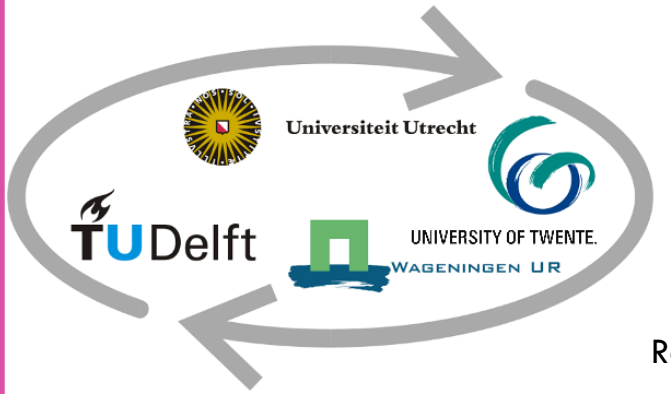
GIMA

Geographical Information Management and Applications

Exploring a Space-Time Autoregressive Moving Average (STARMA) model in spatial crime forecasting

Alkiviadis Rentzelos

[6510620]



MSc Thesis

November 2020

Supervisor: Dr. Ourania Kounadi

Resp. Professor: Prof. Menno-Jan Kraak

Acknowledgments

A very heartfelt thank you goes to Dr. Ourania Kounadi, my supervisor, for her tremendous support and positivity during my thesis research. I really appreciate her incredible generosity and confidence in me.

I would also like to thank Mr. Felix Cheysson, the author of the proposed model for his willingness to respond to several questions.

I gratefully acknowledge Xiaojian Liu from the MSc Geoinformatics of ITC University for his help about R-studio, and finally, I would like to thank my sister and Pavlos Krassakis for their encouragement.

Summary

So far, a variety of methods is used by predictive policing for forecasting crime purposes. Depending on the type of forecasting, some methods work on forecasting the number of crime incidents while others on forecasting whether an area is possible to have crime incidents or not. At the same time, crime is not considered stable in space and most of the popular methods that are used do not take into consideration the parameter of time. This is an important issue not only for police departments, but also for a variety of fields (criminologists, geographers, society and academic society). Hence, the exploration of current Space-time autoregressive models, which are based only on past space and time crime data, becomes more and more necessary.

With the present thesis the importance of a Space-time autoregressive moving average model for forecasting crimes is examined compared to simple methods that have been used till now, based only on past crime data. The study focuses on New York City and three crime types are examined. The selection of current and baseline methods (a conventional Kernel Density Estimation and a naïve approach), their parameters and the proposed method (spatiotemporal autoregressive moving average – STARMA) is done after an extended research literature. Likewise, a research about the evaluation of forecasting performance metrics is done. The selected methods are compared under the same methodological framework, where a threshold value classifies their outputs into two classes; hotspots and non-hotspots.

The proposed method's parameters are experimented to examine how the consideration of space and time affect its performance. In the end, the experiment that excels is compared with the baseline and the conventional method through spatial forecasting accuracy metrics. According to the results, all the three methods present important outputs and each method outperforms for the different examined crime types (all, property and violent). More particularly, the naïve Baseline method outperform for type crime 'all', the conventional KDE excels for crime type 'property' while the proposed STARMA method shows higher performance for the crime type 'violent'.

The study concludes therefore that the current Space-time autoregressive models are quite sensitive to the parameters of space and time and further research needs to be conducted to examine under which spatial and temporal resolution these models could exceed the baseline methods.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background

Predictive policing is defined as a set of analytical and quantitative methods that the police can apply in order to identify areas where a crime is more likely to happen or to prevent it according to statistical predictions (Perry et al., 2018).

During the last decades, predictive policing based on crime forecasts is an important research topic (Perry et al., 2018). Although the definition of crime is possible, prediction of crime events is harder, let alone highly accurate prediction (Grover, Adderley & Bramer, 2007). The fact that crime events do not happen as random but are also not consistent in space and time has led many scientists to reconsider the reliability of existing predictive methods or motivate them to explore new ones (Yu et al., 2011). Nowadays, there is a variety of methods as far as predictive policing is concerned. Some of them are based on approaches to forecast the place and time with high crime occurrence rates, like the object of this research thesis, while, other methods are used to predict offenders, victims or perpetrators' identities (Perry et al., 2018).

A crime forecast that provides information about space and time can be characterized as a valuable source for many applications (Shamsuddin, Ali & Alwee, 2017). First of all, police departments would benefit from accurate short-term crime forecasts and, as a result, their policy would be more effective. For instance, classic control tactics like patrols or crime alerts could take place in particular neighborhoods (Gorr & Harries, 2003). Secondly, geographers and criminologists are able to investigate and explain the geography of crime events and the reasons that trigger them (Li et al., 2014). Furthermore, the society can benefit from accurate crime forecasts; for example, a cooperation between the local authorities and the police in a constructive dialog regarding resources, workloads etc., could decrease crime rate and provide safety to the citizens (Gorr & Harries, 2003).

Till now, many crime forecast methods have been developed. These "conventional" methods can be divided in three main categories depending on how they work and their purpose. The first category concerns methods based on past crime data and try to forecast the location of a potential crime over a specific time frame (e.g. Hotspot Analysis, statistical regressions, Near-repeat Modelling methods, etc.) (Perry et al., 2018). The second category refers to methods that use geography factors associated with

risk for specific types of crime (e.g. Risk Terrain Analysis) (Kennedy, Caplan & Piza, 2011). The last category involves (space) temporal methods, which focus either on potential time or the victims of a crime event, by using victim profiles, like age, race or sex, and connecting them with the place and time of a crime) (Perry et al., 2018).

The previously mentioned (conventional) methods are popular due to some advantages. For instance, Hotspot Analysis methods offer a coherent overview of the study area through producing maps; doing so, specialists can easily read them and make quick decisions (Bachner, 2013). Statistical Regression methods have been used for many years addressing many types of crime and give accurate predictions for short-term forecasts. These methods are based on additional attributes, such as the number of houses, number of unoccupied ones, the number of people that have been convicted for property crimes etc. Forecasts based on near-repeat methods examine crime events that have recently repeatedly occurred (Perry et al., 2018). Risk Terrain Analysis methods, on the other hand, provide results based on a combination of map layers, taking into account multiple factors (e.g. environmental, social, physical and behavioral parameters etc.) (Drawve, Moak, & Berthelot, 2016a). Finally, spatiotemporal analysis methods use various environmental and temporal features for potential crime events (Perry et al., 2018). All of the aforementioned types of crime forecast methods can be applied using different predictive techniques depending on the purpose of the study of crime events.

Contrary to the advantages, there are some drawbacks which increase the doubts about the suitability of these conventional methods. More particularly, regarding Hotspot Analysis methods, they treat spatial and temporal aspects of crime not as one unique entity but separately. Hence, the relative Hotspot techniques ignore the simultaneous interaction of space and time in crime prevalence (Grubesic & MacK, 2008). Traditional Hotspot Analysis relied on the hypothesis that crime remains static. However, in reality, crime is dynamic and this is proved through temporal patterns. This drawback is also valid for Risk Terrain Analysis that considers factors to be static, such as the urban district, which constantly develops (Drawve et al., 2016a). Beyond that, Risk Terrain Analysis creates models pinpointing crime prone areas. Nonetheless, it does not mean that all such areas are dangerous. Therefore, this aspect decreases the level of accuracy of this method, when compared to the Hotspot approach (Drawve et al., 2016a). As far as the Near-repeat methods are concerned, they are based on assumptions that future crimes will occur much close to previous crime events in terms

of time and place. As a consequence, this assumption calls into question their reliability in predictive policing. Finally, regarding the Statistical Regression methods, their prediction also depends on the available data set (e.g. how old the examined data is) while they respond to questions by providing a number (e.g. how many burglaries are expected next week) and not a spot (e.g. neighborhood) (Perry et al., 2018).

As a consequence during the latest years, there is an increasing interest by police departments and scholars in developing techniques using spatial and temporal analysis tools (Shamsuddin et al., 2017) combined with many attempts to develop predictive models of crime events. Although some of these efforts are still in the early stages, there is enough knowledge for exploration them and comparison with the current conventional methods (Groff & La Vigne, 2002).

These efforts also include the Space-Time Autoregressive (ST-AR) models, which can be used for different types of crime and spatial scales according to Shoesmith (2013). Broadly speaking, a space-time model is a time-series model that explicitly takes into account linear dependence between the variables that is lagged both in time and in space (Giacomini & Granger, 2004). Regarding the term "auto-regression", it indicates the linear regression of variables with themselves using past values (Cesario, Catlett & Talia, 2016).

1.2 Problem Statement

Although the relation amongst crime events, human activity and environment is well-known and has been justified by theoretical background, quantifying it still remains a challenging topic. It is quite difficult to answer why a crime event occurred in a specific place in specific time, let alone develop a method that could predict potential crime events with high accuracy (Shamsuddin et al., 2017; Grubesic & Mack, 2008).

There are several reasons that prove the necessity of highly accurate methods for forecasting crime events today. On the one hand, increasing urbanization increases the amount of crime events and makes effective crime prevention more difficult to achieve (Catlett et al., 2019). On the other hand, the above mentioned drawbacks of the simple current methods (e.g. subjective decisions, assumptions, low spatial accuracy, high dependence on the data sets etc.) outline the need for new methods that consider only space and time (crime) data and this, in an attempt to not being affected by any additional parameters (Drawve et al., 2016a; Perry et al., 2018).

This thesis is motivated by these arguments and inspired by Shoesmith (2013) who investigated and compared space – time autoregressive (ST-AR) model with other autoregressive methods (e.g. the Vector auto-regressions and aggregate univariate models) in order to show which method provides higher forecasting accuracy for various types of crime events and in different spatial scales.

In this thesis topic, spatiotemporal autoregressive models is compared with conventional methods for forecasting crime events. Additionally, verifying the performance of any proposed method requires a comparable baseline (Lin, Yen & Yu, 2018). Implemented methods are usually compared with a dummy method or a simplicity (technically) approach in order to prove whether they outperform other forecasting methods. Different data sets and different techniques lead to different alternative methods (Lin et al., 2018). For example, the study of Lin et al. (2018) uses the time series analysis as the baseline method compared to the Deep Neural Networks proposal (DNN). Moreover, the use of a Moving Average (MA) method can improve the results of the above mentioned alternative advanced method. Another study refers to a Machine Learning (ML) algorithm (e.g. Gaussian Naive Bayes, K-Nearest Neighbors and Logistic Regression) as a comparable method (Zhuang, Almeida, Morabito, & Ding, 2017). In contemporary methodologies, data analysis and machine learning models, and more particularly the supervised ones, which can also be used as advanced baseline methods, are quite beneficial for prediction through automated procedures, using only historical (space and time) data. A striking example of this is the Neural Networks, which is quite effective for complex classification topics. However, machine learning methods are often complex and require advanced statistical expertise (Louppe, 2014; Rummens, Hardyns & Pauwels, 2017). Even though Rumments, Hardyns & Pauwels (2017) claim that predictive analyses and big data are recent in the field of criminology, there are studies that are based on machine learning for prediction of crimes (e.g. based on demographics and mobile data) (Bogomolov et al., 2014; Belesiotis et al., 2018).

For this thesis, the term "baseline" refers to either naïve or advanced forecasting methods. Conventional forecasting methods, although are in general baseline methods, in this thesis are mentioned separately.

1.3 Research Scope

The main goal of this thesis is to investigate the potential of a Spatiotemporal Autoregressive (ST-AR) model to forecast crime events and to compare it with baseline and comparable conventional methods. This goal can be divided into two sub-objectives (SO). The first one (SO.1) concerns the implementation of a ST-AR model method, its effectiveness and the impact of input spatiotemporal data on the predictions. The second objective (SO.2) regards the comparison between this model and other methods (baseline and conventional) referring to their accuracy, while the last, RQ (2.2), focuses on the predictive results comparing different crime types. The following sub-section states the SOs and the Research Questions (RQs) related to them.

1.3.1 Scientific Objectives and Research Questions

- **SO.1: Forecast crime events using a spatiotemporal autoregressive model.**

RQ 1.1. How effective is an ST-AR method in predicting crimes in space and time?

RQ 1.2 Are the predictive results sensitive to the parameterization of space and time?

RQ 1.3. How do predictive results vary among different types of crime?

- **SO.2: Comparison of a ST-AR method with baseline and conventional methods.**

RQ 2.1. What methods perform better in terms of the robustness and accuracy of crime forecasts?

RQ 2.2. Is the ranking of performance of the examined methods consistent among different types of crime?

1.3.2 Research Boundaries

There are some limitations and assumptions regarding the study area, the spatial and temporal resolution and the crime types. First of all, the choice of study area is restricted by the availability of data sets. Current options are publicly available repositories for metropolitan cities, such as London[1] or New York[2]. These areas have high crime rates and the spatial extent of a city. In many studies these features are taken into consideration when selecting a study area and thus these study area choices seem to be

[1] Open data about crime and policing in England, Wales and Northern Ireland: https://data.police.uk/
[2] Open data about crime, published by New York City (NYC) agencies and other partners: https://opendata.cityofnewyork.us/

appropriate. The description of the chosen study area is further developed in chapter 3. Secondly, the choice of the spatial scale (e.g. municipality, block, post-code, etc.), and temporal resolution (short-term or long-term forecasting) depend on the properties of the data sets. Lastly, due to time constraints, crime types are grouped and examined into three main categories; namely 'all' crime, 'property' crime (e.g. burglary, larceny, motor vehicle theft), and 'violent' crime (e.g. murder, rape, robbery, assault) following the division of Shoesmith (2013).

1.4 Conceptual Framework

This section provides a conceptual schema (fig. 1) which represents the relationships between the conceptual entities and their features, on which this thesis topic focuses.

Broadly speaking, there are several **problems** that need an effective predictive policing, for example the increase of crime events that tend to follow the growing trend of urbanization. These problems concern a variety of **stakeholders** (criminologists, academic society, etc.) who are working on **crime data**, which, according to Shoesmith (2013) can be divided into different categories (property, violent, etc.). The **conventional methods** (e.g. Hotspot Analysis, Risk Terrain Analysis, etc.) that are used for forecasting or prediction in the field of criminology are associated with some of the problems (i.e. their drawbacks in forecasting). Each conventional method is related to different **forecasting information** (e.g. amount of crimes, crime victims, etc.), depending on the reason they are used. This leads to explore and to propose new **"non-conventional" methods** to forecast crimes, such as the Space-Temporal Autoregressive models, which are **based** only **on** past (historical) crime data. Apart from these methods, **baseline methods**, simple (naïve) or advanced (e.g. the machine learning methods as they described in section 1.2) are used to compare models' effectiveness. In the concept of forecasting methods, three categories are mentioned; conventional methods, baseline and ST-AR models. The ST-AR models can be implemented through various methods (e.g. ST-ARIMA, ST-ARMA etc.). This entity is the one that will be explored in detail in order to be compared with the current forecast methods.

All in all, this reasoning is illustrated in the following figure, where each of the above mentioned concept entities is represented in different box with its details. These concepts are associated to others with lines. The blue line in the figure shows a general connection between concepts and some of their features, the green one shows the concepts that this thesis examines.

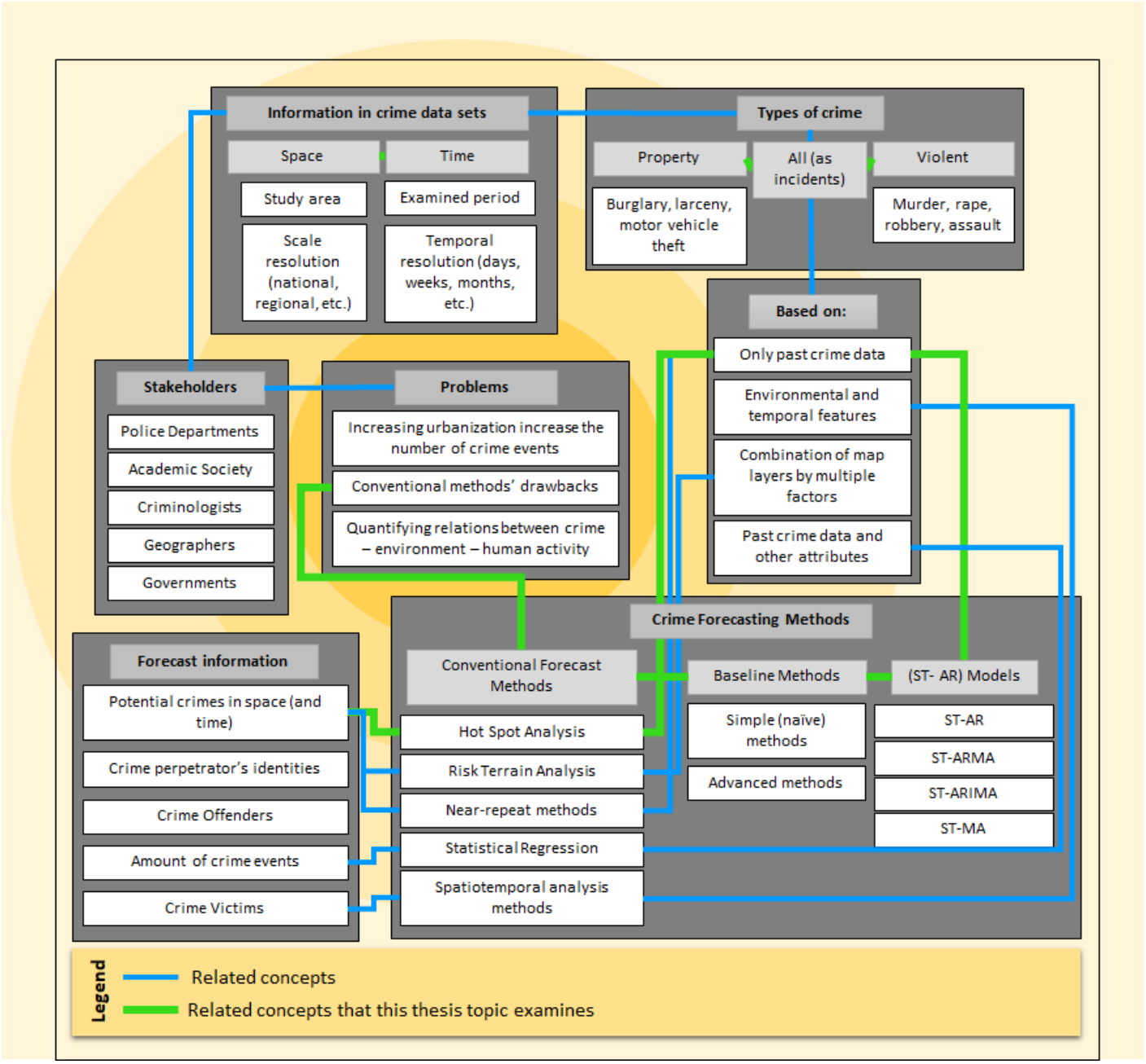


Figure 1: Conceptual Framework – Problems and factors that lead in exploring new crime forecasting methods

2 Theoretical framework

Having set the inspiration, the motivation and the research objectives of this thesis, this chapter makes an extensive analysis of the theoretical framework. Apart from the theoretical background involving predictive policing, the methods that are used for crime event forecasting are described in the next chapters. In the first section, some of the most popular conventional crime forecast methods (Hotspot analysis, near-repeat methods, risk terrain analysis etc.) are mentioned. Among them, the forecast methods that focus on where potential crime can occur are the Hotspot analysis, the Regression methods, the Near-repeat and the Risk Terrain Analysis, while the methods that are only based on past (crime) data are the Hotspot Analysis and the Near-repeat methods (Perry et al., 2018). Additionally, there are methods based on past data, such as spatiotemporal autoregressive (ST-AR) methods, which cover a great range of forecast cases that have also been used also for crime events.

2.1 Crime and Predictive Policing

Criminality is one of the most significant issues in contemporary societies and the accurate prediction of crimes is important but still challenging (Bharati & Sarvanaguru, 2018). Crimes affect the development of a society in different ways: the quality of life, the economic development, population movement, etc. A good illustration of this is that when people need to relocate or visit an area, crime level is something they take into account and is based on actual crime events in these locations or even on rumors (Almanie, Mirza & Lor, 2015; Gonzales, Schofield & Hart, 2005). Due to the effects on the society, many researchers have characterized crime as a social phenomenon and its prevention has been object of study for many years; numerous crime prediction methods are continuously developing. An effective crime prevention policy is imperative due to the rise of urbanization that is linked with increased crime rates. In other words, more citizens means more crime events and it becomes more difficult to predict a potential crime (Catlett et al., 2019).

Crimes are not geographically consistent; they may occur in different places, from small villages to big cities (Bharati & Sarvanaguru, 2018). Moreover, it has been observed that criminality tends to concentrate on specific areas (specific streets or neighborhoods) for different reasons related to the offenders, the victims and certain

circumstances (Chainey, Tompson & Uhlig, 2008). Thus, different types of crime present different spatial relationships and crime allocation (Gonzales et al., 2005).

Additionally, time is a factor that affects criminality; some crimes occur at a specific time of day and, as such, some places are dangerous in only specific hours, for example a night club in a city center. Therefore, on top of spatial relationships, temporal relationships should also be considered. However, this makes the process of predicting crimes more complicated (Gonzales et al., 2005).

Crime patterns vary in space and time and are thus unevenly distributed across different scales. However, since the 1970s, crime mapping appeared as a reliable interpretation through GIS (Geographic Information System) (Khalid et al., 2018). Today, there is a need for technology that could deal better with crime events and their prevention (Bharati & Sarvanaguru, 2018). Hence, predictive policing is developed towards this end.

Predictive policing is an emerging law enforcement approach by using data and making statistical analysis to contribute to crime prevention. This can be done by pinpointing potential crime areas to police forces (Kutnowski, 2017), who are nowadays much more advanced and skilled on programming, information technology and using datasets for further processing compared to the past (Lin et al., 2018). As far as the traditional predictive policing is concerned, this knowledge can provide the police with information through predictive mapping. Doing so, the police can allocate high levels of patrol in specific places or neighborhoods and maybe in a specific time zone. This feature could be very effective for preventing crime since with predictive mapping they can focus on narrowly defined areas (Groff & La Vigne, 2002).

There are various approaches to forecast crime events over a study area and a specific time horizon. However, not all these approaches can be considered together in a certain predictive method. Some methods use a variety of factors and parameters (e.g. behavior of offenders or victims) that lead to a crime event while others are strictly based on historical crime data (e.g. location, time and type of crime) (Groff & La Vigne, 2002). Likewise, some models, like spatiotemporal, rely on past events to predict future ones.

Broadly speaking, predictive modeling is how a model is built in order to make predictions (Bharati & Sarvanaguru, 2018). Predictive modeling is divided into two categories: Regression models and pattern classification. The former analyzes the

relationships between variables and trends aiming to make predictions; the latter aims to assign particular class labels to specific data value according to the prediction results. For example, two labels about weather prediction by a classification model are sunny or rainy day (Bharati & Sarvanaguru, 2018).

2.2 Related Work

Several studies on forecast crime events have been conducted throughout the years. Some of them focus on forecast hotspots by following a binary classification while others follow a regression method to estimate the number of crimes. Most of these studies need a large spatial scale and as such they use either natural geographic boundaries (e.g. census blocks) or technical boundaries, like grids (mostly squares), to divide the study area in many smaller regions / zones. Some studies use only one type of crime whereas others use two or more, depending on their objectives or the data available. USA and London are areas widely used in many relative studies. There are relevant studies that use historic data dating back years to make a forecast about crime events, while those forecasts were either for the following days, weeks, months or years.

Regarding the baseline methods that these studies used in order to compare some methods, some of them used advanced ones (e.g. Zhuang et al. (2017)) or those mentioned in 1.2 section. However, some other use simple naïve models. For example, Shoesmith (2013) compared the STAR method with two baseline naïve model methods; one which is related to the most recent rate of case study events and another one related to the most recent changes in rates for case study events. The following table (table 1) illustrates some of the related work that has been done regarding forecast crime and it involves a variety of features.

Table 1: Overview of forecasting crime studies

	Source	Study Area	Spatial Scale	Sample Period	Crime Types	Forecast	Type of forecast method	Spatial Unit	Temporal Unit of forecasting
1	(Bowen et al., 2018)	DeKalb (USA)	County	2011-2014	1 type (violent)	Hotspots	Classification (binary)	Census blocks	Month
2	(Chainey et al., 2008)	Camden and Islington (London)	Part of capital city	2002-2003	Residential burglary, Street crime, Theft from and of vehicle	Hotspots	Classification	Ellipses and Grids (250 m, 500m)	Many periods (next 1, 2 & 3 days, next 1 & 2 weeks, next 1, 2, 3, 6 & 12 months)
3	(Dash, Safro & Srinivasamurthy, 2019)	Chicago (USA)	City	2011-2015	34 types	Number of crimes	Regression	Communities	Month, Year
4	(Drawve, Moak & Berthelot, 2016)	Little Rock (USA)	City	2008-2009	1 Type (gun crime)	Hotspots	Classification (binary)	Square grid cells (91 m)	6 Months
5	(Gorr & Harries, 2003)	Pittsburg (USA)	City	1991-1998	5 Types	Number of crimes	Regression	Police precincts	Month
6	(Mohler, 2014)	Chicago (USA)	City	2007-2012	2 Types	Hotspots	Classification (binary)	Square grid cells (75 m, 150 m)	Day
7	(Shoesmith, 2013)	USA	Country	1960-2009	2 types (violent, property)	Crime rates	Regression	USA regions	Year
8	(Zhuang et al., 2017)	Portland (USA)	City	2012-2016	All crime	Hotspots	Classification (binary)	Square grid cells (600 ft)	2 weeks

2.3 Hotspot Analysis

In this section, several techniques of hotspot analysis that are conventional forecast approaches are described (i.e. spatial ellipses, heuristic techniques, kernel density estimation). All of these techniques aim to identify areas of high crime concentration (hotspot) and differentiate them from areas of low or non-significant crime concentration. Thus, hotspot analysis performs pattern classification and yields a classification outcome (i.e. high-volume crime versus low volume crime).

The concentration of crime events in identifiable places was first mentioned time in 1982 by Brantingham & Brantingham (2005). This finding is supported by many researchers till today (Chainey et al., 2008). The identification of crime hot spots redefined the importance of spatial features of crime (Anselin et al., 2000). Hotspots are defined as the geographic locations (points, addresses, zip codes, census blocks etc.) within an identifiable boundary (e.g. square grid cells) with a high crime rate compared to the whole region under examination. Therefore, hotspots were the first examples regarding the importance of the area during crime investigation. The first attempts to quantify criminality as a spatial phenomenon started by Sherman, Gartin & Buerger (1989) and it was also supported by Brantingham & Brantingham (1993). Meanwhile, Roncek & Maier (1991), in their study, focusing on city blocks in Cleveland, found that there is a relationship between crime and specific locations. Considering that hotspots are regular, predictable and they do not happen by chance increases the interest to examine whether hotspots are systematic or not. Therefore, their identification should be a careful process with accurate results (Anselin et al., 2000).

Hotspot analysis is based on the premise that retrospective crime patterns are significant indicators for potential crimes (Anselin et al., 2000; Chainey et al., 2008; Khalid et al., 2018; Gonzales et al., 2005). A hotspot approach uses crime data that refer to a period before a fixed date (one has already passed) to create hotspot maps. Gorr & Lee (2015) underline the importance of retrospective data with a new definition, describing crime hotspots as small places in which crime events are frequent and highly predictable, at least over a one-year period. Once these maps are designed, their accuracy is tested. To support that their accuracy is sufficient, comparison with other methods and among different crime types should be made. Crime hotspots can be recognized through visualization (e.g. color-coded maps), statistical identification (e.g.

local autocorrelation) and theoretical prediction (i.e. by applying what is already known) (Khalid et al., 2018).

Regarding the limitations of the hotspot analysis method, it should be mentioned that it is locally descriptive. In other words, not all hotspot techniques can apply to all areas and this is because each of the existing techniques is implemented in a different way (i.e. using different parameters) and for different geographic areas. Furthermore, since the crime hotspot method is only based on historical crime data, the quantity of the used datasets plays a significant role. This means that the more missing values from the data, the less accurate or reliable the result will be. A way that researchers adopt in order to minimize the effect of these limitations of hotspot analysis is projecting the potential crime points-locations into closed spatial features where they believe there is high crime concentration (Gerber, 2014).

Hotspot mapping is the most popular traditional approach that the police and crime reduction units use to forecast crime (Chainey et al., 2008). Hotspot mapping is used to identify and analyze areas with high concentration of criminality. Many techniques can be used in order to achieve this goal. However, many of the techniques have been questioned regarding their reliability. For this reason, they have been subject to several reviews in different ways, for example, either by comparing them or by performing different kind of exercises in order to evaluate their processes. These reviews showed that different hotspot mapping techniques lead to different results (Chainey et al., 2008). The most popular (conventional) techniques for identifying crime hotspots crime are spatial ellipses, grid thematic mapping and Kernel Density Estimation (KDE). The common advantage of these techniques is that they can easily be used and can interpret the results that show potential crime location (Chainey et al., 2008).

Some of the most popular hotspot techniques are described below. Here, hotspot techniques have been described in detail in specific categories having combined written material from several authors. The first is about the spatial or covering ellipses, the second one is about heuristic techniques, like pin and grid thematic maps, and the third is about the Kernel Density Estimation.

2.3.1 (Standard Deviational) Spatial / Covering Ellipses

This technique creates standard deviational ellipses around the locations (i.e. points of crime events) that have been identified as hotspots (Chainey et al., 2008). It is

characterized as an easy technique because it uses mathematical functions to create the ellipses around the clusters of occurrences (Perry et al., 2018).

Spatial ellipses can be created through nearest neighbor hierarchical (Nnh) and the k-means clustering routine. The former uses a nearest neighbor tool to identify groups of minimum number of crimes that are close to each other and it groups the defined points based on a specific criterion, usually through a threshold distance value. The latter performs similarly, by making k number of ellipses and by partitioning the crime points into groupings. Once the best positioning of k centers has been found, centers are replaced by closer points (Gonzales et al., 2005; Perry et al., 2018). Both of these two techniques are able to explore and study spatial data by implementing the spatial ellipses technique but they cannot provide enough aid to hotspot forecasts (Gonzales et al., 2005).

On the one hand, this technique defines hotspot areas without being affected by artificial boundaries because it uses the location of crimes (Perry et al., 2018). However, using ellipses in fact includes a larger area around particular hotspots and hence this is a major drawback of this technique. The provided ellipses are not in line with natural boundaries and can often mislead (Chainey et al., 2008; Gonzales et al., 2005). Producing ellipses that cover smaller areas (in order to follow the natural boundaries) would lead to more efficient results, avoiding the misleading mentioned above. However, some practices showed that initial ellipses' shapes have proved to be helpful in directing policing patrols correctly (Perry et al., 2018). Apart from that, this technique requires to predefine the number of ellipses and its results are sensitive to seasonal phenomena since all the observations are initially equally weighted (Perry et al., 2018). The following figure (fig. 2) illustrates an implementation of the standard deviational spatial ellipses mapping technique (Chainey et al., 2008).

Source: Chainey et al., 2008; p. 4 (edited by author)

Figure 2: Standard deviational spatial ellipses for crime events

2.3.2 Manual Identification of Hotspots on Pin Maps

This approach is a heuristic technique supported by mathematical functions. This kind of techniques work on the assumption that police departments know the study area in advance. Hence, they can easily discern crime indicators (Gonzales et al., 2005). The Manual Identification of Hotspots on Pin Maps is based on personal knowledge and experience involving the examined study area and crime events. Maps are made by subjective judgment and personal criteria about the potential locations of future crime events. These locations are rendered as a single point. This approach does not improve visualization. Although it is not supported by scientists it has been widely used by analysts and police departments because it provides effective results (Gonzales et al., 2005). The following figure (fig. 3) illustrates an implementation of pin mapping technique for crime events.

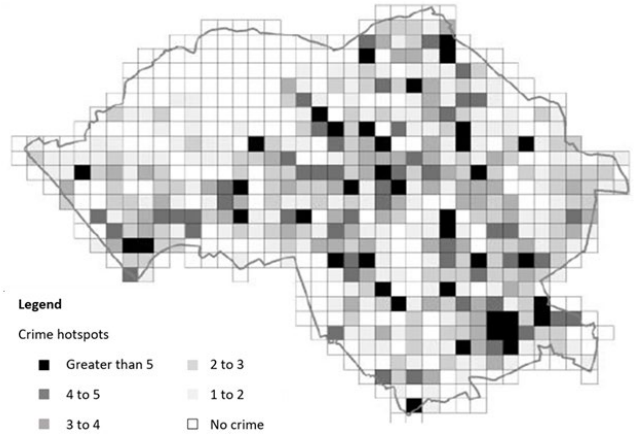
Source: Chainey et al., 2008; p. 4 (edited by author)

Figure 3: Manual Identification of Hot Spots on Pin Maps for crime events

2.3.3 Quadrat / Grid Thematic Mapping

Another heuristic technique is the Quadrat (grid) thematic mapping. This approach is used to identify vulnerable residences by implementing this heuristic technique on a GIS database application (Bowers, Johnson & Pease, 2004). It is also used to map the volume of victims' emergency calls and violent offences in North Carolina. In the study of Chainey et al. (2008) each grid corresponded to a square mile. One of the principles of this approach is the aggregation of crime in cells or grids, usually quadrats, for easy identification and comparison of hotspot areas (Chainey et al., 2008). The produced maps highlight the hotspot areas by using different colors and the choosing color ramps depending on the number of crime events. The equally divided area helps the reader to compare the volume of crimes over the study area (Gonzales et al., 2005; Chainey et al.,

2008). The approach continues by calculating the simple density of the crime points. Beyond the lack of additional parameters this approach presents one more drawback, the subjective decision of crucial locations which is based on visualization (Gonzales et al., 2005). Another drawback is the possibility to lose spatial details because crimes have to conform in each grid cell. Finally, the size of those grids can also influence the provided information in the final hotspot maps (Chainey et al., 2008).The following figure (fig. 4) illustrates an implementation of grid thematic mapping technique for crime events.



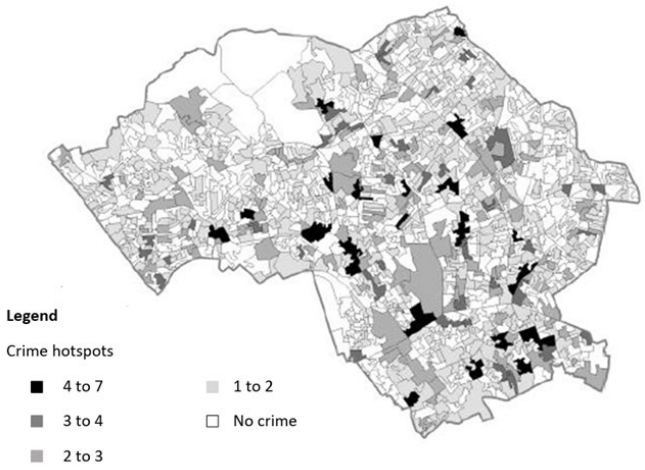
Source: Chainey et al., 2008; p. 4 (edited by author)

Figure 4: Grid Thematic Mapping for crime events

2.3.4 Jurisdiction-Bounded Areas / Thematic Mapping of Geographic Boundary Areas

A third heuristic technique is a widely used approach which portrays the examined study area by settling some boundaries zones (e.g. postal codes) as polygons that appear with different colors according to the level of criminality in each one. On the one hand, this deviation of a study area to many regions helps the police force to work in each separate region since they are familiar with high crime rate areas. More particularly, it has been mentioned that these maps can be made quickly and without high level of technical knowledge. In addition to that, users can easily identify which that concentrate more crime events and are allowed to zoom in for further information. Furthermore, a percentage of criminality can be estimated by using other parameters, such as the population. On the other hand, the main disadvantage of this approach is the separation of the study area into many zones and it has been proved that a differently examined spatial scale can affect or change the predicted results (Gonzales et

al., 2005; Chainey et al., 2008). The following figure (fig. 5) illustrates an implementation of thematic mapping of geographic boundary areas technique for crime events.



Source: Chainey et al., 2008; p. 4 (edited by author)

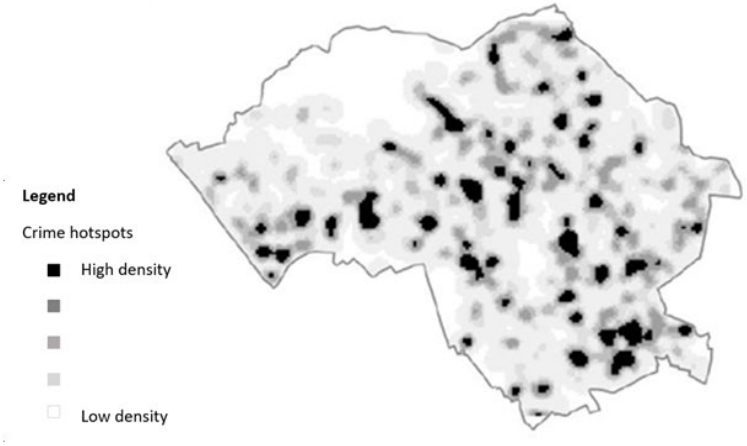
Figure 5: Thematic Mapping of Geographic Boundary Areas for crime events

2.3.5 Kernel Density Estimation (KDE)

Kernel estimation is a natural extension of the quadrat (grid thematic map) heuristic technique. KDE technique is used to examine global trends, like criminality phenomena, based on spatial point data (e.g. locations of crime incidents) (Anselin et al., 2000). KDE is a popular technique because of its appealing way to visualize spatial concentration of data (e.g. crime incidents) (Chainey, 2013). It spreads out the expected contribution of future case study phenomenon risk over the study area by using a mathematical function kernel (Perry et al., 2018). KDE is an interpolation technique that is applied to a continuous surface (Perry et al., 2018). Chainey et al. (2008) proved that KDE was more effective compared to covering ellipses and other hotspot mapping techniques. In addition, Gerber (2014) claims that KDE is the most suitable spatial analysis technique to identify hotspots in a two-dimensional surface.

The main advantage of this technique is that analysts can easily interpret its results. Another advantage of KDE technique is that it can examine a region ignoring the shape of boundaries of hotspots while other techniques are usually affected by barriers (natural, like rivers, or technical, like roads). Furthermore, this technique provides more realistic hotspot contours instead of particular shapes (e.g. ellipses or circles), like the previously mentioned do. KDE also faces the problem of covering ellipses about the initial assumption of number of ellipses for hot spots. Finally, kernel estimation could reveal significant crime patterns over time by comparing to patterns of time periods

(e.g. the same month through the following years) (Anselin et al., 2000). However, a drawback of this technique is related to the choice of thematic range on producing hotspot maps, which is related to the design of the output. Therefore, the way of identification hotspots is directly influenced by the producing maps, without necessarily meaning that there is no statistical robustness. Till now, police agencies have not disputed the validity or the statistical robustness of this technique because it provides reliable and easily read results (Chainey et al., 2008). The following figure (fig. 6) illustrates an implementation of KDE technique for crime events.



Source: Chainey et al., 2008; p. 5 (edited by author)

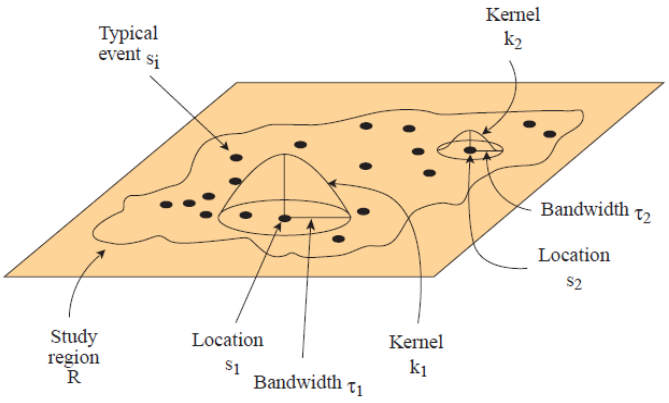
Figure 6: KDE for crime events

In general, the KDE technique can be divided into two main types; single and dual. The former estimates hot spots by using a single variable (e.g. crime incidents) while the latter uses two variables (e.g. crime incidents and population density). Both types produce heat, contour or a surface view map. Threshold value defines from which value and up areas are defined as hot spots (Perry et al., 2018).

The KDE technique is quite robust and it is more and more used today by police practitioners in the field of predictive policing. The requirement inputs for KDE are the cell size and a bandwidth value. Until now, few surveys have been conducted in order to test the ideal values involving these two parameters. Therefore, the effect that different choices of these values have on the results are only vaguely known. The different choices for values of bandwidth or for the specific kernel type (equation) may have small changes regarding results (Perry et al., 2018). The value of bandwidth is significant because it defines the intensity of smoothing in the hot spot maps. These maps look like spatially based histograms, which reflect the point pattern intensity. For

this reason, the ideal value for bandwidth should be neither high nor small but represent the actual distance among the studied points (Anselin et al., 2000). Additionally, it has been observed that lower values create more compact hot spots (Perry et al., 2018).

The following figure (fig. 7) illustrates how a KDE is applied based on case typical events (si) (e.g. crimes) in space (R): moving three-dimensional functions (ki), based on specific searching radius (τi) (bandwidth) scan grids of points. In this process, distances are measured from the center of each grid cell to points that are around that (Anselin et al., 2000).



Source: Anselin et al., 2000; p. 228

Figure 7: Representation of KDE

There are several kernel functions (e.g., normal, uniform, quartic, triangular, etc.) that can be used to study the concentration and the density of crime events (Perry et al., 2018.; Thakali, Kwon & Fu, 2015). It has been observed that different functions provide very similar density values (Chainey, 2013; Thakali et al., 2015). Most kernel functions support that the risk of a future crime event decreases as the distance from crime events increases (Perry et al., 2018). Normal kernel functions use the standard deviation of normal distribution as bandwidth whereas uniform, quartic and triangular functions use the radius of the search to be interpolated. However, the final choice of bandwidth remains a subjective issue (Thakali, Kwon & Fu, 2015).

Cao, Zhang, & Sun (2013) defined KDE technique as a non-parametric way to estimate the distribution of data, using the probability density function f(x) of random x values without initial assumptions. The surface is estimated by the following equation:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \qquad (1)$$

In the above equation Xi refers to any of the data samples, K() is the Kernel function, h is a smoothing parameter that determines the width of kernel function, Kh(X) = (1/h) K(X/h) and n is the sample size.

Likewise, Chainey (2013) defines the function of KDE through the following equation:

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

In equation (2) f(x, y) is the density value at a (x, y) location, n is the number of observations, h is the bandwidth, di is the geographical distance between incident i and location (x, y), and k() is the kernel density function. In other words, in equation (2) di=X-Xi, comparing to the previous one (1). Each (x, y) location is represented spatially as a grid cell (the coordinates referring to the centroid of that cell), with the calculated density value "f" attributed to each cell. The smaller the cell size the more the calculations of "f". The shorter side of the study's areas minimum bounding rectangle should be divided by the arbitrary value of 150 (Khalid et al., 2018). This provides useful starting measure irrespective of any value of cell size. On the other hand, regarding bandwidth (h), Khalid et al. (2018) suggest a specific range for that from calculating "$h=0.68*n^{-0.2}$".

2.3.6 Local Indicators of Spatial Association (LISA)

One of the most popular indexes to examine local spatial autocorrelation is the LISA (Local Indicators of Spatial Association). LISA statistics today are useful for two reasons. Firstly, they are used as methods for local patterns of hotspots, like the Getis-Ord Gi*. Apart from that, they can identify clusters and outliers through specific tools, like Moran scatterplot (Anselin, 1995). For each observation value, LISA provides a significant spatial clustering of similar values around that observation.

One way to implement the LISA method is through the **Getis-Ord Gi*** statistics. Many of the G statistics evaluate the degree of spatial association of a variable within each single point's neighborhood. Gi* is a statistical index that focuses on local patterns (e.g. zip-code areas) of variables' dependence (Getis & Ord, 1992). The Gi* method was introduced by Getis and Ord in order to identify a tendency for spatial clustering and to divide high and low spatial associations (Songchitruksa & Zeng, 2010).

The standardized Gi* local statistical method is calculated through the following equation (3), in which Xi is the attribute value for a feature j, Wij is the spatial weight

between i and j while n symbolizes the total number of features (Jana & Sar, 2016; Songchitruksa & Zeng, 2010).

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_i - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{[\sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2]}{n-1}}} \qquad (3)$$

$$\bar{X} = \frac{\sum_{j=1}^{n} X_j}{n} \qquad (3a) \qquad\qquad S = \sqrt{\frac{\sum_{j=1}^{n} X_j^2}{n} - (\bar{X})^2} \qquad (3b)$$

The Gi* statistic gives a z-score for each feature, so it does not require further tests for statistical significance. According to the outputs a positive value for z indicates a hot spot (clustering) for high values whereas a negative result a cold spot (spatial clustering for low values) and as larger or smaller is the z-score value the more intense is the hot or cold spot, respectively (Anselin, 1995; (Getis & Ord, 1992; Ord & Getis, 1995).

The second way to implement the LISA method is through the **Local Moran's I**. This index is used to test spatial autocorrelation and it has been developed by Moran in 1948 (Anselin, 2007). When spatial autocorrelation is examined for certain geographic area (e.g. city blocks) and input data is only point data (e.g. crime incidents), local Moran's I is an appropriate method to test for clustering (Gonzales et al., 2005). The local Moran's I shows the association of examined features in a region i with their neighbors at j regions regarding their values. More particularly, if I is positive it means that a feature has neighbor features with similarly high (H) or low (L) values and it is part of a cluster. If I is negative it means that the examined feature's neighbors have dissimilar values and the examined feature is part of outlier (Scardaccione et al., 2010). Regarding the values of I, it ranges from -1 to 1; the closer is to one of these values, the stronger is the association between a feature and its neighborhood.

A local Moran statistic for an observation is defined through the following equation (4). In this equation, Zi and Zj are the normalization form of observations Xi and Xj by mean (4a, 4b), their summation underlines that only the neighboring values j (4c), Wij is a weight indexing location of i relative to j, $\bar{X}$ is the mean of Xi and Xj attributes and n is the total number of features i and j (Anselin, 1995).

$$I_i = z_i \sum_{j=1, j\neq i}^{n} w_{ij} z_j \qquad (4)$$

$$z_i = \frac{\mathrm{x_i}\text{-}\bar{X}}{S_i^2} \quad (4a) \qquad z_j = \frac{\mathrm{x_j}\text{-}\bar{X}}{S_j^2} \quad (4b) \qquad S_i^2 = \frac{\sum_{j=1, j\neq i}^{n} (X_j - \bar{X})^2}{n-1} \quad (4c)$$

In order to local Moran's I be interpreted it should be combined by a z-score and a p-value, which show the statistical significant of index values. The lower is the p-value the more is the statistical significance of output cluster or outlier. A value of 95% or upper confidence level for significance is usually used which gives a 0.05 p-value (Scardaccione et al., 2010). Regarding the z-score, the following equation (5) illustrates its implementation, while the equations (5a) and (5b) provide the functions of the expected Moran's I at location i and its overall variance, respectively (Scardaccione et al., 2010). The expected Moran's I (E[I]) is the ideal value by random distribution.

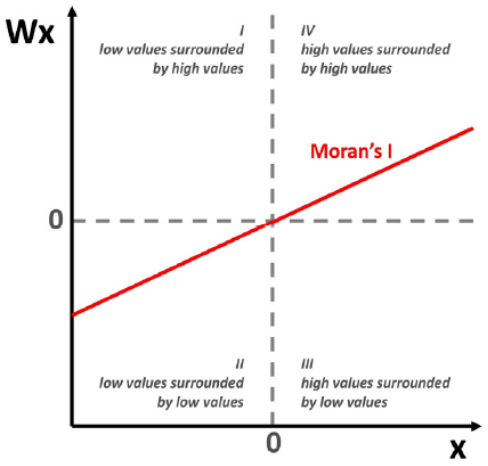
$$z_{It} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \quad (5)$$

$$E[I_i] = -\frac{\sum_{j=1, j\neq i}^{n} w_{ij}}{\text{n-1}} \quad (5a) \qquad V[I_i] = E[I_i^2] - E[I_i]^2 \quad (5b)$$

Apart from z-score and p-value, the Moran's I provides four CO-Types (Clusters and Outliers) depending on the values of I. The first CO-Type refers to significant cluster of high values (HH), which means that a feature and its neighbor features give high values. This first category defines a hot spot area. The second category (cold spot) involves significant cluster with low values (LL). In other words, in this category an examined feature as well its neighbors provide low values. The next two categories address to outliers. More specifically, the third category refers to low values which are surrounded by primary high values (LH), which means low level of similarity. The fourth category refers to high values with low level of similarity with its neighbor values (HL). Another one category (fifth) corresponds to non-significant autocorrelations and it happens when I is very close to zero (Scardaccione et al., 2010).

Local Moran's I is closely related to the Moran's scatterplot, which indicates the presence of local spatial clusters or outliers (Anselin et al., 2000). The graph below (fig. 8) presents the meaning of the Moran's scatterplot, which indicates a better understanding of the spatial distribution of examined features (Anselin, 1995). According to Barreca et al., (2017) the horizontal axis (x) refers to standardized value of variable while the vertical axis (Wx) shows the normalized spatial lag of the same

variable. This graph is highly depended on the study area and according to Scardaccione et al. (2010) the bigger is the scale the better is the performance.



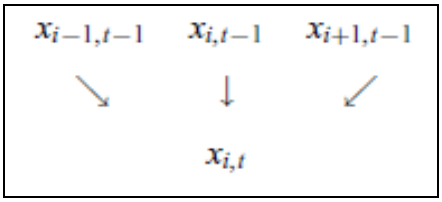
Source: Barreca et al., 2017

Figure 8: Anselin's Moran scatter plot interpretation guide

2.4 Spatiotemporal (ST) Autoregressive (AR) Methods

"A space-time model is a special time series model which is used for the calculation of linear dependencies between the variables in both time and space" (Kurt & Tunay, 2015; p. 3). Generally, case study variable (yit) is obtained from k number of study areas or zones (i =1, 2,…, k) for multiple periods T (t=1, 2,…, T) (Kurt & Tunay, 2015; Giacomini & Granger, 2004). The autoregressive term includes the term of correlation in space and time (Griffith, 2010).

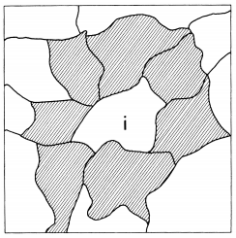
A variable Xt is measured over time t in three neighboring areas i-1, i and i+1, where i-1 and i+1 are the neighbor locations around an crucial location i. Considering that there is dependency between neighborhood areas, the variable Xit will be estimated by the other three values of the same variable in space and time (fig. 9) (Giacomini & Granger, 2004).



Source: Giacomini & Granger, 2004

Figure 9: The idea of spatiotemporal model

An important term that will be referred to below is the lag and it addresses spatial and temporal lag. Spatial lags are related to the potential location of a future crime event and are similar to distribution lags. Spatial lags describe the spatial dependency per distance class while temporal lags, the temporal one (Martin & Oeppen, 1975). The following figure (fig. 10) gives an example of spatial lags for a study area (i) and its neighborhood (Cliff & Ord, 1975). The conditional average of variables (yit) is modelled by each area (i) and a linear function of the past values of the same variables in neighboring areas (Kurt & Tunay, 2015).



Source: Cliff & Ord, 1975

Figure 10: Spatial lags for neighborhood of i

Space-time autoregressive models have been introduced by Cliff & Ord (1975) and generalized by Pfeifer and Deutsch (Giacomini & Granger, 2004). Space-time models rely on the assumption that there are relationships among variables over different areas in the consideration that close variables have higher geographical and, probably, time dependency (Kurt & Tunay, 2015). Recently, interest for space-time modelling is experiencing a resurgence due to the high availability of data sets, the development of programming and technical skills and the necessity for better forecast methods (Griffith, 2010). The most challenging issue concerning space-time modelling is to accurately define the relations between the regions under examination (Kurt & Tunay, 2015). This is a serious problem because the spatial variables can vary through time.

Spatiotemporal models have been widely studied in crime analysis over the latest years (Lin et al., 2018). Below, a couple of spatiotemporal modelling methods are referred to, focusing on their theoretical characteristics, their functions and some relative studies they have been used in.

2.4.1 ST-AR (Spatiotemporal Autoregressive) Models

ST-AR models have clear specifications to describe variables and how they change in space and time (Griffith, 2010). The ST-AR model that is defined in Shoesmith (2013) consists of two parameters, p (autoregressive lags of the region's own log-differenced

crime rate) and q (spatial lags). These parameters can be shown through the following equation (6), in which ci is a constant term for each of i regions, Di is a matrix of covariance of observations, φi (j=1, 2, …, p) coefficients of autoregressive lags (p). The ψl expresses a single coefficient across all n regional equations for each lag i of wik*Δyk,t-l, where wik is a weight matrix (Shoesmith, 2013).

$$\Delta Y_{t+\tau} = \sum_{i=1}^{n} \Delta y_{i,t+\tau}^{star} = \sum_{i=1}^{n} [c_i D_i + \sum_{j=1}^{p} \varphi_i \Delta y_{i,t+\tau-j} + \sum_{k=1}^{n} \sum_{l=1}^{n} \psi_l w_{ik} \Delta_{k,t+\tau-1}] \qquad (6)$$

A simpler form of STAR models that is defined through the following equation (7) has been proposed by Kurt & Tunay (2015) in which l represents time lag and s symbolizes the spatial lag. Temporal and spatial order correspond to p and k, respectively.

$$y_t = \sum_{l=1}^{p} \sum_{s=0}^{k_l} \varphi_{ls} W^{(s)} y_{t-l} + \varepsilon_t \qquad (7)$$

A drawback of STAR models is that they rely either on a single – unique space-time autocorrelation parameter or on two parameters of autocorrelation (one spatial and one temporal). In both cases it remains difficult to capture the high levels of heterogeneity involved in space-time data (Griffith, 2010).

2.4.2 ST-MA (Spatiotemporal Moving Average) Model

While in STAR models space-time autocorrelations tail off both in space and time and partial autocorrelations end after particular temporal and spatial lags, STMA models represent autocorrelations that end after particular lags which decay over time and space (Kamarianakis & Prastacos, 2005).

The following equation (8) (Kurt & Tunay, 2015) illustrates a simple form of STMA models, where now the temporal and spatial order moving average process are symbolized by q and m respectively. A random error (ε) is an uncorrelated in space and time term that not only cannot be avoided but is also as significant as the other parameters (Griffith, 2010) and θ_{ls} is the moving average parameter at temporal lag k and spatial lag I (Phillip E. Pfeifer & Deutsch, 1981).

$$y_t = \varepsilon_t - \sum_{l=1}^{q} \sum_{s=0}^{m_l} \theta_{ls} W^{(s)} \varepsilon_{t-l} \qquad (8)$$

2.4.3 ST-ARMA (Spatiotemporal Autoregressive Moving Average) Model

STARMA models have been proposed by Pfeifer and Deutsch in the early 1980s. Although they have been used since, they had not been investigated in depth due to lack of available datasets. When STARMA models were actually examined in depth, many researchers found that although these models can be estimated by linear and non-linear estimators, they depend on weight parameter values. Although the second type provides more sufficient results, there is not a lot of theoretical background as for its implementation (Kurt & Tunay, 2015).

Compared to the statistical methods and the dynamic space-time models that have been developed, Space-Time Autoregressive Moving Average (STARMA) assumed to perform better (Kurt & Tunay, 2015). What makes the STARMA model more tractable is the use of a grid; this helps the formulation to be well suited to forward simulation in time, leading to a decrease of calculations (Glasbey & Allcroft, 2008). The following equation (9) illustrates the calculation function of the STARMA model (Pfeifer & Deutsch, 1981).

$$z_t = \sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \varphi_{kl} W_l z_{t-k} - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} W_l \varepsilon_{t-k} + \varepsilon_t \qquad (9)$$

In equation (9) p is the autoregressive order, q is the moving- average order, λ_k is the spatial order of the kth autoregressive term, m_k is the spatial order of the moving average term, φ_{kl} and θ_{kl} are parameters; φ_{kl} is the autoregressive parameter at temporal lag k and spatial lag I and θ_{kl} is the moving average parameter at temporal lag k and spatial lag I (Pfeifer & Deutsch, 1981). The first term of the above equation (9) (including the error ε) defines the STAR (p, λ) model. The second term (including an error ε) is the STMA (q, m) model. Finally, W symbol illustrates the weight parameter values, probably related to the distance between the examined district values in space and time. Thus STARMA (p, q, λ, m) is the composition of these two models, as it can be seen in equation 9 (Kurt & Tunay, 2015).

The STARMA model has two special subclasses: Auto-regression (AR) (pλ1, λ2, …, λp) and Moving average (MA) (qm1, m2, …, mq). In other words, when p=0 final model refers to STMA and when q=0 to STAR (Pfeifer & Deutsch, 1981). Equations of Kurt & Tunay, (2015) provide a simple form of STARMA (p, q, k, m) model, as it can be seen below (equation 10), which can be easily read regarding its composition by equations of STAR and STMA, eq. 7 and eq. 8, respectively.

$$y_t = \sum_{l=1}^{p} \sum_{s=0}^{k_l} \varphi_{ls} W^{(s)} y_{t-l} - \sum_{l=1}^{q} \sum_{s=0}^{m_l} \theta_{ls} W^{(s)} \varepsilon_{t-l} + \varepsilon_t \qquad (10)$$

Generally, STARMA models identify the spatiotemporal dependence between different regions and considers the autocorrelation among variables (Zhang et al., 2019). According to that, STARMA models thoroughly examine the autocorrelation between the geographic spatiotemporal data sets and it is suitable for processing spatial-temporal series. In addition to that, STARMA models provide very convenient methods for data analysis (Kurt & Tunay, 2015).

On the other hand, because STARMA models are highly affected by the spatial time series data, when there is a lack of many measurement locations, the model appears too parsimonious (Kamarianakis & Prastacos, 2005).

2.4.4 ST-ARIMA (Spatiotemporal Autoregressive Integrated Moving Average) Model

Having explained the STARMA method, is simplest to explain what makes the difference between this (STARMA) and STARIMA is the term of Integration. This term shows indicates the non-stationary situation of variables. A STARIMA model is defined with the parameters p, d and q. More particularly, p shows the dependent variables, q shows the number of lags of errors while d shows the number of times a non-stationary variable can be stationary. Related to the previous model method, if STARIMA is defined through (p, d, q) parameters, and d=0, then the model becomes STARMA, whereas if d=1 a linear trend guides the model to be stationary. The following equation (11) shows the implementation of STARIMA (p, d, and q), where λ_k is the spatial order of the kth autoregressive term, the m_k is the spatial order of the kth moving average term, the φ_{kl} is the autoregressive parameter at temporal lag k and spatial lag l, the θ_{kl} is the moving average parameter at temporal lag k and spatial lag l, W^l is a weight matrix for spatial order l, ε(t) is the random normally distributed error at time t and the other common symbols / terms with STARMA have explained above (Duan et al., 2016; Pfeifer & Deutsch, 1980).

$$\nabla^d z(t) = -\sum_{k=1}^{p} \sum_{l=0}^{\lambda_k} \varphi_{kl} W^{(l)} \nabla^d z(t-k) + \varepsilon(t) - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} W^{(l)} \varepsilon(t-k) \qquad (11)$$

STARIMA models have been characterized as a useful tool by researchers for dealing with datasets that include large spatial and temporal dimension (Kamarianakis & Prastacos, 2005). STARIMA (p, d, and q) models with $q \neq 0$ are non-linear models.

Therefore, appropriate Taylor's expansions are need to linearize the model (Kamarianakis & Prastacos, 2005).

2.5 Evaluation of Forecasting Performance for Regression

2.5.1 R^2 (R-squared)

The R-squared (R^2) metric was originally developed for linear regression statistical models (e.g. STATA) and its values ranges from 0 to 1 (Cameron & Windmeijer, 1996). The higher is its value the stronger is the relationship between the examine values. The general (unweighted) equation of R-squared tries to show the relationship between two variables, such as incidents and the remaining observations and it is provided through the following equation (12), in which y is the dependent variable with i=1, 2, …, N, the μ is the predictive values of the applied model and ẏ the mean of variables (Cameron & Windmeijer, 1996). In other words, the numerator refers to variance explained by the used model while the denominator the total variance.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{\mu}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \qquad (12)$$

2.5.2 Error Metrics

In this subsection there are two popular metrics which are based on measuring the models' performance errors. These are the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) and the corresponding definitions and equations are provided below in the following two sub-sections.

Before that, it should be clear the importance of errors in a model's performance. As it is defined in Willmott & Matsuura (2005) an error (ei) is the difference between the predicted values (pi) and the observed values (oi), for i= 1, 2, …n (the case incidents) as it is shown by the following equation (13).

$$e_i = p_i - o_i \qquad (13)$$

However, these metrics usually use weights to distribute the potential errors in each value and the following equation (14) illustrates the average model-estimation error. In this equation, the wi is the weight as proportion to the total error and it is scaling assigned to each $|ei|^{\gamma}$. When it is considered as "equal" to all errors (e.g. through the

MAE metric) it is usually equal to 1 and regarding the γ it is an index that should be equal or over 1.

$$\overline{e}_{\gamma} = \left[\frac{\sum_{i=1}^{n} w_i \mid e_i \mid^{\gamma}}{\sum_{i=1}^{n} w_i}\right]^{\frac{1}{\gamma}} \qquad (14)$$

A popular error metric is the **Mean Absolute Error (MAE)**. This metric has been created to compare forecasting performance of all models (e.g. STARMA, ARMA, etc.). Likewise, the calculation of this metric is depicted through the following equation (15) (Chai & Draxler, 2014). It should be mentioned that the use of absolute value is usually a drawback for metrics because it prevents the calculation of other features, such as the gradient or the sensitivity of the MAE metric (Chai & Draxler, 2014).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid e_i \mid \qquad (15)$$

According to Willmott & Matsuura (2005), the MAE metric gives an equal weight to all errors, it is unambiguous and it presents the measure of average error more naturally than the RMSE. However, the MAE is inappropriate when the error distribution is expected to be Gaussian, that is to not get extreme error values (Chai & Draxler, 2014).

Another popular error metric is the **Root Mean Square Error (RMSE)**. The calculation of this index can be summarized in three main steps. The first one concerns the total square error, calculated by summing the individual squared errors. Then, the total square error is divided by the number of samples and finally its square root is calculated. Defining errors as ei (i=1, 2, …, n), where ei is the difference between the variables yi and the relative observations, the RMSE can be calculated by the following equation (16) for n samples (Chai & Draxler, 2014).

$$RMSE = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2\right)} \qquad (16)$$

According to Willmott & Matsuura (2005) RMSE metric does not describe the average model performance error so well probably because this metric is a result of the three above mentioned steps and it does not use the errors weight analogous to each value.

2.6 Evaluation of Forecasting Performance for Classification

2.6.1 Traditional Accuracy Metrics

The performance of classification models is usually evaluated with metrics such as these depicted below. In order to characterize the performance of a model a confusion matrix could be used (fig. 11). *"The standard setup for event-driven classification is to have a so-called "positive" class that represents the target events while the "negative" class represents all non-events"* (Torgo & Ribeiro, 2009; p.6). Row classes refer to the predicted results while the column ones to the actual results (Hossin & Sulaiman, 2015). The True Positive (TP) means that the predicted "positive" agrees with the actual "positive". The True Negative (TN) means that the predicted "negative" agrees with the actual "negative". Likewise, the False Positive (FP) and the False Negative (FN) show "non-agreement" between the predicted and actuals "positive" and "negative", respectively.

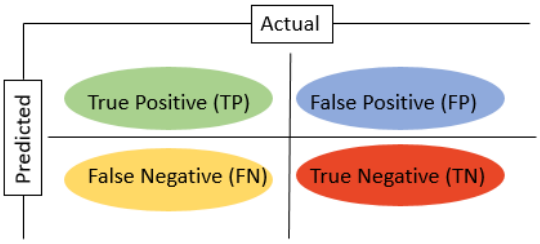


Figure 11: Classification performance basic metrics

A default metric to evaluate the **accuracy** (trueness and precision) binary values or multi-class datasets of a classification method is to calculate the percentage of correctly classifies events (ẏ) out of all events (y). The following equation (17) illustrates this action. The output ranges from 0 to 1 and the higher is the result the higher is the accuracy (ISO, 1994).

$$Accuracy(y, \hat{y}) = \frac{1}{n}\sum_{i=0}^{n-1} \hat{y}_i = y_i \qquad (17)$$

Another classification metric is the **F1**, which consists of **Precision and Recall metrics.** These metrics identify the performance of the applied models on case incidents (e.g. crime events) and ignore the events not included in using classes. Regarding precision, it is defined as the percentage of events of the model which are real events. On the other hand, recall metric, provides the proportion of events occurring in the same domain with the events of models (Torgo & Ribeiro, 2009; Hossin & Sulaiman, 2015).

These two metrics are interconnected and sometimes both are required. Due to that, there are test that give their composition outputs, such as the F1, as it can be seen in the following equation (18) (Torgo & Ribeiro, 2009). This metric represents the harmonic mean between recall and precision values (Hossin & Sulaiman, 2015).

$$F1 = 2 \times \frac{\Pr ecision \times \mathrm{Re} call}{\Pr ecision + \mathrm{Re} call} \quad (18)$$

2.6.2 Crime Forecasting Accuracy Metrics

Two are the most popular crime forecast accuracy metrics, the Hit rate and the Prediction Accuracy Index (PAI).

According to Chainey et al. (2008) the **hit rate** is the percentage of crimes events that occur within the areas where crimes are predicted to occur (hotspots) divided by the total number of crime events of the whole study area. This index may be useful and reliable for a single study area but not for further comparisons.

A more objective index is the **PAI** which is accompanied by the area percentage. This metric (eq. 19) has been proved as appropriate for this purpose by many researchers (Perry et al., 2018; Chainey et al., 2008).

$$PAI = \frac{(\frac{n}{N}) \times 100}{(\frac{a}{A}) \times 100} = \frac{\text{Hit Rate}}{\text{Area Percentage}} \quad (19)$$

In the above equation n is number of crimes in areas where crimes are predicted to occur (e.g. hotspots), N is the number of crimes in the study area, α is the area of areas where crimes are predicted to occur (e.g. area of hotspots) and A is the area of the study area (Chainey, 2013). According to equation (19) if the percentages are equal PAI will be 1. By finding 25% of future crimes (Hit Rate) in 50% of the study area (Area Percentage), PAI will be 0.5. If these percentages correspond to 80 and 40, respectively, PAI will be 2, and so on. As such, the higher the PAI values the greater the predictive accuracy (Chainey et al., 2008).

2.7 Selection of Forecasting Methods and Metrics for Performance Evaluation

This section describes the selected forecasting methods and the selected performance metrics based on their suitability to our research data and scope.

Regarding the hotspot analysis, KDE is selected, which, although is just a simple interpolation method, it is already used by police departments to "forecast" crime areas

(i.e. hotspots) because of its benefits. It is widely used in the field of criminology because it can be interpreted easily and provides better forecasting results than the grid thematic mapping or the spatial covering ellipses techniques. Apart from KDE, a baseline approach is employed too. The approach simply assumes that the most recent situation of crime incidents could be used as forecasting, without any further editing process. This naïve baseline method will be used in order to have a more neutral comparison.

Regarding the Space-time autoregressive methods, STARMA is chosen as more appropriate to create a forecast for potential locations and time periods concerning crime occurrences. In addition to that, STARMA outperforms other Spatiotemporal Autoregressive methods (e.g. STMA, STAR, etc.) in identification space-time dependence between different areas and takes into account the correlation between the using variables (Zhuang et al. 2017).

As far as the performance metrics are concerned, the PAI metric is used as considered to be more appropriate for the scope of this study because it is a crime analysis metric and it is the most used by criminologists. However the Hit rates is also examined (Kounadi, Ristea, Araujo & Leitner, 2020).

The final comparison will occur based on classification outputs (hotspots and non-hotspots). For all the three methods, a threshold value will transform their predicted outputs into the two above mentioned classes. This threshold value is determined after making five quantile classes of outputs and defining the highest one as hotspot, whereas the first four classes are grouped as one “non-hotspot” class. This way is based on methodology that have been used by Chainey et al. (2008) for similar comparison purposes.

3 Methodology

This chapter refers to the methodology that is used for this thesis topic and the steps followed in order to answer the research objectives. More particularly, the following sections and subsections refer to software that is used, which is the mean to carry out the analysis, the study area, the data and, finally, the workflow (fig. 12), which gives a schematic illustration of the methodology.

3.1 Methodological Overview

The first step of the methodology consists of collecting the necessary data (i.e. crime incidents, study area, study zones). The second step is the pre-processing phases, where input datasets are transformed into an appropriate form for the subsequent analysis. This step is explained better through the table 3 in 3.4.4 subsection. The third step is the implementation of current (baseline, conventional) and proposed methods. The last step evaluates the results of the methods applied based on the selected performance metrics and once outputs from applied methods have been classified to hotspots or non-hotspots.

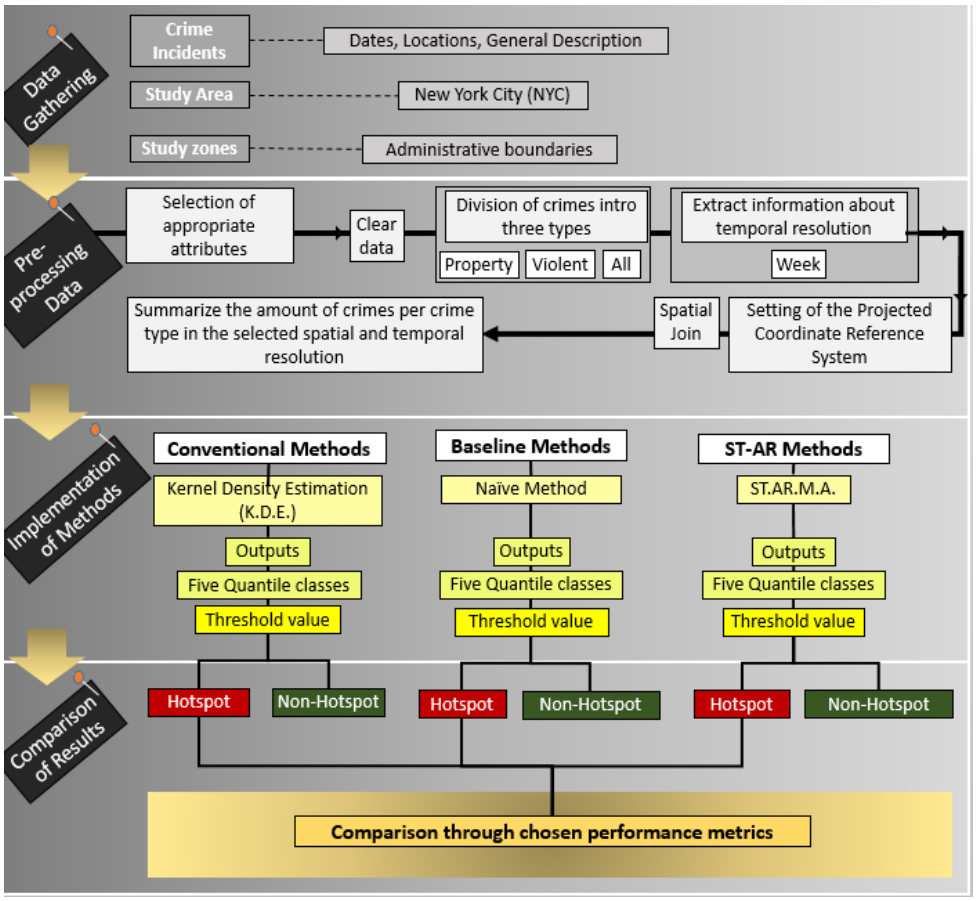


Figure 12: Workflow for evaluation of proposed (ST-AR) method for forecast crime

3.2 Software & Reproducibility

The implementation of the conventional method (i.e. KDE) is done with ArcGIS software while the proposed one (i.e. STARMA[3]) is applied in R-studio interface environment. The implementation of the methods in each respective software concerns also the attempt to make automate process through specific tools (e.g. model builder) or by programming extending code in repeatable loops.

3.2.1 R-studio

Although R-studio is an interface to work in R language, it is referred here as one of the software that was used. In initial processes, it was used to transform default (original) datasets to an appropriate structure. Additionally, this software is used for the proposed method because of the current available packages and libraries. Finally, the evaluation of implemented methods takes place in R-studio environment too, in an automate way. By programming all the processes in this environment ensures the validity of the processes and, at the same time, offer a suitable environment where potential changes could easily occur.

3.2.2 ArcGIS

Geographic Information Systems (GIS) have been widely used as significant tools through their toolbox and the producing maps, in crime investigation under prevention programs (Pawale et al., 2017). A big advantage is that this software provides the ability to create a new tool or modeling a process based on a combination of existing tools from arc toolbox library. When a model builder is ready to run, it can be exported as script and then a useful description can be written in order to guide a potential user more easily by publish it (e.g. in ArcGIS online web site). This software concerns not only the implementation of the KDE method but also the presentation of results for all the three methods in order to have a comparable overview of some specific cases.

3.2.3 GitHub repository

In order to make the workflow of the thesis reproducible, which concerns the pre-processing steps, the implementation of the proposed method, and the evaluation /

[3] Modelling Space Time AutoRegressive Moving Average (STARMA) Processes: https://cran.r-project.org/web/packages/starma/starma.pdf

comparison of all the three methods, a github repository[4] has been created. With the repository involving a guide "readme" file, potential users can be informed about the purpose of the included R-code files. In this way, the processes of the workflow can be repeated and further used in similar researches or policing applications.

3.2.4 Software Limitations

This subsection refers to the limitations that came to deal with. More particularly, working on R-studio for the implementation of the proposed method resulted in serious time computational problems when data of high spatial resolution were used (i.e. 11 Gigabyte). The main cause of the problem appears to be the high dimensionality of weight matrices and consecutively the array for the STARMA model. The higher the weight matrix the bigger the problem and therefore, the option of using many (spatial) neighbor polygons is not feasible. Another option was to run the existing code online, either with increasing the memory of R-studio or with other options, such as "google-collaboratory" or "jupyterhub". Unfortunately, none of these solutions worked and the implementation of all the three methods finally is based on smaller spatial resolution, the second preference, as it is explained in the next sections.

3.3 Study area

New York City (NYC) (fig. 13) is chosen as the case study area of this thesis topic as being a metropolitan area with very high crime rates. Also, the selection of this study area is based on the quantity and the quality (both in space and time) of the data available. More particularly, formal web sources, like the services of New York Police Departments (NYPD), offer a recent and adequate amount of crime incidents in NYC combined with the administrative boundaries validated in the same decade.

[4] Github repository for this thesis:
https://github.com/alkiviadisrentzelos/MScThesis_-Exploring-Space-Time-Autoregressive-models-to-forecast-crime.git

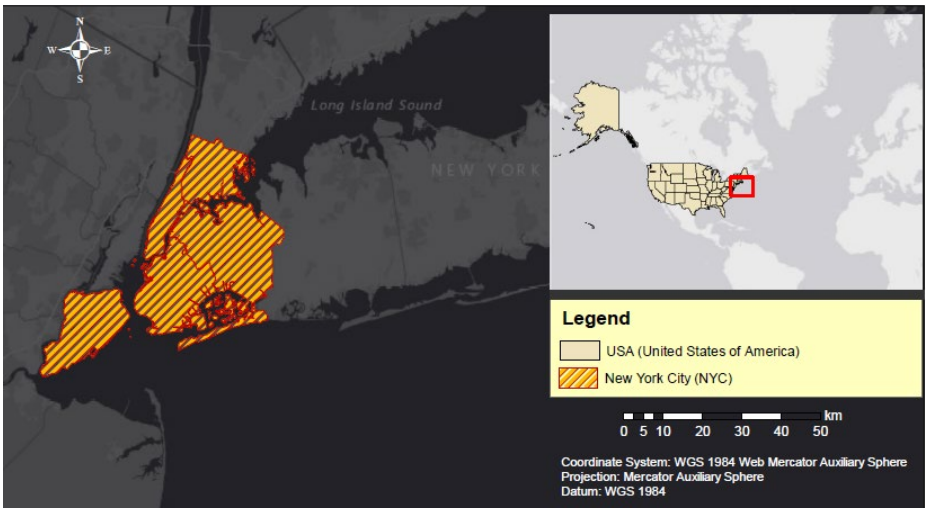


Figure 13: Study area - NYC

3.4 Data

The required data is the crime incidents accompanied by their spatial and temporal reporting. In addition to that, study zones are required to divide the study area in smaller areas (i.e. the spatial units of prediction) and these zones can be administrative boundaries or grid cells. The following subsections illustrate the sources and the features of these main categories of required data.

3.4.1 Datasets for Crime Incidents

Several datasets about crime incidents in NYC with their spatial and temporal features are provided by the NYPD[5]. During the last decade there are the following available datasets: the first one covers the whole of 2019 (the most recent dataset), the second one dates from September 2014 until the end of 2017, while for the year 2018 there is a different format of datasets. Among them, the one that is preferred is the datasets with the more covered period (more than three years). From this datasets, the selected study period will be the year 2016, in order to have available complete observation data before and after in case that an extension of the study period is required.

The selected dataset includes 18 attributes (they can be seen in the external appendix file), some of them being expressed as codes while others providing further details. More particularly, codes divide the crime incidents according to specific categories, like the law or jurisdiction type. Apart from that, there are attributes which

[5] Open data about crime events in New York City by Police Departments (NYPD Arrests Data Historic): https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u

describe the crimes in broad and others in more specific categories. There are also attributes presenting the location and the dates of arrest involving crime incidents and others provide secondary information (e.g. the sex or the age of perpetrators).

Amongst all these attributes, those which are selected are associated with the location of crime incidents, the dates when they occurred and their description. Regarding the description there are a lot of categories. Here, the general description attribute is preferred. Therefore, the division of crimes into two types (property and violent) is based on their general description, whereas the third type (all) includes all the crimes as incidents and also those not belonging to any of the first two types. In order to do this division, formal crime typologies were used, like the formal crime typology of NCJRS[6] (National Criminal Justice Reference Service), as well as studies that provide a division of crimes to 'property' and 'violent' types (Shoesmith, 2013). This process is further explained in pre-processing data subsection (3.4.4) (the detailed division of categories can be seen in external appendix file).

Concerning the location of a crime there are two pairs of attributes in different coordinate systems (projected and geographic). The selected one will concern the projected Coordinate Reference System (CRS), which corresponds to feet as units of measurement and this helps calculate the distances among crime points. Finally, attributes which provide further information, such as the age or the sex of perpetrators, the law type etc., are not used.

For the chosen study period (year 2016), the total number of crime incidents is 312,345 ('all' crimes). Amongst them, 73,194 are 'property' and 74,780 are 'violent' crimes.

An issue that needs exploration is the real location of the original data, before they are spatially joined in zones they belong to. As it can be seen in the following figure (fig. 14), crime points are located in the middle of streets. This means that crime incidents which occurred in neighboring buildings with a significant distance are appeared in the same location (i.e. the closest street). So, even though they marginally belong to a site, their spatial accuracy may be "decreased" analogously to the how big the study zones are.

[6]National Criminal Justice Reference Service: https://www.ncjrs.gov/App/Topics/Topic.aspx?TopicID=56

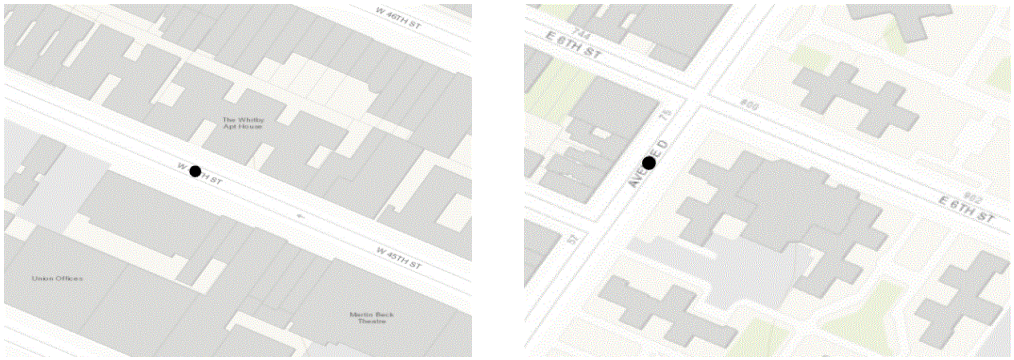


Figure 14: Original locations of crime points

3.4.2 Datasets for Spatial Resolution

Apart from the dataset for crime incidents, a dataset which can define the division of the whole study area into smaller zones is required. This requirement is also referred to as the spatial resolution. This can be done by using either a technical grid (square) cells or administrative boundaries. For this thesis, the former choice is not preferred as a uniform grid cell because it would intersect the existing urban environment "disconnecting" the location of crimes from their physical part of the study area. For example, by using grid cells, it is possible to have some crime incidents occurring in unnatural environments, like property theft on the sea or on rivers.

The most popular datasets available for the selecting boundaries[7] of NYC can be seen in the next figure (fig. 15). The whole NYC which can be represented by one polygon, is divided into 5 boroughs, 77 polygons of NYPP (New York Police Precincts), 248 polygons of postcode (zip-code) areas and finally into 38,799 polygons which are the blocks.

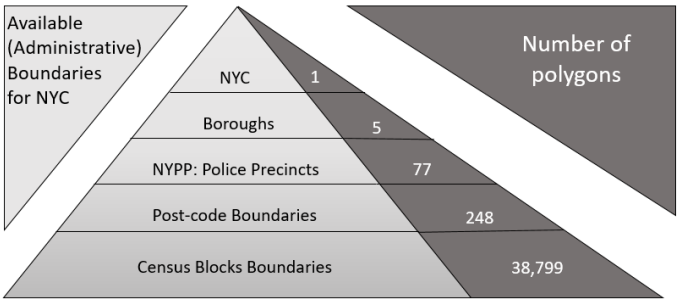


Figure 15: Boundaries available with the number polygons which they consist of

[7] Open data about Administrative and non-boundaries of New York City by the Government: https://data.cityofnewyork.us

According to the related study work, described in subsection 2.2 (table 1), the following table (table 2) presents what other studies chose as regions (zones) related to their study area size. As it has been mentioned in table 1, the division of the study area to smaller regions was based either to the administrative boundaries or to technical grids (rectangle or square).

Table 2: Overview of forecasting crime studies regarding the division of their study area

	Source	Area of the whole study area	Type of study area	Selected administrative boundaries	Technical grids	Average of area of a technical grid cell
1	(Bowen et al., 2018)	701.9 km^2 (~$7x10^9$ ft^2)	County	Census Blocks	-	-
2	(S. Chainey et al., 2008)	36.67 km^2 (~400,000,000 ft^2)	2 Boroughs	-	Rectangle grid cells	125,000 m^2 (~1.5 ft^2)
3	(Dash, Safro & Srinivasamurthy, 2019)	606.1 km^2 (~6.5 x 10^8 ft^2)	City	Communities	-	-
4	(Drawve, Moak & Berthelot, 2016)	316 km^2 (~3.5 x $10^8 ft^2$)	City	-	Square grid cells	8,281m^2 (~90,000 ft^2)
5	(Gorr & Harries, 2003)	151.1 km2 (~1.6 x 10^8 ft2)	City	Police precincts	-	-
6	(Mohler, 2014)	606.1 km^2 (~6.5 x 10^8 ft^2)	City	-	Square grid cells	~15,000 m2 (~160 x 10^6 ft^2)
7	(Shoesmith, 2013)	9,834,000 km^2 (~105 x 10^{12} ft^2)	Country	USA regions	-	-
8	(Zhuang et al., 2017)	375.5 km^2 (~4 x 10^9 ft^2)	City	-	Square grid cells	~360,000 ft^2

Since the above description is not enough to make a decision for what size of study zones should our study area be divided, the following schema (fig. 16) illustrates the information of the table above in a more clear way.

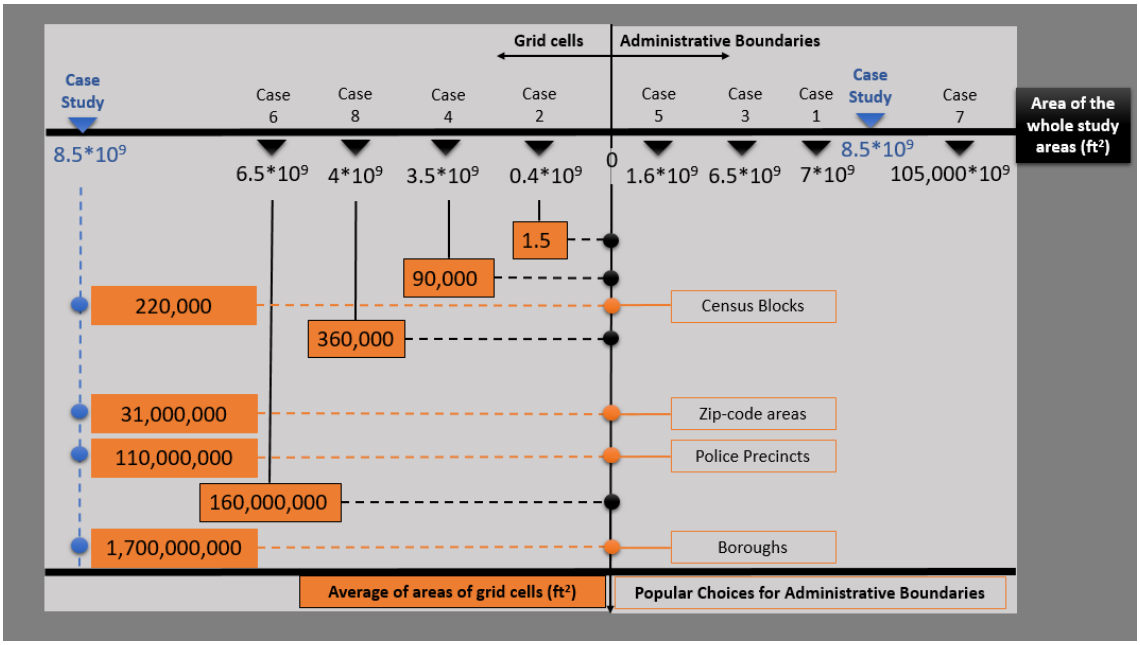


Figure 16: Graphical overview of forecasting crime studies regarding the division of their study area

In the above figure, the eight case studies of table '2' are put on the main horizontal axis. Starting from the zero (center of this axis), studies which used administrative boundaries to examine the study area are located to the right whereas those which used technical grid cells to the left. The cases have been put in an ascending order for both the sides. According to that, the case of NYC is put in the same axis, in both sides, regarding its area (blue one).

The division of the NYC study area into smaller zones is based on how above mentioned studies divided their study areas. Reading upon these studies, it is observed that the smaller the study area the smaller the study zones (spatial resolution) that should be divided. According to that, the case of the NYC should correspond to a large spatial resolution, such as the boroughs or the police precincts. However, this is not necessary because the size of the smaller zones is not the same with the other case studies of the figure above. Furthermore, the selected sample of other studies overview is too limited and thus it would not be completely justified to be based on that for the purpose of this thesis.

In order to select the most appropriate spatial resolution from the available administrative boundaries of NYC, an overview of their area could be more useful. This means that since these boundaries do not have exactly the same area, the similarity of their area for each type should be examined. For this reason, the datasets available for administrative boundaries were imported to ArcGIS software. From this process it

seems that the higher the similarity amongst the spatial polygons the more appropriate they will be.

Based on that, the five Boroughs should be selected as smaller study zones. However, a choice like this would be too broad regarding the scale of the study area. Census Blocks boundaries should be the second option (as they present the second highest similarity amongst their polygons). Zip-code areas seem to be the following option and lastly are the NYPP (these graphs can be seen in external appendix file). So, excluding the five Boroughs for above mentioned reason, the most preferable polygons that could be used as spatial resolution are the Census Blocks and then the Zip-code zones.

The dataset for Census Blocks[8] of NYC is provided through NYC open data and it is the most recent, valid since 2010. It contains 38,799 blocks (polygons) for the whole NYC (fig. 17) with ten attributes: "OBJECTID", "Shape", "bctcb2010", "boro_code", "boro_name", "cb2010", "ct2010", "shape_leng", "Shape_Length" and "Shape_Area". The attribute required from this dataset is their unique code ("bctcb2010") and an attribute which confirm to which of the five boroughs blocks belong (i.e. Manhattan, The Bronx, Brooklyn, Queens and Staten Island). In regard to this attribute, the borough of Manhattan consists of 3,860 blocks, the borough of Bronx of 5,465 blocks, the one of Brooklyn has 9,681 blocks, the borough of Queens includes 14,757 and finally, the Staten Island borough consists of 5,036 blocks.

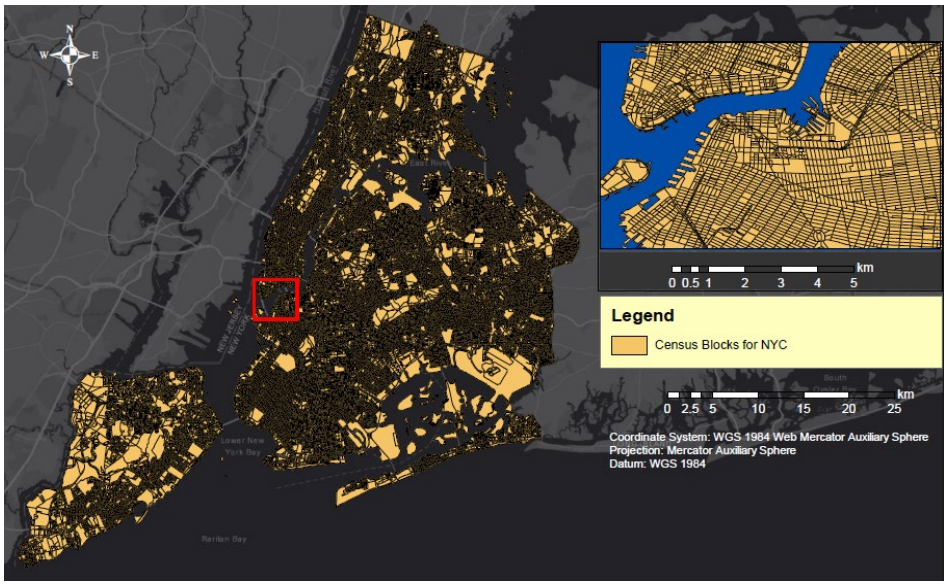


Figure 17: Census Blocks for the study area (NYC)

[8] Open data about census blocks in New York City by Government: https://data.cityofnewyork.us/City-Government/2010-Census-Blocks/v2h8-6mxf

The dataset for Zip-code zones[9] of NYC is provided through NYC open data Government. It contains ten attributes: "Shape", "ZIPCODE", "BLDGZIP", "PO_NAME", "population", "AREA", "STATE", "COUNTY", "ST_FIPS" and "CTY_FIPS". The attribute required from this dataset is their unique code ("ZIPCODE") and the "AREA". The original datasets contain information for the 248 polygons, from which 15 are double subscriptions. So, in case that zip-code administrative boundaries will be used as spatial reference point, the number of unique polygons should be 248. The following figure (fig. 18) illustrates the 248 zip-code zones while the zoom mentions that some of them have quite small size, probably due to overcrowded areas.

Figure 18: Zip-code zones for the study area (NYC)

3.4.3 Datasets for Time Reference Point

One of the main attributes is the "ARREST_DATE" and it is the only attribute which shows when a crime incident occurred. The default time reference corresponds to days. Previous research studies selected to work on days, weeks, 2-weeks, months or years for their time resolution. In this study, all the three methods are examined based on

[9]Open data about zip-code administrative boundaries in New York City by Government: https://data.cityofnewyork.us/Business/Zip-Code-Boundaries/i8iw-xf4u/data?no_mobile=true

weeks for temporal resolution, in order to decrease the volume of the data. Thus, the workload is more manageable and without degrading the time resolution a lot.

The next step is to select the time period for which retrospective data is used for the prediction; in other words how many weeks before the forecast is based on. This, can be roughly answered by examining the temporal autocorrelation of all the weeks (code is included in github repository and also in external appendix file). As it can be seen in the next figure (fig. 19), with one weekly lag (the time lag is the one before the predictive period), the temporal autocorrelation is statistically significant and higher than with two or more weekly lags. More particularly, for the value of zero weekly lag (when the same week is used as predicted period) the autocorrelation is absolutely 1. For the one weekly lag, the temporal autocorrelation decreases to 0.4, for the two weekly lags the value is half (0.2) and so on. Therefore, with one weekly lag it is assumed that the required temporal autocorrelation can be satisfied.

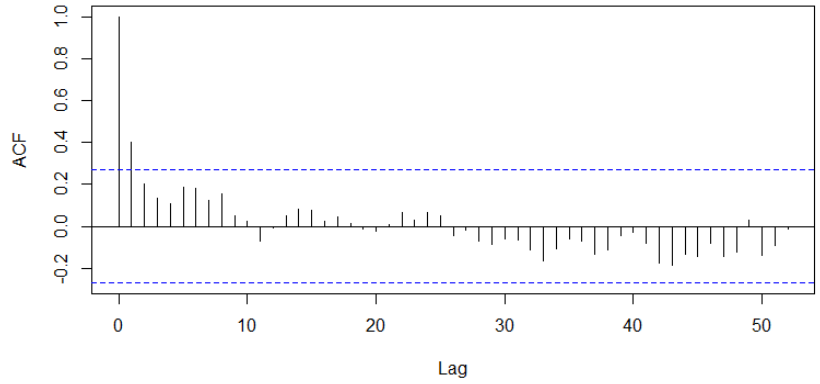


Figure 19: Temporal Autocorrelation of data sets based on week resolution

3.4.4 Pre-processing steps

This subsection includes the pre-processing steps that required for transforming the input datasets into suitable form (the R using code can be seen either in github repository or in the external appendix file).

The pre-processing of input datasets consists of 14 steps which take place in R-studio in order to get reliable and appropriate format of datasets by avoiding manually editing and therefore to eliminate coarse errors. The table below (table 3), represents these steps with the necessary details. To be noted that steps with * concern only the STARMA method.

Table 3: Pre-processes for Crime Incidents Datasets

7 Pre-processing steps for Crime Incidents Datasets
1. Input data
1,048,575 observations (from 9/2014 to 12/2017)
2. Selection of appropriate attributes for both using datasets
For crime incidents: "ARREST_DATE", "OFNS_DESC", "ARREST_BORO", "X_COORD_CD", "Y_COORD_CD" **For study zones:** the attributes related to unique id, geometry and area
3. Selection of the study period
Year 2016 => 314,864 observations
4. Clear data
Delete empty subscriptions (2,708 incidents) and strange descriptions (i.e. "F.C.A.P.I.N.O.S." with 4,402 incidents) => 312,403 observations.
5. Division of crimes to totally three types
58 different crime incidents (all), from which 14 crimes are defined as property (p), 12 as violent (v) and 32 incidents that cannot be characterized either as property or violent.
6. Extract information about temporal resolution
Transform daily information to weekly => (53 in software ~ 52.3 in reality)
7. Setting the appropriate Projected Coordinate Reference System (PCRS)
The coordinates refer to New York State Plane Coordinate System, Long Island Zone, and "NAD 83" with feet as measurement units (FIPS 3104). This Projected Coordinate Reference System corresponds to the code number "102318" according to ESRI Projection[10].
8. Spatial Join
Crime incidents are grouped based on the study zones where they spatially belong.
9. Summarizing the amount of crime incidents per study zones and per crime type
The outputs contain for each week and for each study zone the amount of crime incidents for each type of crimes; property, violent and all as incidents.
10. Replace the N.A. (non-available) values with zero number of incidents
Every study zone should have in numeric format the number of incidents.
11. Replacing of polygons with points regarding the geometry of study zones
The final format of study zones should be represented by their centroids (points) according to how methods require inputs.
12. Make separate files for each type crime*
A division of data per type crime is required.

[10]Spatial Reference of ESRI Projection: https://www.spatialreference.org/ref/esri/102318/

13. Transposition of data *
The step transposes the data in order to have "weeks" as rows and "zip-codes" as columns
14. Extract information for study zones' geometry*
This last step will be used to create the neighborhood of study area

3.5 Implementation of the Naïve Approach (Baseline)

As mentioned in previous chapters, a baseline method is usually used for comparing other methods with a non-extreme situation but with a realistic condition. In this case, the baseline is based on the simple scenario that the data remains the same as one week before. For example, the original data for week t is the same as for week t-1, without any processing. The following figure (fig. 20) illustrates this choice.

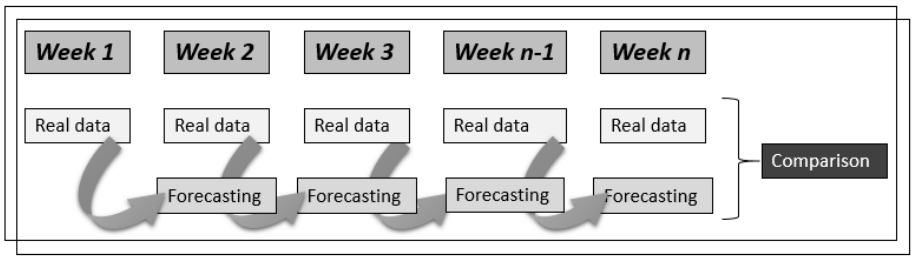
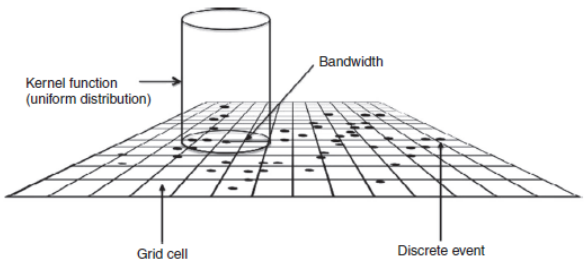


Figure 20: Representation of how the naïve baseline is used

3.6 Implementation of the Kernel Density Estimation (Conventional)

This section provides a clear explanation of the implementation of the method by focusing on both its implementation and technical features (e.g. inputs, outputs and parameters). The implementation of KDE through an ArcGIS software requires three input parameters, namely the bandwidth (search radius), the kernel function (k), t, and the "cell size", related to the resolution of outputs (Hart & Zandbergen, 2014). The following figure (fig. 21) represents the operation of these parameters of KDE.



Source: Hart & Zandbergen, 2014

Figure 21: Visual process of kernel density estimation (KDE)

The following schema (fig. 22) illustrates how a KDE method works. By importing the three input parameters, the KDE command of an ArcGIS software uses the tool "Kernel Density". The outputs of this process is a range of values which represent the density of points in the selected measurement units.

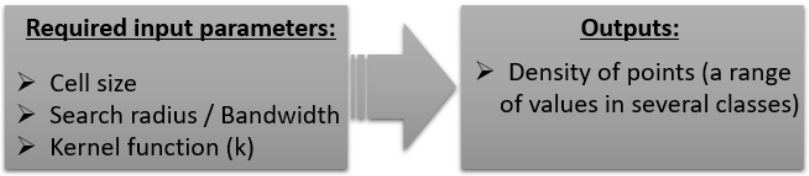


Figure 22: KDE input parameters and outputs

Given that *"KDE is not without its faults"* (Chainey et al., 2008; p. 9), an important issue that discussed by Chainey (2013) is how these two input parameters (cell size and bandwidth) "fit" within the study area; an issue related to the size of the study area and the values of the input parameters. In other words, the outputs from the study area with appropriate values of these two (input) parameters may create small spiky surfaces with the hotspot areas also being very small. This problem could be faced by dividing the implementation of a KDE method into two parts; the first part, in which some key hotspot areas (big size areas) are identified for the whole study area and a second one, in which these key hotspot areas are the study areas and other more specific areas are characterized as hotspots.

Although this technique is theoretically considered to be correct, in practice it is in doubt whether hotspot locations can be identified across the whole study area or not. Furthermore, the GIS software today offers reliable toolsets in order to define the hotspot areas over a threshold value and this makes the previously mentioned technique purely subjective. When the hotspot areas refer to small areas (e.g. blocks) compared to the whole study area (e.g. N.Y.C.), the strange "spiky" surfaces could correspond to the size of blocks.

According to Chainey (2013) the most appropriate values for the two parameters (cell size and bandwidth) can be identified by a number of experiments. However, many researchers make the final selection by looking at the appearance of the mapping output instead of the influence of the selected values to results. Chainey (2013) refers that there is little literature guidance and there is no universal doctrine about the values and under which circumstances these two input parameters are selected (Chainey et al.,

2008). The following table (table 4) summarizes the characteristics of the different options of these two input parameters (cell size and bandwidth).

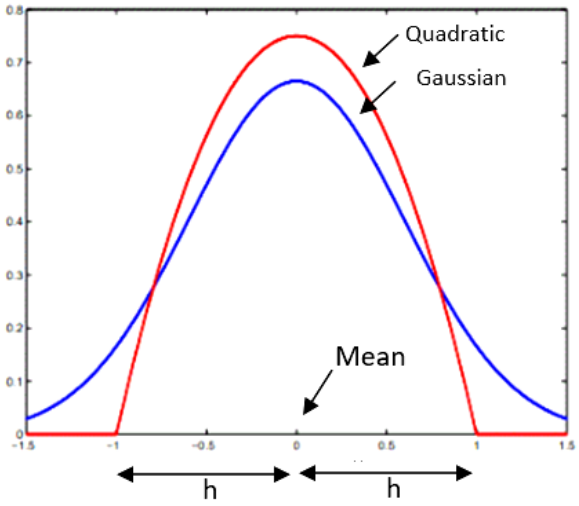
Table 4: Overview of options for two input parameters (cell size and bandwidth) of Kernel Density Estimation

	Options	Implementation	Advantage(s)	Disadvantage(s)	Source
Cell size	**1st option**	Shorter side of the study area divided by 150	Easy to be calculated	It has not been evaluated	(Hart & Zandbergen, 2014)
	2nd option	Default value by the ArcGIS software (the shorter of the width of the output extent in the output spatial reference, divided by 250)	There are not extreme values	Ideal value is unknown	(Hart & Zandbergen, 2014)
Bandwidth	**1st option**	$0.68 \times n^{-0.2}$ (n is number of observed events)	Easy to be calculated	Rough option and does not consider the spatial distribution of the observed events across the study area	Bailey and Gatrell, (1995, p. 86) in (Chainey et al., 2018)
	2nd option	Shorter side of the study area divided by 150 and multiplied by 5	Simple method	It provides only an estimation	Chainey (2011) in (Chainey et al., 2018)
	3rd option	Default value by the ArcGIS software (the smaller of the width of the extent of the input, divided by 30 and reproducing it by multiply the default value for the output cell size parameter by 25/3) based on Silverman's rule of thumb equation: $\left(\frac{4*\sigma^5}{3*n}\right)^{1/5}$ ("σ" is the standard deviation of distribution and "n" is the number of crime events)	ArcGIS 10.2 and over versions has improved the way of calculation of the default value (The lowest possible value is more suitable while the largest value impair KDE purpose)	Lack of scientific justification	Silverman (1986)

It has been proved that even though the parameter of cell size has a little influence on a KDE hotspot mapping for predicting spatial patterns of crime, the smaller the value, the better the visualization of maps. Similarly, for the bandwidth parameter, which does affect the outputs of KDE, the smaller its value, the less the degrading of maps. Finally, the option of adopting the default setting values for these two parameters, is a quite popular approach among most users of KDE for hotspot mapping (Chainey, 2013).

Regarding the third input parameter, there are different types of kernel functions that can be used as input in the implementation of the KDE, such as Gaussian, Epanechnikov, Triangle, etc. (Epanechnikov, 1969).

The following graph (fig. 23) illustrates the Gaussian and the Quadratic model curves. In this figure, the selected sample, provided by Chen and Guo (2016), has bandwidth (h) equal to 1, mean equal to 0 and a variance equal to 0.4 for the Gaussian function and 2.5 for the Quadratic one. What is also different between these two models is that in the Gaussian all the observations are included with a relative importance while the Quadratic model uses fewer observations related to the search radius (bandwidth). This can be seen at the edges of the two curves; the blue one takes place on an unlimited range whereas the red one has specific limitations (Silverman, 2018).



Source: Chen and Guo, 2016 - edited by author

Figure 23: Gaussian and Quadratic kernel functions

The kernel function used in ArcGIS software and then being adopted for this thesis is based on Quadratic model (Silverman, 2018). Even though Gaussian models are more popular for they assume that all random variables are normal random variables sharing the same mean and variance, the Quadratic kernel function can reduce the computational cost, which is higher in exponential functions, such as the Gaussian one (Chen & Guo, 2016).

3.6.1 Implementation of KDE method for NYC

According to what previously mentioned, the implementation of the KDE method takes place through a model builder[11] (see the graphical representation in Appendix A while the respective file is included in the external folder) in order to produce outputs in an automatic way. To be noted, this process (the duration of the model running) for one crime type and for 52 weeks is a time consuming process (~20 hours). Once the datasets with crime incidents have been imported in Kernel density tool and the number of crimes (for each crime type) is set as the population, the selection of defaults parameters and the whole study area as boundary give the first outputs (densities). These are classified into five Quantile classes (Chainey et al., 2008). The next figure (fig. 24) illustrates in brief this procedure.

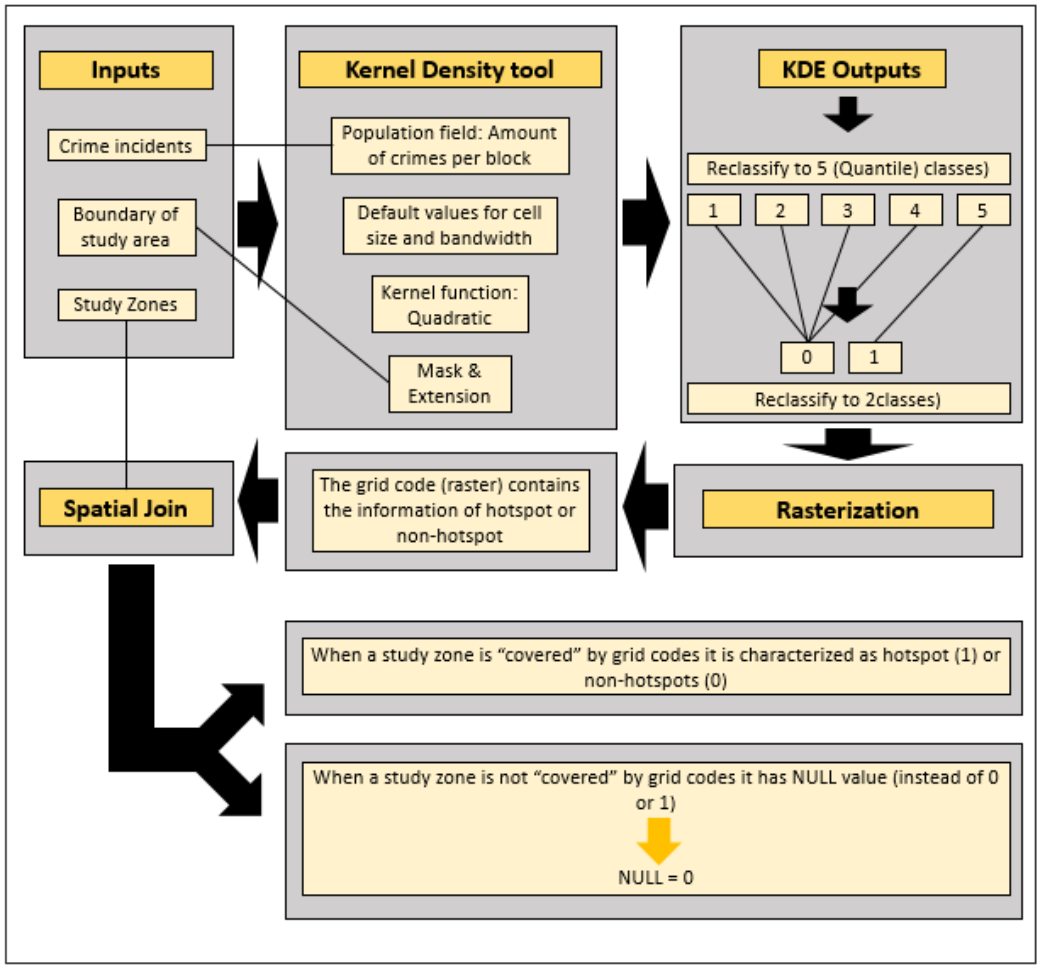


Figure 24: Overview of the implementation of the KDE method

[11]Model Builder available in ArcGIS online by searching "Forecasting crime incidents with KDE": https://uni-utrecht.maps.arcgis.com/home/content.html?view=table&sortOrder=desc&sortField=modified&folder=ded2a011b47a4ffda26acb84f87e0cb4#content

As it can be seen in the above figure, the five Quantile classes are grouped into two classes; non-hotspot (the first four of the five Quantile classes) and hotspot (the fifth class) and they refer to the full study area. This is transformed to raster information (grid codes, see figure 25) in order to continue with some statistical calculations. By spatial joining this raster information with the study zones and computing the mean of values of grid codes per study zone, the latter either characterized as hotspot (1) or non-hotspots (0) or they get "NULL" values. The "NULL" values case occurs because of two reasons: a) the study zones' geometry (schema) is not "covered" by any grid code, for example when the polygons have very small size and b) the polygons are coastal and the majority of the grid codes are not over the land (polygon) but on other areas (e.g. sea or river). In this case, with some tests, it was proved that when the values "NULL" are considered as non-hotspots (0) the forecasting is closer to the reality. The below figure (fig. 25) illustrates these two causes in a zoom area when census blocks are used as study zones.

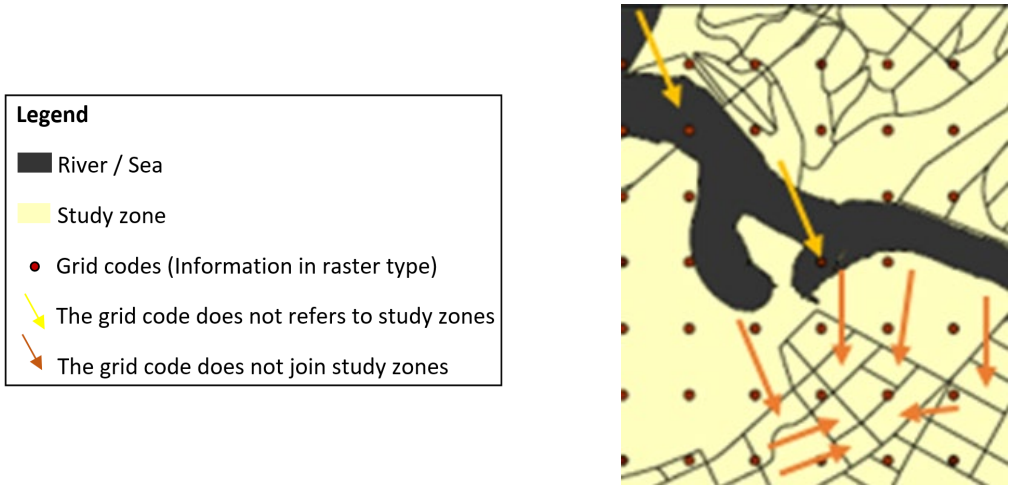


Figure 25: Example of having NULL values in KDE outputs

3.7 Implementation of the Proposed (STARMA) Method

By combining the limited literature and the description of 'starma' package from R-studio, it is concluded that this method consists of seven main steps: a) input datasets, b) creation of neighborhood, c) creation of weight matrices for neighbors, d) simulation, e) identification of process, f) process estimation, and g) diagnose the process. However, in this thesis topic only the three first steps are followed, based on the use of the ready model package and with using its outputs (model residuals) in the same way like other methods (see Appendix B).

In the first step, the working datasets (crime incidents per study zone and per week with a simple geometry feature and study zones as polygons) should be imported in order to get not only the data values but also their geometry features.

In the second step, based on the geometry of study zones (polygons) a type of neighborhood is created, according to a specific search radius in order to find neighbors for each polygon. A first estimation of the search radius could be done by observing the distance of polygons in maps, while the package requires more than one neighbors around the examined polygon in order to work. It should be mentioned that the higher the search radius, the more the neighbors for each polygon and thus the bigger size of the file. On the other hand, if the search radius is very low there is the danger to not get neighbors for some polygons (study zones), which leads to their deletion and therefore to smaller study area. Therefore, a balance should be kept between the value of search radius and the desirable number of neighbors for each polygon (study zone). The following figure (fig. 26) illustrates the positives (with green) and negatives (with red) when low or high radius is used.

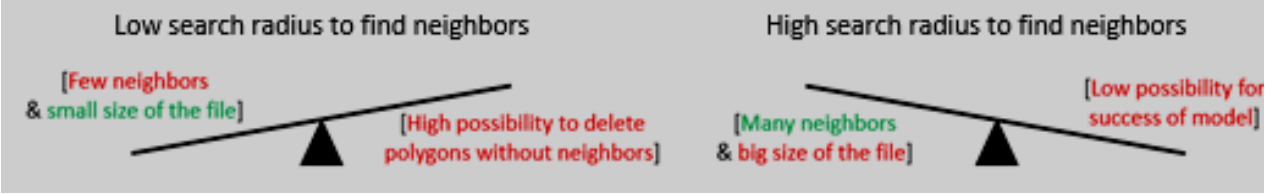


Figure 26: Search radius to find neighbours – positives and negatives for low and high search radius

The third step is strongly associated with the previous one and according to Cheysson (2016) there are three different weight matrices that could be applied for neighbor polygons (study zones) and each of them with different features based on different considerations. The first type of weight that neighbors can have is the "dlist". This type is based on the 100 km distance of neighbors measured from the centroids of polygons. According to Zhang (2012) this weight list is useful for spatial diffusive process (e.g. pollution, temperature, etc.) and it is impacted by the position of the centroids in the study areas. However, in this way big polygons (study zones) isolated and thus they have less impact to their neighbors. The second type is the "klist" which is based on the distance of their centroids. According to Zhang (2012), this weight list is the improved version of the 'dlist' because it controls the lower and the highest possible search radius and the big polygons (study zones) are not isolated. The last type is the "wlist" which presupposes that polygons (study zones) are considered neighbors if they share a

border. According to Zhang (2012) apart from the consideration of common border, this weight list is not based on centroids of polygons and therefore it can be more easily interpreted when modelling relationships between study zones.

Regardless of the type of weight list that is used, the following figure (fig. 27) illustrates the neighbor orders in an ideal environmental. For instance, in the original code, four 'klist' orders of neighbors are created.

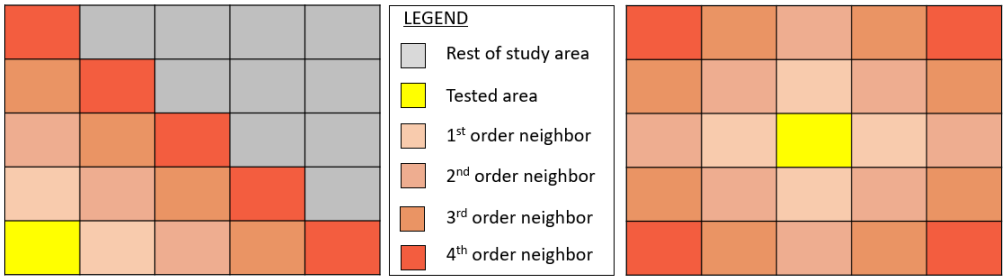


Figure 27: Different orders of neighbours in an ideal environment

Theoretically, in the fourth step, the model is simulated following the equation (9) for specific temporal lags. At the same time, an error parameter, which is in line as for its matrix dimensions with other parameters (i.e. weights), is used. The original code of package is based on the 'klist' weight matrices for neighbor and it creates a matrix with random values for the error parameter which follow the standard normal distribution (mean = 0, standard deviation =1). This means that values of error parameter follows the Gaussian distribution (Cheysson, 2016).

In the fifth step, the model, based on the results from its simulation and the weight neighbour lists, is examined regarding the temporal autocorrelation and partial autocorrelation. The maximum number of t lags can be defined while the default option shows several number of t lags (Cheysson, 2016).

In the pre-last step, the AR (the number of lags of the testing variable or 'φ'-phi-) or / and MA (the number of lags of the error term or 'θ'-theta-) parameters are estimated. After that, a summary of the residuals of the model is available (Cheysson, 2016).

The last step includes the space-time series non correlation test accompanied by X_square test and the relative p-value. This test is an extension of Box-Pierce test statistic in order to reject or not the hypothesis of non-correlation since its temporal and spatial lags and also the degrees of freedom (which should be set equal to the number of parameters of the model) have been set (Cheysson, 2016).

3.7.1 Implementation of STARMA method for NYC

Even though the preferable study zones to implement the STARMA method according to the previous chapter is the census blocks, the effort to do that was not completed due to the high dimensions of matrices for STARMA processes, for the reasons which explained in the subsection 3.2.4. Therefore, like occurred in the other two methods, the STARMA also refers to the zip-code areas as study zones.

The weight that the first order neighbors of study zones affects the forecasting is based on "klist" type for two main reasons. The first one is that there is no limit in the search radius, like the "dlist". So, the selected radius to search for the closest neighbors is 30,000 ft; if it has smaller radius than this, a few study zones do not have neighbor polygons, which leads to their isolation. The second reason for this weight type is that it is not based on sides of polygons but in their centroids. In the case of using the sides there was again the possibility of isolation since two main parts (boroughs) of the whole study area (NYC) are isolated from the rest of the area.

The STARMA method is implemented based on four different experiments regarding the parameters "autoregressive" (AR) (how many lags of the testing variable there are) and "moving average" (MA) (how many lags of the error term are estimated). In the first experiment both of the parameters 'AR' and 'MA' are equal to 1 in order to implement a default model. In the experiment 2, 'AR' is equal to 2 and 'MA' is equal to 1. In the experiment 3, these parameters have inverse values and in the last experiment both of these parameters are equal to 2.

For each experiment and for each type crime, additional parameters are implemented and examined; different number of neighbors and different search radius to find them. For every selected numbers of neighbors, various search radius are used. Every time the first radius to be used is the lowest possible radius followed by an average and a high radius. The lowest radius is the one within which only the exact number of examined neighbors are found (15,800 ft for two neighbors, 16,200 ft for three and four neighbors). The "high" radius is a radius within which many neighbors can be found (20,000 ft for all cases). To be noted, that the lowest radius for the case of two neighbors ("low") cannot be used for the cases of both three and four neighbors. So, the lowest radius for the cases of three and four neighbors is the "average" used for the case of two neighbors.

Table 5: Comparison of four experiments of STARMA for different parameters (type all)

Number of neighbors	Search radius	Experiment 1 (AR=1, MA=1)		Experiment 2 (AR=2, MA=1)		Experiment 3 (AR=1, MA=2)		Experiment 4 (AR=2, MA=2)	
		PAI (weekly average)	Hit Rate (weekly average)	PAI (weekly average)	Hit Rate (weekly average)	PAI (weekly average)	Hit Rate (weekly average)	PAI (weekly average)	Hit Rate (weekly average)
N=2	High	2.419	58.78%	2.408	58.79%	2.422	58.85%	2.422	58.83%
	Average	2.414	58.76%	2.379	58.8%	2.42	58.86%	2.419	58.78%
	Low	2.419	58.79%	2.414	58.85%	2.423	58.87%	2.419	58.78%
N=3	High	2.417	58.78%	2.222	58.33%	2.421	58.86%	2.415	58.79%
	Average	2.423	58.79%	2.404	58.33%	2.425	58.86%	2.427	58.78%
N=4	High	2.423	58.79%	2.346	58.69%	**2.43**	**58.85%**	2.427	58.8%
	Average	2.416	58.7%	2.107	58.24%	2.415	58.82%	2.425	58.8%

As it is shown is the above table, the differences among the results (PAI, Hit rates) are insignificant. However, it is obvious that the experiments 3 and 4 show, in average, higher performance (higher values of PAI and Hit rate) compared to the experiment 1 and 2. The highest performance of STARMA is observed for the third experiment when the higher the number of neighbors and the high search radius are used.

4 Results & Discussion

This chapter shows the results of the three above discussed methods about forecasting crimes. The results are presented, per method for all the three crime types examined, in tables and graphs for the selected metrics and additional information. To be noted that even though two metrics are calculated (PAI, Hit rate), the PAI metric is considered more appropriated for evaluating these forecasting methods (section 2.7). In the last section, the results concerning the comparison amongst the three methods are discussed.

4.1 Results of the implementation of the naïve Baseline method

The forecasting of the naïve baseline method is based on the real crime incidents that occurred the week before, without any processing. Therefore, the threshold value that defines a study zone as a hotspot or non-hotspot, is actually a specific number of crimes every time. The table 6 below illustrates the weekly average of threshold values and number of hotspot study zones for each type of crime for a study area of 248 study zones.

Table 6: Average of threshold values and number of hotspots for the Baseline method

Crime Type	Minimum value	Average Value (weekly average)	Maximum value	Threshold value (weekly average)	Number of Hotspot study zones (weekly average)
All	0	131.4	239	45	51
Property	0	56.8	71	10	53
Violent	0	45.7	72	11	52

According to the above table, the number of hotspot study zones closes range amongst the three crime types showing that 50 of 248 study zones (almost 1/5) are hotspots in weekly average. In addition, this means that the majority of weekly observations have low number of crime incidents.

The table below (table 7) illustrates the average of the weekly PAI values for each crime type, which is between 2.3 and 2.7. Evaluating these results, the naïve baseline method seems to perform well. More particularly, for each crime type the average percentage of

the hit rate is around 60% while the average ratio of size of hotspots regarding the study area is around 25%. To be noted that for each crime type the standard errors and the average of Mean Absolute Deviations (MAD) have been calculated.

Table 7: Results of the naïve Baseline method

Crime types	Hit rate			Hotspots' area / size of study area (average percentage)	PAI		
	Weekly Average of Hit rates	Standard Errors	Mean Absolute Deviation (average)		Weekly average of PAI	Standard Errors	Mean Absolute Deviation (average)
All	59.07%	0.008	0.044	24.47%	2.436	0.02	0.123
Property	64.34%	0.007	0.037	24.27%	2.692	0.01	0.27
Violent	60.98%	0.007	0.039	26.55%	2.318	0.03	0.13

According to the values of table 7, the forecasting for 'property' crimes presents higher PAI values than 'all' crime type and even more than the 'violent' crimes while the last two types range quite close. The average of standard errors gives the lowest value to 'property' type whereas this type has the highest mean absolute deviation. This shows that the forecasting for 'property' type shows a low robustness of the method.

The following graph (fig. 28) summarizes for type 'all' the weekly information for both PAI, Hit rates and also the size ratio between the whole study area and the hotspots' area.

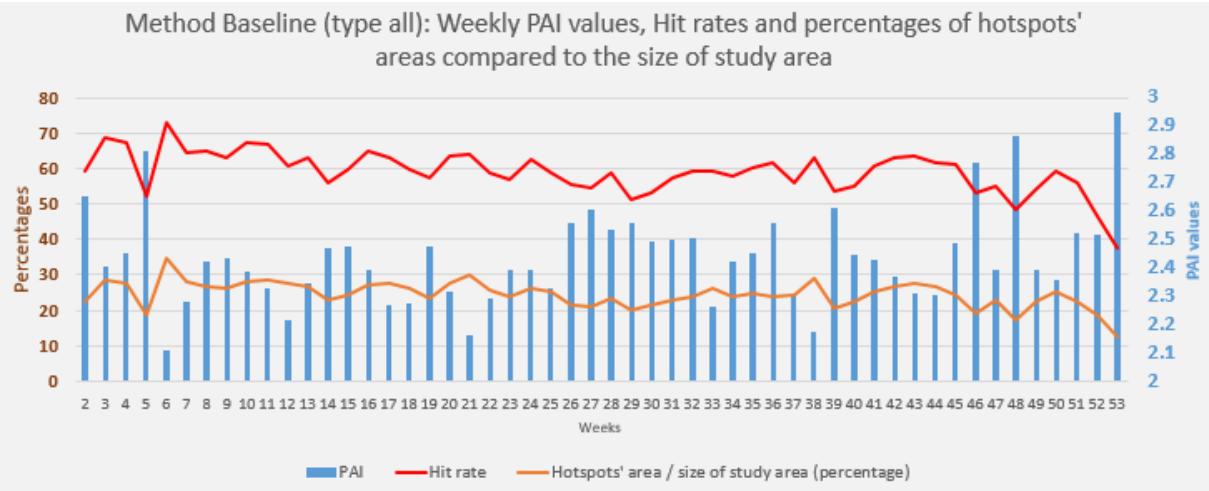


Figure 28: Weekly comparison of PAI values, Hit rates and ratio of areas for the Baseline method and type all

According to the above graph, the PAI metric ranges between 2 and 3 for all the study period (including the last week) and more particularly **from 2.106** (6th week) to 2.863 (48th week) or **to 2.947** (53th week). The hit rate (PAI's numerator) ranges **from 37.40%** (53th week) or from 46.6% (52th week) **to 73.13%** (6th week). For these weeks, the corresponding values for the ratio of the areas (orange line), which represents the denominator of PAI, are **12.69%**, 18.54% and **34.73%**.

For type 'all', the two factors that define the PAI metric (hit rate and ratio of the areas) present similar behavior throughout the whole study period. It is observed that when both of the factors have high values the PAI value is low and vice versa.

On the other hand, the types of 'property' and 'violent' crime include several (weekly) observations in which the factors that define the PAI values do not follow the same trend.

4.2 Results of the implementation of the conventional KDE method

For the conventional KDE method, the threshold values that define a study zone as hotspot or non-hotspot are based on the (kernel) densities (square miles). The table 8 below illustrates for each type crime, the minimum, the maximum and in a weekly average the threshold values, the average of observations (densities) and the number of hotspot study zones for the whole study area (248 zones).

Table 8: Average of threshold values and number of hotspots for the KDE method

Crime Type	Minimum value	Average Value (weekly average)	Maximum value	Threshold value (weekly average)	Number of Hotspot study zones (weekly average)
All	0	183.93	257.24	31.89	60
Property	0	73.41	108.7	6.45	66
Violent	0	29.84	41.66	7.27	61

According to the above table, with 60 – 66 to 248 study zones being hotspots, the KDE method forecasts about ¼ of the study area as hotspot. Even though there is a supremacy of type 'all' regarding the average and maximum values, it forecasts the smallest number of hotspot polygons compared to the other two crime types.

Table 9 shows the performance of the KDE method based on the hit rate and the PAI metric values. In addition, similarly to the previous method, the ratio of between the area of hotspots and the whole study area is calculated.

Table 9: Results of the conventional KDE method

Crime types	Hit rate			Hotspots' area / size of study area (average percentage)	PAI		
	Weekly average of Hit rates	Standard Errors	Mean Absolute Deviation		Weekly average of PAI	Standard Errors	Mean Absolute Deviation
All	39.39%	0.005	0.03	16.33%	2.427	0.03	0.137
Property	50.04%	0.006	0.03	16.28%	3.103	0.01	0.223
Violent	37.48%	0.005	0.03	19.44%	1.938	0.02	0.101

As it can be seen in the above table, only for the crime 'property' the percentage of Hit rate index is around 50% while for the other two types is lower than 40%. The PAI average values show that the 'property' type, which has the highest metric values, has the highest mean absolute deviation. All in all, this method shows higher performance for the 'property' type despite its high MAD value.

The following graph (fig. 29) illustrates for the KDE method and type 'all', the weekly information for both PAI, Hit rates and also the size ratio between the whole study area and the hotspots' area.

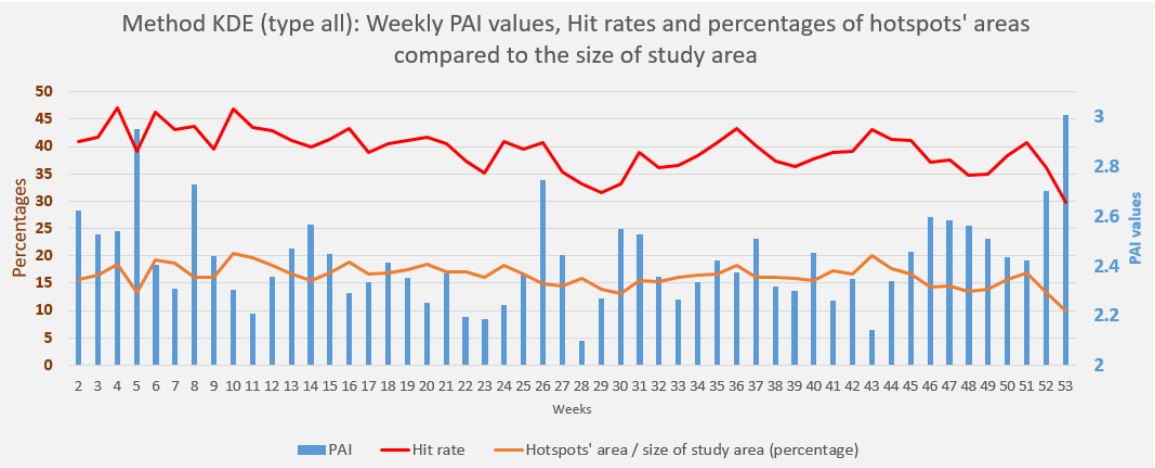


Figure 29: Weekly comparison of PAI values, Hit rates and ratio of areas for the KDE method and all type

According to the above graph, both of the Hit rate and the ratio of the areas' sizes present a fluctuated trend. Since PAI is influenced by both these two values, this explains why its **performance** is quite erratic. As for the type 'all', the nominator of the PAI having an average **39.39%** ranges between 29.72% (week 53) or 31.62% (week 29) and 47.03% (week 4). The corresponding values for the denominator of the PAI, which has average **16.33%**, for these weeks are 9.88% (the lowest), 13.95% (the closest to the second lowest) and 18.52% (very close to the highest 20.37%).

For the property type, the ratio of areas' sizes seem to have similar values, with an average **16.28%** while the hit rates present quite higher values **50.04%**. For violent crimes, the respective percentages are the highest **19.44%** and **37.48%**.

4.3 Results of the implementation of the proposed STARMA method

For the proposed STARMA method the threshold values that define a study zone as hotspot or non-hotspot are based on the sum of the residuals of the model and the normalized real data. The table 10 below illustrates for each type of crime, the minimum, and the maximum and in a weekly average the threshold values, the average of observations (probabilities) and the quantity of the hotspot study zones amongst the total of 248.

Table 10: Average of threshold values and number of hotspots for the KDE method

Crime Type	Minimum value	Average Value (weekly average)	Maximum value	Threshold value (weekly average)	Number of Hotspot study zones (weekly average)
All	-1.15	4.90	12.49	0.8	50
Property	-1.42	7.56	10.53	0.61	50
Violent	-1.63	6.67	12.31	0.75	50

According to the above values, for each crime type, there is equal average number of hotspot study zones, which is around the 1/5 of the study area and the three threshold values seem to not present significant differences in the ranges (minimum – threshold, threshold – maximum) for all the three crime types.

Table 11 presents the metric values with their standard errors and MAD of the STARMA method and also the ratio of areas. The type 'property' is again the crime type

that the method performs better, based on the two metrics. For the two other crime types the PAI values are quite close but the Hit rates present a great difference.

Table 11: Results of the proposed STARMA method

Crime types	Hit rate			Hotspots' area / size of study area (average percentage)	PAI		
	Weekly average of Hit rates	Standard Errors	Mean Absolute Deviation		Weekly average of PAI	Standard Errors	Mean Absolute Deviation
All	58.85%	0.0004	0.02	24.16%	2.43	0.036	0.153
Property	66.74%	0.0005	0.02	23.69%	2.853	0.042	0.242
Violent	64.38%	0.0004	0.02	26.44%	2.446	0.03	0.158

The below graph (fig. 30) illustrates for this proposed method and type 'all' the weekly PAI values, the Hit rates and the ratio of hotspots' areas compared to the study area size. Apart from the last week, because there are only few observations and hence both of the metrics are significantly low, the PAI metric is high and with consistent values.

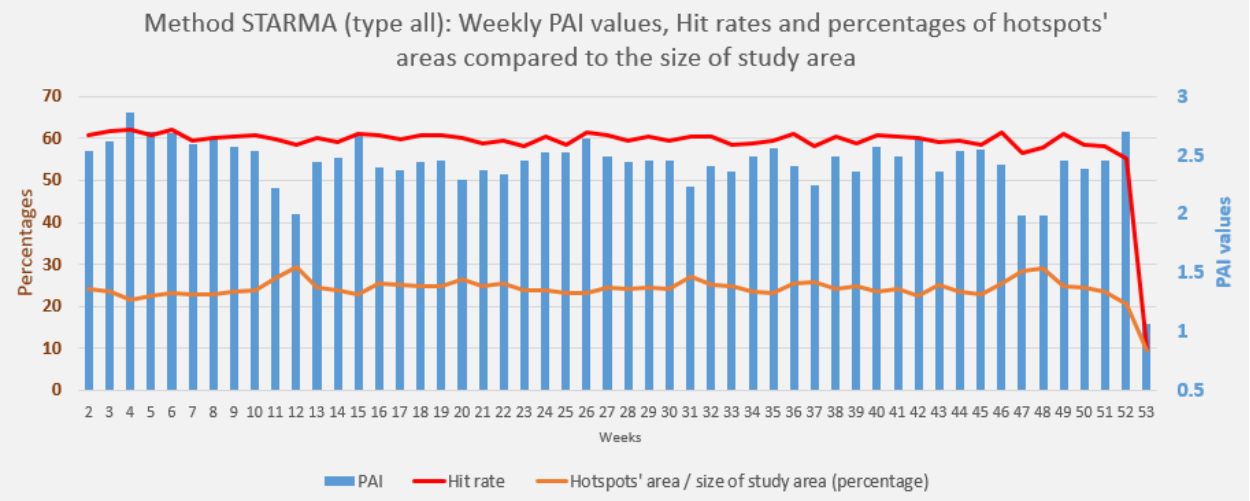


Figure 30: Weekly comparison of PAI values, Hit rates and ratio of areas for the STARMA method and all type

4.4 *Discussion*

This thesis mainly builds on the Shoesmith (2013) study, where an ST-AR model is compared with advanced autoregressive methods (Vector auto-regressions and aggregate univariate models) regarding the forecasting crimes for different types of crimes. Adopting the main idea and same crime types of Shoesmith's study, this thesis

studies an ST-AR method comparing it with simple methods for one spatial resolution (scale). As such, it contributes to the existing literature by highlighting the reliability of the STARMA method to be used for crime forecasting and the importance of taking into consideration both space and time parameters (something that does not apply for the conventional methods).

Based on the results of each applied method, discussed in the previous sections, following can be summarized. For all the three methods the classification of values of crime incidents is done in the same way. However, the difference lies on how these values are represented. For example, while for the naïve Baseline method crime incidents are used without any processing, for the conventional KDE the values are being processed and their densities are used. For the proposed STARMA method, the values are also processed but in a different way, based on the model's residuals and the normalization of the original data. This means that the **threshold values** that are used for each method are based on using different form of input crime data and thus they are not comparable as values. However, a comparable pattern among these methods could be seen by comparing the number of hotspot zones that they define. More particularly, the STARMA and the naïve baseline method forecast the 1/5 of the study area as hotspot whereas for the KDE method the forecast is ¼ of the study area.

Regarding the performance of each method per crime type, what method performs better is based on multiple criteria, such as PAI and Hit rates. For **crime type 'all'**, based on both PAI and Hit rate metrics, the method with the highest performance is the naïve Baseline method. The performance of the STARMA method provides very close values not only for the PAI and Hit rate metrics but also for ratio of the areas (hotspots / study area). According to the above respective graphs (fig. 28 & 30), for type 'all', the STARMA presents a higher level of weekly PAI values (despite the high standard errors and MAD values) while the naïve Baseline method presents a high average PAI due to some high weekly performances. For **crime type 'property'**, based on the PAI metric, the method with the highest performance is the KDE while based on the Hit rate, the STARMA excels. This can be explained due to the high ratio between hotspots' area and the size of the study area for the STARMA method. Although the STARMA also presents higher Hit rate, the denominator of PAI metric (ratio of areas) affects more this metric. In addition to that, the STARMA method again presents higher values as for its standards errors and MAD. For the **crime type 'violent'**, the STARMA

method performs better for both the PAI and Hit rate metrics. The naïve Baseline method has the second performance with significant results. However, its lower Hit rates and its higher ratio of areas seem that subdue its PAI results. What should be noted is the high similarity of their standard errors and MAD values.

In order to compare the performance of these three methods for a same week, three **maps** (see Appendix D) are created for crime type 'all'. The provided information across the three methods refers to the week 47, when, based on the PAI metric the STARMA presents low performance while the naïve Baseline method higher and the KDE the highest. Furthermore, for both the three maps, the information of number of crime incidents is taken into account for each study zone. For this week, the threshold values for the naïve Baseline method is 45, for the KDE 29.63 and for the STARMA 0.385. By dividing the number of real crimes into four classes (0 – 11, 11 – 32, 32 – 58 and 58 – 98), it is observed that the Baseline method forecasts as hotspots all the areas that have the maximum number of crimes, several zones with the third class of crimes and none of the two other classes. The KDE method defines as hotspots the zones that have some crime incidents (low number) and not all the zones that belong to the highest crime class. Regarding the STARMA method, not only it defines all the study zones of highest crime numbers as hotspots, like the Baseline, it also forecasts several study zones with lower number of crimes. Finally, for the examined week the STARMA method shows more spiky distribution of hotspots than the Baseline and even more than the KDE.

5 Conclusions

In this chapter, research questions which were reported in the first chapter are answered, taking into account the aforementioned results. The second section refers to the limitations of this study and the last section to the research directions.

5.1 Answers to research questions

RQ 1.1. *How effective is an ST-AR method in predicting crimes in space and time?*

The results of the proposed ST-AR method (STARMA) showed significant potential in forecasting crimes. Not only the hit rate values, which range from 58% to 67%, but also the PAI values (between 2.4 and 2.9) prove its effectiveness. The exploration of different experiments and by considering different parameters showed that this model presents success in different versions. Without having extreme different results among all the examined cases, when both space and time parameters are considered, the proposed method shows a credible behavior in forecasting.

RQ 1.2. *Are the predictive results sensitive to the parameterization of space and time?*

Taking into account different space and time parameters some slight differences are observed. More particularly, the examination of how the results change in different space parameters, different number of neighbors are tested. However, the parameter of space is quite complex since it is strongly associated with the search radius. The method provides various results among different types of crimes and different experiments. Regarding the time parameter, this is also associated with the results, according to the examined experiments. For example, the scenarios that use only one temporal lag for both the testing variables and error terms, the model presents low results. On the other hand, by using two temporal lags there is higher performance.

Therefore, both the parameters of space and time affect the performance of the model. When there is increased use of lags (time and space) in most of the cases the model presents higher results than using lesser spatial and temporal lags.

RQ 1.3. *How do predictive results vary among different types of crime?*

For all the examined experiments of the STARMA method the results of PAI values is not the same for all the three crime types. More particularly, the type of 'property' crimes for all the cases excels the other two types in a significant level. On the other hand, the other two crime types ('all' and 'violent') present quite similar results.

RQ 2.1. *Which methods perform better in terms of the robustness and accuracy of crime forecasts?*

Robustness is based on how close the PAI or the Hit rate values are among weekly forecasting values and this is expressed through both Mean Absolute Deviation (MAD) and the Standard Errors. According to the results, for the PAI, the three methods provide similar values of MAD and Standard Errors. KDE seems to perform better for crime types 'property' and 'violent' and the Baseline for type 'all'. To be noted that the values of MAD for the STARMA method do not significantly differ. For the Hit rate, in terms of robustness, the STARMA method performs better for all crime types because it presents the lowest MAD and standard errors values.

In terms of accuracy, based on the PAI values, for crime type 'all', the method with the highest performance is the naïve Baseline method (difference with STARMA values is marginal). For crime type 'property' the method that performs better is the KDE while for type 'violent' the STARMA excels. Based on the Hit rates, for type 'all' the naïve Baseline is again the one that performs better whereas for the other two crime types the STARMA method shows the highest performance.

RQ 2.2. *Is the ranking of performance of the examined methods consistent among different types of crime?*

According to the results, there is not one method which perform the best among the different types of crimes. For example, for type all the baseline (naïve) method comes first while the KDE last based on both the PAI and Hit rate metrics. Similarly, again based on both the two metrics, for the crime type 'violent' the STARMA seems that provides the best forecasting contrary to the baseline and even more to the KDE. For type 'property', the PAI and Hit rate metrics indicate different methods to have high

performance. All in all, which method performs better every time depends on the crime type and on the using performance metrics.

5.2 Limitation of current study

During this study, some limitations were faced. These were related mostly to the study area and the spatiotemporal resolution. Concerning the study area, the geography of the NYC consisting of many different parts (e.g. small islands, borders of polygons with rivers) affected the search of neighbours (i.e. isolated blocks) and therefore the study of spatial parameter. Another limitation related to spatial resolution was the high dimensionality of matrices by using census blocks making it impossible to allocate the outputs.

Concerning the temporal resolution, if the study was applied using the original temporal resolution (days) there would be the risk of having files or tables with high dimensions and therefore to deal with the same limitation with using census blocks for spatial resolution.

5.3 Research directions

The above limitations can be perceived as opportunities for future work in order to deal with them and evaluate in a more complete way the proposed method. In this direction, it is recommended for further research to expand present study by analyzing more crime types. In this way, the results of STARMA performance will be validated providing a more robust position in the field of forecasting crimes. Moreover, since the STARMA method is a spatiotemporal model, it is proposed to be tested for different spatiotemporal resolutions in order to see its sensitivity. For instance, different temporal resolutions that could be used are days, 2-weeks period or months. Also, different spatial scale for both study area and study zones could be examined. For example, by using grid boundaries instead of administrative ones or higher spatial resolutions, such as census blocks.

REFERENCES

Almanie, T., Mirza, R., & Lor, E. (2015). Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots. *International Journal of Data Mining & Knowledge Management Process*. https://doi.org/10.5121/ijdkp.2015.5401

Anselin, L., Cohen, J., Cook, D., Gorr, W., & Tita, G. (2000). Spatial analyses of crime. *Criminal Justice*.

Bachner, J. (2013). Predictive Policing. Preventing Crime with Data and Analytics. *Improving Performance Series*.

Belesiotis, A., Papadakis, G., & Skoutas, D. (2018). Tissue Sectioning and Fluorescence Staining for Stereology Download. *ACM Transactions on Spatial Algorithms and Systems*, *3*(4), 1–31. https://doi.org/10.1145/3190345

Bharati, A., & Sarvanaguru, D. (2018). Crime Prediction and Analysis Using Machine Learning. *International Research Journal of Engineering and Technology (IRJET)*.

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014). *Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data*. https://doi.org/10.1145/2663204.2663254

Bowen, D. A., Mercer Kollar, L. M., Wu, D. T., Fraser, D. A., Flood, C. E., Moore, J. C., … Sumner, S. A. (2018). Ability of crime, demographic and business data to forecast areas of increased violence. *International Journal of Injury Control and Safety Promotion*. https://doi.org/10.1080/17457300.2018.1467461

Bowers, K. J., Johnson, S. D., & Pease, K. (2004). Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology*. https://doi.org/10.1093/bjc/azh036

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*. https://doi.org/10.1016/S0272-4944(05)80212-9

Brantingham, P. L., & Brantingham, P. J. (2005). Mobility, Notoriety, and Crime: A Study in the Crime Patterns of Urban Nodal Points. *Journal of Environmental Systems*. https://doi.org/10.2190/dthj-ernn-hvcv-6k5t

Cameron, A. C., & Windmeijer, F. A. G. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business and Economic Statistics*. https://doi.org/10.1080/07350015.1996.10524648

Cao, Y., Zhang, L., & Sun, D. (2013). An air traffic prediction model based on kernel density estimation. *Proceedings of the American Control Conference*. https://doi.org/10.1109/acc.2013.6580831

Catlett, C., Cesario, E., Talia, D., & Vinci, A. (2019). Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. *Pervasive and Mobile Computing*. https://doi.org/10.1016/j.pmcj.2019.01.003

Cesario, E., Catlett, C., & Talia, D. (2016). Forecasting Crimes Using Autoregressive Models. *Proceedings - 2016 IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, DASC 2016, 2016 IEEE 14th International Conference on Pervasive Intelligence and Computing, PICom 2016, 2016 IEEE 2nd International Conference on Big Data*. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.138

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. https://doi.org/10.5194/gmd-7-1247-2014

Chainey, S. (2013). Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime. *BSGLg*.

Chainey, S. P., Curtis-Ham, S. J., Evans, R. M., & Burns, G. J. (2018). Examining the extent to which repeat and near repeat patterns can prevent crime. *Policing*. https://doi.org/10.1108/PIJPSM-12-2016-0172

Chainey, S., Tompson, L., & Uhlig, S. (2008). The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal*. https://doi.org/10.1057/palgrave.sj.8350066

Chen, Y., & Guo, W. A (2016). Local Nonparametric Model For Simultaneous Image Segmentation and Adaptive Smoothing.

Cliff, A. D., & Ord, J. K. (1975). Space-Time Modelling with an Application to Regional Forecasting. *Transactions of the Institute of British Geographers*. https://doi.org/10.2307/621469

Dash, S. K., Safro, I., & Srinivasamurthy, R. S. (2019). Spatio-temporal prediction of crimes using network analytic approach. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*. https://doi.org/10.1109/BigData.2018.8622041

Drawve, G., Moak, S. C., & Berthelot, E. R. (2016a). Predictability of gun crimes: a comparison of hot spot and risk terrain modelling techniques. *Policing and Society*. https://doi.org/10.1080/10439463.2014.942851

Drawve, G., Moak, S. C., & Berthelot, E. R. (2016b). Predictability of gun crimes: a comparison of hot spot and risk terrain modelling techniques. *Policing and Society*. https://doi.org/10.1080/10439463.2014.942851

Duan, P., Mao, G., Zhang, C., & Wang, S. (2016). STARIMA-based traffic prediction with time-varying lags. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. https://doi.org/10.1109/ITSC.2016.7795773

Epanechnikov, V. A. (1969). Non-Parametric Estimation of a Multivariate Probability Density. *Theory of Probability & Its Applications*. https://doi.org/10.1137/1114019

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2014.02.003

Giacomini, R., & Granger, C. W. J. (2004). Aggregation of space-time processes. *Journal of Econometrics*. https://doi.org/10.1016/S0304-4076(03)00132-5

Glasbey, C. A., & Allcroft, D. J. (2008). A spatiotemporal auto-regressive moving average model

for solar radiation. *Journal of the Royal Statistical Society. Series C: Applied Statistics*. https://doi.org/10.1111/j.1467-9876.2007.00617.x

Gonzales, A. R., Schofield, R. B., & Hart, S. V. (2005). Mapping Crime : Understanding Hot Spots. In *National Institute of Justice*.

Gorr, W., & Harries, R. (2003). Introduction to crime forecasting. *International Journal of Forecasting*. https://doi.org/10.1016/S0169-2070(03)00089-X

Gorr, W. L., & Lee, Y. J. (2015). Early Warning System for Temporary Crime Hot Spots. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-014-9223-8

Griffith, D. A. (2010). Modeling spatio-temporal relationships: Retrospect and prospect. *Journal of Geographical Systems*. https://doi.org/10.1007/s10109-010-0120-x

Groff, E. R., & La Vigne, N. G. (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies*. https://doi.org/10.1161/CIR.0b013e31820a55f5

Grover, V., Adderley, R., & Bramer, M. (2007). Review of Current Crime Prediction Techniques. In *Applications and Innovations in Intelligent Systems XIV*. https://doi.org/10.1007/978-1-84628-666-7_19

Grubesic, T. H., & MacK, E. A. (2008). Spatio-temporal interaction of urban crime. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-008-9047-5

Hart, T., & Zandbergen, P. (2014). Kernel density estimation and hotspot mapping: Examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting. *Policing*. https://doi.org/10.1108/PIJPSM-04-2013-0039

ISO. (1994). Accuracy (trueness and precision) of measurement methods and results. *Part 1. General Principles and Definitions*.

Kamarianakis, Y., & Prastacos, P. (2005). Space-time modeling of traffic flow. *Computers and Geosciences*. https://doi.org/10.1016/j.cageo.2004.05.012

Kennedy, L. W., Caplan, J. M., & Piza, E. (2011). Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies. *Journal of Quantitative Criminology*. https://doi.org/10.1007/s10940-010-9126-2

Khalid, S., Shoaib, F., Qian, T., Rui, Y., Bari, A. I., Sajjad, M., … Wang, J. (2018). Network Constrained Spatio-Temporal Hotspot Mapping of Crimes in Faisalabad. *Applied Spatial Analysis and Policy*. https://doi.org/10.1007/s12061-017-9230-x

Kounadi, O., Ristea, A., Araujo, A., & Leitner, M. (2020). A systematic review on spatial crime forecasting. *Crime Science*. https://doi.org/10.1186/s40163-020-00116-7

Kurt, S., & Tunay, K. B. (2015). STARMA Models Estimation with Kalman Filter: The Case of Regional Bank Deposits. *Procedia - Social and Behavioral Sciences*. https://doi.org/10.1016/j.sbspro.2015.06.441

Kutnowski, M. (2017). The ethical dangers and merits of predictive policing. *Journal of Community Safety and Well-Being*. https://doi.org/10.35502/jcswb.36

Li, G., Haining, R., Richardson, S., & Best, N. (2014). Space-time variability in burglary risk: A Bayesian spatio-temporal modelling approach. *Spatial Statistics*. https://doi.org/10.1016/j.spasta.2014.03.006

Lin, Y. L., Yen, M. F., & Yu, L. C. (2018). Grid-based crime prediction using geographical features. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi7080298

Louppe, G. (2014). Understanding Random Forests. *Cornell University Library*.

M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. https://doi.org/10.5121/ijdkp.2015.5201

Martin, R. L., & Oeppen, J. E. (1975). The Identification of Regional Forecasting Models Using Space: Time Correlation Functions. *Transactions of the Institute of British Geographers*. https://doi.org/10.2307/621623

Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2014.01.004

Pawale, P., Bagal, S., Ajabe, S., & Shikalagar, K. (2017). Geo Statistical Approach for Crime Hotspot Detection and Prediction. *International Research Journal of Engineering and Technology*.

Perry, W., McInnis, B., Price, C., Smith, S., & Hollywood, J. (2018). Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations. In *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. https://doi.org/10.7249/rr233

Pfeifer, P. E., & Deutsch, S. J. (1980). A STARIMA model-building procedure with application to description and regional forecasting. *Transactions, Institute of British Geographers*. https://doi.org/10.2307/621846

Pfeifer, Phillip E., & Deutsch, S. J. (1981). Seasonal Space-Time ARIMA Modeling. *Geographical Analysis*. https://doi.org/10.1111/j.1538-4632.1981.tb00720.x

RONCEK, D. W., & MAIER, P. A. (1991). Bars, Blocks and Crimes Revisited: Linking the theory of Routine Activities to the Empiricism of "Hotspots". *Criminology*. https://doi.org/10.1111/j.1745-9125.1991.tb01086.x

Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017). An overview on crime prediction methods. *6th ICT International Student Project Conference: Elevating Community Through ICT, ICT-ISPC 2017*. https://doi.org/10.1109/ICT-ISPC.2017.8075335

SHERMAN, L. W., GARTIN, P. R., & BUERGER, M. E. (1989). Hot spots of Predatory Crime: Routine Activities and the Criminology of Place. *Criminology*. https://doi.org/10.1111/j.1745-9125.1989.tb00862.x

Shoesmith, G. L. (2013). Space–time autoregressive models and forecasting national, regional and state crime rates. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2012.08.002

Silverman, B. W. (2018). Density estimation: For statistics and data analysis. In *Density*

Estimation: For Statistics and Data Analysis. https://doi.org/10.1201/9781315140919

Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., & Abreu, R. (2015). A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. *Proceedings of the 8th International Conference on Educational Data Mining*.

Thakali, L., Kwon, T. J., & Fu, L. (2015). Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. *Journal of Modern Transportation*. https://doi.org/10.1007/s40534-015-0068-0

Torgo, L., & Ribeiro, R. (2009). Precision and recall for regression. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-642-04747-3_26

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*. https://doi.org/10.3354/cr030079

Yu, C. H., Ward, M. W., Morabito, M., & Ding, W. (2011). Crime forecasting using data mining techniques. *Proceedings - IEEE International Conference on Data Mining, ICDM*. https://doi.org/10.1109/ICDMW.2011.56

Zhang, X., Chen, G., Wang, J., Li, M., & Cheng, L. (2019). A GIS-based spatial-temporal autoregressive model for forecasting marine traffic volume of a shipping network. *Scientific Programming*. https://doi.org/10.1155/2019/2345450

Zhuang, Y., Almeida, M., Morabito, M., & Ding, W. (2017). Crime Hot Spot Forecasting: A Recurrent Model with Spatial and Temporal Information. *Proceedings - 2017 IEEE International Conference on Big Knowledge, ICBK 2017*. https://doi.org/10.1109/ICBK.2017.3

APPENDICES

Due to the large content of appendices, the majority of big tables and codes are located at an external file in the submitted folder and here only some of them are presented.

Appendix A: Model builder for KDE – graphical representation

Appendix B: Evaluation of KDE and naïve Baseline methods

```
# The evaluation of the KDE is based on outputs from ArcGIS and more particularly the model builder process.

# The following script is repeated 52 times

#for (XX in 2:53)

kde_trueXX<-as.data.frame(matrix(0,248,7))

for (i in XX:XX) {

 filename <- sprintf("C:\\...all\\allforw%d",i) #this is the path of datasets

 forecastkde <- read.csv(filename, header = TRUE, sep = ",", dec = ".")

 filename2 <- sprintf("C:\\...all\\week%d",i)   #this is the path of datasets

 realkde <- read.csv(filename2, header = TRUE, sep = ",", dec = ".")

 kde_trueXX[,1]<-realkde$XCoord

 kde_trueXX[,2]<-realkde$YCoord

 kde_trueXX[,3]<-realkde$ZIPCODE

 kde_trueXX[,4]<-forecastkde$ZIPCODES_ZONES.SHAPE_AREA

 kde_trueXX[,5]<-realkde[,4]

 kde_trueXX[,6]<-forecastkde[,5]

 kde_trueXX[,7]<-realkde$WK

}

for (i in 1:248){

 kde_trueXX[i,8]<-BZ_def_geo$geometry[i]

}

# After 52 times, all the files are binded

kde_outputs_all<-
rbind(kde_true2,kde_true3,kde_true4,kde_true5,kde_true6,kde_true7,kde_true8,kde_true9,kde_true10,kde_true11,
kde_true12,kde_true13,kde_true14,kde_true15,kde_true16,kde_true17,kde_true18,kde_true19,kde_true20,kde_true21,
kde_true22,kde_true23,kde_true24,kde_true25,kde_true26,kde_true27,kde_true28,kde_true29,kde_true30,kde_true31,
```

```
kde_true32,kde_true33,kde_true34,kde_true35,kde_true36,kde_true37,kde_true38,kde_true39,kde_true40,kde_true41,

kde_true42,kde_true43,kde_true44,kde_true45,kde_true46,kde_true47,kde_true48,kde_true49,kde_true50,kde_true51,
                kde_true52,kde_true53)
# Columns get appropriate names
colnames(kde_outputs_all)[1]<-"X"
colnames(kde_outputs_all)[2]<-"Y"
colnames(kde_outputs_all)[3]<-"ZIPCODE"
colnames(kde_outputs_all)[4]<-"AREA"
colnames(kde_outputs_all)[5]<-"CRIMES"
colnames(kde_outputs_all)[6]<-"FORECAST_HOTSPOT"
colnames(kde_outputs_all)[7]<-"WEEK"
colnames(kde_outputs_all)[8]<-"geometry"
# Assumption that NULL values of KDE outputs are "non-hotspot" areas
sizef <- nrow (kde_outputs_all)
for (k in 1:sizef){
 if (kde_outputs_all[k, 6]=="NULL"){
   kde_outputs_all[k, 6] = 0
 }
}
kde_outputs_all[,6]<- droplevels(kde_outputs_all[,6])
a=0
N=0
n=0
for (i in 1:12896){
 if (kde_outputs_all$WEEK[i]==XX){
   N=N+sum(kde_outputs_all$CRIMES[i])
   if (kde_outputs_all$FORECAST_HOTSPOT[i]==1){
     n=n+sum(kde_outputs_all$CRIMES[i])
```

```
    a=a+sum(kde_outputs_all$AREA[i])
   }
 }
}
# the process is repeated for each type crime------------------------
#-------------------------------------------------------------------------------
# The evaluation of the baseline method is based on output files, which created manually from real data.
# The following script is repeated 52 times
#for (XX in 2:53)
baseline_trueXX<-as.data.frame(matrix(0,248,7))
for (i in 2:2) {
 filename <- sprintf("C:\\...all\\allforw%d",i)
 forecastb <- read.csv(filename, header = TRUE, sep = ",", dec = ".")
 filename2 <- sprintf("C:\\...all\\week%d",i)
 realb <- read.csv(filename2, header = TRUE, sep = ",", dec = ".")
 baseline_trueXX[,1]<-realb$XCoord
 baseline_trueXX[,2]<-realb$YCoord
 baseline_trueXX[,3]<-realb$ZIPCODE
 baseline_trueXX[,4]<-zip_area$ZIPCODES_ZONES.SHAPE_AREA
 baseline_trueXX[,5]<-realb[,4]
 baseline_trueXX[,6]<-forecastb[,4]
 baseline_trueXX[,7]<-realb$WK
 colnames(baseline_trueXX)[1]<-"X"
 colnames(baseline_trueXX)[2]<-"Y"
 colnames(baseline_trueXX)[3]<-"ZIPCODE"
 colnames(baseline_trueXX)[4]<-"AREA"
 colnames(baseline_trueXX)[5]<-"CRIMES"
 colnames(baseline_trueXX)[6]<-"FORECAST_HOTSPOT"
 colnames(baseline_trueXX)[7]<-"WEEK"
}
```

```
# After 52 times, all the files are bind

baseline_outputs_all<-
rbind(baseline_true2,baseline_true3,baseline_true4,baseline_true5,baseline_true6,baseline_true7,base
line_true8,baseline_true9,baseline_true10,baseline_true11,
baseline_true12,baseline_true13,baseline_true14,baseline_true15,baseline_true16,baseline_true17,bas
eline_true18,baseline_true19,baseline_true20,baseline_true21,
baseline_true22,baseline_true23,baseline_true24,baseline_true25,baseline_true26,baseline_true27,bas
eline_true28,baseline_true29,baseline_true30,baseline_true31,

baseline_true32,baseline_true33,baseline_true34,baseline_true35,baseline_true36,baseline_true37,bas
eline_true38,baseline_true39,baseline_true40,baseline_true41,

baseline_true42,baseline_true43,baseline_true44,baseline_true45,baseline_true46,baseline_true47,bas
eline_true48,baseline_true49,baseline_true50,baseline_true51,

                    baseline_true52,baseline_true53)

Percentile_00  = min(baseline_outputs_all[,6])

Percentile_5th  = quantile(baseline_outputs_all[,6], 0.8) #classification based on quantile classes

Percentile_100 = max(baseline_outputs_all[,6])

baseline_outputs_all[,6][baseline_outputs_all[,6] >= Percentile_00 & baseline_outputs_all[,6] <
Percentile_5th]  = 0

baseline_outputs_all[,6][baseline_outputs_all[,6] >= Percentile_5th & baseline_outputs_all[,6] <=
Percentile_100] = 1

k=XX

a_all=0

N_all=0

nall=0

for (i in 1:12896){

 if (baseline_outputs_all$WEEK[i]==k){

   N_all=N_all+sum(baseline_outputs_all$CRIMES[i])

   if (baseline_outputs_all$FORECAST_HOTSPOT[i]==1){

     nall=nall+sum(baseline_outputs_all$CRIMES[i])

     a_all=a_all+sum(baseline_outputs_all$AREA[i])

   }

 }

}

# the process is repeated for each type crime-------------------------
```

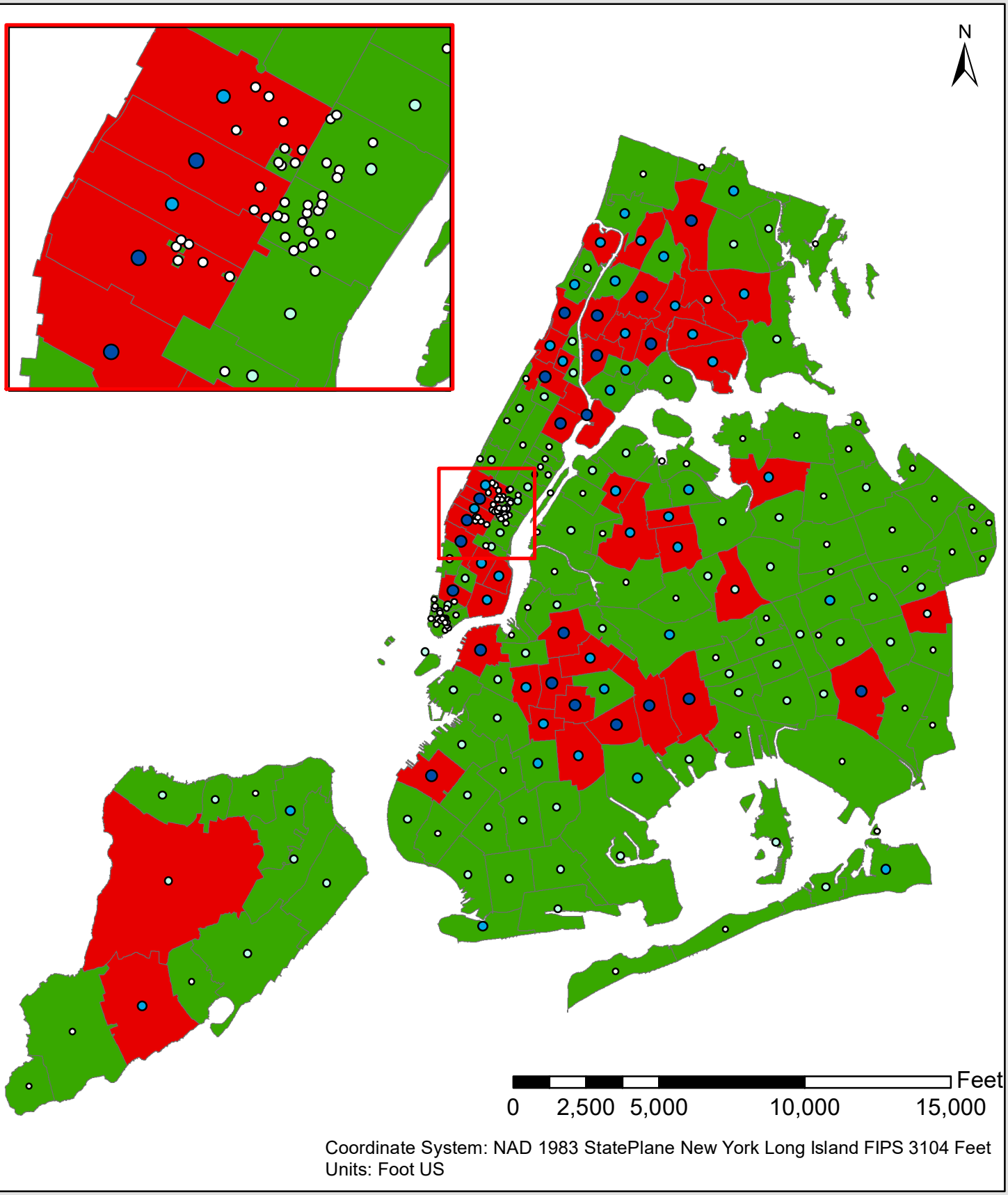
Appendix C: Implementation and evaluation of STARMA method for this thesis

```
# This file contains the implementation of starma based on zip code study zones and its evaluation

# The process is repeated for each crime type:

# Create neighbors based on the study area

knb <- dnearneigh(my_sites_Bz, 0, 15800)  # radius can be changed (cases that examined: r=15800, r=16200 & r=20000)

knb <- nblag(knb, 2) #different number of neighbors are examined (cases that examined: n=2 (only for r=15800), n=3 & n=4)

klist <- list(order0=diag(248), # the number corresponds to the amount of study zones that used

          order1=nb2mat(knb[[1]]),

          order2=nb2mat(knb[[2]]))

         # order3=nb2mat(knb[[3]]))

         # order4=nb2mat(knb[[4]]))

# Normalize the data

all_norm<- stcenter(working_for_Zip_all) #for property type write "working_for_Zip_p" and for violent "working_for_Zip_v"

# Estimate the process

ar<-1 # Different scenarios are examined (ar=1 for scenarios 0 & 2 and ar=2 for scenarios 1 & 3)

ma<-1 # Different scenarios are examined (ma=1 for scenarios 0 & 1 and ma=2 for scenarios 2 & 3)

model <- starma(all_norm, klist, ar, ma)

model

summary(model)

model$residuals #residuals of the model

(max.print = 53) # because of 53 weeks

res_ok_all_0<-model$residuals

suma<-res_ok_all_0 + all_norm #calculation of propabilities based on real data values and their residuals

# The code below refers to its evaluation with PAI and Hit rate metrics

# Classify model

mystar <- data.frame(matrix(NA, nrow = 248, ncol = 53)) # create new empty dataframe

mystar<-as.data.frame(t(suma)) # transpose suma to get weeks as columns and polygons as rows

colnames(mystar) <- paste("week", 1:53, sep = "") #rename
```

```
mystarq<-data.frame(matrix(NA, nrow = 3, ncol = 53)) #making the table to classify values based on quantile classes
for (i in 1:53){
  colnames(mystarq) <- paste("week", 1:53, sep = "")
}
rownames(mystarq)[1]<-"Percentile_00"
rownames(mystarq)[2]<-"Percentile_80" #the fifth class
rownames(mystarq)[3]<-"Percentile_100"
for (i in 1:53){
  mystarq[1,i] = min(mystar[,i]) #it was for suma, but i try for mystar
  mystarq[2,i] = quantile(mystar[,i], 0.8) #get the top 20% of values
  mystarq[3,i] = max(mystar[,i])
}
suma2<-suma #classify the initial outputs to 0 and 1
for (i in 1:53){
  for (j in 1:248){
    if (suma2[i,j] < mystarq[2,i]){
      suma2 [i,j] <- 0
    }
    else if (suma2[i,j] >= mystarq[2,i]){
      suma2 [i,j] <- 1
    }
  }
}
# Create a result file
real_all <- working_for_Zip_all #for property type write "working_for_Zip_p" and for violent "working_for_Zip_v"
colnames(real_all)[1] <- "83" #this information should be reset because the initial format is not numeric
for (i in 1:248){
  if (zip_area$ZIPCODE[i]==colnames(real_all[i])){ # a file with zip codes and their areas has been made "zip_area"
```

```
   real_all[1,i] <- zip_area$ZIPCODES_ZONES.SHAPE_AREA[i]
  }
}
# Calculation of PAI and HIT RATE indexes
Hotspot_star <- data.frame(matrix(0, nrow = 53, ncol = 4))
colnames(Hotspot_star)[1]<-"n" #number of crimes when a study zone is hotspot
colnames(Hotspot_star)[2]<-"N" #total number of crimes in the whole study area
colnames(Hotspot_star)[3]<-"a" #total area of hotspots
colnames(Hotspot_star)[4]<-"A" #total area of the whole study area
for (i in 1:53){
  for (j in 1:248){
    if (suma2[i,j]==1){
      Hotspot_star[i,1] <- Hotspot_star[i,1]+ sum(real_all[i,j])
      Hotspot_star[i,2] <- sum(real_all[i,])
      Hotspot_star[i,3] <- Hotspot_star[i,3]+ sum(real_all[1,j])
      Hotspot_star[i,4] <- sum(real_all[1,])
    }
  }
}
for (p in 1:53){
  Hotspot_star$PAI[p]<- (Hotspot_star[p,1]/Hotspot_star[p,2])/(Hotspot_star[p,3]/Hotspot_star[p,4])
}

Hotspot_star <- cbind(Week = rownames(Hotspot_star), Hotspot_star)
rownames(Hotspot_star) <- 1:nrow(Hotspot_star)
for (i in 1:53){
  Hotspot_star[i,7]<-Hotspot_star[i,2]*100/Hotspot_star[i,3]
}
colnames(Hotspot_star)[7]<-"Hit_rate"
Hotspot_star[1:2,c(2:7)]<-0 # the first two rows (two weeks) should be equal to zero because they are included for forecasting
```

Appendix D: Maps



Legend

Crime Incidents

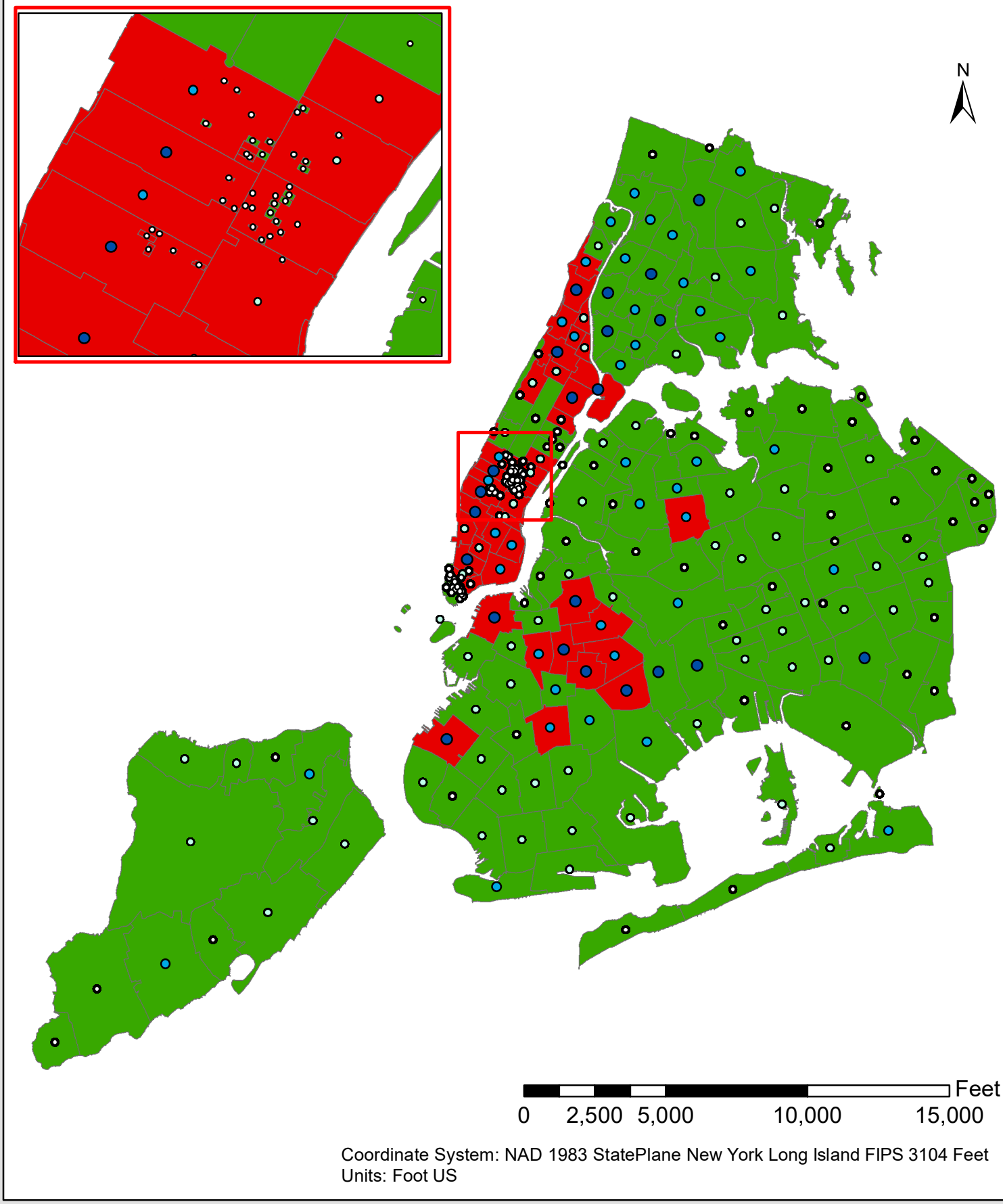
- 0 - 11
- 11 - 32
- 32 - 58
- 58 - 98

Hotspots areas

- 0 (non-hotspot)
- 1 (hotspot)

PAI = 1.989
Hit rate = 56.48%
area of hotspots / size of study area = 28.39%

Map 1: Starma method - type all - week 47



Legend

Crime Incidents

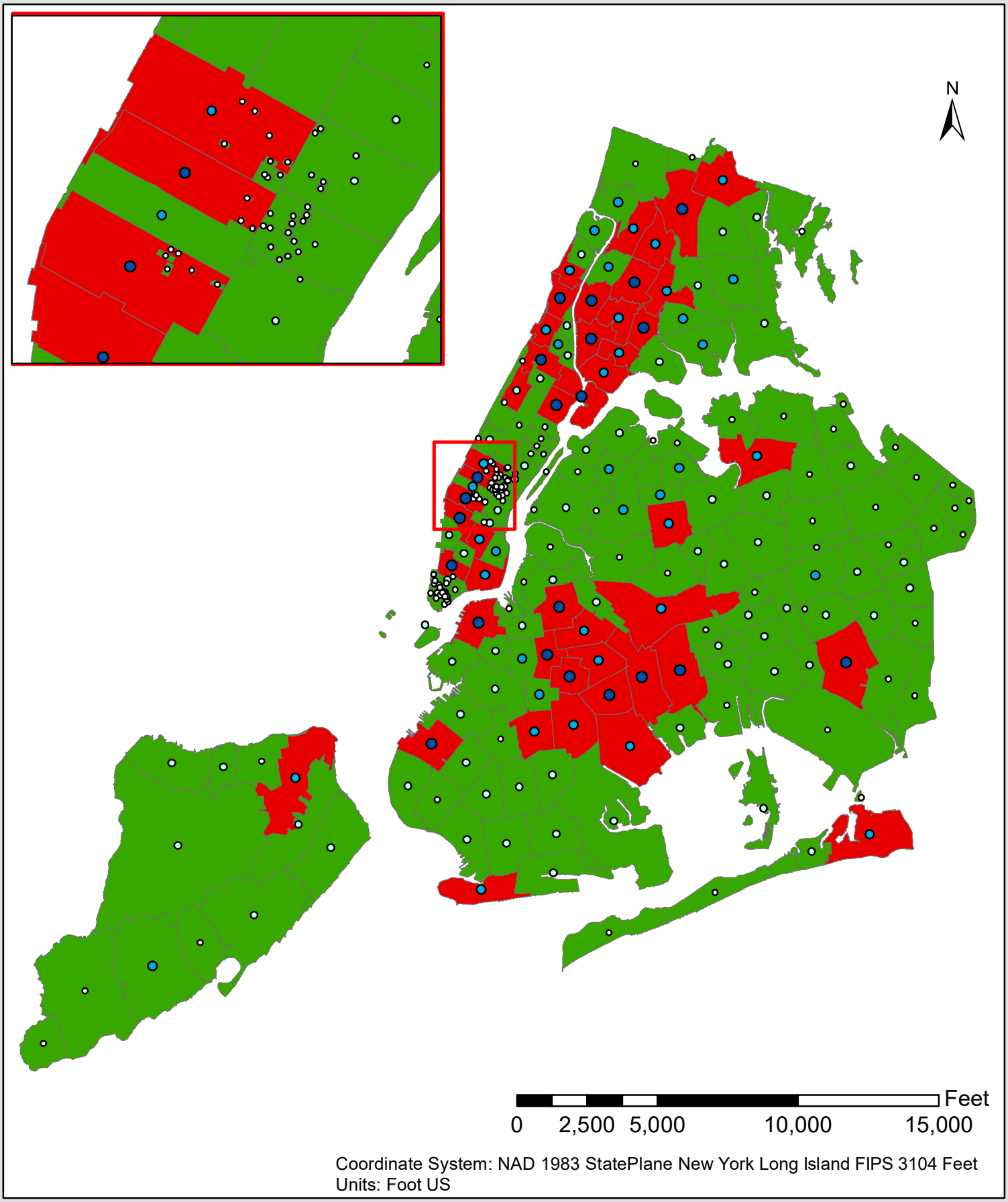
- 0 - 11
- 11 - 32
- 32 - 58
- 58 - 98

Hotspots areas

- 0 (non-hotspot)
- 1 (hotspot)

PAI = 2.58
Hit rate = 37.44%
area of hotspots / size of study area = 14.50%

Map 2: KDE method - type all - week 47



Legend

Crime Incidents

- 0 - 11
- 11 - 32
- 32 - 58
- 58 - 98

Hotspots areas

- 0 (non-hotspot)
- 1 (hotspot)

PAI = 2.391
Hit rate = 54.86%%
area of hotspots / size of study area = 22.94%

Map 3: Naive Baseline method - type all - - week 47