FLSEVIER

Contents lists available at ScienceDirect

Sustainable Cities and Society



journal homepage: www.elsevier.com/locate/scs

A modelling framework for integrated smart city planning and management



Lindsay Westraadt*, André Calitz

Computing Sciences, Nelson Mandela University, South Africa

ARTICLE INFO

Keywords: Smart cities Key performance indicators Integrated decision making Model design

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 (ICMPs). Despite the deluge of data generated by ICMPs and the accompanying growth in computing power, limited research has been conducted on exploring the use of this data to develop quantitative tools for integrated smart city planning and management. In this study, the Design Science Research process was followed to develop and evaluate a modelling framework aimed at exploring the use of ICMP data to identify synergies and dependencies across smart city sectors. This paper provides a summary of framework design and implementation and discusses the Design Science Knowledge gained from the exercise.

1. Introduction

As sustainability issues intensify worldwide (United Nations, 2015), city managers are being called to manage increasingly stressed resources with unprecedented efficiency (IBM, 2010; IEC, 2015). Globally, the Information Technology (IT) industry has stepped up to this challenge, and over the last decade there has been an explosion in smart city solutions (IEC, 2015). While smart city technologies continue to accumulate, the transformation of cities is not following at the anticipated speed and manner (IEC, 2015). It is believed that this discrepancy is due to a lack of a common strategic vision and collaboration across city sectors (IBM, 2010; IEC, 2015). In an effort to bridge this gap, observation of recent smart city trends shows a shift in focus from sector-based interventions towards integrated decision-making informed by Big Data (IBM, 2010; IEC, 2015; Kourtit, Nijkamp, & Steenbruggen, 2017; City Protocol Society, 2015a; Chourabi et al., 2012; Fernández-Güell et al., 2016; Schleicher et al., 2016; Mattoni, Gugliermetti, & Bisegna, 2015, 2017). This move towards integration is evident in the emergence of Integrated City Management Platforms (ICMPs) (City Protocol Society, 2015a, 2016; Cohen, 2014; Huawei, 2018; IBM, 2013; IEC, 2015; Zhuhadar et al., 2017). ICMPs aim to present 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.

Despite the deluge of data generated by ICMPs and the accompanying growth in computing power, limited research has been conducted on exploring the use of this data to develop quantitative tools for integrated smart city planning and management (Lombardi et al., 2012; Mattoni et al., 2015, 2017; Pelorosso, 2020; Schleicher et al., 2016; Westraadt & Calitz, 2018). While prevalent and emerging multi-criteria decision analysis techniques are effective in optimising multi-objective project-level decisions (Greco, Ehrgott, & Figueira, 2016; Gregory et al., 2012) and in solving clearly defined multi-criteria problems (Fujita et al., 2020), they are limited in performing the more strategic task of identifying cross-sector synergies and interdependencies that, once leveraged, are deemed necessary to catapult smart cities to a higher level of efficiency (IBM, 2010; Mattoni et al., 2017; Westraadt & Calitz, 2018).

In this study, the Design Science Research process (Johannesson & Perjons, 2012; Peffers et al., 2008) was followed to develop and evaluate a modelling framework for integrated smart city planning and management. The main aim of the study was to contribute to an understanding of how data generated by ICMPs can be used to identify cross-sector synergies and interdependencies; and how this knowledge can be implemented to improve the efficiency at which smart cities are managed. This paper provides a summary of framework design and implementation and discusses the Design Science Knowledge gained from the exercise.

https://doi.org/10.1016/j.scs.2020.102444

Received 13 December 2019; Received in revised form 3 August 2020; Accepted 9 August 2020 Available online 27 August 2020 2210-6707/© 2020 Elsevier Ltd. All rights reserved.

^{*} Corresponding author at: Department of Physics, PO Box 77000, Nelson Mandela University, Port Elizabeth, 6031, South Africa. *E-mail address*: lindsayw@mandela.ac.za (L. Westraadt).

2. Overview of framework design and implementation

2.1. Rationale and anticipated research contribution

Extensive 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 (Berends, Carrara, & Vollers, 2017; City Protocol Society, 2015b; IEC, 2015; ISO, 2016; ITU, 2016; Manyika et al., 2013; Open Data Charter, 2015; The British Standards Institution, 2014). 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 sectors and their interactions, and on developing globally comparable Key Performance Indicators (KPIs) aimed at setting clear development targets for each city sector (City Protocol Society, 2015a, 2015c; ISO/IEC., 2017; ISO, 2014; ITU, 2018; McCarney, 2015; The British Standards Institution, 2014; U4SSC, 2017).

As such, emerging smart city KPIs provide the common vision necessary for integrated smart city planning and management (City Protocol Society, 2015a, 2015c). The focus of this study was to complement smart city KPI frameworks by developing a quantitative tool aimed at fostering cross-sector collaborations. Specifically, this paper describes the development and evaluation of a modelling framework that fosters cross-sector collaborations by quantifying dependencies between city sectors and identifying common cross-sector goals. To limit the scope of the investigation, the study focused on only one aspect of smart cities, namely crime management.

In this paper, the notion of integrated city planning and management refers to the attainment of synergistic multi-sector solutions to complex smart city challenges (Gregory et al., 2012). This implies the attainment of win-win solutions that efficiently use limited resources to simultaneously benefit more than one city sector. Furthermore, the notion of integrated planning and management implies the identification and accommodation of the social, economic and/or environmental externalities of activities (DEAT, 2004). It follows, that an integrated approach to crime management encompasses the identification of common goals as potential areas of collaboration between city sectors tasked with crime management and other city sectors; and the identification of unintentional externalities from activities in other sectors that may have a negative impact on crime management. The goal of this study was to develop and evaluate a modelling framework that would support this multi-sector approach to crime management.

2.1.1. Framework criteria

The design criteria (Johannesson & Perjons, 2012; Peffers et al., 2008) guiding the development and evaluation of the modelling framework are listed below:

- **Criterion 1:** To identify common goals across multiple city sectors, the modelling framework would need to incorporate the development goals of all relevant city sectors, in addition to the traditional targets for crime management. Furthermore, the framework would need to include a model that determines the relationships between the goals of crime management and those of other city sectors.
- **Criterion 2:** To identify possible negative impacts of activities in other sectors on crime, the modelling framework would need to include a predictive model capable of modelling the dependence of crime on other city sectors. In order to meet this criterion, key features in each sector which have the potential to influence crime would need to be identified.
- Criterion 3: The model(s) used need to leverage the impending deluge of data generated by ICMPs. Currently, it is standard practice to base city planning and management decisions on a heuristic understanding of the behaviour of complex city systems (Gregory et al., 2012; NMBM, 2018; Westraadt & Calitz, 2018). This is in part due to a lack of available data (Westraadt, Calitz, & Cullen, 2019). As ICMPs

and open data portals become common place (Berends et al., 2017; City Protocol Society, 2015b; IEC, 2015; ISO, 2016; ITU, 2016; Manyika et al., 2013; Open Data Charter, 2015; The British Standards Institution, 2014), there is an opportunity to explore the development of data-rich models capable of testing common assumptions about the interdependencies between city sectors.

• **Criterion 4:** The model(s) used would need to be able to accommodate anticipated complex interactions between city sectors and their underlying features (Allen, 1997; Batty & Marshall, 2012; Fernández-Güell et al., 2016).

2.1.2. Proposed modeling framework

It was anticipated that these criteria could be met through the following proposed modelling framework (Fig. 1):

- Input Data: Emerging smart city KPI frameworks (City Protocol Society, 2015c) provide a set of sectoral KPIs and targets, and thereby provide a means of identifying the development goals of all city sectors. Furthermore, a review of key variables influencing crime (Westraadt, 2019) showed that many of these variables overlap with sectoral indicators incorporated into existing city KPI frameworks. It was therefore decided to select a subset of smart city indicators that correspond to commonly used predictors of crime as input to the proposed model. By so doing, the model would meet **Criteria 1 and** 2. Furthermore, this choice of input features would meet **Criterion 3** as ICMP management standards prescribe the collection of KPI data.
- **Predictive Model:** It was anticipated that the modelling requirements of both **Criteria 1 and 2** could be met by developing a single predictive model that predicts crime rates as a function of the input data described above. Since the complexity of these relationships was unknown, it was expected that **Criteria 4** could be met by using an artificial neural network to develop the predictive model.
- Sensitivity Analysis: It was anticipated that synergies between city sectors could be identified by performing a sensitivity analysis on model output to determine the relative sensitivity of crime rates to KPIs in other city sectors. Here, KPIs that strongly influence crime rates may constitute common cross-sector goals (Criterion 1) or, in contrast, may indicate possible externalities that need to be addressed (Criterion 2).

2.1.3. Anticipated research contribution

The goal of this study was to develop and evaluate a modelling framework that would support an integrated approach to crime management. The anticipated research contribution is to evaluate the validity of the modelling framework proposed in Section 2.1.2 and to demonstrate the implementation and anticipated benefits thereof. The main research questions answered by this study (see Sections 3 and 4) are listed below:

RQ 1. Is there sufficient overlap between predictors of crime and smart city KPI frameworks to validate the use of smart city KPIs to predict crime?

RQ 2. Was the model effective at predicting crime?

RQ 3. Was the modelling framework effective in identifying crosssector synergies and externalities?

RQ 4. Did the modelling framework identify any discrepancies in common assumptions about the interdependencies between city sectors?

RQ 5. How can the knowledge gained in this study inform KPI tracking and data collection in smart cities?

2.2. Input data

As indicated in Section 2.1.2, the input data used in the proposed modelling framework was to consist of a subset of smart city indicators



Fig. 1. A modelling framework for integrated crime management in smart cities. (a) KPIs from traditionally isolated management silos are used as input data in (b) a crime prediction model. (c) Sensitivity analysis is carried out on model output to determine the relative influence of sectoral KPIs on crime. Source: Author's own construction.

that correspond to commonly used predictors of crime. Readily available open data for New York City (NYC OpenData) was used to develop and demonstrate the proposed modelling framework. NYC data was chosen since the NYC OpenData portal was one of the most comprehensive city data portals available, covering all major city sectors at spatial resolutions extending beyond the typically reported citywide averages. The indicators used in this study are listed in Table 1. Indicators were selected from the smart city KPI framework proposed by the City Protocol Society (2015a), (City Protocol Society, 2015c)2015c), guided by predictors of crime commonly employed in crime forecasting models (Perry et al., 2013; Westraadt, 2019).

The availability of data at the required spatial unit of modelling was another factor limiting the selection of indicators. City planning most often makes use of citywide annual trends in KPIs and situational indicators to identify challenges and inform decisions (de Blasio, Fuleihan, & Newman, 2018; MOO, 2018; NMBM, 2018). This practice is reflected in the spatial and temporal resolution of data typically available on city open data portals. In contrast, there is a large spatial variation in socio-economic and environmental trends across a city. In order to resolve the diverse spatial pattern of crime and its associated predictors across a city, a smaller spatial unit of analysis was necessary for the modelling framework proposed in this study. Spatially, indicator data was aggregated at the Public Use Microdata Area (PUMA) (Fig. 5) geographic unit. The PUMA (or the associated community district) was one of the most commonly used spatial units of reporting in the NYC open datasets, and was the smallest statistical geographic unit with sufficient annual data to meet the needs of the study.

For each variable (listed in Table 1), a table of annual measures for each PUMA was created for the years 2006–2017. These were then combined into a single table with columns representing each variable. The data table used in the study, therefore, 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 2.3), and predictions were based purely on location features. Furthermore, in order to eliminate bias introduced by the wide range of scales used among the selected input features, each variable was first standardised (Z-score) before proceeding with model development (Hair et al., 2014). Further descriptions of data sources, choice of key performance indicators, data preparation and the associated management challenges and implications are described elsewhere (Westraadt, 2019; Westraadt et al., 2019).

In general, the indicators listed in Table 1 are not exact implementations of City Protocol Society (CPA) indicators (City Protocol Society, 2015a, 2015c), 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 were 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).

To visualise the diversity of domains included in the development of the prototype model, the associated NYC agency is specified in Table 2 for each indicator. The anticipated outcome of model development was to include a diverse range of stakeholders in the decision-making process, thereby fostering synergistic solutions that are often overlooked when problems are solved from within sectoral silos.

2.3. Predictive model

As indicated in Section 2.1.2, the modelling framework prescribed the development of a predictive model that predicts crime rates as a function of the selected smart city KPIs identified in Section 2.2. Since the relationship between KPIs was unknown and expected to be complex (Allen, 1997; Fernández-Güell et al., 2016), artificial neural networks were employed to model the inter-dependencies between variables (Han, Kamber, & Pei, 2012; Tan, Steinbach, & Kumar, 2006). Artificial neural networks are well known for their ability to automatically approximate any function (Han et al., 2012; Tan et al., 2006), and are often used when the relationship between variables is unknown or complex.

Key design parameters that influence how well a trained neural network generalises to new data include the choices of network typology, initial weights and biases, and regularisation constant (Han et al., 2012; Tan et al., 2006). The tuning of these hyperparameters is often performed manually. Consequently, finding the optimal model for a given problem is a trial and error process that is often computationally expensive and time consuming. To circumvent these challenges, software tools exist to automatically optimise network hyperparameters. In this study, one such tool, namely the Model Manager software package developed by Sourmail (2002), (Sourmail, 2004)2004, was used to develop a set of prototype models for crime management in smart cities.

A simple feed-forward neural network with one hidden layer forms the basis of models developed using the Model Manager software package (Sourmail, 2002, 2004). Since the models were to be used to make numeric predictions, only one output unit was used. The main feature of the Model Manager software is its use of a Bayesian learning algorithm to automatically infer the optimal regularisation constant

Table 1

Indicators used in this study.

ID	Agency*	Indicator(s) used	Variable name
1 2	CECM DOF	Events per 100k population Assessed (commercial/ residential) property values relative to citywide average assessed (commercial/ residential) property values	events V1 (residential), V5 (commercial)
3	MOEO	Unemployment rate (%)	unemployment
4	MOEO	Theil's T inequality index (within PUMA**s/between PUMAs) (World Bank, 2014)	ineqT1r (within), ineqT2r (between)
7	DOE	Percentage of population without high school diploma	noHigh
8	DOE	Percentage of population with higher education degrees	degree
11	DYCD	Percentage single female householders	female
12	ACS	Number of credible abuse/ neglect investigations per 100k population	abuse
13	DYCD	Fertility rate per 1000 women aged 15-44	fertility
14	CCRB	Total civilian complaints against uniformed members of the New York City Police Department per 100k population	integrity
15	HRA; DOHMH	Adult New Yorkers without health insurance (%)	insurance
16	DOHMH; NYPD	Drug crimes per 100k population	drugs
20	DHS	Total number of 311 requests related to homeless encampments and panhandling per 100k population	homeless
21	NYCHA	Percentage social housing	socialHousing
22	DOT	Average number of street lights out per day per unit area	SL
24	DCP	Percentage land use type	P1 – P11***
20	DCP	Bank, 2014)	diversity
33	NYPD	Larceny (street/residential/ commercial) per 100k population	larStreet, larCommercial, larResidence
34	NYPD	Robbery (street/residential/ commercial) per 100k population; Assault (street/ residential) per 100k population	robStreet, robCommercial, robResidence, assStreet, assResidence
36	DOT	Pedestrian volume index	pedIndex
3/	DSNY: EDC	orainti reports per 100k	grann
38	DOHMH; DEP	Fine particulate matter (PM2.5) concentration	РМ

* Abbreviations for NYC agencies are defined in Table 2.

** PUMA: Public Use Microdata Area - statistical geographic unit (Westraadt, 2019).

*** Land use codes. See Westraadt (2019) for detail.

from the input data (Shahriari et al., 2016), thereby removing 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 Sourmail, 2002; Sourmail, 2004. The Model Manager training process involves the training of multiple models with different numbers of hidden units (typically 1–25) and different priors on the weights (typically 5). The models are then used to make predictions on an 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 (Sourmail, 2002, 2004).

In addition to automatically inferring the optimal regularisation constant, inherent in the Bayesian optimisation algorithm, is a means of

Table 2

Indicators per NYC agency. *Indicator IDs are specified in Table 1.

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

quantifying the uncertainty in model predictions for a given set of weights (MacKay, 1992). In general, while neural networks provide the best-fit function to a set of data, they do not describe the uncertainties in defining the fitting function in regions of the input space where data are sparse or where the data are noisy (Bhadeshia, 1999; Shahriari et al., 2016). Depending on the chosen hyperparameters, 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). The Bayesian Neural Networks (BNNs) (MacKay, 1992) developed using the Model Manager software (Sourmail, 2002, 2004) allow for the calculation of error bars representing this uncertainty (see dashed lines in Figs. 9–12). This knowledge can further inform decision-making by indicating the reliability of model predications.

In this study, two BNNs were developed for crime management in smart cities. The BNNs were implemented using the Model Manager software package developed by (Sourmail (2004)), and the set of standardised input features listed in Table 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. The rationale behind excluding race from the set of input features is discussed in Section 3.2.

The models were developed to predict street larceny and street robbery in New York City, respectively. The choice of crimes was intentional, with the aim of encapsulating as wide a range of system behaviour as possible. As seen in Figs. 6 and 7, the two crimes exhibit different spatial patterns. They also represent different types of crime, namely property crime (larceny) and violent crime (robbery). The performance of the developed neural networks is described in Section 3.1.

2.4. Sensitivity analysis

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 network 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). In order to identify synergies between city sectors (see Section 2.1), the BNNs were combined with sensitivity analysis (Han et al., 2012) to determine the relative sensitivity of crime rates to KPIs in other city sectors. Typically, as part of a sensitivity analysis, an input variable is varied while the remaining input variables are fixed at some value. The changes in the network output are then observed. The results of the sensitivity analysis are shown in Section 3.3 for each model, together with a demonstration of its anticipated application in practice.

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" (Section 3.2). 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 3.3. Exploratory Factor Analysis was used to identify latent "states" within the input data. The process followed, and the observed states, are discussed in Section 3.2. In summary, an overview of the model development and implementation process is shown in Fig. 2.

3. Results and discussion

3.1. Predictive models

A visualisation of the accuracy and precision of the neural network predictions (see Section 2.3) is shown in Fig. 3. In general, the accuracies of the neural network predictions were comparable to those made by simpler linear regression models. However, the neural networks performed better than the linear models in sparsely populated regions of the input space (that is, where the standardised target values deviated significantly from the norm). The neural networks also predicted crime rates with more precision than the linear models. However, this is likely due to the ensemble method used in the employed software package (Sourmail, 2004). Regardless of the complexity of the chosen model, the good predictive accuracies of the prototype models indicate that the models were effective at predicting crime (RQ 2, see Section 2.1.3) at the temporal and spatial scales used in this study (see Section 2.2). This indicates that there was sufficient overlap between the drivers of crime and the set of indicators used in this study (see Section 2.2) to validate the use of modified smart city KPIs to predict crime (RQ 1, see Section 2.1.3). As stated in Section 2.2, the indicators used in the study (listed in Table 1) are not exact implementations of smart city KPIs, but rather they have been adapted according to closely related crime predictors and available data. Furthermore, 4 out of the 22 indicators used in this study were not associated with a related smart city indicator. In these instances, new indictors were created due to the availability of relevant data for which no smart city indicator exists.

3.2. Exploratory factor analysis

highly correlated groups of variables (factors) latent in the input data:

namely A1, A2, A3 and A4 (Table 3). The extracted factors accounted for 59 % of the total variance (Table 3). The extracted factor loadings are listed in Table 3. 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 Fig. 4. Here, variables with factor loadings of 0.5 or greater (Table 3) were identified as characteristic of a given state (the description of variables are given in Table 1). The four states are further 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 face by many Black communities in NYC. Socioeconomic and demographic indicators such as poverty and race are consequently often used as risk-factors for crime (Sampson, 2006; Taylor, Ratcliffe, & Perenzin, 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.

To visualise attractor basins across NYC, factor scores for each attractor state were calculated for each PUMA in 2017 (see Section 2.2 for a summary of the data preparation and variables used in this study). Factor scores were calculated by averaging over strongly loading input features as delineated in Table 3 (Hair et al., 2014). An average of the following standardised input features was used to calculate 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



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

Exploratory Factor Analysis (EFA) (Hair et al., 2014) revealed four



Fig. 3. Visualisation of the accuracy and precision of neural network predictions compared to those made by linear regression. Source: Author's own construction.

Table 3Extracted factor loadings (n = 660).

	Factors			
% of total variance explained	23.9	14.7	11.7	8.5
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
Р5	-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 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

the state with the highest factor score. The dominate state per PUMA is shown in Fig. 5. Only factor scores greater than 0.5 were used, as these were considered to be sufficiently deviated from the mean. Visual

inspections show good agreement between the crimes associated with the various attractors, and the spatial trends observed in Figs. 6 and 7.

3.3. Sensitivity analysis

3.3.1. Analysis constraints

Concerning sensitivity analysis, there were two anticipated repercussions from the existence of attractor states observed in Section 3.2:

- A The clustering of state variables observed in Section 3.2 is analogous to the concept of basins of attraction in complex dynamic systems. A basin of attraction is a region in state space in which the system tends to remain (Gundry et al., 2011; Sendzimir et al., 2007; Walker et al., 2004; Westley, Patton, & Zimmerman, 2006, 2011; Westley et al., 2015). This equilibrium state can be described as the current system regime, governed by dominant rule-sets emerging from the underlying social and organisational networks and prevailing infrastructures (Westley et al., 2011). Because different state regimes are governed by different "rules" 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 (see Section 2.3), as less data will be available to train the network (Franceschini et al., 2019).

In order to address these anticipated challenges, sensitivity analyses were carried out for the two models developed in Section 3.1, separately for each of the four attractor states identified in Section 3.2. 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 3.2.

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



Fig. 4. Illustration of NYC attractor states. Source: Author's own construction.

citywide averages and ranges for each input feature. Thereby, testing the feasibility of disregarding attractor states.

3.3.2. Demonstration

To demonstrate the application of sensitivity analysis in decisionmaking, three scenarios were 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 are listed in Table 4 and plotted in Fig. 8. The reported values indicate the predicted changes in crime rate per unit change in input feature. Only immediately actionable variables are included in Table 4 and Fig. 8. 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.

Based on an analysis of Table 4 (Westraadt, 2019) and Fig. 8, the agencies deemed to have the most impact on crime in addition to the NYPD, include:

- The Human Resources Administration (HRA) and the Department of Health and Mental Hygiene (DOHMH). 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) could 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.

By quantitatively identifying the most influential agencies in the



Fig. 5. Dominant attractor state per PUMA in 2017. Source: Author's own construction.



Fig. 6. Street robberies per 100k population per PUMA in 2017. Source: Author's own construction.

fight against crime, and their key KPIs impacting crime, the modelling framework provides an effective means of identifying potential areas of synergy between the NYPD and other government sectors (**RQ 3**, see Section 2.1.3). Furthermore, by implementing the modelling framework, the concept of city "states" and their impact on the prediction of crime was identified. The management implications of this concept are further illustrated below. As such, the modelling framework identified

discrepancies in common assumptions about the interdependencies between city sectors, by challenging the notion of a one-size-fits-all approach to crime management (\mathbf{RQ} 4, see Section 2.1.3).

Two key properties of city "states" were demonstrated in the analysis. Table 4 (Fig. 8) illustrates that the response of crime rates to changes in input features is not only dependent on the type of crime, but also on the city "state" under consideration. For example, the effect of



Fig. 7. Street larceny per 100k population per PUMA in 2017. Source: Author's own construction.



Fig. 8. Sensitivity of crime rates to input features for areas in the A1 and A2 attractor states. Variable descriptions are listed in Table 1. Source: Author's own construction.

education had opposite effects on street robbery in the A1 and A2 states. In the A2 state, the number of street robberies increased as the percentage of the population without a high school diploma increased. On the contrary, the number of street robberies decreased as the percentage of the population without a high school diploma increased in the A1 state. The latter is possibly due to an associated decrease in wealthy targets in these high crime areas.

Table 5 illustrates the second observed property of city "states". That is, different states are more sensitive to changes in input features than others. The A1 and A3 states were less influenced by state variables than the A2 and A4 states. This could indicate that these states are more resistance to change or that the key drivers of change in these states have not yet been identified.

To investigate the effect of disregarding attractor states, the sensitivity of crime to input features was investigated in more detail. This is illustrated in Figs. 9-12 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 (see Section 2.3) 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 (Fig. 9) and street lights (Fig. 10). The relatively large error bars for the A2 and A4 states in Fig. 10, indicate that limited data was available for such high rates of street light outages. See Westraadt (2019) for further discussion on the relationship between crime and reported street light outages.

However, citywide trends are not always sufficient to correctly interpret sensitivity analysis results. For street larceny (Fig. 11), 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. This is likely



Fig. 9. Street robbery as a function of the percentage of civilians without health insurance. Source: Author's own construction.



Fig. 10. Street robbery as a function of the number of street lights out reported. Source: Author's own construction.

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 (Fig. 12), 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.

4. Conclusions and recommendations

Despite the deluge of data generated by ICMPs (Section 1), limited research has been done on exploring the use of this data to develop quantitative tools for integrated smart city planning and management. The goal of this paper was to described the development and evaluation



Fig. 11. Street larceny as a function of the unemployment rate. Source: Author's own construction.



Fig. 12. Street robbery as a function of abuse. Source: Author's own construction.

of a modelling framework that fosters cross-sector collaborations by quantifying dependencies between city sectors and identifying common cross-sector goals (Section 2). To limit the scope of the investigation, the study focused on only one aspect of smart cities, namely crime management.

The main premise of this study was that synergistic cross-sector collaborations could be fostered by using data from ICMPs and emerging smart city KPI frameworks to identify common goals between city sectors. It was proposed that, taking smart city KPIs as input, a combined approach employing Bayesian Neural Networks and sensitivity analysis could be used as a tool for identifying these goals (Section 2). To test the feasibility of this supposition, two prototype models were developed and evaluated for integrated crime management in smart cities (Section 2).

The main research questions of this study (see Section 2.1.3) were answered in Section 3. The results of the study showed that the prototype models were effective at predicting crime (**RQ 2**), indicating that there was sufficient overlap between the drivers of crime and the set of indicators used in this study to validate the use of modified smart city

Table 4

Sensitivity of crime rates to input features for areas in the A1 and A2 attractor states. Agency acronym and indictor ID descriptions are listed in Table 2 and Table 1, respectively.

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	0.83	0.58	0.16
		(within)			
		ineqT2r	1.62	0.05	0.29
		(between)			
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;	insurance	0.46	0.17	0.47
	DOHMH				
16	DOHMH;	drugs	0.15	0.23	-0.03
	NYPD				
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;	graffiti	-0.05	-0.04	-0.11
	DSNY; EDC				
38	DOHMH;	PM	0.21	0.08	-0.01
	DEP				

Table 5

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

KPIs to predict crime (**RQ 1**).

As stated in Section 2.2, the indicators used in the study (listed in Table 1) are not exact implementations of smart city KPIs, but rather they have been adapted according to closely related crime predictors and available data. Furthermore, 4 out of the 22 indicators used in this study were not associated with a related smart city indicator. In these instances, new indictors were created due to the availability of relevant data for which no smart city 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). Examination of Table 4 shows that, barring Indicator 37, these indicators played a significant role in predicting crime. It is consequently recommended that the proposed modelling framework be expanded to prescribe the identification of a secondary set of sectoral KPIs that, while traditionally not forming part of a sector's core KPIs, are nevertheless tracked due to their potential impact on neighbouring sectors (RQ 5). This recommendation is similar to the notion of Thresholds of Potential Concern applied in the field of natural resource management (Biggs et al., 2015).

The practical implementation of the modelling framework was demonstrated in Section 3.3. By objectively identifying the most influential agencies in the fight against crime, and their key KPIs impacting crime, the modelling framework provides an effective means of identifying the common goals necessary to support cross-sector collaboration (**RQ 3**). Furthermore, by implementing the modelling framework, the concept of city "states" and their impact on the prediction and management of crime was identified. As such, the modelling framework identified discrepancies in common assumptions about the interdependencies between city sectors, by challenging the notion of a one-size-fits-all approach to crime management (**RQ 4**).

To further test the conclusions made in this study, the concepts need to be tested within other domains (and combinations therefore) such as water or electricity, for example. Furthermore, due to the availability of data, a model for strategic integrated city planning was developed in this study. Smart city activities are dominated by the real-time management of city systems. The application of the proposed modelling framework at the short-term management level therefore should be explored.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Allen, P. M. (1997). Cities and regions as evolutionary complex systems. Geographical Systems, 4, 103–130.
- Bandura, A. (2000). Exercise of human agency through collective efficacy. Current Directions in Psychological Science, 9(3), 75–78.
- Batty, M., Marshall, S., et al. (2012). In J. Portugali (Ed.), 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 (pp. 21–45). Springer.
- 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. ISLJ International, 39 (10), 966–979.
- Biggs, R., 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.
- Browning, C. R. (2002). The span of collective efficacy: Extending social disorganization theory to partner violence. *Journal of Marriage and Family*, 64, 833–850.
- Chourabi, H., et al. (2012). Understanding smart cities: An integrative framework. 45th Hawaii International Conference on System Sciences. https://doi.org/10.1109/ hicss.2012.615.
- City Protocol Society. (2016). Open sensor platform. CPWD-PR_005_Open_Sensor_ Platform.
- City Protocol Society. (2015a). City anatomy: A framework to support city governance, evaluation and transformation. City protocol agreement CPA-I_001-v2.
- City Protocol Society. (2015b). Livable districts and cities. CPC_004_Livable_Districts_and_ Cities.

City Protocol Society. (2015c). City anatomy indicators. CPA-PR_002_Anatomy indicators. Cohen, B. (2014). Methodology for 2014 smart cities benchmarking [cited 2016 29 July];

- Available from: http://www.fastcoexist.com/3038818/the-smartest-cities-in-theworld-2015-methodology.
- de Blasio, B., Fuleihan, D., & Newman, E. W. (2018). Mayor's management report, NYC MOO.
- DEAT. (2004). Overview of integrated environmental management, integrated environmental management, information series 0. Pretoria: Department of Environmental Affairs and Tourism.
- Fernández-Güell, J.-M., et al. (2016). How to incorporate urban complexity, diversity and intelligence into smart cities initiatives. In E. Alba, F. Chicano, & G. Luque (Eds.), Smart cities. Smart-CT 2016. Lecture notes in computer science (Vol. 9704, pp. 85–94). Cham: Springer.
- Franceschini, S., et al. (2019). An ecologically constrained procedure for sensitivity analysis of artificial neural networks and other empirical models. *PloS One*, 14(1), e0211445.
- Fujita, H., et al. (2020). Hypotheses analysis and assessment in counterterrorism activities: A method based on OWA and fuzzy probabilistic rough sets. *IEEE Transactions on Fuzzy Systems*, 28(5), 831–845.
- Greco, S., Ehrgott, M., & Figueira, J. (2016). Multiple criteria decision analysis: State of the art surveys.
- Gregory, R., et al. (2012). Structured decision making: A practical guide to environmental management choices. Wiley-Blackwell.
- Gundry, L. K., et al. (2011). Entrepreneurial Bricolage and innovation ecology:
- Precursors to social innovation? Frontiers of Entrepreneurship Research, 31(19), 3.
 Hair, J. F., et al. (2014). Multivariate data analysis (7th edition). Essex, England: Pearson Education Limited.
- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (Third edition). Waltham. MA, USA: Morean Kaufmann Publishers.
- Huawei. (2018). Intelligent operation center solution, 27 June 2018]; Available from: htt ps://e.huawei.com/za/solutions/industries/smart-city/ioc.
- 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.
- IEC. (2015). White paper: Orchestrating infrastructure for sustainable smart cities. Geneva, Switzerland: International Electrotechnical Commission.
- ISO. (2016). ISO 37101:2016, sustainable development in communities Management system for sustainable development - Requirements with guidance for use.

L. Westraadt and A. Calitz

Sustainable Cities and Society 63 (2020) 102444

ISO. (2014). ISO 37120:2014, sustainable development of communities - Indicators for city services and quality of life.

ISO/IEC. (2017). ISO/IEC 30182:2017, smart city concept model - Guidance for establishing a model for data interoperability.

- ITU. (2016). Shaping smarter and more sustainable cities: Striving for sustainable development goals. Geneva, Switzerland.
- ITU. (2018). KPIs on smart sustainable cities, 25 June 2018]; Available from: https: //www.itu.int/en/ITU-T/ssc/Pages/KPIs-on-SSC.aspx.
- Johannesson, P., & Perjons, E. (2012). A design science primer (book no longer available -Now an introduction to design science). Lexington: CreateSpace.
- Kourtit, K., Nijkamp, P., & Steenbruggen, J. (2017). The significance of digital data systems for smart city policy. Socio-Economic Planning Sciences, 58, 13-21.
- Lombardi, P., et al. (2012). Modelling the smart city performance. Innovation: The European Journal of Social Science Research, 25(2), 137-149.
- MacKay, D. J. C. (1992). A practical Bayesian framework for backpropagation networks. Neural Computation, 4, 448-472.
- Manyika, J., et al. (2013). Open data: Unlocking innovation and performance with liquid information. McKinsey & Company.
- 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., et al. (2017). A quantitative evaluation of the mutual influences among smart strategies applied at district level. IEEE.
- 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. MOO. (2018). Social indicators report: Update to the 2016 report.
- NMBM. (2018). Integrated development plan (IDP) (Second edition). 2017/18-2021/22. Open Data Charter. (2015). International Open data charter.
- Peffers, K., et al. (2008). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45-77.
- Pelorosso, R. (2020). Modeling and urban planning: A systematic review of performancebased approaches. Sustainable Cities and Society, 52.
- Perry, W. L., et al. (2013). Predicitve policing: The role of crime forecasting in law enforcement operations. RAND Corporation.
- Sampson, R. J. (2006). Collective efficacy theory: Lessons learned and directions for future inquiry. In F. T. Cullen, J. P. Wright, & K. R. Blevins (Eds.), Taking stock: The Status of criminological theory (p. 149). New Brunswick, N.J: Transaction Publishers.
- Schleicher, J. M., et al. (2016). A holistic, interdisciplinary decision support system for sustainable smart city design. In E. Alba, F. Chicano, & G. Luque (Eds.), Smart cities. Smart-CT 2016. Lecture notes in computer science (Vol. 9704, pp. 1-10). Cham: Springer.
- Sendzimir, J., et al. (2007). Assessing the resilience of a river management regime: Informal learning in a shadow network in the Tisza River Basin. Ecology and Society, 13(1), 11.

- Shahriari, B., et al. (2016). Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE, 104(1), 148-175.
- Sourmail, T. (2002). Simultaneous precipitation reactions in creep-resistant austenitic stainless steels. Department of materials science and metallurgy. University of Cambridge.

Sourmail, T. (2004). Softwares for Bayesian neural networks training [cited 2018 16

- December]; Available from: http://thomas-sourmail.net/neural_networks.html. 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(5), 635–657.
- The British Standards Institution. (2014). PAS 181:2014, smart city framework Guide to establishing strategies for smart cities and communities.
- U4SSC. (2017). Collection methodology for key performance indicators for smart sustainable cities. Geneva, Switzerland.
- United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development (A/RES/70/1).
- Walker, B., et al. (2004). Resilience, adaptability and transformability in social-ecological systems. Ecology and Society, 9(2), 5.
- Westley, F., Patton, M. Q., & Zimmerman, B. (2006). Getting to maybe: How the world is changed. Toronto: Random House Canada.
- Westley, F., et al. (2015). Social innovation lab guide. The Rockefeller Foundation.
- Westley, F., et al. (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.
- Westraadt, L. (2019). A model for crime management in smart cities, in computing sciences. Port Elizabeth, South Africa: Nelson Mandela University.
- Westraadt, L., & Calitz, A. (2018). A gap analysis of new smart city solutions for integrated city planning and management. SAICSIT 2018 - Annual Conference of the South African Institute of Computer Scientists and Information Technologists.
- Westraadt, L., Calitz, A. P., & Cullen, M. (2019). Guidelines for managerial decision making in smart cities: A South African perspective. International Business Conference (IBC 2019).
- World Bank. (2014). Introduction to poverty analysis (English). Washington, DC: World Bank Group.
- Zhuhadar, L., et al. (2017). The next wave of innovation-Review of smart cities intelligent operation systems. Computers in Human Behavior, 66, 273-281.