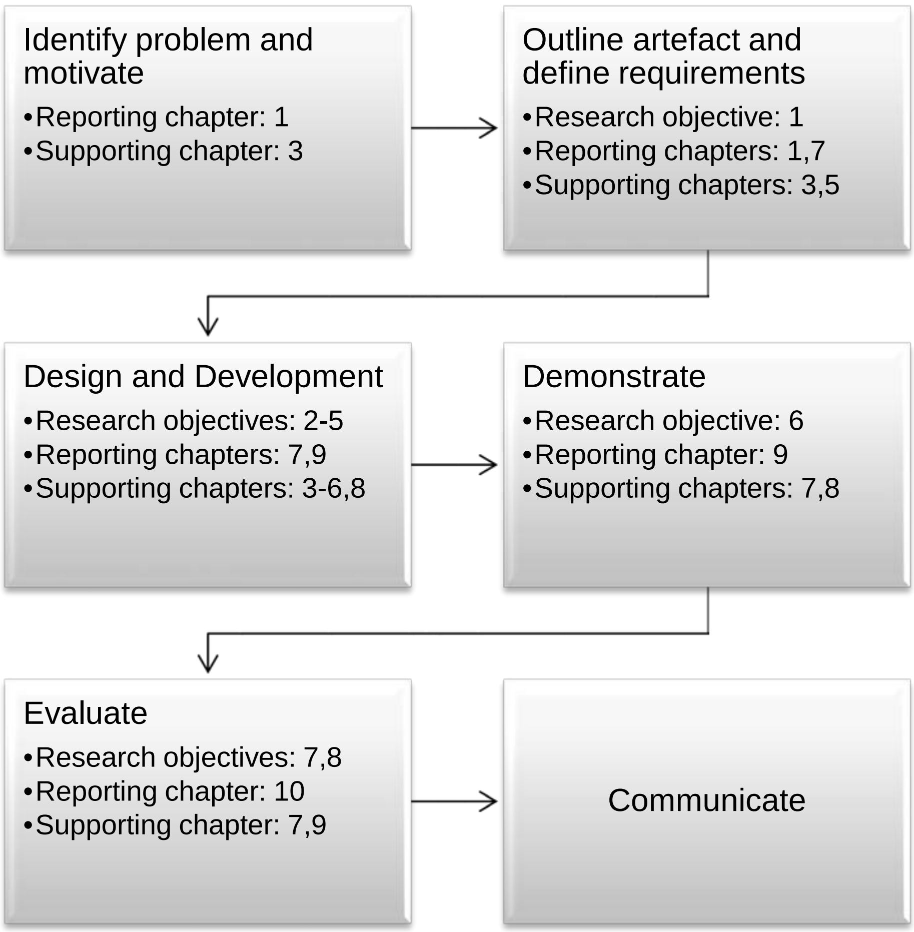


Figure 11-1: DSR process followed in this study.

11.2 Achievement of research objectives

The research objectives stated in Section 1.5 were achieved as follows:

- **RO1: Identify the functional, construction and environmental requirements of an effective model**

Chapter 3 provided the supporting literature necessary to formulate the research problem stated in Section 1.2, and to identify the requirements of a potential solution.

This was achieved by performing a gap analysis of current city planning practices in South Africa, and by reviewing proposed information system solutions to the identified challenges.

It was shown that smart city data can most effectively be used within an integrated adaptive decision-making framework, supported by a comprehensive set of Key Performance Indicators (KPIs) and tools for quantitative scenario analysis (Section 3.8). Chapter 3 addressed RO_1 by defining the functional, construction and environmental requirements of an artefact that could achieve these specifications (Section 3.8).

It was supposed that a predictive model that incorporates all stakeholder and sustainability considerations could serve as a reliable tool for systems-level scenario analysis. It was anticipated that such a model could aid cross-sector collaboration (see Section 3.7), thereby minimising unexpected externalities (see Section 3.4), and fostering synergistic solutions that are often overlooked when problems are solved from within sectoral silos.

In order to limit the scope of the study, the study focused on only one aspect of smart cities, namely crime management. To this end, the solution requirements identified in Section 3.8 were further developed in Chapter 5 within the context of crime management. Model requirements identified in Chapters 3 and 5 were consolidated in Chapter 7 (Section 7.2).

- **RO_2: Identify relevant input and output parameters**
- **RO_3: Identify and characterise available data sources**
- **RO_4: Identify the modelling technique to be used to develop the model**

The purpose of this study was to test the feasibility of the proposed model by developing and evaluating a prototype model for crime management in smart cities. Current smart city activities were reviewed in Chapter 4, and potential smart city design solutions aimed at meeting the model requirements specified in Section 3.8 were proposed. The design solutions proposed in Section 4.6 addressed $RO_{2\text{-}4}$ by indicating smart city KPI standards, open data and machine learning as potential solutions to parameter identification, data requirements and choice of modelling approach, respectively.

Despite the deluge of data generated in smart cities and the accompanying growth in computing power, limited research has been done on exploring the use of this data to develop objective quantitative tools for project prioritisation and scenario analysis at the system-of-systems level (Lombardi *et al.*, 2012; Mattoni *et al.*, 2015; Schleicher *et al.*, 2016; Mattoni *et al.*, 2017) (see Section 4.2.3). In this study, it was proposed that a predictive model for whole-system scenario analysis could be developed by building upon emerging smart city management solutions.

It was supposed that this could be achieved by using smart city KPI frameworks to represent data from traditionally isolated management silos as a set of sectoral KPIs (Figure 1-5a). It was envisaged that the inter-dependencies between sectoral KPIs could be encapsulated in an artificial neural network (Figure 1-5b), which could be used for prediction and simulation at the system-of-systems level (Figure 1-5c).

Design elements identified in Section 4.6 were further developed in Chapters 5 and 6, and were consolidated in Section 7.3. Regarding RO_2, the prototype model was to incorporate as input a subset of the City Protocol Society's *City Anatomy Indicators* (Section 4.4.2); and the commonly employed predictors of crime summarised in Section 5.3.4 were to be used as a guide when selecting KPIs from the CPA indicator framework. The indicators used in this study are summarised in Section 7.5.

Regarding RO_3, due to the limited accessibility of South African data at the time of this study (Section 7.4), readily accessible open data for New York City was used to develop and demonstrate the prototype model. The data sources used in this study are summarised in Section 7.6.

Lastly, Regarding RO_4, the model was to predict the relative influence of selected KPIs on crime. It was proposed that a combined approach employing Bayesian

Neural Networks (Chapter 6) and sensitivity analysis (Section 6.5.1) could be used to predict the relative influence of input features on crime.

- **RO_5: Develop the model**
- **RO_6: Demonstrate the application of the model**

A prototype model for crime management was developed and demonstrated in Chapter 9 according to the design specifications laid out in Chapter 7, thereby addressing $RO_{5\text{-}6}$. Two prototype models were developed for demonstration purposes, one for street larceny and one for street robbery. It was shown how a combined approach employing Bayesian Neural Networks and sensitivity analysis could be used as a tool for systems-level scenario analysis, and for identifying the key government agencies at play in the fight against crime (Section 9.6.2).

From the model implementation, it was determined that Bayesian Neural Networks were effective in handling the anticipated complexity of city interactions. This was evident in the good predictive accuracies (Section 9.4) of the prototype models. During the initial stage of model exploration, it was found that there was a high degree of correlation among input features (Section 9.5). Further investigation showed that the input data tended to cluster together to represent different system "states" (Section 9.5).

Exploratory Factor Analysis (EFA) was used to identify latent "states" within the input data, and revealed four highly correlated groups of variables (factors) latent in the input data. These factors were interpreted as city "states", characterised by the types of crime and socio-economic variables that tended to cluster together. Due to the clustering in input space, care needed to be taken when fixing variables for sensitivity analysis. It was consequently shown that the proposed modelling approach could most effectively be implemented when latent system states were first identified with the use of EFA, followed by contextualised sensitivity analyses within the respective city states.

Furthermore, the measure of uncertainty inherent in Bayesian Neural Networks proved useful in identifying sparsely populated regions of input space, thereby identifying potentially unreliable predictions (Section 9.6.2). While the reliability of model predictions can only be learned in practice; the model development process itself, proved to be an invaluable communication tool capable of explicating implicit assumptions, and identifying gaps in understanding.

- **RO_7: Evaluate the efficacy of the model**
- **RO_8: Develop a set of implementation guidelines for the South African context based on knowledge derived from the development and evaluation of the prototype model**

The main objective of Chapter 10 was to address RO_7 and RO_8. The prototype model was evaluated based on the model requirements specified in Section 7.2. Due to the limited accessibility of South African data at the time of this study, readily accessible open data for New York City was used to develop and demonstrate the prototype model (Section 7.4). The prototype model was consequently evaluated *ex ante* within the South African context by way of a mixed-method case study.

Nelson Mandela Bay Municipality (NMBM) was selected as the most feasible metro to investigate as research agreements were already in place between the Nelson Mandela University and the NMBM. The efficacy of the model was evaluated by way of an interview with NMBM employees, and a document study of their IDP and performance management systems.

It was found that the prototype model successfully integrated data from traditionally isolated management silos, thereby providing a means for synergistic cross-sector collaboration. By building on emerging smart city KPI frameworks, it was shown how a combined approach employing Bayesian Neural Networks and sensitivity analysis could effectively be used as a tool for systems-level scenario analysis, and for identifying the key government agencies at play in the fight against crime (Section 9.6.2).

Furthermore, it was shown that comparable KPI frameworks to the one used in this study were already implemented in South African IDPs. It was anticipated therefore, that the demonstrated modelling approach could supplement existing IDP processes with relatively little effort and disruption of existing management processes. Guidelines for the implementation of the developed modelling approach within the South African context were developed based on the implementation and evaluation of the prototype model, and are listed in Section 10.4.

As shown in Chapter 4, the world is moving towards integrated decision-making informed by Big Data. This move towards integration is evident in the emergence of Integrated City Management Platforms, which aim to orchestrate smart city infrastructure at a system-of-systems level (IEC., 2015). Despite the fact that smart city solutions in South Africa are in their infancy, South Africa already has the policy framework in place to support integrated decision-making at the planning level through the implementation of IDPs. In this study, it was shown how emerging trends in smart city integration can complement existing IDP practices in South Africa, with the potential of transforming the smart city status of South African cities.

11.3 Research contributions

11.3.1 Practical contributions

In this study, a prototype model that effectively integrates data from traditionally isolated management silos was successfully developed and demonstrated for crime management in smart cities. Such a model has applications in systems-level scenario analysis, and effectively supports cross-sector collaboration (see Section 3.7), thereby fostering synergistic solutions that are often overlooked when problems are solved from within sectoral silos. The practical implementation and anticipated use of the developed modelling approach is described in Chapter 9.

Furthermore, guidelines for the implementation of the developed modelling approach within the South African IDP context were developed, and are reported in Section 10.4. It was found that comparable KPI frameworks to the one used in this study are already implemented in South African IDPs, and it is anticipated that the demonstrated modelling approach could supplement existing IDP processes with relatively little effort and disruption of existing management processes. The developed modelling approach therefore successfully met the goal of this study, by showing how existing, and rapidly emerging smart city data can be integrated and utilised to more effectively support planning in South African cities.

11.3.2 Theoretical contributions

11.3.2.1 A modelling paradigm for effective integrated decision-making in smart cities

The main focus of this study was the development of an effective modelling paradigm for integrated decision-making in smart cities. As the smart city concept matures, attention is shifting from domain-specific ICT solutions, towards an interconnected and synergistic approach to solving urban challenges (Fernández-Güell et al., 2016) (see Section 4.3). However, limited progress has been observed in this regard (Fernández-Güell et al., 2016). While frameworks for the management and use of Big Data in smart cities have been proposed (Batty, 2013; Hashem *et al.*, 2016; Pan *et al.*, 2016; Rathore *et al.*, 2016; Silva *et al.*, 2017; Thakuriah *et al.*, 2017), these have focused on architecture considerations. Limited attention has been given to the use of data for integrated decision-making at the system-of-systems level (Lombardi *et al.*, 2012; Mattoni *et al.*, 2015; Schleicher *et al.*, 2016; Mattoni *et al.*, 2017).

As the call for integration intensifies, city scientists and managers will need to develop their capacity to process and analyse large amounts of data from disparate sources. Conversely, while ICT vendors have the required hacking skills for managing and analysing Big Data, they may not have the domain knowledge held by urban planners and city managers. Data Science, therefore, holds great value in bridging the gap between disciplines, and may lead to effective solutions in the new era of integrated smart cities (see Section 4.2.3).

In an attempt to bridge the gap between the different domains (as illustrated in Figure 4-2), an approach to integrated decision-making that incorporates solutions from both city planning and smart city perspectives was developed and evaluated in this study. The study showed how emerging trends in smart city solutions can complement current planning practices in South Africa (see article in Appendix 2). By developing a prototype model for systems-level scenario analysis, this study demonstrated a modelling paradigm for effective integrated decision-making in smart cities.

11.3.2.2 A tool for predictive policing

Although not explicitly stated, the place-based predictive policing techniques discussed in Section 5.3.3.1 all produce actionable predictions at the tactical level (Perry et al., 2013) i.e. predictions are made in blocks covering police patrol beats. In contrast, the prototype model developed in this study made predictions at the strategic level, and is more suited to decision-making at the precinct level (see Section 8.2.1.4).

The prototype model therefore differs from prevalent crime prediction models in terms of the spatial and temporal resolution of predictions. The model consequently extends the focus of current predictive policing practice by developing a tool for strategic decision-making necessary for the adaptive management practices described in Section 5.3.5.3. The modelling approach therefore provides an effective means of shifting the focus of predictive policing activities from an over emphasis on numbers to meeting community needs.

11.3.2.3 A tool for change research

The clustering of state variables observed in Section 9.5 is analogous to the concept of basins of attraction in complex dynamic systems (Walker *et al.*, 2004; Westley *et al.*, 2011). Change research into complex socio-ecological systems studies the nature and behaviour of attractor basins. Particular attention is focused on understanding the underlying interactions causing the emergence and dissolution of attractor states. Much work has been done on applying the concepts of attractor basins and the associated resilience theory to real-world problems (Walker *et al.*, 2004; Westley *et al.*, 2006; Sendzimir *et al.*, 2007; Gundry *et al.*, 2011; Westley *et al.*, 2011; Westley *et al.*, 2015). Further exploration of the approach used in this study to identify attractor basins (Section 9.5) may prove useful in studies aimed at modelling complex systems.

11.4 Limitations and challenges

Due to the limited accessibility of South African data at the time of this study, readily accessible open data for New York City was used to develop and demonstrate the prototype model. The efficacy of the prototype model was consequently evaluated *ex ante* for the South African context by means of a case study of the Nelson Mandela Bay Municipality. Further work is therefore required to refine the model for South African applications.

In order to limit the scope of the investigation, the study focused on only one aspect of smart cities, namely crime management. In order to test the generality of the conclusions made in this study, the concepts need to be tested within other domains (and combinations therefore).

A model for integrated decision-making at the long-term planning level was developed in this study. Smart city activities are dominated by the real-time management of city systems. The application of the proposed modelling approach at the short-term management level therefore should be explored.

Lastly, predictive models are fundamentally limited by sampling bias (see Sections 5.3.5.2 and 9.3.2.3). Caution therefore needs to be taken when interpreting model predictions (see Section 6.5.2). Despite this inherent challenge, the model development process itself is an invaluable communication tool capable of explicating implicit assumptions, and identifying gaps in understanding.

11.5 Future research

Based on the above discussions, the following future research activities are proposed:

- Further refine the model within the South African context by following the implementation guidelines specified in Section 10.4. This includes developing a data landscape for NMBM.

- Applying the concepts developed in this study to other sectors such as water or electricity, for example.
- Applying the concepts developed in this study to the real-time management of city infrastructure.
- Further explore how the modelling approach proposed in this study can support change research (see Section 11.3.2.3).

11.6 Summary

Observation of global smart city trends shows a shift in focus from sector-based interventions towards integrated decision-making informed by Big Data. This move towards integration is evident in the emergence of Integrated City Management Platforms, which aim to orchestrate smart city infrastructure at a system-of-systems level. Despite the fact that smart city solutions in South Africa are in their infancy, South Africa already has the policy framework in place to support integrated decision-making at the planning level through the implementation of IDPs.

The main research problem addressed in this study is that South African cities are not effectively integrating and utilising available, and rapidly emerging smart city data sources for planning and management. To this end, it was proposed that a predictive model, that assimilates data from traditionally isolated management silos, could be developed for prediction and simulation at the system-of-systems level. As proof of concept, the study focused on only one aspect of smart cities, namely crime management. Subsequently, the main objective of this study was to develop and evaluate a predictive model for crime management in smart cities that effectively integrated data from traditionally isolated management silos.

In this chapter, the achievement of research objectives was explained (Section 11.2) and the research contributions of the study were highlighted (Section 11.3). Study limitations and challenges were noted in Section 11.4, and future research was suggested in Section 11.5. It was successfully shown how emerging trends in smart city integration can complement existing IDP practices in South Africa. The practical contributions of this study was the development of a prototype model for integrated decision-making in smart cities, and the associated guidelines for the implementation of the developed modelling approach within the South African IDP context.

Theoretically, this work contributed towards the development of a modelling paradigm for effective integrated decision-making in smart cities. This work also contributed towards developing strategic-level predictive policing tools aimed at proactively meeting community needs, and contributed to the body of knowledge regarding complex systems modelling.

REFERENCES AND APPENDICES

References

Ahvenniemi, H., Huovila, A., Pinto-Seppä, I. & Airaksinen, M. 2017. What are the differences between sustainable and smart cities? *Cities,* 60**,** 234-245.

Aklilu, A., Belete, A. & Moyo, T. 2014. Analysing Community Participation in the Municipal Integrated Development Planning Processes in Limpopo Province, South Africa. *Mediterranean Journal of Social Sciences,* 5**,** 257-262.

Al Nuaimi, E., Al Neyadi, H., Mohamed, N. & Al-Jaroodi, J. 2015. Applications of big data to smart cities. *Journal of Internet Services and Applications,* 6**,** 25.

Albino, V., Berardi, U. & Dangelico, R. M. 2015. Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *Journal of Urban Technology,* 22**,** 3-21.

Alfreds, D. 2015. Tshwane Safety App Overcomes Language Barriers. *fin24tech* [Online]. Available: http://www.fin24.com/Tech/News/Tshwane-safety-app-overcomes-language-barriers-20150324 [Accessed 16 August 2016].

Alfreds, D. 2016. Tshwane Launches City-wide Crime Safety App. *fin24tech* [Online]. Available: http://www.fin24.com/Tech/News/tshwane-launches-city-wide-crime-safety-app-20160421 [Accessed 16 August 2016].

Allen, P. M. 1997a. Cities and Regions as Evolutionary Complex Systems. *Geographical Systems,* 4**,** 103 – 130.

Allen, P. M. 1997b. *Cities and Regions as Self-Organizing Systems: Models of Complexity*, International Ecotechnology Research Centre, Cranfield University, England.

American National Standards Institute. 2018. *Smart and Sustainable Cities* [Online]. Available: https://webstore.ansi.org/smart-cities/default.aspx [Accessed 27 June 2018].

Angelidou, M. 2015. Smart cities: A conjuncture of four forces. *Cities,* 47**,** 95-106.

Anže Žitnik. 2019. *Stopping Crime Before it Happens: Predicitive Policing* [Online]. MEDI@4SEC. Available: http://media4sec.eu/blog-18/ [Accessed 18 January 2019].

ArcGIS. 2018. *USA ZIP Code Areas* [Online]. Available: https://www.arcgis.com/home/item.html?id=8d2012a2016e484dafaac0451f9aea24 [Accessed 30 October 2018].

Azavea. 2015. HunchLab: Under the Hood.

Azavea. 2017. A Citizen's Guide to HunchLab.

Babuta, A. 2017. *Occasional Paper: Big Data and Policing: An Assessment of Law Enforcement Requirements, Expectations and Priorities*, Royal United Services Institute for Defence and Security Studies.

Bachner, J. 2013. Predictive Policing: Preventing Crime with Data and Analytics. IBM Center for the Business of Government.

Bandura, A. 2000. Exercise of Human Agency Through Collective Efficacy. *Current Directions in Psychological Science,* 9**,** 75-78.

Bathembu, C. 2016. *Municipal ward committees: What you need to know* [Online]. Available: https://www.vukuzenzele.gov.za/municipal-ward-committees-what-you-need-know [Accessed 30 June 2018].

Batty, M. 2013. *Urban Informatics and Big Data: A Report to the ESRC Cities Expert Group*.

Batty, M. & Marshall, S. 2012. The Origins of Complexity Theory in Cities and Planning. *In:* Portugali, J., Meyer, H., Stolk, E. & Tan, E. (eds.) *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning.* Springer.

Bennett Moses, L. & Chan, J. 2016. Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society***,** DOI: 10.1080/10439463.2016.1253695.

Berends, J., Carrara, W. & Vollers, H. 2017. Analytical Report 6: Open Data in Cities 2. European Data Portal.

Bhadeshia, H. K. D. H. 1999. Neural Networks in Materials Science. *ISIJ International,* 39**,** 966-979.

Biggs, R., Rhode, C., Archibald, S., Kunene, L. M., Mutanga, S. S., Nkuna, N., Ocholla, P. O. & Phadima, L. J. 2015. Strategies for managing complex social-ecological systems in the face of uncertainty: examples from South Africa and beyond. *Ecology and Society,* 20**,** 52.

Bird, J. A. 2015. Open Data Standards. Available: https://www.w3.org/blog/data/2015/02/26/open-data-standards/ [Accessed 9 February 2019].

Bondy, M., Roth, S. & Sager, L. 2018. Crime is in the Air: The Contemporaneous Relationship between Air Pollution and Crime. IZA – Institute of Labor Economics.

Bowers, K. J., Johnson, S. D., Guerette, R. T., Summers, L. & Poynton, S. 2011. Spatial displacement and diffusion of benefits among geographically focused policing initiatives: a meta-analytical review. *Journal of Experimental Criminology,* 7**,** 347-374.

Braga, A. A. 2001. The Effects of Hot Spots Policing on Crime. *The Annals of the American Academy of Political and Social Science,* 578**,** 104-125.

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. J. & Brantingham, P. L. 1984. Patterns in crime. New York: Macmillan.

Brayne, S., Rosenblat, A. & Boyd, D. 2015. Predictive Policing. datacivilrights.org.

Brooks, A. 2018. Market Assessment: Big Data and Visualization Solutions in Law Enforcement. IDC White Paper.

Browning, C. R. 2002. The Span of Collective Efficacy: Extending Social Disorganization Theory to Partner Violence. *Journal of Marriage and Family,* 64**,** 833-850.

Bushman, B. J., Wang, M. C. & Anderson, C. A. 2005. Is the Curve Relating Temperature to Aggression Linear or Curvilinear? Assaults and Temperature in Minneapolis Reexamined. *Journal of Personality and Social Psychology,* 89**,** 62–66.

Caplan, J. M., Kennedy, L. W. & Piza, E. L. 2013. Joint Utility of Event-Dependent and Environmental Crime Analysis Techniques for Violent Crime Forecasting. *Crime & Delinquency,* 59**,** 243 –270.

Carrara, W., Nieuwenhuis, M. & Vollers, H. 2016. Open Data Maturity in Europe 2016: Insights into the European State of Play. European Data Portal.

Center for Government Excellence. 2017. *Open Data Standards Directory* [Online]. Available: https://datastandards.directory/ [Accessed 9 February 2019].

Chen, M., Mao, S. & Liu, Y. 2014. Big Data: A Survey. *Mobile Networks and Applications,* 19**,** 171-209.

Chicago Police Department. 2016. CPD Welcomes the Opportunity to Comment on Recently Published RAND Review. News release.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A. & Scholl, H. J. 2012. Understanding Smart Cities: An Integrative Framework. *45th Hawaii International Conference on System Sciences. DOI: 10.1109/hicss.2012.615.*

Chutel, L. 2016. Cops on Demand: An “Uber for Police” has been Launched in South Africa. *Quartz* [Online]. Available: http://qz.com/682636/an-uber-for-police-has-been-launched-in-south-africa/ [Accessed 16 August 2016].

Cilliers, D. P. & Retief, F. 2016. The extent and status of environmental management frameworks (EMFs) in South Africa, 2006–2015. *South African Geographical Journal,* 99**,** 283-300.

City of Chicago. 2017. *Strategic Subject List* [Online]. Available: https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np [Accessed 18 January 2019].

City of Johannesburg. 2011. Joburg 2040: Growth and Development Strategy.

City of Johannesburg. 2015. 2012/16 Integrated Development Plan: 2015/16 Review.

City Protocol Society. 2015a. City Anatomy Indicators. CPA-PR _002_Anatomy Indicators.

City Protocol Society. 2015b. City Anatomy: A Framework to support City Governance, Evaluation and Transformation. City Protocol Agreement CPA-I_001-v2.

City Protocol Society. 2015c. Livable Districts and Cities. CPC_004_Livable_Districts_and_Cities.

City Protocol Society. 2016. Open Sensor Platform. CPWD-PR_005_Open_Sensor_Platform.

Cohen, B. 2012. *What Exactly Is A Smart City?* [Online]. Available: http://www.fastcoexist.com/1680538/what-exactly-is-a-smart-city [Accessed 29 July 2016].

Cohen, B. 2014. *Methodology for 2014 Smart Cities Benchmarking* [Online]. Available: http://www.fastcoexist.com/3038818/the-smartest-cities-in-the-world-2015-methodology [Accessed 29 July 2016].

Collis, J. & Hussey, R. 2014. *Business Research: A Practical Guide for Undergraduate and Postgraduate Students - Fourth Edition,* New York, Palgrave Macmillan.

Conway, D. 2010. *The Data Science Venn Diagram* [Online]. Available: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram [Accessed 14 September 2016].

Corman, H. & Mocan, N. 2002. Carrots, Sticks and Broken Windows. *NBER Working Paper No. 9061.* Cambridge, MA: National Bureau of Economic Research.

D'Orsogna, M. R. & Perc, M. 2015. Statistical physics of crime: a review. *Physics of Life Reviews,* 12**,** 1-21.

de Blasio, B., Fuleihan, D. & Newman, E. W. 2018. Mayor’s Management Report, NYC MOO.

DEA. 2014. *Environmental Impact Assessment and Management Strategy for South Africa (Draft).* Pretoria, Department of Environmental Affairs.

DEAT. 1998. National Environmental Management Act 107 of 1998.

DEAT. 2004. *Overview of Integrated Environmental Management, Integrated Environmental Management, Information Series 0,* Pretoria, Department of Environmental Affairs and Tourism.

Everatt, D., Marais, H. & Dube, N. 2010. Participation … for what Purpose? Analysing the Depth and Quality of Public Participation in the Integrated Development Planning Process in Gauteng. *Politikon,* 37**,** 223-249.

Ferguson, A. G. 2012. Predictive Policing and Reasonable Suspicion. *Emory Law Journal,* 62**,** 259-325.

Ferguson, A. G. 2017a. Policing Predictive Policing. *Washington University Law Review,* 94**,** 1109-1189.

Ferguson, A. G. 2017b. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement,* New York, New York University Press.

Fernández-Güell, J.-M., Guzmán-Araña, S., Collado-Lara, M. & Fernández-Añez, V. 2016. How to Incorporate Urban Complexity, Diversity and Intelligence into Smart Cities Initiatives. *In:* Alba, E., Chicano, F. & Luque, G. (eds.) *Smart Cities. Smart-CT 2016. Lecture Notes in Computer Science, vol 9704.* Cham: Springer.

fin24tech. 2017. *Smart cities will now have their own nervous system* [Online]. Available: https://www.fin24.com/Tech/Opinion/smart-cities-will-now-have-their-own-nervous-system-20171116 [Accessed 27 June 2018].

Fischer, S. N., Shinn, M., Shrout, P. & Tsemberis, S. 2008. Homelessness, mental illness, and criminal activity: examining patterns over time. *American Journal of Community Psychology,* 42**,** 251-265.

Foresee, F. D. & Hagan, M. T. 1997. Gauss-Newton approximation to Bayesian learning. *Proceedings of the International Joint Conference on Neural Networks*.

Frost & Sullivan. 2013. Strategic Opportunity Analysis of the Global Smart City Market.

Gandomi, A. & Haider, M. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management,* 35**,** 137-144.

Gasco-Hernandez, M. 2018. Building a Smart City: Lessons from Barcelona *Communications of the ACM,* 61**,** 50-57.

Gorr, W. & Harries, R. 2003. Introduction to crime forecasting. *International Journal of Forecasting,* 19**,** 551-555.

Groff, E. R. & Vigne, N. G. L. 2002. Forecasting the Future of Predictive Crime Mapping. *Crime Prevention Studies,* 13**,** 29-57.

Gundry, L. K., Kickul, J. R., Griffiths, M. D. & Bacq, S. C. 2011. Entrepreneurial Bricolage and Innovation Ecology: Precursors to Social Innovation? *Frontiers of Entrepreneurship Research,* 31**,** 3.

Hagan, J. & Peterson, R. D. 1995. *Crime and Inequality*, Stanford University Press.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. 2014. *Multivariate Data Analysis – 7th Edition,* Essex, England, Pearson Education Limited.

Haken, H. 2012. Complexity and Complexity Theories: Do These Concepts Make Sense? *In:* Portugali, J., Meyer, H., Stolk, E. & Tan, E. (eds.) *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning.* Springer.

Han, J., Kamber, M. & Pei, J. 2012. *Data Mining: Concepts and Techniques,* Waltham, MA, USA, Morgan Kaufmann Publishers.

Happimo. 2016. *Namola* [Online]. Available: http://namola.com [Accessed 16 August 2016].

Hardin, G. 1968. The Tragedy of the Commons. *Science,* 162**,** 1243-1248.

Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J. & Williams, P. 2010. Foundations for Smarter Cities. *IBM Journal of Research and Development,* 54**,** 1–16.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E. & Chiroma, H. 2016. The role of big data in smart city. *International Journal of Information Management,* 36**,** 748-758.

Hassani, H., Huang, X., Silva, E. S. & Ghodsi, M. 2016. A Review of Data Mining Applications in Crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal,* 9**,** 139-154.

Hastie, T., Tibshirani, R. & Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* New York, Springer-Verlag.

Head, T. 2017. *The future is on its way: SA tech firm using AI to predict crime before it happens* [Online]. Available: https://www.thesouthafrican.com/the-future-is-on-its-way-sa-tech-firm-using-ai-to-predict-crime-before-it-happens/ [Accessed 11 December 2018].

Hevner, A. & Chatterjee, S. 2010. *Design Research in Information Systems: Theory and Practice,* New York, Springer.

Hinton, G. 2013. *CSC321: 2011 - Introduction to Neural Networks and Machine Learning. Lecture 11: Bayesian learning continued.* University of Toronto.

HMIC. 2017. PEEL: Police Effectiveness 2016: A National Overview. London: Her Majesty's Inspectorate of Constabulary.

Huawei. 2018. *Intelligent Operation Center Solution* [Online]. Available: https://e.huawei.com/za/solutions/industries/smart-city/ioc [Accessed 27 June 2018].

Humby, T.-L. 2018. South Africa Mid-Term Report 2016-2018 (Year 1). Open Government Partnership.

Hummelbrunner, R. & Jones, H. 2013. *A guide for planning and strategy development in the face of complexity*, Background note, Overseas Development Institute.

Hunt, P., Saunders, J. & Hollywood, J. S. 2014. Evaluation of the Shreveport Predictive Policing Experiment. RAND Safety and Justice Program.

IBM. 2010. *The world's 4 trillion dollar challenge: Using a system-of-systems approach to build a smarter planet*, IBM Institute for Business Value.

IBM. 2013. *IBM Intelligent Operations Center for Smarter Cities*, IBM Corporation.

IBM. 2018. *Smarter Cities* [Online]. Available: https://www.ibm.com/smarterplanet/us/en/smarter_cities/overview [Accessed 30 June 2018].

IEC. 2015. *White Paper: Orchestrating Infrastructure for Sustainable Smart Cities,* Geneva, Switzerland, International Electrotechnical Commission.

ISO. 2014. ISO 37120:2014, Sustainable development of communities - Indicators for city services and quality of life.

ISO. 2016. ISO 37101:2016, Sustainable development in communities - Management system for sustainable development - Requirements with guidance for use.

ISO/IEC. 2017. ISO/IEC 30182:2017, Smart city concept model - Guidance for establishing a model for data interoperability.

ITU. 2016a. *Shaping Smarter and More Sustainable Cities: Striving for Sustainable Development Goals,* Geneva, Switzerland.

ITU. 2016b. *Unleashing the Potential of the Internet of Things,* Geneva, Switzerland.

ITU. 2018. *KPIs on Smart Sustainable Cities* [Online]. Available: https://www.itu.int/en/ITU-T/ssc/Pages/KPIs-on-SSC.aspx [Accessed 25 June 2018].

Joh, E. E. 2017. Feeding the Machine: Policing, Crime Data, & Algorithms. *William & Mary Bill of Rights Journal,* 26**,** 287-302.

Johannesson, P. & Perjons, E. 2012. *A Design Science Primer (Book no longer available - now An Introduction to Design Science)*, CreateSpace: Lexington.

Kacfah Emani, C., Cullot, N. & Nicolle, C. 2015. Understandable Big Data: A survey. *Computer Science Review,* 17**,** 70-81.

Kehl, D., Guo, P. & Kessler, S. 2017. Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing *Responsive*

Communities Initiative, Berkman Klein Center for Internet & Society. Harvard Law School.

Kelly, M. 2000. Inequality and Crime. *Review of Economics and Statistics,* 82**,** 530-539.

Kennedy, L. W., Caplan, J. M. & Piza, E. L. 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.

Kleinhans, R., Van Ham, M. & Evans-Cowley, J. 2015. Using Social Media and Mobile Technologies to Foster Engagement and Self-Organization in Participatory Urban Planning and Neighbourhood Governance. *Planning Practice & Research,* 30**,** 237-247.

Kourtit, K., Nijkamp, P., Franklin, R. S. & Rodriguez-Pose, A. 2014. A blueprint for strategic urban research: the urban piazza. *Town Planning Review,* 85**,** 97-126.

Kourtit, K., Nijkamp, P. & Steenbruggen, J. 2017. The significance of digital data systems for smart city policy. *Socio-Economic Planning Sciences,* 58**,** 13-21.

Kuyper, T. 2016. *Smart City Strategy & Upscaling: Comparing Barcelona and Amsterdam.* Master Thesis, Universitat Pompeu Fabra & Barcelona School of Management.

Lämmerhirt, D., Rubinstein, M. & Montiel, O. 2017. The State of Open Government Data in 2017: Creating meaningful open data through multi-stakeholder dialogue. Open Knowledge International.

Langton, L., Berzofsky, M., Krebs, C. & Smiley-McDonald, H. 2012. National Crime Victimization Survey: Victimizations Not Reported to the Police, 2006-2010. U.S. Department of Justice.

Lombardi, P., Giordano, S., Farouh, H. & Yousef, W. 2012. Modelling the smart city performance. *Innovation: The European Journal of Social Science Research,* 25**,** 137-149.

Mabuza, E. 2016. Tshwane Metro Police Happy with Safety App. *Times Live* [Online]. Available: http://www.timeslive.co.za/thetimes/2016/05/12/Tshwane-metro-police-happy-with-safety-app [Accessed 16 August 2016].

MacKay, D. J. C. 1992. A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation,* 4**,** 448-472.

MacKay, D. J. C. 1995. Probable networks and plausible predictions: a review of practical Bayesian methods for supervised neural networks. Unpublished.

Making Sense, Balestrini, M., Bejtullahu, S., Bocconi, S., Boerwinkel, G., Boonstra, M., Boschman, D.-S., Camprodon, G., Coulson, S., Diez, T., Fazey, I., Hemment, D., Horn, C. v. d., Ilazi, T., Jansen-Dings, I., Kresin, F., McQuillan, D., Nascimento, S., Pareschi, E., Pólvora, A., Salaj, R., Scott, M., Seiz, G. & Woods, M. 2018. *Citizen Sensing: A Toolkit*, European Commission.

Manyika, J., Chui, M., Groves, P., Farrell, D., Kuiken, S. V. & Doshi, E. A. 2013. Open data: Unlocking innovation and performance with liquid information. McKinsey & Company.

Mapou, A. E. M., Shendell, D., Ohman-Strickland, P., Madrigano, J., Meng, Q., Whytlaw, J. & Miller, J. 2017. Environmental Factors and Fluctuations in Daily Crime Rates. *Journal of Environmental Health,* 80**,** 8-22.

Marais, L., Human, F. & Botes, L. 2008. Measuring what? The utilisation of development indicators in the integrated development planning process. *Journal of Public Administration,* 43**,** 376-400.

Marais, M., Retief, F. P., Sandham, L. A. & Cilliers, D. P. 2014. Environmental management frameworks: results and inferences of report quality performance in South Africa. *South African Geographical Journal,* 97**,** 83-99.

MathWorks. 2018. MATLAB and Statistics Toolbox Release R2018b. Natick, Massachusetts, United States: The MathWorks, Inc.

Mattoni, B., Gugliermetti, F. & Bisegna, F. 2015. A multilevel method to assess and design the renovation and integration of Smart Cities. *Sustainable Cities and Society,* 15**,** 105-119.

Mattoni, B., Nardecchia, F., Benelli, A., Buscaglione, S., Pagliaro, F. & Burattini, C. 2017. A quantitative evaluation of the mutual influences among Smart strategies applied at district level. *IEEE*.

Mautjana, H. M. & Mtapuri, O. 2014. Integrated Development Plans without Development Indicators: Results from Capricorn District Municipalities in South Africa. *Mediterranean Journal of Social Sciences,* 5**,** 474-483.

McCarney, P. 2015. The Evolution of Global City Indicators and ISO37120: The First International Standard on City Indicators. *Statistical Journal of the IAOS,* 31**,** 103–110.

McCarthy, B. & Hagan, J. 1991. Homelessness: A Criminogenic Situation? *The British Journal of Criminology,* 31**,** 393-410.

MEDIA4SEC. 2016. Report on State of the Art Review.

Memeburn. 2015. *'Uber for Cops' Hopes to Fight Crime* [Online]. Available: http://www.enca.com/technology/namola-app-hopes-fight-crime [Accessed 16 August 2016].

Misuraca, G. 2007. Cape Town's "Smart City" Strategy in South Africa. *E-Governance in Africa, from Theory to Action: A Handbook on ICTs for Local Governance.* Africa World Press & International Development Research Centre.

Mnguni, L. 2016. *Why SA's municipalities are failing and how to fix them* [Online]. Available: https://www.businesslive.co.za/rdm/politics/2016-03-03-why-sas-municipalities-are-failing-and-how-to-fix-them/ [Accessed 29 June 2018].

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L. & Brantingham, P. J. 2016. Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association,* 110**,** 1399-1411.

MOO. 2018. Social Indicators Report: Update to the 2016 Report.

Morgan, S. L. & Winship, C. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research,* New York, Cambridge University Press.

Morrison-Saunders, A., Pope, J., Gunn, J. A. E., Bond, A. & Retief, F. 2014. Strengthening impact assessment: a call for integration and focus. *Impact Assessment and Project Appraisal,* 32**,** 2-8.

Neal, R. M. 1995. *Bayesian Learning for Neural Networks.* Doctor of Philosophy, University of Toronto.

Neal, R. M. 2004. *Software for Flexible Bayesian Modeling and Markov Chain Sampling* [Online]. Available: https://www.cs.toronto.edu/~radford/fbm.software.html [Accessed 10 February 2019].

Neal, R. M. 2012. *Bayesian Learning for Neural Networks*, Springer New York.

Neirotti, P., De Marco, A., Cagliano, A. C., Mangano, G. & Scorrano, F. 2014. Current trends in Smart City initiatives: Some stylised facts. *Cities,* 38**,** 25-36.

New York City Department of City Planning. 2010. New York City PUMAs and Community Districts.

Ngamlana, N. & Eglin, R. 2015. *Learning Brief #1: Spatial Planning and Integrated Development Planning*, afesis-corplan.

Nisbet, R., Elder, J. & Miner, G. 2009. *Handbook of Statistical Analysis and Data Mining Applications,* Canada, Academic Press.

NMBM. 2017. 2016/17 Annual Report.

NMBM. 2018a. 2017/18 Mid-term Budget and Performance Report.

NMBM. 2018b. Integrated Development Plan (IDP) - Second Edition 2017/18 - 2021/22.

NMBM. 2018c. *Performance Agreements* [Online]. Available: http://www.nelsonmandelabay.gov.za/Documents.aspx?catID=10&pageID=224 [Accessed 1 January 2019].

O’Neil, C. & Schutt, R. 2014. *Doing Data Science: Straight Talk from the Frontline,* Sebastopol, CA, O’Reilly Media, Inc.

OECD. 2009. Applications of Complexity Science for Public Policy: New Tools for Finding Unanticipated Consequences and Unrealized Opportunities.

Open Data Census. 2017. US City Open Data Census. Open Knowledge International.

Open Data Charter. 2015. International Open Data Charter.

Open Data Institute. 2019. *Open Standards for Data* [Online]. Available: http://standards.theodi.org/ [Accessed 9 February 2019].

Open Government Partnership. 2018. *South Africa* [Online]. Available: www.opengovpartnership.org/countries/south-africa [Accessed 27 November 2018].

Pan, Y., Tian, Y., Liu, X., Gu, D. & Hua, G. 2016. Urban Big Data and the Development of City Intelligence. *Engineering,* 2**,** 171-178.

Pearl, J. 2009. *Causality: Models, Reasoning, and Inference (Second Edition),* Cambridge, University Press.

Pease, K. 1999. A Review of Street Lighting Evaluations: Crime Reduction Effects. *In:* Tilley, P. a. (ed.) *Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention, Crime Prevention Studies, 10, 47-76.*

Peffers, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. 2008. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems,* 24**,** 45-77.

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C. & Hollywood, J. S. 2013. Predicitve Policing: The Role of Crime Forecasting in Law Enforcement Operations. RAND Corporation.

Pike, A. & Schulz, G. W. 2014. *Hollywood-style surveillance technology inches closer to reality* [Online]. Available: https://www.revealnews.org/article-legacy/hollywood-style-surveillance-technology-inches-closer-to-reality/ [Accessed 12 December 2018].

Pramanik, M. I., Lau, R. Y. K., Yue, W. T., Ye, Y. & Li, C. 2017. Big data analytics for security and criminal investigations. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 7**,** e1208.

PredPol Inc. 2018. *PredPol* [Online]. Available: https://www.predpol.com/ [Accessed 18 January 2019].

President's Task Force on 21st Century Policing. 2015. Final Report of the President's Task Force on 21st Century Policing. Washington, DC: Office of Community Oriented Policing Services.

Project Everyone. 2016. *The 17 Goals* [Online]. Available: https://www.globalgoals.org/ [Accessed 12 November 2016].

Provost, F. & Fawcett, T. 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data,* 1**,** 51-59.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rao, Y. R. 2017. Automatic Smart Parking System using Internet of Things (IOT). *International Journal of Engineering Technology Science and Research,* 4**,** 225-228.

Ratcliffe, J. H. 2016. *Intelligence-Led Policing - Second Edition*, Routledge.

Rathore, M. M., Ahmad, A., Paul, A. & Rho, S. 2016. Urban planning and building smart cities based on the Internet of Things using Big Data analytics. *Computer Networks,* 101**,** 63-80.

Ritchie, A. & Thomas, R. 2009. *Sustainable Urban Design: An Environmental Approach 2nd Edition*, Taylor & Francis.

Robinson, D. & Koepke, L. 2016. Stuck in a Pattern: Early evidence on "predictive policing" and civil rights. Upturn.

Rockefeller Foundation. 2016. *100 Resilient Cities* [Online]. Available: http://www.100resilientcities.org [Accessed 25 July 2016].

Rodriguez, P. P. & Gianola, D. 2018. *brnn: Bayesian Regularization for Feed-Forward Neural Networks. R package version 0.7.* [Online]. Available: https://CRAN.R-project.org/package=brnn [Accessed 10 February 2018].

Roeder, O., Eisen, L.-B. & Bowling, J. 2015. What caused the crime decline? : Brennan Center for Justice at New York University School of Law.

RSA. 2000. Local Government: Municipal Systems Act 32 of 2000.

RSA. 2011. *Guidelines for the Development of Spatial Development Frameworks - Version 8*, Department of Rural Development and Land Reform.

RSA. 2012a. *Companion to the Environmental Management Impact Assessment Regulations 2010 (IEM Guideline Series, Guideline 5)*, Department of Environmental Affairs.

RSA. 2012b. *Environmental Management Framework Regulations 2010 (IEM Guideline Series, Guideline 6)*, Department of Environmental Affairs.

RSA. 2017. *Amendments to the Environmental Impact Assessment Regulations 2014*, Department of Environmental Affairs.

RSA. n.d. *IDP Guide Pack*, Department of Provincial and Local Government.

Rutgers Center on Public Security. 2019. *Risk Terrain Modeling* [Online]. The State University of New Jersey. Available: http://www.riskterrainmodeling.com/ [Accessed 10 February 2019].

Ruwanza, S. & Shackleton, C. M. 2015. Incorporation of environmental issues in South Africa's municipal Integrated Development Plans. *International Journal of Sustainable Development & World Ecology,* 23**,** 28-39.

SALGA. 2013. *Municipal Barometer* [Online]. Available: http://www.municipalbarometer.co.za/ [Accessed 14 September 2016].

SALGA. 2015. Smart Cities. *in.KNOW.vation.* Pretoria: South African Local Government Association.

Sampson, R. J. 2006. Collective Efficacy Theory: Lessons Learned and Directions for Future Inquiry. *In:* Cullen, F. T., Wright, J. P. & Blevins, K. R. (eds.) *Taking Stock: The Status of Criminological Theory.* New Brunswick, N.J: Transaction Publishers.

Sampson, R. J. & Raudenbush, S. W. 2004. Seeing Disorder: Neighborhood Stigma and the Social Construction of “Broken Windows”. *Social Psychology Quarterly,* 67**,** 319–342.

SANRAL. 2018. *i-TRAFFIC* [Online]. Available: https://www.i-traffic.co.za/ [Accessed 26 November 2018].

Saunders, J., Hunt, P. & Hollywood, J. S. 2016. Predictions put into practice: a quasi-experimental evaluation of Chicago’s predictive policing pilot. *Journal of Experimental Criminology,* 12**,** 347-371.

Saunders, M., Lewis, P. & Thornhill, A. 2009. *Research Methods for Business Students - Fith Edition,* Essex, Pearson Education Limited.

Schleicher, J. M., Vögler, M., Inzinger, C. & Dustdar, S. 2015. Towards the Internet of Cities: A Research Roadmap for Next-Generation Smart Cities. UCUI’15, Melbourne, Australia. DOI: http://dx.doi.org/10.1145/2811271.2811274.

Schleicher, J. M., Vogler, M., Inzinger, C., Fritz, S., Ziegler, M., Kaufmann, T., Bothe, D., Forster, J. & Dustdar, S. 2016. A Holistic, Interdisciplinary Decision Support System for Sustainable Smart City Design. *In:* Alba, E., Chicano, F. & Luque, G. (eds.) *Smart Cities. Smart-CT 2016. Lecture Notes in Computer Science, vol 9704.* Cham: Springer.

Sendzimir, J., Magnuszewski, P., Flachner, Z., Balogh, P., Molnar, G., Sarvari, A. & Nagy, Z. 2007. Assessing the Resilience of a River Management Regime: Informal Learning in a Shadow Network in the Tisza River Basin. *Ecology and Society,* 13**,** 11.

Shan, C., Wang, H., Chen, W. & Song, M. 2015. *The Datascience Handbook: Advice and Insights from 25 Amazing Data Scientists*, Data Science Bookshelf.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. 1989. Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology,* 27**,** 27-55.

Silva, B. N., Khan, M. & Han, K. 2017. Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making. *Wireless Communications and Mobile Computing***,** DOI: 10.1155/2017/9429676.

Simmons, R. 2016. Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System. *Michigan State Law Review*, 947-1017.

Sloman, S. A. 2005. *Causal Models: How People Think About the World and Its Alternatives,* Oxford, Oxford University Press.

Smith, G. J. D., Bennett Moses, L. & Chan, J. 2017. The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach. *The British Journal of Criminology,* 57**,** 259-274.

Sourmail, T. 2002. *Simultaneous Precipitation Reactions in Creep-Resistant Austenitic Stainless Steels.* Doctor of Philosophy, University of Cambridge.

Sourmail, T. 2004. *Softwares for Bayesian neural networks training* [Online]. Available: http://thomas-sourmail.net/neural_networks.html [Accessed 16 December 2018].

StatsSA. 2018. *Statistics South Africa* [Online]. Available: http://www.statssa.gov.za/ [Accessed 10 February 2019].

Steenbruggen, J., Tranos, E. & Nijkamp, P. 2015. Data from mobile phone operators: A tool for smarter cities? *Telecommunications Policy,* 39**,** 335-346.

Tan, P.-N., Steinbach, M. & Kumar, V. 2006. *Introduction to Data Mining,* Boston, Pearson Education, Inc.

Taylor, R. B., Ratcliffe, J. H. & Perenzin, A. 2015. Can We Predict Long-term Community Crime Problems? The Estimation of Ecological Continuity to Model Risk Heterogeneity. *Journal of Research in Crime and Delinquency,* 52**,** 635-657.

Thakuriah, P., Tilahun, N. & Zellner, M. 2017. *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics,* Switzerland, Springer.

The British Standards Institution. 2014. PAS 181:2014, Smart city framework – Guide to establishing strategies for smart cities and communities.

The British Standards Institution. 2017. PAS 183:2017, Smart cities – Guide to establishing a decision-making framework for sharing data and information services.

The Economist. 2009. *Triple Bottom Line* [Online]. Available: http://www.economist.com/node/14301663 [Accessed 26 July 2016].

The Leadership Conference on Civil and Human Rights, 18 Million Rising, American Civil Liberties Union, Brennan Center for Justice, Center for Democracy & Technology, Center for Media Justice, Color of Change, Data & Society Research Institute, Demand Progress, Electronic Frontier Foundation, Free Press, Media Mobilizing Project, NAACP, National Hispanic Media Coalition, Open MIC (Open Media and Information Companies Initiative), Open Technology Institute at New America & Public Knowledge. 2016. Predictive Policing Today: A Shared Statement of Civil Rights Concerns.

The Rockefeller Foundation | Arup. 2015. City Resilience Framework.

The Rockefeller Foundation | Arup. 2018. City Resilience Index: Understanding and measuring city resilience.

Towers, S., Chen, S., Malik, A. & Ebert, D. 2018. Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective. *PLoS One,* 13**,** e0205151.

Tran, D., Kucukelbir, A., Dieng, A. B., Rudolph, M., Liang, D. & Blei, D. M. 2016. Edward: A library for probabilistic modeling, inference, and criticism. arXiv preprint arXiv:1610.09787.

U4SSC. 2017a. *Collection Methodology for Key Performance Indicators for Smart Sustainable Cities,* Geneva, Switzerland.

U4SSC. 2017b. *Implementing Sustainable Development Goal 11 by Connecting Sustainability Policies and Urban-Planning Practices Through ICTs,* Geneva, Switzerland.

U.S. Department of Justice Federal Bureau of Investigation. 2018. *UCR Offense Definitions* [Online]. Available: https://www.ucrdatatool.gov/offenses.cfm [Accessed 2 Novemer 2018].

United Nations. 2018. *World Urbanization Prospects: The 2018 Revision, Country Profiles* [Online]. Department of Economic and Social Affairs, Population Division. Available: https://population.un.org/wup/Country-Profiles/ [Accessed 25 November 2018].

United Nations. 2015. *Transforming our world: the 2030 Agenda for Sustainable Development (A/RES/70/1)*.

United States Census Bureau. 2018a. *American Community Survey (ACS): Concept & Definitions* [Online]. Available: https://www.census.gov/programs-surveys/acs/geography-acs/concepts-definitions.html [Accessed 30 October 2018].

United States Census Bureau. 2018b. *Decennial Census and the American Community Survey (ACS)* [Online]. Available: https://www.census.gov/programs-surveys/decennial-census/about/census-acs.html [Accessed 1 November 2018].

United States Census Bureau. 2018c. *Public Use Microdata Areas (PUMAs)* [Online]. Available: https://www.census.gov/geo/reference/puma.html [Accessed 30 October 2018].

United States Census Bureau. 2018. *Understanding and Using American Community Survey Data: What All Data Users Need to Know,* Washington, DC, U.S. Government Printing Office.

Vaishnavi, V. & Kuechler, B. 2004. *Design Science Research in Information Systems* [Online]. Available: http://www.desrist.org/design-research-in-information-systems/ [Accessed 19 August 2016].

Vaishnavi, V. K. & Kuechler, W. 2015. *Design Science Research Methods and Patterns: Innovating Information and Communication Technology* Boca Raton, FL, Taylor and Francis Group.

Venktess, K. 2017. *SA firm to use AI to predict crime before it happens* [Online]. Available: https://www.fin24.com/Tech/News/sa-firm-to-use-ai-to-predict-crime-before-it-happens-20170529 [Accessed 11 December 2018].

Walker, B., Holling, C. S., Carpenter, S. R. & Kinzig, A. 2004. Resilience, adaptability and transformability in social–ecological systems. *Ecology and Society,* 9**,** 5.

Wang, X., Ning, Z., Hu, X., Ngai, E. C. H., Wang, L., Hu, B. & Kwok, R. Y. K. 2018. A City-Wide Real-Time Traffic Management System: Enabling Crowdsensing in Social Internet of Vehicles. *IEEE Communications Magazine,* 56**,** 19-25.

Weisburd, D., Groff, E. R. & Yang, S.-M. 2009. Understanding Developmental Crime Trajectories at Places: Social Disorganization and Opportunity Perspectives at Micro Units of Geography. Report to the National Institute of Justice.

Weller, C. 2016. *A California police department is using software to decide if you're about to commit a crime* [Online]. Business Insider. Available: https://www.businessinsider.com/intrado-beware-system-tracks-threat-levels-2016-1?IR=T [Accessed 19 January 2019].

Welsh, B. P. & Farrington, D. C. 2008. Effects of Improved Street Lighting on Crime. *Campbell Systematic Reviews,* 13.

Westley, F., Laban, S., Rose, C., McGowan, K., Robinson, K., Tjornbo, O. & Tovey, M. 2015. Social Innovation Lab Guide. The Rockefeller Foundation.

Westley, F., Olsson, P., Folke, C., Homer-Dixon, T., Vredenburg, H., Loorbach, D., Thompson, J., Nilsson, M., Lambin, E., Sendzimir, J., Banarjee, B., Galaz, V. & van der Leeuw, S. 2011. Tipping towards sustainability: emergent pathways of transformation, Working Paper No. 3. Prepared for the "3rd Nobel Laureate Symposium on Global Sustainability: Transforming the World in an Era of Global Change", in Stockholm, 16-19 May 2011. Stockholm Resilience Centre, the Royal Swedish Academy of Sciences, the Stockholm Environment Institute, the Beijer Institute of Ecological Economics and the Potsdam Institute for Climate Impact Research.

Westley, F., Patton, M. Q. & Zimmerman, B. 2006. *Getting to maybe: How the world is changed,* Toronto, Random House Canada.

Wikipedia. 2018a. *Administrative divisions of New York (state)* [Online]. Available: https://en.wikipedia.org/wiki/Administrative_divisions_of_New_York_(state) [Accessed 30 October 2018].

Wikipedia. 2018b. *Boroughs of New York City* [Online]. Available: https://en.wikipedia.org/wiki/Boroughs_of_New_York_City [Accessed 30 October 2018].

Wikipedia. 2018c. *Community boards of New York City* [Online]. Available: https://en.wikipedia.org/wiki/Community_boards_of_New_York_City [Accessed 30 October 2018].

Wikipedia. 2018d. *Complex system* [Online]. Available: https://en.wikipedia.org/wiki/Complex_system [Accessed 28 June 2018].

Wikipedia. 2018e. *Environmental Impact Assessment* [Online]. Available: https://en.wikipedia.org/wiki/Environmental_impact_assessment [Accessed 29 June 2018].

Wikipedia. 2018f. *Neighborhoods in New York City* [Online]. Available: https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City [Accessed 30 October 2018].

Wikipedia. 2018g. *Organization of the New York City Police Department* [Online]. Available: https://en.wikipedia.org/wiki/Organization_of_the_New_York_City_Police_Department#Police_precincts [Accessed 30 October 2018].

Wikipedia. 2018h. *ZIP Code* [Online]. Available: https://en.wikipedia.org/wiki/ZIP_Code [Accessed 30 October 2018].

Willis, J. J. 2011. First-Line Supervision under Compstat and Community Policing: Lessons from Six Agencies. Office of Community Oriented Policing Services.

Witten, I. H., Frank, E. & Hall, M. A. 2011. *Data Mining: Practical Machine Learning Tools and Techniques,* Burlington, MA, USA, Morgan Kaufmann Publishers.

World Bank. 2014. *Introduction to poverty analysis (English),* Washington, DC, World Bank Group.

World Commission on Environment and Development. 1987. *Our Common Future*, Oxford University Press.

World Wide Web Foundation. 2017. Open Data Barometer Global Report - Fourth Edition.

Xenos, M., Vromen, A. & Loader, B. D. 2014. The great equalizer? Patterns of social media use and youth political engagement in three advanced democracies. *Information, Communication & Society,* 17**,** 151-167.

Zhuhadar, L., Thrasher, E., Marklin, S. & de Pablos, P. O. 2017. The next wave of innovation—Review of smart cities intelligent operation systems. *Computers in Human Behavior,* 66**,** 273-281.

Appendix 1. Meeting Documentation

NELSON MANDELA
UNIVERSITY

Computing Sciences

SMART CITY DISCUSSION

Date and time: 20 November 2018 at 2 pm
Venue: MIS Charles Babbage Room

AGENDA

1. **Attendance Register**

2. **Background to the Project**
   For noting.

3. **NYC Indicators and Data Sources (Doc 1)**
   For noting.

4. **NMBM Indicators**
   For discussion:
   - Citywide
   - Departmental
   - Methods of reporting

5. **NMBM Data Sources**
   For discussion:
   - Availability
   - Access (Open Data / Interoperability)
   - Spatial resolution
   - Temporal resolution and period

NYC INDICATORS (Doc 1A)								Data Source (spatial resolution)				
ID	Domain	Agency	Indicator	CPA	MMR	Social	Current Reporting Scale	City	Borough	PUMA	Point	Zipcode
1	Culture	Office of Citywide Event Coordination and Management (CECM)	Events	Performing arts shows per 1000 population (Functions SI)	-	-	-	14	14	14	14	
2	Economy	Department of Finance (DOF)	Assessed value of residential properties as a percentage of total assessed value of all residential properties	Assessed value of commercial and industrial properties as a percentage of total assessed value of all properties (Economy CI)	-	-	-	15	15	15	15	
4	Economy	Department of Finance (DOF)	Income distribution	Gini Index (Economy CI)	-	Income Distribution	Citywide / Annual	1;2;3	1;2;3	4;5	-	
6	Economy	Department of Youth and Community Development (DYCD)	Youth disconnectivity	Youth unemployment rate (Functions SI)	-	Disconnected Youth, 16-24 Not at Work & Not in School	Borough / Annual	18	18	-	-	
5	Economy	Mayor's Office for Economic Opportunity	Percentage of city population living in poverty	Percentage of city population living in poverty (Economy CI)	Total SNAP households (000)	Number of New Yorkers in or Near Poverty; Total Supplemental Nutrition Assistance Program Recipients	CD / Annual	1;2;3	1;2;3	4;5	-	
3	Economy	New York State Department of Labor	Unemployment rate	City's unemployment rate (Functions CI); Percentage of employed population (Economy CI)	New York City unemployment rate (%)	Unemployment Rate; Labor Force Participation Rate	Borough / Annual	1;2;3	1;2;3	4;5	-	
7	Education	Department of Education (DOE)	Positive Learning Environment	Primary education student/teacher ratio (Functions CI); Students per teacher in mandatory education (Functions SI)	-	Four-Year College Readiness; Four-Year High School Graduation Rate	Citywide; Borough / Annual	31; 32	31; 32	31; 32	31; 32	
8	Education	Department of Education (DOE)	Number of higher education degrees per 100k population	Number of higher education degrees per 100k population (Functions SI)	-	Number of NYC Public School Students Attaining Associate's or Bachelor's Degrees	Citywide / Annual	1;2;3	1;2;3	4;5	-	
9	Environment	Office of Emergency Management (OEM)	Weather: Light levels	Global solar irradiance yearly average (W/m2) (Environment SI)	-	-	-	13	13	-	-	
10	Environment	Office of Emergency Management (OEM)	Weather: Wind speed	Average wind speed (km/h) (Environment SI)	-	-	-	13	13	-	-	
12	Family unit	Administration for Children's Services (ACS)	Number of abuse/neglect investigations	-	Number of State Central Register consolidated investigations; Investigations that found credible evidence of abuse or neglect (%)	-	Citywide / Annual	16	16	16	-	
13	Family unit	Department of Health and Mental Hygiene (DOHMH)	Fertility rate: Annual number of live births per 1000 women aged 15-49 years	Fertility rate: Annual number of live births per 1000 women aged 15-49 years (Citizens CI)				36	36	36	-	
11	Family unit	Department of Youth and Community Development (DYCD)	Female householders	-	-	-	-	1;2;3	1;2;3	4;5	-	
25	Governance	Board of Elections (BOE)	Voter participation	Number of registered voters as a percentage of the voting age population (Citizens SI); Voter participation in last city election (as a percentage of eligible voters) (Citizens CI)	Total active voters (000); Total registered voters (000); Voter turnout - general election (000)	Eligible Voter Registration Rate; Turnout Among Voting Age Population (percent); Turnout Among Registered Voters (percent)	Borough / Annual	40	40	-	-	
14	Governance	Civilian Complaint Review Board (CCRB)	Officer integrity	Number of convictions for corruption and/or bribery by city officials per 100k population (Government SI)	Total civilian complaints against uniformed members of the New York City Police Department	-	Citywide / Annual	29	29	29	-	
16	Health	Department of Health and Mental Hygiene (DOHMH) / NYPD	Drug activity	-	Narcotics arrests; Deaths from unintentional drug overdose (CY)	-	Citywide / Annual	10; 12	10; 12	10; 12	10; 12	

NYC INDICATORS (Doc 1A)

ID	Domain	Agency	Indicator	CPA	MMR	Social	Current Reporting Scale	Data Source (spatial resolution): City	Borough	PUMA	Point	Zipcode
15	Health	Human Resources Administration (HRA) / Department of Health and Mental Hygiene (DOHMH)	Health insurance coverage	Public expenditure on health per capita (Functions CI)	Medicaid enrollees administered by HRA (000) / Adult New Yorkers without health insurance (%) (CY)	-	Citywide / Annual	1;2;3	1;2;3	4;5	-	
17	Housing	Department of City Planning (DCP)	Density housing	Density housing (Built Domain CI)	-	-	-	15	15	15	15	
18	Housing	Department of Finance (DOF)?	Percentage of empty housing	Percentage of empty housing (Functions SI)	-	-	-	1;2;3	1;2;3	4;5	-	
19	Housing	Department of Finance (DOF)?	Percentage of housing for rent	Percentage of housing for rent (Functions SI)	-	-	-	1;2;3	1;2;3	4;5	-	
20	Housing	Department of Homeless Services (DHS)	Number of homeless per 100k population	Number of homeless per 100k population (Functions SI)	Average number of adult families in shelters per day; Average number of families with children in shelters per day; Average number of single adults in shelters per day; Unsheltered individuals who are estimated to be living on the streets, in parks, under highways, on subways, and in the public transportation stations in New York City	Homeless: Average Daily Census	Citywide / Annual	33; 34; 35	33; 35	33	12	
21	Housing	New York City Housing Authority (NYCHA)	Percentage of social housing	Percentage of social housing (Functions CI)	Applicants placed in public housing; Section 8 occupied units (vouchers); Apartments (000)	-	Citywide / Annual	41	41	41	-	
23	Infrastructure	Department of Information Technology & Telecommunications (DoITT)	New York City Households with Internet Access	Number of internet connections per 100k population (Infrastructure CI)	-	New York City Households with Internet Access	Borough / Annual	40	40	-	-	
22	Infrastructure	Department of Transportation (DOT)	Average time to repair street lights	Average length of electrical interruptions (in hours)	Average time to repair street lights	-	Citywide / Annual	12	12	12	12	12
24	Land use	Department of City Planning (DCP)	Land use (area per 100k population)	Green area (hectares) per 100k population (Built Domain CI); Neighborhood Homogeneity (Built Domain CI); Industrial availability: Space density (Built Domain SI); Percentage parking places off the road (Infrastructure CI); Areal size of mix-use developments as a percentage of city total built area (Built Domain CI)	-	-	-	15	15	15	15	
26	Population	Department of City Planning (DCP)	Cultural diversity	Cultural diversity (Citizens CI)	-	-	-	1;2;3	1;2;3	4;5	-	
27	Population	Department of City Planning (DCP)	Population density	Population density (Citizens CI)	-	-	-	1;2;3;7	1;2;3;7	4;5;8	-	
28	Public safety	Administration for Children's Services (ACS)	Total admissions to detention	-	Total admissions to detention	-	Citywide / Annual	17	17	-	-	17
29	Public safety	Department of Correction (DOC)	Admissions to Department of Correction per 1000	-	Admissions	Admissions to Department of Correction per 1000	Borough / Annual	37; 40	40	-	-	
30	Public safety	Department of Probation (DOP)	Department of Probation Population per 1000	-	Adult investigation reports completed - total; Juvenile investigation reports completed; Juvenile supervision - Intake cases received; Adult supervision cases - end of period; Juvenile supervision cases - end of period; Adult initial risk assessments completed; Juvenile initial risk assessments completed	Department of Probation Population per 1000	Borough / Annual	38;39;40	38;39;40	-	-	

NYC INDICATORS (Doc 1A)								Data Source (spatial resolution)				
ID	Domain	Agency	Indicator	CPA	MMR	Social	Current Reporting Scale	City	Borough	PUMA	Point	Zipcode
31	Public safety	New York State Division of Criminal Justice Services	Number of police officers per 100k population	Number of police officers per 100k population (Functions CI)	-	-	-	30	-	-	-	
32	Public safety	NYPD	Crimes (all police agencies) per 100k population	Crimes (all police agencies) per 100k population (Functions SI)	-	-	-	10	10	10	10	
33	Public safety	NYPD	Crimes against property per 100k population	Crimes against property per 100k population (Functions SI)	Major felony crime: Burglary; Grand larceny; Grand larceny auto	-	Citywide / Annual	10; 11	10; 11	10	10	
34	Public safety	NYPD	Violent crime rate per 100k population	Violent crime rate per 100k population (Functions SI); Number of homicides per 100k population (Functions CI)	Major felony crime: Murder and non-negligent manslaughter; Forcible rape; Robbery; Felonious assault	Violent Crime per 1000	Borough / Annual	10; 11	10; 11	10	10	
35	Transportation	Department of City Planning (DCP)	Mean travel time to work	Number of personal automobiles per capita (Infrastructure CI); Annual number of public transport trips per capita (Functions CI); Percentage of commuters using a transportation mode to work other than a personal vehicle (Functions SI)	-	Mean Travel Time to Work	NTA / Annual	1;2;3	1;2;3	4;5	-	
36	Transportation	Department of Transportation (DOT)	Pedestrian volume index	Surface of pedestrian priority areas and streets / Total street area (Built Domain CI); Kilometres of bicycle paths and lanes per 100k population (Infrastructure SI)	Pedestrian space installed (square feet); Pedestrian volume index; Citi Bike annual membership: Trips (000); Bicycle lane miles installed	-	Citywide / Annual	19	19	-	19	

NYC DATA SOURCES (Doc 1B)

ID	Source Agency	Data Description	Type	URL	Update Frequency	Data dictionary	Spatial Scale	Temporal Scale	Start Year	End Year	Period	Comments
1	United States Census Bureau	American Community Survey	Files	https://www1.nyc.gov	Annually	Yes	Borough	1-Year Estimates			2008-2009; 2011-2016	
2	United States Census Bureau	American Community Survey	Files	https://www1.nyc.gov	Annually	Yes	Borough	3-Year Estimates			2007-2009; 2011-2013	
3	United States Census Bureau	American Community Survey	Files	https://www1.nyc.gov	Annually	Yes	Borough	5-Year Estimates			2010; 2012-2016	
4	United States Census Bureau	American Community Survey	Files	https://www1.nyc.gov	Annually	Yes	PUMA	3-Year Estimates			2007-2009; 2011-2013	
5	United States Census Bureau	American Community Survey	Files	https://www1.nyc.gov	Annually	Yes	PUMA	5-Year Estimates			2010; 2014-2016	
6	United States Census Bureau	American Community Survey	Files	https://www1.nyc.gov	Annually	Yes	NTA	5-Year Estimates			2010; 2012-2016	
7	United States Census Bureau	Decennial Census - Census 2010: Total Population and Persons Per Acre, 2000-2010	File	https://www1.nyc.gov	Decennially	NA	Borough	Year			2000; 2010	
8	United States Census Bureau	Decennial Census - Census 2010: Total Population and Persons Per Acre, 2000-2010	File	https://www1.nyc.gov	Decennially	NA	NTA	Year			2000; 2010	
9	United States Census Bureau	PEP: Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2017	Dataset	factfinder.census.gov	Annually	Yes	Borough	Year	2010	2017		
10	NYPD	NYPD Complaint Data Historic	Dataset	opendata.cityofnewyo	Annually	Yes	Borough; Police Precinct; Coordinates	Date&Time	2006	2017		
11	New York State Division of Criminal Justice Services	Index Crimes by County and Agency: Beginning 1990	Dataset	data.ny.gov	Annually	No	Borough	Year	1990	2017		
12	311	311 Service Requests from 2010 to Present	Dataset	opendata.cityofnewyo	Daily	Yes	Borough; CD; Address; ZIP Code; Coordinates	Date&Time	2010	Present		
13	National Centers for Environmental Information: National Oceanic and Atmospheric Administration	Local Climatological Data (LCD)	Datasets	https://www.ncdc.noa	Daily	Yes	Borough	Date&Time	2006	2017		NY CITY CENTRAL PARK, NY US; JFK
14	Office of Citywide Event Coordination and Management (CECM)	NYC Permitted Event Information - Historical	Dataset	opendata.cityofnewyo	As needed	Yes	Borough; CD; Police Precinct; Street Address	Start Date&Time; End Date&Time	2008	Present		
15	Department of City Planning (DCP)	Archived Primary Land Use Tax Lot Output (PLUTO)	Datasets	https://www1.nyc.gov	Annually	Yes	Borough; CD; Police Precinct; School District; Council District; Health Area; ZIP Code; Tax Lot; Census Tract; Address; Coordinates	-	2006	Present		Data for 2008 unavailable.
16	Administration for Children's Services (ACS)	Abuse/Neglect by Community District (CD)	Files	opendata.cityofnewyo	Annually	Yes	Borough; CD	Year	2010	2017		
17	Administration for Children's Services (ACS)	Detention and Placement Demographic reports	Files	opendata.cityofnewyo	Annually	Yes	ZIP Code	Year	2015	2017		
18	Mayor's Office of Operations (OPS)	Social Indicators Report Data - Citywide	Dataset	opendata.cityofnewyo	Annually	Yes	Borough	Year	2005	2016		
19	Department of Transportation (DOT)	Bi-annual pedestrian counts	Shapefile	opendata.cityofnewyo	Bi-Annually	Yes	Borough; Street; Coordinates	Month	2007	2017		May/Sep
20	Department of City Planning (DCP)	2010 Census Tracts	Shapefile	opendata.cityofnewyo	As needed	Yes	-	-	-	-		
21	Department of City Planning (DCP)	Borough Boundaries	Shapefile	opendata.cityofnewyo	As needed	Yes	-	-	-	-		
22	Department of City Planning (DCP)	Community Districts	Shapefile	opendata.cityofnewyo	As needed	Yes	-	-	-	-		
23	Department of City Planning (DCP)	Neighborhood Tabulation Areas	Shapefile	opendata.cityofnewyo	As needed	Yes	-	-	-	-		
24	Department of City Planning (DCP)	Police Precincts	Shapefile	opendata.cityofnewyo	As needed	Yes	-	-	-	-		
25	Department of City Planning (DCP)	Public Use Microdata Areas (PUMA)	Shapefile	opendata.cityofnewyo	As needed	Yes	-	-	-	-		
26	Department of Information Technology & Telecommunications (DoITT)	Zip Code Boundaries	Shapefile	opendata.cityofnewyo	As needed	Yes	-	-	-	-		
27	United States Census Bureau	County	Shapefile	https://www.census.g	As needed	Yes	-	-	-	-		Cartographic Boundary Shapefiles
28	United States Census Bureau	Public Use Microdata Areas (PUMA)	Shapefile	https://www.census.g	As needed	Yes	-	-	-	-		Cartographic Boundary Shapefiles
29	Civilian Complaint Review Board (CCRB)	Where Incidents That Led To A Complaint Took Place By Precinct	Datasets	opendata.cityofnewyo	Historic	Yes	Police Precinct	Year	2005	2009		
30	New York State Division of Criminal Justice Services	Law Enforcement Personnel by Agency: Beginning 2007	Dataset	data.ny.gov	Annually	Yes	Citywide	Year	2007	2017		
31	Department of Education	2005 - 2017 School Quality Review Ratings	Dataset	opendata.cityofnewyo	Annually	Yes	School	Year	2005	2017		
32	Department of Education	School locations	Datasets	opendata.cityofnewyo	Annually	Yes	All	Year	2012	Present		

NYC DATA SOURCES (Doc 1B)

ID	Source Agency	Data Description	Type	URL	Update Frequency	Data dictionary	Spatial Scale	Temporal Scale	Start Year	End Year	Period	Comments
33	Department of Homeless Services (DHS)	Case Census by Borough, Community District, and Facility Type	Dataset	opendata.cityofnewyo	Monthly	Yes	Borough; CD	Month			2018	July-Sep
34	Department of Homeless Services (DHS)	DHS Daily Report	Dataset	opendata.cityofnewyo	Daily	Yes	Citywide	Day	2013	Present		
35	Department of Homeless Services (DHS)	Directory Of Homeless Population By Year	Dataset	opendata.cityofnewyo	Historic	Yes	Borough	Year	2009	2012		
36	Department of Health and Mental Hygiene (DOHMH)	Fertility rate per 1000 women aged 15-44	Files	https://www1.nyc.gov	Annually	No	CD	Year	2006	2016		
37	Department of Correction (DOC)	Inmate Admissions	Dataset	opendata.cityofnewyo	Monthly	Yes	Citywide	Date&Time	2014	Present		
38	Department of Probation (DOP)	DOP Adult Cases Snapshot by Calendar Year	Dataset	opendata.cityofnewyo	Annually	Yes	Borough	Year	2006	2017		
39	Department of Probation (DOP)	DOP Juvenile Cases Snapshot by Calendar Year	Dataset	opendata.cityofnewyo	Annually	Yes	Borough	Year	2006	2017		
40	Mayor's Office of Operations (OPS)	Social Indicator Report Data	Dataset	opendata.cityofnewyo	Annually	Yes	Borough	Year	2000	2017		
41	New York City Housing Authority (NYCHA)	NYCHA Development Data Book	Dataset	opendata.cityofnewyo	As needed	Yes	Borough; CD; Street	-	-	-		

SMART CITY DISCUSSION

Attendance Register

Date and time: 20 November 2018 at 2 pm

Venue: MIS Charles Babbage Room

Name	Surname	Job Title	Affiliation	Email	Contact number	Signature
Lindsay	Westraadt	PhD Candidate	NMU, Computing Sciences	lindsayw@mandela.ac.za	041-504 4764	L. Westraadt
André	Calitz	Prof	NMU, Computing Sciences	Andre.Calitz@mandela.ac.za	041-504 2639	AC.
Anthea	Van der Hoogen	Lecturer	" "	anthea.vanderhoogen.ac.za	041-5042758	
Louise	Gaston	Systems Specialist	NMBM	lgaston@mandelametro.gov.za	041 5061529	
RAJAN	NAIDOO	DEPUTY DIRECTOR	NMBM	RNAIDOO@MANDELAMETRO.GOV.ZA	041-5061321	
CHRIS	VAN LEYVELD	DD: APP DEV	NMBM	CVANLEYVELD@ ~	0415061218	
Linda	Szuhanyi	APP DEV	NMBM	lszuhanyi@mandelametro.gov.za		

Appendix 2. Peer-reviewed Conference Paper

A Gap Analysis of New Smart City Solutions for Integrated City Planning and Management

L. Westraadt
Department of Computing Sciences
Nelson Mandela University
P.O. BOX 77000
South Africa
Lindsay.Westraadt@mandela.ac.za

A. P. Calitz
Department of Computing Sciences
Nelson Mandela University
P.O. BOX 77000
South Africa
Andre.Calitz@mandela.ac.za

ABSTRACT

Currently, much attention is being given to integrating smart city solutions aimed at eliminating systemic inefficiencies with a focus on integrated decision-making, interoperability and collaboration. However, the concepts of integrated decision-making have preceded the recent wave of integration by half a century. There are concerns that the hype in smart city integration is detached from existing integrated planning and management practices, which are well-rooted in complexity science and built upon decades of practice. The contribution of this paper is to contextualize new smart city solutions for integrated city planning and management within the framework of South African policy tools, with the aim of identifying synergies and gaps for future research. This is achieved by carrying out a literature review of existing and emerging tools for integrated decision-making in South African cities and performing a subsequent gap analysis.

CCS CONCEPTS

• **Information systems** → Information Systems Applications → Decision support systems

KEYWORDS

Integrated Development Plan, Integrated Environmental Management, System of Systems, Smart Cities.

1 INTRODUCTION

Over the last decade there has been an explosion in smart city solutions and a deluge in data generated from these activities. This has led to radically reduced inefficiencies in all sectors, including smart buildings, transportation and energy [1, 2]. However, the transformation of cities is not following as expected by city management and it is being proposed that the silo approach to smart cities development is reaching its limits [3]. Practitioners are calling for a coordinated approach and recently, research is being conducted on developing appropriate frameworks aimed at eliminating systemic inefficiencies and fostering integrated decision-making, interoperability and collaboration [3, 4].

This paper aims to summarize the latest developments in smart city integration in the context of existing policy tools aimed at integrated planning and management in cities. The call for integrated resources management has preceded the recent wave in smart city integration by half a century, with the first Environmental Impact Assessments being performed in the USA in the 1960's [5], the publication of the Tragedy of the Commons in 1968 [6] and the call for Sustainable Development in 1987 [7]. There is thus 50 years of research and practice in integrated resources management that can be built upon and learned from [8].

The paper investigates and revisits the need for integrated resources management (Section 3) and summarizes the key principles and requirements for effective and efficient management of complex systems in Section 3.3. Section 4 provides an overview of existing and emerging tools for integrated decision-making in cities. Specifically, a review of existing legislative and policy tools for integrated decision-making in South Africa is provided (Section 4.1.1), along with a summary of the challenges experienced in practice (Section 4.1.2) and the proposed Information Systems solutions (Section 4.1.3). This is followed by an overview of the state-of-the-art in Smart Cities system of systems solutions (Section 4.2). Section 5 consolidates the reviews on existing and emerging tools for integration, identifying synergies and gaps for future research. A summary of conclusions and recommendations for future work is presented in Section 6.

2 RESEARCH PROBLEM AND OBJECTIVE

Currently, much attention is being given to integrating smart city solutions aimed at eliminating systemic inefficiencies [3, 4], with a focus on integrated decision-making, interoperability and collaboration. However, concepts of integrated decision making have preceded the recent wave of integration by half a century, and there is concern that the hype in smart city integration is detached from existing integrated planning and management

SAICSIT '18, September 26–28, 2018, Port Elizabeth, South Africa

ACM ISBN 978-1-4503-6647-2/18/09...$15.00
https://doi.org/10.1145/3278681.3278699

practices, which are well-rooted in complexity science and built upon decades of practice [8].

The research objective of this study is to contextualize new smart city solutions for integrated city planning and management within the framework of South African policy tools, with the aim of identifying synergies and gaps for future research. This is achieved by carrying out a literature review of existing and emerging tools for integrated decision-making in South African cities and performing a subsequent detailed gap analysis.

3 THE CITY AS A COMPLEX SYSTEM AND THE NEED FOR INTEGRATED PLANNING AND MANAGEMENT

3.1 The Goal: Sustainable and Resilient Cities

Cities have a major impact on human quality of life and the natural environment. Although different cities have different challenges [3], the driving force behind all city planning and management is to create sustainable and resilient cites. The term "sustainable development" was originally coined in 1987 by the World Commission on Environment and Development and refers to "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [7]. Practically, this has come to entail development that synergistically meets basic human needs, promotes job creation and economic growth, reduces inequality and respects the natural environment [9, 10]. City managers, therefore aim to address socio-economic challenges, such as service delivery, transportation, education, job creation, food security, public safety, healthcare services and quality of life. This must be achieved against the pressures of urbanization, aging infrastructure, fiscal and socio-political constraints and limitations in the sink and source capacity of the environment.

While sustainability indicators focus on the social, economic and environmental performance of cities [11], urban resilience focuses on the capacity of individuals and systems within a city to survive, adapt and thrive in the face of chronic stresses and acute shocks [12]. City resilience focuses on emergency response, reliable communication and mobility, as well as inclusive, integrated planning and management informed by data [13].

3.2 The City as an Interdependent System of Systems

In light of the different agendas at play within a city, cities are often seen as a multifaceted system of systems, where each sub-system is composed of an integrated web of stakeholders from both the public and private sectors; all trying to, independently, optimize their respective interactions within a particular domain [3, 4]. Despite the fact that the large majority of economic decisions are made locally at the sub-system level [3, 4, 14], city systems are highly interdependent. In fact, economists estimated in 2010 that, on average almost 50 percent of economic outputs generated by a particular sub-system relies on input from other systems [4]. There exist a number of delineations of core city systems [4, 15, 16]. Fig. 1 depicts the city as a system of systems and illustrates the interdependencies that exist between the sub-systems.

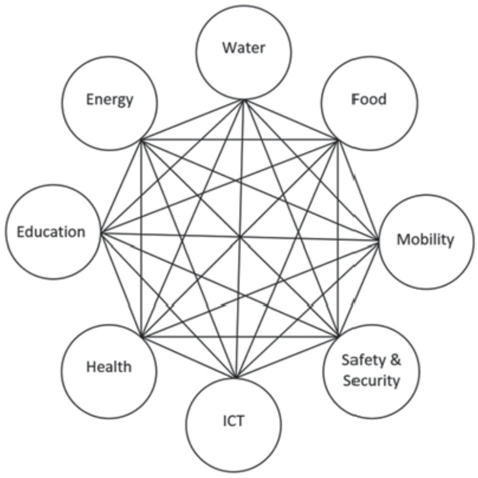


Figure 1. Depiction of the city as a system of systems and the interdependencies that exist between sub-systems. Source: Author's own construction.

3.3 Implications of Complexity for Management

Complex systems are defined to be systems that consist of many interdependent components and have distinct properties arising from these interdependencies, such as adaptability, emergence, self-organization, attractors, chaos and non-linearity [17-19]. These properties make complex systems intrinsically difficult to model and manage [17-19]. Cities, by their very nature, are complex systems [17, 18, 20-22] and it has long been known that in order to effectively and efficiently manage cities, they need to be managed at the systems level.

Managing complex city systems independently from within sectoral silos can result in poor performance, specifically when interdependencies and feedback loops result in unexpected externalities and outcomes. A number of management paradigms and principles have been developed to take into account the inherent uncertainty in complex socio-ecological systems [18, 23, 24]. These principles can and are, applied to the management of complex city systems. In order to limit externalities, a key principle in managing complex socio-ecological systems is to promote integrated decision-making. Integration in this context refers to the vertical (within sectoral silos) and horizontal (inter-silo) integration of all stakeholders and sustainability concerns (social, economic and ecological) into all decision making processes across the full life-cycle of activities [25]. The desired outcomes of integrated decision-making is sustainability-led optimization of financial and environmental resources and the limitation of socio-economic and ecological externalities. Another key principle in managing complex socio-ecological systems is adaptive management [23]. In order to manage for unexpected outcomes resulting from feedback loops and other properties of complex systems, adaptive management methods aim to identify a set of system performance indicators, and then monitor the impact of interventions to ensure the desired outcome of decisions.

At the operations level, the need for systems-level optimization is also apparent [3, 4, 6]. In 2010, IBM economists estimated that inefficiencies were costing the world US$15 trillion annually in waste and lost resources [4]; with more than 50% of food supplied never being eaten, nearly 35% of water designated for agricultural use being wasted, 25% of electricity generated never being used, and inefficiencies of over 35% in the Healthcare, Government and Safety and Education sectors. Over and above the direct costs, [4] the indirect costs associated with these inefficiencies had secondary effects in other areas such as consumer spending, pollution and quality of life indicators. These huge levels of inefficiency were attributed to managing city systems from within sectoral silos and while some level of inefficiency is inescapable, economists estimated an annual global savings potential of US$4 trillion should a systems approach be adopted [4].

Based on the above analysis, three key management principles, namely *effective stakeholder engagement, adaptive management* and *sustainability-led systems-level optimization*, have been identified as vital for effective and efficient integrated city planning and management across the full activity life-cycle. These principles, as implemented in Table 1, will be used as a reference framework when reviewing the effectiveness of existing and emerging tools for integration.

4 EXISTING AND EMERGING TOOLS FOR INTEGRATION

4.1 Environmental Assessment and Integrated Decision Making in South Africa

4.1.1 Legislation and Policy Tools. In South Africa, integrated city planning and management is supported mainly through the implementation of Integrated Development Plans [26] and their supporting Spatial Development Frameworks [27] required by the Local Government: Municipal Systems Act (32 of 2000) [28] and by the application of Integrated Environmental Management principles and tools, promoted and regulated through the National Environmental Management Act (107 of 1998) [24].

The main aim of an Integrated Development Plan (IDP) is to accelerate service delivery in municipalities and to deliver the spatial, social, ecological and economic urban patterns that are in line with the country's democratic and sustainability visions [29, 30]. An IDP is a principal strategic development plan prepared by local government for a five year period, which guides and informs all planning, budgeting, management and decision-making in a municipality [26]. The purpose of an IDP is to identify the development needs and concerns in a municipality and then formulate and prioritise possible intervention programmes and projects in an inclusive and integrated manner, taking into accout Key Performance Indicators (KPIs) and targets, stakeholder concerns, spatial development considerations, budgeting constraints and legislative and sectoral considerations [26].

A Spatial Development Framework (SDF) is a core component of an IDP and consists of an in-depth mapping of the bio-physical, socio-economic and built environment status quo, patterns and trends in a municipality [27]. It quantifies the spatial development needs (e.g. housing and mobility) and infrastructure capacities within the municipality and qualitatively assesses the performance of the municipality against desired spatial form and principles [27]. The purpose of a SDF is to identify the location of spatial tools (e.g. nodes, corridors, infill and densification, containment, protection and growth areas) that are required to meet a municipality's priorities [27]. It is imperative that both the IDP and the SDF are guided by comprehensive stakeholder engagement programmes [26, 27].

Integrated Environmental Management (IEM) is a South African term that is equivalent to the globally applied term Environmental Assessment and Management [25]. IEM aims to support sustainable development through the use of a wide range of environmental assessment and management tools – such as Environmental Impact Assessments and Environmental Management Systems – throughout the full activity life cycle and by all sectors of society [25]. An overview of commonly used IEM tools at each stage in the activity life cycle is shown in Fig. 2 [25].

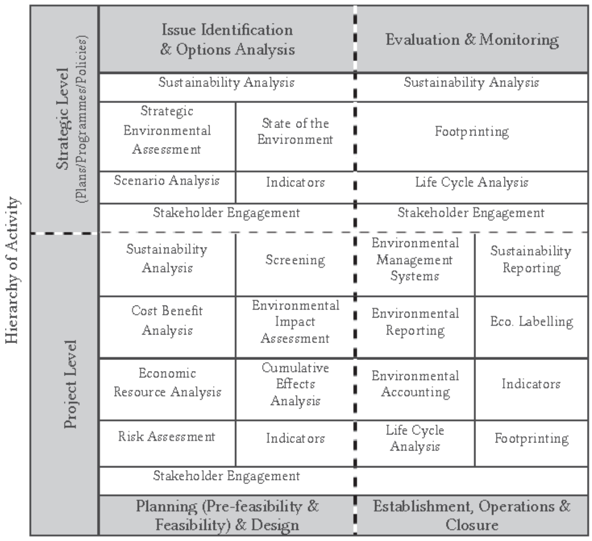


Figure 2. An overview of commonly used IEM tools at each stage in the activity life cycle. Source: Adapted from [25].

Although the National Environmental Management Act (NEMA) (107 of 1998) [24] promotes the application of a wide range of IEM tools [24, 25], project-level Environmental Impact Assessment (EIA) [31, 32] and more recently, strategic level Environmental Management Frameworks (EMF) [33] are the only regulated tools for IEM in South Africa [34].

Project-level EIA requires that the social, biophysical and other impacts associated with certain listed activities be identified, predicted, evaluated and mitigated prior to major decisions being made [5, 32]; while strategic-level EMFs are aimed at providing context for project-level environmental authorizations [34, 35].

EIAs are prepared by independent consultants, namely Environmental Assessment Practitioners (EAPs), who typically compile a report based on a number of specialist impact studies identified in the scoping stage of the EIA [32]. Listed activities may only proceed once the relevant government Competent Authority approves the EIA and its accompanying Environmental Management Plan (EMP) aimed at avoiding and mitigating potential impacts.

The purpose of the EMF is to function as a support mechanism in the EIA process and to inform decision-making regarding land-use planning applications in IDPs and SDFs [33]. By performing a status quo assessment (including sensitivity analysis, environmental opportunities and constraints), the EMF makes significant and detailed spatial information available about a specific geographical area. In addition to the status quo, EMFs identify the desired state of the environment and propose the way forward by identifying specific management zones and management guidelines. Management guidelines include limits and cumulative impacts, the identification of existing impacts to be addressed, the identification of the scope of potential impacts and information needs of EIAs and the identification of activities requiring EIAs in delineated geographical areas.

4.1.2 Challenges in Practice. Despite the positivity embodied in the introduction of the key policy tools aimed at achieving sustainable socio-economic reform in South Africa, there exists a chronic disconnect between planning and implementation [29, 34]. The current IEM system is often criticized as being inefficient and ineffective [34] and IDPs have consistently failed to achieve transformation [29]. The number of civic protests in South Africa is on the increase, with over 100 protest incidents every year [30]. The level of violence associated with these protests is also rising [30]. The main reason cited for violent protests has been dissatisfaction with municipal service delivery [30].

Effective service delivery has been adversely affected by a crippling lack of capacity in local government [30], staggering estimates of corruption and fruitless expenditure [30], political agendas and factional battles within governing parties [30]. Although to a lesser degree, similar concerns regarding quality and ethics have plagued the implementation of EIA and EMF [34, 36].

Across all activity levels there is a general lack of understanding and internalization of sustainability principles [34, 37]. *Sustainability-led development* aims to shift the focus of environmental management away from impact mitigation and instead, aims to focus on synergistic solutions that optimize positive socio-economic and ecological outcomes. However, it was found that the main focus of environmental intervention in South Africa is on impact identification, mitigation and compliance and sustainability-led development has not been embraced [34, 37]. This is made evident by the limited scope of KPIs used in IDPs [38, 39] and the limited focus on environmental issues in IDPs [37].

Another concern is the lack of *integration* and *effective cooperative governance.* Sector plans are still largely drawn up independently when preparing IDPs and "the level of integration is determined by the degree to which municipal departments talk to one another" [29]. In addition, there is a lack of coordination across the spheres of government, resulting in fragmented and sometimes conflicting, planning and implementation [29]. At project-level EIA, fragmentation and duplication of authorization processes often lead to frustration [34]. Even in EIA reporting, there is a lack of integration, with numerous specialists' reports being prepared independent of each other [40].

A major flaw in IEM in South Africa is the absence of *adaptive management* and *monitoring* [34]. The IDP has been criticized for its failure to acknowledge limited foresight and to deliver a framework that fosters adaptive planning approaches [29]. Such a framework would need to facilitate responsive scenario analysis and ongoing monitoring, learning and adaptation [41].

Meaningful public engagement is one of the cornerstones of sustainable development [42], fundamental in achieving bottom-up sustained transformation. However, a large body of literature records the failure of IDP and IEM processes to achieve meaningful stakeholder engagement [29, 34, 42, 43]. Failure has been attributed to the non-functionality of ward committees aimed at facilitating engagement [29, 43], coupled with a low sense of ownership of development initiatives and capacity of citizens to participate [43].

Although the IDP only facilitates citizen participation in the planning phase of development, there has been a call for participatory monitoring and evaluation aimed at involving citizens and other stakeholders in the implementation of IDPs [29]. It is envisaged that such an initiative will facilitate accountability, co-creation and networked solutions [29].

4.1.3 Proposed Information Systems Solutions. In 2014, after reviewing the state of IEM in South Africa, the Environmental Impact Assessment and Management Strategy (EIAMS) was developed, delineating the pillars of effective IEM [34]. It was suggested in the EIAMS that the generic guiding principles of sustainability described in NEMA be elucidated through the development of *clear sustainability objectives, indicators and targets* at national and local government levels [34]. It is hoped that clear sustainability targets will provide the necessary strategic vision to promote sustainability-led development, and the required framework for cooperative and efficient decision-making, performance monitoring and adaptive management [34].

In addition, the EIAMS [34] also called for *effective environmental information management*. Reliable, current, publicly available and understandable environmental information is an essential requirement for effective decision-making and participation by all role-players in the IEM processes. Effective environmental information management is therefore critically important to the success of IEM. While a wide variety of data and information is being generated by various government departments, industry, non-governmental organizations and consultants, there are no information systems fully implemented to collate the data and make it publicly available [34]. The Department of Environmental Affairs has taken steps to address this gap, including the Environmental GIS website (http://egis.environment.gov.za), but the solutions are not yet adequate [34]. The EIAMS [34] calls for a centrally maintained catalogue of available information, and the development of appropriate data standards.

Table 1 indicates how current policy tools are supporting the three key management principles aimed at effective and efficient integrated city planning and management (identified in Section 3.3) at each stage of the activity life cycle; and indicates how the information systems interventions proposed in the EIAMS [34] will meet current shortfalls in the implementation of these principles.

Table 1: Current policy tools and environmental information needs for IEM across the development cycle.

Stage in the Development Cycle	Key Management Principle	Current Policy Tool	Proposed Supporting Environmental Information Needs Identified in the EIAMS for South Africa**
Strategic- & Project-level Planning	Stakeholder Engagement	*Strategic-level:* IDP, SDF, EMF *Project-level:* EIA supported by EMF	**Spatial data***: • current state and trends of the environment • desired state of the environment, including sustainability objectives, indicators and targets * Data to be continuously updated with feedback from monitoring and new information acquired during impact assessment studies. **Tools for scenario analysis**
	Adaptive Management		
	Sustainability-led Systems-level Optimization		
Implementation, Monitoring, Auditing, Enforcement & Feedback	Stakeholder Engagement	-	**Community feedback:** • Community and other environmental monitoring forums • Registering of complaints and whistleblowing
	Adaptive Management	EIA & accompanying Env. Management Plan indicating mitigation requirements Optional implementation of Env. Management Systems (ISO 14001)	**Documents***: • EIA and EMP reports • Environmental Authorization and associated conditions of approval • Compliance notices • Norms, standards and legislative requirements • Site plans and layout plans • Baseline monitoring data *Information generated is often in the form of difficult-to-access reports. Standardizing the requirements for digital formats will facilitate accessibility to documents. **Routine monitoring data** **Reporting:** • trend identification • non-compliance and incident reporting • Need for adaptive management intervention
	Systems-level Optimization	-	-

**EIAMS is the Environmental Impact Assessment and Management Strategy developed for South Africa in 2014 [34].

4.2 System of Systems Thinking in Smart Cities

4.2.1 The IoT, Big Data and Smart Cities. There is some ambiguity in the definition of smart cities. SALGA [14] quoted Joe Bignan from the Economist in saying that "different industries approach the subject from their comfort zones. IT companies define a smart city through a technology lens; developers concentrate on physical infrastructure; utilities insist it is about sustainable energy; and the green lobby champions the environment. Smart Cities are all of the above." After reviewing 23 definitions of smart cities [44] from a range of fields, including architecture, urban planning, engineering, business and information management, it is evident that most define smart cities as those whose goals are to achieve sustainability – social, economic, and environmental – and who make use of technology to achieve these goals.

Connectivity is fundamental to smart cities. Smart people are connected through smart phones and infrastructure and the urban environment are connected through the Internet of Things (IoT). The IoT is the name given to the growing trend in which large numbers of networking sensors are embedded into various devices, enabling information-gathering and control functions [45]. This ubiquitous connectivity allows for real-time monitoring and management of infrastructure and citizens.

The world is undergoing a paradigm shift in its view and use of data [46]. For the first time, as a result of on-line and off-line datafication, massive amounts of data is available about many aspects of citizen's lives and the computing power to analyze this data has become readily available, even in the cloud (cloud computing). Big Data is the name given to the unprecedented amount of data being generated from the datafication of many aspects of our lives and is defined in terms of the challenges that these new data streams pose. Big Data is frequently defined with reference to the main "3Vs" of volume, variety, and velocity [47] and the two additional "Vs" of value [45] and variability [1, 47].

By combining real-time monitoring, event management, data analytics and advanced citizen engagement, cognitive technologies are leveraging the IoT and Big Data to radically reduce inefficiencies in all government sectors including smart buildings, healthcare, education, emergency management, public safety, city planning and operations, government administration, water, transportation and energy [1, 2].

4.2.2 System of Systems Thinking. Despite the explosion in smart city solutions over the last decade, the transformation of cities is not following as expected [3]. Optimization and integration in historical verticals (sectoral silos) is the core of today's smart cities projects and it is being proposed that this silo approach to smart cities development is reaching its limits [3]. Practitioners are calling for a coordinated approach and recently, much work is being done on developing appropriate frameworks aimed at eliminating systemic inefficiencies and fostering integrated decision-making, interoperability and collaboration [3, 4].

4.2.3 Standards. As the concepts and practices of smart cities emerge and mature, standards developing organizations (SDOs) are developing standards to promote best practice. Initially, standards have been focusing on sectoral best practice such as building, energy, wastewater, smart grids and intelligent transportation systems standards [48]. Emerging standardization activities, however, are focusing on the 'bigger picture' [3, 48], and a number of SDOs are developing specifications aimed at fostering a system of systems approach to management [3, 49-52].

The International Electrotechnical Commission (IEC) [3] identified three key standard requirements to orchestrate smart city infrastructure: namely, data standards focusing on data format and security [53]; technical standards [54, 55] focusing on the integration of interoperable infrastructures and services; and management standards.

Management standards are aimed at fostering integrated decision-making and collaboration between stakeholders by providing mutual communication tools necessary for creating a common vision, benchmarking, knowledge transfer, quality assurances, project assessments and collaboration between different operators and service providers [3]. Thus far, management standards have focused on developing conceptual models of the city as a system of systems [15, 16, 50] and performance metrics [11, 56-59].

4.2.4 Conceptual Models. The purpose of conceptual models is to create a common visual understanding of the core components of a city, and their interactions [15, 16, 50]. Fig. 3 is an illustration of a conceptual model developed by the City Protocol Society [15] termed the "City Anatomy". "The City Anatomy offers a common language describing the city ecosystem as: a set of physical structures, the living entities that make up a city's society, and the flow of interactions between them" [15]. The City Anatomy is an example of a conceptual model aimed at creating a common visual understanding of the core components of a city and their interactions.

4.2.5 Performance Metrics. Now, more than ever, sustainable urban planning and management is dependent upon evidence-based decision-making [11]. Big Data generated through sensing platforms contains massive amounts of city information at several levels [57]. However, the effective use of this data in decision-making has been hampered by a lack of integration and a clear vision [3, 57].

Recently, attention has been given to developing KPIs aimed at steering city planning and management activities towards meeting sustainable development goals [9, 60] and measuring and monitoring smart city performance with respect to these goals [11, 56-59]. In addition to setting clear development targets and creating a framework for prioritizing city challenges, globally comparable KPIs are essential for comparative learning across cities and evaluating the impact of interventions [11, 57].

Typically, standards focus on developing a comprehensive set of KPIs to measure a city's social, economic, environmental and governance performance and resilience [11, 56-59]. An example of a set of KPIs is the City Protocol Society's "City Anatomy Indicators" [57] shown in Figure 4. The City Anatomy Indicators are an expansion of the indictors proposed by the International Organization for Standardization [56] and are aimed at assessing the various sub-systems in the City Anatomy framework [15].

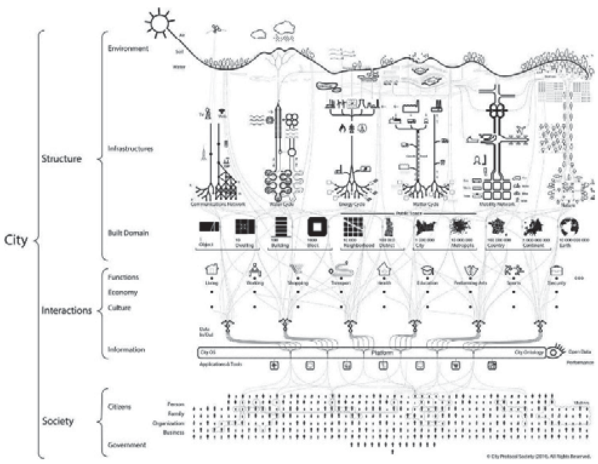


Figure 3. Bird's-eye view of the City Protocol Society's "City Anatomy". Source: [15]. The City Anatomy is an example of a conceptual model aimed at creating a common visual understanding of the core components of a city and their interactions. To download a readable copy of the "City Anatomy", visit: cityprotocol.org/publications/

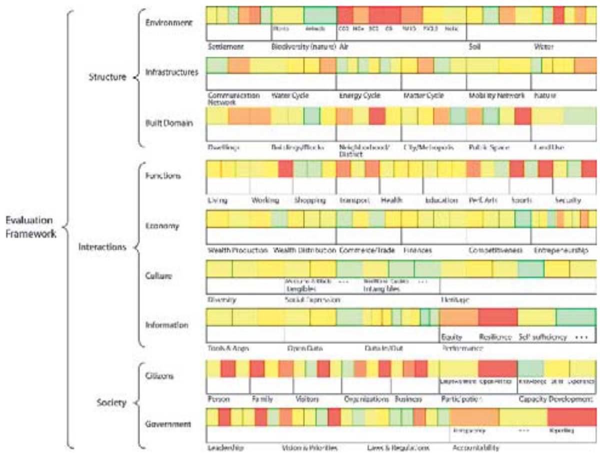


Figure 4. Dashboard view of the City Protocol Society's "City Anatomy Indicators". Source: [57]. To download a readable copy of the "City Anatomy Indicators", visit: cityprotocol.org/publications/

KPIs allow the linking of short-term and long-term goals through metrics [3, 57]. The City Anatomy Indicators are calculated using real-time data generated from city management platforms [55] (see Section 4.2.6), and allow the evaluation of city metrics either from a short-term operations perspective (e.g., in emergency management situations) or from a long-term strategic perspective. Fig. 4 depicts the high-level evaluation framework of the City Anatomy Indicators as a dashboard view where city functioning and status quo is visualized using green, yellow and red indicators for the various systems and subsystems that form a city [57]. Anatomy Indicators provide an integrated suite of indicators and targets for city evaluation and transformation. The evaluation framework with the top level indicators is shown in Fig. 4.

4.2.6 Integrated City Management Platforms. A fundamental component of the City Anatomy [57] introduced in Section 4.2.4 is the informational or systems platform aimed at facilitating knowledge acquisition and information transfer between trans-disciplinary systems. These systems platforms, also known as integrated city management platforms [3], support integrated decision-making through the automated monitoring of city performance indicators for situational analysis; and tools and applications for system-level data analysis and representation, decision support and management actions [15, 55]. Example integrated city management platforms are the Intelligent Operations Centre solutions of IBM [61, 62] (Fig. 5) and Huawei [63, 64].

Figure 5. IBM Intelligent Operations Center. Source: [61].

5 GAP ANALYSIS

5.1 Analysis of Existing Tools for Integration

Analysis of the existing framework for environmental assessment and integrated decision making in South Africa (Section 4.1) showed a significant lack of vision and coordination in the IDP and IEM processes, which are the primary tools for sustainable socio-economic development in the country. Further challenges include a chronic lack of capacity within local government, corruption and poor stakeholder engagement processes.

Recently however, the Environmental Impact Assessment and Management Strategy (EIAMS) was developed, delineating the pillars of effective IEM [34] and proposing the way forward with regards to improving IEM practices. Table 1 summarizes the environmental information needs identified in the EIAMS report [34]. The information needs for each stage of the activity life-cycle are indicated in Table 1 along with existing policy tools implemented at each level. The information is tabulated with reference to the three key management principles identified in Section 3.3, namely *effective stakeholder engagement*, *adaptive management* and *sustainability-led systems-level optimization*. A review of the information needs recommended in the EIAMS strategy [34] indicates that the proposed interventions will significantly alleviate the lack of vision and coordination inherent in the IEM system, by the introduction of clear performance indicators and targets and a framework for adaptive management and monitoring.

Two possible criticisms of the EIAMS specifications [34], however, is the limited emphasis on stakeholder engagement tools at the planning level and the absence of recommendations for visualization tools aimed at fostering operations-level systems optimization among stakeholders. The latter is the back-bone of modern smart city solutions and a likely fixture in future city operations, yet IEM remains limited in its adoption of operations-level environmental management tools.

A major weakness in the IDP and IEM processes is the inability to achieve meaningful citizen engagement. As mentioned in Section 4.1.2, failure has been attributed to the non-functionality of ward committees aimed at facilitating engagement [29, 43], coupled with a low sense of ownership of development initiatives and capacity of citizens to participate [43]. Ward committees, chaired by the ward councilors, are made up of members representing various ward interests [65] and are the major legislative vehicle through which citizens can participate in the development of IDPs [29]. Ward committees, however, are highly politicized and fraught with power struggles [29, 30] and in their current state, are not effective platforms for engagement [29, 30]. In order to achieve effective public engagement, citizens need to be appropriately capacitated with meaningful information regarding their rights, and sustainability and development issues [34, 42]. Appropriate smart stakeholder engagement tools could be very effective in transforming the state of stakeholder engagement in South Africa. Particularly in light of substantial efforts by local government to provide free Wi-Fi access at strategic locations across most of the country's major cities [14].

5.2 Analysis of Emerging Tools for Integration

It is evident that recent developments in smart city conceptual frameworks and performance metrics (Section 4.2) are in direct alignment with the EIAMS call for a common vision and coordination framework [34]. With reference to Table 1, it is apparent that smart city system-of-systems solutions are effectively geared toward meeting all the proposed environmental information needs specified in the EIAMS report [34], as well as meeting the gaps identified in this research (Section 5.1).

What is not clear, however, is how effective the implementation of proposed conceptual models and performance metrics will be in practice. Sustainability KPIs are notoriously difficult to quantify [38] and caution needs to be taken with one-size-fits-all approaches. Another area requiring further research is the extent to which Big Data analytics can play a role in improving assessment and simulation tools used in decision-making. There is a rich history of urban computational modeling [66-68] that is already leveraging Big Data [8], much of it based on complexity science and focused on simulating urban form and flows. These models have applications in activities such as Spatial Development Frameworks and can greatly enrich the strategic planning process. However, there is significantly less focus on the development of computational multi-criterion evaluation methods [69-72] and the subsequent exploration of what Big Data has to offer with regards to sustainability-led systems-level optimization.

6 CONCLUSIONS AND FUTURE RESEARCH

The research objective of this study was to contextualize new smart city solutions for integrated city planning and management within the framework of South African policy tools, with the aim of identifying synergies and gaps for future research. This was achieved by carrying out a literature review of existing and emerging tools for integrated decision-making in South African cities and performing a subsequent gap analysis.

Analysis of existing tools (Section 5.1) showed a significant lack of vision and coordination in the Integrated Development Planning and Integrated Environmental Management processes in South Africa which are the primary tools for sustainable socio-economic development in the country. Further challenges included chronic lack of capacity within local government, corruption, and poor stakeholder engagement processes.

Recently, however, the Environmental Impact Assessment and Management Strategy (EIAMS) was developed, aimed at improving IEM practices [34]. A review of the environmental information systems recommendations made in the EIAMS strategy [34] indicate that the proposed interventions will significantly alleviate the lack of vision and coordination inherent in the IEM system, by the introduction of clear performance indicators and targets and a framework for adaptive management and monitoring. Concerns, however, were noted with regards to the limited emphasis on stakeholder engagement tools at the planning level, and the absence of recommendations for visualization tools aimed at fostering operations-level systems optimization among stakeholders. The latter being the back-bone of modern smart city solutions.

Analysis of emerging tools for integration (Section 5.2) showed that smart city systems solutions are effectively geared toward meeting all the proposed environmental information needs specified in the EIAMS report [34], as well as meeting the gaps identified in this research (Section 5.1).

What is not clear, however, is how effective the implementation of proposed conceptual models and performance metrics will be in practice, and the extent to which Big Data analytics can play a role in improving assessment and simulation tools used in decision-making. The answers to both these questions will likely be found in the context of existing practices for integrated planning and management which are well-rooted in complexity science and built upon decades of practice [8].

REFERENCES

[1] E. Al Nuaimi, *et al.* 2015. Applications of big data to smart cities. *Journal of Internet Services and Applications* 6, 1, 25.
[2] IBM. *Smarter Cities*. [Accessed: 30 June 2018] Available from: https://www.ibm.com/smarterplanet/us/en/smarter_cities/overview.
[3] IEC. 2015. *White Paper: Orchestrating Infrastructure for Sustainable Smart Cities*. Geneva, Switzerland: International Electrotechnical Commission.
[4] IBM. 2010. *The world's 4 trillion dollar challenge: Using a system-of-systems approach to build a smarter planet*. IBM Institute for Business Value.
[5] Wikipedia. 2018. *Environmental Impact Assessment*. [Accessed: 29 June 2018] Available from: https://en.wikipedia.org/wiki/Environmental_impact_assessment.
[6] G. Hardin. 1968. The Tragedy of the Commons. *Science* 162, 1243-1248.
[7] World Commission on Environment and Development. 1987. *Our Common Future*. Oxford University Press.
[8] M Batty. 2013. *Urban Informatics and Big Data: A Report to the ESRC Cities Expert Group*.
[9] United Nations. 2015. *Transforming our world: the 2030 Agenda for Sustainable Development (A/RES/70/1)*.
[10] The Economist. 2009. *Triple Bottom Line*. [Accessed: 26 July 2016] Available from: http://www.economist.com/node/14301663.
[11] P. McCarney. 2015. The Evolution of Global City Indicators and ISO37120: The First International Standard on City Indicators. *Statistical Journal of the IAOS* 31, 103–110.
[12] Rockefeller Foundation. 2016. *100 Resilient Cities*. [Accessed: 25 July 2016] Available from: http://www.100resilientcities.org.
[13] The Rockefeller Foundation | Arup. 2015. *City Resilience Framework*. [Accessed: 25 July 2016] Available from: http://www.100resilientcities.org/page/-/100rc/Blue%20City%20Resilience%20Framework%20Full%20Context%20v1_5.pdf.
[14] SALGA. 2015. *Smart Cities*. in.KNOW.vation [Accessed: March 2015] Available from: http://www.cpsi.co.za/wp-content/uploads/2015/05/FINALSALGAPUBLICATIONLOWRES.pdf.
[15] City Protocol Society. 2015. City Anatomy: A Framework to support City Governance, Evaluation and Transformation (City Protocol Agreement (CPA-I_001-v2)).
[16] ISO/IEC. 2017. ISO/IEC 30182:2017, Smart city concept model - Guidance for establishing a model for data interoperability.
[17] H. Haken. 2012. *Complexity and Complexity Theories: Do These Concepts Make Sense?* In *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning*, 7-20. J. Portugali, H. Meyer, E. Stolk, and E. Tan, Editors. Springer.
[18] OECD. 2009. Applications of Complexity Science for Public Policy: New Tools for Finding Unanticipated Consequences and Unrealized Opportunities.
[19] Wikipedia. 2018. *Complex system*. [Accessed: 28 June 2018] Available from: https://en.wikipedia.org/wiki/Complex_system.
[20] K. Kourtit, *et al.* 2014. A blueprint for strategic urban research: the urban piazza. *Town Plan Rev* 85, 1, 97-126.
[21] M. Angelidou. 2015. Smart cities: A conjuncture of four forces. *Cities* 47, 95-106.
[22] M. Batty and S. Marshall. 2012. *The Origins of Complexity Theory in Cities and Planning*. In *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning*, 21-45. J. Portugali, H. Meyer, E. Stolk, and E. Tan, Editors. Springer.
[23] R. Biggs, *et al.* 2015. Strategies for managing complex social-ecological systems in the face of uncertainty: examples from South Africa and beyond. *Ecology and Society* 20, 1, 52.
[24] DEAT. 1998. National Environmental Management Act 107 of 1998.
[25] DEAT. 2004. *Overview of Integrated Environmental Management, Integrated Environmental Management, Information Series 0*. Pretoria: Department of Environmental Affairs and Tourism (DEAT).
[26] RSA. *IDP Guide Pack*. Department of Provincial and Local Government.
[27] RSA. 2011. *Guidelines for the Development of Spatial Development Frameworks - Version 8*. Department of Rural Development and Land Reform.
[28] RSA. 2000. Local Government: Municipal Systems Act 32 of 2000.
[29] N. Ngamlana and R. Eglin. 2015. *Learning Brief #1: Spatial Planning and Integrated Development Planning*. afesis-corplan.
[30] L. Mnguni. 2016. *Why SA's municipalities are failing and how to fix them*. Rand Daily Mail [Accessed: 29 June 2018] Available from: https://www.businesslive.co.za/rdm/politics/2016-03-03-why-sas-municipalities-are-failing-and-how-to-fix-them/.
[31] RSA. 2012. *Companion to the Environmental Management Impact Assessment Regulations 2010 (IEM Guideline Series, Guideline 5)*. Department of Environmental Affairs.

[32] RSA. 2017. *Amendments to the Environmental Impact Assessment Regulations 2014*. Department of Environmental Affairs.
[33] RSA. 2012. *Environmental Management Framework Regulations 2010 (IEM Guideline Series, Guideline 6)*. Department of Environmental Affairs.
[34] DEA. 2014. *Environmental Impact Assessment and Management Strategy for South Africa (Draft)*. Pretoria: Department of Environmental Affairs (DEA).
[35] D.P. Cilliers and F. Retief. 2016. The extent and status of environmental management frameworks (EMFs) in South Africa, 2006–2015. *South African Geographical Journal* 99, 3, 283-300.
[36] M. Marais, *et al.* 2014. Environmental management frameworks: results and inferences of report quality performance in South Africa. *South African Geographical Journal* 97, 1, 83-99.
[37] S. Ruwanza and C.M. Shackleton. 2015. Incorporation of environmental issues in South Africa's municipal Integrated Development Plans. *International Journal of Sustainable Development & World Ecology* 23, 1, 28-39.
[38] L. Marais, F. Human, and L. Botes. 2008. Measuring what? The utilisation of development indicators in the integrated development planning process. *Journal of Public Administration* 43, 3, 376-400.
[39] H.M. Mautjana and O. Mtapuri. 2014. Integrated Development Plans without Development Indicators: Results from Capricorn District Municipalities in South Africa. *Mediterranean Journal of Social Sciences* 5, 8, 474-483.
[40] A. Morrison-Saunders, *et al.* 2014. Strengthening impact assessment: a call for integration and focus. *Impact Assessment and Project Appraisal* 32, 1, 2-8.
[41] R. Hummelbrunner and H. Jones. 2013. *A guide for planning and strategy development in the face of complexity*. Background note, Overseas Development Institute.
[42] D. Everatt, H. Marais, and N. Dube. 2010. Participation ... for what Purpose? Analysing the Depth and Quality of Public Participation in the Integrated Development Planning Process in Gauteng. *Politikon* 37, 2-3, 223-249.
[43] A. Aklilu, A. Belete, and T. Moyo. 2014. Analysing Community Participation in the Municipal Integrated Development Planning Processes in Limpopo Province, South Africa. *Mediterranean Journal of Social Sciences* 5, 25, 257-262.
[44] V. Albino, U. Berardi, and R.M. Dangelico. 2015. Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *Journal of Urban Technology* 22, 1, 3-21.
[45] M. Chen, S. Mao, and Y. Liu. 2014. Big Data: A Survey. *Mobile Networks and Applications* 19, 2, 171-209.
[46] C. O'Neil and R. Schutt. 2014. *Doing Data Science: Straight Talk from the Frontline*. Sebastopol, CA: O'Reilly Media, Inc.
[47] A. Gandomi and M. Haider. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35, 2, 137-144.
[48] American National Standards Institute. 2018. *Smart and Sustainable Cities*. [Accessed: 27 June 2018] Available from: https://webstore.ansi.org/smart-cities/default.aspx.
[49] ISO. 2016. ISO 37101:2016, Sustainable development in communities - Management system for sustainable development - Requirements with guidance for use.
[50] The British Standards Institution. 2014. PAS 181:2014, Smart city framework – Guide to establishing strategies for smart cities and communities.
[51] City Protocol Society. 2015. Livable Districts and Cities (CPC_004_Livable_Districts_and_Cities).
[52] ITU. 2016. *Shaping Smarter and More Sustainable Cities: Striving for Sustainable Development Goals*. Geneva, Switzerland.
[53] The British Standards Institution. 2017. PAS 183:2017, Smart cities – Guide to establishing a decision-making framework for sharing data and information services.
[54] ITU. 2016. *Unleashing the Potential of the Internet of Things*. Geneva, Switzerland.
[55] City Protocol Society. 2016. Open Sensor Platform (CPWD-PR_005_Open_Sensor_Platform).
[56] ISO. 2014. ISO 37120:2014, Sustainable development of communities - Indicators for city services and quality of life.
[57] City Protocol Society. 2015. City Anatomy Indicators (CPA-PR _002_Anatomy Indicators).
[58] ITU. 2018. *KPIs on Smart Sustainable Cities*. [Accessed: 25 June 2018] Available from: https://www.itu.int/en/ITU-T/ssc/Pages/KPIs-on-SSC.aspx.
[59] U4SSC. 2017. *Collection Methodology for Key Performance Indicators for Smart Sustainable Cities*. Geneva, Switzerland.
[60] U4SSC. 2017. *Implementing Sustainable Development Goal 11 by Connecting Sustainability Policies and Urban-Planning Practices Through ICTs*. Geneva, Switzerland.
[61] IBM. 2013. *IBM Intelligent Operations Center for Smarter Cities*. IBM Corporation.
[62] L. Zhuhadar, *et al.* 2017. The next wave of innovation—Review of smart cities intelligent operation systems. *Computers in Human Behavior* 66, 273-281.
[63] Huawei. 2018. *Intelligent Operation Center Solution*. [Accessed: 27 June 2018] Available from: https://e.huawei.com/za/solutions/industries/smart-city/ioc.
[64] fin24tech. 2017. *Smart cities will now have their own nervous system*. [Accessed: 27 June 2018] Available from: https://www.fin24.com/Tech/Opinion/smart-cities-will-now-have-their-own-nervous-system-20171116.
[65] C. Bathembu. 2016. *Municipal ward committees: What you need to know*. vukuzenzele (Sep 2016 - 2nd edition) [Accessed: 30 June 2018] Available from: https://www.vukuzenzele.gov.za/municipal-ward-committees-what-you-need-know.
[66] M. Batty. 2013. *The New Science of Cities*. MIT Press.
[67] M. Batty. 2012. Building a science of cities. *Cities* 29, S9-S16.
[68] Vahid Moosavi. 2015. Computational Urban Modeling: From Mainframes to Data Streams. In *Artificial Intelligence for Cities Workshop in AAAI Conference*. Austin, Texas.
[69] B. Mattoni, F. Gugliermetti, and F. Bisegna. 2015. A multilevel method to assess and design the renovation and integration of Smart Cities. *Sustainable Cities and Society* 15, 105-119.
[70] Benedetta Mattoni, *et al.* 2017. A quantitative evaluation of the mutual influences among Smart strategies applied at district level. *IEEE*.
[71] J.M. Schleicher, *et al.* 2016. *A Holistic, Interdisciplinary Decision Support System for Sustainable Smart City Design*. In *Smart Cities. Smart-CT 2016. Lecture Notes in Computer Science, vol 9704*, 1-10. E. Alba, F. Chicano, and G. Luque, Editors. Springer: Cham.
[72] J.-M. Fernández-Güell, *et al.* 2016. *How to Incorporate Urban Complexity, Diversity and Intelligence into Smart Cities Initiatives*. In *Smart Cities. Smart-CT 2016. Lecture Notes in Computer Science, vol 9704*, 85-94. E. Alba, F. Chicano, and G. Luque, Editors. Springer: Cham.