universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Exploring the predictive performance of space-time auto-regressive models in different temporal resolutions“

verfasst von / submitted by

Daniel Larcher BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 856
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Kartographie und Geoinformation
Betreut von / Supervisor:	Ass.-Prof. Dr. Ourania Kounadi, BSc MSc

Contents

List of Figures

List of Tables

Kurzfassung

Die räumliche Verteilung der Kriminalität war Gegenstand einer der ersten wissenschaftlichen Studien in der Kriminologie. Heute bildet die Umweltkriminologie einen wichtigen theoretischen Rahmen in der aktuellen kriminologischen Theorie. Der wesentliche Punkt der Umweltkriminalität ist, dass die Kriminalität nicht zufällig in Raum und Zeit verteilt ist. Die Einbeziehung von Raum und Zeit in die Kriminalitätsvorhersage hat sich als wertvolle Erkenntnis für viele verschiedene Forschungsbereiche erwiesen. In den letzten 30 Jahren haben Fortschritte in der Informationstechnologie, insbesondere geografische Informationssysteme, die Einführung von Kriminalitätsprognosen in den Polizeidienststellen erleichtert. Die Verwendung von Prognoseinstrumenten durch die Polizei ist jedoch heute stark in die Kritik geraten. Die Verwendung voreingenommener Variablen hat dazu geführt, dass bestimmte Bevölkerungsgruppen in den Vereinigten Staaten weiterhin zu Opfern werden. Die größte Herausforderung, vor der die Strafverfolgungsbehörden heute stehen, ist jedoch die effiziente und genaue Analyse der wachsenden Datenmengen. Daher werden Methoden benötigt, die keine verzerrten Daten verwenden, Daten effizient nutzen und genaue Vorhersagen machen.

Eine Methode, die alle diese Anforderungen erfüllt, sind die Modelle des Space-Time Auto-Regressive Moving Average (STARMA). Obwohl sie nachweislich Daten effizient nutzen und ähnliche Methoden übertreffen, sind sie im Bereich der Kriminalitätsvorhersage noch nicht eingehend untersucht worden. Diese Studie zielt darauf ab, die Forschungslücke im Bereich der Kriminalitätsvorhersage zu schließen, indem sie an frühere Studien anknüpft und STARMA-Modelle mit ihrem nicht-räumlichen Gegenstück (ARIMA) in verschiedenen Zeiträumen und Kriminalitätsarten vergleicht. Frei verfügbare Daten und Software wurden verwendet, um die Replikation dieser Analyse zu erleichtern, und der programmierte R-Code wurde veröffentlicht.

In der Studie werden STARMA- und ARIMA-Modelle für drei Arten von Verbrechen (alle, Gewalt- und Eigentumsdelikte) in fünf Zeiträumen (wöchentlich, monatlich, vierteljährlich, halbjährlich und jährlich) erstellt. Die Modelle werden dann mit ihren Fehlermetriken (MSE, RMSE und R2), den abgebildeten Residuen des Modells und den LISA-Karten der Modellresiduen verglichen.

Die wichtigsten Ergebnisse dieser Studie sind, dass STARMA-Modelle besser abschneiden, wenn die Straftaten im Untersuchungsgebiet mindestens ein Moran's I von $0,2$ auf-

weisen. Wenn die Straftaten jedoch zufällig im Raum verteilt sind, schneiden die ARIMA-Modelle besser ab. Die besten STARMA-Modelle wurden für den vierteljährlichen Zeitraum gefunden, und der wichtigste ist der autoregressive Parameter erster Ordnung.

Abstract

The spatial distribution of crime was the subject of one of the first scientific studies in criminology. Today, environmental criminology forms an important theoretical framework in current criminological theory. The essential point of environmental criminology is that crime is not randomly distributed in space and time. The incorporation of space and time into crime prediction has proven to be a valuable insight for many different areas of research. Over the past 30 years, advances in information technology, particularly geographic information systems, have facilitated the implementation of crime forecasting in police departments. However, the use of forecasting tools by police today has come under severe criticism. The use of biased variables has resulted in the continued victimization of certain populations in the United States. However, the greatest challenge facing law enforcement agencies today is how to efficiently and accurately analyze the growing volumes of data. Therefore, methods are needed that do not use biased data, use data efficiently, and make accurate predictions.

Space-Time Auto-Regressive Moving Average (STARMA) models are one method that meets all of these requirements. Although they have been shown to use data efficiently and outperform similar methods, they have not been studied in depth in the field of crime prediction. This study aims to fill the research gap in crime forecasting by following up on previous studies and comparing STARMA models with their non-spatial counterpart (ARIMA) in different time periods and crime types. Freely available data and software were used to facilitate replication of this analysis, and programmed R code was published.

The study builds STARMA and ARIMA models for three types of crimes (all, violent, and property) in five time periods (weekly, monthly, quarterly, semiannual, and annual). The models are then compared with their error metrics (MSE, RMSE, and R2), the mapped residuals of the model, and the LISA maps of the model residuals.

The main findings of this study are that STARMA models perform better when the crimes in the study area have at least a Moran's I of 0.2. However, when the crimes are randomly distributed in space, the ARIMA models perform better. The best STARMA models were found for the quarterly period, and the most important is the first-order autoregressive parameter.

Preface

> Everything is related to everything else, but near things are more related than distant things.
>
> *first law of geography*
> *W. R. Tobler*

The first law of geography formulated by Tobler (1970) was a constant companion during my study. Understanding spatial relationships has always fascinated me, which was the reason I became a geographer. While studying geography, it became clearer to me every day that using geographic information technology to extract knowledge from data and display the results on maps in order to communicate the knowledge was the path I wanted to take. Therefore, the decision to study cartography and geoinformation was the logical step after completing my geography degree.

When I approached my supervisor with some ideas for my master's thesis, she introduced me to a method that predicts crimes based on space and time of past criminal events. The ability to predict something before it happened intrigues me, not to mention the fact that it can be done with a minimal number of variables.

I would like to thank Dr. Kounadi for her introduction to this topic, her constant support, and her repeated motivation throughout the process.

I would also like to thank all the fellow students I met during my studies. It was truly a great experience that I would not want to miss in my life.

A special thanks also goes to all the friends I met in Vienna who helped me grow as a person and who are always there for me when I need them.

I would also like to thank my family, who never questioned how long the path I was on would last.

And finally, I would like to thank Eva for her caring support during this challenging period.

Chapter 1

Introduction

This chapter serves as an introduction to the thesis. First, the background for the choice of the topic of this thesis is presented. A brief historical overview is also given in the next section. Then, a section explains the problem and the reason why the analysis conducted is necessary. Then, the scope of the research is outlined. The scientific objectives and the research questions that this thesis aims to answer are provided. In the same section, the limitations of the research are also established. Finally, the related work in the context of this thesis is presented.

1.1 Background

Geography has become progressively more relevant in law enforcement and crime prevention (J. Cohen et al., 2007), although the histories of criminology and geography are closely intertwined. The history of criminology dates back to the early 19th century in France, where the first studies of the spatial distribution of criminals were conducted after the French government published data on crime and the prison system (Burinsma & Johnson, 2018). Based on this data, André-Michel Guerry and Adriano Balbi created the first crime map showing the relationship between education level and two types of crime: Violent and property crimes (Hunt, 2019).

Today environmental criminology is a prominent theoretical framework in current criminological theory (Andresen, 2020). While mainstream criminology is primarily concerned with why a crime is committed, the spatial distribution of criminal events is the primary concern of environmental criminology (Burinsma & Johnson, 2018). The Crime Pattern Theory serves as a metatheory of environmental criminology (P. J. Brantingham & Brantingham, 2008), summarizing the three major theories in the field:

- the **Routine Activity Theory** by L. E. Cohen and Felson (1979) addresses differences or shifts in the social environment that are indicative of changes in crime rates;

- the **Geometry Of Crime** by P. J. Brantingham and Brantingham (1981) deals with the constructed environment and how it shapes the geographic pattern of crime;
- and the **Rational Choice Theory** by Cornish and Clarke (1986) looks at with the cognitive environment that determines the decision-making processes of potential criminals (Andresen, 2020).

These theories combined under the Crime Pattern Theory by P. J. Brantingham and Brantingham (1984) give an explanation on why crimes occur in certain places at specific times. Moreover, the theories lead to the notion that place, rather than people, is the critical element of crime (Hunt, 2019). Therefore, it should be possible to predict crime using space as a variable (P. J. Brantingham & Brantingham, 1981).

The theories of environmental criminology have been scientifically tested, and in the last decade various research fields have made great strides in developing methods for spatiotemporal prediction of crime (Kounadi et al., 2020). The integration of space and time information in crime forecast can be used as valuable knowledge for many purposes (Shamsuddin et al., 2017). For example, methods that incorporate space and time can help law enforcement target persistent crime hotspots (Gorr & Harries, 2003). In addition, these methods can assist policymakers plan for safer public spaces (Perry et al., 2013). Another example would be to help researchers understand and explain the geography of risk (Li et al., 2014). The spatial forecast of crime-related information is called spatial crime forecasting (SCF) (Kounadi et al., 2020).

Perry et al. (2013) define three categories of predictive methods based on the context and goals of predictive methods:

- Hot-spot analysis, statistical regression, near-repetition, and data mining methods typically identify *where* a crime will occur over a given time period (*when*) and consequently identify *who* is probably to be a victim.
- Temporal and spatial methods are used to determine *when* a crime is most likely to occur. They also determine *who* will be victims because they take into account the local population.
- Risk-terrain analysis addresses the spatial factors that lead to an increased risk of crime and, therefore, *where* certain crimes might occur more frequently.

Most crime forecasting methods involve examining data on past crimes and victims, whether from a tactical (short-term) or strategic (long-term) perspective (Perry et al., 2013). Depending on the purpose for law enforcement, Gorr and Harries (2003) classifies the duration of a forecast as short-term (tactical deployment), medium-term (resource allocation), or long-term (strategic planning). The main challenge for law enforcement agencies today is to efficiently and accurately analyze the growing amount of crime data (Malleson et al., 2010).

Advances in geographic information technologies in the 1990s facilitated the introduction of crime forecasting into police departments (Gorr & Harries, 2003). Today, the use of forecasting methods by police is known as predictive policing. Predictive policing is the application of analytical techniques to identify likely targets for police action and prevent crime through statistical prediction (Perry et al., 2013). In recent years, however, predictive policing has come under heavy criticism for targeting certain groups in society (McGrory & Bedi, 2020; Uberti, 2021).

The main objective of this thesis is to explore the potential of Space-Time Auto-Regressive (STAR) models for long-term crime forecasting. The method is derived from the STARIMA (Space-Time Auto-Regressive Integrated Moving Average) model proposed by Pfeifer and Deutsch (1980). An autoregressive model predicts the variable of interest using a linear combination of its past variables (Cesario et al., 2016). A space-time model accounts for the linear dependence between time- and space-lagged variables (Giacomini & Granger, 2004).

The main reason for the interest in the STAR methods are the findings of Shoesmith (2013), which found that STAR models use data efficiently and outperform other similar models. Yet, his study remains the only paper in the field of spatial crime forecasting, which explored this method.

The study area of this work will be New York City, following the study of Rentzelos (2020), as the city provides open data on criminal incidents and has a high crime rate.

The next section elaborates on some of the problems just described in the area of spatial crime prediction.

1.2 Problem Statement

The main issue concerning the STAR method in SCF is that it has been little researched. The work of Shoesmith (2013) is the only peer-reviewed work that examines the STAR method for forecasting crimes. While STAR models are relatively unexplored in SCF, there are other scientific fields where STAR models have been explored more extensively, such as economics (Nurhayati et al., 2012; Pfeifer & Bodily, 1990), business (Borovkova et al., 2008; Ohtsuka et al., 2010), disease (Gottwald et al., 1992; Reynolds & Madden, 1988), and environmental (Deutsch & Pfeifer, 1981; Ip & Li, 2017).

As mentioned in the previous section, the biggest challenge for law enforcement agencies is to efficiently and accurately analyze the increasing data. Because STAR models have been shown to use data efficiently and often outperform other models (Shoesmith, 2013), they have potential for implementation in law enforcement agencies.

With increasing urbanization bringing significant social and economic changes, city governments face several challenges to ensure public safety (Catlett et al., 2019). As described in the previous section, long-term forecasts are used for strategic planning. To assist city

governments, further long-term SCF research is needed to facilitate strategic decisions for the future.

Predictive policing has experienced a backlash in recent years (Uberti, 2021). For example, McGrory and Bedi (2020) reported on an intelligence program developed by the Pasco County Sheriff with the goal of preventing crimes before they occur. However, over the years, a system was built that continuously monitored and harassed residents based on arrest histories, unspecified information, and arbitrary decisions by police analysts. The inclusion of biased information in predictive methods, as well as a general criticism of police after the murder of George Floyd, prompted several police departments in the United States to ban predictive policing methods (Uberti, 2021). This recent criticism calls for more research on effective methods that eliminate as many confounding variables as possible. Because STAR models use only spatial and temporal information about crimes, they are of interest to the field of predictive policing.

The next section presents the research scope, scientific objectives, and research questions.

1.3 Research Scope

The aim of this paper is to make a further contribution to the field of SCF by investigating the little explored STAR method. Different temporal resolutions are investigated to accommodate different application purposes, but the focus is on long-term forecasts.

Many studies have found variations in the spatial distribution between different types of crimes (Balbi & Guerry, 1829; R. Harries, 2003; Shoesmith, 2013). For example, R. Harries (2003) found that predictions of property crimes are generally more accurate than violent crimes. Therefore, following the classification of Shoesmith (2013) and Rentzelos (2020), crime data are divided into three categories: **all**, **property**, and **violent** crimes.

To verify the performance of a proposed method, a comparable baseline is needed (Lin et al., 2018). The baseline method in this work will be the non-spatial counterpart of STARIMA, the ARIMA models. Research in the field of crime prediction suggests that ARIMA models have high predictive accuracy Cesario et al. (2016), Chen et al. (2008), and Gorr et al. (2003). In addition, the forecast[1] R package allows for quick and easy modeling and prediction of ARIMA models. The comparison between the methods will help you determine if the integration of the space is worth the extra effort.

The different methods are compared at a range of time steps and for differing crime types. The comparison between the models is performed using different error metrics for regressions as well as a visual inspection with maps of the model errors.

In the systematic review of SCF, Kounadi et al. (2020) notes that current studies often lack reporting of study experiments, making them difficult to comprehend. Therefore, the process from data preparation to implementation and evaluation of the models is written

[1] https://cran.r-project.org/web/packages/forecast/

in R, a programming environment for data manipulation, computation, and graphing that many people use as a statistical system (Venables et al., 2021). The R code is published in the appendix of the paper and on GitHub[2]. In addition, the analysis is performed using freely available data and software.

1.3.1 Scientific Objectives and Research Questions

As mentioned earlier, there is a need in SCF for further research on STARIMA methods and their ability to make long-term predictions compared to traditional forecasting methods. This work is intended to help answer the following study objectives (SO) and research questions:

SO 1: To explore the long-term forecasting performance using the STARMA method.

- 1.1 How does the performance of STARMA models vary with increasing levels of time lags?
- 2.2 What is the effect of the model's parametrization on the predictive results?

SO 2: To compare the STARMA method with its ARIMA counterpart.

- 2.1 What is the added value or limitations when using the STARMA method compared to the ARIMA method?
- 2.2 Which models performs best and is this dependent on the time lag?

SO 3: To identify if the STARMA method's performance differs between types of crime.

- 3.1 Are best performing time lags dependent on the crime type, and if yes which time lags are more adequate for each crime type?

1.3.2 Research Boundaries

In this work, freely available data and software are used to be able to reproduce the analysis. Therefore, there are restrictions on which data can be used. A detailed overview of the available options and the selected data can be found in section 3.3. In addition, the R packages are a strong limitation for the analyses performed, since the modeling is done with the currently available packages. A detailed presentation of the software and packages used can be found in section 3.1.

Since exploring different spatial scales (e.g., blocks, zip codes, police precincts) is not part of the research questions, an appropriate spatial scale is chosen based on the selected study area and data, but is not explored further. Also, the selected study area will be treated as an isolated area due to the limitations in terms of harmony of data and time constraints of the thesis.

[2] https://github.com/baccanazzo/ExploringCrimeForecastPerformance_STARM_ARIMA

1.4 Related work

STAR models are relatively unexplored in the field of SCF. While STAR models are relatively unexplored in SCF, there are other scientific fields where STAR models have been explored, such as economics (Nurhayati et al., 2012; Pfeifer & Bodily, 1990), business (Borovkova et al., 2008; Ohtsuka et al., 2010), disease (Gottwald et al., 1992; Reynolds & Madden, 1988), and environment (Deutsch & Pfeifer, 1981; Ip & Li, 2017).

Although STAR models are relatively unexplored in SCF, extensive literature on predicting crime in space and time uses regression models. Table 1.1, based on the table layout of Kounadi et al. (2020), provides a quick comparison of the selected related scientific work. Due to the limited length of the thesis, not every paper can be presented in detail, but three of the most influential studies for this thesis are presented in the following subsections.

Most of the selected studies on crime have their study area in the U.S. at the city level. With its extensive history in criminology and its vast, freely available data, it is generally where most of the studies are conducted. Although most selected studies examine crime at the city level, the spatial units vary in size and shape.

Another factor that varies widely across the selected studies is time. Some studies use a sample period of 12 months, while others use 600 months. However, most studies use a period between 60 and 100 months. The temporal unit also has a wide range: from days to weeks to months to years, most temporal units in the selected studies range.

Regarding crime data, most of the selected studies distinguish between the different types of crimes, mostly between property and violent crimes, due to the different nature of these crimes. The sample size is often not specified but also varies from study to study.

In terms of prediction, most studies use regression-based methods, as regression-based methods are also examined in this thesis. The conclusion of these methods, i.e., what the methods predict, is usually the number of crimes.

In the following subsections, some of these studies are presented to examine what and how the researchers conducted their studies.

1.4.1 Gorr, Olligschlaeger, and Thompson (2003)

Gorr et al. (2003) investigated whether it is possible to accurately predict crime in small areas, such as police districts, one month in advance. They compared simple univariate methods with simple naïve models commonly used by police.

Validation of the forecasting models was based on splitting the data samples into estimation and denial samples to measure out-of-sample forecast accuracy. A rolling time horizon was used to produce 36 forecasts with one month lead time for each precinct during 1996-1998. For each monthly forecast, the immediately preceding 60-month period was used to estimate the forecast model. The forecast was then compared to the actual closure month to calculate the resulting error. Three different methods were used to measure fore-

Table 1.1: Adapted from Kounadi et al. (2020): Overview of related spatial crime forecasting studies

Source	Space		Time		Crime data		Forecasting			
	Study Area	Scale	Sampling period	Months	Type	Samples	Inference	Task	Spatial Unit	Temporal unit
Brown and Oxford (2001)	Richmond, VA, USA	City	1994–1999	72	Breaking and entering	$\lesssim$24.000	Number of crimes	Regression	$\approx 0.64mi^2$ ($\approx 1,65km^2$) grid cells	Week, Month
Cesario et al. (2016)	Chicago, IL, USA	Part of City	2001–2014	168	All crimes	174.403	Number of crimes	Regression	Area of Chicago ($10.869km^2$)	Week
J. Cohen et al. (2007)	Pittsburgh, PA, USA	City	1991-1998	96	Violent and property	$\approx$1,3 million	Number of crimes	Regression	$4.000ft$ ($\approx 1.220m$) sq. grid cells	Month
Dash et al. (2018)	Chicago, IL, USA	City	2011–2015	60	34 crime types	6,6 million	Number of crimes	Regression	Communities	Month, Year
Gorr et al. (2003)	Pittsburgh, PA, USA	City	1991-1998	96	5 crime types	$\approx$1 million	Number of crimes	Regression	Police precincts	Month
Ivaha et al. (2007)	Cardiff, UK	City	2001-2003	26	Criminal damage	N.A.	% of crime in clusters	Regression	Clusters of varying size	Month
Kadar and Pletikosa (2018)	New York, NY, USA	City	2014–2015	24	All and 5 crime types	174.682	Number of crimes	Regression	Census tract	Year
Liesenfeld et al. (2017)	Pittsburgh, PA, USA	City	2008–2013	72	All crimes	9.936	Number of crimes	Regression	Census tract	Month, Year
Rentzelos (2020)	New York, NY, USA	City	2016	12	All, violent and property	312.403	Hot spots	Regression	Zip codes	Week
Rodriguez et al. (2017)	San Francisco, CA, USA	City	2003–2013	120	Burglary	N.A.	Properties of clusters	Regression	Clusters	Day
Shoesmith (2013)	USA	Country	1960-2009	600	Violent and property	N.A.	Crime rate	Regression	Regions of USA	Year
Zhao and Tang (2017)	New York, NY, USA	City	2012-2013	12	N.A.	N.A.	Number of crimes	Regression	$2km$ sq. grid cell	Day, Week
Zhuang et al. (2017)	Portland, OR, USA	City	2012-2016	58	All crime	N.A.	Hot spots	Binary classification	$600ft$ ($\approx 180m$) sq. grid cells	2 Weeks

cast accuracy: mean absolute error (MAE), mean squared error (MSE) and mean absolute percentage error (MAPE).

Gorr et al. (2003) found that the number of violations must be on the order of 30 or more per spatial unit to achieve an absolute forecast error of 20% or less. A second important finding was that virtually any model-based forecasting approach is far more accurate than current police practice. Gorr et al. (2003) also point out that the length of the planning horizon often classifies forecasting and decision problems. For example, short-term crime forecasts are used for tactical crime control measures, medium-term forecasts are used for resource allocation, and long-term forecasts are used for strategic planning. Therefore, the use of medium- and long-term forecasts for police operations should be considered when making spatial unit decisions for this study.

1.4.2 Shoesmith (2013)

Although STAR models were introduced over 30 years ago, the work of Shoesmith (2013) is the only work published in SCF that explore STAR models. Shoesmith (2013) used the STAR model to predict violent and property crime in the United States at the regional and state levels. He then compared the results of univariate AR models, vector AR (VAR) and Bayesian VAR, and two naïve models.

The regional crime rates for each year were calculated by summing the number of crimes in each state in each region and dividing by the region's total population in that year, expressed as crimes per 100.000 population. The nine regional projections for each year were then weighted by population to obtain the total national crime rate. The actual regional populations for each year were then used in the forecast intervals to focus the analysis on the predicted crime rates. The predicted national crime rates were then expressed in natural logs and compared to the actual log values of the individual crime rates at each forecast step. Therefore, the forecast errors are roughly equivalent to the percentage errors. The measure of forecast accuracy was the root mean squared error (RMSE), which retains units and can also be interpreted as the percent forecast error (Shoesmith, 2013).

The study concluded that STAR models outperformed five other aggregate forecasting approaches in predicting violent crime, and particularly property crime. Shoesmith (2013) concludes that the use of group estimates and spatial lags improves forecast accuracy at all levels, from long-term crime forecasts at the national and state levels to one-month patrol district forecasts aimed at crime prevention.

1.4.3 Rentzelos (2020)

Another academic study on STAR models is the master's thesis by Rentzelos (2020). He compared the space-time autoregressive moving average (STARMA) model with a kernel density estimation (KDE) and a naïve model for property, violent, and total crime.

Rentzelos (2020) experimented with the STARMA parameters to examine how accounting for space and time affects the model's performance. The best experiment was compared with the baseline and conventional methods, dividing the model results into two classes (hotspots and non-hotspots) based on a threshold, and then using evaluation metrics for spatial prediction accuracy.

The results showed that all three methods produce significant results, and each method performed better for the different crime types studied (all, property, and violent). In particular, the naive baseline method performed better for the "all" crime type and the conventional KDE method for the "property" crime type. In contrast, the proposed STARMA method performed better on the "violent" crime type. Therefore, he concludes that the current STAR models are pretty sensitive to spatial and temporal parameters. Moreover, further research needs to be conducted to investigate how these models' spatial and temporal resolution could outperform the baseline methods (Rentzelos, 2020).

1.5 Structure of the thesis

This section outlines the structure of the thesis. After the brief introduction to the topic of this paper, the theoretical framework of this is provided in chapter 2. First, a brief overview of the history of crime forecasting and the role of cartography and GIS in the field is outlined. Theories within environmental criminology and the factor of space in crime are reviewed in section 2.2. Important measures to account for spatial correlation and how to represent a neighborhood relationship in an equation are presented in section 2.3. Section 2.4 introduces the different categories of prediction methods. Finally, the prediction methods studied in this work are discussed in detail in section 2.5 and the performance metrics used to compare the methods can be found in section 2.6.

Chapter 3 contains the methodology of the work. First, section 3.1 introduces the software that was used to explore, transform, analyze, and map the data. The study area is described in section 3.2. The crime data used for the analysis are presented in section 3.3, along with the spatial resolution decisions, temporal resolution selection, and processing steps required to transform the data for the models. The modeling workflow is then outlined in Figure 3.1, from data exploration to the creation of spatial weights to finally model the methods.

The results are presented and discussed in chapter 4. First, some general observations are made. Then, the results are compared based on temporal resolution. Finally, the research questions are answered.

The final chapter of this thesis summarizes the main findings and provides specific answers to the research questions. It also highlights the limitations of the study and makes suggestions for future work.

In the Appendix tables (Appendix A) that are too long to include in the main part and

the final R Code (Appendix B) are attached.

Chapter 2

Theoretical Framework

The second chapter of this thesis presents the theoretical framework of this work. First, the history from mapping crime to predicting crime is briefly reviewed. Then, the main theories of environmental criminology are provided. Then, important metrics for measuring spatial autocorrelation are introduced and methods to incorporate spatial dependencies into an equation. Crime prediction methods are then explained, followed by the presentation of the prediction methods used in this thesis. The last section of the second chapter presents the performance metrics used to compare the analyzed methods.

2.1 From mapping crimes to forecasting them

This section introduces the three schools that have significantly influenced environmental crime, and briefly traces the path from crime mapping to understanding criminal patterns to predicting crime.

2.1.1 The Cartographic School

Nearly 200 years ago, the first formal study of crime and space began, led by social ecologists André-Michel Guerry and Lambert-Adolphe Quetelet (Ferguson, 2011). Balbi and Guerry (1829) produced one of the earliest crime maps (Figure 2.1) showing the relationship between educational attainment (bottom map), violent crime (top left), and property crime (top right) in France (Hunt, 2019). If we compare the three maps in Figure 2.1, we can see that crimes against property are more frequent in regions with low education levels. In contrast, crimes against persons are more frequent in regions with a high level of education.

Quetelet (1842), on the other hand, added statistics to his maps to show spatial differences in France and its social groups. He also noted that crimes against persons reach a maximum in summer and crimes against property in winter (Quetelet, 1842). Guerry and Quentelet are considered the founders of the **Cartographic School** (Chainey & Ratcliffe, 2013). With their followers, these early pioneers were the first document and map an em-

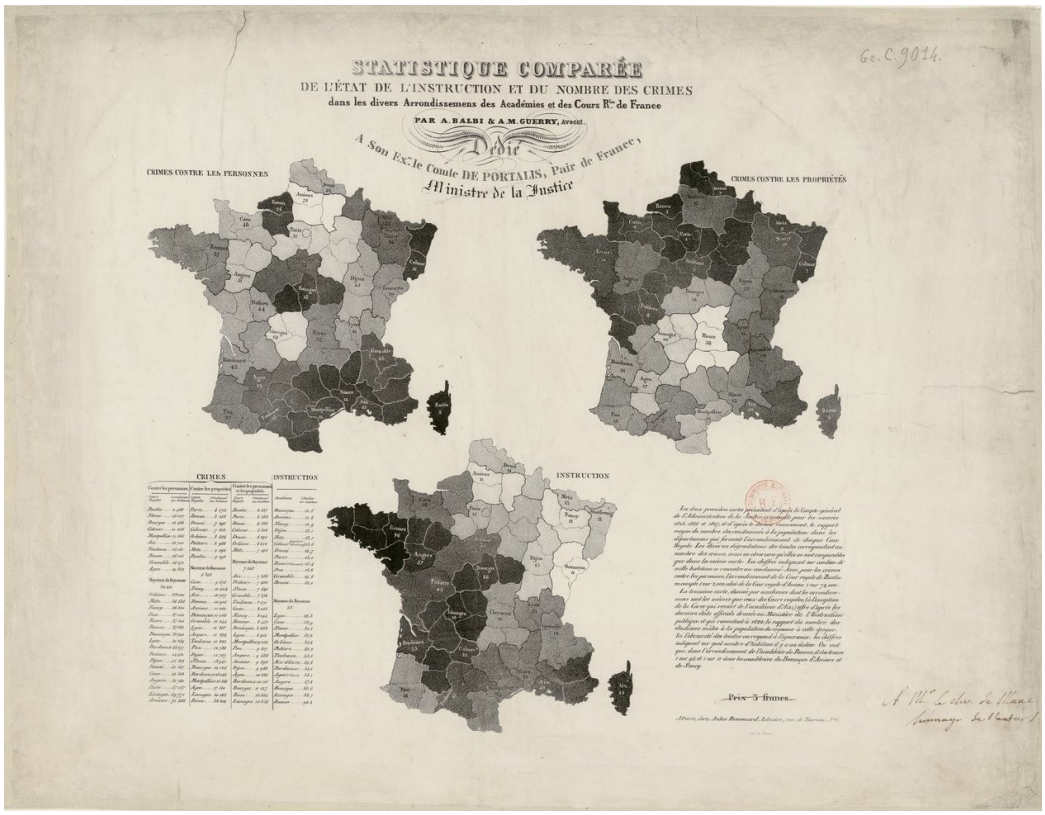


Figure 2.1: Balbi and Guerry (1829): Distribution of education level and crime in France.

pirical regularity of crime. These early pioneers, along with their followers, were the first to document and map an empirical regularity of crime (Ferguson, 2011).

2.1.2 The Chicago School

In the early 20th century, the **Chicago School** added new ideas to the field of environmental crime. The Chicago School refers to a particular group of sociologists at the University of Chicago who focused on the city as a social laboratory in the first half of the 20th century (Lutters & Ackerman, 1996).

Drawing on the human ecology theory, which views the city like the natural ecological communities of animals and plants (Park & Burgess, 1925), Burgess (1925) developed the concentric zone model (Figure 2.2). Burgess' model assumes that the city is surrounded by five concentric circles, each of which expresses a different degree of development of the city (Chainey & Ratcliffe, 2013). As each person belongs to a particular community and thus to a particular geographic area, relocation and dispersal occur until he or she finds a place where he or she can shelter and contribute to the community (Park & Burgess, 1925). In Transition Zone II, the zone with the highest mobility, the "ghetto" exists as a place where social ties are strong and social disorganization is not the same as outside the ghetto, resulting in areas of high crime (Jørgensen, 2010).

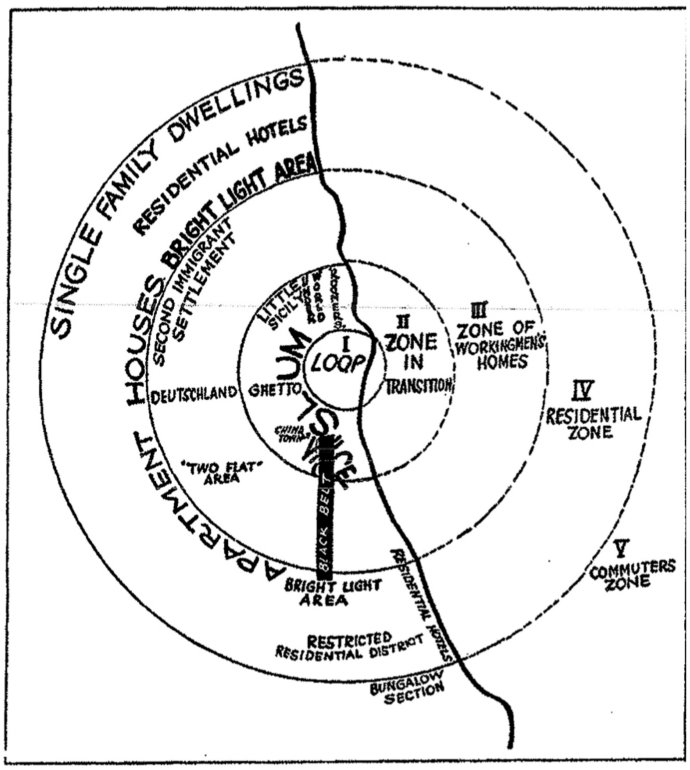


Figure 2.2: Burgess (1925): Concentric model of Chicago.

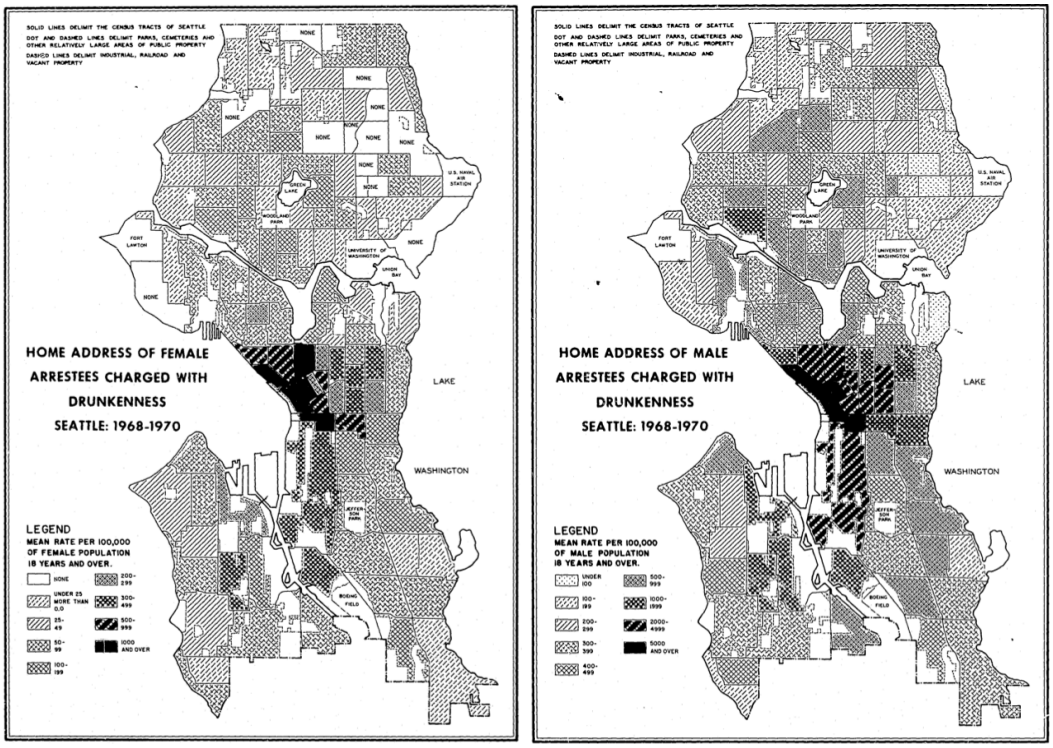
Shaw and McKay (1931, 1942, 1969) extended human ecology theory and the concentric zone model with the development of **social disorganization theory**. They mapped the residences of juveniles facing court and found that the distribution of criminals in the city conformed to a systemic pattern, with high crime rates concentrated in Zone II during the transition period (Akers, 2012). Shaw and McKay also discovered that the spatial arrangement of high-crime areas remained the same over time, although the overall ethnic and racial makeup of residents changed (Kikuchi, 2010). Thus, they argued that the spatial distribution of juvenile crime was due to environmental characteristics such as racial heterogeneity rather than the personal characteristics of individuals (Kikuchi, 2010). These findings are very interesting given recent criticisms of predictive policing. The inclusion of space, according to Shaw and McKay's findings, is not a biased variable, but rather a symptom of deeper socioeconomic inequalities (Uberti, 2021).

Shaw and McKay's work formed the basis for much of U.S. criminology, but was not applicable in Europe due to more extensive urban development (Chainey & Ratcliffe, 2013).

The complex link between urban ecosystems and social relations was the driving force behind most of the sociological work of the Chicago School (Jørgensen, 2010). Their findings inspired criminology and spawned new theories (section 2.2), which over time led to interest in studying crime with newly developed GIS technologies (Ferguson, 2011).

2.1.3 The GIS School

While theories explaining the relationship between criminal behavior and space took off in the second half of the 20th century, crime forecasting itself was made possible by advances in information technologies, particularly geographic information systems (GIS) (Gorr & Harries, 2003). Although the first use of computers in crime mapping dates back to the mid-1960s in St. Louis, it took until the late 1990s for computers to become affordable, facilitating the entry of GIS into crime analysis (K. Harries, 1999). An example of early GIS maps are the maps in Figure 2.3 of Schmid and Schmid (1972). These show the different spatial distribution of home addresses of female (map 2.3a) and male (map 2.3b) arrestees charged with drunkenness in Seattle between 1968 and 1970.



(a) Female arrestees. (b) Male arrestees.

Figure 2.3: Schmid and Schmid (1972): home address of arrestees charged with drunkenness Seattle 1968-1970.

GIS made it possible to list reported crimes in a comprehensive database, allowing law enforcement to examine historical and current crime patterns in any location at any time (Ferguson, 2011). GIS has enabled environmental criminologists to explore the spatial dimensions of crime in new and compelling ways, forming the **GIS School** (Chainey & Ratcliffe, 2013).

The history of predictive policing shows us the importance of maps and GIS in this field. Especially how maps can convey information in an accessible way and how GIS has opened

up new possibilities in data analysis.

2.2 Theories of environmental criminology

Environmental criminology theories form the basis of the SCF. The following theories explain why offenders and victims concentrate in certain locations, forming spatial patterns. These theories are an important reason why incorporating space into crime prediction is important.

2.2.1 Routine Activity Theory

The **Routine Activity Theory** (RAT) of L. E. Cohen and Felson (1979) states that most criminal acts require the spatial and temporal convergence of the following three elements:

- motivated offenders,
- appropriate targets,
- and the absence of capable guards.

The central premise of the theory is that crimes can be prevented if any of these three elements do not spatially converge (L. E. Cohen & Felson, 1979). L. E. Cohen and Felson (1979) also argues that we should be able to predict crime because of the consistency in our routines.

Routine Activity Theory (RAT 1, Figure 2.4) has evolved since its initial introduction (Eck & Madensen, 2015). Based on the work of Hirschi (1969), Felson (1986) further developed the basic theory (RAT 2, Figure 2.4) and introduced another controller: the handler. Handlers strive to keep potential offenders from getting into trouble through emotional and social ties (Eck & Madensen, 2015).

Although both RAT 1 and 2 suggest that spatial proximity is required, the study of Sherman et al. (1989) showed that location is an essential concept of the RAT. Their study analyzed calls to police to locate hotspots in Minneapolis and found that 50% of calls were to 3% of all possible locations (Sherman et al., 1989). They provide, among other (Cromwell et al., 1995; Kennedy & Forde, 1990; Messner & Tardiff, 1985), empirical validation of Routine Activity Theory.

Because both offenders and victims have a controller, Eck (1994) suggested that persons who own property or are authorized by the owner are place controllers. Finally, Madensen (2007) proposed a theory of place management that shows how it works to prevent crime and what factors influence place management, leading to RAT 3 (Figure 2.4).

Sampson et al. (2010) provided another and since today final extension of the theory to answer the question of why crime occurs even in the presence of three controllers. They

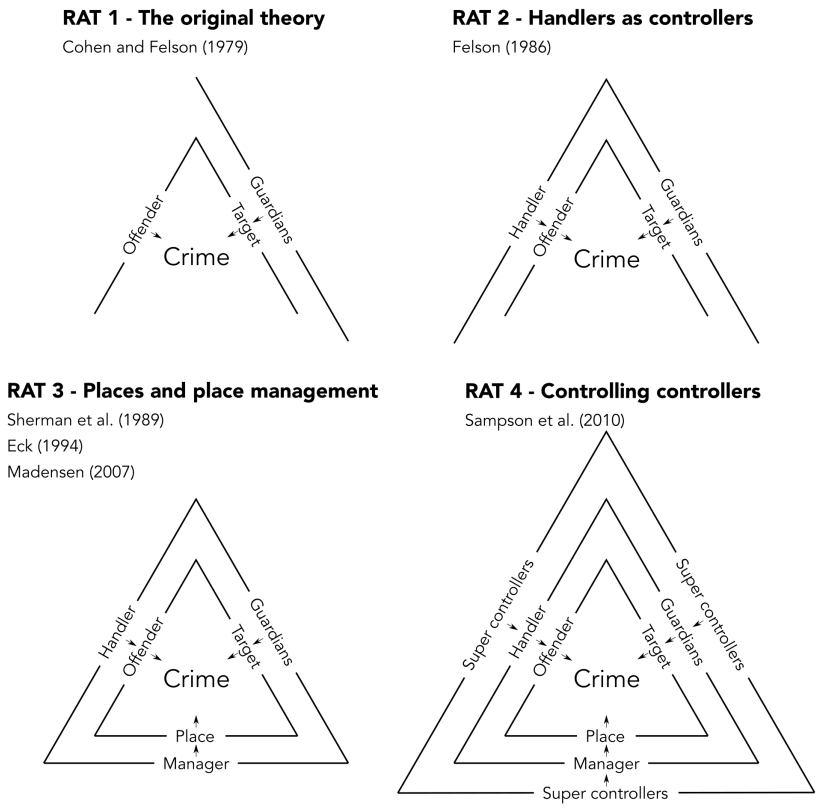


Figure 2.4: Adapted from Eck and Madensen (2015): The development of Routine Activity Theory.

introduced supercontrollers (RAT 4, Figure 2.4): People, organizations, and institutions that create incentives for controllers to prevent or facilitate crime (Sampson et al., 2010).

Rational choice theory and its complements show the complexity of a criminal event and how offenders, targets, locations, and controllers interact. Before all the additions to the RAT were proposed, other criminologists attempted to understand the spatial dimension of crime and put their findings into a theoretical framework.

2.2.2 Geometry Of Crime

The **Geometry Of Crime**, introduced by P. L. Brantingham and Brantingham (1993) and P. J. Brantingham and Brantingham (1981), seeks to understand the spatial dimension of crime by considering the concepts of edges, activity nodes, and paths (Song et al., 2017). Based on the work of Lynch (1960), who sought to understand the role and impact of urban planning, the geometry of crime examines how the spatio-temporal dimension of a criminal event interacts with the perpetrator and the target (Andresen, 2020).

P. L. Brantingham and Brantingham (1993) and P. J. Brantingham and Brantingham

(1981) introduced the term environmental backcloth to refer to the movement and change in our environment. For example, we may feel safer in a square that is busy during the day than alone in the same square late at night. Thus, although the environment remains the same, the environmental framework changes our perception when the temporal dimension of the same place changes (Andresen, 2020). This change in the temporal dimension can range from seconds in the case of hooligans leaving the stadium after a game to decades in the case of the gentrification process of a city.

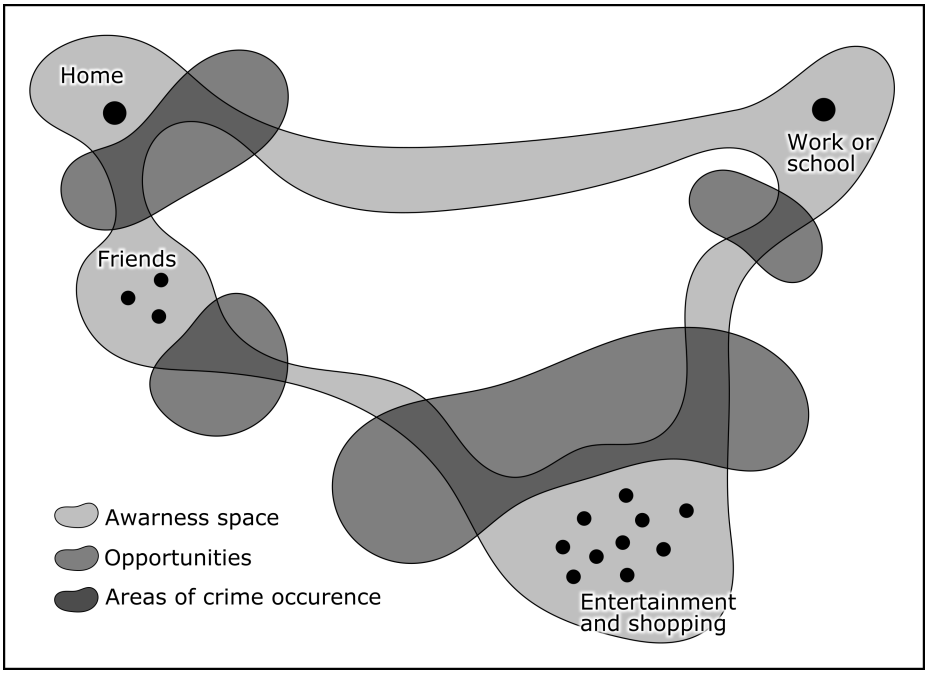


Figure 2.5: Hypothetical model adapted from Chainey and Ratcliffe (2013): Where offender awareness space and opportunities overlap, areas of criminal occurrence are formed.

In our daily routine activities, such as walking from home to work, to school, or to the grocery store, we create a cognitive map of paths, places, and even social and economic infrastructures where we feel comfortable over time (Chainey & Ratcliffe, 2013). As a result, these places (also called activity nodes) and the paths between them become our *awareness spaces* (P. J. Brantingham & Brantingham, 1981). Criminals also develop spaces of awareness in their environment, and some of them become spaces of opportunity because of the absence of guards (Figure 2.5). One of the reasons why crime is concentrated in certain places is that the nature of our built environment and the convergence of activity spaces make up only a relatively small portion of a city's land area (Song et al., 2017).

Another exciting concept introduced by the Brantinghams are edges: These can be physical edges such as a river, a railroad line, or a change in land use, or they can be subtle edges, changes that are felt even though no physical boundary is crossed (Andresen, 2020). Along these edges, P. L. Brantingham and Brantingham (1975, 1978) found, burglary rates were far higher than in the interior of neighborhoods. Anonymity is present at both

subtle and physical edges, allowing for situational assessment of criminal opportunity that is helpful to crime (Song et al., 2017).

The Brantinghams have made an important contribution to explaining why space and crime are intertwined. The next subsection presents the theory of the decision-making process of criminals and why the decision to commit a crime is associated with space.

2.2.3 Rational Choice Theory

The **Rational Choice Theory** by Cornish and Clarke (1986) has at its core the ideas of choice and decision-making and the centrality of the crime event to ongoing criminal activity: while failure leads to a reduction or even discontinuation of crime, success in committing crime drives the development of a criminal lifestyle (Cornish & Clarke, 2008). The foundation of the theory lies in the expected utility principle in economic theory, which states that someone makes rational decisions based on the expectation of minimizing costs or losses and maximizing gains or benefits (Akers, 2012). The decision-making process of the Rational Choice Theory consists of three fundamental choices when it comes to crime:

1. the decision to be committed to a crime,
2. the decisions required for particular criminal events,
3. and the decision to cease from crime (Andresen, 2020).

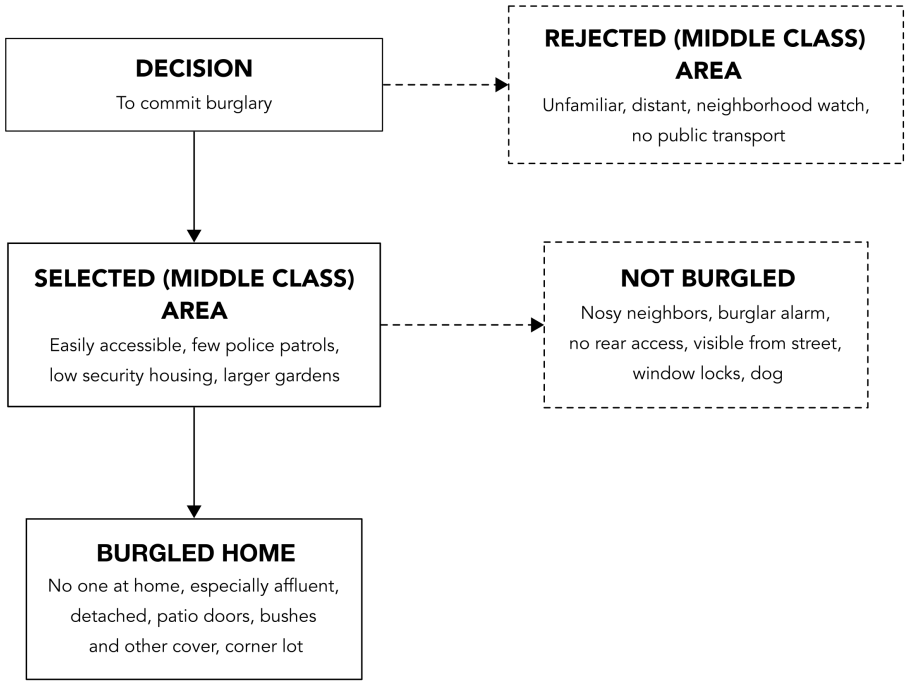


Figure 2.6: Adapted from Andresen (2020): Decision-making process of a burglary.

It is the second point that is of greatest interest to environmental criminologists (Andresen, 2020). For example: a person has decided to commit a burglary (Figure 2.6). The perpetrator then makes rational decisions about the area and house of interest. The geographic characteristics of the crime scene play an important role in his decision-making process. Thus, if the crime scene was an excellent rational choice for the burglar, it might also be an excellent rational choice for another burglar.

2.2.4 Crime Pattern Theory

The **Crime Pattern Theory** is a metatheory that combines the three main theoretical perspectives on environmental crime (P. J. Brantingham & Brantingham, 2008) discussed in the previous subsections:

- the Routine Activity Theory addresses differences or shifts in the social environment that are indicative of changes in crime rates,
- the Geometry Of Crime deals with the constructed environment and how it shapes the geographic pattern of crime,
- and the Rational Choice Theory looks at with the cognitive environment that determines the decision-making processes of potential criminals (Andresen, 2020).

Individually, each theory provides a solid understanding of crime, but taken together, they can provide a meaningful account of the environment in which crime occurs within the framework of the Crime Pattern Theory (Andresen, 2020). Thus, the theories of environmental criminology provide us with a theoretical understanding of why crimes occur in certain places and at certain times. Moreover, the theories lead to the notion that place, rather than people, is the critical element of crime (Hunt, 2019). Therefore, it should be possible to predict crime using space as a variable (P. J. Brantingham & Brantingham, 1981).

The following section examines how to identify and measure the patterns and how to model the spatial component in an equation to predict crime.

2.3 Measuring spatial distribution

The history and theories in environmental criminology showed us that there are spatial patterns in crime. But how can we measure the spatial distribution of crimes? And how can model the spatial distribution into a equation? This section will shortly answer those questions and present some key spatial statistics.

2.3.1 Centrographic Statistics

Centrographic statistics are the most basic type of descriptors for spatial distribution of crime incidents (Levine, 2015). The term centrographic represents the group of geographical studies in the field of two-dimensional statistical analysis (Sviatlovsky & Eells, 1937). They allow us to measure and assess a phenomenon's average location, dispersion, movements, and directional change over time (LeBeau, 1987). The field of centrographic statistics (Ebdon, 1977; Furfey, 1927; Lefever, 1926; Neft, 1966) includes:

- Mean center
- Median center
- Center of minimum distance
- Standard deviation of X and Y coordinates
- Standard distance deviation
- Standard deviational ellipse

These statistics were applied to crime analysis, for example, by LeBeau (1987) who found that different classes of sexual offenders have relatively different spatial distributions and that the spatial distribution of offender classes is not uniform over time. Centrographic statistics can be an excellent tool to quickly get a sense of the spatial distribution of the data.

2.3.2 Spatial Autocorrelation

If the relative location of crimes can provide information about the variation in the spatial pattern of the data, we can say that the data also shows **spatial autocorrelation** (Cliff et al., 1975). Spatial autocorrelation is one of the most well-known concepts in spatial statistics and implies the spatial dependence of the data set (Levine, 2015). For example, if high crime values in one location are associated with high crime values in a neighboring location (i.e., spatially clustered), there is positive spatial autocorrelation (Figure 2.7). In contrast, if all values are randomly distributed, there is no spatial autocorrelation. And when high and low crime values alternate, there is negative spatial autocorrelation. An example of negative spatial autocorrelation is when the fluctuating values of nearby crime locations become increasingly similar as the distance between crime locations increases (Leitner et al., 2021).

Spatial autocorrelation can be measured with statistics like Moran's I (Moran, 1950), Geary's c (Geary, 1954) or Getis-Ord G (Getis & Ord, 1992). These statistics provide a value for the entire study area and are also referred to as global spatial autocorrelation indicators. In contrast, the local indicators of spatial association (LISA) proposed by Anselin (1995),

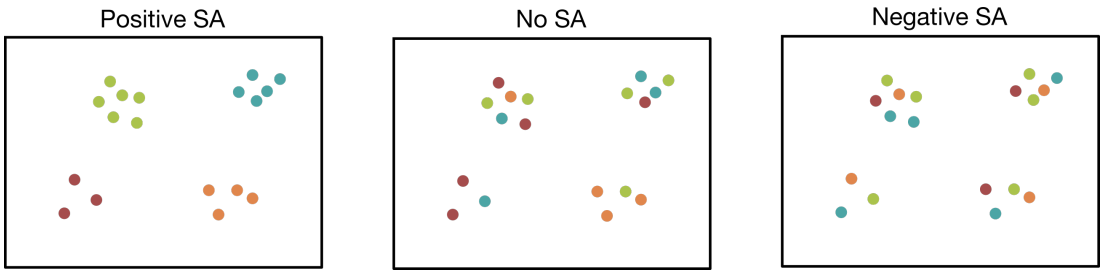


Figure 2.7: An illustration of the three types of spatial autocorrelation.

indicate variation in spatial autocorrelation within and across the study area (Leitner et al., 2021). Since LISAs are adaptations of global spatial autocorrelation indicators that establish a proportional relationship between the sum of local statistics and a corresponding global statistic, there are as many corresponding LISAs as there are global indicators (Anselin, 2020).

2.3.3 Assessing spatial weights

The basis of spatial analysis is Tobler's first law of geography: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). But what is the distance of near things? Conversely, how distant are distant things? One of the most important questions that arises with spatially dependent data is how to weight the distance between neighbors.

Moran (1948) introduced the concept of the weighting (or weights) matrix. **Spatial weights** are a crucial component in the specification of spatial variables in the models used in this thesis (Anselin & Rey, 2014).

The weights represent the neighbor structure between the observations as a $n \times n$ matrix W in which the elements w_{ij} of the matrix are the spatial weights (Anselin & Rey, 2014):

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \tag{2.1}$$

If the spatial unit j is a neighbor of the unit i, the spatial weight $w_{ij} \neq 0$, otherwise, when $w_{ij} = 0$ they are not neighbors (Kelejian & Piras, 2017). The matrix expresses the neighbor relation in a binary form (1 or 0) in its most simplistic form. A row i represents each spatial unit in the matrix and the possible neighbors by a column j.

Two of the most commonly used constructions of spatial weights are neighborhood relations based on **contiguity** and **distance** measure (Anselin & Rey, 2014). Anselin and Rey (2014) define contiguity as two spatial units that share a common border of non-zero length. Additionally, named after the allowed moves of the named chess pieces, the conti-

guity can be distinguished if the common border, vertex, or both are defined as neighbors (Figure 2.8).

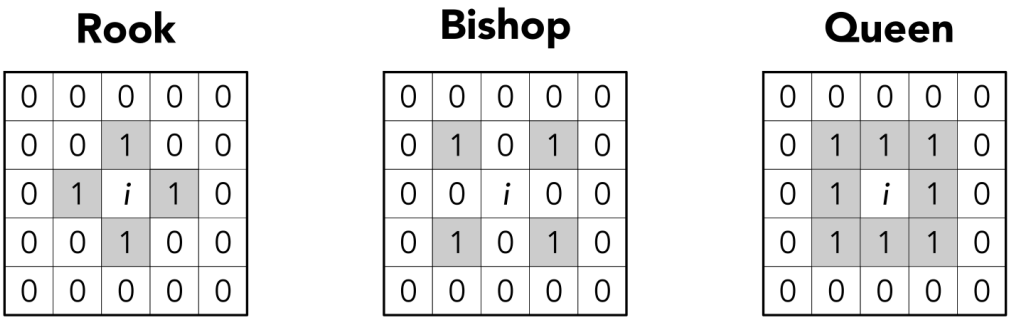


Figure 2.8: The different criteria of contiguity in a neighbor relation of area i, expressed in binary form.

Since now, we have only considered direct neighbors, also called first-order neighbors, but one can define k-neighbors. So, for example, a second order neighbor would be a first-order neighbor to any observation that is already a first-order neighbor of i. To avoid duplications, it is only applying to locations that are not already first-order neighbors. Figure 2.9 illustrates the third order neighbor structure of i in a rook case.

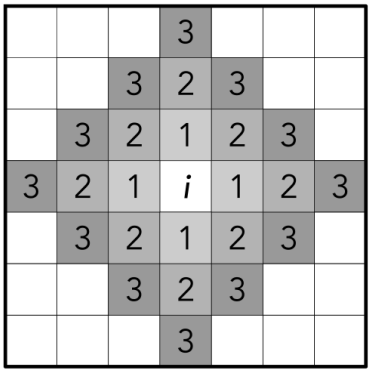


Figure 2.9: First, second and third order neighbors of area i in rooks case.

While contiguity measures calculate the distance between polygons, distance measures calculate the distance between points. Similar to the contiguity measures, there are different methods to calculate the distance when assessing spatial weights. One of the most used distance statistics is the nearest neighbor method developed by Cliff and Ord (1981). This method considers $w_{ij} = 1$ if they are the nearest of the kth observations or if the observations i and j are within a given distance defined by the researcher ($d_{ij} \leq \alpha$), or otherwise assign $w_{ij} = 0$ (Militino et al., 2004). Another prominent practice to define weights is through an inverse polynomial of the distance between each pair of observations ($w_{ij} = 1/d_{ij}$) (Kurt & Tunay, 2015).

Regardless of which method is used, if the weight matrix is not defined correctly, it can lead to inconsistencies in the coefficient estimates and reduce the forecasting accuracy of the models (Anselin, 1988). Therefore, spatial weights are selected before the implementation of the models and should reflect the geographical features of the study area (Kurt & Tunay, 2015).

2.4 Predictive methods

In order to predict crime, two things had to happen: the establishment of theories within environmental criminology and the advances of GIS, specifically the use of GIS to map crime in police departments (Gorr & Harries, 2003). The following subsections present the various categories of crime prediction techniques that have evolved to date.

2.4.1 Hot Spot Analysis

Hot Spot Methods are used to predict areas of increased crime risk based on past crime data (Perry et al., 2013). Sherman (1995) defines hot spots "as small places in which the occurrence of crime is so frequent that it is highly predictable, at least over a 1-year period." Since no numerical threshold defines the number of crimes that make an area a hot spot, they are relative to their study area (Chainey & Ratcliffe, 2013).

Figure 2.10 shows different hot spot maps of criminal activity in New York City (NYC) during New Year's Eve from 2006 to 2020. To compare the different techniques and settings, crime counts of arrests are classified in the same way for all maps (except Map A). Figure 2.10 shows the impact of the different settings on the size of the hot spots. Looking at all maps in Figure 2.10, it is clear that the largest crime hot spot on New Year's Eve is downtown Manhattan. This high concentration of crime is related to the annual New Year's Eve celebration in Times Square, which is typically attended by 58.000 people (Ly & Hanna, 2021). However, the size of the hotspot changes depending on the method and settings. We will encounter the same problem later when we evaluate the spatial weighting of the models under study.

The following points briefly explain the different methods for creating hot spot maps:

- **Point maps** are the most common methods to display geographic patterns of crime (Jefferis, 1999). They are popular because they resemble the familiar and traditional way of placing pins onto a wall map (Chainey & Ratcliffe, 2013). However, the disadvantage of these maps can be seen in Map A in Figure 2.10: it is difficult to identify the location, number, and shape of criminal hot spots when points overlap.

- **Geographic boundary maps** are maps where crime events are aggregated to administrative or statistical boundaries like states, counties, or police precincts. The counts of crimes then can be used to create thematic maps (also called choropleth

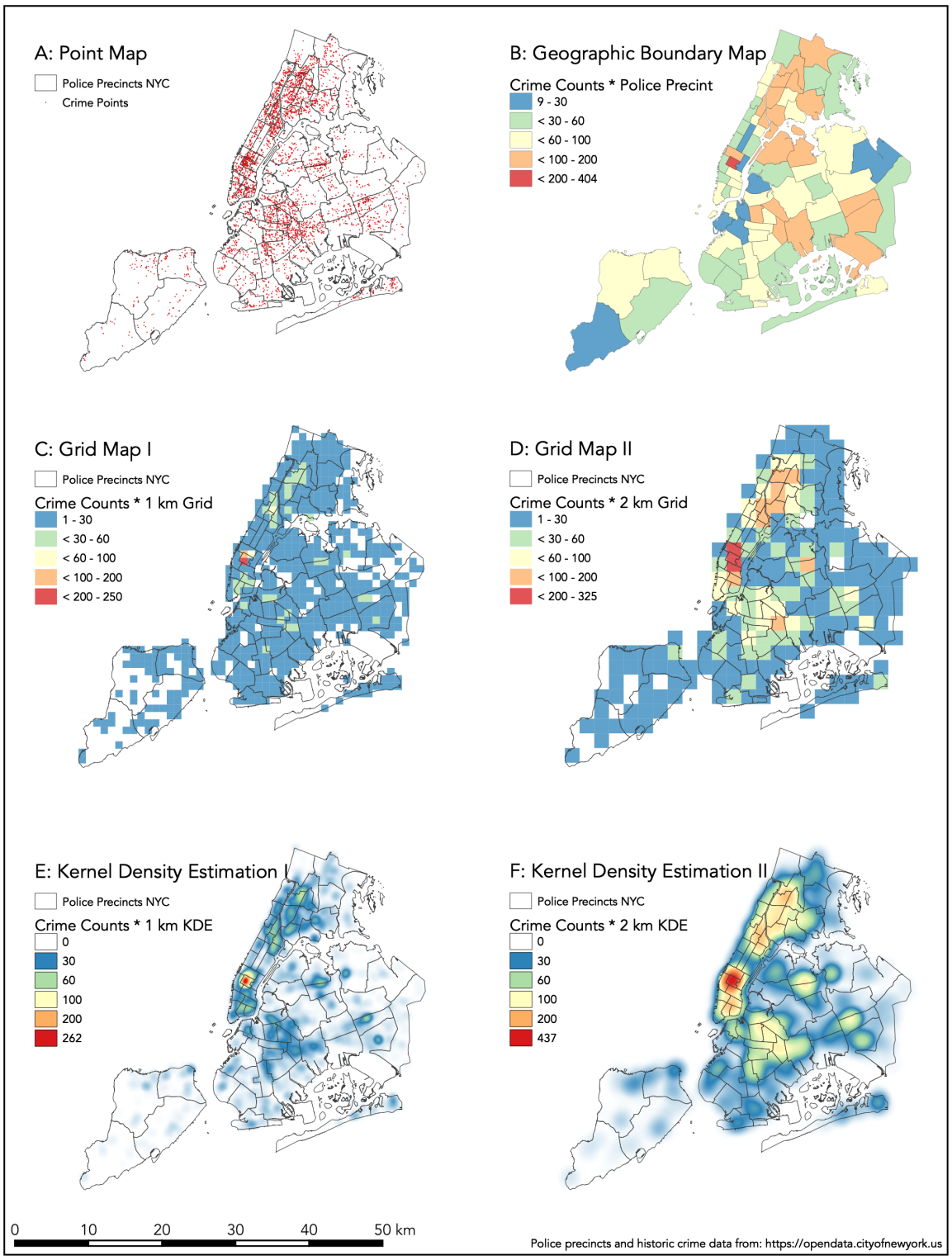


Figure 2.10: Hot Spot Maps of criminal activity on December 31st from 2006 to 2020 in New York City.

maps - see Map B in Figure 2.10) that display the criminal distribution in the study area (Eck et al., 2005).

The downside of this technique is that it can be misleading: naturally, the map reader is drawn to extensive, heavily shaded areas (MacEachren, 1995; Monmonier, 1991). Furthermore, the aggregation of points shades the whole area but often does not represent the actual spatial pattern of the criminal events and gives the impression that crime spreads over the whole area (Chainey & Ratcliffe, 2013). Although not done by many crime mappers, the appropriate approach would be to divide the number of crimes by some appropriate denominator, like the number of residents in the specific area. Additionally, the researcher defines the breaks in the classification, which can be subjective and modified to enforce its point.

- **Grid maps** can be used to overcome the downside of varying sizes and shapes of geographic areas (see Map C and D in Figure 2.10). First, a quadratic grid gets laid on top of the study area, and then the crime events get aggregated into the overlapping grid cell. While grid maps tend to better represent the spatial pattern of crime, they have similar problems as geographic boundary maps in terms of accurately assigning hot spots. Map D, for example, has a "red" hot spot eight times larger than Map C's hot spot.

- **Clustering Methods** usually set a starting point and then find the point farthest from the first point, so they are divided into two groups (K. Harries, 1999). After that, the distances from each starting point to other points are calculated repeatedly, and clusters are formed based on new starting points to minimize the sums of distances within the clusters (K. Harries, 1999).

- **Continuous surface smoothing methods** like the kernel density estimation (KDE), create a smooth continuous surface to express the density of crimes spread across the study area (Chainey & Ratcliffe, 2013). The KDE method is explained by Eck et al. (2005) in the following steps:

  1. "A fine grid is generated over the point distribution [...];
  2. A moving three-dimensional function of a specified radius visits each cell and calculates weights for each point within the kernel's radius. Points closer to the center receive a higher weight and therefore contribute more to the cell's total density value.
  3. Final grid cell values are calculated by summing the values of all circle surfaces for each location."

  Map E and F in Figure 2.10 show how different search radii (Map E 1 km, Map F 2 km) impact the results. Chainey et al. (2008) compared KDE with other hot spot techniques and found that the KDE has the highest prediction accuracy index.

Hot spot analysis methods provide a overview of the study area in the form of hot spot maps, allowing professionals to read them easily and make quick decisions (Bachner, 2013). However, hot spot techniques ignore the simultaneous interaction of space and time in crime prevalence (Grubesic & Mack, 2008). Because traditional hot spot analysis is based on the hypothesis that crime remains constant (Rentzelos, 2020), it has clear disadvantages for making long-term predictions.

2.4.2 Regression Methods

For Young (2018), a: "regression equation describes how the mean value of a y-variable (also called the response or dependent variable) relates to specific values of the x-variable(s) (also called the predictor(s) or independent variable(s)) used to predict y. A regression model also incorporates a measure of uncertainty or error." Contrary to the hot spot analysis, regression models can incorporate various explanatory independent variables. The selection of these independent variables can be made with different methods like:

- a **manual selection** of correlated variables;
- **forward/stepwise regression** that iteratively adds additional variables to build the statistically "best" regression model;
- or **mathematical optimization methods** that place penalties on variables and solve complicated optimization problems to fit the "best" overall model (Perry et al., 2013).

While these methods often lead to accurate models, leading indicators of analytic policing enable the transition from reactive to proactive (Perry et al., 2013). A leading indicator of crime can be a variable that indicates the direction in which crime will move in the future. For example, Gorr and Olligschlaeger (2002) found that simple time series, regressions that consider only the change in crime during a period outperform more complex regressive methods.

2.4.3 Data Mining Techniques

Cleve and Lämmel (2020) define **data mining** as: "the extraction of knowledge from data." A more precise definition can be found in the online dictionary Merriam-Webster (2021) which defines data mining as: "the practice of searching through large amounts of computerized data to find useful patterns or trends." Regression models are a subset of data mining, as are the following methods:

- **Classification methods** predict a category for a result (e.g., "There is a 90% chance of a burglary in this area next week"), rather than a continuous number, as in regression (e.g., "There is an average of 2.75 burglaries here next month");

- **Clustering methods** divide crime records into groups of records that are similar mathematically (e.g., "This neighborhood show similar attributes to the neighborhoods labeled as high-crime");
- **Ensemble methods** combine several simple predictive methods to render a concluding overall prediction (Perry et al., 2013).

Depending on the definition, many methods can be considered as data mining techniques. Overall, these techniques are widely used in crime pattern analysis thanks to the development of computing power and the existence of a considerable amount of data.

2.4.4 Near-Repeat Methods

Near-Repeat Methods assume that some future crimes will be very close in time and location to current crimes-that is, areas with recently higher crime rates are likely to experience higher crimes in the immediate future (Perry et al., 2013). For example, Bernasco and Nieuwbeerta (2005) found that burglars always attack groups of nearby targets because offenders are well aware of local vulnerabilities. Similarly, Tita and Ridgeway (2007) found that a shooting committed by a gang can trigger waves of retaliatory violence in the local area (territory) of the rival gang. These findings prompted Mohler et al. (2011) to investigate a self-exciting process used in seismology to show that these methods are well suited for criminological applications.

The Santa Cruz Police Department in California used a version of Mohler's algorithm for a six-month test run that began in July 2011, and after the algorithm proved successful, it was adopted into operations (Thompson, 2011). Comparing crime statistics for the first half of 2012 with statistics for the same period in 2011, the agency reports that property thefts decreased by 19 percent without additional officers or shifts (Jonesi, 2012).

Another method for predicting burglaries is ProMap, which uses recent burglaries and a simple mathematical model to determine which areas are at the highest risk for burglaries (Johnson et al., 2009).

2.4.5 Risk Terrain Modelling

The **Risk Terrain Modeling** (RTM) developed by Caplan and Kennedy (2010) builds on traditional hot spot techniques by incorporating measures that reflect the physical and social environment of the study area (Marchment & Gill, 2021). RTM consist of techniques that:

1. attempt to recognize geographic characteristics that add to crime risk (e.g., bars, liquor stores, certain types of major roads), and
2. make predictions about crime risk based on the proximity of specific locations to these risky features (Perry et al., 2013).

While studies have shown that RTM is an effective forecasting method for identifying hazardous locations and can be a valuable tool for targeting responses, the limitations of RTM are that it does not account for temporal variations in crime locations and may identify areas where crime never occurs (Marchment & Gill, 2021).

2.4.6 Spatio-Temporal Analysis

As we have learned in section 2.1 and 2.2, crime patterns change over time. So far, the presented forecasting models have focused on the crime event, while **Spatio-Temporal Analysis** examines the relationship between crime and the environment. According to Wikle et al. (2019) there are two approaches in spatio-temporal statistical modeling:

- The **descriptive approach** attempts to characterize the spatio-temporal process in terms of its mean and covariance functions. This approach is advantageous when we do not fully understand the mechanism driving the modeled spatio-temporal phenomenon or when we want to understand how the covariates in a regression affect the phenomenon (Wikle et al., 2019).
- The **dynamic approach** assumes that knowing the past and then models how the past statistically evolves into the present and predicts how it will look in the future (Wikle et al., 2019).

Recalling the findings of (Quetelet, 1842) on seasonal crime patterns in subsection 2.1.1, we note that none of the previous methods have considered this phenomenon. The theoretical basis for the seasonality of crime is provided by Routine Activity Theory, in which L. E. Cohen and Felson (1979) found that changes in our routine activities affect crime rates. Depending on the season, our routine activities change to some degree (Andresen & Malleson, 2013). For example, people are more likely to spend time outdoors and leave their homes unguarded during the summer months. Van Koppen and Jansen (1999) analyzed daily, weekly, and seasonal variations in commercial robbery data in the Netherlands and found that there were more commercial robberies during the winter months. In addition, J. Cohen et al. (2003) analyzed quadratic grids across Pittsburgh, Pennsylvania, for different types of violent and property crimes and found that seasonality varied significantly across the city. Spatio-temporal analysis helps to include these seasonal effects in the regressions, which is essential to eliminate known causes of change in crime patterns and lead to better predictions (Perry et al., 2013).

The prediction methods compared in this paper are, at their core, regressive methods. Due to the inclusion of space in ARIMA models, it can be argued that STARIMA models also belong to spatio-temporal analysis. In the following sections, both methods are presented in detail.

2.5 Predictive models explored in this thesis

As we have learned so far, space and time are critical factors in the analysis of crime patterns. In addition, this paper examines methods that use no variables other than the crimes themselves, space, and time. One method that combines all of these variables for prediction is the **Space-Time Autoregressive Integrated Moving Average** (STARIMA) method introduced by Cliff et al. (1975) and generalized by Pfeifer and Deutsch (1980). This method is based on the **Autoregressive Integrated Moving Average** (ARIMA) method introduced by Box and Jenkins (1970) and used for time series forecasting. The STARIMA method combines the following methods to achieve better prediction accuracy:

- **Autoregressive** methods (AR) are regressions that predict the value of a variable (i.e., crime) based on its own past values.
- **Moving Average** methods (MA) are used to smooth the data, so that single outliers are balanced (Shumway & Stoffer, 2000).
- **Integrated** methods (I) refer to a non-stationary trend component in the data (Shumway & Stoffer, 2000).
- **Space-Time** methods (ST) are special time series methods that explicitly account for the linear dependencies between the spatially and temporally lagged variables (Giacomini & Granger, 2004).

The name of the model used reflects the individual components used in the model. For example, STAR refers to a model that uses the space-time and autoregressive part, STARMA adds the moving average, and STMA removes the autoregressive part in the model. The following subsection describes in more detail the individual components that can be used in a STARIMA model.

2.5.1 Autoregressive models

The idea of **autoregressive** models is that present value can be explained as a function of its past values (Shumway & Stoffer, 2000). In the simplest form of regression, a linear regression, the dependent variable y_t is regressed on a vector of the independent variables x_t:

$$y_t = x_t\beta + \varepsilon_t \tag{2.2}$$

In the Equation 2.2, β contains a vector of parameters to be estimated, and ε_t are random errors that are independent and identically distributed (i.i.d) (Kazar et al., 2004). Here, the dependent variable is influenced only by the current value of the independent variable. However, in the case of a time series, it is desirable that the dependent variable can be

influenced by its own past values (Shumway & Stoffer, 2000). A time series is defined as a sequence of observations that follow each other in time (Box et al., 2008).

Autoregressive models are based on the notion that the current value of the time series y_t can be explained as a function of p past values $(y_{t-1}, y_{t-2}, \dots y_{t-p})$, where p indicates the number of steps into the past required to predict the current value. An autoregressive model of order p for a time series y_t, abbreviated as AR(p), can be described as follows in the case of first order AR(1):

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{2.3}$$

where the autoregressive coefficient ϕ is put $|\phi| < 1$ to ensure stationarity (Cliff & Ord, 1981). The AR(p) process can also be written as:

$$\phi(B) y_t = \varepsilon_t \tag{2.4}$$

2.5.2 Moving Average models

Moving Average models help to obtain a smoother time series. Unlike the AR representation, which assumes that the y_t on the left side of the equation are linearly combined, the moving average model of order q, abbreviated MA(q), assumes that the error term ε_t on the right side of the defining equation is linearly combined to form the observed data (Shumway & Stoffer, 2000). The MA model can be written as follows:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{2.5}$$

where there are q lags $(j = 1, \dots, q)$ in the moving average and $\theta_1, \theta_2, \dots, \theta_q$ are parameters that determine the overall pattern of the process. The MA(q) process can also be written as:

$$y_t = \theta(B) \varepsilon_t \tag{2.6}$$

2.5.3 Integrated models

Many empirical time series (e.g., stock prices) behave as if they do not have a fixed mean, which raises the need for models that describe such homogeneous non-stationary behavior (Box et al., 2008). The **Integrated** model can be obtained by assuming some appropriate differences of the stationary processes of AR and MA (Box et al., 2008).

2.5.4 Space-Time models

A **Space-Time** model is a special time series model that explicitly accounts for the linear dependencies between the spatially and temporally lagged variables (Giacomini & Granger, 2004). In the case of spatial data, following the first law of geography (Tobler, 1970), the

assumption that observations are i.i.d. can not be made when spatial autocorrelation is present (Anselin, 1988). Leaving out spatial dependency between observations in a linear regression would be mean that the ϵ_i are not independent of one another and lead to weak models with low prediction accuracy (Shekhar et al., 2002).

When we combine the previously mentioned methods, we obtain an ARIMA model. This model considers only the temporal component. To account for spatial dependence in the data (subsection 2.3.3), we need to include coefficients of spatial dependence in the autoregressive and moving average coefficients (Giacomini & Granger, 2004).

The integration of the spatial dependencies results in the STARIMA model, which is expressed by Pfeifer and Deutsch (1980) as follows:

$$\bigtriangledown^d z(t) = \sum_{k=1}^{p} \sum_{l=1}^{\lambda_k} \phi_{kl} W^{(l)} \bigtriangledown^d z(t-k) + \varepsilon(t) - \sum_{k=1}^{q} \sum_{l=0}^{m_k} \theta_{kl} W^{(l)} \varepsilon(t-k) \tag{2.7}$$

where:

- $z(t)$ is the observation of the random variable at site $i, i = 1, 2, \ldots N$, and time t as a weighted linear combination of past observations and errors, which can be both spatially and temporally lagged;
- $\bigtriangledown$ is the $N \times N$ difference operator matrix;
- p is the autoregressive order;
- d is the number of differences;
- q is the moving average order;
- λ_k is the spatial order of the k^{th} autoregressive term;
- m_k is the spatial order of the k^{th} moving average term;
- ϕ_{kl} is the autoregressive parameter at temporal lag k and spatial lag l;
- θ_{kl} is the moving average parameter at temporal lag k and spatial lag l;
- $W^{(l)}$ is the $N \times N$ matrix of weights for spatial order l;
- and $\varepsilon(t)$ is the random normally distributed error vector at time t.

Different parts of this equation can be set to 0 to remove one of the components: for example, if $d = 0$ the STARIMA model becomes a STARMA model, if $q = 0$ only the autoregressive term remains, leaving a STAR model, and so on. The magnitude of the parameters is written in round brackets next to the name to get a quick overview of the model's parameters. For example, if a STARIMA model (p, q, d) does not use the integrated part, it can be written as $STARIMA(1, 1, 0)$ or as $STARMA(1, 1)$.

In the next subsections the models used for comparison will be presented in detail.

2.5.5 STARMA

In theory the comparison would be ideal between ARIMA and STARIMA. But since there is only an R packages available for STARMA modeling, the integrated part has to be dropped from the spatiotemporal models. This would result in the following adapted equation for STARMA models by Kurt and Tunay (2015):

$$y_t = \sum_{l=1}^{p} \sum_{s=0}^{k_l} \phi_{ls} W^{(s)} y_{t-l} - \sum_{l=1}^{q} \sum_{s=0}^{m_l} \theta_{ls} W^{(s)} \varepsilon_{t-l} + \varepsilon_t \tag{2.8}$$

where l symbolizes the time lag and s the spatial lag. Again $W^{(s)}$ is the spatial weight matrix at the spatial lag s, ϕ_{ls} is the autoregressive and θ_{ls} the moving average parameter at the respective time lag l and space lag s. A closer look at Equation 2.8 helps identifying the structure of the equation: the first half revolves around the AR parameter ϕ_{ls} and the second part around the moving average parameter θ_{ls}. In both half the spatial relationship is represented by the weight matrix $W^{(s)}$.

STARMA models generally identify spatiotemporal dependence between different regions and account for autocorrelation between variables (Zhuang et al., 2017). When the data have not only temporal but also spatial dependencies, STARMA models prove to be a very practical method (Kurt & Tunay, 2015). However, because STARMA models are strongly influenced by the spatial time series data, the model seems to be too parsimonious when there is a lack of many measurement locations (Kamarianakis & Prastacos, 2005).

2.5.6 ARIMA

The ARIMA method was first described by (Box & Jenkins, 1970). Since the previously mentioned STARMA method has an advantage when the data have spatial autocorrelation and there is an R package that allows automatic modeling of the ARIMA method, the comparison is made with STARMA versus ARIMA. The inclusion of the integrated method is sometimes needed when a time series requires a differencing transformation to stabilize the mean of the time series and thus eliminate or reduce trend and seasonality (Cesario et al., 2016).

Assuming a time series $(y_t : t = 1 \dots n)$, where y_t is the value of the time series in the time period t, an $ARIMA(p, d, q)$ model is expressed as follows:

$$y_t^{(d)} = c + \phi_1 y_{t-1}^{(d)} + \cdots + \phi_p y_{t-p}^{(d)} + \theta_1 \varepsilon_{t-1}^{(d)} + \cdots + \theta_q \varepsilon_{t-q}^{(d)} + \varepsilon_t \tag{2.9}$$

where:

- $y_t^{(d)}$ is the d^{th}-differenced series of y_t;
- $\phi_1, \dots, \phi_p$ are the coefficients of the AR part;
- $\theta_1, \dots, \theta_q$ are the coefficients of the MA part;

- $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ are lagged errors;
- ε_t is the white noise and takes the forecast error into account;
- c is a correcting factor (Cesario et al., 2016).

Since automatic forecasts of time series are often needed in business, there have been several attempts to automate ARIMA modeling (Hyndman & Khandakar, 2008). In this work, the *auto.arima*[1] function of the R package **forecast** is used to automatically model ARIMA models for each time period and crime type.

2.6 Performance metrics for regressions used in the thesis

Once the models have delivered their predictions, we need a way to measure their performance. Among the essential means of evaluating model performance in environmental sciences remain statistical comparisons of the model predictions $(P_i; i = 1, 2, \ldots, n)$ with presumably reliable and pairwise matched observations $(O_i; i = 1, 2, \ldots, n)$ (Willmott & Matsuura, 2005). The comparison between predictions P_i and observations O_i leads to an error e_i that can be expressed at the individual level as follows:

$$e_i = P_i - O_i \tag{2.10}$$

When assessing an error on the whole model, not just the individual predictions, there are several metrics that can be used, with each of them having different strengths and weaknesses. The most common metrics are presented in the following subsections.

2.6.1 Mean Absolute Error and Root Mean Square Error

Two popular metrics in evaluating regression results in the field of spatial crime forecasting are the **Mean Absolute Error** (**MAE**) and the **Root Mean Square Error** (**RMSE**) (Kounadi et al., 2020).

The MAE (Equation 2.11) is calculated by adding the absolute values of the model errors ε to obtain a total error, and then dividing the total error by n, assuming all $w_i = 0$ (Willmott & Matsuura, 2005).

$$MAE = \left[n^{-1} \sum_{i=1}^{n} |e_i| \right] \tag{2.11}$$

While the MAE gives equal weight to all errors, the RMSE penalizes variance by giving more weight to errors with larger absolute values than errors with smaller absolute values (Chai & Draxler, 2014). This is because the RMSE (Equation 2.12) is calculated by summing the individual squared errors (not the absolute values) to get the total squared error,

[1] https://www.rdocumentation.org/packages/forecast/versions/8.13/topics/auto.arima

then dividing the total squared error by n (obtaining the mean square error), and finally taking the square root (Willmott & Matsuura, 2005).

$$RMSE = \left[n^{-1} \sum_{i=1}^{n} |e_i|^2 \right]^{1/2} \tag{2.12}$$

Absolute values are highly undesirable in many mathematical calculations, which is a distinct advantage of RMSEs over MAEs (Chai & Draxler, 2014).On the other hand, according to Willmott and Matsuura (2005), the RMSE metric does not describe the average error of model performance well because it does not use the weighting of errors analogous to each value. While RMSE is more appropriate for representing model performance when a normal error distribution is assumed, a combination of metrics including, but by no means limited to, RMSE and MAE is often required to evaluate model performance (Chai & Draxler, 2014).

2.6.2 Coefficient of determination or *R*-squared

Although MAE and RMSE are helpful, they have a common drawback: since their values can range from zero to infinity, a single value from them does not tell us much about the performance of the regression in terms of the distribution of the ground truth elements (Chicco et al., 2021). On the other hand, the R-squared (R^2, see Equation 2.13) outputs a value between $-\infty$ ($R^2 < 0$ indicating a worse fit than the average regression line) and one (i.e., a perfect fit), generating a high score that allows better comparability between models (Chicco et al., 2021). The R^2 or coefficient of determination introduced by Wright (1921) is obtained by dividing the squared difference between the predicted and average values by the squared difference between the actual and average values:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)}{\sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)} \tag{2.13}$$

Chapter 3

Methodology

The third chapter contains the methodology of the work. First, all the software used for the analysis and for the presentation of the results is presented. The study area chosen for the analysis is introduced in the following section, as well as its geography shortly described. Next, the data that was used to model the methods under study are introduced. The same section explains the spatial and temporal resolution decision and the processing steps required to provide the data for the models. The data are then examined to gain insight into the characteristics of the data. In the following section, various spatial weighting matrices are analyzed with the data obtained from the processing steps to help select the appropriate weighting matrix for the models. The modeling workflow of ARIMA and STARIMA is also provided in this section. Finally, the metrics used to compare the models is explained.

To get a better overview of the methodology, the workflow of the analysis is shown in Figure 3.1. The workflow is divided into four main steps: Data gathering, data processing, implementation and comparison of the models.The R code produced as part of this work is also divided into these categories, with the exception of the data gathering step, as this step does not require automation. This subdivision should help further studies find the code for the required step more quickly. All steps are described in detail in this chapter.

3.1 Software

To make the analysis repeatable, all software used in this work is freely available. Data processing, implementation and comparison of the models are performed using RStudio. Maps are created in QGIS to display informative results. GeoDa is used to explore the spatial relationship of the data. Finally, a GitHub repository is created to review the code, contribute, or replicate the workflow of the thesis.

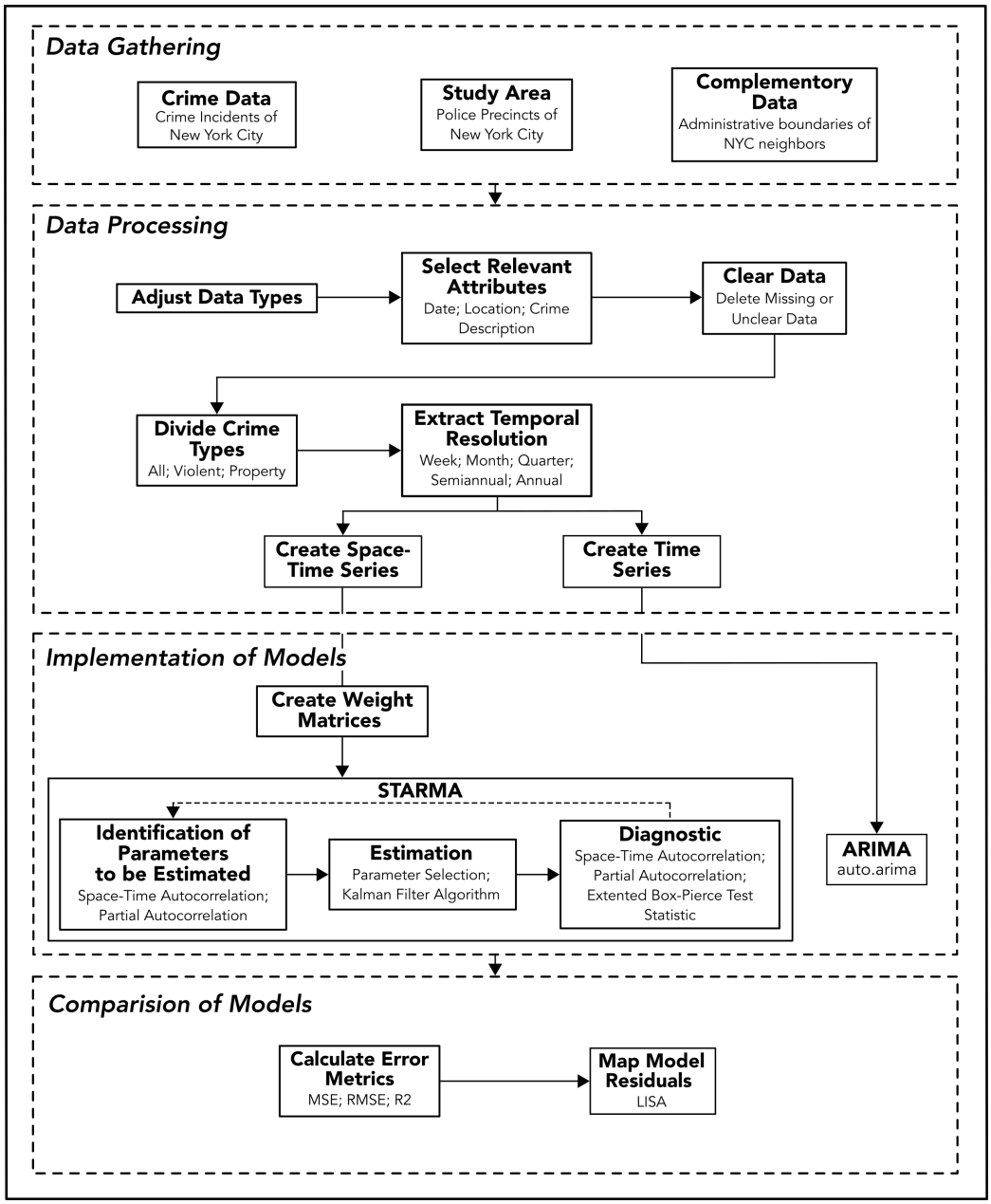


Figure 3.1: Procedure of the analysis part of the thesis.

3.1.1 R-Studio

R[1] is a programming environment for data manipulation, calculation, and graphical representation, which many people use as a statistical system (Venables et al., 2021). **RStudio**[2] is an integrated development environment that provides an interface with many helpful features and tools (Ismay & Kim, 2020).

The decision to use R as the primary programming language for data processing and

[1] https://www.r-project.org
[2] https://www.rstudio.com/products/rstudio/

model implementation is due to the available **STARMA**[3] package and the variety of libraries in its environment that are needed to process the data. Moreover, once the code is written, each step of the process can be reconstructed and replicated. Additionally, RStudio was chosen because it is freely available and facilitates coding in R.

3.1.2 QGIS

QGIS[4] is a free open-source GIS that provides standard functions and features found in other GIS (the QGIS Development Team, 2021). GIS have proven to be fundamental tools in crime prevention programs through their toolbox and map creation (Pawale et al., 2017). QGIS is used to create the maps found in this thesis, with the exception of the residual maps in chapter 4, which were created using the **tmap**[5] package in R, and the LISA cluster maps, which were created using GeoDa.

3.1.3 GeoDa

GeoDa[6] is a free and open source software tool developed to gain new insights from data analysis by examining spatial patterns, developed by Anselin (2017). GeoDa is used in this thesis to study and analyze neighborhood relationships because it is a simple and easy-to-use toolset for exploration. The workflow for exploring and selecting the best representation of the neighborhood relationship is explained in subsection 3.5.1. As mentioned in the previous subsection, GeoDa is used to create LISA cluster maps to facilitate comparison of model outputs.

3.1.4 GitHub

GitHub[7] is an open source version control system (VCS) called Git (GitHub, 2021). In the systematic review of SCF, Kounadi et al. (2020) note that current studies often lack reporting of study experiments, making them difficult to follow. For this reason, a GitHub repository[8] is created for this work. This repository contains all the R code to reproduce the data processing, model implementation, and comparison. In addition, the repository helps to understand the steps required to implement STARMA models and may enhance further research in this area.

[3] https://cran.r-project.org/web/packages/starma/
[4] https://qgis.org
[5] https://cran.r-project.org/web/packages/tmap/
[6] https://geodacenter.github.io
[7] https://github.com
[8] https://github.com/baccanazzo/ExploringCrimeForecastPerformance_STARMA_ARIMA

3.2 Study area

The study area for this thesis is the New York City (NYC) metropolitan area. NYC is located in New York State on the northeast coast of the United States of America. Since the early nineteenth century, it has been the most populous and economically powerful city in the United States (US) (Jacobs et al., 1999). With 8,8 million inhabitants, it is the largest city in the US (the U.S. Census Bureau, 2020). With an area of $778km^2$, NYC has a population density of about 11.300 inhabitants per km^2, which also makes NYC the densest city in the US (the U.S. Census Bureau, 2021).

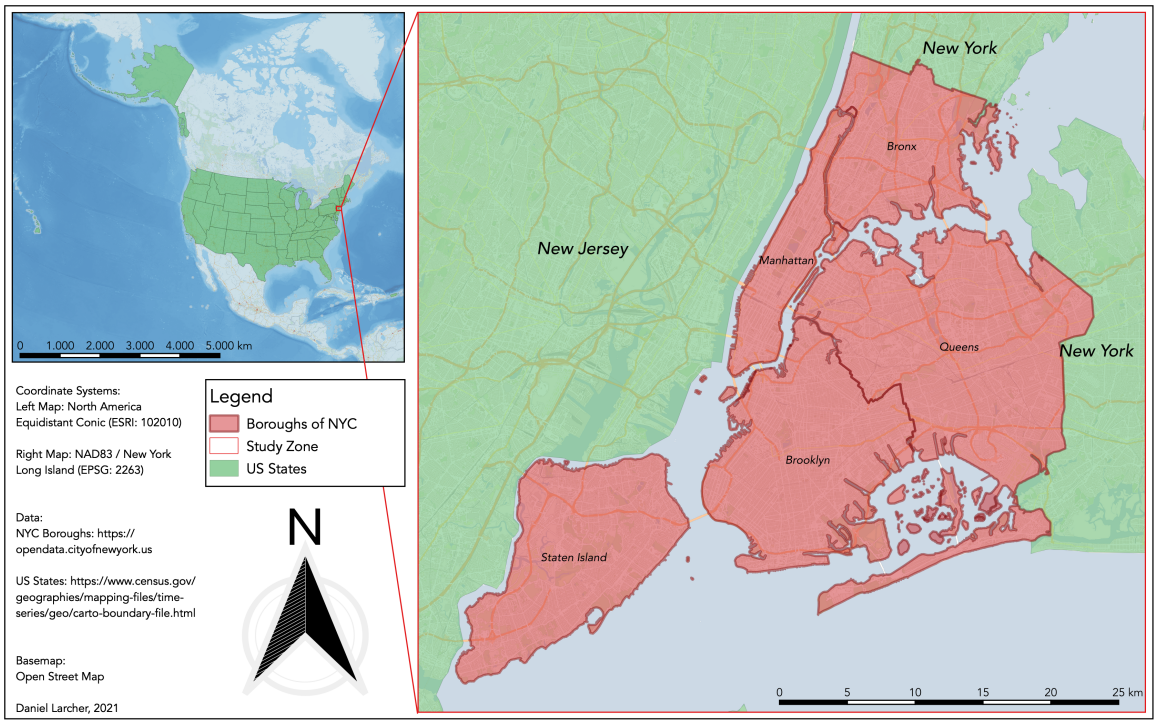


Figure 3.2: Location of the study area and its boroughs.

NYC consists of five boroughs: Manhattan, Bronx, Queens, Brooklyn and Staten Island (Figure 3.2). NYC is located at the mouth of the Hudson River, which separates Manhattan from New Jersey. The East River separates Manhattan and the Bronx from Brooklyn and Queens. Staten Island is also separated from New Jersey and the other boroughs by water. NYC has a border with the U.S. mainland in the Bronx and the only other land border is in Queens with New York State. The unique geography of New York City ensures that all boroughs except Brooklyn and Queens are connected only by bridges or tunnels.

Although the crime rate in NYC has been declining since the 1990s, it is still high enough to provide valuable results for the models studied (Levitt, 2004). Therefore, NYC is the ideal study area for examining long-term spatial crime forecasting using freely available crime data and administrative data.

For the purposes of this thesis, it is assumed that New York City is an isolated city with

no outside influence.

3.3 Data

This section introduces the data used in this work. First, the criminal record dataset is presented, then the spatial resolution of the study is clarified, and finally, the temporal resolutions of the study are revealed.

3.3.1 Crime Records

To implement the models under study, we need to know three things: the type, location, and time of the crimes. The New York Police Department (NYPD) publishes a variety of datasets with incident-level data for NYC [9]: arrest, summons, shooting, and complaint data, both historically (updated annually) and for the current year (updated quarterly). These datasets are reviewed by the Office of Management Analysis and Planning and published annually on the NYC OpenData[10] website (the City of New York, 2021b).

Although an arrest does not automatically mean that the crime occurred, the **NYPD Arrest Data (Historic)**[11] dataset was selected for this study because from the available datasets it best represents the actual crime records. For the purposes of this thesis, the arrest numbers are also considered actual crime numbers. The dataset contains approximately 5.15 million arrests with 19 variables from January 1, 2006 to December 31, 2020. All variables and their descriptions can be found in the Appendix A - Table A.1.

The variables relevant for the analysis are:

- The exact date of arrest for the reported event (Arrest_Date);
- the description of the type of crime (OFNS_Desc);
- and the coordinates of the arrest (X_COORD_CD, Y_COORD_CD).

It should be noted that there is a variable called ARREST_PRECINCT that indicates in which precinct the arrest occurred. When examining the data, it was discovered that some arrests occur outside of the respective police precinct. Therefore, the coordinates of the arrest are used to establish in which precinct the arrest took place.

The dataset contains 88 different types of crimes. In general, the forecasts for property crimes are more precise than for violent crimes (R. Harries, 2003). To better distinguish the different types of crimes, they are divided into three categories: Violent, property, and all crimes. The classification of the different types of crimes into property and violent crimes follows the classification of Rentzelos (2020), which was based on the formal crime typologies of the National Criminal Justice Reference Service[12]. The "all crimes" category

[9] https://www1.nyc.gov/site/nypd/stats/crime-statistics/citywide-crime-stats.page
[10] https://opendata.cityofnewyork.us
[11] https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u
[12] https://www.ojp.gov/taxonomy/term/4426

includes all records in the arrest dataset. The detailed classification can be found in the GitHub repository or in the Appendix B - R Code.

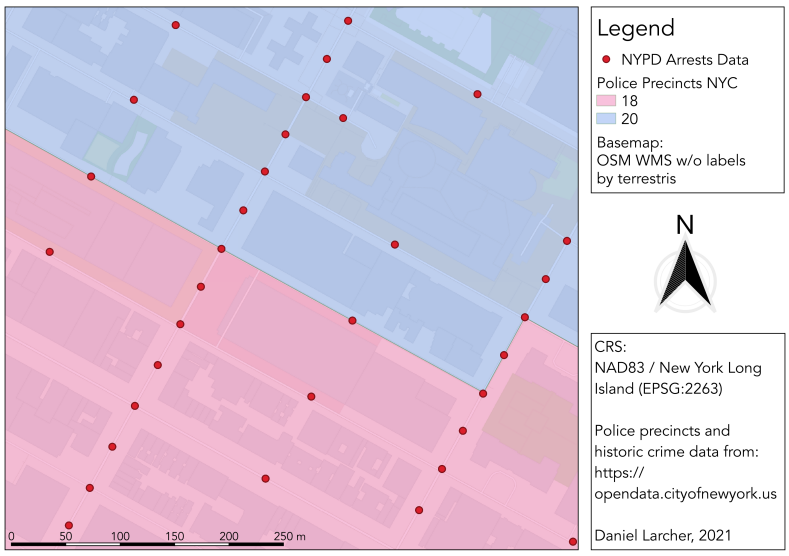


Figure 3.3: Location of crime points in the dataset.

The dataset contains two coordinate pairs, one pair in a projected coordinate system (PCS)[13] and the other in a geographic coordinate system (GCS)[14]. In the PCS of the dataset, distances are measured in feet, while in the GCS, distances are measured in degrees. Since it is necessary to calculate distances between points, the PCS is chosen because distances in feet are easier to handle than in degrees. It should be noted that these coordinates are not the exact coordinates of the arrest, but the coordinates of the center of the nearest street segment or intersection (Figure 3.3), whichever is closer to the actual point of arrest (the City of New York, 2021a). It should also be noted that when crime points are aggregated into police precincts, some of the actual arrests are aggregated in the wrong area. For example, if a precinct's boundary runs along a street and arrests are centered on that street segment, arrests on one side of the street will be summarized in the wrong area, as shown in Figure 3.3. Since the aggregation of points into spatial units is always fraught with problems, this problem can be considered but not circumvented.

In the next subsection, the various available spatial resolutions of NYC are examined and a decision is made as to which spatial resolution should be chosen as the study resolution.

3.3.2 Spatial Resolution

As shown in Figure 2.10, the spatial resolution of the boundaries to which the points are aggregated directly affects the map's message. Similarly, the spatial resolution chosen to

[13] https://epsg.io/2263
[14] https://epsg.io/4326

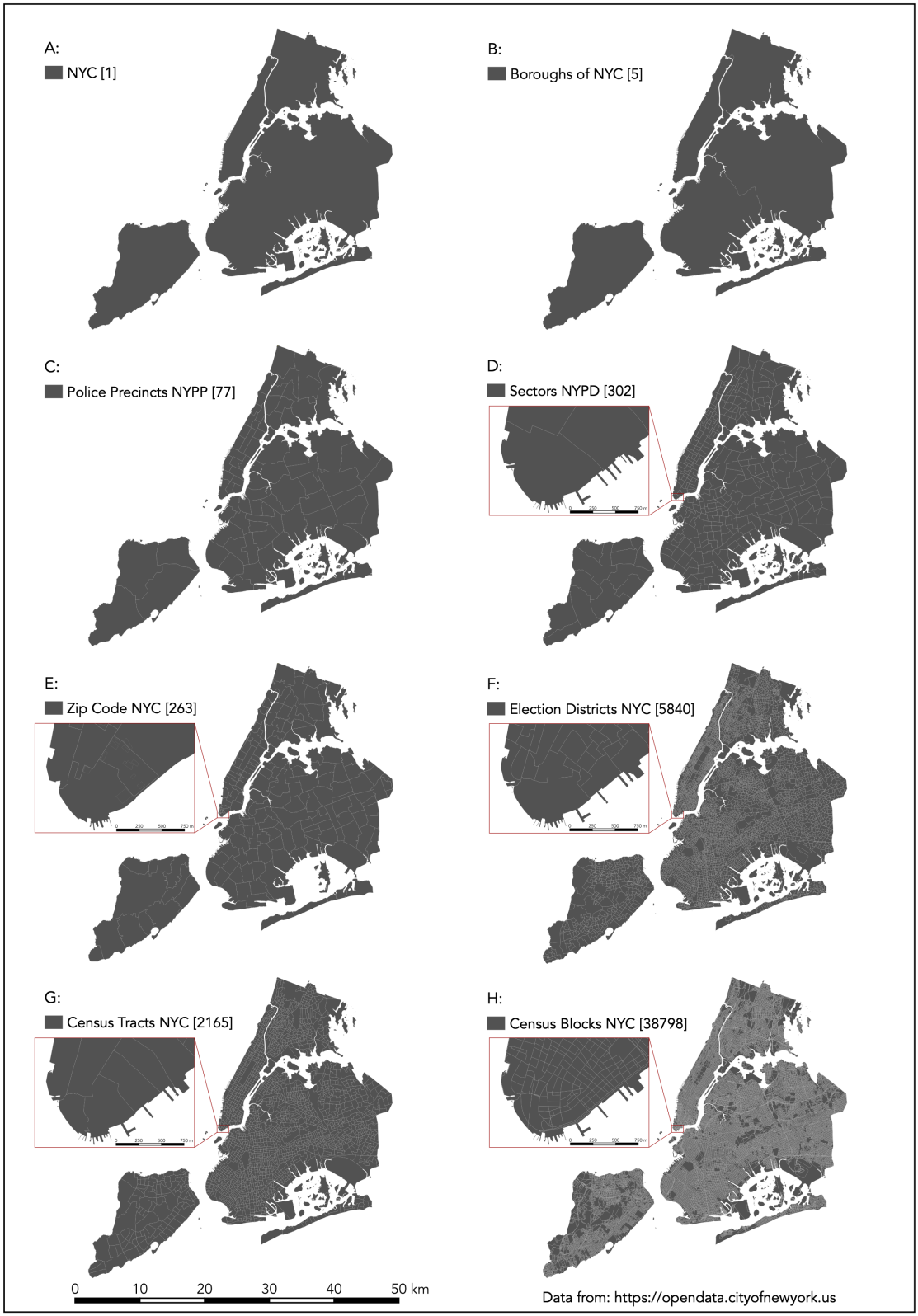


Figure 3.4: Different administrative boundaries of NYC and the NYPD.

aggregate the points for the models affects the results. The use of raster maps is employed to compensate for the different shapes and sizes of administrative boundaries. However, they also have a disadvantage. A raster cell has no connection to the real world, and if a raster is placed over, i.e., NYC, some of the raster areas are over water or actual boundaries such as the city limits. Since the goal of long-term forecasting in this thesis is to help policy makers make strategic decisions, the use of administrative boundaries is preferred.

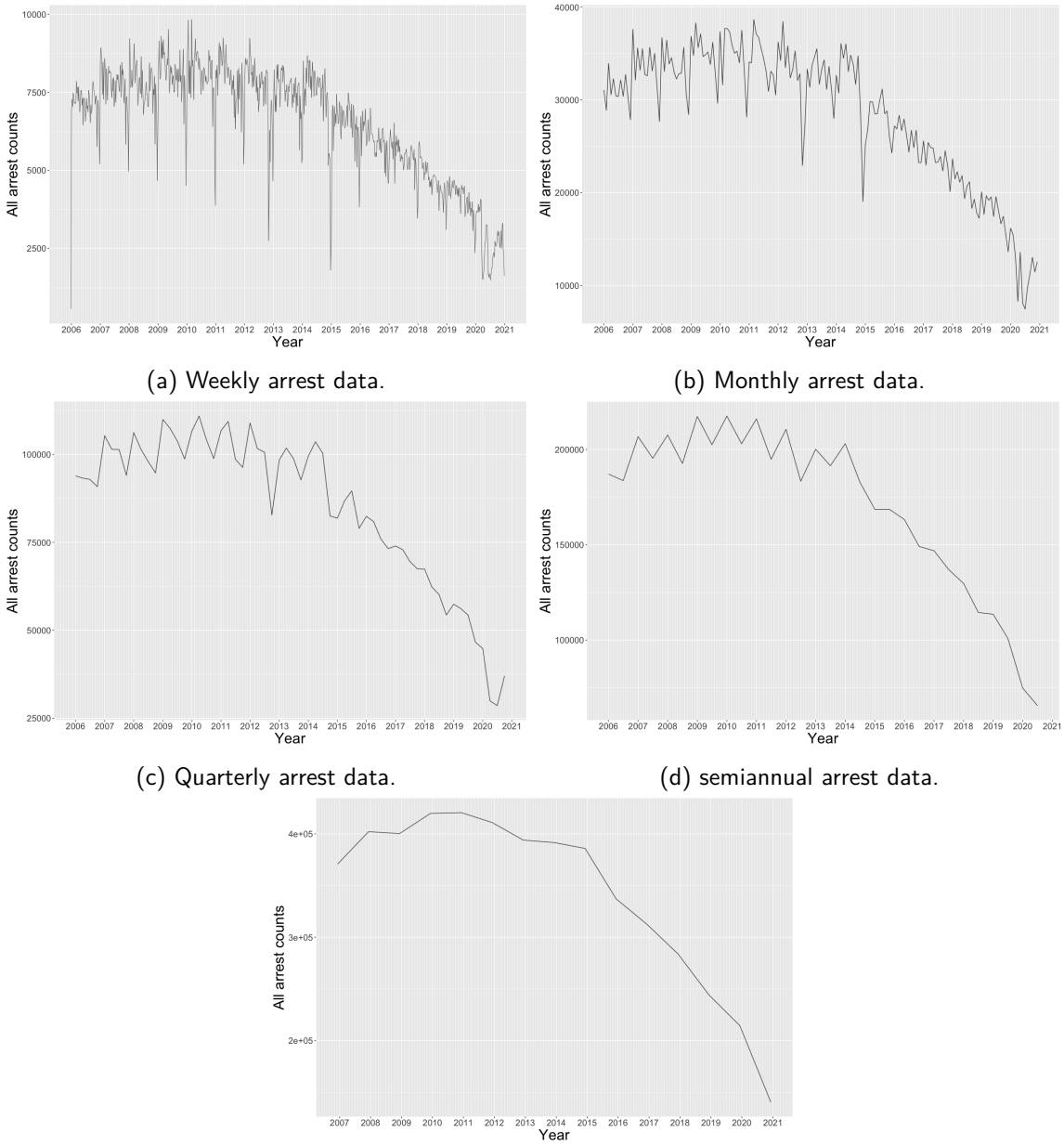
NYC has several administrative boundaries of varying sizes, the most important of which are shown in Figure 3.4. The city boundaries (map A) and boroughs (map B) are not used for the study because of their size and the resulting lack of information gain. Ideally, one would think that the smaller the resolution, the more accurately one could determine the high crime areas. However, there are some computational limitations to using STARMA models. This limitation is due to the structure of the input tables: Each row represents a crime data set, and each column represents a spatial unit. Using the census blocks (map H) would show the blocks with high crime rates, but the high number of almost 39.000 polygons would make it impossible to compute the models due to the size of the input table. Rentzelos (2020) also employed STARMA models in his study and used zip codes as the spatial resolution. Map E of Figure 3.4 shows that NYC has "only" 263 zip codes, and Rentzelos already reported computational problems. His findings leave only three options for the spatial resolution of the study: police precincts (map C), NYPD sectors (map D), and zip codes (map E).

Long-term crime forecasts are used for strategic planning (Perry et al., 2013). Hence, the choice of spatial resolution should ideally fall on boundaries that are useful for police enforcement. Consequently, the boundaries used by the NYPD seem to be the best choice. The choice between police precincts and sectors falls on police precincts first, as sectors, with their 302 areas, are likely to cause computational problems in the analysis. Furthermore, as noted in section 1.4, Gorr et al. (2003) found that a crime count of 30 or more is needed to achieve an absolute forecast error of 20% or less. After pre-processing the crime data, the number of monthly violent and property arrests was below 30 in only a few police precincts, further solidifying the selection because a lower resolution would result in a high number of areas with no arrests.

3.3.3 Temporal Resolution

As Gorr and Harries (2003), and Perry et al. (2013) note, long-term forecasts are necessary for strategic planning. But neither they nor other authors define what unit of time is short- or long-term. As is well known, time is relative, even in crime forecasting. From the extensive literature review, it can be concluded that most authors define the short term as less than one month, while most long-term forecasts have a temporal resolution of up to one year. In order to compare the models comprehensively, it is important to consider the temporal resolution of all categories. This led to the selection of the following temporal

resolutions, with the effects on the data shown graphically in Figure 3.5:



(a) Weekly arrest data.

(b) Monthly arrest data.

(c) Quarterly arrest data.

(d) semiannual arrest data.

(e) Yearly arrest data.

Figure 3.5: Effects of the temporal agglomeration on the data.

(a) One week: representing short-term predictions

(b) One month: representing medium-term predictions

(c) Quarterly (three months): representing long-term predictions

(d) Semiannual (six months): representing long-term predictions

(e) One year: representing long-term predictions

Agglomeration directly affects forecasts because each method is modeled on the underlying data. As the time lag increases, more individual observations are added to a single one, reducing the variation and number of observations in the data. Because of this smoothing with increasing time lag, the overall trend is easier to follow, but may hide some changes in the data. An example of this phenomenon is the last year of data that was affected by the COVID-19 pandemic. In Figure 3.5a, 3.5b, and 3.5c, the data show a rapid decline in arrests at the beginning of 2020, followed by an increase in the later half of the year. This trend is not evident in the semiannual and annual arrest data, as shown in Figure 3.5d and 3.5e. This phenomenon means that the model also does not "see" the trend and therefore may produce less accurate forecasts.

3.3.4 Processing steps

In order to process the data using the models to be analyzed, the crime data must be processed in a specific way (Figure 3.6b) so that the temporal observations are row-by-row and each column corresponds to a location (Cheysson, 2016). In addition, the data must be divided into the categories of violent and property crimes. In this subsection, all necessary steps from the original datasets to the modelable data-set are explained. Figure 3.6 illustrates the difference between the input and the output data-set of all crimes after the processing steps.

	ARREST_KEY	ARREST_DATE	PD_CD	PD_DESC	KY_CD	OFNS_DESC
1	32311380	06/18/2007	511	CONTROLLED SUBSTANCE, POSSESSION 7	235	DANGEROUS DRUGS
2	192799737	01/26/2019	177	SEXUAL ABUSE	116	SEX CRIMES
3	193260691	02/06/2019	*NA*	*NA*	*NA*	*NA*
4	149117452	01/06/2016	153	RAPE 3	104	RAPE
5	190049060	11/15/2018	157	RAPE 1	104	RAPE
6	24288194	09/13/2006	203	TRESPASS 3, CRIMINAL	352	CRIMINAL TRESPASS
7	189182271	10/24/2018	153	RAPE 3	104	RAPE
8	196324211	04/23/2019	157	RAPE 1	104	RAPE
9	196785901	05/04/2019	175	SEXUAL ABUSE 3,2	233	SEX CRIMES
10	197554056	05/23/2019	175	SEXUAL ABUSE 3,2	233	SEX CRIMES

(a) Extract of original NYC Arrest Data.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	300	458	290	219	311	270	316	1087	148	374	310	171	61	534	185	662	216	559	366	504
2	335	433	290	247	281	283	284	958	155	331	282	157	54	482	216	619	191	563	332	470
3	356	476	323	282	422	277	348	1088	173	349	346	197	96	529	251	783	225	540	453	518
4	333	444	355	268	308	283	335	1146	144	332	277	178	72	542	220	634	150	583	382	404
5	329	449	386	249	347	292	349	1168	159	399	281	182	111	538	286	602	160	524	384	501
6	341	399	310	259	347	292	331	1179	157	480	324	175	91	467	232	590	196	533	362	494
7	337	441	378	199	245	253	377	1215	172	418	297	132	57	496	210	592	243	491	358	528
8	355	447	387	303	355	335	374	1222	162	423	293	159	60	542	223	708	198	534	424	457
9	344	413	338	249	373	405	334	1167	157	319	321	171	78	447	253	622	229	560	371	490
10	348	420	350	282	385	320	340	1354	159	377	327	180	51	528	328	706	212	542	387	602

(b) Extract of data-set ready for STARMA.

Figure 3.6: Difference between the original and the processed data set.

There are nine processing steps required to transform the original datasets. All steps are written in R to avoid manual processing errors and to facilitate replication of the analysis. As mentioned earlier, the detailed R code can be found in the GitHub repository or in Appendix B. The following is a brief overview and explanation of each step using the monthly category as an example:

Step 1: **Input data**
In this step the original datasets (Arrest Data = 5.153.369 obs.; Police Precincts = 77) are loaded into the workspace.

Step 2: **Select relevant attributes**
The relevant variables (type of offense, date and location) for the analysis are selected. Also, some data manipulations on the data types are required.

Step 3: **Clear N.A. and strange data**
Missing data or unclear crime descriptions (N.A. and F.C.A. P.I.N.O.S.) are deleted (4.529.353 obs.).

Step 4: **Spatial join of crime data & police precincts**
The police precincts are joined to the arrest data. Since some entries were not joined into police precincts, the variable ARREST_PRECINCT from the arrest data was used to fill the missing values.

Step 5: **Categorize arrest data into crime types**
Depending on the offense description (71 unique offense descriptions, 16 in the property offense category, 18 in the violent offense category), a binary code is assigned in the appropriate column.

Step 6: **Extract information about temporal resolution**
The daily data is transformed into weekly, monthly, quarterly, semiannual and annual data.

Step 7: **Group & sum the data**
Arrests for each individual police precinct and month are summed and arrests for each crime type are then totaled (11,088 observations).

Step 8: **Transpose data for the implementation into STARMA**
In order to have "months" as rows and "police precincts" as columns (144 obs.).

Step 9: **Replace geometry of polygons with their centroid point**
To speed up the calculation when calculating neighborhood relationships.

Once processed, the data is ready for exploration and modeling of the methods to be analyzed.

3.4 Data exploration

In this section, we briefly discuss the data sets used for crime forecasting. First, we consider the number of crimes in the entire study area. Using Figure 3.7, we can compare the monthly arrest counts and annual averages for the crime types "all" (Figure 3.7a), "property" (Figure 3.7b), and "violence" (Figure 3.7c). A comparison of the three charts shows that the number of all crimes and that of violent crimes follow a similar trend. Both types of crime increase slowly until the end of 2010 and then decline steadily until 2020. The number of property crimes, on the other hand, rises steadily until it reaches a peak in October 2014 and then falls again. An interesting observation is that in all three annual averages, there is a significant decline in 2015 and 2020. While the decline in 2020 is due to the COVID-19 pandemic and subsequent restrictions in March 2020 (Francescani, 2020), the decline in arrests in 2015 is due to the murder of two NYPD officers on December 20, 2014, and officers' resulting fear for their own safety (Bertrand, 2015). This underscores the fact that the data used in this analysis are not the actual crimes themselves, but the arrests recorded by the NYPD. While this is not perfect, it is the best representation of actual crimes available for this study.

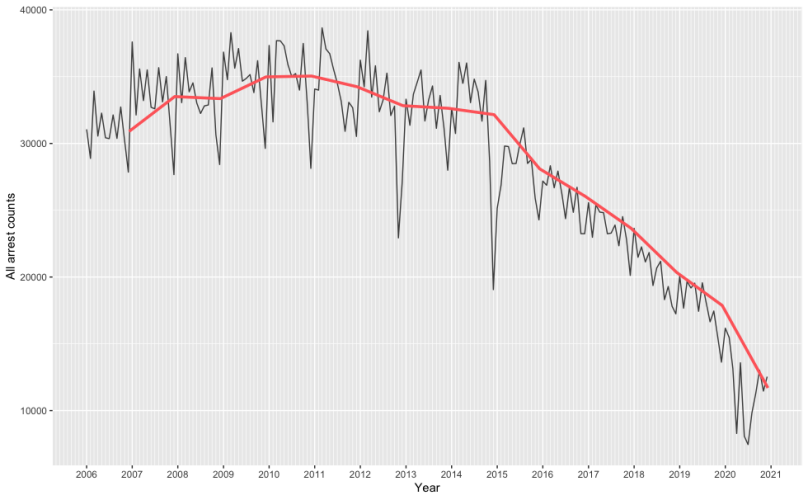
Looking at the monthly arrest counts in the three charts in Figure 3.7, we see that the crime figures for the different types of crime develop differently over the year. While for both violent and property crimes, the number of crimes is lowest at the end of the year, the number of violent crimes is higher in the first six months of the year than in the last. For property crimes, on the other hand, the number of crimes is highest at the beginning and middle of the year. When looking at the monthly crime numbers for all crimes, it is difficult to see a seasonal trend.

3.5 Modeling workflow

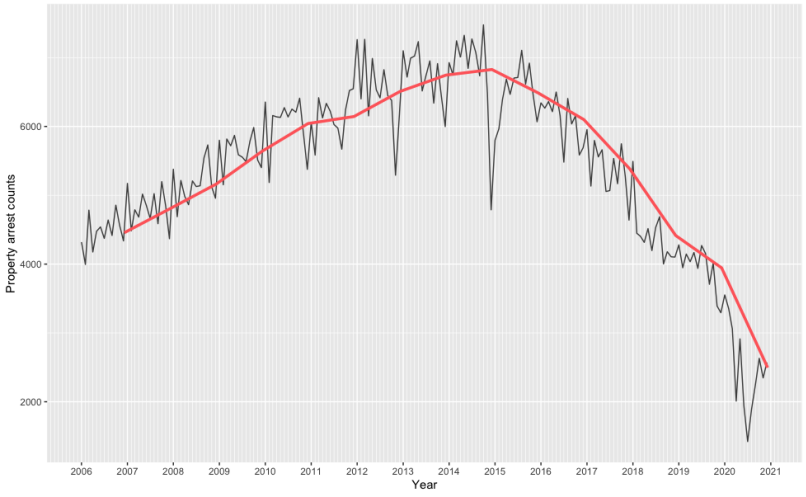
In this section, the modeling workflow is presented. First, the spatial weighting matrices must be defined. To do this, several options must be considered and analyzed for how accurately they represent the spatial relationships of the data. The spatial weighting matrices can be used to model the STARMA models, using the three-step iterative procedure. Finally, the workflow to automatically estimate the ARIMA models is described.

3.5.1 Creation of neighborhood relationships and spatial weights

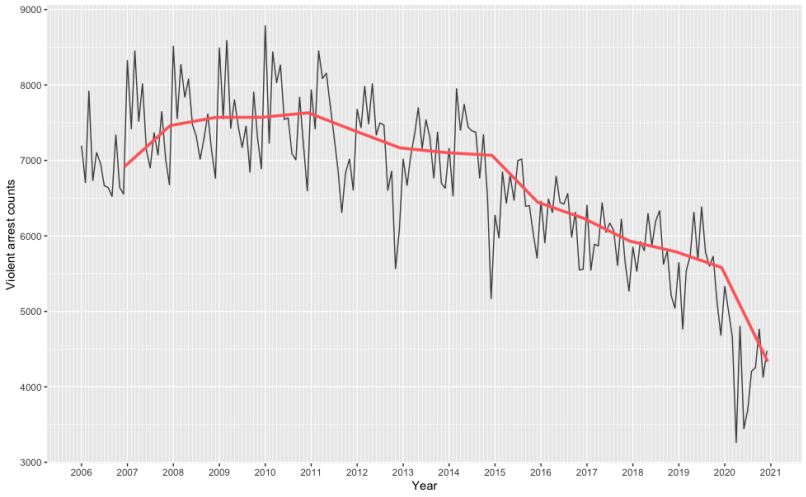
Spatial weight matrices are an essential part of the STARMA method. As described in subsection 2.3.3, when creating neighborhood relationships and spatial weights, there is no one-size-fits-all solution. In this subsection, the initial selection of potential neighborhood relations is presented and the process of selecting the final neighborhood relations is described. Then, the process of creating the spatial weighting matrices is presented.



(a) All monthly arrest counts and yearly means of NYC 2006-2020.



(b) Property monthly arrest counts and yearly means of NYC 2006-2020.



(c) Violent monthly arrest counts and yearly means of NYC 2006-2020.

Figure 3.7: Comparison of monthly arrest counts and yearly means per crime type in NYC 2006-2020.

Using GeoDa and the datasets created in the processing steps, various neighborhood relationships will be explored to determine which is the best representation for the underlying study data. Since the boundaries of the police precincts are on land, precincts that are separated by water but connected by a bridge would not be considered neighbors when using the contiguity-based method. Therefore, police precincts that are connected by bridges in the real world were also connected in the digital file using satellite imagery in QGIS. Then, three different methods were used to create the neighborhood relationships and several different settings. Table 3.1 provides an initial selection of potential neighborhood relationships, as well as some statistics on the settings. This initial selection was made by looking at the study area and trying to represent all possible real neighborhood relationships.

Table 3.1: Selection of potential neighborhood relationships

#	Method	Settings	Number of neighbors				% non zero
			min	max	mean	median	
1	Distance based	27.783 ft	1	33	19,64	22	25,5%
2		30.000 ft	2	38	22	25	28,57%
3		40.000 ft	2	52	33,04	36	42,91%
4	k-nearest Neighbor	1	1	1	1	1	1,3%
5		2	2	2	2	2	2,6%
6		3	3	3	3	3	3,9%
7		4	4	4	4	4	5,19%
8	Contiguity based	1st order queen	1	7	4,7	5	6,11%
9		2nd order queen	4	22	12,73	13	16,53%
10		1st order rook	1	7	4,57	5	5,94%
11		2nd order rook	4	22	12,47	12	16,19%

The distance-based method uses the centroid of each precinct to calculate the distance between them. Since each precinct must have at least one neighbor, the minimum distance that can be used is the minimum distance between the two farthest centroids. In the spatial dataset used in this work, the minimum distance between the two farthest centroids is 27.783 feet and represents the distance between the centroids of police precincts 121 and 123 on Staten Island. Looking at the police precinct in Map C of Figure 3.4 or Moran's I values in Table 3.2, it is clear that a purely distance-based relationship is problematic in this case. Because police districts have a high variance in terms of size, large districts have a small number of neighbors while small districts have a large number of neighbors. Even using the minimum distance, the average number of neighbors in the districts is almost 20. Also, the high number of neighbors negatively affects the spatial autocorrelation of the distance-based relationship, as shown in Table 3.2. The distance-based methods have the lowest Moran's I value among all methods. This can be explained by the fact that the high number of neighbors ensures that there are many observations with different values that lower the Moran's I value, which means that there is less (or no, if $I = 0$) spatial

autocorrelation.

Table 3.2: Spatial autocorrelation of potential neighborhood relationships of each analyzed dataset (all, property (p) and violent (v) crime) in the first week (w), first month (1m), first quarter (3m), first half (6m) and the whole year (y) of 2011.

Case	**Moran's *I* of method #**										
study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
w all	0,079	0,086	0,042	0,189	**0,203**	0,201	0,201	0,195	0,123	0,202	0,128
w p	**0,061**	0,050	0,034	-0,101	0,045	0,021	0,026	0,022	0,010	0,026	0,008
w v	0,189	0,174	0,094	0,341	0,323	0,321	0,370	0,376	0,314	**0,380**	0,322
1 m all	0,078	0,089	0,039	0,243	**0,262**	0,255	0,248	0,246	0,145	0,252	0,113
1 m p	0,053	0,056	0,039	-0,039	**0,092**	0,060	0,085	0,081	0,028	0,087	0,026
1 m v	0,221	0,204	0,097	0,450	0,426	0,447	0,439	0,454	0,364	**0,459**	0,167
3 m all	0,090	0,100	0,041	0,261	**0,278**	0,277	0,260	0,260	0,162	0,264	0,366
3 m p	0,048	0,054	0,040	-0,046	**0,093**	0,062	0,083	0,081	0,027	0,083	0,027
3 m v	0,219	0,206	0,098	**0,484**	0,455	0,463	0,468	0,481	0,375	0,483	0,380
6 m all	0,075	0,085	0,030	**0,288**	0,287	0,278	0,264	0,260	0,154	0,265	0,161
6 m p	0,041	0,046	0,039	0,007	**0,124**	0,092	0,122	0,100	0,061	0,104	0,063
6 m v	0,216	0,204	0,094	**0,490**	0,460	0,463	0,469	0,473	0,366	0,477	0,370
y all	0,064	0,073	0,021	**0,292**	0,290	0,278	0,265	0,253	0,145	0,259	0,151
y p	0,040	0,046	0,037	-0,011	**0,113**	0,081	0,109	0,095	0,046	0,098	0,047
y v	0,199	0,188	0,081	**0,477**	0,440	0,443	0,450	0,448	0,340	0,451	0,344

A method that is also based on distance but restricts the number of selected neighbors is the k-nearest neighbor method (# four to seven in Table 3.1 and 3.2). Given the chosen number of neighbors (k), this method selects the k-nearest neighbors based on the distance of the centroids. Table 3.2 shows that the k-nearest neighbor method is the one that has the highest number of Moran's I values closest to one (values highlighted in bold). The setting with two nearest neighbors has the most highest Moran's I values, followed by the setting with one nearest neighbor. Interestingly, the nearest neighbor setting is the only setting in the comparison that produces negative Moran's I values, indicating a scattered pattern. As learned in section 2.3, Moran's I values near 0 are generally indicative of spatial randomness. When the number of nearest neighbors is increased to three or more, the spatial autocorrelation decreases, as can be seen in Table 3.2.

The contiguity-based method (# eight to eleven in Table 3.1 and 3.2) does not use distance to determine neighbors, but whether they share a common boundary. As with the previously used methods, the setting with more neighbors selected tends to result in lower spatial autocorrelation values than the methods with a smaller number. The contiguity-based method with the first-order tower setting yields the highest spatial autocorrelation value in the violent crime category for the weekly and monthly time periods.

It is important to note that the Moran's I values of Table 3.2 represent only one time period of the entire time period. As mentioned earlier, the method with the highest Moran's

I values is the k-nearest neighbor method. The crime category with the highest spatial autocorrelation is violent crime, while property crime has the lowest spatial autocorrelation. In general, the spatial autocorrelation is weak across all different time periods and crime types, especially for all crimes and property crimes, suggesting that the selection of police districts as study areas may be too large. Since the Moran's I calculated in Table 3.2 measures global spatial autocorrelation, the high variance in spatial autocorrelation between study areas could be another reason for the low SA.

Since the 2−nearest neighbor method consistently yielded a high Moran's I among all neighborhood relationships studied, this method is selected as the neighborhood relationship for this study. As mentioned in subsection 2.3.3, the neighbors must be weighted to create a spatial weight matrix. For this study, all neighbors are weighted equally.All neighborhood relations and weights studied in GeoDa are created with the **spdep**[15] package in R and used in the implementation of STARMA models. The package contains several functions for creating spatial weights. The function *dnearneigh* creates a neighborhood relation based on distance, while *poly2nb* creates a neighborhood relation based on contiguity. The *knearneigh* function can be used to specify which k number of neighbors should be considered as neighbors.

3.5.2 Modeling of STARMA

Up to this point, the data has been processed and the spatial weights created. Now a STARMA model must be fitted for each crime type and time period. This is done using the aforementioned three-step iterative model building process developed by Box and Jenkins (1970) and extended to space-time modeling by Pfeifer and Deutsch (1980), using the STARMA package by Cheysson (2016).

In order to apply the three-step iterative model building process for STARMA models, the space-time series obtained after the processing steps must be centered and scaled so that its mean is 0 and its standard error is 1. The space-time series must be centered because the *starma* function does not estimate an intercept coefficient (Cheysson, 2016). The centering and scaling is done with the function *stcenter*.

Then, the three-step iterative model-building process is repeated until the best-fit model is found for each type of crime and time period:

1. Identification: Using the space-time autocorrelation function (*stacf*, for the AR parameter ϕ) and partial autocorrelation function (*stpacf*, for the MA parameter θ) to identify the parameters to be estimated.

2. Estimation: The parameters are estimated with the *starma* function.

3. Diagnostic: The *stacf*, *stpacf* and *stcor.test* functions can be used to determine if the residuals resemble white noise. The function *stcor.test* computes an expansion of

[15] https://cran.r-project.org/web/packages/spdep/

the Box-Pierce test statistic to accept or reject the non-correlation of the individual observations Cheysson (2016).

These three steps are repeated until the residuals ideally resemble white noise, i.e., they are no longer contiguous. The spatial weighting matrix best found in Table 3.1 is used to model STARMA.

The STARMA package uses a Kalman filter algorithm by Cipra and Motyková (1987) where the parameters are specified as the state vector of the state space system, which results in the iteration of the algorithm directly estimating the parameters. This makes the algorithm extremely efficient in terms of time and computational capacity, since no optimization routine is required (Cheysson, 2016). Since AR and MA orders can be defined with a binary matrix, the user can estimate parameters even at high temporal and spatial lags, which is a strength of the estimation function (Cheysson, 2016). The following Figure 3.8 shows an example of a matrix used to define the AR parameters of the weekly STARMA model of violent crimes. This matrix would indicate to the algorithm that the AR parameters at spatial lag zero (i.e., the area of prediction) would be of order one through ten (except for order seven), and at spatial lag one (i.e., the neighboring areas of prediction) a first-order AR parameter should be estimated.

	V1	V2
1	1	1
2	1	0
3	1	0
4	1	0
5	1	0
6	1	0
7	0	0
8	1	0
9	1	0
10	1	0
11	0	0
12	0	0
13	0	0
14	0	0

Figure 3.8: Matrix of AR parameters to be estimated at spatial lags 0 (V1) and 1 (V2) and temporal lags 1 to 14.

After estimating the parameters of the model, the error metrics of the obtained model are calculated. The calculation of the error metrics is described in section 3.6, while the next subsection presents the modeling process of ARIMA.

3.5.3 Modeling of ARIMA

The modeling of ARIMA is performed with the *auto.arima* function of the R package **forecast**[16]. This function fits the best ARIMA model to a univariate time series (Hyndman & Khandakar, 2008). The algorithm uses the Akaike's Information Criterion (AIC) or the Bayesian information criterion (BIC) to determine which model is the best-fit model (Hyndman et al., 2021). The AIC was developed by Akaike (1974), while the BIC was developed by Schwarz (1978). Both criteria are an extension of maximum likelihood estimation, with the difference between AIC and BIC being the multiplication of the dimension (Schwarz, 1978). By minimizing the BIC and AIC, the best parameter values are determined (Cesario et al., 2016). The STARMA package also computes the BIC for each model. Stoica and Selen (2004) found that the BIC is generally a better fit than the AIC.

Since the estimation of the parameters for the ARIMA models is done automatically, it is much faster than the manual estimation of the parameters. To allow comparability, the input data are centered and scaled like the input data of the STARMA models. Then the space-time series is split into 77 time series, one time series for each study area. This step is necessary because an ARIMA model is designed for time series. To allow further comparability, the maximum order option of the *auto.arima* function is set to the maximum order used in the corresponding STARMA model. In this way, 77 ARIMA models are created and their residuals are collected in a table to obtain the same output table as the STARMA model and make them comparable.

3.6 Comparison of the models

After the parameters of STARMA and ARIMA are estimated, the residuals or errors of the model for each police precinct and time period are written to a table. This table has the same layout as the input table with the observed values. When the residuals of the models are added to the actual values, the predicted values are obtained. With these two tables containing the actual and predicted values for each police precinct and time period, the error metrics mentioned in section 2.6 can be calculated.

Before the final computation of the error metrics, some of the observed and predicted values must be deleted, more precisely the first observations for which the model does not have enough past data to compute them correctly. For example, if the highest AR parameter of a model is of order ten (ϕ_{10}), the first ten (temporal) observations are excluded from the calculation of the error metrics. This ensures that the error of each model is based on the selected parameters and not on the missing data. Since there may be different maximum orders of the ARIMA models, the maximum order allowed in the *auto.arima* settings is used to subtract the number of observations. This ensures that the final error metrics of the STARMA and ARIMA models are calculated with the same number of observations.

[16] https://cran.r-project.org/web/packages/forecast/

Finally, the error rates and their standard deviations are calculated for each police precinct, and the mean of the calculated metrics is written to a final table to compare the models. Typically, the model that produces lower error metrics outperforms the other, but it is also optimal if these errors are not clustered in space.

To further compare the models, the standard deviation (s.d.) of their residuals is mapped. Since these maps can only be interpreted visually and could be subjective to the interpreter, a LISA cluster map is also computed for each model analyzed. Spatial weights are required to compute the LISA cluster map. Since the $2-$nearest neighbor method was chosen because it best represents the spatial dependencies in the data (subsection 3.3.2), it is also chosen as the weighting matrix for the LISA cluster maps. The LISA cluster maps are based on the local Moran value and were created using GeoDa.

All STARMA models are compared to the ARIMA model using three scoring methods: the error metrics, the s.d. of residuals, and the LISA cluster maps. Each of these methods counts as one point for the model that outperforms the others.

Chapter 4

Results and discussion

In this chapter, the results of the analysis of this thesis are presented and discussed. First, a general observation will be given. Then, for a better overview, the chapter is divided into sections representing the individual periods. Finally, the research questions are answered.

4.1 General observations after modeling

Estimating the parameters for the STARMA model proved to be more difficult than expected. The main challenge was to fit the residuals of all models to resemble white noise. Because the overall trend of the data fluctuates throughout the study period, estimating the STARMA parameters was very time consuming.

As seen in Figure 3.7 and explained in section 3.4, the general trend in the arrest data in New York is upward in the early years of the study period, followed by a steady decline and a significant drop in the trend in 2015 and 2020. In diagnosing the process of the STARMA models, the residuals of each model showed a significant negative spatio-temporal autocorrelation at a little bit over half of the time steps (values outside the dotted blue line in Figure 4.1). This time step corresponds to the first significant decline in arrests in 2015. It is important to remember that the models examined in this study are essentially equations that attempt to best represent the underlying data. Therefore, high variance in the data leads to high variance in the errors of the models. One solution would be to cut the data set at certain points, such as at 2015, to get a more linear trend. However, this solution would not be applicable to the longer time spans of six-month and annual models because there would not be enough study data to make valuable forecasts. Since the ARIMA models have the same problem in parameter estimation, it was decided to keep the entire study data for all models.

To further improve the parameters of STARMA, the BIC of the model was used. In general, the model with the lower BIC is preferred. In addition, the error metrics of STARMA were calculated and compared with the error metrics of the corresponding ARIMA model. The goal was to estimate the parameters of STARMA until they had similar error

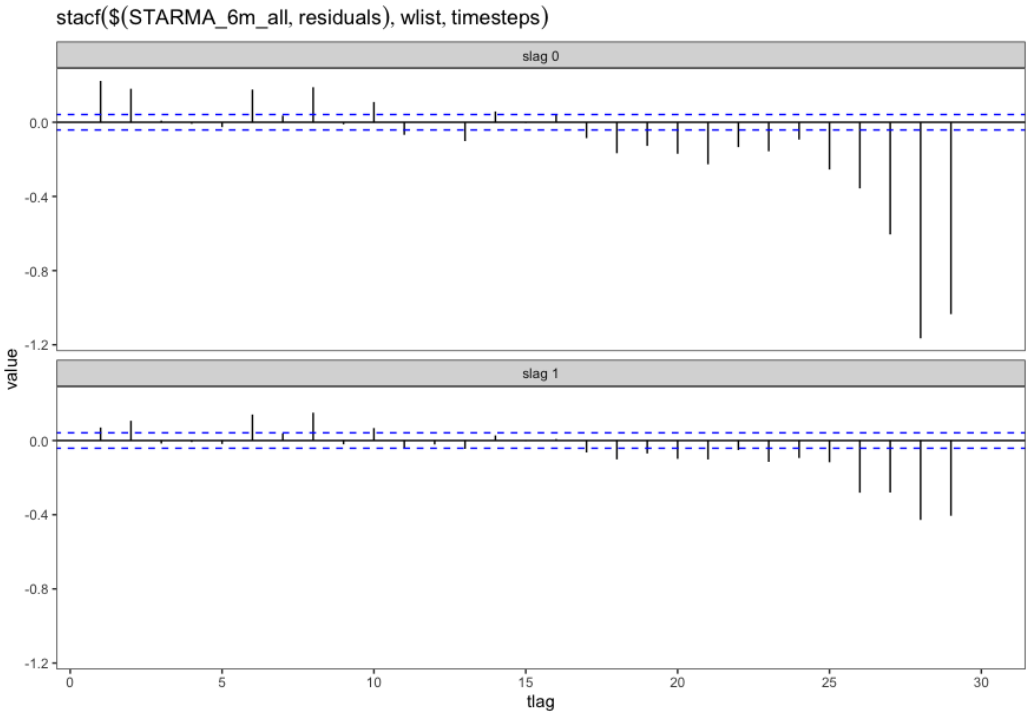


Figure 4.1: Spatio-temporal autocorrelation function of STARMA 6m all residuals.

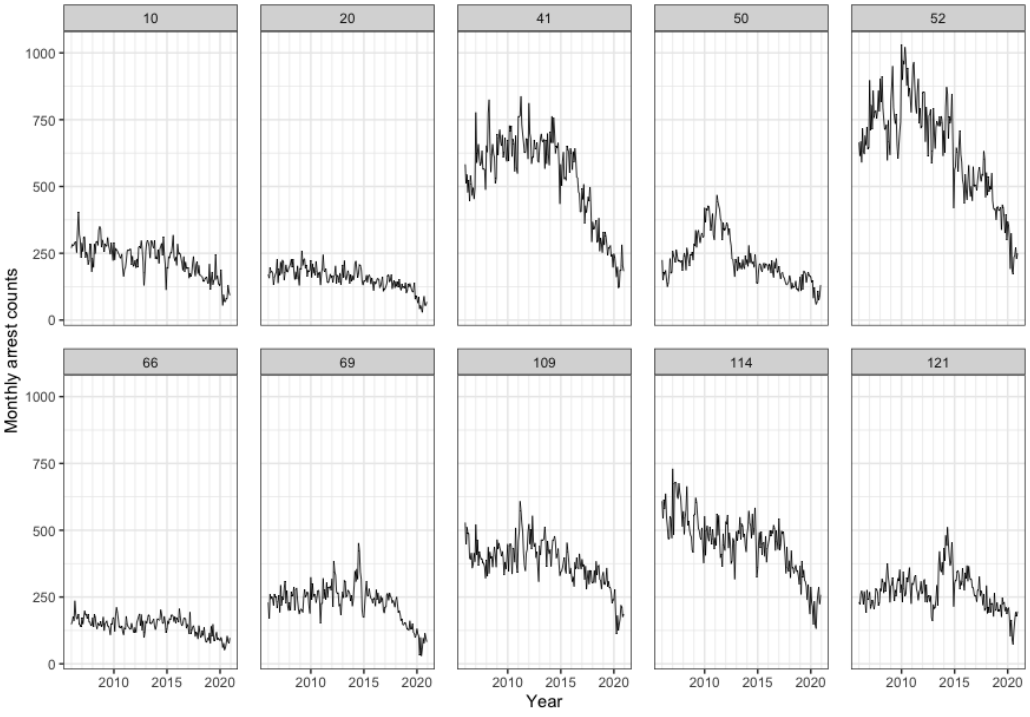


Figure 4.2: Monthly arrest counts of ten police precincts.

metrics. Because the ARIMA approach calculates a model for each police precinct, whereas the STARMA approach creates a model for the entire study area, it was difficult to achieve

the same or better error metrics with the STARMA models. When we look at the study data in the different police precinct (Figure 4.2), we can observe different trends in the precincts. This could be one reason why ARIMA models, since they fit a model to each study area, might achieve better error metrics than the STARMA models.

In the next sections, the results of the analysis are considered in detail, broken down by time period.

4.2 Weekly performance evaluation

The shortest time span in the analysis is one week. This time period represents short-term forecasts. The study period of 2006-2020 includes 784 weekly observations. Due to the large number of observations, the weekly time category has the most estimated parameters of all time periods. Table 4.1 shows that 14 parameters were estimated for the total crime category, of which nine are AR parameters and five are MA parameters. In contrast, the property and violent crime categories have fewer total parameters, and only the violent crime category has a parameter that represents the neighborhood relationship (ϕ_{11}). All three models also have the highest BIC values compared to the other time periods, with the w-all model having the lowest value among the three weekly models.

Table 4.1: Selection of weekly STARMA parameters with the corresponding BIC

Case study	Parameters	BIC
w all	$\phi_{10}, \phi_{20}, \phi_{30}, \phi_{40}, \phi_{50}, \phi_{70}, \phi_{80}, \phi_{100}, \phi_{130}, \theta_{10}, \theta_{20}, \theta_{40}, \theta_{80}, \theta_{130}$	36.244
w p	$\phi_{10}, \phi_{20}, \phi_{30}, \phi_{40}, \phi_{50}, \phi_{60}, \phi_{80}, \phi_{100}, \phi_{110}, \theta_{10}, \theta_{40}, \theta_{80}$	64.227
w v	$\phi_{10}, \phi_{11}\phi_{20}, \phi_{30}, \phi_{40}, \phi_{50}, \phi_{60}, \phi_{80}, \phi_{90}, \phi_{100}, \theta_{10}, \theta_{40}, \theta_{80}$	79.286

Table 4.2 shows the estimated parameters of the weekly STARMA models and the corresponding significance code. The parameter with the highest value in all three models is the 1st order AR parameter (ϕ_{10}). This means that the crime scores from one week earlier are weighted the most. An interesting difference emerges when looking at the second highest AR parameters in Table 4.2. While for the "all" and "property" crime categories the second highest AR parameter concerns crime values from four weeks ago (ϕ_{40}), the second highest parameter in the violent crime model ϕ_{20} concerns values from two weeks ago. The parameters accounting for the neighboring values are all low except of the weekly all crime model (ϕ_{11}) Low values do not mean that they are not important to the overall model, only that the data used to predict the values have less influence on the final result than parameters with high values.

As for MA parameters, all except θ_{20} are negative values. The purpose of MA models, as learned in subsection 2.5.2, is to create a smoother time series. The highest absolute values are found at θ_{40} for the weekly all and property model. For the violence model, the highest value is at θ_{10}, while θ_{40} is not significant for this model. Nevertheless, the parameter was

included because the error metrics were better with it than without it.

Table 4.2: Parameters of weekly STARMA models

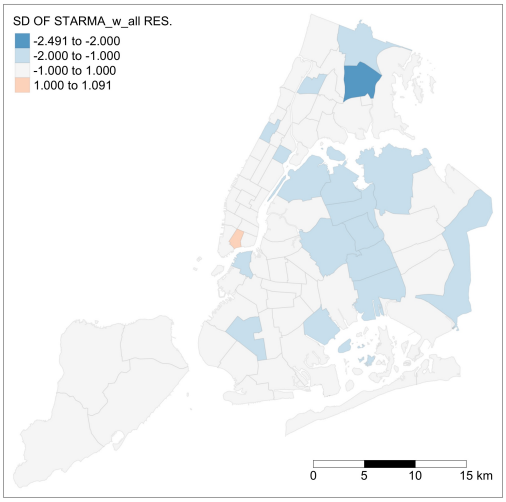
Parameter	w all		w p		w v	
	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
ϕ_{10}	0,544	***	0,478	***	0,629	***
ϕ_{11}	-0,241	***			0,012	***
ϕ_{20}			0,049	***	0,161	.
ϕ_{30}	0,117	***	0,038	***	0,022	***
ϕ_{40}	0,499	***	0,368	***	0,098	*
ϕ_{50}	-0,048	***	-0,012		0,018	*
ϕ_{60}	-0,020	**	-0,023	***	0,028	***
ϕ_{70}	0,020	***				
ϕ_{80}	0,042	*	0,161	***	0,216	***
ϕ_{90}			-0,022	***	-0,022	***
ϕ_{100}	-0,049	***	-0,037	***	-0,017	***
ϕ_{110}			-0,008	.		
ϕ_{130}	0,132	***				
θ_{10}	-0,230	***	-0,293	***	-0,461	***
θ_{20}	0,297	***				
θ_{40}	-0,340	***	-0,313	***	-0,039	
θ_{80}	-0,042	***	-0,142	***	-0,189	***
θ_{130}	-0,113	***				

A look at the error metrics in Table 4.3 shows that both STARMA and ARIMA models have similar error metrics for the weekly period. Highlighted are the "best" values, i.e., the lowest MSE, RMSE, and standard deviations, and the highest R^2. It is interesting to note that as R^2 increases, both MSE and RMSE increase. In addition, the model with the lowest BIC, w-all, has the lowest MSE and RMSE, but also the lowest R^2. To gain a better insight into the residuals of the models, they were mapped (Figure 4.3).

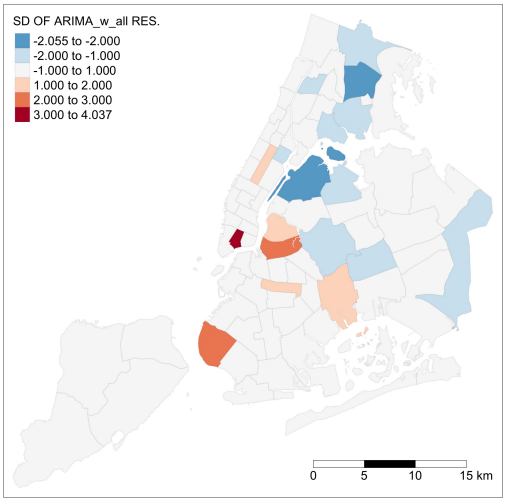
Table 4.3: Error metrics and their standard deviation for weekly models

Case	STARMA						ARIMA					
study	MSE	s.d.	RMSE	s.d.	R2	s.d.	MSE	s.d.	RMSE	s.d.	R2	s.d.
w all	**0,2246**	**0,0834**	**0,2965**	**0,1147**	0,8924	**0,0082**	0,2246	0,0856	0,2967	0,1171	**0,9005**	0,0129
w p	0,2842	**0,1126**	0,3786	0,1542	**0,9320**	**0,0244**	**0,2810**	0,1130	**0,3751**	**0,1533**	0,9320	0,0336
w v	0,3346	**0,1166**	0,4306	**0,1567**	0,9379	**0,0244**	**0,3317**	0,1184	**0,4266**	0,1571	**0,9470**	0,0333

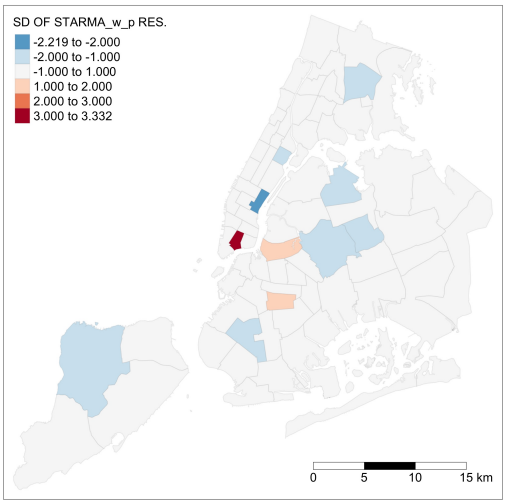
Figure 4.3 shows the standard deviation of the model residuals at the last observed time step (week 784). The blue areas indicate overprediction, while the red areas indicate underprediction. Comparing the 4.3a map with the 4.3b map, we see that the STARMA residuals in the Queens police precincts exhibit overprediction of up to 2 s.d.. In contrast,



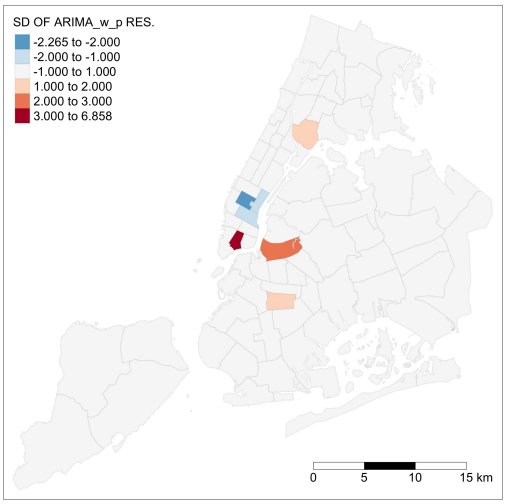
(a) s.d. of w all STARMA residuals.



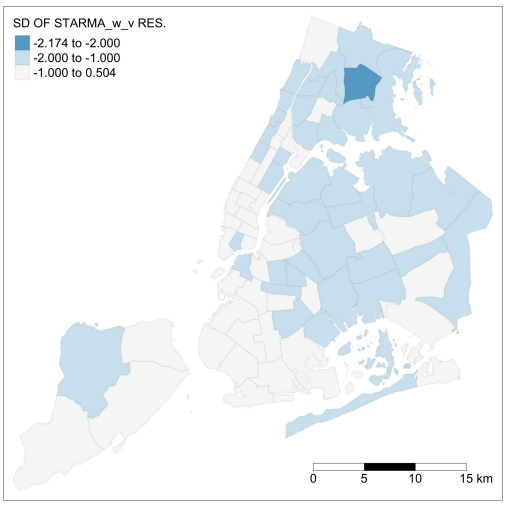
(b) s.d. of w all ARIMA residuals.



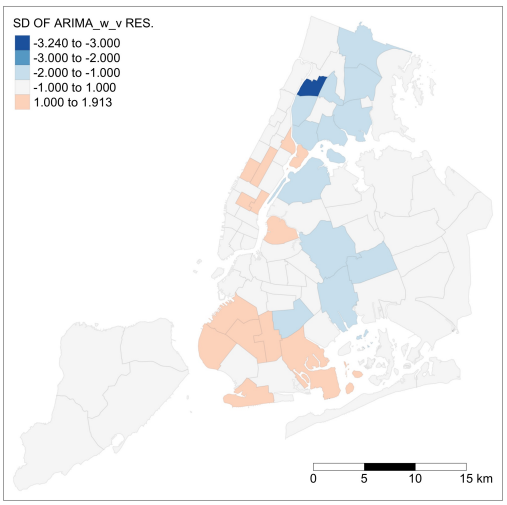
(c) s.d. of w p STARMA residuals.



(d) s.d. of w p ARIMA residuals.



(e) s.d. of w v STARMA residuals.



(f) s.d. of w all ARIMA residuals.

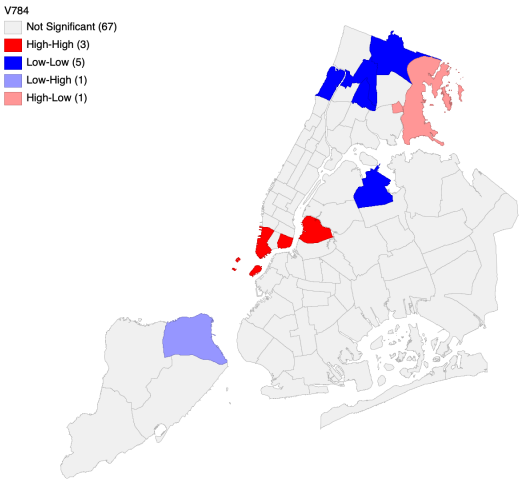
Figure 4.3: Residuals of time period 784 of the weekly models

the ARIMA residuals in the 4.3b map exhibit more extreme s.d. in both overprediction and underprediction. Between the two maps, it is apparent that the STARMA model has less extreme error, but the areas of significant s.d. are clustered. It appears that the STARMA models smooth the errors to the neighborhoods of the police precincts that produce high s.d. of the ARIMA residuals. This observation can also be made with the 4.3c and 4.3d maps, as well as with the 4.3e and 4.3f maps. A look at the three maps of ARIMA residuals shows a similar spatial distribution for the overall and violent crime categories, with cold spots around the Bronx and hot spots in Brooklyn and the southern tip of Manhattan. In contrast, the ARIMA models for property crime appear to be concentrated in Manhattan, with alternating over- and underestimates.

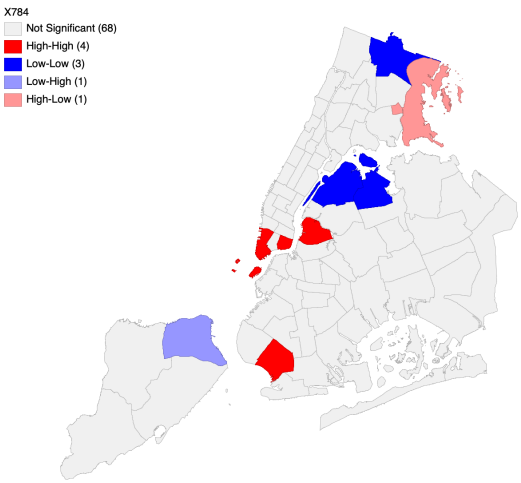
To obtain a statistical evaluation of the remaining spatial autocorrelation of the residuals of the models, they are mapped in Figure 4.4. In the "Weekly All" category, the ARIMA model (Figure 4.4b) has one more police precinct in the "Not Significant" category than the corresponding STARMA model (Figure 4.4a). This means that one more area is not spatially autocorrelated. The differences in the LISA cluster map of the weekly all-crime models are that the STARMA model generates two more areas in the "Low" category, meaning that low-crime areas are correlated with neighboring low-crime areas. The weekly ARIMA model, on the other hand, generates one more area in the High-High category, meaning that high-crime areas are correlated with neighboring high-crime areas. Both High-High and Low-Low categories indicate spatial clusters, while the High-Low and Low-High categories indicate spatial outliers.

The STARMA (Figure 4.4c) and ARIMA (Figure 4.4d) weekly property crime models yield the same number of clusters and outliers in the same location. As with the weekly violent crime models, the errors in the STARMA model (Figure 4.4e) produce 2 areas of spatial clusters and outliers, meaning that the residuals of the model produce fewer areas of spatial autocorrelation than the corresponding ARIMA model (Figure 4.4f).

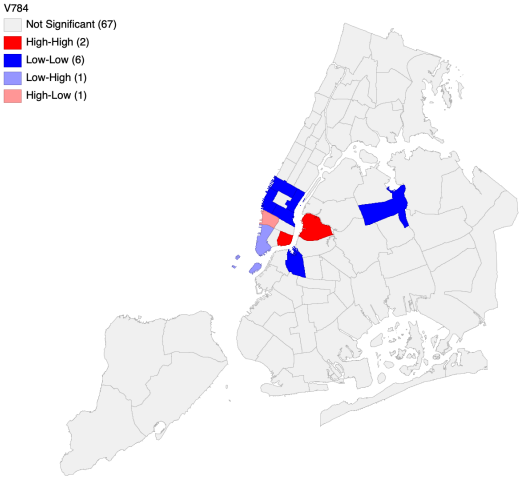
In general, the data sample of 784 weeks may not be ideal for modeling ARIMA and STARMA due to the changing trend in the data. The error metrics of STARMA and ARIMA were similar, but thanks to the mapping of the residuals of the model, the differences between the two approaches can be seen. The comparison of the LISA cluster maps showed that the residuals from both methods produced similar spatial clusters.



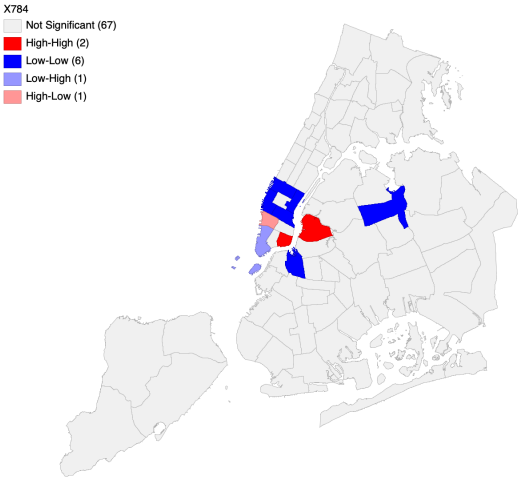
(a) LISA of w all STARMA residuals.



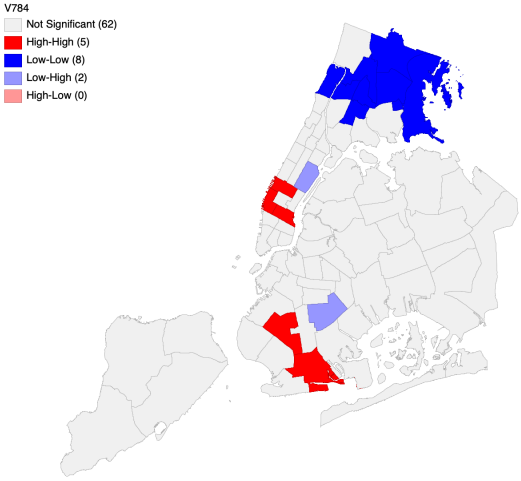
(b) LISA of w all ARIMA residuals.



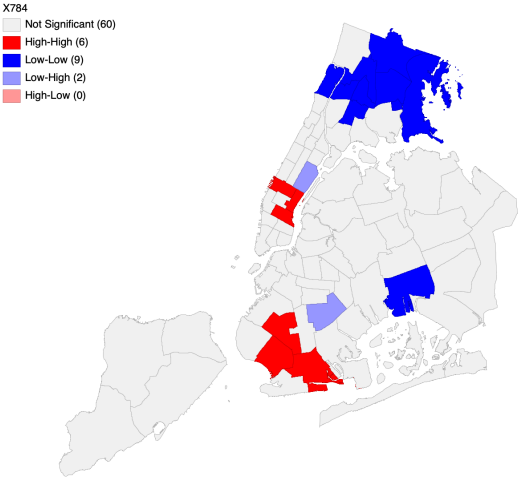
(c) LISA of w p STARMA residuals.



(d) LISA of w p ARIMA residuals.



(e) LISA of w v STARMA residuals.



(f) LISA of w all ARIMA residuals.

Figure 4.4: LISA of time period 784 of the weekly models.

4.3 Monthly performance evaluation

The monthly category consists of 180 temporal observations and represents medium-term forecasts. Looking at the BIC of the STARMA parameters in Table 4.4, the selected parameters fit the data better than the weekly models. The most important change in parameter selection is the absence of MA parameters in the all and violent crime categories, which essentially makes them STAR models.

Table 4.4: Selection of monthly STARMA parameters with the corresponding BIC

Case study	Parameters	BIC
1 m all	$\phi_{10}, \phi_{20}, \phi_{30}, \phi_{50}, \phi_{70}, \phi_{100}, \phi_{120}$	821
1 m p	$\phi_{10}, \phi_{20}, \phi_{21}, \phi_{30}, \phi_{70}, \phi_{80}, \phi_{120}, \phi_{121}, \theta_{20}$	3.246
1 m v	$\phi_{10}, \phi_{20}, \phi_{30}, \phi_{40}, \phi_{80}, \phi_{100}, \phi_{120}, \phi_{121}$	7.126

A closer look at the parameters of the monthly model (Table 4.5) mirrors the results of the parameters of the weekly model in that the values one time period prior are the most important values for predicting future arrests. Both the property and violent crime models have the second highest second-order parameter (ϕ_{20}), while for the overall model ϕ_{120}, the parameter for values twelve months ago, is the second highest. The model is the only monthly model that has estimates for the neighbors, neither of which are significant, but still improve the error metrics. The property model is also the only monthly model where an MA parameter shows an improvement in the models.

Table 4.5: Parameters of monthly STARMA models

Parameter	1m all		1m p		1m v	
	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
ϕ_{10}	0,425	***	0,395	***	0,315	***
ϕ_{20}	0,174	***	0,375	***	0,229	***
ϕ_{21}			0,049			
ϕ_{30}	0,063	***	0,057	***	0,099	***
ϕ_{40}					0,004	
ϕ_{50}	0,068	***				
ϕ_{70}	0,013		0,052	***		
ϕ_{80}			0,009		0,023	**
ϕ_{100}	0,021	*	-0,022	***	0,118	***
ϕ_{120}	0,230	***	0,098	***	0,189	***
ϕ_{121}			0,001		0,023	***
θ_{20}			-0,196	***		

The 1m-all STAR model has better overall error metrics than its ARIMA counterpart Table 4.6. The same is true for the 1m-v STAR model, except for the lower R^2. The 1-m-p

STARMA model has slightly worse or equal error metrics, but better s.d. of the errors.

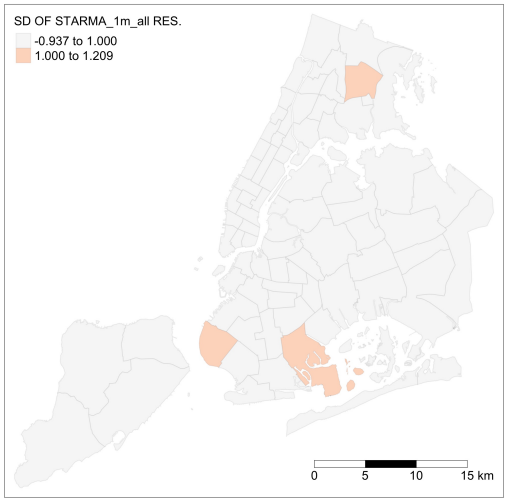
Table 4.6: Error metrics and their standard deviation for monthly models

Case	STARMA						ARIMA					
study	MSE	s.d.	RMSE	s.d.	R2	s.d.	MSE	s.d.	RMSE	s.d.	R2	s.d.
1 m all	**0,1500**	**0,0700**	**0,2083**	**0,0922**	**0,9076**	**0,0168**	0,1632	0,0735	0,2131	0,0970	0,9002	0,0196
1 m p	0,1856	0,0820	0,2394	**0,1063**	0,8927	**0,0143**	**0,1827**	**0,0818**	**0,2366**	0,1069	**0,8999**	0,0315
1 m v	**0,2090**	**0,0832**	**0,2668**	**0,1087**	0,9009	**0,0166**	0,2108	0,0931	0,2692	0,1204	**0,9142**	0,0352

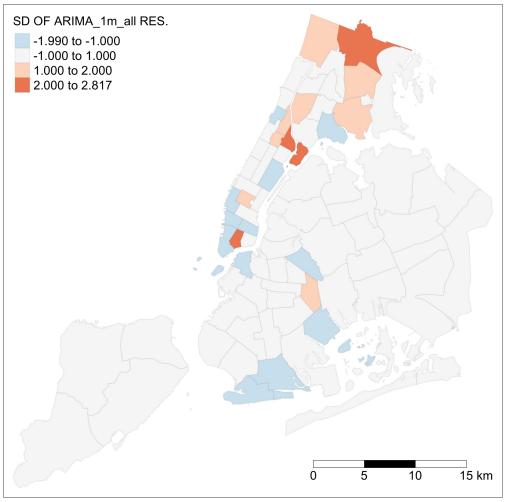
The good metrics of the monthly STAR models for all and violent crime are also evident in the residual maps 4.5a and 4.5e compared to their ARIMA counterparts. Both models overestimate or underestimate the prediction less than the ARIMA models. However, there still seems to be improvements in the parameter estimation of violent crime, as there are still some clusters. For the property category (maps 4.5c and 4.5d), the STARMA model has less extreme prediction error in fewer police precincts. Overall, the STARMA models produce prediction errors in fewer areas and also do not produce extreme errors. In addition, the residuals are not as spatially clustered as in the ARIMA models.

To gain further insight into which police precincts are still spatially clustered or represent outliers, the local Moran's of the residuals is calculated and mapped in Figure 4.6. Both the monthly ARIMA and STARMA models in the total and property crime category have the same proportion of non-spatially clustered areas. The only difference between these models is the location of the spatially clustered areas and the outliers. The weekly ARIMA model for property crime (Figure 4.6d) produces one more area that is not spatially correlated and therefore performs better than its STARIMA counterpart.

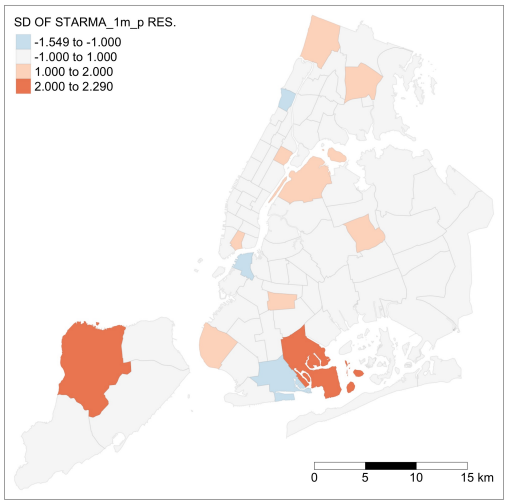
In the monthly category, the STARMA models performed better than the ARIMA models for both all and violent crime categories. The mapped residuals of the STARMA models also looked promising compared to the ARIMA models, but the LISA maps did not show less spatially correlated areas.



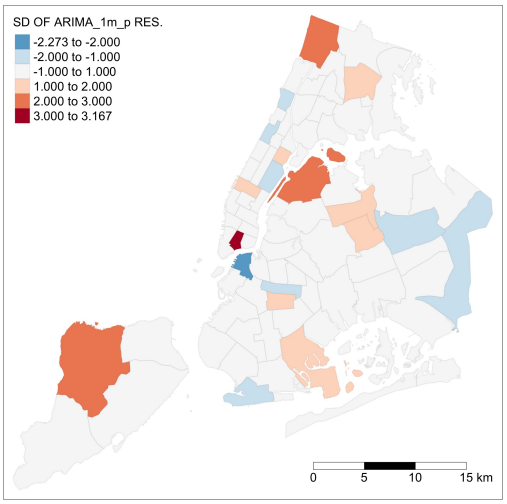
(a) s.d. of 1m all STARMA residuals.



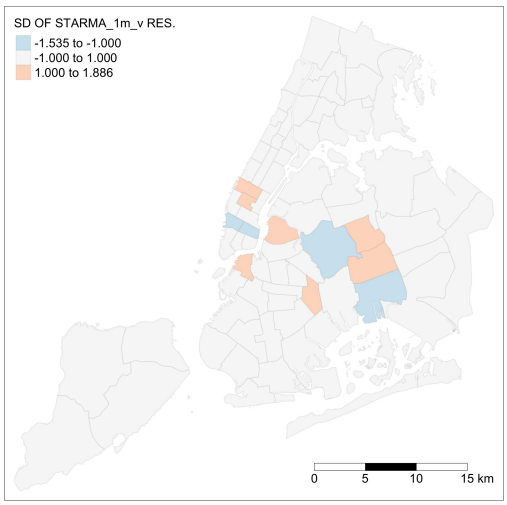
(b) s.d. of 1m all ARIMA residuals.



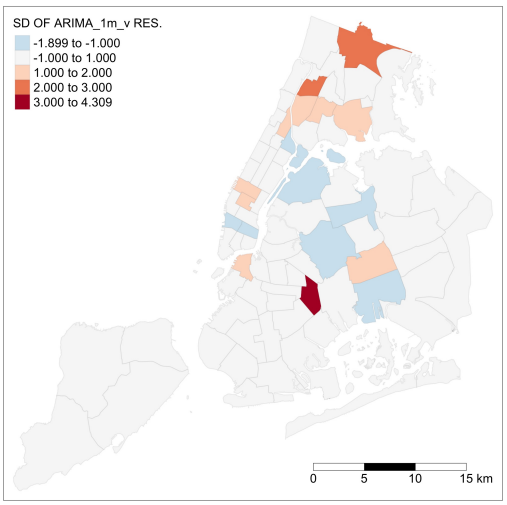
(c) s.d. of 1m p STARMA residuals.



(d) s.d. of 1m p ARIMA residuals.

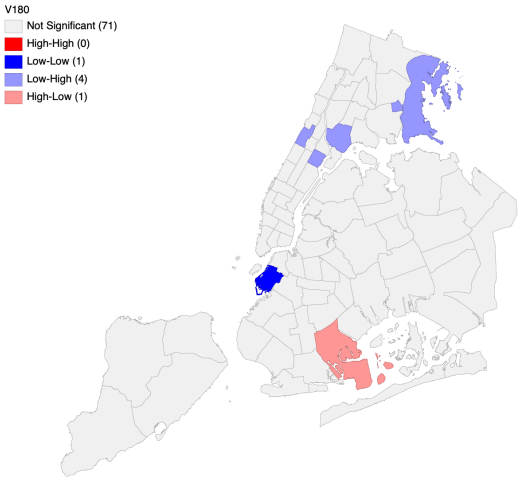


(e) s.d. of 1m v STARMA residuals.

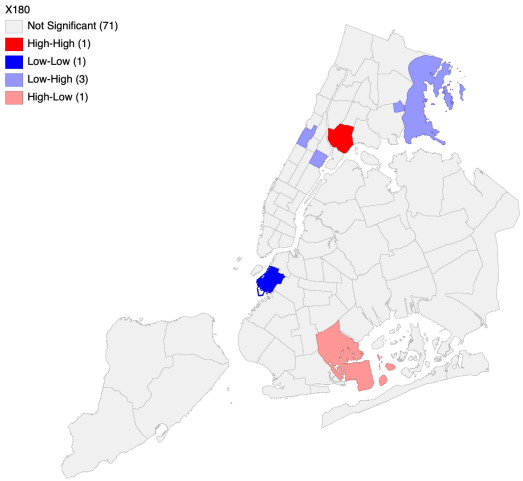


(f) s.d. of 1m v ARIMA residuals.

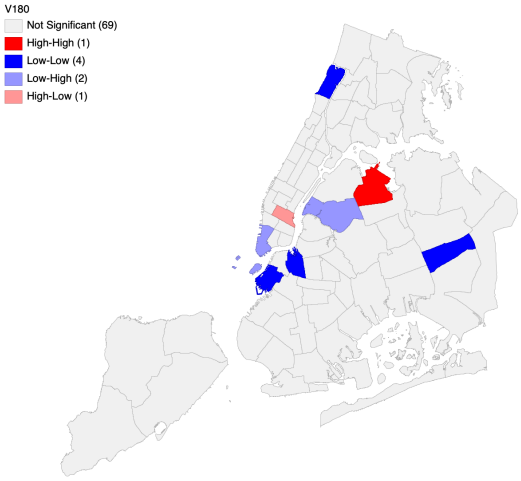
Figure 4.5: Residuals of time period 180 of the monthly models.



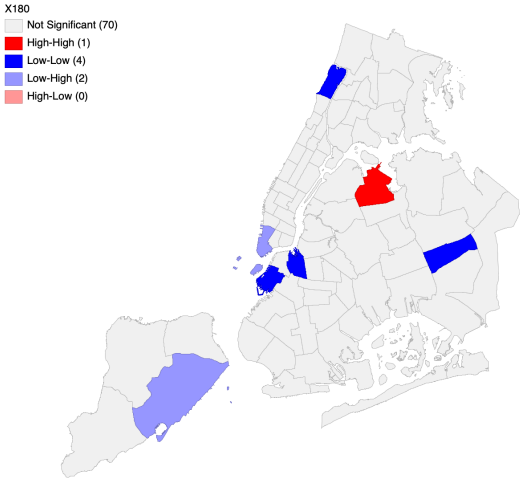
(a) LISA of 1m all STARMA residuals.



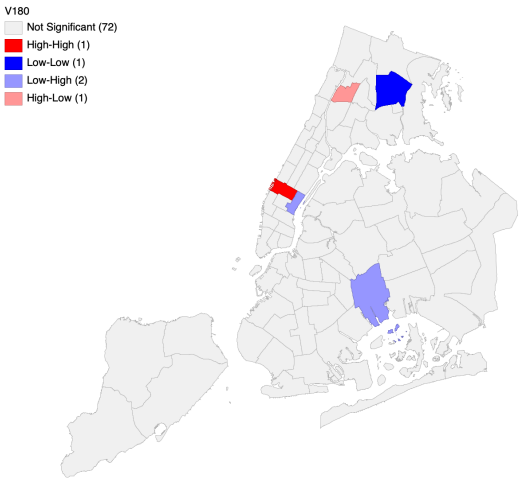
(b) LISA of 1m all ARIMA residuals.



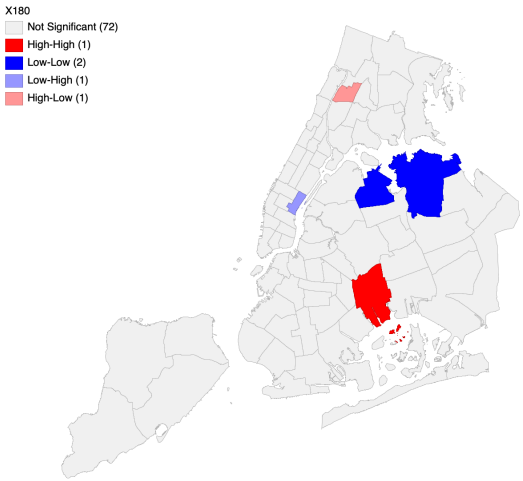
(c) LISA of 1m p STARMA residuals.



(d) LISA of 1m p ARIMA residuals.



(e) LISA of 1m v STARMA residuals.



(f) LISA of 1mw all ARIMA residuals.

Figure 4.6: LISA of time period 180 of the monthly models.

4.4 Quarterly performance evaluation

The quarterly category consists of 60 observations and represents one of the long-term periods. A look at the parameters in Table 4.7 shows that the model for all and the model for violent crime do not use MA parameters, as did the corresponding models in the monthly category. In addition, the STARMA models in this temporal category have the smallest BIC of all models.

Table 4.7: Selection of quarterly STARMA parameters with the corresponding BIC

Case study	Parameters	BIC
3 m all	$\phi_{10}, \phi_{20}, \phi_{30}, \phi_{40}$	284
3 m p	$\phi_{10}, \phi_{20}, \phi_{30}, \phi_{40}$	123
3 m v	$\phi_{10}, \phi_{30}, \phi_{40}, \phi_{80}, \phi_{81}$	1.772

A closer look at the parameters thanks to Table 4.8 shows the importance of the last period before the prediction. For all three models, the parameter with the highest values is ϕ_{10}. The parameter related to the values of the previous year (ϕ_{40}) has the second highest value. In addition, the monthly violence model has estimated parameters in the eighth order, both from their own and neighbors' values.

Table 4.8: Parameters of quarterly STARMA models

Parameter	3m all		3m p		3m v	
	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
ϕ_{10}	0,623	$***$	0,673	$***$	0,498	$***$
ϕ_{20}	0,051	$*$	0,091	$***$		
ϕ_{30}	0,034		0,099	$***$	0,142	$***$
ϕ_{40}	0,282	$***$	0,117	$***$	0,248	$***$
ϕ_{80}					0,079	$***$
ϕ_{81}					0,018	$*$

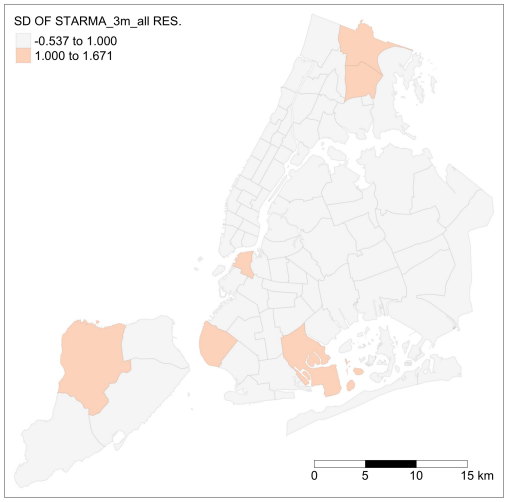
Looking at 4.9, the ARIMA models have the best MSE and RMSE. The STARMA all and violent crime models have the best R^2, but all error metrics and their s.d. are close. Again, mapping the residuals will provide more insight into which model performs better.

Table 4.9: Error metrics and their standard deviation for quarterly models

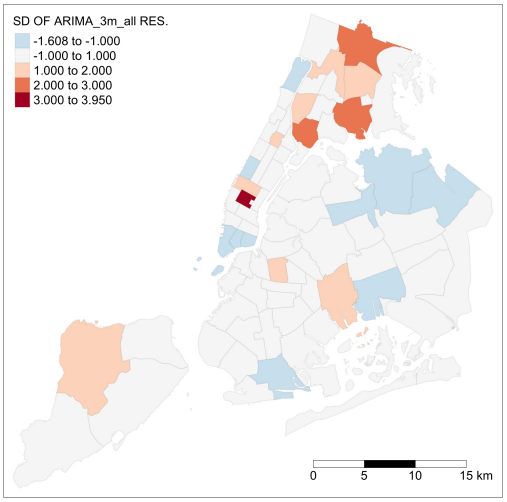
Case	STARMA						ARIMA					
study	MSE	s.d.	RMSE	s.d.	R2	s.d.	MSE	s.d.	RMSE	s.d.	R2	s.d.
3 m all	0,1456	0,0699	0,1855	0,0897	**0,9137**	**0,0220**	**0,1390**	**0,0698**	**0,1775**	**0,0878**	0,9084	0,0268
3 m p	0,1510	**0,0738**	0,1948	**0,0972**	0,8627	**0,0268**	**0,1478**	0,0787	**0,1900**	0,0987	**0,8775**	0,0360
3 m v	0,1597	**0,0775**	0,2039	**0,0999**	**0,8977**	**0,0289**	**0,1531**	0,0776	**0,1965**	0,1005	0,8927	0,0389

Looking at the maps in Figure 4.7, the residuals of the STARMA models are less spatially connected than their ARIMA counterpart. As with the monthly STARMA model, the STARMA violent crime model has room for more parameters to account for the spatial autocorrelation. Most of the residuals outside the first s.d. of the STARMA model are under-predicted. All in all, the results of the quarterly STARMA models look very promising.

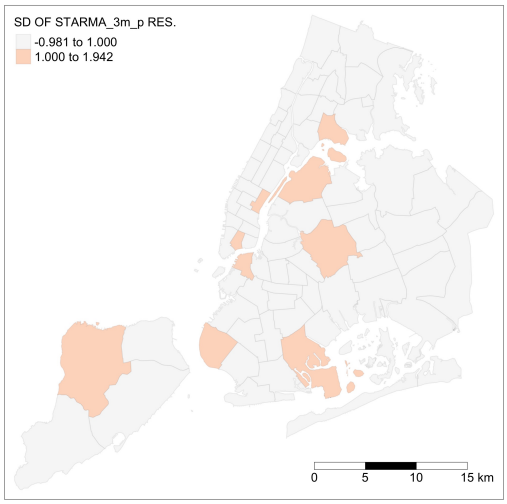
The promising residuals of the STARMA models are also reflected in the LISA cluster maps of the STARMA models for violent crime (Figure 4.8e) and for all crime (Figure 4.8a). Both generate four more areas that do not resemble any kind of spatial autocorrelation. In contrast, the monthly STARMA model for property crime (Figure 4.8c) performs worse than its ARIMA counterpart (Figure 4.8f).



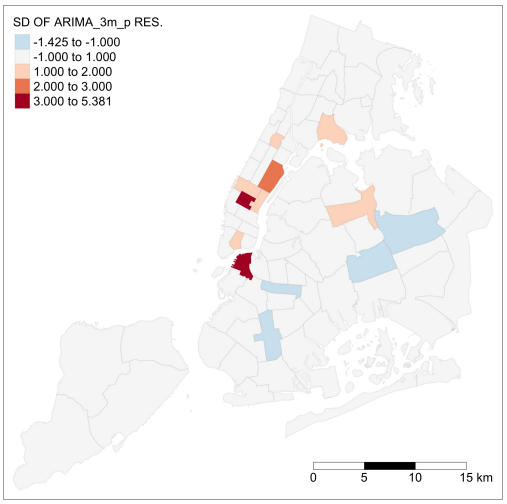
(a) s.d. of 3m all STARMA residuals.



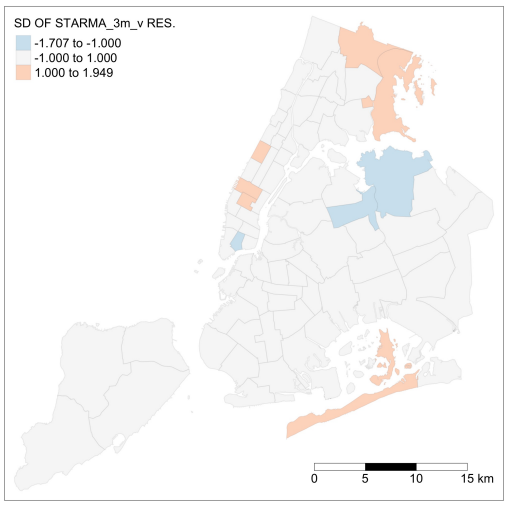
(b) s.d. of 3m all ARIMA residuals.



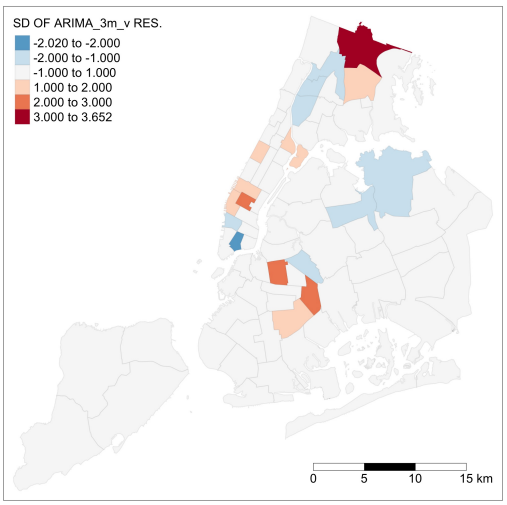
(c) s.d. of 3m p STARMA residuals.



(d) s.d. of 3m p ARIMA residuals.

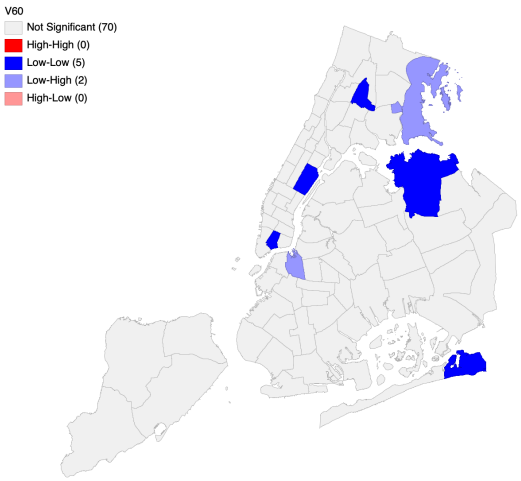


(e) s.d. of 3m v STARMA residuals.

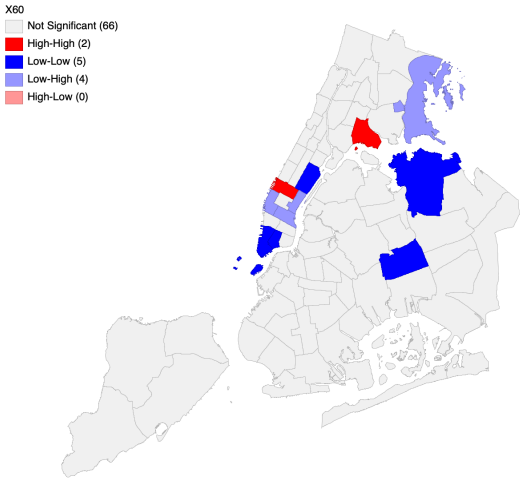


(f) s.d. of 3m v ARIMA residuals.

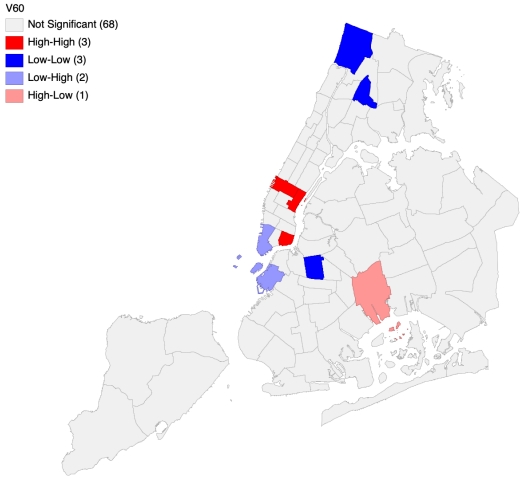
Figure 4.7: Residuals of time period 60 of the quarterly models.



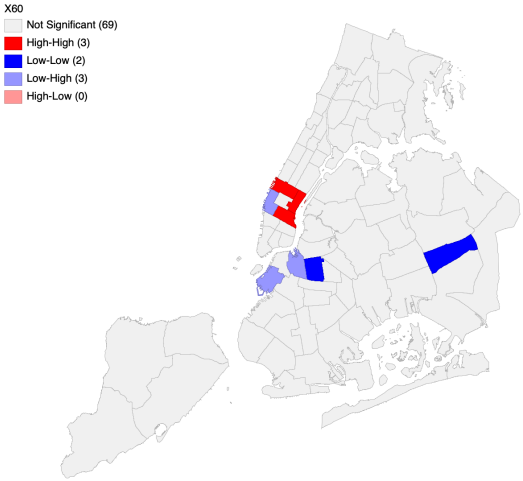
(a) LISA of 3m all STARMA residuals.



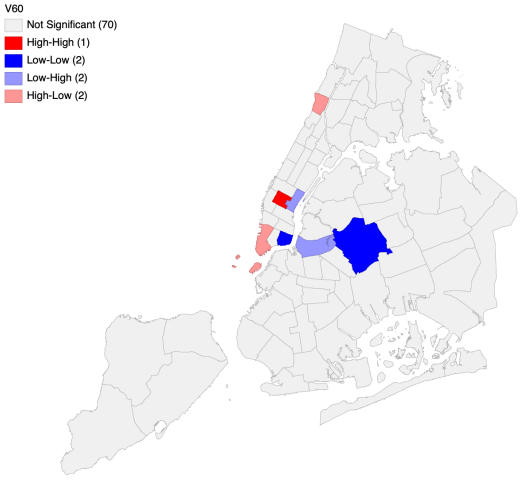
(b) LISA of 3m all ARIMA residuals.



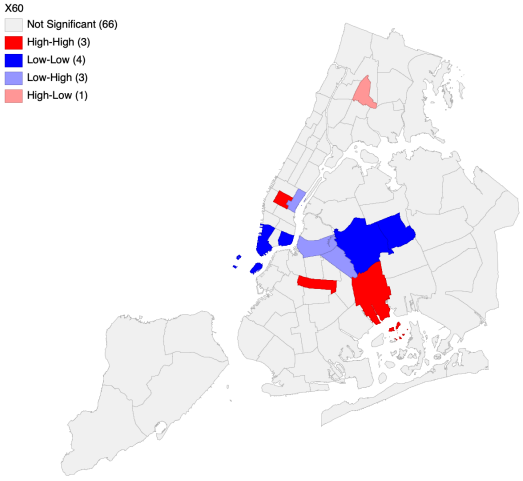
(c) LISA of 3m p STARMA residuals.



(d) LISA of 3m p ARIMA residuals.



(e) LISA of 3m v STARMA residuals.



(f) LISA of 3m all ARIMA residuals.

Figure 4.8: LISA of time period 60 of the quarterly models.

4.5 Semiannual performance evaluation

The semiannual category consists of 30 observations and is the second of the three long-term periods. Using Table 4.10, we can see that the property and violent crime model has no MA parameters. Moreover, this is the only category in which all three STARMA models have one or more parameters related to their neighbor. The BIC of the STARMA models is higher compared to the BIC of the quarterly models, with the exception of the BIC of the violent model.

Table 4.10: Selection of semiannual STARMA parameters with the corresponding BIC

Case study	Parameters	BIC
6 m all	$\phi_{10}, \phi_{20}, \phi_{41}, \theta_{30}, \theta_{40}, \theta_{41}$	606
6 m p	$\phi_{10}, \phi_{20}, \phi_{30}\phi_{41}$	377
6 m v	$\phi_{10}, \phi_{11}, \phi_{20}, \phi_{21}, \phi_{40}$	1.159

Using Table 4.11, the most influential parameter of the semiannual month is again ϕ_{10}, followed by ϕ_{20}. Interestingly, the STARMA model parameters have very small values compared to the other two models. Both the violence and total crime models have significant parameters that include neighborhood values. To see if the parameters have an effect on the spatial distribution of the residuals, a close look at the maps is needed.

Table 4.11: Parameters of semiannual STARMA models

Parameter	6m all		6m p		6m v	
	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
ϕ_{10}	0,688	$***$	0,087	$***$	0,494	$***$
ϕ_{11}					-0,079	$*$
ϕ_{20}	0,275	$***$	0,067	$***$	0,401	$***$
ϕ_{21}					0,102	$***$
ϕ_{30}			-0,011	$***$		
ϕ_{40}					0,061	$*$
ϕ_{41}	0,001		-9,122e-5			
θ_{30}	-0,056					
θ_{40}	0,177	$***$				
θ_{41}	0,105	$***$				

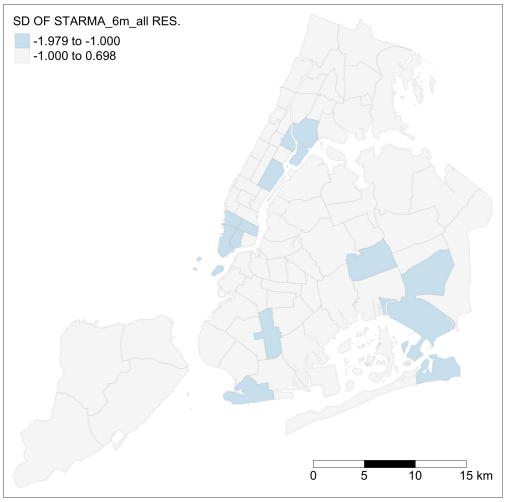
The error metrics of the semiannual models in Table 4.12 are almost the same and follow the same distribution as the error metrics of the quarterly models. The only difference is that the standard deviation of R^2 of the semiannual STARMA model is worse than that of the ARIMA model. In general, the standard deviation of R^2 is higher for all models than for the quarterly counterpart.

Table 4.12: Error metrics and their standard deviation for semiannual models

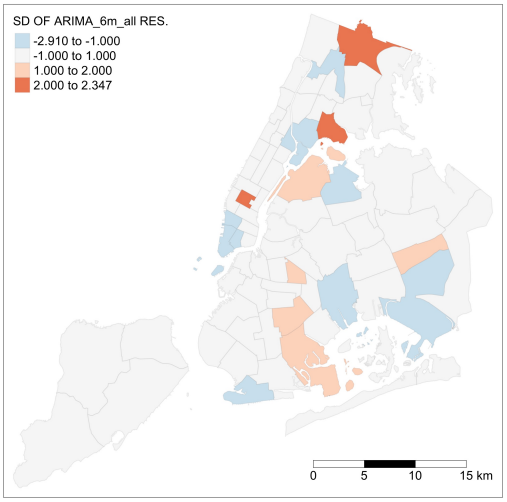
Case	STARMA						ARIMA					
study	MSE	s.d.	RMSE	s.d.	R2	s.d.	MSE	s.d.	RMSE	s.d.	R2	s.d.
6 m all	0,1450	0,0632	0,1795	0,0775	**0,9119**	0,0475	**0,1296**	**0,0615**	**0,1636**	**0,0788**	0,9115	**0,0321**
6 m p	0,1470	**0,0790**	0,1879	**0,1040**	0,8482	**0,0421**	**0,1432**	0,0850	**0,1806**	0,1056	**0,8856**	0,0482
6 m v	0,1523	**0,0816**	0,1901	**0,0997**	**0,8918**	**0,0296**	**0,1413**	0,0844	**0,1783**	0,1030	0,8861	0,0465

The residuals of the semi-annual STARMA model look consistent compared to their ARIMA counterparts in Figure 4.9. Again, almost all of the STARMA model residuals are overestimates. In addition, the residuals in all STARMA models appear to resemble some sort of spatial autocorrelation, but again, they look very promising compared to the errors in the ARIMA models.

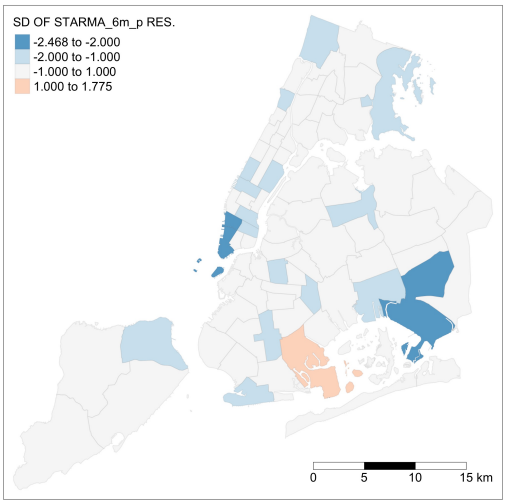
Compared to the LISA cluster maps for the quarterly models, the semiannual models (Figure 4.10) yield fewer areas of nonsignificant spatial autocorrelation. The semiannual ARIMA model (Figure 4.10b) for all crimes performs best overall, followed by the STARMA model for property (Figure 4.10d) and violent crimes (Figure 4.10f).



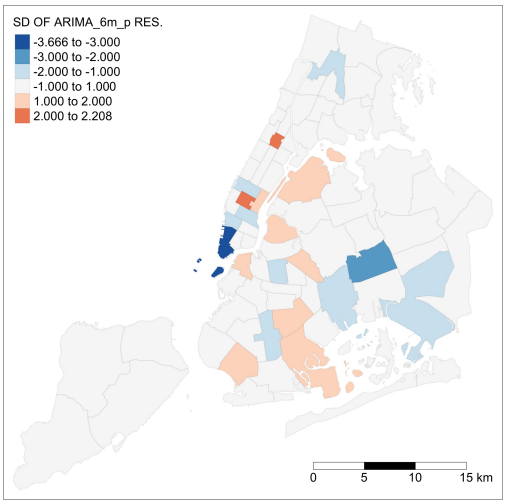
(a) s.d. of 6m all STARMA residuals.



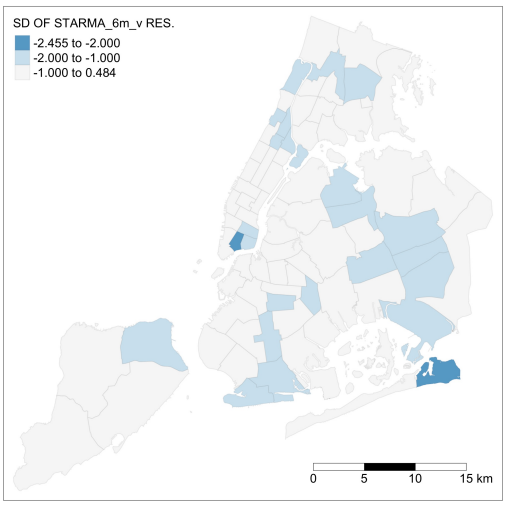
(b) s.d. of 6m all ARIMA residuals.



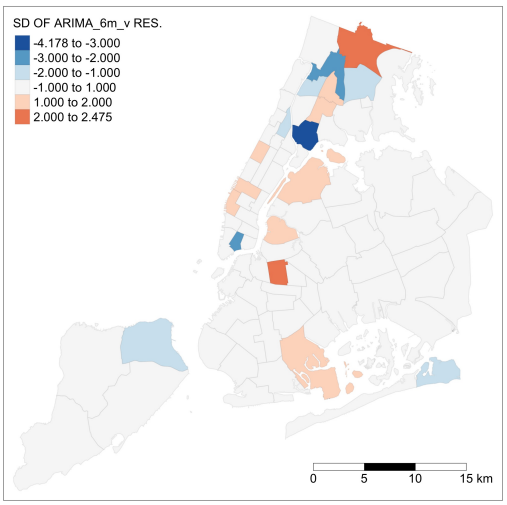
(c) s.d. of 6m p STARMA residuals.



(d) s.d. of 6m p ARIMA residuals.

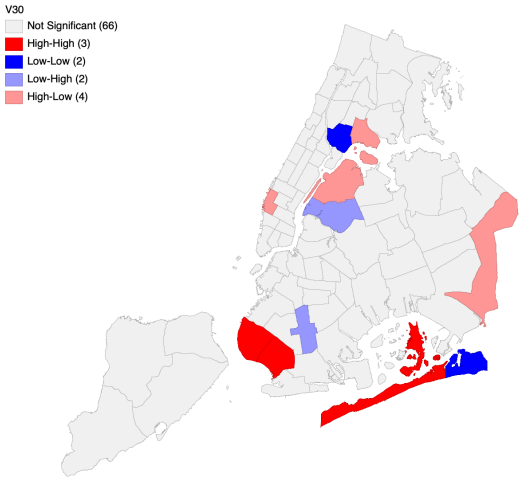


(e) s.d. of 6m v STARMA residuals.

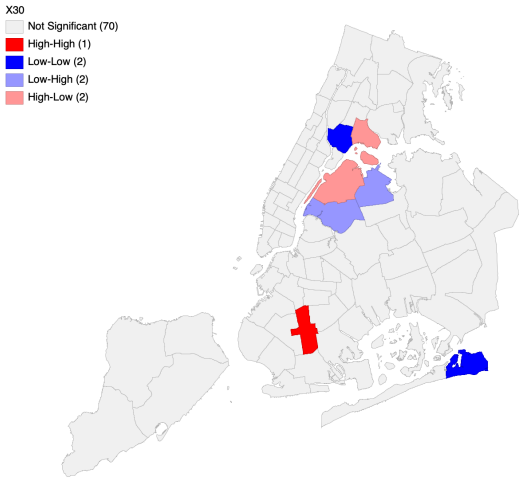


(f) s.d. of 6m v ARIMA residuals.

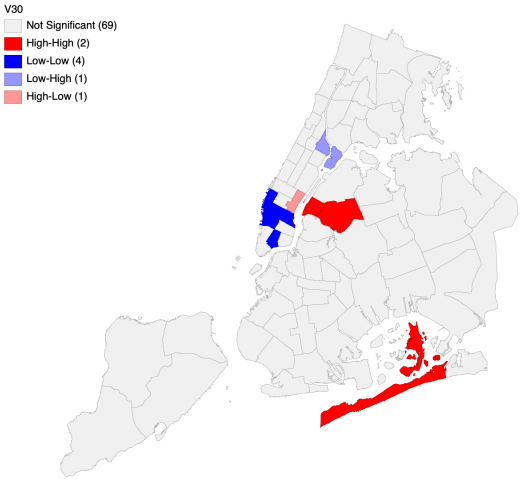
Figure 4.9: Residuals of time period 30 of the semiannual models.



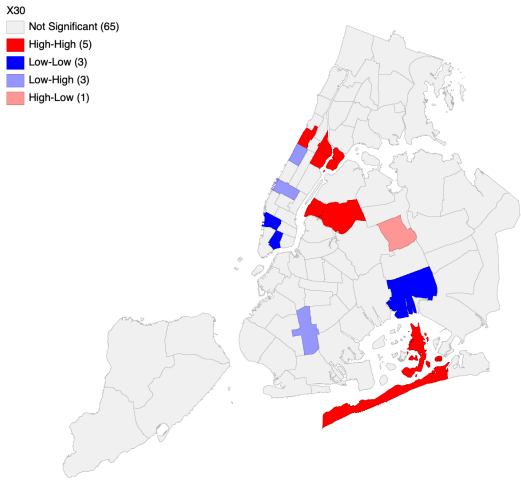
(a) LISA of 6m all STARMA residuals.



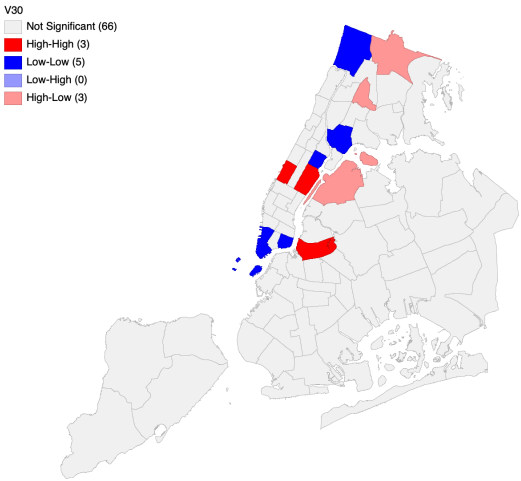
(b) LISA of 6m all ARIMA residuals.



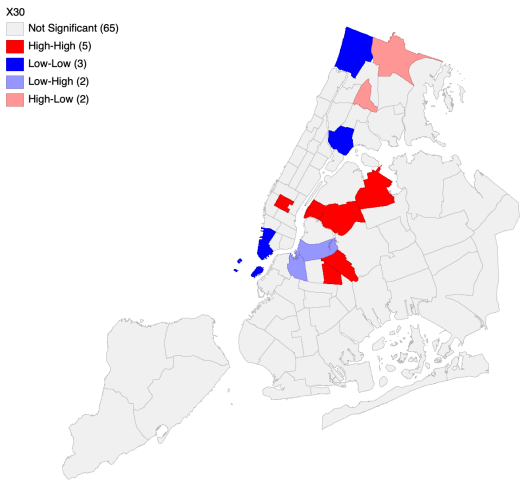
(c) LISA of 6m p STARMA residuals.



(d) LISA of 6m p ARIMA residuals.



(e) LISA of 6m v STARMA residuals.



(f) LISA of 6m all ARIMA residuals.

Figure 4.10: LISA of time period 30 of the semiannual models.

4.6 Annual performance evaluation

The last temporal category of the study is annual with 15 observations. Due to the small number of observations, the STARMA models in the annual category have the smallest number of parameters (Table 4.13). Since all three models do not use an MA parameter, all three models can also be considered STAR models. The BIC of the annual models also increases compared to the semi-annual models.

Table 4.13: Selection of annual STARMA parameters with the corresponding BIC

Case study	Parameters	BIC
y all	ϕ_{10}, ϕ_{11}	779
y p	ϕ_{30}	2.194
y v	ϕ_{30}	2.191

Because there are few parameters in the annual models, they all have high values except ϕ_{11} (Table 4.14). The annual property and violent crime models are the first in which the first-order AR parameter was not estimated because ϕ_{10} yielded higher BIC and error metrics.

Table 4.14: Parameters of annual STARMA models

Parameter	y all		y p		y v	
	Estimate	Signif.	Estimate	Signif.	Estimate	Signif.
ϕ_{10}	0,968	$***$				
ϕ_{11}	0,19	.				
ϕ_{30}			0,825	$***$	0,864	$***$

Almost all error metrics favor the annual ARIMA model, with the exception of the R^2 of the STARMA model for all and property crimes (Table 4.15). While in the previous periods the error metrics were quite similar, this is not the case for the annual models.

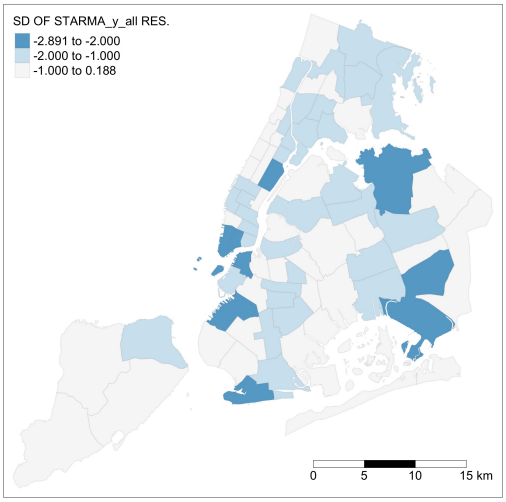
Table 4.15: Error metrics and their standard deviation for annual models

Case	STARMA						ARIMA					
study	MSE	s.d.	RMSE	s.d.	R2	s.d.	MSE	s.d.	RMSE	s.d.	R2	s.d.
y all	0,1646	0,0870	0,2031	0,1017	**0,9315**	**0,0304**	**0,1420**	**0,0774**	**0,1795**	**0,0953**	0,9080	0,3444
y p	0,3129	0,2618	0,3704	0,2965	**0,9006**	**0,0470**	**0,1605**	**0,1153**	**0,1991**	**0,1406**	0,8966	0,0737
y v	0,2552	0,1295	0,3026	0,1523	0,8757	0,0969	**0,1425**	**0,0919**	**0,1797**	**0,1143**	**0,8813**	**0,0773**

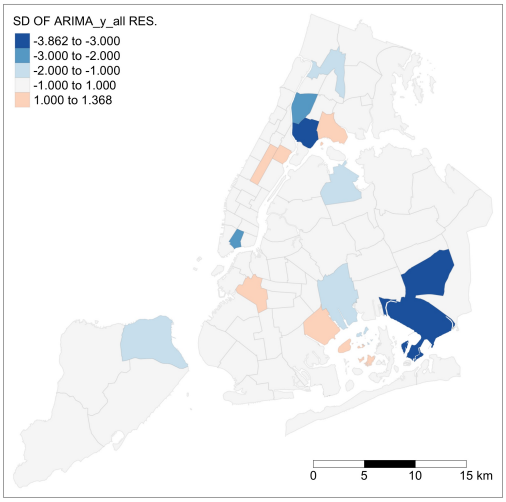
In mapping the residuals of the last period (Figure 4.11), the annual STARMA models all overpredict a large number of areas. This general over-prediction could because the last year in the data was 2020, which had a drastic drop in arrests due to the COVID-19

pandemic. For the ARIMA models, there are fewer areas outside the first s.d., but also greater diversity in the areas that are outside.

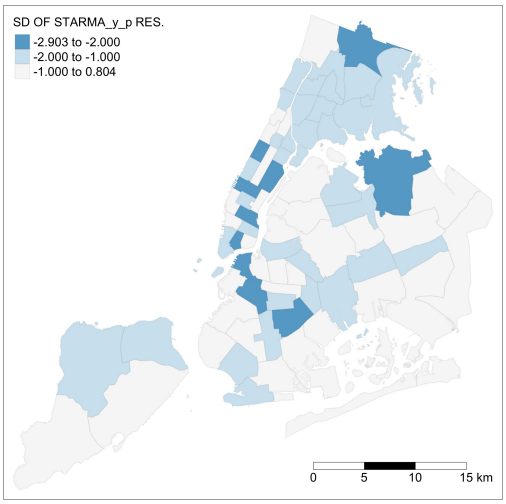
The LISA cluster maps of the annual Figure 4.12 models yield similar amounts of spatially nonautocorrelated areas. Interestingly, the location of areas with spatial clusters and outliers varies greatly between the STARMA and ARIMA models of the property and violent crime category. All in all, the STARMA model (Figure 4.12a) and the violent crime model (Figure 4.12e) perform better than their ARIMA counterpart, and the ARIMA property crime model (Figure 4.12d) performs better than its STARMA counterpart.



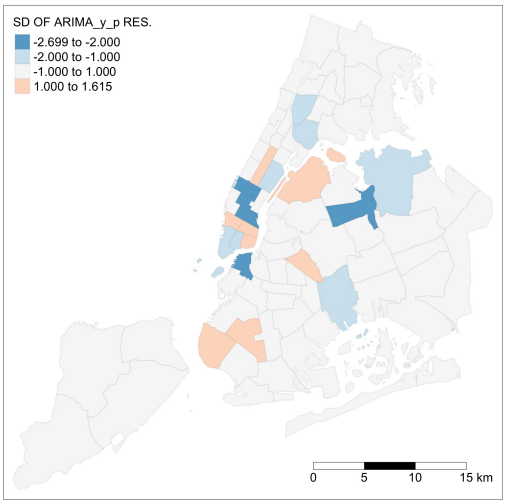
(a) s.d. of y all STARMA residuals.



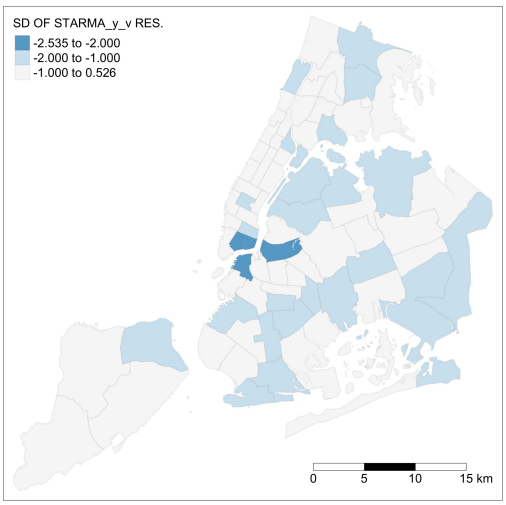
(b) s.d. of y all ARIMA residuals.



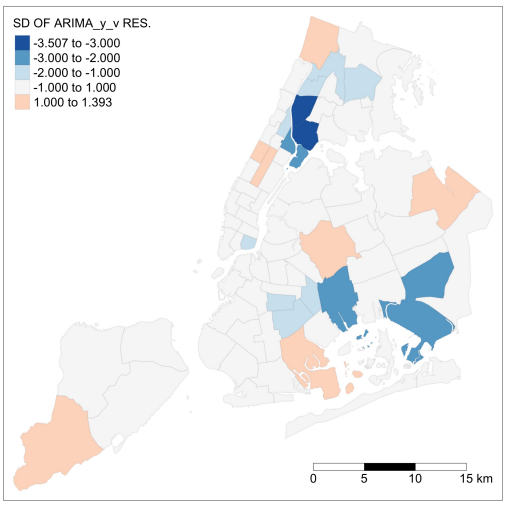
(c) s.d. of y p STARMA residuals.



(d) s.d. of y p ARIMA residuals.

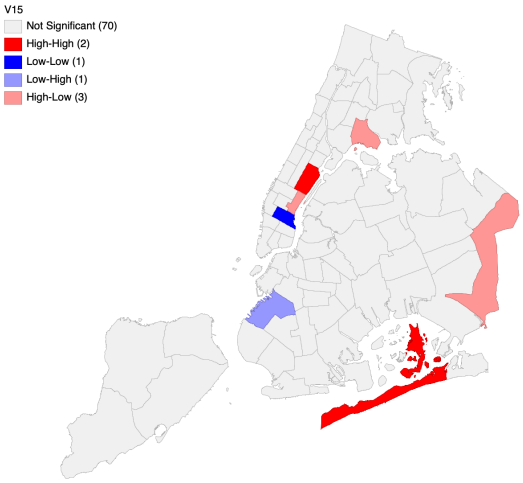


(e) s.d. of y v STARMA residuals.

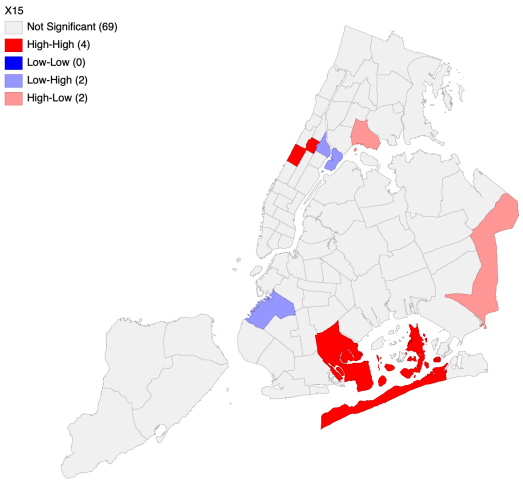


(f) s.d. of y v ARIMA residuals.

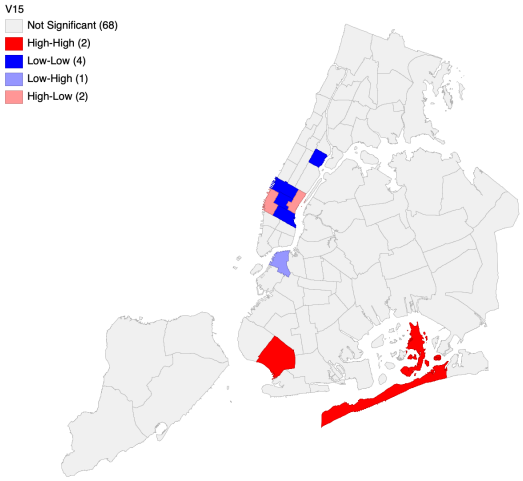
Figure 4.11: Residuals of time period 15 of the annual models.



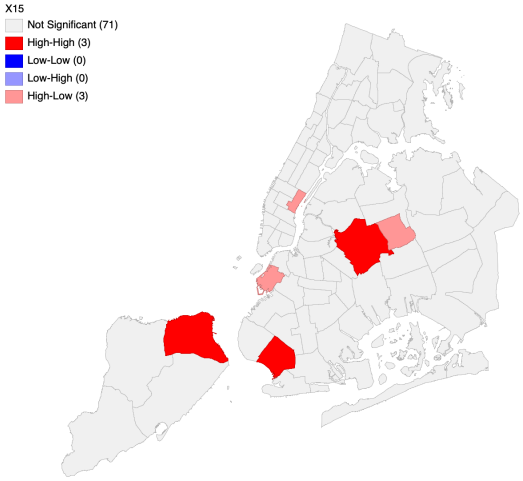
(a) LISA of y all STARMA residuals.



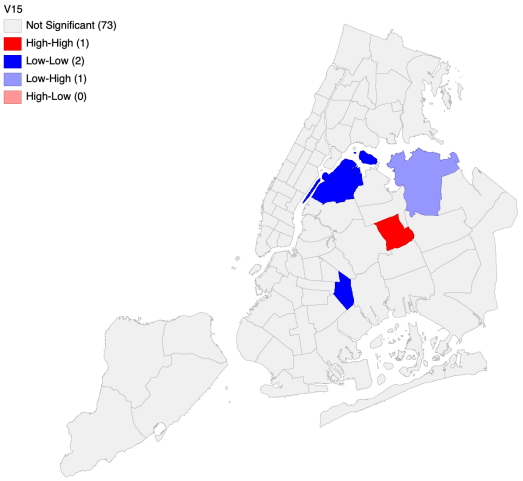
(b) LISA of y all ARIMA residuals.



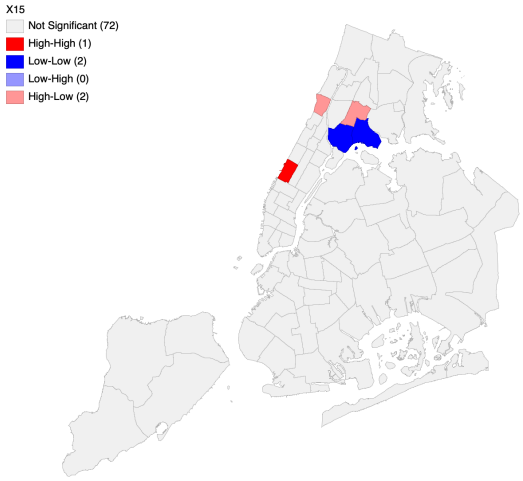
(c) LISA of y p STARMA residuals.



(d) LISA of y m p ARIMA residuals.



(e) LISA of y v STARMA residuals.



(f) LISA of y all ARIMA residuals.

Figure 4.12: LISA of time period 15 of the annual models.

4.7 Model evaluation

To get an overview of which model performs best on which performance metric, each analyzed model is evaluated against its counterpart. For each temporal and criminal category, the performance metrics (error metrics, residual maps, and LISA cluster maps) of the STARMA and ARIMA models are compared, and the model that outperforms the other receives one point for each performance metric. If neither model outperforms the other, for example because they have the same spatial distribution in the LISA maps or because it cannot be attributed to a particular model to have better residual maps than the other, both models receive half a point. When evaluating the error metrics, the standard deviations of the errors are not counted whether one model performs better than the other. Interpretation of the residual maps is purely visual, which can be subjective. The model whose residuals yielded fewer errors outside the first s.d. or that had less variation between over- and under-prediction was considered better. For the LISA cluster maps, the model that yielded more nonsignificant spatially correlated regions than the other performed better. How each model compared to each other per category can be seen in Table 4.16.

Table 4.16: Summary of the evaluation of the STARMA and ARIMA models for each case study. The error metrics, residual map, and LISA map of the STARMA and ARIMA models for each case study are compared. The method that performs best receives a 1. If both methods perform similarly, they both receive 0,5.

Case	STARMA				ARIMA			
study	Error metrics	Residuals	LISA	$\sum$	Error metrics	Residuals	LISA	$\sum$
w all	1	1	–	**2**	–	–	1	1
w p	1	–	0,5	1,5	–	1	0,5	1,5
w v	–	1	1	**2**	1	–	–	1
1 m all	1	1	0,5	**2,5**	–	–	0,5	0,5
1 m p	–	1	–	1	1	–	1	**2**
1 m v	1	1	0,5	**2,5**	–	–	0,5	0,5
3 m all	–	1	1	**2**	1	–	–	1
3 m p	–	1	–	1	1	–	1	**2**
3 m v	1	1	1	**3**	–	–	–	0
6 m all	–	1	–	1	1	–	1	**2**
6 m p	–	1	1	**2**	1	–	–	1
6 m v	1	1	1	**3**	–	–	–	0
y all	–	0,5	1	1,5	1	0,5	–	1,5
y p	–	0,5	–	0,5	1	0,5	1	**2,5**
y v	–	0,5	1	1,5	1	0,5	–	1,5
$\sum$	6	**12,5**	**8,5**	**27**	**9**	2,5	6,5	18

Overall, the STARMA methods performed better than the ARIMA method. The main driver of this difference is found in the mapped residuals of the models. The STARMA

model residuals consistently yielded fewer residuals outside of the first s.d.. However, it is important to note that the mapped residuals are from the last temporal observation, not for the total period. It could be that the ARIMA residuals from a different temporal observation would perform better than the STARMA residuals. However, not every time period can be analyzed due to time constraints. Moreover, the difference in the overall performance of the residuals is measured in the error metrics of the models. Here, the ARIMA models perform slightly better than their STARMA counterparts. The difference here lies in the automatic estimation of the ARIMA parameters and in the fact that a separate ARIMA model is estimated for each constituency. Estimating the STARMA parameters proved to be an arduous task because there are many different parameter combinations possible. Certainly, there are some undiscovered parameter combinations in the STARMA models. Nevertheless, most STARMA models yielded similar error metrics as the ARIMA models. When comparing the local spatial autocorrelation of the residuals, the STARMA models slightly outperformed the ARIMA models.

The STARMA method outperformed the ARIMA method in every temporal category except the annual one. This could be related to the small number of temporal observations (15) in the annual category and some parameters of the STARMA models that were not estimated. In particular, the monthly, quarterly, and semiannual STARMA models performed very well. For the weekly models, the number of observations was definitely not ideal, but still the weekly models yielded good performance metrics.

When looking at the crime categories, the STARMA and ARIMA methods performed almost equally well. While the ARIMA models performed best in the property crime category, the STARMA models performed best in the violent crime category. In the all crimes category, both methods performed similarly, with a slight advantage for the STARMA method. This is undoubtedly related to the measured spatial correlation of the different crime types. If we recall Table 3.2, the property crime category yields the lowest values of all categories, which are close to zero, i.e., spatially randomly distributed. For all crime categories, Moran's I values were around $\approx 0,2$, also not a good indicator of spatial autocorrelation. The violent crime category had the highest Moran's I values ($\approx 0,4$), which is why the STARMA method significantly outperforms the ARIMA method in this category.

Chapter 5

Conclusion

The main objective of this thesis is to further investigate the promising STARMA models in the field of spatial crime forecasting. To determine if the objectives of the study were met, the research questions posed at the beginning of the thesis are answered in this chapter. There were several limitations during the study that are also reported in this chapter. At the end of this chapter and thesis, future research directions are identified based on the results of this thesis.

5.1 Answering the research question

1.1 *How does the performance of STARMA models vary with increasing levels of time lags?*

The performance of the STARMA models first increases up to a time lag of three months (quarterly) and then decreases again. This variation can be attributed to the number of temporal observations available to estimate the parameter. For example, for the annual models, only 15 temporal observations were available for each county, which limits the number of parameters that can be estimated. This resulted in more incorrectly predicted areas. On the other hand, the weekly models had a large number of temporal observations (784), which made it difficult to estimate parameters and therefore resulted in weaker performance of the weekly models compared to the monthly, quarterly, and semi-annual models.

2.2 *What is the effect of the model's parametrization on the predictive results?*

The parameterization of the model has a great influence on the prediction result. In this study, the most influential parameter was the first-order AR parameter (ϕ_1), which was estimated in all but two year models. Because this parameter considers the last observation before the value to be predicted, it can be concluded that the number of arrests is strongly influenced by the number of arrests in the previous period. But only one parameter was not sufficient for the models. A variety of different AR parameters were found to further improve the prediction results. In contrast, the spatially lagged

parameters (e.g., ϕ_{11}) did not have a large impact on the error metrics, but reduced the spatial autocorrelation of the model residuals.

The MA parameters (θ) had a smaller impact on the prediction results than the AR parameters. For most of the best performing STARMA models, estimating an MA parameter decreased the performance of the model. Again, the inclusion of an MA parameter for the neighbors had little to no effect on the prediction results.

2.1 *What is the added value or limitations when using the STARMA method compared to the ARIMA method?*

The greatest added value of STARMA methods was found when the underlying data exhibited spatial autocorrelation. In this study, the models estimated on data that had a Moran's I value of $0,2$ or more outperformed the ARIMA models. On the other hand, the limitations of STARMA become apparent when the data are randomly distributed in space. In this case, the ARIMA models generally perform better.

Another important limitation of the STARMA method is the lack of packages to predict an estimated model or automatic estimation of the parameters. The implementation of the ARIMA models was really fast and easy thanks to the automatic estimation of the parameters. This makes them the first choice when the data have spatial randomness.

Because the STARMA models estimated one model for the entire study area, unlike the 77 ARIMA models, the mapping errors of the STARMA models had less variation and extreme standard deviations. This could be advantageous in predicting values for many areas in a region because the errors of the models are similar and therefore can be better managed than a high variation of error deviations.

2.2 *Which models performs best and is this dependent on the time lag?*

The best model is the quarterly STARMA model for violent crime, followed by the semiannual STARMA model for violent crime, and the monthly STARMA model for violent and all crime. All STARMA models with the best performance were estimated with data that had spatial autocorrelation over the study area and had a number of observations between 30 and 180. Therefore, it can be assumed that the performance of the STARMA model depends more on the spatial autocorrelation and the number of observations than on a specific time period.

3.1 *Are best performing time lags dependent on the crime type, and if yes which time lags are more adequate for each crime type?*

The best-performing time delays do not depend on the type of crime. For the best performing time delays, monthly, quarterly, and semiannual delays, the STARMA model performs best in the all and violent crimes categories. The ARIMA model for property crimes, on the other hand, performs best in these time lags. This shows that the type

of crime has an effect on which models perform better than others, but does not affect whether one time lag performs better than another. The number of observations could have a greater impact on the performance of certain time lags.

5.2 Limitations of the study

There were several limitations in conducting this study. The first was the crime data used for the models. As described in section 3.3, the data used for the crime counts were NYPD arrest records. Upon examination of the data, it became apparent that the data may not represent actual crimes, but rather police activity. In addition, the spatial autocorrelation of the property crimes indicated spatial randomness, while the violent crimes had some spatial clusters, which is inconsistent with findings in the criminology literature.

This raised the question of whether the spatial resolution of police precincts used was ideal for property crime data. The choice was justified by reported computational problems with approximately 300 different areas. In this study, modeling STARMA with the 77 police precincts was not a computational problem.

Another limitation was the lack of predictive capability of the R package *starma*. This package is a great help in creating STARMA models, but lacks the ability to predict values with the model. Also, the package is no longer supported.

5.3 Future research directions

The limitations of this study open new possibilities for scientific work in the field of spatial crime forecasting and the use of STARMA models:

1. The use of police arrest data has been found to be limited, since they are a reflection of police activity. To address any bias in crime data, future research should examine data not directly related to police activity. An example of such data are emergency call locations. It would also be interesting to explore new ways to accurately locate criminal events using new technologies such as the Internet of Things.

2. The effect of different spatial resolutions of the same data on STARMA models would also be an interesting research topic. For example, which spatial resolutions perform better and how many observations per spatial resolution are ideal for STARMA models could be investigated. In addition, methods for determining ideal spatial representations for the data based on spatial autocorrelation identifiers should be further investigated. Ways to automate the spatial resolution selection process or weighting are also interesting research directions.

3. Future studies could also analyze the effect of different spatial scales on the computational effect of STARMA models. This study could help to further automate the

STARMA process. For example, it could be analyzed at which dimensionality of the input matrix the STARMA R package reaches its limits. This study could improve the current or future R packages regarding the STARMA method.

4. Since this study used always the whole data set (2006-2020) the exploration of different temporal resolutions with the same set of observations could be further investigated. In this case, each case study would have the same number of temporal observations and therefore would allow a further comparison between the different forecasting periods.

5. Due to the widespread use of ARIMA models, there have already been several successful attempts to automate the parameter estimation process of ARIMA models. Further research on the STARMA method may reveal the potential of the method. An automated STARMA estimation process as well as a STARMA prediction tool would further advance research on the STARMA method. Further automation processes in determining the best spatial resolution or spatial weighting for each model would further enhance the use of STARMA models. The main challenge is that the definition of space is very different compared to time. While time is well defined in terms of its length and intervals, defining the "right" spatial resolution is more complex. But this challenge is also the reason why spatial studies are so fascinating. Given the constant technological development, I am hopeful that one day it will be as easy to create a STARMA model as an ARIMA model.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Akers, R. L. (2012). *Criminological theories: Introduction and evaluation* (2nd ed.). Routledge. https://doi.org/10.4324/9781315062723

Andresen, M. A. (2020). *Environmental criminology: Evolution, theory, and practice* (2nd ed.). Routledge.

Andresen, M. A., & Malleson, N. (2013). Crime seasonality and its variations across space. *Applied Geography*, *43*, 25–35. https://doi.org/https://doi.org/10.1016/j.apgeog.2013.06.007

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic.

Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L. (2017). *The geoda book: Exploring spatial data*. GeoDa Press LLC.

Anselin, L. (2020). *Local spatial autocorrelation (1): Lisa and local moran*. Retrieved January 26, 2022, from https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html#references

Anselin, L., & Rey, S. J. (2014). *Modern spatial econometrics in practice: A guide to geoda, geodaspace and pysal*. GeoDa Press LLC.

Bachner, J. (2013). *Predictive policing: Preventing crime with data and analytics* (Improving Performance Series). IBM Center for The Business of Government. Retrieved January 17, 2022, from https://www.businessofgovernment.org/sites/default/files/Predictive%20Policing.pdf

Balbi, A., & Guerry, A.-M. (1829). *Statistique comparée de l'état de l'instruction et du nombre des crimes dans les divers arrondissements des académies et des cours royales de france*. Retrieved July 10, 2021, from https://gallica.bnf.fr/ark:/12148/btv1b53093802z#

Bernasco, W., & Nieuwbeerta, P. (2005). How do residential burglars select target areas? : A new approach to the analysis of criminal location choice. *The British Journal of Criminology*, *45*(3), 296–315. https://doi.org/10.1093/bjc/azh070

Bertrand, N. (2015, January 5). There's been a staggering drop in arrests in new york for the second week in a row. *Business Insider*. Retrieved December 8, 2021, from https://www.businessinsider.com/theres-been-a-drastic-drop-in-nypd-arrests-2015-1

Borovkova, S., Lopuhaä, H. P., & Ruchjana, B. N. (2008). Consistency and asymptotic normality of least squares estimators in generalized star models. *Statistica Neerlandica*, *62*(4), 482–508. https://doi.org/10.1111/j.1467-9574.2008.00391.x

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: Forecasting and control* (4th ed.). Wiley. https://doi.org/10.1002/9781118619193

Brantingham, P. L., & Brantingham, P. J. (1975). Residential burglary and urban form. *Urban Studies*, *12*(3), 273–284. https://doi.org/10.1080/00420987520080531

Brantingham, P. L., & Brantingham, P. J. (1978). A topological technique for regionalization. *Environment and Behavior*, *10*(3), 335–353. https://doi.org/10.1177/0013916578103004

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, *13*(1), 3–28. https://doi.org/10.1016/S0272-4944(05)80212-9

Brantingham, P. J., & Brantingham, P. L. (1981). Notes on the geometry of crime. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental criminology* (pp. 27–54). Waveland Press.

Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in crime*. Macmillan.

Brantingham, P. J., & Brantingham, P. L. (2008). The geometry of crime and crime pattern theory. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (Second Edition, pp. 98–115). Taylor & Francis Group.

Brown, D. E., & Oxford, R. B. (2001). Data mining time series with applications to crime analysis. *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*, *3*, 1453–1458. https://doi.org/10.1109/ICSMC.2001.973487

Burgess, E. W. (1925). The growth of the city. In R. E. Park & E. W. Burgess (Eds.), *The city*. University of Chicago Press.

Burinsma, G. J. N., & Johnson, S. D. (2018). Environmental criminology: Scope, history, and state of the art. In G. J. N. Burinsma & S. D. Johnson (Eds.), *The oxford handbook of environmental criminology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190279707.013.38

Caplan, J. M., & Kennedy, L. W. (2010). *Risk terrain modeling manual*. Rutgers Center on Public Security.

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*, *53*, 62–74. https://doi.org/https://doi.org/10.1016/j.pmcj.2019.01.003

Cesario, E., Catlett, C., & Talia, D. (2016). Forecasting crimes using autoregressive models. *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. 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*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chainey, S., & Ratcliffe, J. (2013). Spatial theories of crime. *Gis and crime mapping*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118685181

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, *21*(1), 4–28. https://doi.org/10.1057/palgrave.sj.8350066

Chen, P., Yuan, H., & Shu, X. (2008). Forecasting crime using the arima model. *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, *5*, 627–630. https://doi.org/10.1109/FSKD.2008.222

Cheysson, F. (2016, February 11). *Modelling space time autoregressive moving average (starma) processes*. Version 1.3. Retrieved November 11, 2021, from https://cran.r-project.org/web/packages/starma/starma.pdf

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Cipra, T., & Motyková, I. (1987). Study on kalman filter in time series analysis. *Commentationes Mathematicae Universitatis Carolinae*, *28*(3), 549–563.

the City of New York. (2021a). *Nypd arrest incident level data footnotes*. Retrieved October 30, 2021, from https://data.cityofnewyork.us/api/views/8h9b-rp9u/files/1e24ad70-9ad6-449f-8bae-8e05f3d50533?download=true&filename=NYPD_Arrest_Incident_Level_Data_Footnotes.pdf

the City of New York. (2021b). *Nypd arrests data (historic)*. Retrieved October 29, 2021, from https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u

Cleve, J., & Lämmel, U. (2020). *Data mining* (3rd ed.). De Gruyter Oldenbourg. https://doi.org/10.1515/9783110676273

Cliff, A. D., Haggett, P., Ord, J. K., Bassett, K. A., & Davies, R. B. (1975). *Elements of spatial structure: A quantitative approach*. Cambridge University Press.

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes : Models & applications*. Pion.

Cohen, J., Gorr, W., & Durso, C. (2003). *Estimation of crime seasonality: A cross-sectional extension to time series classical decomposition* (H. John Heinz III Working Paper No. 2003-18). Carnegie Mellon University. Pittsburgh, PA.

Cohen, J., Gorr, W. L., & Olligschlaeger, A. M. (2007). Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. *Geographical Analysis*, *39*(1), 105–127. https://doi.org/10.1111/j.1538-4632.2006.00697.x

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American sociological review*, *44*(4), 588–608. https://doi.org/10.2307/2094589

Cornish, D. B., & Clarke, R. V. (1986). *The reasoning criminal: Rational choice perspectives on offending*. Springer.

Cornish, D. B., & Clarke, R. V. (2008). The rational choice perspective. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (Second Edition, pp. 29–61). Taylor & Francis Group.

Cromwell, P., Dunham, R., Akers, R., & Lanza-Kaduce, L. (1995). Routine activities and social control in the aftermath of a natural catastrophe. *European Journal on Criminal Policy and Research*, *3*(3), 56–69. https://doi.org/10.1007/BF02242928

Dash, S. K., Safro, I., & Srinivasamurthy, R. S. (2018). Spatio-temporal prediction of crimes using network analytic approach. *2018 IEEE International Conference on Big Data (Big Data)*, 1912–1917. https://doi.org/10.1109/BigData.2018.8622041

Deutsch, S. J., & Pfeifer, P. E. (1981). Space-time arma modeling with contemporaneously correlated innovations. *Technometrics*, *23*(4), 401–409. https://doi.org/10.1080/00401706.1981.10487686

Ebdon, D. (1977). *Statistics in geography : A practical approach*. Blackwell.

Eck, J. E. (1994). *Drug markets and drug places: A case-control study of the spatial structure of illicit drug dealing* (Doctoral dissertation). Department of Criminology and Criminal Justice, University of Maryland.

Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M., & Wilson, R. E. (2005). *Mapping crime: Understanding hot spots*. U.S. Department of Justice Office of Justice Programs. Retrieved September 15, 2021, from https://www.ojp.gov/pdffiles1/nij/209393.pdf

Eck, J. E., & Madensen, T. D. (2015). Meaningfully and artfully reinterpreting crime for useful science: An essay on the value of building with simple theory. In M. A. Andresen & G. Farrell (Eds.). Palgrave Macmillan UK. https://doi.org/10.1057/9781137391322_5

Felson, M. (1986). Linking criminal choices, routine activities, informal control, and criminal outcomes. In D. B. Cornish & R. V. Clarke (Eds.), *The reasoning criminal: Rational choice perspectives on offending* (pp. 119–128). Springer.

Ferguson, A. G. (2011). Crime mapping and the fourth amendment: Redrawing high-crime areas. *The Hastings Law Journal*, *63*(1), 179–232.

Francescani, C. (2020, June 17). Timeline: The first 100 days of new york gov. andrew cuomo's covid-19 response. *abc News*. Retrieved December 8, 2021, from https://abcnews.go.com/US/News/timeline-100-days-york-gov-andrew-cuomos-covid/story?id=71292880

Furfey, P. H. (1927). A note on lefever's "standard deviational ellipse". *American Journal of Sociology*, *33*(1), 94–98. https://doi.org/10.1086/214336

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, *5*(3), 115–146. https://doi.org/10.2307/2986645

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Giacomini, R., & Granger, C. W. (2004). Aggregation of space-time processes. *Journal of Econometrics*, *118*(1-2), 7–26. https://doi.org/10.1016/S0304-4076(03)00132-5

GitHub. (2021). *Set up git*. Retrieved October 20, 2021, from https://docs.github.com/en/get-started/quickstart/set-up-git

Gorr, W., & Harries, R. (2003). Introduction to crime forecasting. *International Journal of Forecasting*, *19*(4), 551–555. https://doi.org/10.1016/S0169-2070(03)00089-X

Gorr, W., & Olligschlaeger, A. (2002). *Crime hot spot forecasting: Modeling and comparative evaluation, summary* (Research report). Submitted to the U.S. Department of Justice.

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting*, *19*(4), 579–594. https://doi.org/https://doi.org/10.1016/S0169-2070(03)00092-X

Gottwald, T. R., Reynolds, K. M., Campbell, C. L., & Timmer, L. W. (1992). Spatial and spatiotemporal autocorrelation analysis of citrus canker epidemics in citrus nurseries and groves in argentinia. *Phytopathology*, *82*(8), 843–851.

Grubesic, T. H., & Mack, E. A. (2008). Spatio-temporal interaction of urban crime. *Journal of Quantitative Criminology*, *24*(3), 285–306. https://doi.org/10.1007/s10940-008-9047-5

Harries, K. (1999). Mapping crime. principle and practice. *National Institute of Justice*, *164*(3). Retrieved July 10, 2021, from https://www.ojp.gov/pdffiles1/nij/178919.pdf

Harries, R. (2003). Modelling and predicting recorded property crime trends in england and wales - a retrospective. *International Journal of Forecasting*, *19*(4), 557–566. https://doi.org/10.1016/S0169-2070(03)00090-6

Hirschi, T. (1969). *Causes of delinquency*. University of California Press. https://doi.org/10.4324/9781315081649

Hunt, J. (2019). From crime mapping to crime forecasting: The evolution of place-based policing a brief history. *NIJ Journal*, *281*. Retrieved July 10, 2021, from https://nij.ojp.gov/topics/articles/crime-mapping-crime-forecasting-evolution-place-based-policing

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., & Yasmeen, F. (2021, June 1). *Forecasting functions for time series and linear models*. Version 8.15. Retrieved January 3, 2022, from https://cran.r-project.org/web/packages/forecast/forecast.pdf

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, *27*(3).

Ip, R. H., & Li, W. (2017). On some matérn covariance functions for spatio-temporal random fields. *Statistica Sinica*, *27*(2), 805–822. https://doi.org/10.5705/ss.202015.0037

Ismay, C., & Kim, A. Y. (2020). *Statistical inference via data science: A moderndive into r and the tidyverse*. Chapman & Hall/CRC.

Ivaha, C., Al-Madfai, H., Higgs, G., & Ware, A. (2007). The dynamic spatial disaggregation approach: A spatio-temporal modelling of crime. *Lecture Notes in Engineering and Computer Science*, *2166*, 961–966. Retrieved November 3, 2021, from http://www.iaeng.org/publication/WCE2007/WCE2007_pp961-966.pdf

Jacobs, J. B., Friel, C., & Raddick, R. (1999). *Gotham unbound: How new york city was liberated from the grip of organized crime*. New York University Press. https://doi.org/doi:10.18574/9780814737965

Jefferis, E. (1999). A multi-method exploration of crime hot spots: A summary of findings. *Crime Mapping Research Center. National Institute of Justice*.

Johnson, S. D., Bowers, K. J., Birks, D. J., & Pease, K. (2009). Predictive mapping of crime by promap: Accuracy, units of analysis, and the environmental backcloth. In D. Weisburd, W. Bernasco, & G. J. N. Bruinsma (Eds.), *Putting crime in its place: Units of analysis in geographic criminology* (pp. 171–198). Springer. https://doi.org/10.1007/978-0-387-09688-9_8

Jonesi, M. (2012, October 8). Predictive policing a success in santa cruz, calif. *Government Technology*. Retrieved September 29, 2021, from https://www.govtech.com/public-safety/predictive-policing-a-success-in-santa-cruz-calif.html

Jørgensen, A. (2010). The sense of belonging in new urban zones of transition. *Current Sociology*, *58*(1), 3–23. https://doi.org/10.1177/0011392109348542

Kadar, C., & Pletikosa, I. (2018). Mining large-scale human mobility data for long-term crime prediction. *EPJ Data Science*, *7*(1), 26. https://doi.org/10.1140/epjds/s13688-018-0150-z

Kamarianakis, Y., & Prastacos, P. (2005). Space–time modeling of traffic flow. *Computers & Geosciences*, *31*(2), 119–133. https://doi.org/10.1016/j.cageo.2004.05.012

Kazar, B. M., Shekhar, S., Lilja, D. J., Vatsavai, R. R., & Pace, R. K. (2004). Comparing exact and approximate spatial auto-regression model solutions for spatial data analysis. In M. J. Egenhofer, C. Freksa, & H. J. Miller (Eds.), *Giscience 2004. third international conference on geographic information science* (pp. 140–162). Springer. https://doi.org/10.1007/978-3-540-30231-5_10

Kelejian, H., & Piras, G. (2017). Spatial models: Basic issues. *Spatial econometrics*. Elsevier. https://doi.org/10.1016/B978-0-12-813387-3.00001-9

Kennedy, L. W., & Forde, D. R. (1990). Routine activities and crime: An analysis of victimization in canada. *Criminology*, *28*(1), 137–152. https://doi.org/10.1111/j.1745-9125.1990.tb01321.x

Kikuchi, G. (2010). *Neighborhood structures and crime : A spatial analysis*. LFB Scholarly Publishing LLC.

van Koppen, P. J., & Jansen, R. W. J. (1999). The time to rob: Variations in time of number of commercial robberies. *Journal of Research in Crime and Delinquency*, *36*(1). https://doi.org/10.1177/0022427899036001003

Kounadi, O., Ristea, A., Leitner, M., & Jr., A. A. (2020). A systematic review on spatial crime forecasting. *Crime science*, *9*(7), 1–22. 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*, *195*, 2537–2547. https://doi.org/10.1016/j.sbspro.2015.06.441

LeBeau, J. L. (1987). The methods and measures of centrography and the spatial dynamics of rape. *Journal of Quantitative Criminology*, *3*(2), 125–141. https://doi.org/10.1007/BF01064212

Lefever, D. W. (1926). Measuring geographic concentration by means of the standard deviational ellipse. *American Journal of Sociology*, *32*(1), 88–94. https://doi.org/10.1086/214027

Leitner, M., Glasner, P., & Kounadi, O. (2021). Geographies of crime. In J. C. Barnes & D. R. Forde (Eds.), *The encyclopedia of research methods in criminology and criminal justice* (pp. 60–63). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119111931.ch12

Levine, N. (2015). *Crimestat: A spatial statistics program for the analysis of crime incident locations (v 4.02)*. Ned Levine & Associates, Houston, Texas, and the National Institute of Justice, Washington, D.C.

Levitt, S. D. (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. *The Journal of Economic Perspectives*, *18*(1), 163–190. http://www.jstor.org/stable/3216880

Li, G., Haining, R., Richardson, S., & Best, N. (2014). Space-time variability in burglary risk: A bayesian spatio-temporal modelling approach. *Spatial Statistics*, *9*, 180–191. https://doi.org/10.1016/j.spasta.2014.03.006

Liesenfeld, R., Richard, J. F., & Vogler, J. (2017). Likelihood-based inference and prediction in spatio-temporal panel count models for urban crimes. *Journal of Applied Econometrics*, *32*(3), 600–620. https://doi.org/10.1002/jae.2534

Lin, Y. L., Yen, M. F., & Yu, L. C. (2018). Grid-based crime prediction using geographical features. *ISPRS International Journal of Geo-Information*, *7*(8). https://doi.org/10.3390/ijgi7080298

Lutters, W. G., & Ackerman, M. S. (1996). An introduction to the chicago school of sociology. *Interval Research, Proprietary.*

Ly, L., & Hanna, J. (2021, December 23). Times square new year's eve celebration will be scaled back, city says. *CNN*. Retrieved January 17, 2022, from https://edition.cnn.com/2021/12/23/us/nyc-times-square-nye-2022/index.html

Lynch, K. (1960). *The image of the city*. MIT Press.

MacEachren, A. M. (1995). *How maps work: Representation, visualization and design*. Guilford Press.

Madensen, T. D. (2007). *Bar management and crime: Toward a dynamic theory of place management and crime hotspots* (PhD dissertation). University of Cincinnati.

Malleson, N., Heppenstall, A., & See, L. (2010). Crime reduction through simulation: An agent-based model of burglary. *Computers, Environment and Urban Systems*, *34*(3), 236–250. https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2009.10.005

Marchment, Z., & Gill, P. (2021). Systematic review and meta-analysis of risk terrain modelling (rtm) as a spatial forecasting method. *Crime Science*, *10*(1), 12. https://doi.org/10.1186/s40163-021-00149-6

McGrory, K., & Bedi, N. (2020, September 3). Targeted: Pasco's sheriff created a futuristic program to stop crime before it happens. it monitors and harasses families across the county. *Tampa Bay Times*. Retrieved July 10, 2021, from https://www.pulitzer.org/cms/sites/default/files/content/targeted-tampabaytimes-story1.pdf

Merriam-Webster. (2021). *Data mining*. Retrieved September 12, 2021, from https://www.merriam-webster.com/dictionary/data%20mining

Messner, S. F., & Tardiff, K. (1985). The social ecology of urban homicide: An application of the routine activities approach. *Criminology*, *23*(2), 241–268. https://doi.org/10.1111/j.1745-9125.1985.tb00336.x

Militino, A. F., Ugarte, M. D., & García-Reinaldos, L. (2004). Alternative models for describing spatial dependence among dwelling selling prices. *Journal of Real Estate Finance and Economics*, *29*(2), 193–209. https://doi.org/10.1023/B:REAL.0000035310.20223.e9

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108. https://doi.org/10.1198/jasa.2011.ap09546

Monmonier, M. (1991). *How to lie with maps*. University of Chicago Press.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society: Series B (Methodological)*, *10*(2), 243–251. https://doi.org/10.1111/j.2517-6161.1948.tb00012.x

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1-2), 17–23. https://doi.org/10.1093/biomet/37.1-2.17

Neft, D. S. (1966). *Statistical analysis for areal distributions*. Regional Science Research Inst.

Nurhayati, N., Pasaribu, U. S., & Neswan, O. (2012). Application of generalized space-time autoregressive model on gdp data in west european countries (S.-c. Chow, Ed.). *Journal of Probability and Statistics*, *2012*. https://doi.org/10.1155/2012/867056

Ohtsuka, Y., Oga, T., & Kakamu, K. (2010). Forecasting electricity demand in japan: A bayesian spatial autoregressive arma approach. *Computational Statistics & Data Analysis*, *54*(11), 2721–2735. https://doi.org/10.1016/j.csda.2009.06.002

Park, R. E., & Burgess, E. W. (1925). *The city*. University of Chicago Press.

Pawale, P., Bagal, S., Ajabe, S., & Shikalagar, K. (2017). Geo statistical approch for crime hotspot detection and prediction. *International Research Journal of Engineering and Technology*, *4*(5), 2703–2706.

Perry, W. L., McInnis, B., Price, C. C., Smith, S., & S.Hollywood, J. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. RAND Corporation. https://doi.org/10.7249/RR233

Pfeifer, P. E., & Bodily, S. E. (1990). A test of space-time arma modelling and forecasting of hotel data. *Journal of Forecasting*, *9*(3), 255–272. https://doi.org/10.1002/for.3980090305

Pfeifer, P. E., & Deutsch, S. J. (1980). A three-stage iterative procedure for space-time modeling. *Technometrics*, *22*(1), 35–47. https://doi.org/10.1080/00401706.1980.10486099

the QGIS Development Team. (2021, October 20). *Qgis user guide*. Retrieved October 20, 2021, from https://docs.qgis.org/3.16/en/docs/user_manual/preamble/foreword.html

Quetelet, L.-A.-J. (1842). Of the developement of the propensity to crime. *A treatise on man: And the development of his faculties*. William; Robert Chambers.

Rentzelos, A. (2020). *Exploring a space-time autoregressive moving avarage (starma) model in spatial crime forecasting* (Master's thesis). GIMA - Geographical Information Management and Applications.

Reynolds, K. M., & Madden, L. V. (1988). Analysis of epidemics using spatio-temporal autocorrelation. *Phytopathology*, *78*(2), 240–246.

Rodriguez, C. D., Gomez, D. M., & Rey, M. A. (2017). Forecasting time series from clustering by a memetic differential fuzzy approach: An application to crime prediction. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 3372–3379. https://doi.org/10.1109/SSCI.2017.8285373

Sampson, R., Eck, J. E., & Dunham, J. (2010). Super controllers and crime prevention: A routine activity explanation of crime prevention success and failure. *Security Journal*, *23*(1), 37–51. https://doi.org/10.1057/sj.2009.17

Schmid, C. F., & Schmid, S. E. (1972). *Crime in the state of washington* (Statistics). Washington Law and Justice Planning Office.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

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)*, 1–5. https://doi.org/10.1109/ICT-ISPC.2017.8075335

Shaw, C. R., & McKay, H. D. (1931). *Social factors in juvenile delinquency*. U.S. Government Printing Office.

Shaw, C. R., & McKay, H. D. (1942). *Juvenile delinquency and urban areas: A study of rates of delinquency in relation to differential characteristics of local communities in american cities*. University of Chicago Press.

Shaw, C. R., & McKay, H. D. (1969). *Juvenile delinquency and urban areas: A study of rates of delinquency in relation to differential characteristics of local communities in american cities* (revised ed.). University of Chicago Press.

Shekhar, S., Schrater, P., Vatsavai, R. R., WuLi, W., & Chawla, S. (2002). *Spatial contextual classification and prediction models for mining geospatial data* (Technical Report 02-008). Department of Computer Science and Engineering. University of Minnesota. Retrieved September 7, 2021, from https://hdl.handle.net/11299/215512

Sherman, L. W. (1995). Hot spots of crime and criminal careers of places. In J. E. Eck & D. Weisburd (Eds.), *Crime and place* (pp. 35–52). Criminal Justice Press.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, *27*(1). 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*, *29*, 191–201. https://doi.org/10.1016/j.ijforecast.2012.08.002

Shumway, R. H., & Stoffer, D. S. (2000). *Time series analysis and its applications*. Springer.

Song, J., Andresen, M. A., Brantingham, P. L., & Spicer, V. (2017). Crime on the edges: Patterns of crime and land use change. *Cartography and Geographic Information Science*, *44*(1), 51–61. https://doi.org/10.1080/15230406.2015.1089188

Stoica, P., & Selen, Y. (2004). Model-order selection: A review of information criterion rules. *IEEE Signal Processing Magazine*, *21*(4), 36–47. https://doi.org/10.1109/MSP.2004.1311138

Sviatlovsky, E. E., & Eells, W. C. (1937). The centrographical method and regional analysis. *Geographical Review*, *27*(2), 240–254. https://doi.org/10.2307/210093

Thompson, K. (2011). The santa cruz experiment: Can a criminal act be prevented before it begins? by turning its crime problem into a data problem, one city is reinventing police work for the 21st century. *Popular science*, *279*(5).

Tita, G., & Ridgeway, G. (2007). The impact of gang formation on local patterns of crime. *Journal of Research in Crime and Delinquency*, *44*(2), 208–237. https://doi.org/10.1177/0022427806298356

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, *46*, 234–240. https://doi.org/10.2307/143141

Uberti, D. (2021, July 8). After backlash, predictive policing adapts to a changed world. *Wall Street Journal*. Retrieved July 10, 2021, from https://www.wsj.com/articles/after-backlash-predictive-policing-adapts-to-a-changed-world-11625752931

the U.S. Census Bureau. (2020). *Quickfacts: New york city, new york* [Population, census, april 1, 2020]. Retrieved October 28, 2021, from https://www.census.gov/quickfacts/fact/table/newyorkcitynewyork/PST045219

the U.S. Census Bureau. (2021). *Gazetteer files: New york*. Retrieved October 28, 2021, from https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2021_Gazetteer/2021_gaz_place_36.txt

Venables, W. N., Smith, D. M., & the R Core Team. (2021, August 10). *An introduction to r: Notes on r: A programming environment for data analysis and graphics*. Version 4.1.1. Retrieved October 20, 2021, from https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

Wikle, C. K., Zammit-Mangion, A., & Cressie, N. (2019). *Spatio-temporal statistics with r*. Chapman & Hall/CRC.

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*, *30*(1), 79–82. https://doi.org/10.3354/cr030079

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*(7), 557–585.

Young, D. S. (2018). *Handbook of regression methods*. Chapman & Hall/CRC. https://doi.org/10.1201/9781315154701

Zhao, X., & Tang, J. (2017). Modeling temporal-spatial correlations for crime prediction. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 497–506. https://doi.org/10.1145/3132847.3133024

Zhuang, Y., Almeida, M., Morabito, M., & Ding, W. (2017). Crime hot spot forecasting: A recurrent model with spatial and temporal information. *2017 IEEE International Conference on Big Knowledge (ICBK)*, 143–150. https://doi.org/10.1109/ICBK.2017.3

Appendix A

Tables

Table A.1: Column info of NYPD Arrest Data (Historic)

Column Name	Column Description
ARREST_KEY	Randomly generated persistent ID for each arrest
ARREST_DATE	Exact date of arrest for the reported event
PD_CD	Three digit internal classification code (more granular than Key Code)
PD_DESC	Description of internal classification corresponding with PD code (more granular than Offense Description)
KY_CD	Three digit internal classification code (more general category than PD code)
OFNS_DESC	Description of internal classification corresponding with KY code (more general category than PD description)
LAW_CODE	Law code charges corresponding to the NYS Penal Law, VTL and other various local laws
LAW_CAT_CD	Level of offense: felony, misdemeanor, violation
ARREST_BORO	Borough of arrest. B(Bronx), S(Staten Island), K(Brooklyn), M(Manhattan), Q(Queens)
ARREST_PRECINCT	Precinct where the arrest occurred
JURISDICTION_CODE	Jurisdiction responsible for arrest. Jurisdiction codes 0(Patrol), 1(Transit) and 2(Housing) represent NYPD whilst codes 3 and more represent non NYPD jurisdictions
AGE_GROUP	Perpetrator's age within a category
PERP_SEX	Perpetrator's sex description
PERP_RACE	Perpetrator's race description
X_COORD_CD	Midblock X-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
Y_COORD_CD	Midblock Y-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
Latitude	Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
Longitude	Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)

Table A.1: (continued)

Column Name	**Column Description**
Lon_Lat	Latitude and Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)

Appendix B

R Code

B.1 R code for data processing

```
library(here)
library(tidyverse)
library(readr)
library(lubridate)
library(sf)
library(sp)
library(reshape2)
library(spdep)


# Step 1: Input data
crime_data <- read_csv("/Users/dnlrc/Documents/Uni/Masterarbeit/Data/
    NYPD_Arrests_Data__Historic_.csv") # read data
admin_data <- read_sf("/Users/dnlrc/Documents/Uni/Masterarbeit/Data/NY_
    PP/nypp.shp") #the chosen file should be as shape file (".shp"),
     this refers to precincts as study zones

# Step 2: Select relevant attributes
crime_data$DATE <- as.Date(crime_data$ARREST_DATE, format = "%m/%d/%Y")
    #change ARREST_DATE from string to date
cr_sel <- crime_data %>%
          dplyr::select(OFNS_DESC, ARREST_PRECINCT ,X_COORD_CD, Y_COORD_CD
              , DATE)
pp <- admin_data[,c(1,3,4)]# selection of useful attributes (at least
    precinct id, area and their geometry)
pp_list <- as.list(pp$Precinct)

# Step X: Selection of the study period
#cr_sel_start <- cr_sel %>%
```

```
        filter(cr_sel$DATE >= "2006-01-01")
#cr_sel_end <- cr_sel_start %>%
        filter(cr_sel_start$DATE <= "2006-12-31")

# Step 3: Clear N.A. and strange data
clr_temp_1 <- cr_sel  #use cr_sel_end if a subset of date is used
clr_temp_2 <- clr_temp_1 %>%
        drop_na() # delete empty rows
cr_clr <- clr_temp_2[(!(clr_temp_2$OFNS_DESC=="F.C.A. P.I.N.O.S.")),] #
    delete strange data

# Step 4: Spatial Join of crime data & Police Precincts
cr_sf <- cr_clr %>%
        st_as_sf(coords=c("X_COORD_CD", "Y_COORD_CD"), crs=2263, remove=
            FALSE) #create sf objects for spatial join
pp_sf <- pp  %>%
        st_as_sf(wkt="geometry", crs=2263, remove=FALSE)

cr_pp_sj <- st_join(cr_sf, pp_sf, join = st_intersects) #spatial join

cr_pp_sj$Precinct <- ifelse(is.na(cr_pp_sj$Precinct), cr_pp_sj$ARREST_
    PRECINCT, cr_pp_sj$Precinct) #replace not joined point with ARREST_
    PRECINCT info

cr_pp_nc <- cr_pp_sj[,c(1,5,7)] #dropping coordinates and geometry for
    faster processing
cr_pp_ng <- st_drop_geometry(cr_pp_nc)

# Step 5: Categorize arrest data into crime types
div_temp <- cr_pp_ng
div_temp$all <- 1
size <- nrow(div_temp)

for(i in 1:size){
        if(div_temp[i,1]=="ABORTION"){
                div_temp$p[i]<-0 # property = p
                div_temp$v[i]<-0 # violent = v
        }else if(div_temp[i,1]=="ADMINISTRATIVE CODE"){
                div_temp$p[i]<-0
                div_temp$v[i]<-0
        }else if(div_temp[i,1]=="ADMINISTRATIVE CODES"){
                div_temp$p[i]<-0
                div_temp$v[i]<-0
        }else if(div_temp[i,1]=="AGRICULTURE & MRKTS LAW-UNCLASSIFIED"){
```

```
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="ALCOHOLIC BEVERAGE CONTROL LAW"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="ANTICIPATORY OFFENSES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="ARSON"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="ASSAULT 3 & RELATED OFFENSES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="BURGLAR'S TOOLS"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="BURGLARY"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="CHILD ABANDONMENT/NON SUPPORT"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="CHILD ABANDONMENT/NON SUPPORT 1"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="CRIMINAL MISCHIEF & RELATED OF"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="CRIMINAL MISCHIEF & RELATED OFFENSES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="CRIMINAL TRESPASS"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="DANGEROUS DRUGS"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="DANGEROUS WEAPONS"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="DISORDERLY CONDUCT"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
```

```
}else if(div_temp[i,1]=="DISRUPTION␣OF␣A␣RELIGIOUS␣SERV"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="DISRUPTION␣OF␣A␣RELIGIOUS␣SERVICE"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="ENDAN␣WELFARE␣INCOMP"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="ESCAPE␣3"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="FELONY␣ASSAULT"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
}else if(div_temp[i,1]=="FOR␣OTHER␣AUTHORITIES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="FORCIBLE␣TOUCHING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
}else if(div_temp[i,1]=="FORGERY"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="FRAUDS"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="FRAUDULENT␣ACCOSTING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="GAMBLING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="GRAND␣LARCENY"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="GRAND␣LARCENY␣OF␣MOTOR␣VEHICLE"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="HARRASSMENT"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
}else if(div_temp[i,1]=="HARRASSMENT␣2"){
        div_temp$p[i]<-0
```

```
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="HOMICIDE-NEGLIGENT-VEHICLE"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="HOMICIDE-NEGLIGENT,UNCLASSIFIE"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="HOMICIDE-NEGLIGENT,UNCLASSIFIED"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="INTOXICATED & IMPAIRED DRIVING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="INTOXICATED/IMPAIRED DRIVING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="JOSTLING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="KIDNAPPING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="KIDNAPPING & RELATED OFFENSES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="KIDNAPPING AND RELATED OFFENSES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="LOITERING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="LOITERING FOR DRUG PURPOSES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="LOITERING,BEGGING"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="LOITERING/GAMBLING (CARDS, DIC"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="LOITERING/GAMBLING (CARDS, DICE, ETC)")
      {
        div_temp$p[i]<-0
        div_temp$v[i]<-0
```

```
}else if(div_temp[i,1]=="MISCELLANEOUS PENAL LAW"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="MOVING INFRACTIONS"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="MURDER & NON-NEGL. MANSLAUGHTE"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
}else if(div_temp[i,1]=="MURDER & NON-NEGL. MANSLAUGHTER"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
}else if(div_temp[i,1]=="NEW YORK CITY HEALTH CODE"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="NYS LAWS-UNCLASSIFIED FELONY"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="NYS LAWS-UNCLASSIFIED VIOLATION"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="OFF. AGNST PUB ORD SENSBLTY &"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="OFF. AGNST PUB ORD SENSBLTY & RGHTS TO
    PRIV"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="OFFENSES AGAINST MARRIAGE UNCLASSIFIED"
    ){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="OFFENSES AGAINST PUBLIC ADMINI"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="OFFENSES AGAINST PUBLIC ADMINISTRATION"
    ){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="OFFENSES AGAINST PUBLIC SAFETY"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
}else if(div_temp[i,1]=="OFFENSES AGAINST THE PERSON"){
        div_temp$p[i]<-0
```

```
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="OFFENSES INVOLVING FRAUD"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="OFFENSES RELATED TO CHILDREN"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="OTHER OFFENSES RELATED TO THEF"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="OTHER OFFENSES RELATED TO THEFT"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="OTHER STATE LAWS"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="OTHER STATE LAWS (NON PENAL LA"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="OTHER STATE LAWS (NON PENAL LAW)"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="OTHER TRAFFIC INFRACTION"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="PARKING OFFENSES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="PETIT LARCENY"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="POSSESSION OF STOLEN PROPERTY"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="POSSESSION OF STOLEN PROPERTY 5"){
        div_temp$p[i]<-1
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="PROSTITUTION & RELATED OFFENSES"){
        div_temp$p[i]<-0
        div_temp$v[i]<-0
    }else if(div_temp[i,1]=="RAPE"){
        div_temp$p[i]<-0
        div_temp$v[i]<-1
    }else if(div_temp[i,1]=="ROBBERY"){
```

```
            div_temp$p[i]<-0
            div_temp$v[i]<-1
        }else if(div_temp[i,1]=="SEX␣CRIMES"){
            div_temp$p[i]<-0
            div_temp$v[i]<-1
        }else if(div_temp[i,1]=="THEFT␣OF␣SERVICES"){
            div_temp$p[i]<-1
            div_temp$v[i]<-0
        }else if(div_temp[i,1]=="THEFT-FRAUD"){
            div_temp$p[i]<-1
            div_temp$v[i]<-0
        }else if(div_temp[i,1]=="UNAUTHORIZED␣USE␣OF␣A␣VEHICLE"){
            div_temp$p[i]<-1
            div_temp$v[i]<-0
        }else if(div_temp[i,1]=="UNAUTHORIZED␣USE␣OF␣A␣VEHICLE␣3␣(UUV)")
            {
            div_temp$p[i]<-1
            div_temp$v[i]<-0
        }else if(div_temp[i,1]=="UNDER␣THE␣INFLUENCE,␣DRUGS"){
            div_temp$p[i]<-0
            div_temp$v[i]<-0
        }else if(div_temp[i,1]=="UNLAWFUL␣POSS.␣WEAP.␣ON␣SCHOOL"){
            div_temp$p[i]<-0
            div_temp$v[i]<-0
        }else if(div_temp[i,1]=="UNLAWFUL␣POSS.␣WEAP.␣ON␣SCHOOL␣GROUNDS"
            ){
            div_temp$p[i]<-0
            div_temp$v[i]<-0
        }else if(div_temp[i,1]=="VEHICLE␣AND␣TRAFFIC␣LAWS"){
            div_temp$p[i]<-0
            div_temp$v[i]<-0
        }
}

cr_div <- div_temp

# Step 6: Extract information about temporal resolution +
        # Step 7: group & sum the data +
                # Step 8: Transpose data for model implementation

# 1 week
cr_pp_w <- cr_div
cr_pp_w$tp <- floor_date(cr_pp_w$DATE, unit = "week", 1)
```

```
cr_pp_sum_w <- cr_pp_w %>%
        group_by(tp, Precinct) %>% # group by the unique code of
            precincts & time period
                summarize(sumall = sum(all), sump = sum(p), sumv = sum(v
                    )) #sum the crime

cr_pp_all_w <- cr_pp_sum_w[,c(1,2,3)]
cr_pp_p_w <- cr_pp_sum_w[,c(1,2,4)]
cr_pp_v_w <- cr_pp_sum_w[,c(1,2,5)]

cr_pp_sts_w_all <- dcast(cr_pp_all_w, tp~Precinct, value.var = "sumall")
    #transpose for crime type all
cr_pp_sts_w_all[is.na(cr_pp_sts_w_all)] <- 0 #replace NA with 0
cr_pp_sts_w_p <- dcast(cr_pp_p_w, tp~Precinct, value.var = "sump") #
    transpose for crime type p
cr_pp_sts_w_p[is.na(cr_pp_sts_w_p)] <- 0
cr_pp_sts_w_v <- dcast(cr_pp_v_w, tp~Precinct, value.var = "sumv") #
    transpose for crime type v
cr_pp_sts_w_v[is.na(cr_pp_sts_w_v)] <- 0


# 1 month
cr_pp_1m <- cr_div
cr_pp_1m$tp <- floor_date(cr_pp_1m$DATE, unit = "month")

cr_pp_sum_1m <- cr_pp_1m %>%
        group_by(tp, Precinct) %>% # group by the unique code of
            precincts & time period
                summarize(sumall = sum(all), sump = sum(p), sumv = sum(v
                    )) #sum the crime

cr_pp_all_1m <- cr_pp_sum_1m[,c(1,2,3)]
cr_pp_p_1m <- cr_pp_sum_1m[,c(1,2,4)]
cr_pp_v_1m <- cr_pp_sum_1m[,c(1,2,5)]

cr_pp_sts_1m_all <- dcast(cr_pp_all_1m, tp~Precinct, value.var = "sumall
    ") #transpose for crime type all
cr_pp_sts_1m_p <- dcast(cr_pp_p_1m, tp~Precinct, value.var = "sump") #
    transpose for crime type p
cr_pp_sts_1m_v <- dcast(cr_pp_v_1m, tp~Precinct, value.var = "sumv") #
    transpose for crime type v

# 3 months
cr_pp_3m <- cr_div
```

```
cr_pp_3m$tp <- floor_date(cr_pp_3m$DATE, unit = "quarter")

cr_pp_sum_3m <- cr_pp_3m %>%
        group_by(tp, Precinct) %>% # group by the unique code of
            precincts & time period
                summarize(sumall = sum(all), sump = sum(p), sumv = sum(v
                    )) #sum the crime

cr_pp_all_3m <- cr_pp_sum_3m[,c(1,2,3)]
cr_pp_p_3m <- cr_pp_sum_3m[,c(1,2,4)]
cr_pp_v_3m <- cr_pp_sum_3m[,c(1,2,5)]

cr_pp_sts_3m_all <- dcast(cr_pp_all_3m, tp~Precinct, value.var = "sumall
    ") #transpose for crime type all
cr_pp_sts_3m_p <- dcast(cr_pp_p_3m, tp~Precinct, value.var = "sump") #
    transpose for crime type p
cr_pp_sts_3m_v <- dcast(cr_pp_v_3m, tp~Precinct, value.var = "sumv") #
    transpose for crime type v

# 6 months

cr_pp_6m <- cr_div
cr_pp_6m$tp <- floor_date(cr_pp_6m$DATE, unit = "halfyear")

cr_pp_sum_6m <- cr_pp_6m %>%
        group_by(tp, Precinct) %>% # group by the unique code of
            precincts & time period
                summarize(sumall = sum(all), sump = sum(p), sumv = sum(v
                    )) #sum the crime

cr_pp_all_6m <- cr_pp_sum_6m[,c(1,2,3)]
cr_pp_p_6m <- cr_pp_sum_6m[,c(1,2,4)]
cr_pp_v_6m <- cr_pp_sum_6m[,c(1,2,5)]

cr_pp_sts_6m_all <- dcast(cr_pp_all_6m, tp~Precinct, value.var = "sumall
    ") #transpose for crime type all
cr_pp_sts_6m_p <- dcast(cr_pp_p_6m, tp~Precinct, value.var = "sump") #
    transpose for crime type p
cr_pp_sts_6m_v <- dcast(cr_pp_v_6m, tp~Precinct, value.var = "sumv") #
    transpose for crime type v

# 1 year
cr_pp_y <- cr_div
cr_pp_y$tp <- format(as.Date(cr_div$DATE, format="%Y-%m-%d"),"%Y")
```

```
cr_pp_sum_y <- cr_pp_y %>%
        group_by(tp, Precinct) %>% # group by the unique code of
            precincts & time period
                summarize(sumall = sum(all), sump = sum(p), sumv = sum(v
                    )) #sum the crime

cr_pp_all_y <- cr_pp_sum_y[,c(1,2,3)]
cr_pp_p_y <- cr_pp_sum_y[,c(1,2,4)]
cr_pp_v_y <- cr_pp_sum_y[,c(1,2,5)]

cr_pp_sts_y_all <- dcast(cr_pp_all_y, tp~Precinct, value.var = "sumall")
    #transpose for crime type all
cr_pp_sts_y_p <- dcast(cr_pp_p_y, tp~Precinct, value.var = "sump") #
    transpose for crime type p
cr_pp_sts_y_v <- dcast(cr_pp_v_y, tp~Precinct, value.var = "sumv") #
    transpose for crime type v

# Step 9: Replace geometry of polygons with their centroid point
pp_centr <- st_centroid(pp) #get centroid of police precincts
pp_geom<- pp_centr [,c(1,3)]
pp_XY <- pp_centr %>%
        st_coordinates() %>% as.data.frame() # get X and Y coordinates
            of the centroids of police precincts
pp_XY_mtx <- as.matrix(pp_XY)

cr_pp <- merge(cr_pp_sum,pp_geom, by="Precinct")
st_write(cr_pp, "/Users/dnlrc/Documents/Uni/Masterarbeit/Data/R_Outputs/
    cr_pp.shp", append=FALSE)
```

B.2 R code for implementing the models

```
library(spdep)
library(starma)
library(tmap)
library(janitor)
library(Metrics)
library(stats)
library(forecast)


### Create the spatial weight matrix ###
# load shape of police precicts with bridges
```

```
admin_data_b <- read_sf("/Users/dnlrc/Documents/Uni/Masterarbeit/Data/NY
    _PP/nypp_bridges.shp")
pp_b_sf <- pp_b  %>% st_as_sf(wkt="geometry", crs=2263, remove=FALSE)
pp_b <- admin_data_b[,c(1,3,4)]
names(pp_b)[names(pp_b) == 'POLY_ID'] <- 'Precinct'

# transpose data for input in GeoDa (row-wise are precincts, column time
    )
cr_pp_all_ts_y <- dcast(cr_pp_all_y, Precinct~tp, value.var = "sumall")
    #transpose for crime type all
cr_pp_all_ts_y$Precinct <- 1:77
cr_pp_p_ts_y <- dcast(cr_pp_p_y, Precinct~tp, value.var = "sump") #
    transpose for crime type p
cr_pp_p_ts_y$Precinct <- 1:77
cr_pp_v_ts_y <- dcast(cr_pp_v_y, Precinct~tp, value.var = "sumv") #
    transpose for crime type v
cr_pp_v_ts_y$Precinct <- 1:77

# write output file
cr_pp_poly <- merge(cr_pp_all_ts_y,pp_b, by="Precinct")
cr_pp_poly <- merge(cr_pp_p_ts_y,pp_b, by="Precinct")
cr_pp_poly <- merge(cr_pp_v_ts_y,pp_b, by="Precinct")
st_write(cr_pp_poly, "/Users/dnlrc/Documents/Uni/Masterarbeit/Data/R_
    Outputs/GeoDa_Xplore/cr_pp_w.shp", append=FALSE)


k2_nb <- knearneigh(pp_XY_mtx, k=2)
k2_nb <- knn2nb(k2_nb)
k2_nb <- nblag(k2_nb, 2)
k2_list <- list(order0=diag(77), # the number corresponds to the amount
    of study zones that used
order1=nb2mat(k2_nb[[1]], zero.policy=TRUE),
order2=nb2mat(k2_nb[[2]], zero.policy=TRUE))

### STARMA modeling ###

# The process is repeated for each crime type:

# Normalize the data
input_data_starma <- cr_pp_sts_w_p [,c(2:78)] #paste the transpose data
    frame here, the date column will be dropped
colnames(input_data_starma) <- c(1:77) #rename the police precincts into
     numbers from 1 to 77
cr_norm <- stcenter(input_data_starma) #stcenter centers and scales the
```

```
    space-time series data such that its mean is 0 and its standard
    error 1.
timesteps <- nrow(input_data_starma)

# insert spatial weights list here
wlist <- k1_list

# 1 Identification: Using stacf and stpacf, the user should try to
    identify which parameters should be estimated.
I_stacf <- stacf(cr_norm, wlist)
I_stpacf <- stpacf(cr_norm, wlist)

# 2 Estimation: Use starma to estimate the parameters.
# set AR and MA parameters to 1 for the first run
ar <- 1
ma <- 1

# AR parameters
ar <- matrix(0, 12, 2) #row -th tlag #col -th slag
ar[1,1] <- 1 #set AR parameter of spatial lag 0
ar[1,2] <- 1 #set AR parameter of spatial lag 1

# MA parameters
ma <- matrix(0, 12, 2) #row -th tlag #col -th slag
ma[3,1] <-1 #set MA parameter of spatial lag 0
ma[1,2] <- 1 #set AR parameter of spatial lag 1

# Run the Kalman filter algorithm
model <- starma(cr_norm, wlist, ar, ma, iterate=5)
model
summary(model)
D_stpacf <- stpacf(model$residuals, wlist, tlag.max = 53)

# 3 Diagnose the process. Go back if the residuals show autocorrelation
D_stacf <- stacf(model$residuals, wlist)
D_stpacf <- stpacf(model$residuals, wlist)
D_stcor_test <- stcor.test(model$residuals, wlist)

#Calculate error metrics

res_starma <- model$residuals
cr_actl <- t(cr_norm)
cr_prdct <- res_starma + cr_norm
cr_prdct <- t(cr_prdct)
```

```
cr_actl <-  cr_actl[,c(5:timesteps)]
cr_prdct <- cr_prdct[,c(5:timesteps)]
acc_starma <- data.frame(matrix(NA, ncol = 3))
colnames(acc_starma)<- c("RMSE","MAE","R2")

#Function for R squared
rsq <- function (x, y) cor(x, y) ^ 2

for(i in 1:nrow(cr_prdct))  {
        temp_actl <- cr_actl [i,]
        temp_prdct <- cr_prdct [i,]
        temp_RMSE <- rmse(temp_actl, temp_prdct)
        temp_MAE <- mae(temp_actl, temp_prdct)
        temp_R2 <- rsq(temp_actl, temp_prdct)
        temp_tbl <- cbind(temp_RMSE, temp_MAE, temp_R2)
        colnames(temp_tbl)<- c("RMSE","MAE","R2")
        acc_starma <- rbind(acc_starma, temp_tbl)
        rm(temp_actl,temp_prdct,temp_RMSE,temp_R2,temp_tbl)
}
acc_starma <- acc_starma [c(2:78),]
rownames(acc_starma) <- 1:77

acc_starma_sum <- data.frame(matrix(NA, ncol = 6))
colnames(acc_starma_sum)<- c("RMSE","sd_RMSE","MAE","sd_MAE","R2","sd_R2
    ")
acc_starma_sum$RMSE <- mean(acc_starma[,1])
acc_starma_sum$sd_RMSE <- sd(acc_starma[,1])
acc_starma_sum$MAE <- mean(acc_starma[,2])
acc_starma_sum$sd_MAE <- sd(acc_starma[,2])
acc_starma_sum$R2 <- mean(acc_starma[,3])
acc_starma_sum$sd_R2 <- sd(acc_starma[,3])

# Save model & data
STARMA_w_p <- model
AR_w_p <- ar
MA_w_p <- ma
stacf_w_p <- D_stacf
stpacf_w_p <- D_stpacf
stcor_test_w_p <- D_stcor_test
acc_starma_w_p <- acc_starma
acc_starma_sum_w_p <- acc_starma_sum

### ARIMA modeling ###
```

```
input_data_arima <- cr_pp_sts_y_v [,c(2:78)] #paste the transpose data
    frame here, the time column will be dropped
colnames(input_data_arima) <- c(1:77) #rename the police precincts into
    numbers from 1 to 77
input_data_arima<- stcenter(input_data_arima) #stcenter centers and
    scales the space-time series data such that its mean is 0 and its
    standard error 1.
timesteps <- nrow(input_data_arima)

acc_arima <- data.frame(matrix(NA, ncol = 3))# create a data frame for
    the calculate error metrics
colnames(acc_arima)<- c("RMSE","MAE","R2")
res_arima <- data.frame(matrix(NA, ncol = timesteps))

for(i in 1:ncol(input_data_arima)) {          #function that calculates
    an ARIMA model for all police precincts and then calculates the
    error metrics
        temp <- data.frame(input_data_arima[,i])
        colnames(temp) <- colnames(input_data_arima)[i]
        temp <- ts(temp, start = 1, end = timesteps, frequency = 1)
        temp_arima <- auto.arima(temp, max.order= 3)
        temp_res <- temp_arima$residuals
        temp_prdct <- temp_arima$residuals + temp
        temp_actl <- as.vector(temp)
        temp_prdct <- as.vector(temp_prdct)
        temp_actl <-  temp_actl[c(4:timesteps)]
        temp_prdct <- temp_prdct[c(4:timesteps)]
        temp_RMSE <- rmse(temp_actl, temp_prdct)
        temp_MAE <- mae(temp_actl, temp_prdct)
        temp_R2 <- rsq(temp_actl, temp_prdct)
        temp_tbl <- cbind(temp_RMSE, temp_MAE, temp_R2)
        colnames(temp_tbl)<- c("RMSE","MAE","R2")
        acc_arima <- rbind(acc_arima, temp_tbl)
        res_arima <- rbind(res_arima, temp_res)
        rm(temp,temp_arima,temp_prdct,temp_RMSE,temp_R2,temp_tbl,temp_
            res)
}

# Calculate ARIMA error metrics
acc_arima <- acc_arima [c(2:78),]
rownames(acc_arima) <- 1:77
res_arima <- res_arima [c(2:78),]
rownames(res_arima) <- 1:77
```

```
acc_arima_sum <- data.frame(matrix(NA, ncol = 6))
colnames(acc_arima_sum)<- c("RMSE","sd_RMSE","MAE","sd_MAE","R2","sd_R2"
    )
acc_arima_sum$RMSE <- mean(acc_arima[,1])
acc_arima_sum$sd_RMSE <- sd(acc_arima[,1])
acc_arima_sum$MAE <- mean(acc_arima[,2])
acc_arima_sum$sd_MAE <- sd(acc_arima[,2])
acc_arima_sum$R2 <- mean(acc_arima[,3])
acc_arima_sum$sd_R2 <- sd(acc_arima[,3])

# Save model & data
acc_arima_y_v <- acc_arima
acc_arima_sum_y_v <- acc_arima_sum
res_arima_y_v <- res_arima
view(acc_arima_sum_y_v)
```

B.3 R code to visualize the results

```
library(ggplot2)


### Code to create Figure 4.2 ###
cr_all <- cr_pp_all
cr_all$tp <- paste("01",cr_all$tp ,sep = "/")
cr_all$tp <- as.Date(cr_all$tp, format="%d/%Y/%m")
count_pp <- unique(cr_all$Precinct)
count_pp_s <- c(10,20,41,50,52,66,69,109,114,121)
cr_max_pp_s <- filter(cr_all, Precinct %in% count_pp_s)

cr_max_TSP <- ggplot(cr_max_pp_s) +
geom_line(aes(x = tp, y = sumall), size=.3) +
facet_wrap(~Precinct, ncol = 5) +
xlab("Year") +
ylab("Monthly arrest counts") +
theme_bw() +
theme(panel.spacing = unit (1, "lines")) +
scale_x_date(date_labels = "%Y", date_minor_breaks = "1 year",date_
    breaks = "5 years")

print(cr_max_TSP)

### Code to create Figure 3.7 ###
cr_max_1w <- cr_pp_sum_w
cr_max_agg_1w <- aggregate(cr_max_1w["sumall"], by=cr_max_1w["tp"], sum)
```

```
cr_max_agg_1w$tp <- as.Date(cr_max_agg_1w$tp, format="%Y-%m-%d")

cr_max_1m <- cr_pp_sum_1m
cr_max_agg_1m <- aggregate(cr_max_1m["sumall"], by=cr_max_1m["tp"], sum)
cr_max_agg_1m$tp <- as.Date(cr_max_agg_1m$tp, format="%Y-%m-%d")

cr_max_3m <- cr_pp_sum_3m
cr_max_agg_3m <- aggregate(cr_max_3m["sumall"], by=cr_max_3m["tp"], sum)
cr_max_agg_3m$tp <- as.Date(cr_max_agg_3m$tp, format="%Y-%m-%d")

cr_max_6m <- cr_pp_sum_6m
cr_max_agg_6m <- aggregate(cr_max_6m["sumall"], by=cr_max_6m["tp"], sum)
cr_max_agg_6m$tp <- as.Date(cr_max_agg_6m$tp, format="%Y-%m-%d")

cr_max_y <- cr_pp_sum_y
cr_max_agg_y <- aggregate(cr_max_y["sumall"], by=cr_max_y["tp"], sum)
cr_max_agg_y$tp <- as.Date(cr_max_agg_y$tp, format="%Y")

cr_max_TSP_p1 <- ggplot() +
#geom_line(data=cr_max_agg_1w, aes(x = tp, y = sumall), color="grey30",
    size=.3) +
#geom_line(data=cr_max_agg_1m, aes(x = tp, y = sumall), color="grey30",
    size=.5) +
#geom_line(data=cr_max_agg_6m, aes(x = tp, y = sumall), color="grey30",
    size=.5) +
#geom_line(data=cr_max_agg_3m, aes(x = tp, y = sumall), color="grey30",
    size=.5) +
geom_line(data=cr_max_agg_y, aes(x = tp, y = sumall), color="grey30",
    size=.5) +
xlab("Year") +
ylab("All arrest counts") +
theme(axis.title.y = element_text(size = 15)) +
theme(axis.title.x = element_text(size = 15)) +
scale_x_date(date_labels = "%Y",date_minor_breaks = "1 month", date_
    breaks = "1 year")

print(cr_max_TSP_p1)


### Code to create residual maps of chapter 4 ###
pp_sf_poly <- mutate(pp, POLY_ID = row_number())

# STARMA
res_starma <- STARMA_y_all$residuals
```

```
starma_breaks <- scale(res_starma)
starma_breaks <- as.data.frame(t(starma_breaks))
res_g <- mutate(starma_breaks, POLY_ID = row_number())
res_g <- merge(y=res_g, x=pp_sf_poly, by="POLY_ID")
res_sf <- res_g %>%
          st_as_sf(wkt="geometry", crs=2263, remove=FALSE)
st_write(res_sf, "/Users/dnlrc/Documents/Uni/Masterarbeit/Data/R_Outputs
    /res_sf.shp", append=FALSE)
starma_max <- max(starma_breaks$V15)
starma_min <- min(starma_breaks$V15)
starma_breaks <- c(starma_min,-2,-1,starma_max)
mypal_starma <- c('#2166ac','#67a9cf','#d1e5f0','#f7f7f7','#fddbc7','#
    ef8a62','#b2182b')
mypal_starma <- c('#67a9cf','#d1e5f0','#f7f7f7')

tm_shape(res_sf) +
tm_fill("V15", title = "SD␣OF␣STARMA_y_all␣RES.", style = "fixed",
    breaks = starma_breaks, midpoint = 0 ,palette = mypal_starma) +
tm_borders(alpha = 0.1) +
tm_scale_bar(lwd = 0.5,text.size = 1) +
tm_layout(legend.position = c("left", "top"), legend.title.size = 1.2,
    legend.text.size = 1)

tmap_save(tm_shape(res_sf) +
tm_fill("V15", title = "SD␣OF␣STARMA_y_all␣RES.", style = "fixed",
    breaks = starma_breaks, midpoint = 0 ,palette = mypal_starma) +
tm_borders(alpha = 0.1) +
tm_scale_bar(lwd = 0.5,text.size = 1) +
tm_layout(legend.position = c("left", "top"), legend.title.size = 1.2,
    legend.text.size = 1),
"/Users/dnlrc/Documents/Uni/Masterarbeit/Data/R_Outputs/Graphics/STARMA_
    y_all.jpg")


### ARIMA
arima_breaks <- scale(res_arima_w_p)
arima_breaks <- as.data.frame(arima_breaks)
res_g <- mutate(arima_breaks, POLY_ID = row_number())
res_g <- merge(y=res_g, x=pp_sf_poly, by="POLY_ID")
res_sf <- res_g %>%
          st_as_sf(wkt="geometry", crs=2263, remove=FALSE)
arima_max <- max(arima_breaks$X15)
arima_min <- min(arima_breaks$X15)
arima_breaks <- c(arima_min,-3,-2,-1,1,arima_max)
```

```
mypal_arima <- c('#2166ac','#67a9cf','#d1e5f0','#f7f7f7','#fddbc7','#
    ef8a62','#b2182b')
mypal_arima <- c('#2166ac','#67a9cf','#d1e5f0','#f7f7f7','#fddbc7')

tm_shape(res_sf) +
tm_fill("X15", title = "SD OF ARIMA_y_all RES.", style = "fixed", breaks
     = arima_breaks, midpoint = 0 ,palette = mypal_arima) +
tm_borders(alpha = 0.1) +
tm_scale_bar(lwd = 0.5,text.size = 1) +
tm_layout(legend.position = c("left", "top"), legend.title.size = 1.2,
    legend.text.size = 1)

tmap_save(tm_shape(res_sf) +
tm_fill("X15", title = "SD OF ARIMA_y_all RES.", style = "fixed", breaks
     = arima_breaks, midpoint = 0 ,palette = mypal_arima) +
tm_borders(alpha = 0.1) +
tm_scale_bar(lwd = 0.5,text.size = 1) +
tm_layout(legend.position = c("left", "top"), legend.title.size = 1.2,
    legend.text.size = 1),
"/Users/dnlrc/Documents/Uni/Masterarbeit/Data/R_Outputs/Graphics/ARIMA_y
    _all.jpg")

### STARMA residual output to create LISA cluster maps in GeoDa
res_starma <- STARMA_w_p$residuals
starma_breaks <- res_starma
starma_breaks <- as.data.frame(t(starma_breaks))
res_g_starma <- mutate(starma_breaks, POLY_ID = row_number())
res_g_starma <- merge(y=res_g_starma, x=pp_sf_poly, by="POLY_ID")
res_sf_starma <- res_g_starma %>%
        st_as_sf(wkt="geometry", crs=2263, remove=FALSE)
st_write(res_sf_starma, "/Users/dnlrc/Documents/Uni/Masterarbeit/Data/R_
    Outputs/res_sf_starma.shp", append=FALSE)

### ARIMA residual output to create LISA cluster maps in GeoDa
arima_breaks <- res_arima_w_p
arima_breaks <- as.data.frame(arima_breaks)
res_g_arima <- mutate(arima_breaks, POLY_ID = row_number())
res_g_arima <- merge(y=res_g_arima, x=pp_sf_poly, by="POLY_ID")
res_sf_arima <- res_g_arima %>%
        st_as_sf(wkt="geometry", crs=2263, remove=FALSE)
st_write(res_sf_arima, "/Users/dnlrc/Documents/Uni/Masterarbeit/Data/R_
    Outputs/res_sf_arima.shp", append=FALSE)
```

Eidesstattliche Erklärung

Ich versichere:

- dass ich die Masterarbeit selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.
- dass alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Publikationen entnommen sind, als solche kenntlich gemacht sind.
- dass ich dieses Masterarbeitsthema bisher weder im In- noch im Ausland (einer Beurteilerin/ einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe.
- dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

11/02/2022

Datum Unterschrift