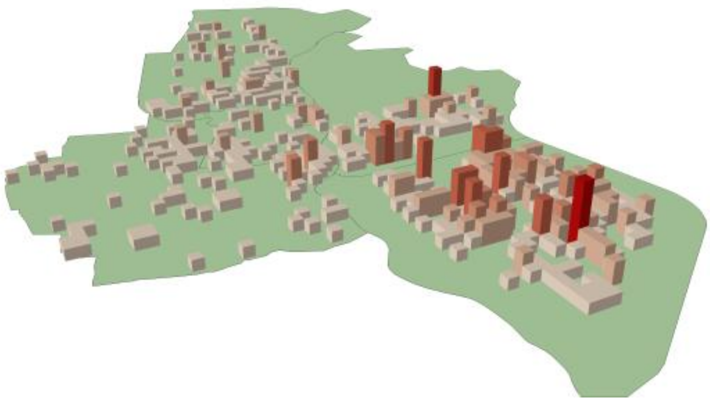
Geographical Information Management and Applications

Smart Cities and The Spatiotemporal Analysis of Residential Burglaries



Author: Roderick Duinker
Date: May 2016
Supervisor: Prof. dr. Stan Geertman

Universiteit Utrecht
TU Delft
University of Twente.
Wageningen UR

Acknowledgments

I wrote this thesis for the Master of Science degree program 'Geographical Information Management and Applications'. I would like to thank several people who helped me in the process of writing this thesis.

Firstly, I would like to thank my supervisor prof. dr. Stan Geertman of the Faculty of Geosciences, Utrecht University. He helped me set out my research and gave directions along the way.

Secondly, I would like to thank Vicrea Solutions for giving me an internship at their company and bringing me in close contact with the municipality of Haarlem. In particular, I would like to thank Flip van der Valk who was my supervisor with Vicrea. He gave me the inspiration for the research subject, but also helped me see my own potential.

Thirdly, I would like to thank the municipality of Haarlem, and in particular Jan Koers, for giving me an internship and providing me with the necessary data for my research. I would also like to thank Rolph de Jong who helped me to get on my way in Haarlem and all his colleagues who invested time and effort to help me out with my research.

Finally, I would like to thank all who I consulted during my research. Special thanks to prof. dr. Wim Bernasco, Senior researcher at the Netherlands Institute for the Study of Crime and Law Enforcement and dr. Derek Karssenberg and dr. Marco Helbich, both with Utrecht University.

I put a significant amount of time and effort into writing this thesis and I certainly learned a lot during the process. I hope that you, as the reader, learn something as well and will have an enjoyable read.

Roderick Duinker
Zeewolde, 21 May 2016

Summary

The smart city is about achieving sustainable urban growth through a participatory government that is investing in human and social capital, traditional (transport) and modern (ICT) infrastructure. It uses modern information and communication technology like sensor networks, data mining and cloud computing for the automatic discovery of data patterns, rules and knowledge to ultimately help solve urban issues. But there is also critique stating that smart cities are just a hype fueled by place marketing motives and that they violate the privacy of citizens. Is the smart city simply a hype or are there any real and practical benefits to applying smart city concepts?

One common urban issue is public safety. The public safety and perceived level of safety are greatly influenced by a specific type of crime: residential burglaries. Burglaries are often not evenly distributed across both space and time: there are often concentrations in certain areas and during certain seasons. What are the explanations for these observed spatial and temporal patterns of burglaries?

This study combines the theoretical concepts of smart cities and the real world social issue of residential burglaries. The goal is to assess the practical usefulness of smart city concepts by finding a suitable method for explaining the structural spatiotemporal patterns of burglaries and applying this method in the Dutch municipality of Haarlem.

In short, the method used for the spatiotemporal analysis of burglaries involves the following steps: identifying the potential risk factors; defining the study area and operationalizing the risk factors; building a negative binomial regression model while keeping in mind issues like multicollinearity, spatial autocorrelation and seasonality; and finally creating a seasonal risk terrain surface.

To identify the risk factors, the assumption is made that crime events can be traced back to a combination of three components: a target, an offender and the setting. For each of these components, theories from spatial criminology were consulted. These theories included 'rational choice' and 'optimal foraging', 'awareness space' and 'offender neighborhood', and 'social disorganization' and 'environmental design'. Based on these theories 30 potential risk factors for burglaries are identified.

These risk factors were all operationalized to a spatial grid of 100 by 100 meter cells of the municipality of Haarlem, including the number of burglaries

per cell. Next, a negative binomial regression model for each season is created. Negative binomial regression is suitable for finding the structural causes of crime and include significance tests to assess and quantify the explanatory power of multiple independent variables. Spatial autocorrelation, a common effect in spatial statistics where outcome values of one cell are affected by other cells close by, is corrected for by adding a spatial lag variable to the model.

From the 30 risk factors that entered the model building process, 16 risk factors turned out to be significant in explaining variations in burglary rates in at least one season. The most consistent risk factors turned out to be: the distance to areas with high ethnic heterogeneity, the building density within an area and the share of risky properties (detached, semi-detached and corner properties) in an area. The assumption that there are three types of risk factor (target, offender and setting) responsible for burglaries seems to hold true, although there seem to be some seasonal effects in play as well. Most burglaries occur during the winter. The significant risk factors and the underlying theories differ from season to season. These differences in risk factor types per season can hint towards different types of offenders which are active during different seasons. Target risk factors are usually more associated with rational offenders, where offender and setting risk factors are more associated with irrational offenders.

Based on the results of the negative binomial regression models, a risk terrain surface was created for each season. Risk terrain maps assist in strategic decision making and tactical action by showing where conditions are ideal for events to occur in the future. Separate map layers representing the presence, absence, or intensity of each significant risk factor at every place throughout a terrain is created, and then all map layers are combined using weights determined by the regression analysis to produce a composite 'risk terrain' map with attribute values that account for all risk factors at every place throughout the geography.

The persistent hot spot in the southeastern district called *Schalkwijk* can largely be explained by the presence of multiple offender risk factors. Many characteristics of an offender neighborhood are found in *Schalkwijk*: many people receiving welfare benefits, a high ethnic heterogeneity, many people in the demographic risk group (males aged 15 to 24) and a high percentage of rental properties. Potential future hot spots, based on a combination of several risk factors, can be found in the neighborhoods *Slachthuisbuurt*, *Spaarndam* and *Koninginnebuurt*.

The spatiotemporal analysis of burglaries in Haarlem showed that smart city concepts are not only a hype and can benefit local governments and citizens in multiple ways. The smart city can act as facilitator, motivator and promoter.

The smart city as facilitator provides the necessary data-related conditions for the successful analysis of social issues like crime by vertically and horizontally integrating datasets through spatial data infrastructures and to publish this data as open data. The vertical and horizontal integration of datasets from different fields and subjects allows researchers to combine datasets and to possibly find causal relationships that could not be found otherwise. Smarter governance and the development of spatial data infrastructures can lead to higher quality data and higher quality data can benefit all types of spatiotemporal analysis, including those of burglaries and other crime events. And because these integrated and high quality datasets are increasingly published as open data, anyone has access to it. In this way value can be added to the data and it could potentially spur economic development as opportunities arise for the development of new analysis methods and reporting tools.

The smart city as motivator motivates city and local governments to look at existing urban issues differently by demonstrating the power of modern information and communication technologies in for example massive and complex calculations, data mining, and analysis; which helps in the automatic discovery of patterns, rules and knowledge, and provides remote monitoring, control and feedback to the real world for intelligent city management and informed decision-making.

And finally, the smart city as promoter generates awareness among city and local governments, as well as among businesses and citizens. The alleged buzzword 'smart city' is not necessarily a bad thing as it has the potential to reach a larger public by informing people about the potential benefits and draw in funds and subsidies. As such, it can help to promote smart city concepts from the drawing board to reality. Obviously, it is important to acknowledge that the term smart city can evoke negative reactions, for example concerning the privacy of citizens, so it is important to address how the privacy of citizens can be respected while focusing on the intention of improving the quality of life.

The case study of the spatiotemporal analysis of burglaries in Haarlem demonstrates the added value of smart city concepts by increasing accountability, transparency and effectiveness of city and local governments. This also shows that being a smart city should never be the ultimate goal, but that being a smart city can ultimately result into more efficient and effective management of a city or local government.

Contents

Chapter 1

Introduction

The inspiration for this thesis subject stems from two observations on two different subjects that might seem unrelated at first: smart cities and spatiotemporal patterns of residential burglaries.

1. **Smart cities** are a hot topic. It seems like governments and private companies all want to contribute to smarter cities, albeit with different motivations. It is hard to describe what a smart city is or does, but two of its many goals are to make cities safer and its governance more efficient. Information technology seems to be a main facilitator of smart cities, with associated terms like 'sensors', 'big data' and 'open data'. Are smart cities just a hype fueled by commercial interests of city marketeers and private companies or can the concepts behind it actually improve cities for its residents in a meaningful way?

2. **Spatiotemporal patterns of residential burglaries**. Residential burglaries are a type of crime that is relatively common, but at the same time can have a great negative impact on victims. Burglaries can make residents feel unsafe in their own neighborhood, even when they are not victims themselves. Meanwhile, the risk of becoming a victim of burglary does not seem to be equal across space. Reports of burglaries are often higher in certain areas and lower in other. To be able to bring down the number of burglaries and create a safer environment for residents to live in, policy makers want to know where these burglary 'hot spots' are and why they are there.

The first section of this introduction (section 1.1) elaborates further on these two subjects. The aim is to not go into too much detail yet, but to focus on the societal and scientific relevance and the interrelationship between the two subjects.

Based on this, the second section (section 1.2) defines the scope of this research by presenting the main research question and a set of sub-questions. In addition to the research questions, the study area is presented in a subsection.

Finally, section 1.3 explains the structure and the contents of this report.

1.1 Smart cities and burglaries

This first section introduces the two main subjects of this study: smart cities and spatiotemporal patterns of residential burglaries. First, these two subjects are discussed individually, focusing on why it is deemed relevant to study them. Second, the focus is on the relationship between the two subjects. This serves to make clear why it makes sense to view and study two seemingly unrelated subjects like smart cities and burglaries in relationship to each other.

1.1.1 Smart cities

During the past few years, the term 'smart cities' has become more and more popular within urban development theories. In the essence, the smart city concept aims at sustainable urban growth, including coping with economic, environmental and social issues like crime (Gruen, 2013). It borrows concepts from more traditional urban growth theories and integrates these using modern information and communication technologies (ICT). Therefore, ICT is the binding factor in the smart city theory, tying together data streams and systems between different sources and research fields.

Smart cities are hard to define as so many definitions and interpretations exist. So at this point, rather than trying to give a concise definition that covers all of the related subjects, Moss Kanter and Litow (2009) present a very tangible, albeit somewhat idealized, image of the smart city.

> A smarter city infuses information into its physical infrastructure to improve conveniences, facilitate mobility, add efficiencies, conserve energy, improve the quality of air and water, identify problems and fix them quickly, recover rapidly from disasters, collect data to make better decisions and deploy resources effectively, and share data to enable collaboration across entities and domains. Its operations are instrumented and guided by performance metrics, with interconnections across sectors and silos (p. 2).

Based on quotes like these, the relevance of the smart city seems apparent to local governments. Who would not want a more efficient city by utilizing a physical information structure that is often already there? It is therefore no surprise that several cities and municipalities in the Netherlands are trying to catch the 'smart city train'. To get a better idea of smart cities and related initiatives, lets take a look at a few examples from local governments.

The municipality of Amsterdam is part of 'Amsterdam Smart City' (ASC) which is a partnership between companies, governments, knowledge institutions and the people of Amsterdam. The ASC describes a smart city as a city where social and technological infrastructures and solutions facilitate and accelerate sustainable economic growth (see figure 1.1). Their ultimate goal is to have a habitable city where it is pleasant to both live and work (ASC, 2015).

Two related examples of current projects carried out by ASC are ‘Smart Light’ and ‘Flexible street lighting’ which focus both on smarter public lighting (see figure 1.2). The aim of these projects is to enable the control of street lights individually. Lights can be dimmed or the color can be adjusted depending on the weather or to control pedestrians. For example, street lights can shine brighter when someone is near it. There are also ideas to equip light posts with Wi-Fi antennas and sensors to measure air quality. The ultimate goals are to save energy and to improve the perceived safety of residents.

But not only the municipality of Amsterdam partakes in smart city initiatives. Other examples include the municipality of Eindhoven where they installed sound sensors to monitor the level of noise in the *Stratumseind*: a street in the city center of Eindhoven with much nightlife activity. Residents get access to this data to monitor the noise levels themselves to get more objective measures. This initiative resulted in fewer noise complaints.

In Heerlen they want to create an online map with empty buildings to show residents and entrepeneurs where there is opportunity for new urban developments (Ministerie van Infrastructuur en Milieu, 2014). And in Groningen they are working on a smart energy grid which can deliver power and heat efficiently. Domestic activities which require much energy are scheduled to be carried out when energy is readily available and the least expensive, for example during periods with high winds when wind mills produce the most electricity [1].

And this is just a small sample of current smart city initiatives in Dutch municipalities demonstrating the popularity of the concept of smart cities.

The potential benefits of smart cities are clear for governments and residents. Smart city concepts can potentially increase the effectiveness and efficiency of governments and create more habitable environments for residents. But there is also considerable criticism towards smart cities and the associated concepts. Critique is generally focused on smart cities being a hype and on potential privacy violations.

Hollands (2008) for example criticizes the use of smart cities as a label for place marketing purposes. Cities want to be smart from the perspective of city marketing, rather than having the sole purpose of improving the quality of life. Therefore, some believe smart cities to be a buzzword and just another empty hype.

Martinez-Balleste et al. (2013) emphasize that residents need to be aware that smart cities have the ability to silently gather a variety of information about them. These concerns regarding privacy are also key in the news article *Hoe de slimme stad een dom idee kan worden* (‘How the smart city can turn into a dumb idea’; van Noort (2015)). This article describes how the smart city

[1] URL: `http://www.powermatchingcity.nl/site/pagina.php?`

can be an attractive target for hackers, illustrated by the example of a burglar using data from smart energy meters to deduct if residents are at home or not.

To conclude, the popularity of smart cities among local governments and researchers can be explained by the potential benefits they offer. Information and communication technology can be utilized to improve the effectiveness and efficiency of governments and the quality of life of residents. Meanwhile, there are plenty of concerns regarding privacy and the actual added value of the smart city concept. It is therefore fair to ask the question: is the concept of the smart city just an empty hype or are there any real benefits to it?

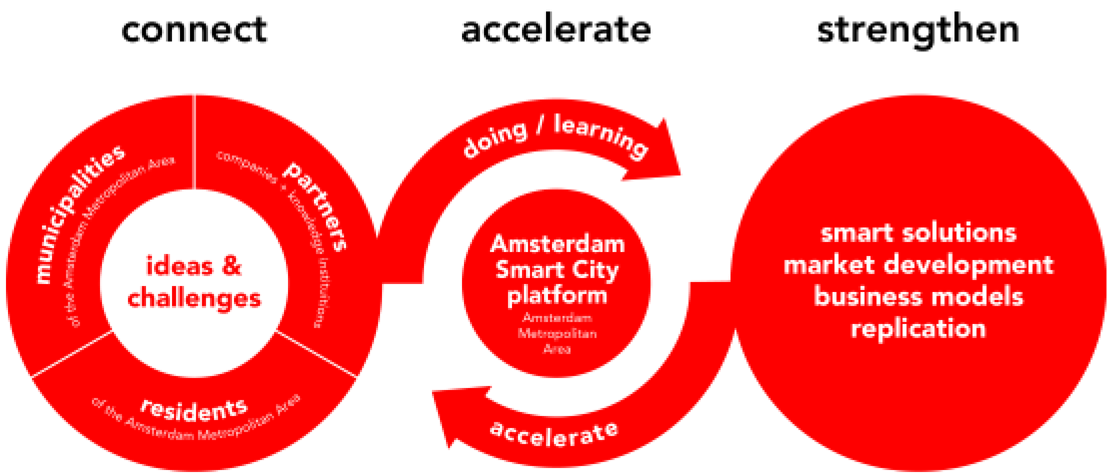


Figure 1.1: The goals of the Amsterdam Smart City partnership.

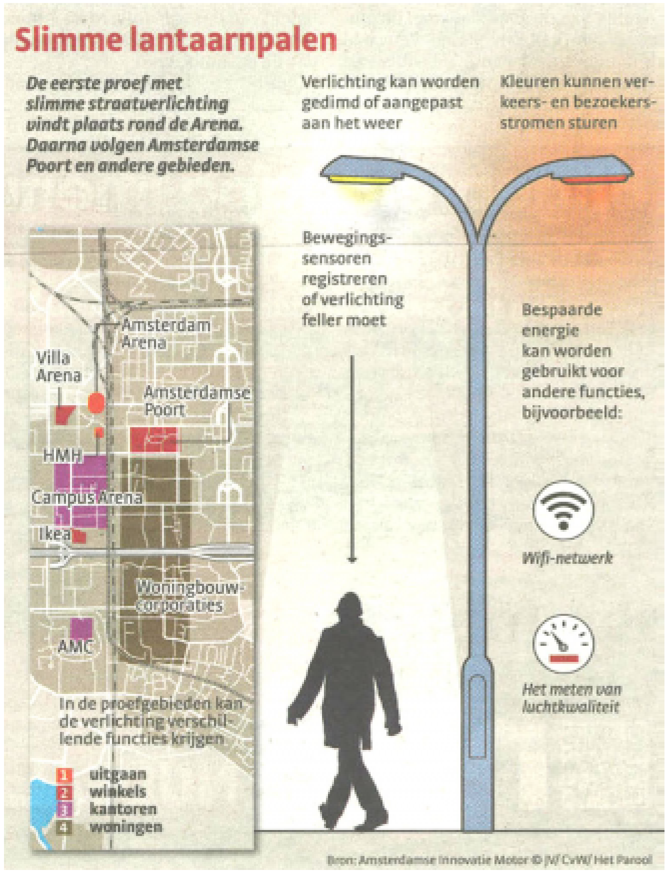


Figure 1.2: Infographic about smart street lights (in Dutch).

1.1.2 Spatiotemporal analysis of residential burglaries

The second subject in this thesis is the spatiotemporal patterns of residential burglaries. Before further introducing this subject, it is important to define what is exactly meant with the term 'residential burglary'. The following definition is used here:

> The stealing of someone else's goods or money from a dwelling by someone who accessed the dwelling without the knowledge of or permission from the rightful owner (Klein Haneveld et al., 2012).

During this research, the term 'residential burglaries' is often simply referred to as 'burglaries'.

Residential burglary is a social phenomenon that receives much attention in society. This can be exemplified by taking a look at some news articles published in the past years in the Netherlands.

The article *Burgers in actie tegen inbraken* ('Citizens in action against burglaries'; NOS (2013)) describes an initiative where citizens are asked to submit their plans against burglaries. The ultimate goal of the initiative is to have a single day without any burglaries occurring.

Vinexwijk trekt inbrekers ('Vinexwijk attracts burglars'; NOS (2014)) states that the chances of burglaries in 'Vinex neighborhoods'[2] are 15 percent higher than in other neighborhoods. This figure is based on reports by insurance companies. The increased risk is allegedly due to less social control and the presence of more dual income households. This news article further mentions the differences in burglary rates between Dutch provinces.

Finally, the news article *Aantal inbraken daalt, maar niet overal* ('Number of burglaries is declining, but not everywhere'; RTLNieuws (2015)) points out that although the number of burglaries on average declined in the Netherlands, this trend is not consistent across space as some municipalities saw a sharp rise in burglaries instead.

Based on these news articles it is clear that there is societal relevance to the analysis of spatiotemporal patterns of burglaries. It could help to find explanations for variations in burglary rates and what can possibly be done to influence these patterns.

The societal relevance of spatial patterns of burglaries can also be demonstrated by looking at the impact on the perceived level of safety, an impact that has been generally acknowledged for a long time. Already in 1980, Maguire (1980) summarized multiple scientific articles to show that 73% of burglary victims expressed 'considerable fear' of a repeat, that over 40% of female victims were afraid to be alone in their houses for some weeks afterwards, and that 70% of victims of a selection of mainly property crimes were 'very distressed'. Not much has changed in more recent years.

The *Veiligheidsmonitor 2014* ('Safety Monitor 2014') report by the CBS[3] states that residents view the chance of becoming a victim of burglary the highest in comparison to other types of crime (p. 49). There seems to be a trend of increasing fear of burglaries over the past ten years (see figure 1.3). Moreover, the report states that 3 percent of all Dutch residents were confronted with burglary or attempted burglary in 2014 alone.

The *Veiligheidsmonitor* also points out the strong relationship between the occurrence of burglaries and how residents score their own neighborhood. Residents feel less safe in areas with many burglaries. Moreover, 46% of the residents who already became victim of a burglary considered the risk of 'revictimization' within twelve months as very high. From the residents who were not a victim of burglary before, 10% considered this risk to be very high (Akkermans et al. (2015)). It can be concluded that burglaries affect the perceived level of safety of residents considerably, especially when someone recently became a victim of burglary.

[2] Vinex neighborhood is a term used in the Netherlands that refers to newly developed neighborhoods at the edge of larger cities to facilitate and regulate urban growth.

[3] *Centraal Bureau voor de Statistiek* or Central Bureau for Statistics.

In addition to the findings from the *Veiligheidsmonitor 2014*, the Dutch Police also states that burglaries typically have a large impact on the level of perceived safety, while being a fairly common type of crime (Klein Haneveld et al., 2012). Alongside their usual efforts to fight this type of crime, the Dutch police launched a website in October 2013 that shows the number of residential burglaries and attempted burglaries in the last three months for all Dutch neighborhoods[4] (see figure 1.4). The police hopes that this website helps to raise awareness among residents about the risk of burglaries (Politie, 2013).

Together, the numerous news articles and reports, evidence found in scientific literature and the statements from the Dutch police, help to demonstrate the societal relevance of studying residential burglaries. More knowledge about this subject could help citizens, local governments and law enforcement to improve their measures and policies against burglaries. This could lead to fewer burglaries and improve both the actual level of safety and the perceived level of safety.

But analyzing the spatial patterns of crime is not only of societal relevance, but of scientific relevance too. As already highlighted by the news articles about higher rates of burglaries in particular neighborhoods and differing rates between different provinces: burglaries do not tend to spread evenly across space. Place matters when trying to understand the perceived patterns of burglaries and their causes. This notion corresponds with ideas from the theoretic field of 'spatial criminology'. Spatial criminology research describes place as one of the essential cornerstones in understanding crime. Thus, spatial criminology focuses on the role of geography in studying the observed and expected spatial patterns of crime (Ratcliffe, 2010).

Although studying crime is traditionally seen as the domain of disciplines like sociology and psychology, geography and criminology have proved to be a viable combination. Crime clearly has a spatial component in that both crimes and offenders typically have a location associated with them (Chainey and Ratcliffe, 2005). It therefore makes sense to pay special attention to these spatial components of crime.

During the 1970s and 1980s the research into this spatial dimension of crime got a stimulus from the developments in information technology in general and Geographical Information Systems (GIS) in particular. The most important factor were the lower costs of computer hardware. Chainey and Ratcliffe (2005) explain: *"These reductions in the cost of computer hardware were complemented by improved operating systems, electronic storage media and developments in computer software, and have had a wide and significant impact in introducing GIS technologies to new areas, such as policing and crime reduction"* (p. 2).

So these developments allowed researchers in spatial criminology to find concentrations and more complex patterns in crime occurrence, to study the relationships between crime and environmental and socioeconomic factors, and to assess the effectiveness of police actions and anti-crime measures on a spatial

[4] URL: `http://www.politie.nl/misdaad-in-kaart`.

scale (Chainey and Ratcliffe, 2005). These types of analysis were unavailable or less accessible before this digital revolution.

In addition, the developments in information technology allow researchers to not only assess the spatial component of crime, but also the temporal aspect of crime. Crime does not only seem to vary across space, but also across time. For example, a study by Sorensen (2004) shows that burglaries occur more often in the autumn and winter than in the spring and summer. It also turns out that there is an increased number of burglaries during holidays like Christmas. Patterns in burglaries cannot only vary over large periods of time, i.e. years or seasons, but also over smaller periods, i.e. day and night.

So to sum it up: finding explanations for the spatiotemporal patterns of burglaries is of both societal and scientific relevance. There is great value in defining a suitable and generic analysis method for finding these explanations. Better understanding of spatiotemporal patterns of burglaries can improve measures and policies and ultimately lead to a safer environment for people to live in.

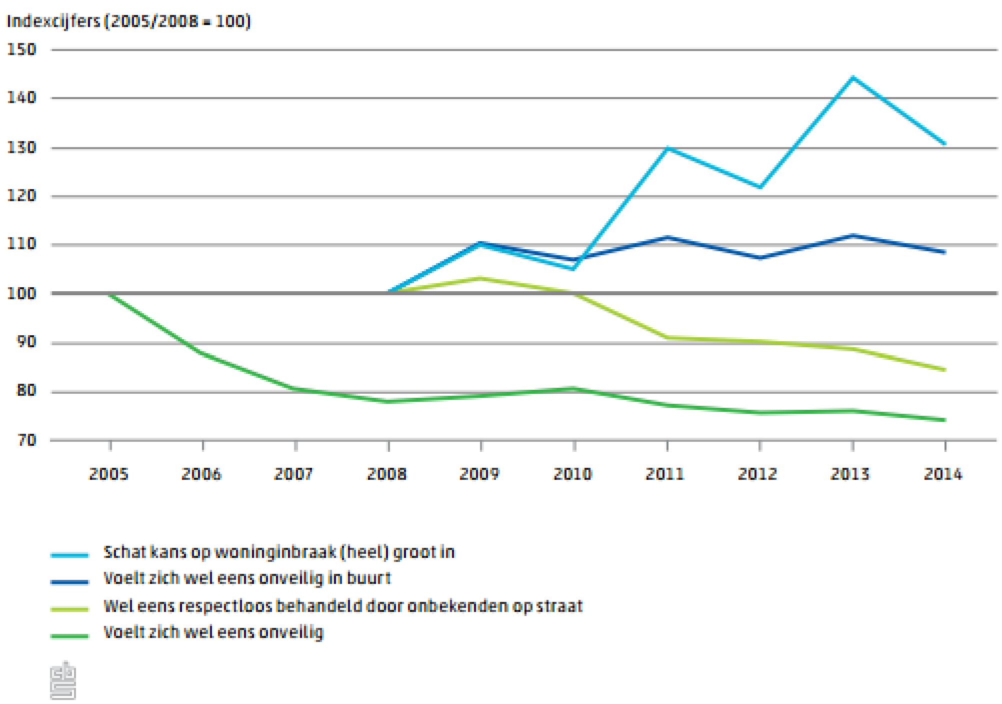


Figure 1.3: This graph shows the trends of the perception of safety among Dutch residents. The perceived risk of burglary is on the rise (light blue line). Source: Akkermans et al. (2015).

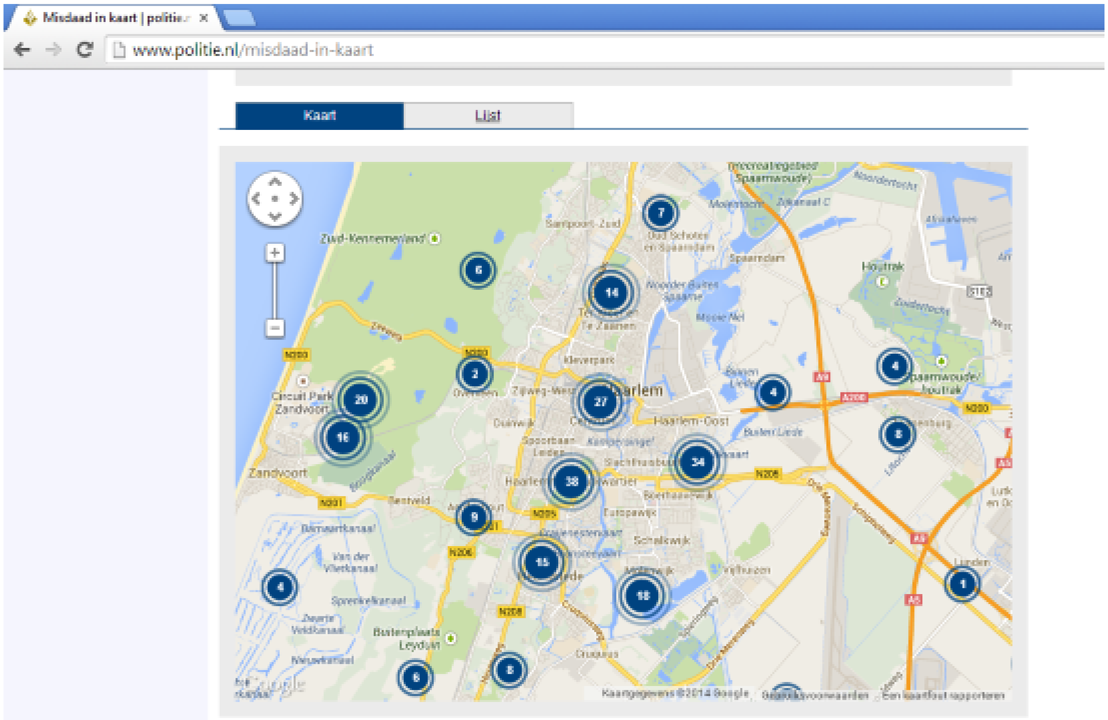


Figure 1.4: An example of the website 'Misdaad in kaart' showing burglaries in and around Haarlem.

1.1.3 How do smart city concepts and burglaries relate?

After the first introduction to the subjects of smart cities and spatiotemporal patterns of burglaries, this section argues the added value of considering these subjects in relationship to each other. Three arguments are presented.

Sharing the same goals

The ultimate goal of smart cities is to increase the quality of life of residents. Public safety can be seen as an important aspect of quality of life. As became apparent in the previous section, burglaries can have a great impact on the (perceived) level of safety. A better understanding of the spatiotemporal patterns of crime can help to predict or prevent burglaries. There is an overlap in the goals of smart cities and the spatiotemporal analysis of burglaries: both aim to improve the quality of life of residents. Concepts and ideas from the smart city might help in the spatiotemporal analysis of burglaries. Therefore, the smart city is used as a theoretical framework for analyzing the spatiotemporal patterns of residential burglaries.

The role of information and communication technology

Modern information and communication technology plays a large role in both smart cities and the spatiotemporal analysis of burglaries. Within smart city concepts, the exchange of (big) data from different sources and sensors is one of the key factors in making a city smarter. Meanwhile, the spatiotemporal analysis of burglaries has greatly benefited from the advancements in GIS, allowing access to more advanced spatial analysis methods. Here again, concepts from the smart city (like sensor data, big data, and open data) can be complementary to the spatiotemporal analysis of burglaries. Sensor, big and open data could serve as input data for this type of analysis.

Spatiotemporal analysis of burglaries as test case for applying smart city concepts

Finally, the spatiotemporal analysis of burglaries can be seen as a test case for applying smart city concepts. How do the mostly theoretical concepts of the smart city hold up in practice? How can concerns about smart cities, for example with regard to privacy, be mitigated or resolved? How can relatively generic smart city concepts aid in fighting specific social problems like crime, and burglaries in particular? In short, the spatiotemporal analysis of burglaries based on concepts from the smart city theory, can demonstrate the usefulness or shortcomings of the smart city concepts in practice. Results of this 'test case' can ultimately help to improve the feasibility of smart city concepts from a practical point of view.

1.2 Research questions

Based on the previous sections the main research question and the subquestions can be defined. The main goal of this research is twofold:

1. to assess the practical usefulness of smart city concepts,
2. by finding a suitable method for explaining the structural spatiotemporal patterns of burglaries and applying this method in practice.

The main research question is:

> *How can the application of smart city concepts help improve public safety by decreasing burglary rates?*

To answer this main question several subquestions are formulated. These questions are listed below together with a short description.

1. *What is a smart city and what are the underlying theoretical concepts?* Here, the focus is on the theory of smart cities. How did they originate and how can the smart city be defined? As smart cities are multifaceted, the aim is also to identify the most relevant elements and concepts within

the broad smart city spectrum. Therefore, the goal is to answer this subquestion by starting out with a broad view of smart cities and narrowing it down to the most essential concepts with the practical goal of decreasing burglary rates.

2. *What methods are available for the spatiotemporal analysis of crime data?*
   Here different methods for spatiotemporal analysis are explained and compared. What analysis methods are available and what method is deemed most suitable in the context of this research? The answer to this subquestion results in a method that can be used to identify and find the explanations of the spatiotemporal patterns of burglaries.

3. *What are the possible risk factors for residential burglaries?*
   The aim of this subquestion is to discuss the relevant spatial criminology theories concerning spatiotemporal patterns of burglaries and to identify possible risk factors based on these theories to serve as input for the spatiotemporal analysis of burglaries.

4. *What are the results of the spatiotemporal analysis of residential burglaries?*
   This subquestion is all about the results of the spatiotemporal analysis. What variables can be used to explain the patterns of burglaries, what is the influence of time and place and how do these results relate to the underlying spatial criminology theories?

Together, the answers to these subquestions form the answer to the main research question which is presented in the conclusion of this report.

1.2.1 Study area

The study area for the spatiotemporal analysis of burglaries is the Dutch municipality of Haarlem (see figure 1.5), an urban municipality in the *Provincie Noord-Holland* (Province of North-Holland) with just over 150 000 inhabitants[5]. There are two main reasons for this decision, one from a societal viewpoint and one from a more scientific viewpoint.

Firstly, local research by the municipality of Haarlem showed that residents want more attention from local government and police directed towards the problem of burglaries (van der Werff, 2013). Considering different types of crime, in four of the five city districts residents see fighting residential burglaries as the highest priority. Only in the *Centrum* district, residents are more concerned about vandalism. So, generally speaking, most Haarlem residents consider burglaries as the most important crime-related problem and therefore want a proportional amount of resources to go towards fighting and preventing them.

[5] https://www.haarlem.nl/feiten-en-cijfers/

Fortunately, the municipality of Haarlem has recognized this and is already investing more in the area of public safety. A good example of this is the *Overlastmonitor* (Nuisance Monitor; see figure 1.6). This information technology solution provides a digital representation of all reported accounts of nuisance and crime incidents in the municipality of Haarlem (E-overheid, 2013). Therefore, it has to combine different sources of data: crime incident data from the police and nuisance incident data from the municipality. This combination of data can provide new insights into the spatial and temporal patterns of nuisance and crime. The integrated data can be viewed by a selection of local government employees to help them monitor crime and nuisance and help them develop policies.

The concerns from residents regarding burglaries coupled with the possibilities offered by the availability of crime data through the Overlastmonitor, provides an essential and solid basis for further investigation into the connection between crime and place. It demonstrates the eligibility of the municipality of Haarlem as the study area for this research.

The choice for Haarlem can also be explained from a scientific viewpoint. As was already argued, residential burglaries are typically not distributed evenly across space. This can also be demonstrated empirically by taking a closer look at the situation in the municipality of Haarlem. By conducting an initial hot spot analysis of the residential burglary incidents in Haarlem in 2011 (see figure 1.7) it becomes apparent that residential burglaries in Haarlem are not evenly distributed across space. They rather seem to cluster in the southeastern districts of Haarlem. This finding supports the assumptions that were made earlier in this introduction regarding the uneven spatial distribution of burglaries and crime in general.

When the residential burglaries of the same year are plotted against time, it becomes clear that there is also a distinct temporal pattern (see figure 1.8). It seems that residential burglaries in 2011 most often occurred during the winter. As finding a method to explain these spatial and temporal patterns of burglaries is a main goal of this research, Haarlem is suitable as a study area as these patterns can clearly be observed here.

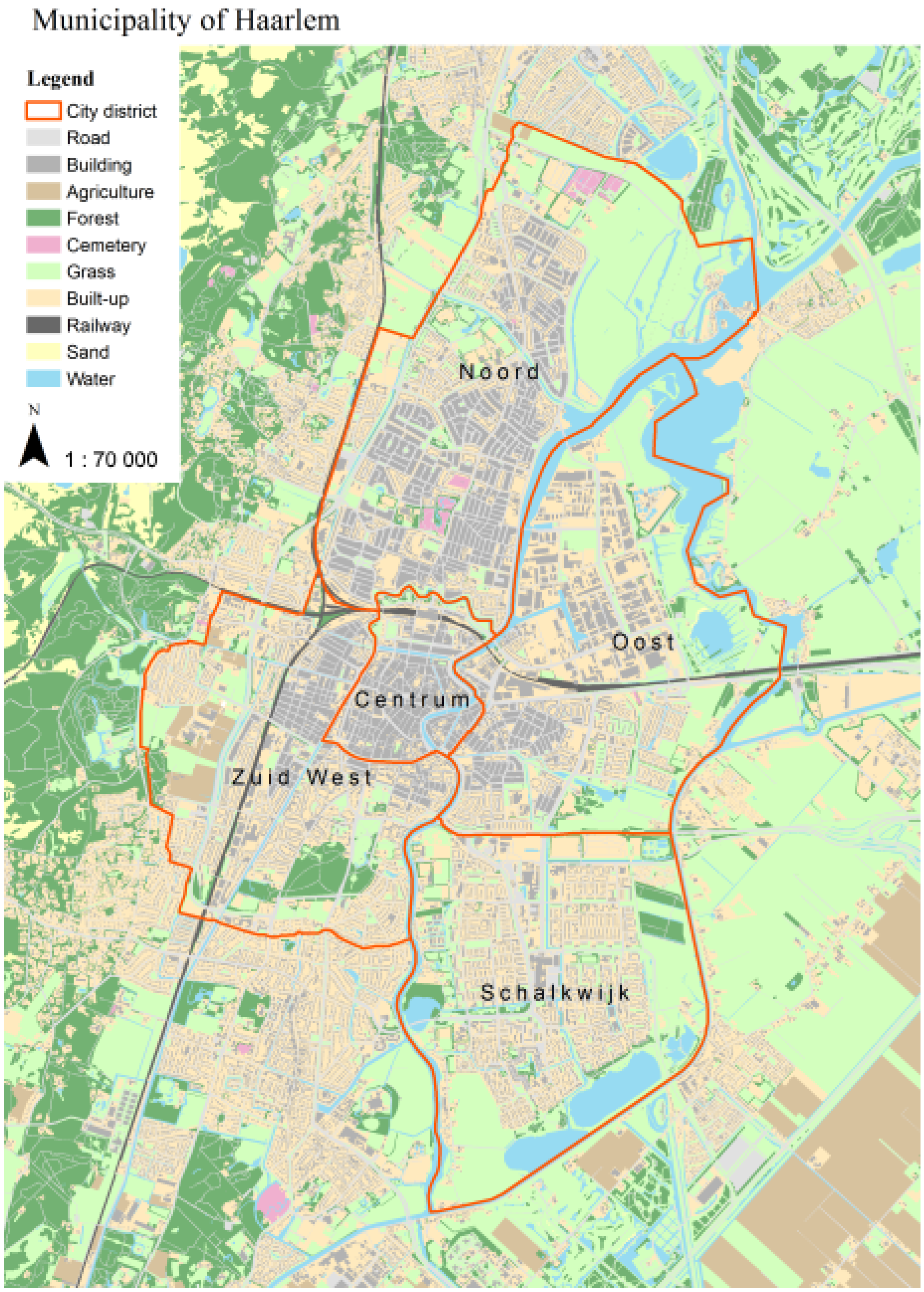


Figure 1.5: The municipality of Haarlem in 2014 marked by a red outline. The geographic data used to construct this map is derived from the *Basisregistratie Topografie* (Key Register Topography). The district geometries are provided by the municipality Haarlem directly.

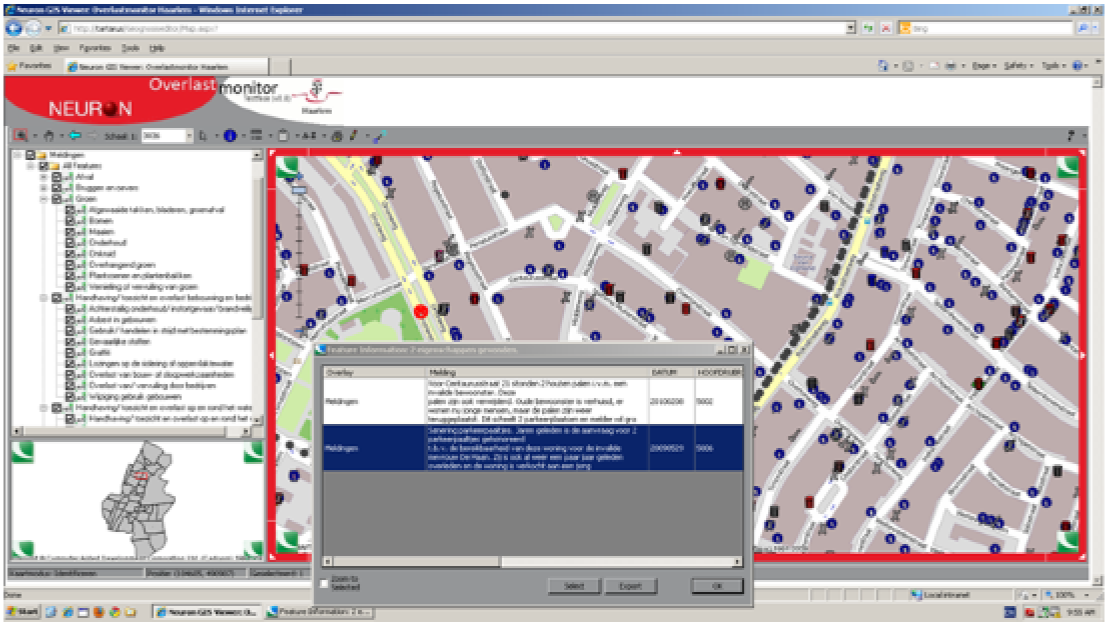


Figure 1.6: A screenshot of the *Overlastmonitor*.

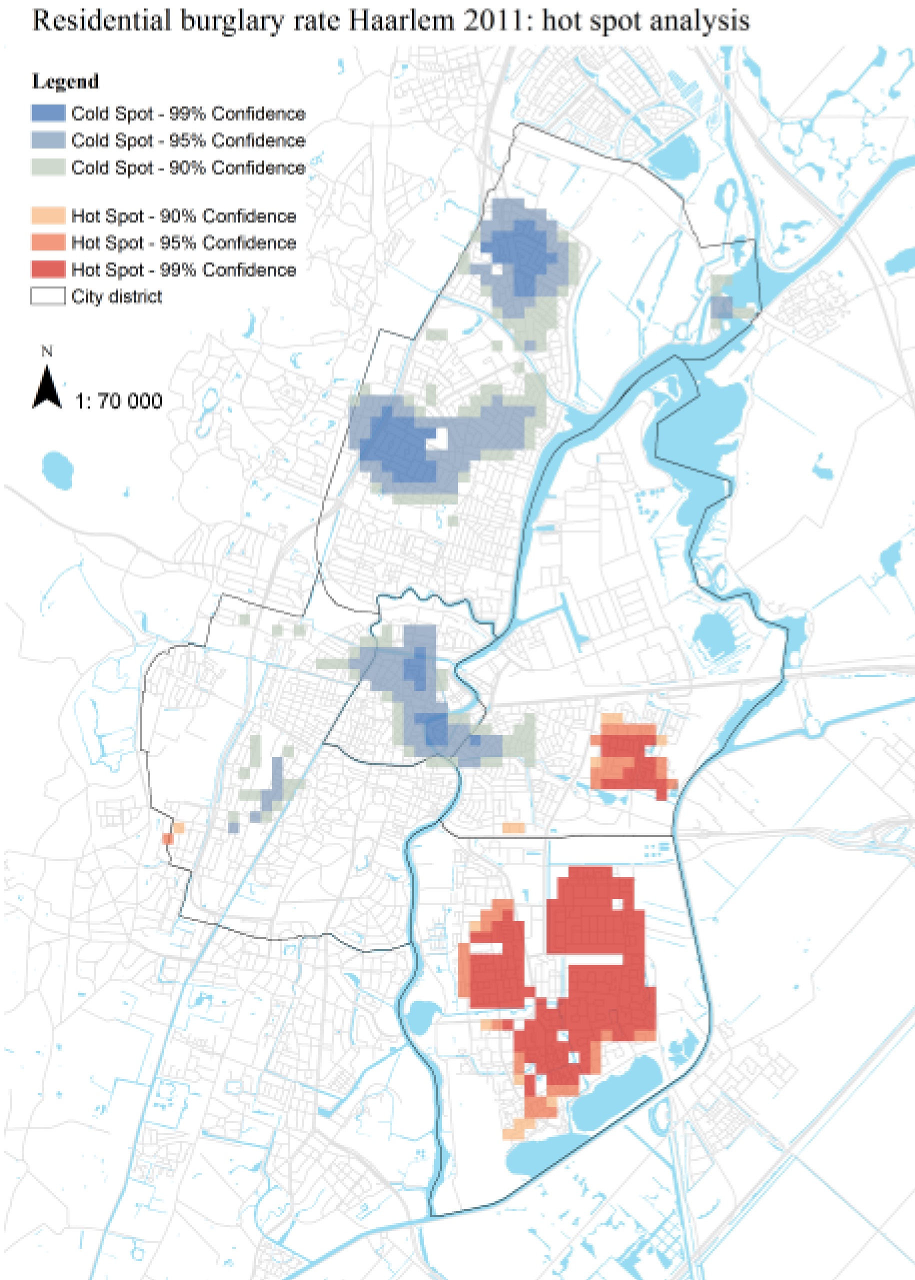


Figure 1.7: This map of the municipality of Haarlem shows the results of the hot spot analysis of residential burglary rates (number of burglaries divided by the number of residential addresses) in 2011. The hot spots are calculated based on burglary rates that are aggregated to 100 by 100 meter cells.

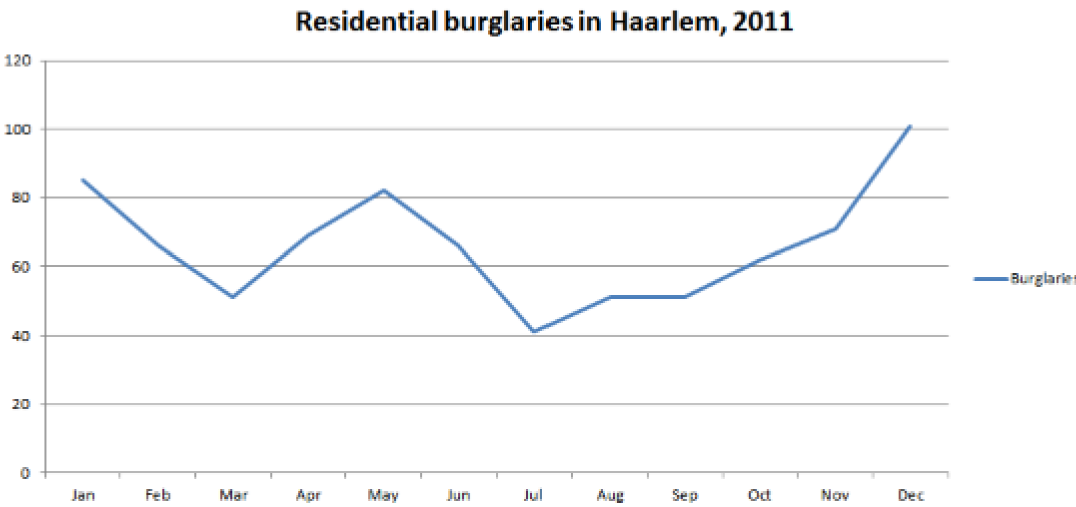


Figure 1.8: This graph shows the recorded burglary incidents in Haarlem for the year 2011 per month.

1.3 Structure of the report

The structure of this report follows the order of the research questions. Chapter 2, 'Smart cities', starts with laying down the broad theoretical base for this research by answering the question: what is a smart city and what are the underlying theoretical concepts? It elaborates on the smart city concepts and shows how the study of burglaries fits well within this theoretical framework.

Chapter 3, 'Methods for the spatiotemporal analysis of crime data', focuses on answering the question: what methods are available for the spatiotemporal analysis of crime data? It outlines the available methods for the spatiotemporal analysis of crime data and derives a suitable method that can be applied to analyze the spatiotemporal patterns of burglaries.

Chapter 4, 'Risk factors for residential burglaries', is a logical continuation of the previous chapter as the method defined in chapter 3 needs input variables to work. The subquestion answered in this chapter is: what are the possible risk factors for residential burglaries? It takes a closer look at the spatial criminology literature to find the most common theories that can be used to explain the occurrence of burglaries. From these theories the relevant burglary risk factors are derived.

Chapter 5, 'Burglary analysis results', shows the process of carrying out the spatiotemporal analysis and presents the results. These results are then compared to the theories used to find the risk factors of burglaries. Do the findings from Haarlem confirm or disprove these theories and the associated assumptions?

Chapter 6, 'Conclusion and discussion', is the final chapter and presents the conclusions to the research questions and a discussion of the results.

Besides these regular chapters, there are three appendices. These appendices serve to elaborate on certain subjects and are referred to within the regular chapters. The appendices are not essential for the average reader, but contain additional information, generally on a higher level of detail than is necessary in the regular chapters.

The first appendix, 'Input data', provides detailed information about the input data that are used in this research. It shows the steps that were followed to operationalize the risk factors identified in chapter 4. Moreover, the underlying assumptions of how independent variables relate to the dependent variable are discussed including references to the theory and the source of the data.

The second appendix, 'Data sources', provides more in-depth information about the data sets and data registrations that are used as the source for the data in this research.

The third and final appendix, 'Model results', provides a complete overview of the models. These include extra information about the model results presented in chapter 5.

1.4 Conclusion

To conclude this introductory chapter: what makes this research unique and what is the added value?

- First of all and perhaps most importantly, this study gives a 'real world' example of an **application of smart city concepts**. Much is written about smart cities, mostly from a conceptual point of view, where the smart city is often presented as an abstract and almost utopian vision on the future of cities and its management. This study wants to explore how smart city principles can be applied in the here and now, using methods, data and infrastructure that is already available.

- This study tests hypotheses about spatiotemporal patterns and risk factors of burglaries from different established spatial criminology viewpoints, e.g. from rational, ecological, demographic and environmental design viewpoints. Thus, this study proposes a **holistic approach** to finding explanations for burglary patterns, rather than focusing on one spatial criminology viewpoint. That also means that many different data from many different data sources has to be accessible, such as data about properties, people, roads or neighborhood statistics. The aim is to extract the necessary data from **generic and standard datasets**. The main advantage of using common and standardized data repositories is that, as the source data and information model are always the same, a generic method can be developed for accessing, processing, analyzing and presenting the data. Other advantages are that the available data are being reused and value is added to it, and that the data are as up-to-date as possible.

- A final unique aspect of this study is the proposed level of detail of the spatial analysis. In many similar studies where a spatial analysis of crime rates is conducted, the neighborhood is generally the smallest spatial unit used. Here, the aim is to use a much **larger spatial scale**, allowing for more differentiation in the results of the analysis: the spatial patterns can be analyzed in more detail. Moreover, potential **temporal effects** are taken into account as well.

Chapter 2

Smart cities

The spatiotemporal analysis of residential burglaries can be nested within the theory about 'smart cities'. This chapter aims at explaining this connection and aims at answering the research subquestion: *what is a smart city and what are the underlying theoretical concepts?*

It starts with the more general theoretical concepts of smart cities and tries to come to a workable definition which can be used in this study. Secondly, the broad view of smart cities is narrowed down to the subfields of smart governance and smart living. Within these subfields the most relevant concepts of smart cities for this study are distinguished. The added value of smart governance and smart living are demonstrated by looking at related concepts: e-government, spatial data infrastructures, problem-oriented policing and predictive policing. To conclude, the most relevant aspects of smart cities are summed up to construct a theoretical framework for the spatiotemporal analysis of burglaries.

2.1 What is a smart city?

2.1.1 Urbanization

Nowadays, more than half of the world population is living in cities (Bettencourt et al., 2007). This trend of urbanization is expected to continue and by the year 2050, the share of people living in cities is expected to reach 70%. Meanwhile, cities cover just 2% of the earth's surface, but consume about 75% of the world's resources (Gruen, 2013).

This trend of urbanization is also clearly observable in the Netherlands, which is expected to develop at higher rates than both Europe's and Western Europe's averages (see figure 2.1[1]). The exponential growth of cities and its population can have negative impacts on the environment. In addition to these environmental impacts, urbanization poses new challenges to policymakers, concerning for example social equity and economic development.

[1] Image URL: `http://tinyurl.com/lrue4v7`.

To successfully manage urban growth, Gruen (2013) states that policymakers have to find harmony between spatial, social, economical and environmental aspects of a city and its inhabitants. Harmony between these aspects is based on three pillars: earth environment, economic development and social equity. These pillars can be balanced through sustainability. One way of achieving sustainable urban growth is through the development of 'smart cities'.

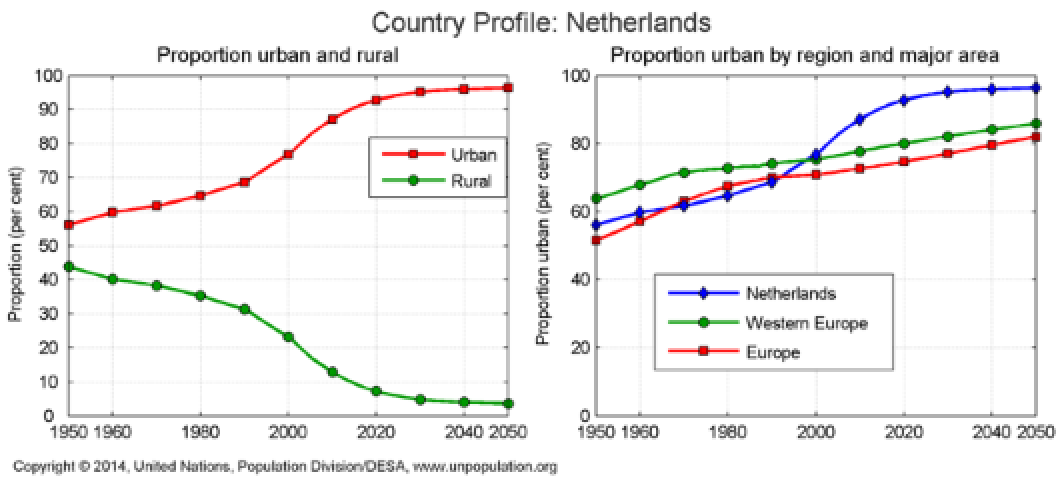


Figure 2.1: Urbanization 'Country Profile' of the Netherlands (data source: UN (2014))

2.1.2 Definition

There is no straightforward definition of smart cities, but there are many different understandings of the concept, partially explained by its interdisciplinary character. Different fields use their own definition. To further complicate matters, the term smart city is closely related to and sometimes used interchangeably with similar terms, like: intelligent, innovative, knowledge, wired, digital, creative and cyber cities (Hollands, 2008). Moreover, the term smart city is often used for the purpose of 'place marketing', the promotion of cities or districts to attract businesses and human capital, further obfuscating its meaning.

Caragliu et al. (2011) conducted a literature review to find a proper definition of a smart city. They distilled six common characteristics of the smart city, reflecting the different disciplinary backgrounds mentioned before. These six characteristics are summarized below.

1. The *"utilization of networked infrastructure to improve economic and political efficiency and enable social, cultural, and urban development"* (Hollands, 2008). Here, the term 'infrastructure' refers to infrastructures such as transport, business services, housing and a range of public and private

services, but in particular to 'new' infrastructures such as information and communication networks. This first point demonstrates the idea of **connectivity** as the main source of economic growth; an idea which is also captured in the concept of the 'wired city'.

2. An *"underlying emphasis on business-led urban development"* (Hollands, 2008). Business-led urban development refers to the efforts of cities to attract and facilitate businesses by creating favorable policies, while urban governance is shifting from a managerial form to entrepreneurial forms. Evidence shows that cities that are **business-oriented** score well on socioeconomic performance (Caragliu et al., 2011).

3. A strong focus on achieving the **social inclusion** of various urban residents in public services. This characteristic of smart cities revolves around the idea that all social classes should benefit equally from urban growth (Caragliu et al., 2011).

4. A stress on the crucial role of **high-tech and creative industries** in long-run urban growth. Many of the ideas behind this are based on the 'creative cities' of Richard Florida, where cities are believed to prosper if they succeed in attracting the creative class, a skilled workforce of: *"people in design, education, arts, music and entertainment, whose economic function is to create new ideas, new technology and/or creative content"* (Florida, 2004).

5. Profound attention to the role of **social and relational capital** in urban development. *"People need to be able to use technology in order to benefit from it"* (Caragliu et al., 2011). In order for a city to be smart, it is important that all citizens have the necessary level of education and access to be able to benefit from the knowledge-based economy (Coe et al., 2001).

6. **Social and environmental sustainability** is a major strategic component of cities. This last point refers to the environmental impact of smart cities. When demands on scarce resources are increasing, their exploitation must guarantee the safe and renewable use of natural resources (Caragliu et al., 2011).

Caragliu et al. (2011) incorporated all aforementioned characteristics of smart cities in one concise definition.

> "We believe a city to be smart when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance" (p. 70).

This definition of a smart city is adhered to in the remainder of this research.

As many of the underlying characteristics and concepts of smart cities are not necessarily new in the field of urban growth and development theories, what does the smart city concept add exactly? In the smart city theory, **information and communication technologies** are identified as the main drivers for sustainable urban growth and they are an essential constituent of smart cities. Therefore, information and communication technologies can be considered to be the facilitator that connects the different more traditional concepts of urban growth. Modern ICT can facilitate smart informational and cognitive processes, such as information collection and processing, real-time alerts, forecasting, learning, collective intelligence and distributed problem solving (Gruen, 2013).

Li et al. (2013) provide more examples of how information technologies can support smart city management. They describe for example how ubiquitous sensor networks can be used for real-time sensing, measuring, and data transmitting from still or moving objects. Such a ubiquitous sensor network is possible due to the recent developments in information technology. These developments facilitate the widespread use of sensors, for example in mobile phones, cameras or even street lights as was described in the introduction (Dodgson and Gann, 2011). Furthermore, cloud computing can be deployed for massive and complex calculations, data mining, and analysis; which helps in the automatic discovery of patterns, rules and knowledge, and provides remote monitoring, control and feedback to the real world for intelligent city management and public services (Li et al., 2013).

2.1.3 Operationalization

The definition presented in the previous section is still rather theoretical in nature. How can this theoretical notion of smart cities be translated to smart cities in practice?

In a report by the Centre of Regional Science at the Vienna University of Technology, Giffinger et al. (2007) attempt to rank medium-sized European cities on their 'smartness'. Therefore, they operationalized the term smart city by creating a framework of six key elements and several subelements (see figure 2.2). The six key elements show much similarity with the theoretical definition of a smart city by Caragliu et al. (2011) and their six underlying characteristics of smart cities.

Giffinger et al. (2007) provide a concise description of the key elements. These are summed up below.

- **Smart economy** includes elements related to economic competitiveness like innovation, entrepreneurship, trademarks, productivity and flexibility of the labor market as well as the integration in the (inter-)national market.

- The **smart people** element does not only include the level of education of the citizens but also the quality of social interactions regarding integration and public life and the 'open-mindedness' of the citizens.

- **Smart governance** comprises aspects of political participation, services for citizens as well as the functioning of the administration.
- Local and international accessibility are important aspects of **smart mobility** as well as the availability of information and communication technologies and modern and sustainable transport systems.
- **Smart environment** is scored by looking at attractive natural conditions (climate, green space, etc.), pollution, resource management and also by efforts towards environmental protection.
- Finally, **smart living** comprises various aspects of quality of life such as culture, health, safety, housing and tourism.

The final ranking by Giffinger et al. (2007) shows that in practice cities tend to score well on only some elements of smart cities. This already became apparent in the introduction, where some examples were given of smart city initiatives. These initiatives are often based on one element of the smart city, for example smart mobility, without necessarily taking a holistic approach. Even the city with the best overall smart city score, Luxembourg, scores relatively low on smart environment. Eindhoven, the first Dutch city on the ranking, scores particularly high on the elements smart mobility and smart economy, but lower on other. Therefore, the conclusion can be drawn that in practice cities tend to focus on some elements of the smart city. There is not yet such a thing as a fully developed smart city.

One of the main goals of this research, finding a method to explain spatiotemporal patterns of burglaries, relates particularly well to two elements of the smart city: 'smart governance' and 'smart living'. Smart governance often refers to the usage of new channels of communication with citizens, e.g. 'e-governance' or 'e-democracy' (Giffinger et al., 2007). Smart living refers to the goal of achieving a high quality of life of which public safety is an integral part. These specific fields of smart cities and the most relevant associated concepts are discussed in more detail in section 2.2.

SMART ECONOMY
(Competitiveness)

- Innovative spirit
- Entrepreneurship
- Economic image & trademarks
- Productivity
- Flexibility of labour market
- International embeddedness
- *Ability to transform*

SMART PEOPLE
(Social and Human Capital)

- Level of qualification
- Affinity to life long learning
- Social and ethnic plurality
- Flexibility
- Creativity
- Cosmopolitanism/Open-mindedness
- Participation in public life

SMART GOVERNANCE
(Participation)

- Participation in decision-making
- Public and social services
- Transparent governance
- *Political strategies & perspectives*

SMART MOBILITY
(Transport and ICT)

- Local accessibility
- (Inter-)national accessibility
- Availability of ICT-infrastructure
- Sustainable, innovative and safe transport systems

SMART ENVIRONMENT
(Natural resources)

- Attractivity of natural conditions
- Pollution
- Environmental protection
- Sustainable resource management

SMART LIVING
(Quality of life)

- Cultural facilities
- Health conditions
- Individual safety
- Housing quality
- Education facilities
- Touristic attractivity
- Social cohesion

Figure 2.2: These six key elements and subelements form the framework for the indicators for and the assessment of a city's performance as smart city (Giffinger et al., 2007).

2.2 Smart city concepts

From the broad view on smart cities in the previous section, the focus is shifted towards two specific aspects of the smart city that are most relevant for the objective of this research: smart governance and smart living. After further introducing these two smart city elements, particular concepts of both these fields are discussed, which can prove to be relevant for the spatiotemporal analysis of burglaries.

2.2.1 Smart governance

Lynn et al. (2000) define governance in general as:

> "regimes of laws, administrative rules, judicial rulings, and practices that constrain, prescribe, and enable government activity, where such activity is broadly defined as the production and delivery of publicly supported goods and services" (p. 235).

Governance becomes 'smart' when ICT is used to improve it (Chourabi et al., 2012). Nam and Pardo (2011) further note that smart governance does more than simply regulate the outputs of economic and societal systems. They state that: *"Smarter government means collaborating across departments and with communities, to become more transparent and accountable, to manage resources more effectively, and to give citizens access to information about decisions that affect their lives"* (p. 287).

Using ICT to increase the accountability, transparency and ultimately the efficiency of a governmental organization can be seen as the general goals of smart governance. Chourabi et al. (2012) discern several characteristics that smart governance should have based on a review of literature.

- **Collaboration**: smarter governments need to collaborate across departments and with communities (Nam and Pardo, 2011). The goal is to make public administration more accessible, effective and transparent (Schaffers et al., 2012).
- **Participation and partnership**: refers to the participation of citizens in the decision-making process (Giffinger et al., 2007). Therefore, the participation of citizens is closely related to the transparency of decision-making (Odendaal, 2003). The benefits of public participation include gaining trust from citizens towards local government, gaining legitimacy of decisions and ultimately better policy overall (Irvin and Stansbury, 2004).
- **Communication**: refers to the efficient communication both within the governmental organization and with citizens and stakeholders outside the organization (Odendaal, 2003).

- **Data-exchange**: the sharing of data across organizational boundaries but also publication of datasets to make them available to everyone as open data. *"Open Data is becoming a new impulse for competitiveness for the region providing the data"* (Hielkema and Hongisto, 2013). Open data can be used to create value and to ultimately benefit a region's development.

- **Service and application integration**: refers to the organizational integration of different departments within local governments. Integration or partial integration of departments can prevent resources to be wasted through duplication (Odendaal, 2003).

- **Leadership and champion**: research shows that the presence of a 'champion' can be crucial in developing a smarter governance (Lam, 2005; Mooij, 2003). A champion can be someone from higher management who ensures sufficient funds, commitment and resources to be made available or a champion can take the role of an architect who is involved in the operational aspects of a project (Lam, 2005).

E-government and SDIs

Applications of the smart governance principles are found in the concepts of 'e-government' and 'spatial data infrastructures'.

Marche and McNiven (2003) describe e-government as: *"the provision of routine government information and transactions using electronic means, most notably those using internet technologies, whether delivered at home, at work, or through public kiosks"* (p. 75).

Layne and Lee (2001) defined four developmental stages of e-government, where each phase incorporates the previous phase(s) (figure 2.3):

1. cataloging,
2. transaction,
3. vertical integration, and
4. horizontal integration.

The first stage is focused on establishing an online presence for the government. This is usually a web portal where citizens can find information and documents. Hence the naming of this stage as **cataloging**. This stage is characterized by one-way communication: from government to citizen.

In the second stage, **transaction**, the focus is on connecting the internal government system to online interfaces and allowing citizens to transact with a government electronically (Layne and Lee, 2001). In this stage one-way communication is followed up by two-way communication: citizens can for example not only download a permit request online, but can also submit the request online.

The third stage comprises **vertical integration**. What characterizes this stage is that e-government is no longer just an electronic presentation or outlet of the regular government, but that e-government also means a change in the internal structure of government processes. This can be the integration of databases of local governments and national government. A good example of this in the Netherlands is the integration of local geographical information from the BAG (*Basisadministratie Adressen en Gebouwen*; 'Key Register of Adresses and Buildings') in a centralized national data repository[2]. More detail on the Dutch key registers is provided in the next section.

The fourth and final stage of e-government is **horizontal integration**. This is a continuation of the integration process of the previous stage. With horizontal integration, there is not only integration of processes between different levels of government, but also an integration of processes between different functional departments of a government. Layne and Lee (2001) state: *"Horizontal integration refers to system integration across different functions in that a transaction in one agency can lead to automatic checks against data in other functional agencies"* (p. 133). For example, a resident files a building permit at a local government. This request is automatically checked with a database of building regulations and the tax department is notified about a possible reevaluation of the property value.

The vertical and horizontal integration within governments is closely related to the concept of 'Spatial Data Infrastructures' (SDIs). Budhathoki et al. (2008) explain: "*These infrastructures are created to facilitate the coordinated production, access, and use of geospatial data among producers and users in an electronic environment. SDIs use electronic media to connect distributed repositories of geospatial information (GI) and make these available to users through a single entry point often called 'geoportal'*" (p. 149). The ultimate goal of SDIs is to add economic and social value to geospatial data, through the better management of spatial data sets and by allowing broader access to spatial data. Rajabifard and Williamson (2001) add: "*By reducing duplication and facilitating integration and development of new and innovative business applications, SDIs can produce significant human and resource savings and returns*" (p. 2).

Kiehle et al. (2006) argue that the next step of SDIs is the standardized geoprocessing of spatial data to turn data into information. "*This process involves the acquisition of problem-specific data, the application of specific computations (e.g. normalized differentiated vegetation index (NDVI, spatial intersection, spatial buffering, etc.), and the visualisation of results (usually as maps or maplike presentations)*" (p. 273).

In the Netherlands, the national government and local governments have already implemented many elements related to e-government and SDI concepts [3]. National and local governments have their own website where they interact with their residents, but the Dutch government is moving further and further

[2]See for example `https://bagviewer.kadaster.nl/`.
[3]URL: `http://www.e-overheid.nl/english`.

towards the latest stages of e-government, for example by tying together information flows from different governmental levels (vertical integration) as well as between different departments (horizontal integration). Many of the governmental data sets are also being published as open data[4]. A major initiative that is a good example of this, while also being very relevant for this research, is the creation of a *stelsel van basisregistraties* or 'system of key registers' (Wetenschappelijke Raad voor het Regeringsbeleid, 2011) (see figure 2.4).

A key register is a central high quality dataset on a particular subject. For example, there are key registers of addresses and buildings, persons and topography. These registers are maintained by local governments and combined on a national scale: a good example of vertical integration. Governmental organizations are obliged to use these central datasets as the basis for their activities. The idea is that data are gathered once and used multiple times, meaning there is only a single 'truth'. The system of key registers should ensure that the different key registers are connected to each other, tying together information from different government departments, thus spurring horizontal integration and collaboration (E-overheid, 2014). An increasing number of these key registers is being published as open data. Therefore, the system of key registers can be seen as a crucial element in the national spatial data infrastructure of the Netherlands.

The relevance of the smart governance concept for the analysis of burglaries is a result of the smart city principles of collaboration, communication, data-exchange and service and application integration: the integration of data and information between different departments of government and sharing these data. The key registers allow data to be collected, to be exchanged and in many cases also to be used by third parties as open data. This makes these datasets available for the use in analyses, like for example the one conducted in this research. A smart governance enables access to open and structured high quality data, which can be used for different types of analysis, which can ultimately result in more informed and transparent decision making.

[4]For example through `https://www.pdok.nl/en/node`.

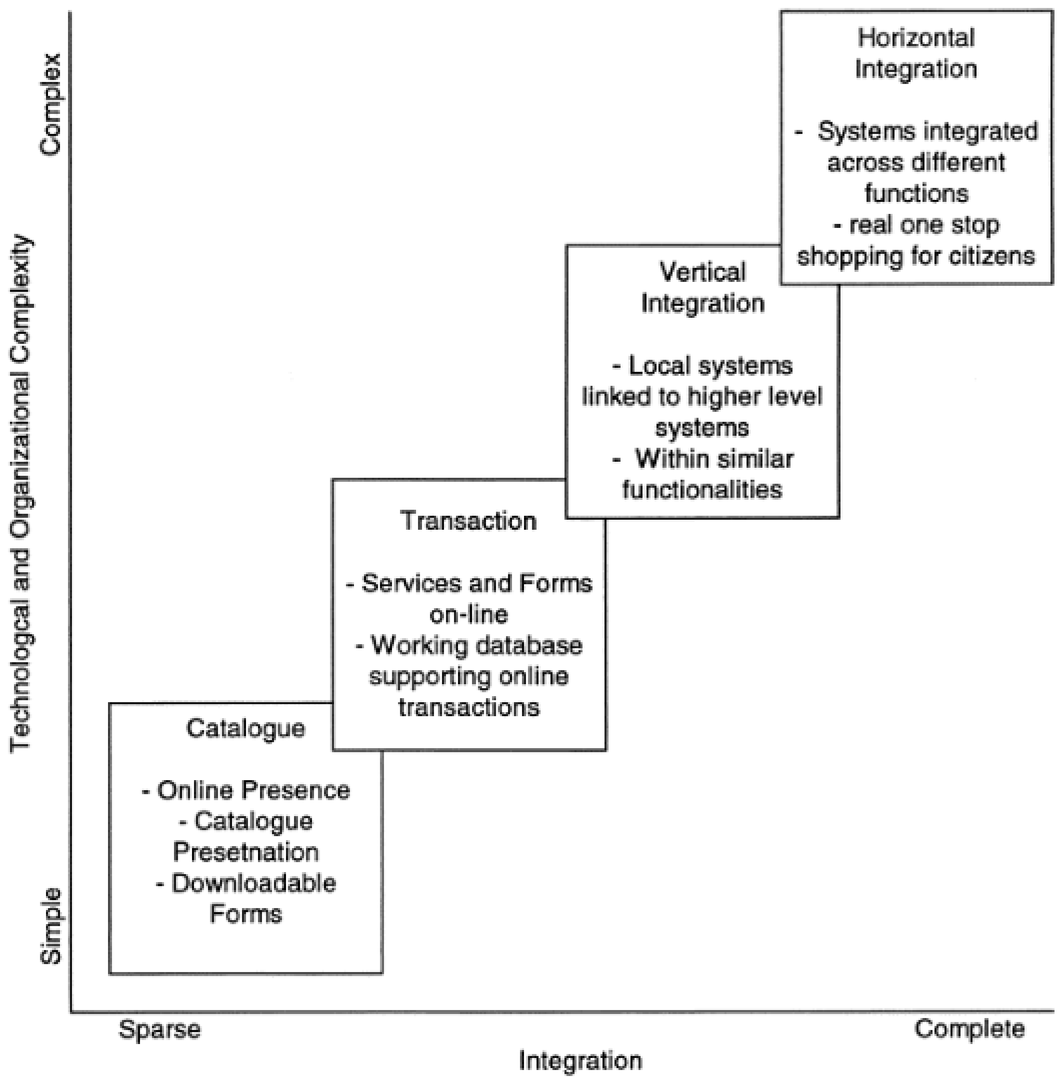


Figure 2.3: The four stages of e-government development (Layne and Lee, 2001).

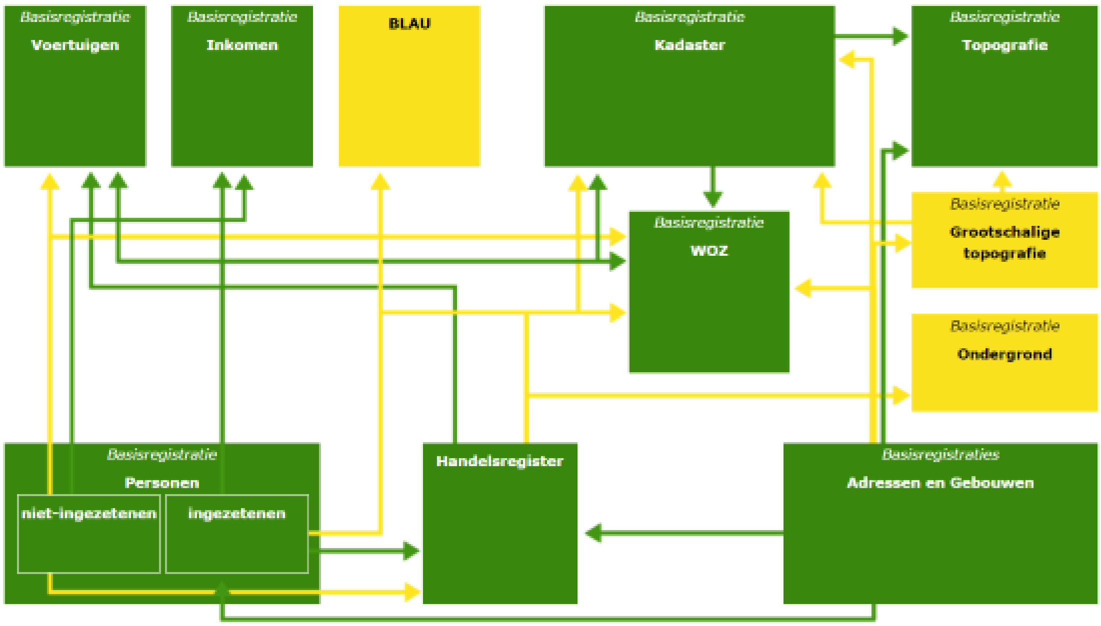


Figure 2.4: The system of key registers (in Dutch). Finished registers are displayed in green and unfinished registers in yellow (source: `http://tinyurl.com/nl2hhct`).

2.2.2 Smart living

Smart living is the other relevant smart city element for this study. Due to the trend of globalization and the rise of the service economy, capital and labor are more mobile than ever before. Because of this volatility, cities try to be as competitive as possible to attract or retain this capital. It is against this background that cities try to distinguish themselves from other places by focusing on maximizing the 'quality of life' (Rogerson, 1999). The concept of smart living is based on using ICT to maximize the quality of life.

In their effort of ranking cities by their 'smartness', Giffinger et al. (2007) mention culture, health, safety, housing and tourism as some aspects of quality of life. Rogerson (1999) performed a literature study to find the different factors that determine quality of life. What becomes apparent is that there are many factors that can influence quality of life. The most common are: environment and pollution, housing costs and access, health care and public health, education provision and levels, art and cultural diversity, and crime and public safety. For this research, the focus is on this last aspect: crime and public safety. How can ICT and smart governance improve the quality of life with regard to crime and public safety?

Problem-oriented policing

The practical concept of Problem-Oriented Policing (POP) is a policing strategy that is developed as a reaction on the estrangement of the police from the community observed in the 1960s in the United States. This period was characterized by urban riots and rising crime rates. The activities of the police were mainly focused on law enforcement and maintaining order (Jenkins, 2014). POP advocates a pro-active strategy against crime, instead of the reactive and incidence-based approach that was common at the time (Spelman and Eck, 1987). It focuses on long-term analysis of enduring problems causing crime as opposed to focusing on only the most recent occurrences of crime (Tilley, 2003). Therefore, POP is about being close to the community, knowing what is going on under the surface, and reducing the fear of crime of citizens(Jenkins, 2014). Boba (2003) provides a useful definition of problem-oriented policing.

> "An approach/method/process conducted within the police agency in which formal criminal justice theory, research methods, and comprehensive data collection and analysis procedures are used in a systematic way to conduct in-depth examination of, develop informed responses to, and evaluate crime and disorder problems."

POP and the related data collection and analyses benefit from developments in ICT, facilitating for example: finding patterns and concentrations in crime, modeling crime occurrence and making forecasts. Weisburd et al. (2010) assessed the effectiveness of POP in reducing crime and disorder and they found that POP interventions have a statistically significant effect on the outcomes examined. Problem-Oriented Policing can be seen as an example of using ICT to improve the quality of life, i.e. by improving the personal safety of citizens within a smart city.

The principles of POP have been applied in a case study of burglaries of single-family houses in Savannah, Georgia (USA) (Scott, 2004). This research project was undertaken by the Savannah Police Department. The data for the analysis of burglaries came from case files, environmental surveys of burglary sites and interviews with and surveys of police specialists, offenders, citizens and victims. By performing spatiotemporal analysis the researchers found out that a relatively high percentage (36%) of burglaries occurred near a school. Furthermore, they found out that many burglaries took place during school hours. These observations led, together with interviews with offenders, to the conclusion that daytime burglaries are closely related to truancy. Based on this conclusion, the researchers recommended to improve the prevention and control of truancy to help decrease burglary rates.

This example shows how POP tries to find the underlying causes of crime by making use of different kinds of analysis. This allows the discovery of relationships between different factors and crime occurrence and possible explanations of crime. Moreover, the case study shows that the POP-approach and analyses can culminate in clear recommendations for strategic levels of police departments and (local) governments.

Smart governance concepts can potentially improve the methodology used in Savannah. For example, many of the environmental data used, lighting conditions, size of the lot and distance from roadways and neighbors, were obtained using time consuming environmental surveys. As these types of data are typically available within other governmental organizations, a horizontal integration of governmental organizations (the fourth stage of e-government development (Layne and Lee, 2001)), can facilitate easy access to these data. Therefore, a smart city can be a facilitator of problem-oriented policing.

Predictive policing

Another practical smart living concept, closely related to problem-oriented policing, is 'predictive policing'. "*Predictive policing is the application of analytical techniques, particularly quantitative techniques, to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions*" (Perry, 2013). Predictive policing needs relevant input data for its analyses to function. These can for example be data on crime incidents, infrastructure, demographics, types of businesses, socioeconomic statistics et cetera. The results of these analyses can be helpful in allocating limited police resources to the areas that need them the most (Willems and Doeleman, 2014).

The predictive policing methods can be divided in four categories, based on a study of predictive policing by Perry (2013) (p. xiv).

1. Methods for predicting crimes. These approaches are used to forecast places and times with an increased risk of crime.

2. Methods for predicting offenders. These approaches identify individuals at risk of offending in the future.

3. Methods for predicting perpetrators' identities. These techniques are used to create profiles that accurately match likely offenders with specific past crimes.

4. Methods for predicting victims of crimes. Similar to those methods that focus on offenders, crime locations, and times of elevated risk, these approaches are used to identify groups or, in some cases, individuals who are likely to become victims of crime.

An example of an application of predictive policing in the Netherlands is the *Criminaliteits Anticipatie Systeem* (CAS; Crime Anticipation System). This system is developed by the Amsterdam Police Department together with the VU University Amsterdam. The system gathers many types of data over three years with a two week time interval, including data about crime incidents, distance to known suspects, distance to the nearest highway ramp, type and number of businesses, demographics and socioeconomic statistics. These data are assigned to a grid of areas measuring 125 by 125 meters. A neural network, a data mining technique, is then used to find patterns in this large data set. Finally, the outcomes of this analysis are displayed as a heat map of Amsterdam, featuring the risk of future crime incidents per area (see figure 2.5). These heat maps can be used to better allocate police resources.

The results of CAS are positive. Crime incidents decreased where extra resources were allocated based on CAS. Also the CAS resulted in better justification of the allocation of police resources: decision making is better informed and not just based on a ‘hunch’ (Willems and Doeleman, 2014). Other local governments are eager to put CAS to the test in their cities (van Dijk, 2015).

Both the smart living concepts of problem-oriented policing and predictive policing show how ICT and geographic information systems can play a major role in fighting crime and increasing the public safety, factors that play an important part in the overall quality of life in a city. It is also abundantly clear how a smart governance is complementary to the smart living concepts described above. A smart governance has the potential to lay the foundation for easy access to high quality and structured data to serve as input for problem-oriented policing and predictive policing analysis methods. Meanwhile, the smart living concepts of problem-oriented and predictive policing can improve communication with citizens and make a public administration more accessible, effective and transparent, which are all goals of a smart governance.

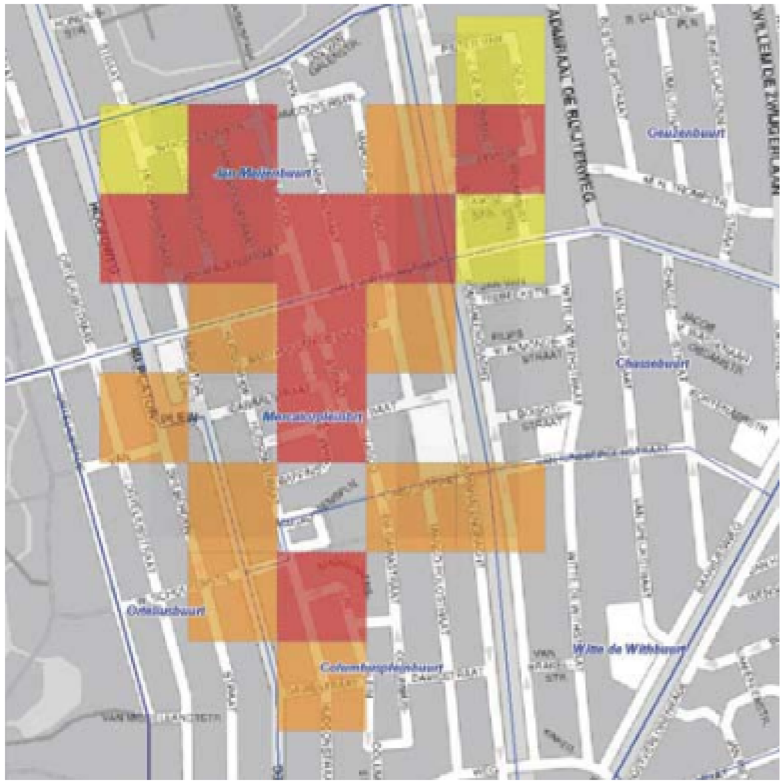


Figure 2.5: A heat map from Amsterdam showing the results from the *Criminaliteits Anticipatie Systeem* (Willems and Doeleman, 2014).

2.3 Smart city critiques

Where the previous section mostly emphasized the advantages and opportunities of the smart city, this image is balanced by discussing some of the critique that is being expressed with regard to smart cities.

In their discussion of smart city characteristics, Caragliu et al. (2011) already warn for the excluding effect that the implementation of smart city concepts can have. Not everyone can be expected to have the same level of education and access to services. In addition to this, smart city concepts could spur the socioeconomic development in one area, but can leave other areas in decline. This is due to the volatile nature of human and social capital. An extreme example of this is 'brain drain': the selective migration of higher-educated people (Beine et al., 2001). The warning here is that smart city concepts have a tendency of looking at a city as a closed system, ignoring its local embeddedness in a broader system of cities, villages and rural areas. The local and regional effects of the development of smart cities should be taken into consideration.

Another point of critique is put forward by Hollands (2008) with regard to the business-led nature of many smart city initiatives. Examples are provided where local governments work closely together with private companies to gather the necessary funds for smart city developments. This can be in the form of public-private partnerships or within more vague terms of business 'co-operation' or 'consultation'. These kinds of constructions can cause potentially conflicting interests and contradictions with respect to for example social and environmental sustainability.

Furthermore, Hollands (2008) states that the term smart city is often used as place marketing, without referring to actual infrastructural change or evidence of workable and effective IT policies. City marketeers try to use 'smart' terms to attract businesses and higher-educated workers, without taking a holistic approach towards the concept or making any structural changes.

In line with the previous critique, in his book 'Against The Smart City', Greenfield (2013) has critique on the so-called 'turnkey smart cities': "*the proposition that a smart enough city, for example one built from scratch with a single dominant supplier or alliance of suppliers and no existing infrastructure or accredited urban culture to deal with, will provide perfect knowledge of the needs of its citizens and be able to meet them perfectly*" (Griffiths, 2013). Greenfield (2013) contests the notion of being able to develop smart cities anywhere you like by following a standard recipe. Every city differs by its specific geographies, social milieus and inhabitants, and developments towards a smart city should account for these differences.

Hollands (2008) adds that "*smart cities have to be more than just broadband networks*" and that "*being connected is no guarantee of being smart*" (p. 310). He refers to the South American city of Lima where "*despite increasing rates of telecommunication diffusion, in 1990 less than half of all households in the city had a phone and only seven per cent had access to the internet, with the poorest 50 times less likely to have the internet*" (p. 310).

And finally, and perhaps the most important source of critique: violation of privacy. As already mentioned in the introduction, Martinez-Balleste et al. (2013) emphasize that residents need to be aware that smart cities have the ability to silently gather a variety of information about them.

The concept of predictive policing introduced earlier, is a good example to demonstrate the privacy concerns surrounding smart cities. Bits of Freedom is a Dutch digital rights organization, focusing on privacy and communications freedom in the digital age [5]. Every year they hand out the 'Big Brother Award' to the person, company or governmental organization that violates the privacy of citizens the most. In 2015 this award was rewarded to the Dutch minister Plasterk (Minister of Home Affairs) and the head of the Dutch National Police for their efforts regarding predictive policing (NU.nl, 2015). The critique focuses on the goal of acquiring information about citizens on a large scale and the use

[5] For more information `https://www.bof.nl/home/english-bits-of-freedom/`.

of a national network of sensors to predict crime events in the near future and to act proactively. The fear is that citizens are marked as suspects based on smart algorithms and sensors, rather than because of actually being guilty of committing a crime (Big Brother Awards, 2015).

Wim Elfrink, head of the smart cities team at the private ICT-company Cisco, states that privacy must play an instrumental role in any smart city strategy, otherwise citizens might fear the introduction of other innovative technology. *"Having security policies, having privacy policies is a given. I think you have to first give the citizens the right to opt-in or opt-out. Then all these policies, longer term, security and privacy are going to be the biggest imperatives. If we don't solve this, people will opt-out more"* (Datoo, 2014). Martinez-Balleste et al. (2013) also emphasize the importance of the preservation of privacy in order for smart cities to succeed. They state that legislation is essential to guarantee the achievement of privacy in smart cities.

This section demonstrates that although smart cities are a promising concept, at the same time there are many concerns that should be addressed when developing or implementing smart city strategies or policies. For example, the proposed analysis of burglaries will have to find a balance between privacy and level of detail. These types of considerations and trade-offs are discussed in more detail in section 3.4.

2.4 Conclusion

As became clear in this chapter, the literature on smart cities and the associated topics are very broad. In this concluding section the most relevant concepts from the smart city literature are summed up to provide a more focused conceptual background that forms the basis for the spatiotemporal analysis of burglaries (see figure 2.6). The smart city theoretical framework used in this research is based on the following starting points.

- Using ICT as a mean to reach the goals. Thanks to developments in ICT, whether these are expanding sensor networks, advancements in data mining techniques or developments in GIS, previously impossible or complex types of analyses are made easier and more accessible.

- Focus on improving personal safety and perceived safety. This is based on the smart city element of smart living and its focus on improving the quality of life.

- Focus on structural factors causing crime. Not just looking at single crime events, but trying to find explanations for the spatial or temporal patterns in aggregated data. This characteristic is based on notions from the concepts of problem-oriented policing and predictive policing.

- Using integrated and structured datasets. This characteristic relates directly to the vertical and horizontal integration and the development of

SDIs within smart governance. The spatiotemporal analysis of burglaries aims at using data on varying subjects and from different sources, mostly key registers, for the analysis of burglaries. A good way of creating relationships between different datasets, is through using geography. By incorporating a spatial component in each dataset, these datasets can be compared and analyzed to find relationships and patterns. Moreover, because of the use of data from the Dutch key registers, the analysis method can be standardized to allow it to be applied in different Dutch municipalities: applying standardized geoprocessing of spatial data as a way of taking SDIs to the next level.

To conclude, within this study the term smart city refers to the use of modern information and communication technologies and integrated datasets in a systematic way to conduct in-depth examination of, develop informed and transparent responses to, and evaluate crime and disorder problems to improve the personal safety and perceived safety and ultimately improve the quality of life of citizens.

The smart informational and cognitive processes associated with smart cities, such as information collection and processing, real-time alerts, forecasting, learning, collective intelligence and distributed problem solving, can be applied through the use of the problem-oriented policing and predictive policing concepts.

Especially, the notion of the smart city about integrated and open datasets, can be used to feed crime analysis and to ultimately improve the quality of life of citizens. Meanwhile, problem-oriented and predictive policing can improve communication with citizens and make a public administration more accessible, effective and transparent. In this way, smart city concepts are used as the theoretical background for the spatiotemporal analysis of burglaries, keeping the critique on smart cities in mind.

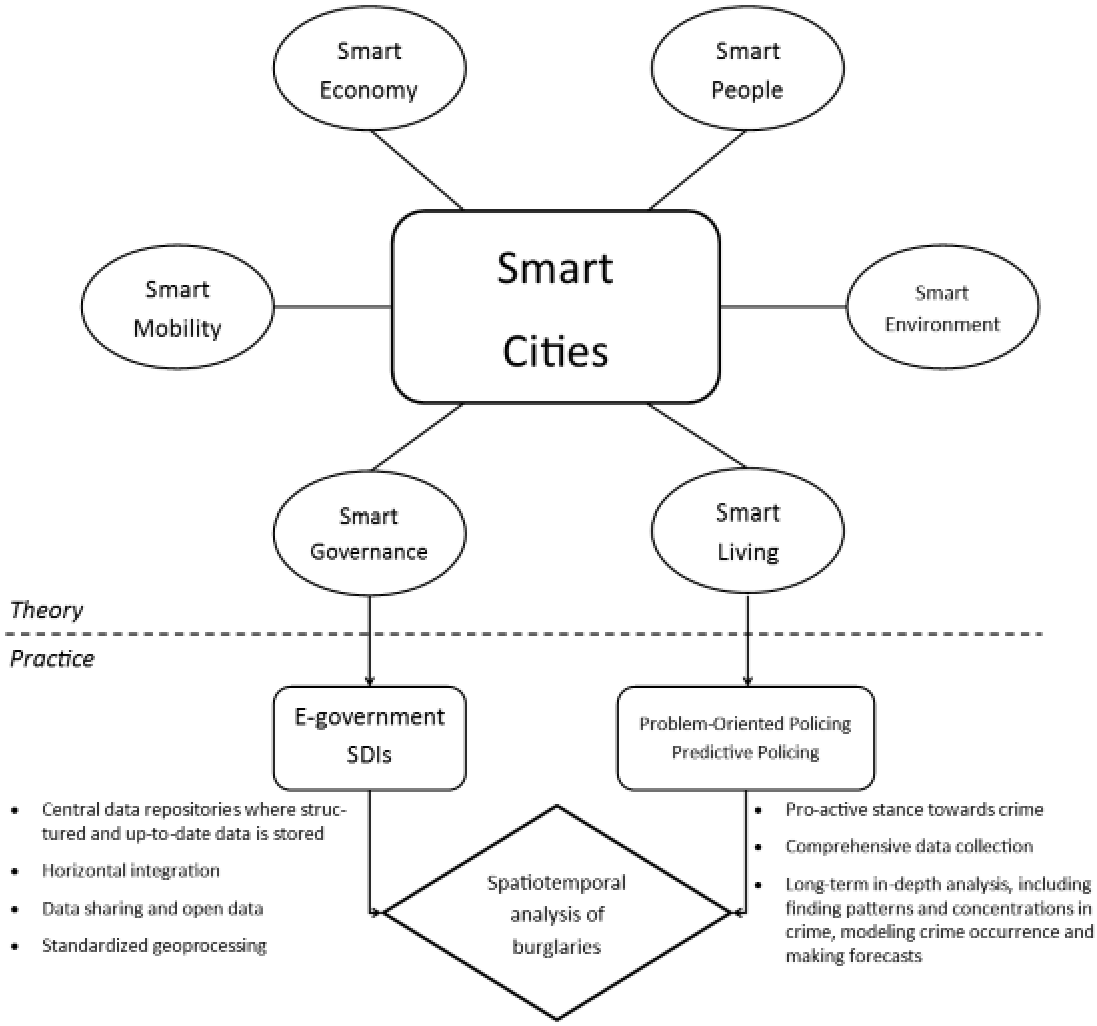


Figure 2.6: Conceptual model for the application of smart city concepts in the spatiotemporal analysis of burglaries.

Chapter 3

Methods for the spatiotemporal analysis of crime data

This chapter focuses on the subquestion: *what methods are available for the spatiotemporal analysis of crime data?* First, a general overview of common methods for spatiotemporal analysis of crime is provided. This overview is based on a review of literature. Then, based on these general methods, the specific method for this study is presented, which combines elements from the different more common methods. The chapter concludes with an overview of the proposed method in the form of a flow diagram.

3.1 Spatiotemporal analysis of crime

The spatiotemporal analysis of crime patterns is not something new. That is why this section takes a look at the existing methods for the spatiotemporal analysis of crime. These methods all have a distinct spatial component and utilize geographic information systems. Therefore, they are often also referred to as methods of crime mapping.

Groff and La Vigne (2002) present a categorization and discussion of methods for analyzing current patterns of crime and for predicting future patterns. They present the different characteristics and advantages and disadvantages of these methods. Their categorization forms the basis of the categorization used here and includes:

- hot spot and near repeat methods;
- grid and raster methods; and
- univariate and multivariate regression methods.

These broad categories, specific methods and their characteristics are discussed next.

Hot spots and near repeats

Possibly the most well-known method of spatiotemporal analysis in criminology is the hot spot analysis (see figure 3.1). Hot spots can be described as areas with a high concentration of crime. Eck et al. (2005) provide the following definition: *"a hot spot is an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization"* (p. 2). Cool spots are the opposite of hot spots: areas where crime is below average.

Hot spot methods are often used not only to get an insight in the current spatiotemporal patterns of crime, but also to get an idea of future patterns of crime. The assumption is made that hot spots of today will persist in the future (Perry, 2013). Hot spots can be analyzed over different time intervals. Hot spots tend to be more mobile when shorter time intervals (weeks) are considered, but are much more consistent across longer time intervals (months or years). To find the structural spatial patterns of crime, a longer time interval has to be used for the hot spot analysis (Groff and La Vigne, 2002). Because hot spot methods only use crime data, it is hard to explain why hot and cold spots occur in a certain place (Perry, 2013).

Hot spot analyses are relatively easy and fast to conduct using GIS applications. This partly explains the popularity of the method in the practical field. Hot spot analyses can be conducted continuously to get an almost live image of where crimes tend to occur.

Most GIS software packages are able to perform some form of hot spot analysis by looking at the density of crime incidents. There are also some specialized packages available like STAC[1] and CrimeStat[2].

Apart from the practicability, hot spot analyses are also popular because they are easy to interpret (Groff and La Vigne, 2002). Therefore, hot spot methods are suitable for communicating crime patterns to a broad audience.

The theory of 'near repeats' is based on the hot spot analysis and is specifically aimed at burglaries. The assumption of the near repeats concept is that burglaries occur closer to each other in both space and time than can be expected based on chance (Johnson et al., 2007). To be more specific, dwellings within 400 meters of a burgled home, run an increased risk of being burgled within the next fourteen days (Johnson, 2008). Other researchers even indicate a period of increased risk of two months (Bowers et al., 2004).

The concept of near repeats is derived from theories about the spreading of infectious diseases. Explanations for the near repeat phenomenon include *"burglars coming back for items they left the first time (early repeats), burglars coming back to steal replacement items (delayed repeats), and burglars telling*

[1] See for more information `http://tinyurl.com/q97era9`.
[2] See for more information `http://tinyurl.com/pyheua5`.

other burglars that a particular residence or area is a suitable target" (Moreto et al., 2013).

Just like with hot spot methods, near repeat methods only use crime data to predict where crime might occur in the near future. Therefore, also near repeat methods are not catered to finding explanations for the spatiotemporal patterns found.

The main advantage of hot spot and near repeat analyses is that the results are easy to interpret. This makes these methods a strong tool for presenting crime patterns (see also figure 3.2).

The most important disadvantage is that hot spots and near repeats do not tell anything about the structural causes of the observed patterns. This makes it hard to create strategies to address these underlying issues that cause crimes (Groff and La Vigne, 2002). With respect to making forecasts, a disadvantage of near repeat and hot spot methods is that they need initial crimes to function, as this method does not use structural explanatory variables as inputs. If there are few or no initial crimes to begin with, hot spot and near repeat analyses provide no accurate results or are not possible to conduct at all (Caplan and Kennedy, 2010).

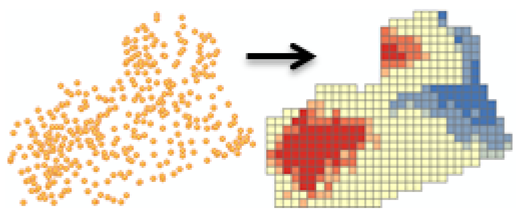


Figure 3.1: Example of a hot spot analysis based on point data input (image source: `http://tinyurl.com/ocxgjpo`).

Figure 3.2: The resulting image of a hot spot analysis based on burglaries in Haarlem. It demonstrates how this type of analysis can produce rather dramatic images to deliver a certain message.

Grid and raster methods

Grid and raster methods divide the study area into a series of equally sized cells. This grid or raster approach is used to combine different data layers and values into one opportunity surface (see figure 3.3).

Each data layer represents a risk factor for crime that is operationalized to values for each cell of the grid or raster. Besides the inclusion of risk factors that increase the risk of crimes, also factors that reduce the risk of crime can be incorporated. The relevant risk factors are derived from scientific literature. Map algebra is used to combine the grid cell values of each layer to create one risk surface. The grid cell values are based on the presence or absence of a risk factor and are operationalized with a value of 0 or 1 (Groff and La Vigne, 2001, 2002; Caplan and Kennedy, 2010).

By default, the different data layers have the same weight, but different weights can be assigned to them based on a regression analysis. Regression analysis can be used to observe the presence or strength of a relationship between an independent and dependent variable, in this case between a risk factor and the occurrence of crime.

The advantages of cell-based methods are the transparency of the straightforward process and an easy visualization of the results. Grid and raster based methods have the potential to be easily accessible to the average law enforcement analyst and are suitable for use as decision support tools (Groff and La Vigne, 2002; Caplan and Kennedy, 2010). But there are also some essential disadvantages to this method.

First, the amount of data available has a direct effect on the level of detail that can be incorporated in the model (Groff and La Vigne, 2002). The size of the cells and therefore the accuracy of the method depends on the availability and spatial resolution of the input data.

A second disadvantage is in the operationalization of risk factors to values of 1 or 0. While this operationalization is very straightforward and simple, at the same time much detail in the data is lost due to this simplification of reality. More gradual spatial patterns, like distances or densities, are less accurately modeled using only two possible values.

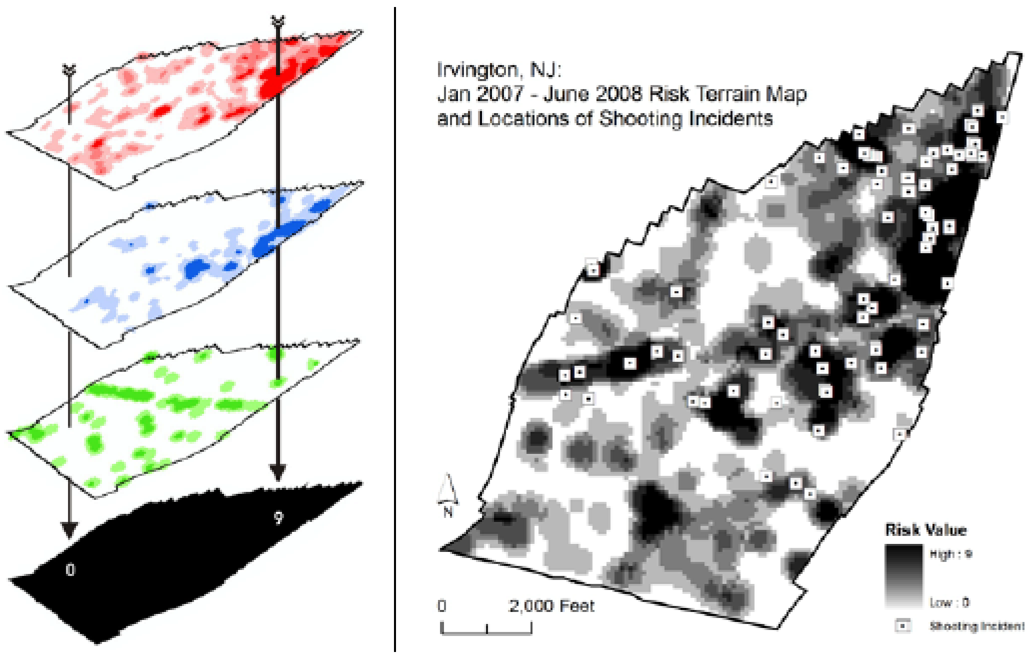


Figure 3.3: An example of a raster method where a composite risk terrain map combines different risk factor data layers (Caplan and Kennedy, 2011).

Univariate and multivariate regression methods

The last category of methods for analyzing spatiotemporal patterns of crime consists of univariate and multivariate regression methods.

Univariate regression methods use the previous value of a variable to predict the future value (Gorr and Olligschlaeger, 2001; Groff and La Vigne, 2002). Therefore they are similar to hot spot methods, but the latter is inherently spatial where univariate regression methods are more temporal in nature. Because univariate regression methods only use one variable the data requirements are minimal. Besides that, univariate regression methods are atheoretical, meaning it is not necessary to consider which variables have to be included in the model. Within the univariate regression methods, a distinction can be made based on complexity. Complexity of the methods refers to the software employed and the skills required to use them.

The less complex methods are the 'random walk' and the 'naive lag 12' methods. The random walk method uses crime data of the previous month to make predictions for the next month. The naive lag 12 method takes the seasonality of crime into consideration and uses for example data of July to predict crime rates for July next year (Gorr and Olligschlaeger, 2001; Groff and La Vigne, 2002). While these methods are less complex univariate regression methods, they are also less accurate in their predictions (Groff and La Vigne, 2002).

An example of a more complex method is 'exponential smoothing'. Smoothing models assign more weight to more recent data points and this weight decreases exponentially over time. Predictions with the exponential smoothing method are based on a longer period of time than with less complex methods. This causes random errors in the data to be evened out (Gorr and Olligschlaeger, 2001). That partly explains why exponential smoothing is an accurate method for the prediction of small to medium changes in crime patterns (Groff and La Vigne, 2002).

Multivariate regression methods use more than two variables. An example is the 'leading indicator' method. This method uses current and past values of various independent variables to predict the values for the dependent crime variable. The method uses literature to identify the independent variables that can be used for predicting crimes, grounding this method in theory.

In order to develop robust model parameters, it is essential that the spatial units are large enough to include enough observations. Moreover, it is important that the independent data has the same spatial and temporal scale as the dependent data.

Multivariate regression methods perform well at the prediction of large changes in crime patterns and they can therefore be considered an addition to univariate regression methods. A major advantage is that multivariate regression methods allow the inclusion of spatial and temporal lags in the model (Groff and La Vigne, 2002). Groff and La Vigne (2002) state that: *"Spatial lags allow the explicit modeling of the effects of the values in neighboring cells on the value of the subject cell. Temporal lags allow the modeling of the effects during previous time periods on the study time period"* (pp. 44–45).

A disadvantage of this method is that, in order to apply it, significant expertise on the part of the end user is required in the field of multivariate statistical methods (Groff and La Vigne, 2002).

Another promising multivariate regression method uses 'artificial neural networks'. This method is used by the *Criminaliteits Anticipatie Systeem* discussed earlier (see section 2.2.2). Although there are multiple types of neural networks, the multi-layer feed-forward networks with backpropagation is the most studied type and is therefore used here as an example. As Olligschlaeger (1997) states: *"regardless of type, all artificial neural networks consist of a number of processing units that send signals to one another via a large number of weighted connections"* (p. 322). More specifically, *"feed-forward networks with backpropagation 'learn' to map the input units to the output units by adjusting the weights on the connections in response to error signals transmitted back through the network"* (p. 325). This process is repeated until the margin of error is minimized.

The advantage of this method is that the resulting model scores relatively well on predictive accuracy. But there are also some important disadvantages. The model does not include any tests of significance, making it very hard to define which input variables have a significant effect on the outcome event. Furthermore, the method is more or less atheoretical, because theory does not play a central role in selecting the relevant input variables. Another downside of neural networks is that its use needs a lot of technical expertise, making this method less accessible for mainstream users. Moreover, neural networks require significant computing power and time (Groff and La Vigne, 2002).

Method selection

In order to select the most suitable method, several selection criteria are defined based on the goals of this study.

- This study focuses on the **structural causes** of burglaries based on notions from problem-oriented policing, instead of focusing on short-term variations. This also means **significance tests** should be included to assess the significance of the identified risk factors in explaining the outcome event of burglaries.
- **Spatial effects** should be taken into account. The effect of 'nearness' should be quantifiable and controlled for.
- The method that is used should be able to support decision making processes. This means that the method should be as **transparent** as possible, allowing a justification of the results found and the methods used. Additionally, the results of the analysis should be **easy to present and interpret** by a wide audience.

Considering these points, multivariate regression methods, complemented with elements from grid and raster methods, form the basis for this study. These methods are suitable for finding the structural causes of crime and include significance tests to assess and quantify the explanatory power of multiple independent variables. For example, hot spot methods lack this ability and are more directed towards identifying spatiotemporal patterns of crime, instead of explaining them. Neural networks lack the necessary significance tests.

Multivariate regression methods are also suitable for quantifying spatial effects in the model, for example via the inclusion of a spatial lag variable.

Finally, multivariate regression methods are suitable for use in decision support because of their transparency. Neural networks, although promising in their predictive qualities, are much less transparent. It is difficult to oversee what neural networks exactly do and how they work, making them act like a black box (Olligschlaeger, 1997). This is certainly not a method characteristic that matches the requirement of transparency.

As the results of multivariate regression methods often include a number of statistical measures, they are not always easy to interpret and present to a wide audience. Elements of the grid and raster methods can help here. The concept of a combined risk surface from the grid and raster methods can be a good addition in presenting the results of the multivariate regression analysis. The areas with the highest risk of burglaries can be identified based on the significant explanatory risk factors.

To conclude, the proposed method for the spatiotemporal analysis of burglaries includes two major steps.

1. Applying a multivariate regression method to find the significant variables in explaining the spatiotemporal patterns of burglaries and defining their explanatory power relative to each other.

2. Using a risk terrain model to present the results of the regression analysis and to identify the areas with the highest risk of burglaries.

The next sections go deeper into the multivariate regression and raster method used for the spatiotemporal analysis of burglaries.

3.2 Multivariate regression methods

3.2.1 Dependent data descriptive statistics

After the decision to use a multivariate regression for the spatiotemporal analysis of burglaries, there are still different multivariate methods to consider. At the basis of all multivariate regression models is the assumption that there is a causal relationship between a dependent variable Y and every independent variable X_i (de Vocht, 2008). The decision for a suitable multivariate regression method is largely based on the characteristics of the data of the phenomenon that is being explained. Here this phenomenon is residential burglaries. So let us first take a look at the burglary data of the study area Haarlem.

Crime figures are commonly expressed as crime rates: the total count of crimes in a certain period of time per spatial unit (Osgood, 2000; Helbich and Jokar Arsanjani, 2014). Therefore, crime data is also referred to as count data. When shorter time periods or smaller spatial units are used to aggregate data, crime rates are relatively low and will often contain many 0 values at locations where no crimes occurred. This results in a very distinct histogram of crime rates.

This can be demonstrated by taking a look at the data distribution for burglary counts in 2010 in Haarlem aggregated to 100 meter square cells. Figure 3.4 shows a map of the burglary rates for Haarlem in 2010, figure 3.5 shows the frequency table and figure 3.6 shows the same data in a frequency bar chart. It must be noted that the burglary rate is expressed here as the number of burglaries per 1000 residential addresses. This is to correct for the number of addresses per cell.

The map demonstrates again how burglaries tend to concentrate in certain areas: the initial observation that motivated this research. Furthermore, the frequency bar chart clearly shows the distinct pattern for crime data: many areas with no incidents and a few areas with many incidents. This distinct data pattern is key in selecting a suitable multivariate regression model.

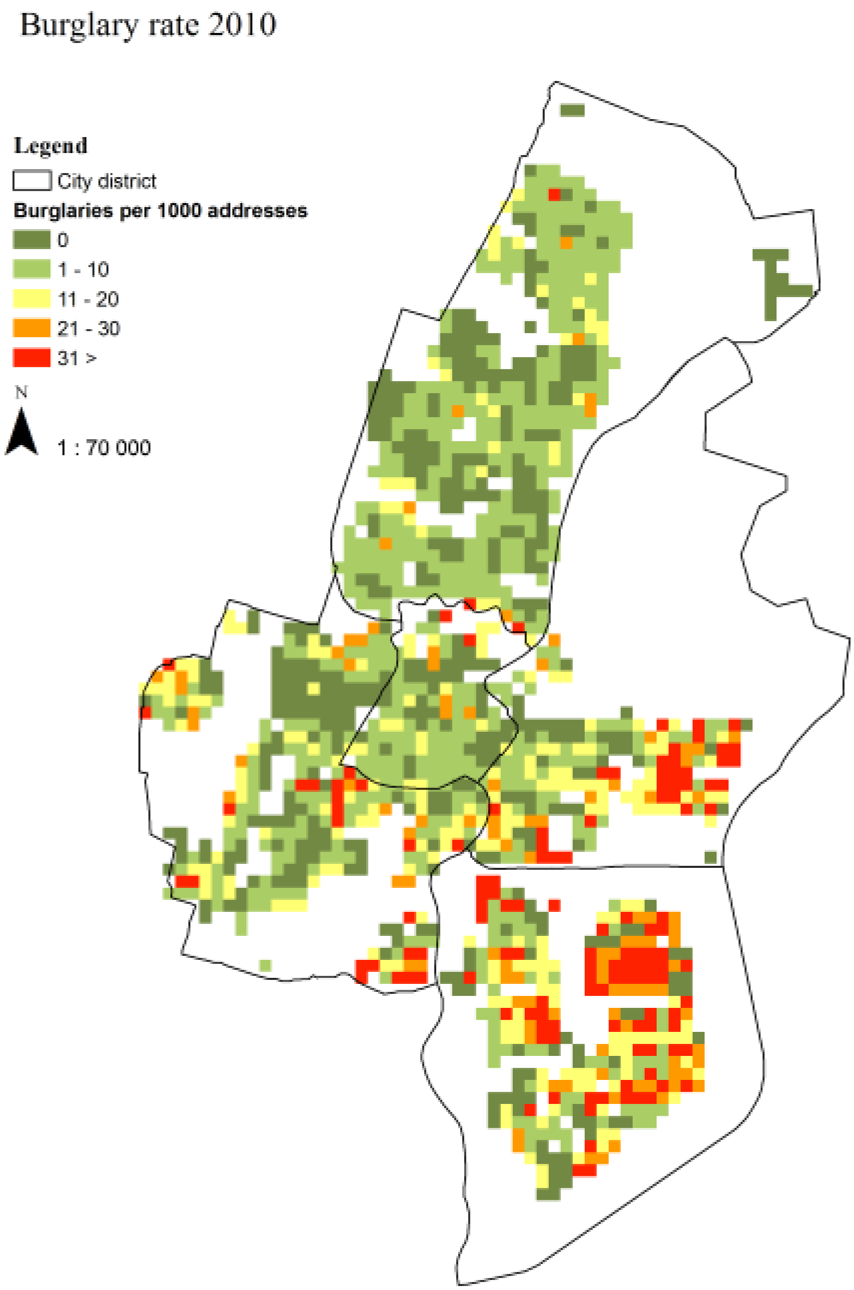


Figure 3.4: Map showing the number of burglaries per 1000 residential addresses for Haarlem in 2010. The grid cells used are 100 by 100 meter.

Burglary rate	*Frequency*	*Percent*	*Cumulative percent*
0 -13	1083	74,18	74,18
14 - 26	212	14,52	88,70
27 - 40	87	5,96	94,66
41 - 53	30	2,05	96,71
53 - 66	21	1,44	98,15
67 - 79	12	0,82	98,97
80 - 92	7	0,48	99,45
93 - 106	4	0,27	99,73
107 - 119	3	0,21	99,93
120 - 132	1	0,07	100,00
Total	1460	100,00	

Figure 3.5: Frequency table of burglary rates for Haarlem in 2010.

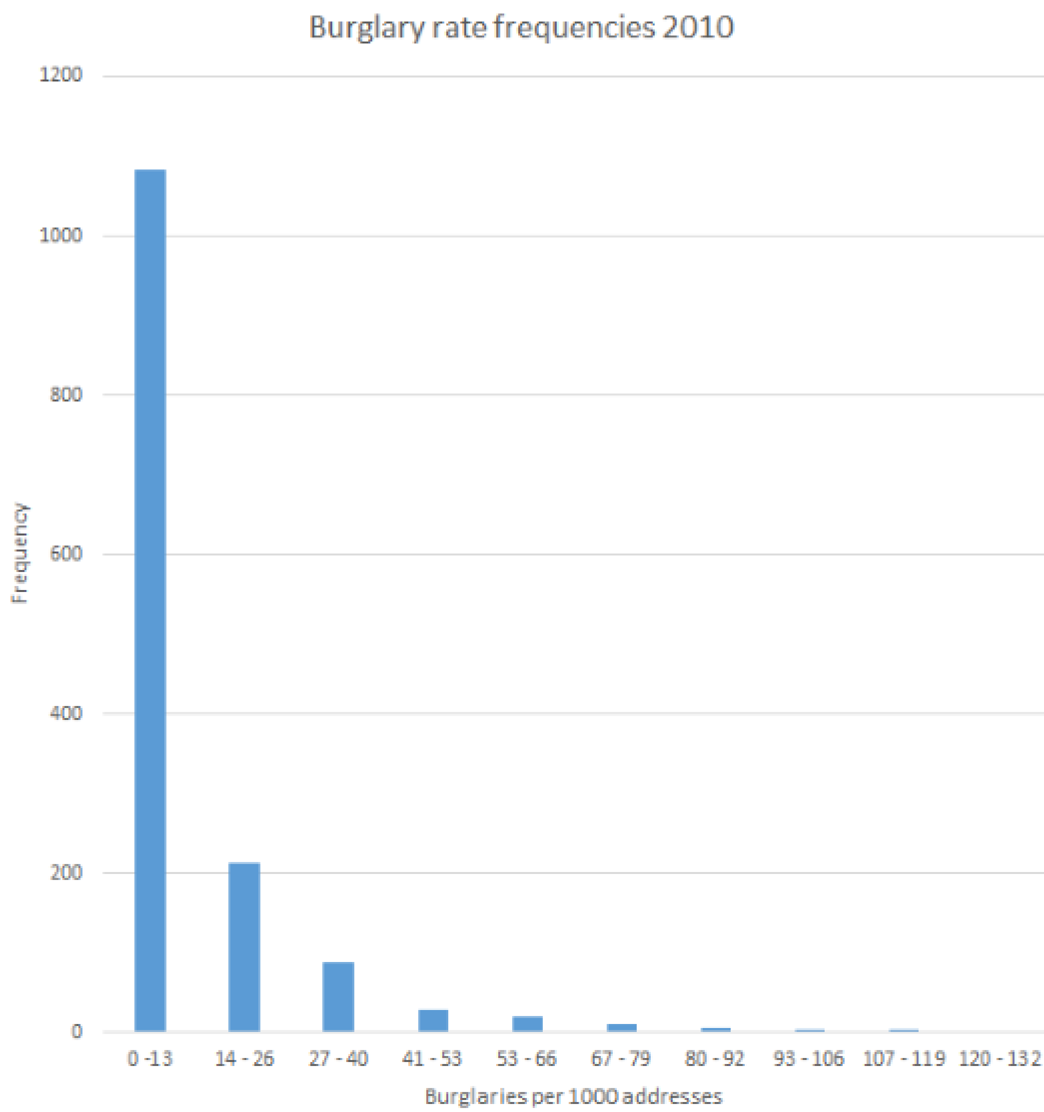


Figure 3.6: Bar chart showing the frequencies of burglary rates for Haarlem in 2010 based on 100 meter square grid cells. The crime rates are reclassified to ten classes of equal size.

3.2.2 Linear regression

Probably the most common multivariate regression method is linear regression using ordinary least squares (OLS). This type of regression is based on a gaussian, or normal, distribution of the dependent data. The bar chart clearly shows that the burglary count data does not follow the typical bell-shaped gaussian distribution, but is positively skewed (see figure 3.6). This makes the use of gaussian models unsuitable (Helbich and Jokar Arsanjani, 2014; Gardner et al., 1995; Osgood, 2000).

A common strategy for addressing this problem is to transform the data so they become less skewed (Osgood, 2000; Helbich and Jokar Arsanjani, 2014). Osgood (2000) states that although a logarithmic transformation renders the data more suitable for linear regression analysis, other issues are introduced.

For example, a logarithmic transformation requires adding a constant to values of zero. Often a value of 1 is added, but this choice is very arbitrary and it may drastically affect the results (Osgood, 2000). This ultimately makes the analysis results less reliable.

Another method to make the data less skewed is to aggregate the data. This means using a lower spatial resolution for the analysis, decreasing the number of zero values. For this study, this option is undesirable because details on local patterns of burglaries will be lost.

Another reason why linear models are unsuitable is that they allow for the prediction of negative numbers, while crime counts can by definition only be positive integers (Helbich and Jokar Arsanjani, 2014; Osgood, 2000).

All in all, it shows that the commonly applied linear regression is an unsuitable method for modeling crime counts. Ordinary least squares regression models, with or without log transformation, are being discouraged with count data and it is being recommended to explicitly consider the nature of this data type (Helbich and Jokar Arsanjani, 2014). But which multivariate regression method does meet these requirements?

3.2.3 Poisson regression

A good alternative for crime count data are Poisson-based regression models (see figure 3.7)(Osgood, 2000; Gardner et al., 1995). These models can solve the problems described above because they recognize the dependence of crime rates on counts of crimes (Osgood, 2000). The Poisson distribution has been useful in many problems in criminology and criminal justice, for example in assessing the potential for selective incapacitation, projecting prison populations, and estimating the size of the criminal population. Osgood (2000) even states that *"Poisson originally derived the distribution for analyzing rates of conviction in France during the 1820s"* (p. 23).

Osgood (2000) provides a useful description of the main features of Poisson distributions (p. 23):

> The Poisson distribution characterizes the probability of observing any discrete number of events (i.e., 0, 1, 2, ...), given an underlying

> mean count or rate of events, assuming that the timing of the events is random and independent. For instance, the Poisson distribution for a mean count of 4.5 would describe the proportion of times that we should expect to observe any specific count of robberies (0, 1, 2, ...) in a neighborhood, if the "true" (and unchanging) annual rate for neighborhood were 4.5, if the occurrence of one robbery had no impact on the likelihood of the next, and if we had an unlimited number of years to observe.

This quote from Osgood (2000) immediately shows that there are some assumptions related to Poisson-based models. One assumption is that the fitted value λ_i is the true rate for each case, which implies that the explanatory variables account for all of the meaningful variation among the aggregate units (Osgood, 2000). Osgood (2000) already states that it is very unlikely that this assumption will be valid, for there is no more reason to expect that a Poisson regression will explain all of the variation in the true crime rates than to expect that a linear regression would explain all variance other than error of measurement. It is very likely that this is also true for the analysis of burglaries in Haarlem, as it is not realistic to expect that all relevant explanatory variables for this complex social phenomenon can be identified.

The other assumption is that there is no dependency among individual crime events (Osgood, 2000; Huang and Cornell, 2012). This assumption is not met too. Osgood (2000) mentions possible sources of dependency in crime data: *"individual offending at a high rate over a brief period until being incarcerated, multiple offenders being arrested for the same incident, and offenders being influenced by one another's behavior"* (p. 28). For burglaries specifically, as became clear from the theory related to near repeats (see section 3.1), dwellings within 400 meters of a burgled home run an increased risk of being burgled within the next fourteen days (Johnson, 2008).

Because these assumptions are not met, 'overdispersion' in which residual variance exceeds λ_i, is ubiquitous in analyses of crime data (Osgood, 2000). Osgood (2000) explains that *"applying the basic Poisson regression model to such data can produce a substantial underestimation of standard errors of the β's, which in turn leads to highly misleading significance tests"* (p. 28). β refers to the regression coefficient. That is why the alternative 'negative binomial regression' comes into play.

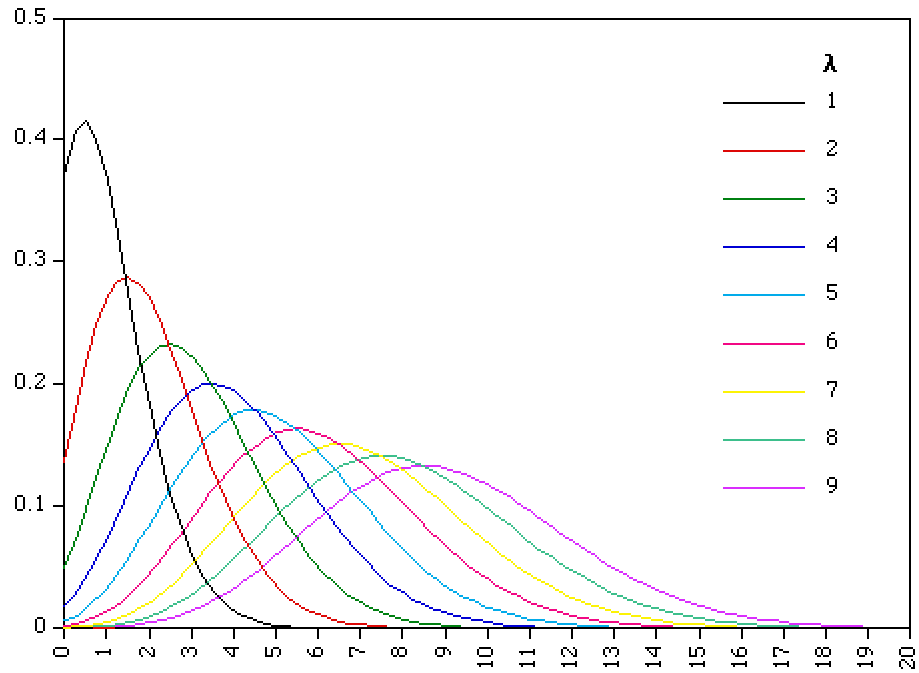


Figure 3.7: Several poisson distributions are portrayed with different mean counts λ (source: `http://www.umass.edu/wsp/images/poisson3.gif`).

3.2.4 Negative binomial regression

The negative binomial model incorporates a source of overdispersion in the probability model by adding a case-specific residual term to the regression model, comparable to the error term in OLS regression (Osgood, 2000). Osgood (2000) further explains that *"negative binomial regression combines the Poisson distribution of event counts with a gamma distribution of the unexplained variation in the underlying or true mean event counts, λ_i"* (p. 29).

Negative binomial regression is considered as the most suitable multivariate regression method for the spatiotemporal analysis of burglaries in Haarlem. The advantages of this method are summed up below.

- The specific **nature of the data**, counts of events, is respected. Negative binomial regression is, as it is based on the Poisson distribution, specifically developed for integer positive data values. Therefore, it can handle the skewed nature of count data caused by the many zeros. Linear regression for example does not respect the nature of the count data.
- Negative binomial regression allows **dependencies among individual**

burglary events. This is important because theory shows that burglary events influence other burglary events both spatially and temporally. Therefore, negative binomial regression is preferred over the Poisson regression that assumes that there is no dependence among individual crime events.

- Finally, there is **empirical evidence** that supports the use of negative binomial regression for analyzing crime count data and obtaining reliable and accurate results. Osgood (2000) shows a comparison between applying a standard OLS model, a log transformed OLS model, a basic Poisson model and a negative binomial model to crime rate data. The outcome is that the choice of an appropriate model, i.e. a model that fits the data, seriously affects the significance tests. OLS models wrongly find significant relationships between variables where the negative binomial model does not. For this study, these significance tests are key in determining whether or not there is a statistically significant relationship. Furthermore, the mean squared error was lowest for the negative binomial model and the R^2 value, the percentage of explained variance by the model, was higher than for the OLS models (Osgood, 2000). Also Huang and Cornell (2012) demonstrate how a negative binomial regression model fitted with overdispersed count data outperforms OLS regression. Using the best fitting and theoretically justifiable statistical model, i.e. the negative binomial model, improves the chances of drawing the right conclusions from the spatiotemporal analysis of burglaries.

3.3 Risk terrain modeling

The grid or raster method 'risk terrain modeling' (RTM) is used to visually present the results of the negative binomial regression and to identify risk areas (see figure 3.8). These risk areas can be used to target anti-burglary measures and to allocate police resources. RTM is a relatively simple method for assessing how geospatial factors contribute to crime risk. It is developed by Joel Caplan and his associates at Rutgers University (Perry, 2013). Caplan and Kennedy (2010) (p. 23) describe RTM as:

> (...) an approach to risk assessment that standardizes risk factors to common geographic units over a continuous surface. Separate map layers representing the presence, absence, or intensity of each risk factor at every place throughout a terrain is created in a geographic information system (GIS), and then all map layers are combined to produce a composite "risk terrain" map with attribute values that account for all risk factors at every place throughout the geography. Risk terrain maps assist in strategic decision making and tactical action by showing where conditions are ideal for events to occur in the future.

Traditionally, measures against crime are focused on crime hot spots. This can introduce the problem of the 'whack-a-mole' effect, where measures reduce crime rates in one area, but increase crime rates elsewhere. Because RTM produces an overview of high risk areas based on structural characteristics, this effect can be mitigated (Groff and La Vigne, 2002). RTM creates maps that show those areas with the highest risk of crime, not because a police report showed that most past crimes occurred in these areas, but because social and situational characteristics create favorable conditions for crime. Caplan and Kennedy (2010) state that "*the advantage of RTM is that it provides a picture of a landscape in terms of factors that contribute to negative events, such as crime, that are more enduring than just the characteristics of the people who frequent these places*" (p. 25).

So to summarize, risk terrain modeling can be used to combine the significant risk factors found during the negative binomial regression analysis into one weighted risk surface. The weights of the different layers can be derived from the regression coefficients. This risk surface can be used to identify high risk areas, based on structural risk factors rather than on just previous occurrences of crime. Therefore, risk terrain modeling can be a powerful tool for presenting the results of the regression analysis clearly and to aid in decision making concerning anti-crime measures and the allocation of police resources.

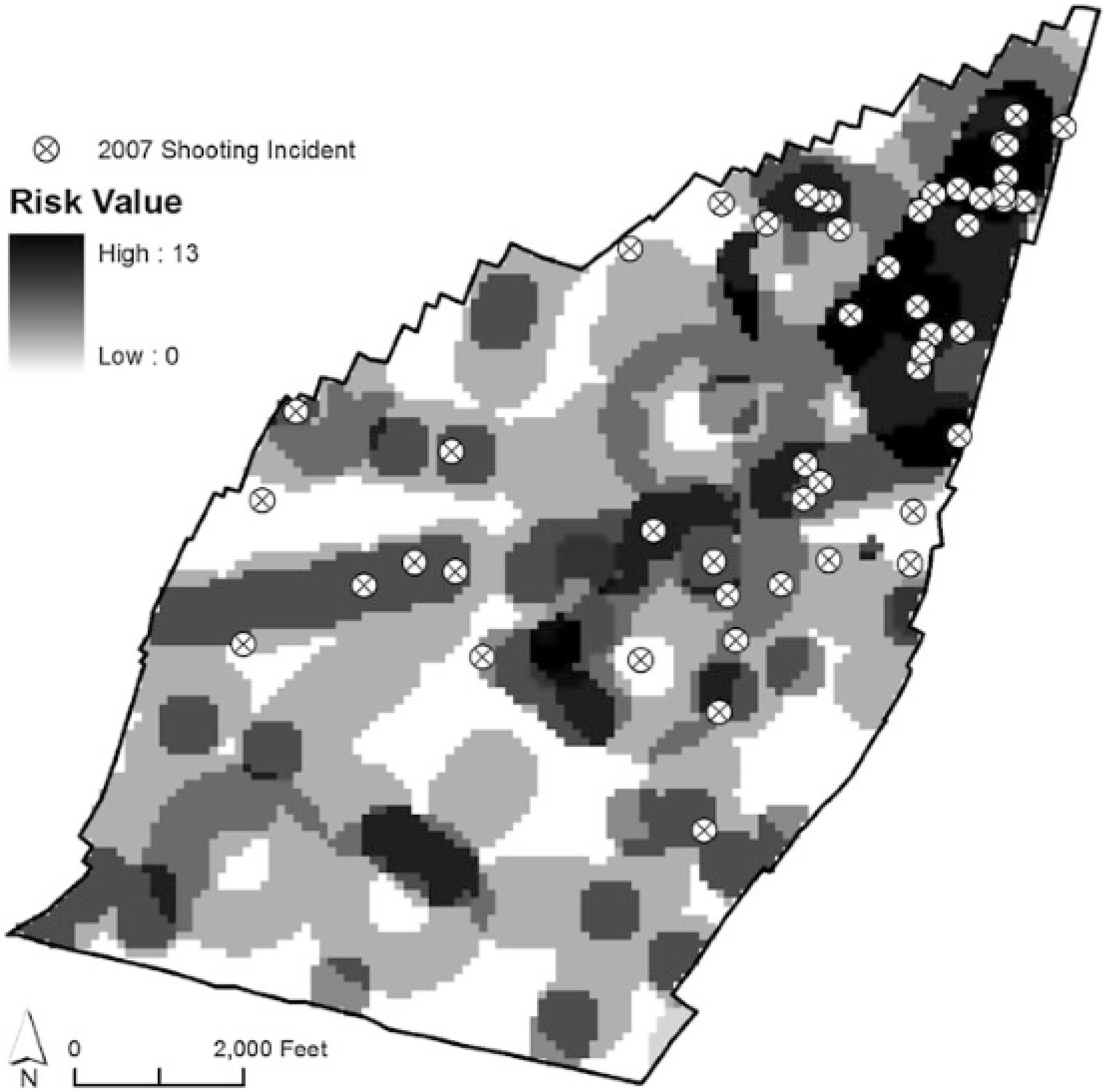


Figure 3.8: Weighted risk terrain of gun shootings (Caplan and Kennedy, 2014).

3.4 Analysis steps

This section discusses the analysis steps and modeling considerations needed to analyze the spatial and temporal patterns of burglaries in Haarlem. It includes all proposed steps, from data input and preparation to the actual modeling and the visualization of the results. All of this is summed up and presented in a flow diagram at the end of this chapter.

3.4.1 Identifying the potential risk factors

The very first step is to identify the potential risk factors of burglaries. This is based on a literature study (which is performed in chapter 4 where different theories and the possible risk factors are discussed in detail). The risk factors found in the literature study are operationalized to a grid of the study area.

3.4.2 Study area

The study area outlined in the introduction (see figure 1.5) has to be converted to spatial units that can be used for the regression analysis. This conversion is based on notions from grid and raster methods and uses a grid to aggregate the burglary incidents (Caplan and Kennedy, 2010).

The choice of grid cell size is a trade-off between accuracy and privacy. On the one hand, a high spatial resolution, meaning smaller spatial units, allows for more detailed spatial patterns to be modeled. But on the other hand, a high spatial resolution might interfere with the privacy of citizens. This is also one of the main concerns regarding the smart city concept. In the Netherlands, it is governmental policy to ensure that public data cannot be traced back to individuals. That is why, as a rule of thumb, governmental organizations typically aggregate this data over at least ten addresses. This means that the cell size cannot be too small as well.

Here 100 by 100 meter cells are considered to strike a balance between accuracy and privacy. At this spatial resolution it is expected that there will still be plenty of detail left in the spatial patterns to draw meaningful conclusions, while the number of addresses within this area is likely to be above ten.

The first step is to overlay the study area with a square grid with 100 meter cells. Cells that are completely outside the municipality of Haarlem are omitted. Also cells that contain less than ten addresses are removed (see figure 3.9). Then the data about the risk factors needs to be operationalized to the grid cells.

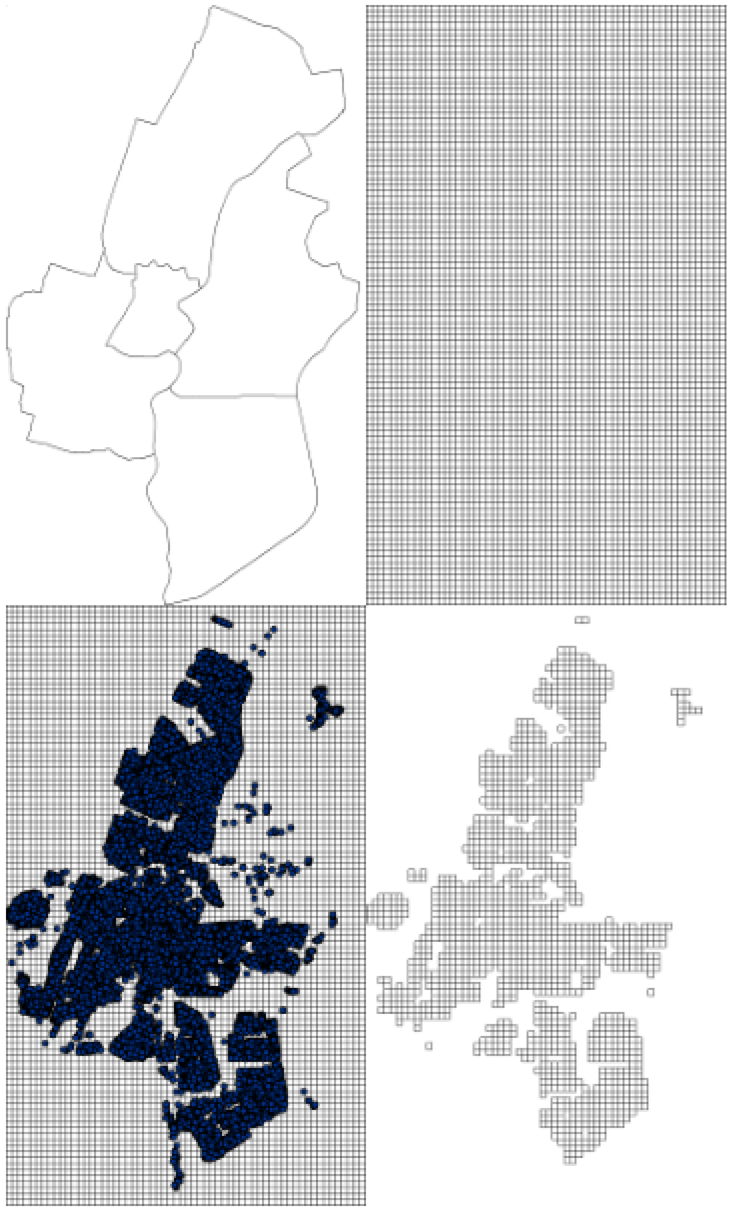


Figure 3.9: Operationalization of the study area. The first image (top left) shows the municipality and city districts. The second image (top right) shows the conversion to a grid with 100 meter square cells. The third image (bottom left) shows the grid with an overlay of all the residential addresses at the end of 2010 based on BAG data. Finally, the fourth image (bottom right) shows the final operational study area where the 1460 cells with at least ten residential addresses are included.

3.4.3 Model building

During the model building process, several possible issues have to be considered together with some modeling choices. Here, four model building considerations are discussed:

1. multicollinearity,
2. variable selection,
3. spatial autocorrelation, and
4. seasonality.

Multicollinearity

The first step of building the regression model is checking for 'multicollinearity' between the explanatory variables. Multicollinearity refers to correlation among explanatory variables that is caused when two variables are measures of the same phenomenon (de Vocht, 2008). This might happen with explanatory variables related to for example income: such as 'household income' and 'cars per household'. It is likely that households with more cars also have a higher income.

Multicollinearity should be avoided, because it causes problems for estimating the structural relationships through the use of regression techniques. *"Attempts to apply regression techniques to highly multicollinear independent variables generally result in parameter estimates that are markedly sensitive to changes in model specification and to sample coverage"* (Farrar and Glauber, 1967)(pp. 93–94).

Here, multicollinearity is tested by calculating the bivariate correlations between pairs of all independent variables. Multicollinearity is detected when Spearman's r, a correlation coefficient, of two independent variables is equal to or larger than 0.9, or $r \geq 0.9$. When this occurs, the variable with the least explanatory power is removed. Spearman's correlation coefficient is preferred here, because it does not assume linearity between the independent variables, where for example Pearson's correlation coefficient does (de Vocht, 2009). Because of the different nature of the many independent variables, linearity between all pairs of them is not assumed here.

Variable selection

It is likely that there are many risk factors involved in explaining spatiotemporal patterns of burglaries, resulting in many potential independent variables included in the regression model. Researchers often follow the statistical principle of parsimony, striving for a 'minimal adequate model' (Guthery et al., 2005; Whittingham et al., 2006). This is largely based on the classical scientific notion of 'Ockham's razor': *"a hypothesis with the lowest tally of assumptions is more likely to be true than alternative hypotheses"* (Guthery et al., 2005).

Having many explanatory variables in a model can introduce the risk of 'overfitting'. Overfitting occurs when a statistical model describes random error or noise instead of the underlying relationship. This could occur when the number of observations is too low relative to the number of explanatory variables (The University of Texas, 2012). Babyak (2004) notes that "*overfitting yields overly optimistic model results: "findings" that appear in an overfitted model don't really exist in the population and hence will not replicate*" (p. 411).

To come to a more parsimonious model, often some type of stepwise regression is applied: adding or removing explanatory variables until only the statistically significant variables remain (de Vocht, 2008). However, there are also downsides to using stepwise regression methods. Whittingham et al. (2006) for example state that the selected 'best' model is influenced by the algorithm used, the order of variable entry and the number of candidate variables. This introduces sensitivity in the regression model selection procedure. They further note that identifying a single best model can suggest a level of confidence in the final model that is not justified by the data, focusing all further analysis and reporting on that single model (Whittingham et al., 2006).

Flom and Cassell (2007) summarize the problems with stepwise methods: *"parameter estimates are likely to be too far away from zero, variance estimates for those parameter estimates are not correct either so confidence intervals and hypothesis tests will be wrong and there are no reasonable ways of correcting these problems"* (p. 1). They further describe the essential problem as applying methods intended for one test to many tests.

Flom and Cassell (2007) provide some alternatives to stepwise regression methods. One of these alternatives is to simply use a full model, which means leaving non-significant variables in the model. The idea here is that if theory suggests that variables will be significant, then a small and non-significant result is still of interest (Flom and Cassell, 2007). This is in accordance with the idea that research should guide the statistical analysis and not the other way around.

This study follows that idea. All included explanatory variables are derived from a study of literature, where a wide range of researchers assumed their significance. That is why no stepwise regression is applied here and all independent variables are deemed relevant. To prevent overfitting, there should be at least ten to fifteen observations per explanatory variable included in the regression model (Babyak, 2004).

Spatial autocorrelation

An issue that often occurs in spatial analysis is 'spatial autocorrelation'. Spatial autocorrelation is the *"coincidence of value similarity with local similarity"* (Anselin and Bera, 1998). Dormann et al. (2007) state that *"analysis of spatial data is complicated by a phenomenon known as spatial autocorrelation"* (p. 610). Also Helbich and Jokar Arsanjani (2014) note that *"the estimation of Poisson and negative binomial models (NBM) is complicated by spatial autocorrelation"* (p.1).

Spatial autocorrelation occurs when the values of variables sampled at nearby locations are not independent from each other (Dormann et al., 2007). This can cause issues because one of the assumptions of count regressions is residual independence (Helbich and Jokar Arsanjani, 2014). This means that the fitted model should perform equally well over space. But unfortunately, it often happens that the fitted model is underpredicting or overpredicting in certain areas. This can happen when for example some important spatial independent variable is not included in the model (Dormann et al., 2007).

Spatial correlation can be detected using a common statistical measure in spatial statistics: the Moran's I value. This value can be used to classify a spatial pattern as clustered, dispersed or random (see figure 3.10). Moran's I also includes a z-score and p-value to assess the significance (Esri, 2013).

When spatial autocorrelation is detected in the residuals of the fitted model, this has to be corrected for to come to reliable analysis results. If not corrected for, spatial autocorrelation can cause type 1 errors, meaning the detection of effects that are not present (Dormann et al., 2007). Helbich and Jokar Arsanjani (2014) add: *"standard errors and estimated coefficients may be biased as well as inconsistent, risking erroneous conclusions on the basis of a misspecified regression model"* (p. 1). There are several methods available for correcting for spatial autocorrelation.

A promising method is that of 'eigenvector spatial filtering'. Helbich and Jokar Arsanjani (2014) show how this spatial filter efficiently absorbs spatial autocorrelation from the variable's actual effect in linear and nonlinear negative binomial models, resulting in well-specified regressions that assure model assumptions. Eigenvector spatial filtering aims to extract eigenvectors from a transformed spatial neighborhood matrix, which describes the spatial arrangement and connectivity between entities of spatial systems (Helbich and Jokar Arsanjani, 2014)[3]. An important characteristic of eigenvector spatial filtering is that it is topology-based, which means that all of the spatial units should be adjacent to each other (Helbich and Jokar Arsanjani, 2014). The study area of Haarlem includes some 'island cells' (see figure 3.9) violating this pre-condition and making this spatial filtering technique unsuitable.

A more common alternative method for handling spatial autocorrelation is adding a spatially lagged dependent variable as an additional predictor to the spatial model. This variable is based on a spatial weights matrix which expresses for each observation (row) those locations (columns) that belong to its neighborhood set as nonzero elements (Anselin and Bera, 1998). Therefore, spatial weights are a measure of potential interaction between a pair of observations (Anselin and Bera, 1998). Which observations are considered as neighbors is arbitrary. Common methods are for example queen (adjacent units sharing an edge or node) and rook (adjacent units sharing an edge) contiguities (Helbich and Jokar Arsanjani, 2014). Here, based on near repeats theory, the neighbor-

[3] A more detailed description of this relatively complex method can be found in articles by Helbich and Jokar Arsanjani (2014) and Diniz-Filho and Bini (2005).

hood is defined as all observations within 400 meters, as this seems to be the maximum area of influence of a burglary event (Johnson, 2008). The values of neighbors are then multiplied with their spatial weights to calculate a spatial lag for each observation (see also figure 3.11). This methodology is also applied in the spatiotemporal analysis of burglaries in Haarlem.

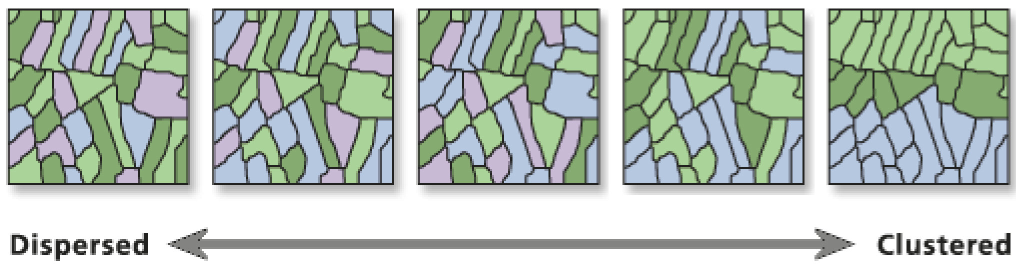
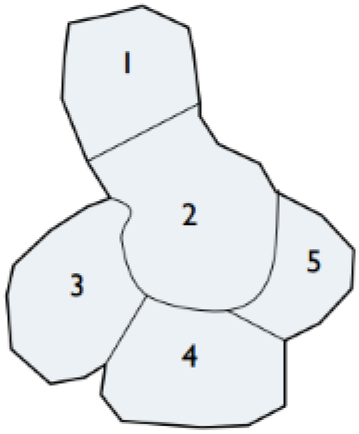


Figure 3.10: Moran's I classifies the spatial pattern that can be found in data from dispersed to random to clustered (source: `http://tinyurl.com/mtpmopj`).



$$W = \begin{matrix} 0 & 1 & 0 & 0 & 0 \\ .25 & 0 & .25 & .25 & .25 \\ 0 & .5 & 0 & .5 & 0 \\ 0 & .33 & .33 & 0 & .33 \\ 0 & .5 & 0 & .5 & 0 \end{matrix}$$

Area	y	Wy
1	5	(1*7)=7
2	7	(.25*5)+(.25*9)+(.25*12)+(.25*11)=9.25
3	9	(.5*7)+(.5*12)=9.5
4	12	(.33*7)+(.33*9)+(.33*11)=8.91
5	11	(.5*7)+(.5*12)=9.5

Figure 3.11: This figure shows how the spatial lag model works. The first image shows the different spatial units numbered 1 to 5. To the right of it is the spatial weights matrix. This shows the relationship between the spatial units. The weights decrease when there are multiple neighbors. The bottom table shows how a spatial lag variable is calculated based on the spatial weights and the neighboring values (source: `http://tinyurl.com/po9st47`).

Seasonality

The final modeling consideration is seasonality. Time variables like time of day, the day of the week and the season can have an influence on the burglary rate. This is related to the daily routines of people and the chance that they are at home within certain time intervals or on certain days (Caplan and Kennedy, 2011). As the focus of this study is on structural spatiotemporal patterns of burglaries, seasonality is the main time variable that is considered here. Many studies show that seasonality is a factor that should be recognized when analyzing crime in general and residential burglaries more specifically (Brown and Altman, 1983; Farrell and Pease, 1994; Yan, 2004; Caplan and Kennedy, 2011; Coupe and Blake, 2006; Perry, 2013; Sorensen, 2004).

A study by Sorensen (2004) shows that burglaries occur more often in the autumn and winter than in the spring and summer. He suggests that this temporal pattern is caused by *"a combination of temperature and hours of daylight since the warmer, lighter seasons are characterized by the informal surveillance of garden users, the difficulty of judging home occupancy on the basis of interior lighting, and the lack of cover of darkness"* (p. 13). It also turns out that there is a higher number of burglaries around Christmas, so holidays can be a risk factor as well.

Brown and Altman (1983) state that seasonal and lighting variations cause differences in visual accessibility, street activity and other factors that can influence the offender's judgment of vulnerability. But Perry (2013) suggests that burglary rates might increase in the summer when there is no school. Also Sorensen (2004) states that the summer can provide favorable settings for burglars, as doors and windows are more likely to be left open and unlocked.

So although it is not straightforward how seasonality exactly affects the burglary rate in an area, at least there seems to be consensus among criminologists that seasonality is a factor to be reckoned with. The goal is to find out whether subsets of burglaries for different seasons produce different spatial patterns and also different explanations for these patterns. In order to test this, the burglaries over the four years of analysis (2010-2013) are aggregated per season. The independent variables are also aggregated over the four years and then averaged. Based on these figures, four negative binomial regression models are fitted for each season. Based on these models, conclusions can be drawn about the effects of different seasons on the explanations for burglaries.

3.4.4 Risk terrain surface

The final step of the analysis is the interpretation and presentation of the results by developing a risk terrain surface of the study area.

For each season, the significant risk factors of burglaries are derived from the respective negative binomial regression model. Each of these risk factors are standardized to individual risk layers by identifying the 'areas of highest risk' based on that variable. This standardization is necessary to be able to combine the different risk layers to one risk surface. Risk terrain modeling uses a binary differentiation of the risk layer: areas with the highest risk are assigned a value of 1 and all other areas are valued 0 (Caplan and Kennedy, 2010).

This process is relatively straightforward for distance based risk factors, where being close to a certain risk factor increases or decreases the risk of burglary. Areas within a certain threshold are assigned a value of 1 and areas outside the threshold a value of 0 (or the other way around based on the direction of the relationship between the risk factor and the burglary rate).

Identifying high risk areas for risk layers based on the presence of a certain risk factor, for example the presence of a certain demographic group, is less straightforward. For this type of risk layers it is necessary to first create density maps based on the count or presence of a certain risk factor. Standard deviation is a statistical measure that can be used to identify the high risk areas based on the densities found (Caplan and Kennedy, 2010). For example, areas with a density of a certain risk factor of more than two standard deviations from the mean, i.e. the highest or lowest 2.5%, can be reclassified to value 1 and the other areas with value 0.

Here, a different classification is used. Instead of values of 0 and 1, risk factors are reclassified to values of 1, 2 and 4 to allow for more nuances of risk. This also avoids defining distance thresholds for distance-based risk factors, which is often an arbitrary process. The risk values are defined by using nested means classification on the data set of each individual risk layer.

Nested means classification is a general objective method for the calculation of class intervals for statistical maps, specifically aimed at constructing map classes for non-uniform distributions (Scripter, 1970). Scripter (1970) explains that nested means classification *"divides a numerical array into two classes and the means of each of these two map classes yield four map classes with smaller intervals. Repeating the process yields additional means and additional classes with smaller intervals"*.

The nested means classification is used here to divide each risk data layer into eight classes. The first four classes, which are all below the mean value, are assigned with risk value 1 which stands for 'low risk'. Risk value 2 is assigned to the first three classes above the mean value, meaning 'medium risk'. Finally, risk value 4 is assigned to the highest class. Value 4 is used to really elevate the 'high risk' areas from the other areas.

The resulting standardized risk layers are then combined using a weighted map overlay, weighing each layer based on the standardized beta coefficients that resulted from the regression analysis (figure 3.12).

Standardized beta coefficients account for the varying means and variances of the independent variables. Therefore, standardized beta coefficients enable the comparison of coefficients from different independent variables. The formula to calculate standardized beta coefficients from beta coefficients is:

$$\beta_1^* = \beta_1 \frac{S_{x1}}{S_y} \tag{3.1}$$

where β_1^* is the standardized regression coefficient, β_1 the beta coefficient, S_{x1} the standard deviation of the independent variable and S_y the standard deviation of the dependent variable (Bring, 1994).

The weighted map can help visualize the results and identify the areas with the highest risk of burglaries. To make it easier to interpret the resulting composite map, the risk cell values can be divided by the lowest risk cell value found. Through this method, risk values start with one and higher risk values represent how much more risk there is compared to the lowest value (Caplan et al., 2013). For example a value of five means that the risk of burglaries within that specific cell is five times higher than the minimum risk.

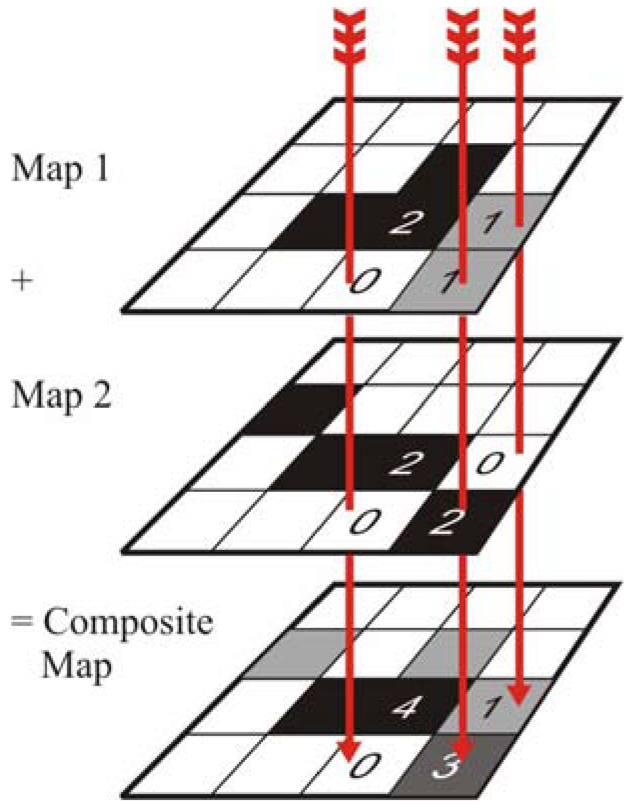


Figure 3.12: Combining different risk layers to produce a composite risk map (Caplan et al., 2013).

3.5 Conclusion

This chapter started with general methods of crime mapping: hot spots and near repeats, grid and raster methods and univariate and multivariate regression methods. A comparison of these methods with the selection criteria defined for this study showed that a multivariate regression method, complemented with the grid and raster method, is considered to be the most suitable method. It focuses on identifying the structural causes, it can take spatial effects into account, the process is transparent and the results are relatively easy to interpret and present.

From the multitude of available multivariate regression models, the negative binomial regression model is selected. This regression model acknowledges the specific nature of data of crime rates: skewed data with many zeros and only positive and integer values.

The final section presented the full methodology. It started with an operationalization of the study area and the input variables. After that, the actual regression model buiding process is discussed addressing issues of multicollinearity, variable selection, spatial autocorrelation and seasonality. The final step discussed was the presentation of the analysis results in one comprehensive burglary risk terrain map showing the areas with the highest risk of burglaries.

This complete process is schematically summarized in figure 3.13.

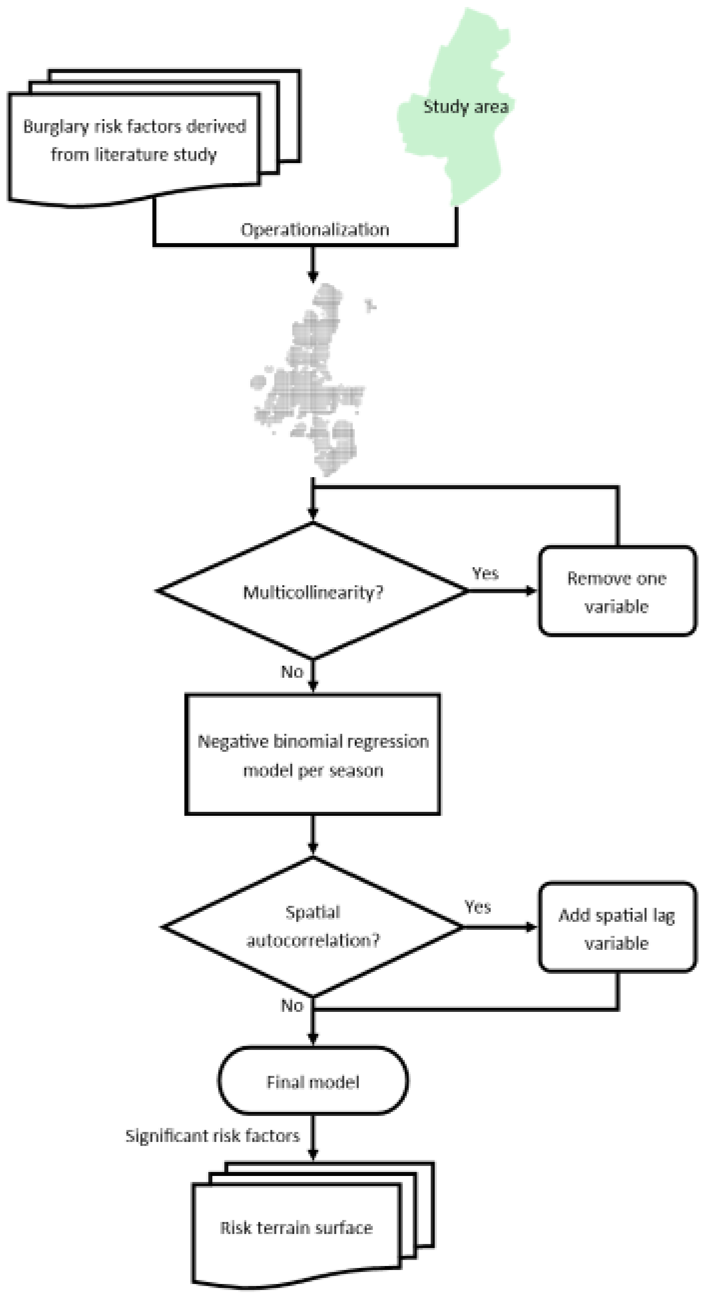


Figure 3.13: Schematic overview of the methodology for the spatiotemporal analysis of residential burglaries in Haarlem.

Chapter 4

Risk factors of residential burglaries

This chapter focuses on the first step in the spatiotemporal analysis of burglaries: identifying the risk factors. It aims at answering the research subquestion: *what are the possible risk factors for residential burglaries?* To find an answer to this question, theories from the field of spatial criminology are consulted: a scientific field that is concerned with the spatial and temporal patterns of crime (Pauwels et al., 2012). The possible risk factors of burglaries found in this literature study are used as input explanatory variables for the regression analysis.

Brantingham and Brantingham (1995) discern three components of crime. Crime events are most likely to occur when these components come together (see figure 4.1). The three components of crime are:

- potential targets,
- potential offenders, and
- settings.

The first component of crime consists of the potential targets. The focus is on where potential targets are located. When considering residential burglaries, the potential targets are the homes of people.

The second component, potential offenders, is about the potential locations of offenders. As offenders are mobile, not only their place of residence is relevant, but also the locations that are within their space of daily activities and even the routes that connect these locations.

The last component of crime is the setting. The setting determines whether a combination between a potential target and a potential offender results in the occurrence of crime. The setting component of crime includes risk factors that influence the likelihood that a burglary will be successful. Favorable settings for burglaries increase the risk of burglaries.

These three components are used to find auxiliary risk factors of burglaries. For each of the components, the relevant theories are discussed and the relevant risk factors are identified. The focus is on the spatial patterns of the associated risk factors. Especially areas where risk factors from different crime components overlap, are assumed to be at increased risk.

Figure 4.1: The three components of crime based on Brantingham and Brantingham (1995).

4.1 Target

The theories and risk factors discussed here are related to the characteristics of dwellings and their residents: the potential targets. Two relevant theories are presented here: the 'rational choice' and the 'optimal foraging' theory. Based on these theories, the potential risk factors are identified.

4.1.1 Target theories

Rational choice

Within the discipline of criminology there traditionally has been a distinction between two types of offenders: the rational and the irrational offender (Walsh, 1986). The goal of the crime and the preparations involved are the main differences between these types of offender. For example, armed robbers, terrorists and in some cases burglars too, are considered as rational. On the other hand, drunks, vandals, shoplifters and joyriders are considered irrational. There is an implicit distinction between ‘professionals’ and ‘thrill seekers’. Professionals typically have a certain predefined goal, where thrill seekers act on opportunity and instinct.

Burglars can be related to both types of offenders depending on the level of organization and preparation of the burglary. The rational burglar commits organized burglaries, where the irrational burglar commits opportunity burglaries (Klein Haneveld et al., 2012). These types of burglary have different characteristics: burglars use different methods, choose different types of dwellings as their target and choose different kinds of neighborhoods. For example, organized burglaries occur more often in wealthy neighborhoods and near the edges of neighborhoods, allowing for a fast getaway. Opportunity burglaries are typically committed by local offenders, who have first-hand knowledge of the neighborhood and its characteristics. Klein Haneveld et al. (2012) state that research indicates that only 7% of burglaries are ‘pure’ opportunity burglaries and that most burglaries are planned to some degree.

The rational choice theory can be applied to the rational type of burglars. This theory assumes that the decision process is guided by the wish to maximize the goal. With regard to burglaries, this translates to maximizing the return of a burglary together with a minimization of the risk of getting caught (Bernasco, 2010; Chainey and Ratcliffe, 2005; Gale, 2013). The risk of getting caught is largely determined by the setting of a potential target, for example the level of social control in a neighborhood. This setting component of crime is discussed in section 4.3. Here, the focus is on the maximization aspect of the rational choice theory.

Optimal Foraging

The optimal foraging theory is similar to the rational choice theory. The optimal foraging theory is based on observations from wildlife. When predators choose their prey, they instinctively weigh the potential nutritive value against the efforts and risks of finding, attacking and eating the prey. The hypothesis is that the animal’s pattern of choice of food type is based on maximizing the net rate of energy intake (Pyke et al., 1977).

Burglars are assumed to behave in a similar fashion. They choose those targets where the expected gains outweigh the perceived risks (Bernasco, 2010). Based on the theories of rational choice and optimal foraging, several potential risk factors can be deduced which influence the attractiveness of a target.

4.1.2 Target risk factors

From the viewpoint of maximization several variables can be identified that act as indicators for the expected returns of a burglary (for a summary of the risk sectors see table 4.1 at the end of this section). These variables are based on a review of literature. Many of these variables are directly or indirectly related to the income of a household.

The potential target risk factors are introduced below. A distinction is made between variables that are related to individuals and variables that are related to objects, i.e. dwellings.

Person-related factors

The average **household income** is an obvious risk factor. When the household income is higher, it is more likely that these households own more expensive goods (Bernasco, 2006; Liu and Brown, 2003; Malczewski and Poetz, 2005). This assumption is also at the basis of most of the potential target risk factors.

Another risk factor is **unemployment**. The assumption here is that employment figures can give an indication for the average income level in an area: areas with high unemployment have a lower average income. From this assumption follows that areas with lower unemployment run a higher risk of burglaries (Deadman, 2003).

Level of education can also be seen as an indication of income level. The assumption is that a higher average level of education corresponds with a higher average household income in an area (Chainey and Ratcliffe, 2005; Wilsem et al., 2006). Likewise, the level of **consumption** and the number of **cars per household** are indicators of wealth in an area too (Deadman, 2003; Liu and Brown, 2003; Bernasco, 2006).

Object-related factors

The average **property value** in an area is an obvious indicator of wealth. Many studies have identified property value as a potential explanatory variable of burglaries (Bernasco and Luykx, 2003; Bernasco and Nieuwbeerta, 2005; Bernasco, 2006, 2010; Malczewski and Poetz, 2005).

Several indicators are available that are directly linked with the property value. These indicators are the **property size** and the share of **detached properties** (Bernasco, 2006; Wilsem et al., 2006). Larger properties and detached properties usually have higher values.

Another indicator of average income is **homeownership**. Homeownership is often expressed as the percentage of homeowners in an area. The assumption made is that a higher percentage of homeownership indicates a wealthier neighborhood making it a more suitable target for the rational burglar (Bernasco and Luykx, 2003; Bernasco, 2006; Groff and La Vigne, 2001; Malczewski and Poetz, 2005; Wilsem et al., 2006).

Finally, a relatively high number of dwellings within a spatial unit increases the likelihood of finding a suitable dwelling from the offender's perspective (Bernasco and Nieuwbeerta, 2005; Perry, 2013). Therefore, a higher **building density** is considered a risk factor.

Target risk factors
Person-related
Household income
Unemployment
Level of education
Consumption
Cars per household
Object-related
Property value
Property size
Detached properties
Homeownership
Building density

Table 4.1: List summarizing the target risk factors of residential burglaries based on a study of literature.

4.2 Offender

Potential offenders are the second component of crime. Two relevant theories from the spatial criminology discipline are discussed: the 'awareness space' theory and the 'offender neighborhood' theory. After that, the associated potential risk factors are identified.

4.2.1 Offender theories

Awareness space

The concept of awareness space is very common in the spatial criminology literature and its origins can be traced back to research by Brantingham and Brantingham (Brantingham and Brantingham, 1993; Bernasco and Nieuwbeerta, 2005; Caplan and Kennedy, 2011; Moreto et al., 2013; Nubani and Wineman, 2005; Pauwels et al., 2012; Taylor, 2003). Awareness space is the familiar physical space where people perform their daily activities. This space is also referred to as the 'routine activity space' and it includes places and routes that are familiar to a subject. The awareness space differs per individual, changing for example with the age of people.

The awareness space includes nodes, a number of central locations, and paths, the routes connecting the nodes (see figure 4.2). The nodes can be someone's home, shopping areas, city centers, work, school, sporting facilities, parks and recreation centers. Paths can be the connecting roads or public transport connections.

The concept of awareness space can be used to identify potential risk factors of crime and burglaries. The central hypothesis is that offenders are more likely to commit burglaries in places that are familiar to them. There are two arguments to support this assumption.

First, familiar areas are preferable because offenders are less likely to be considered as strangers. Thus, offenders are less likely to stand out and draw attention.

Second, offenders have better knowledge of the physical infrastructure in a familiar area. This increases the likelihood that offenders can enter and leave their target without being spotted.

The preference of offenders for familiar and nearby areas has been shown by different studies. Researchers have observed a distance decay pattern: the risk of burglaries decreases with the increasing distance from nodes and paths within the awareness space of offenders (Bernasco and Nieuwbeerta, 2005).

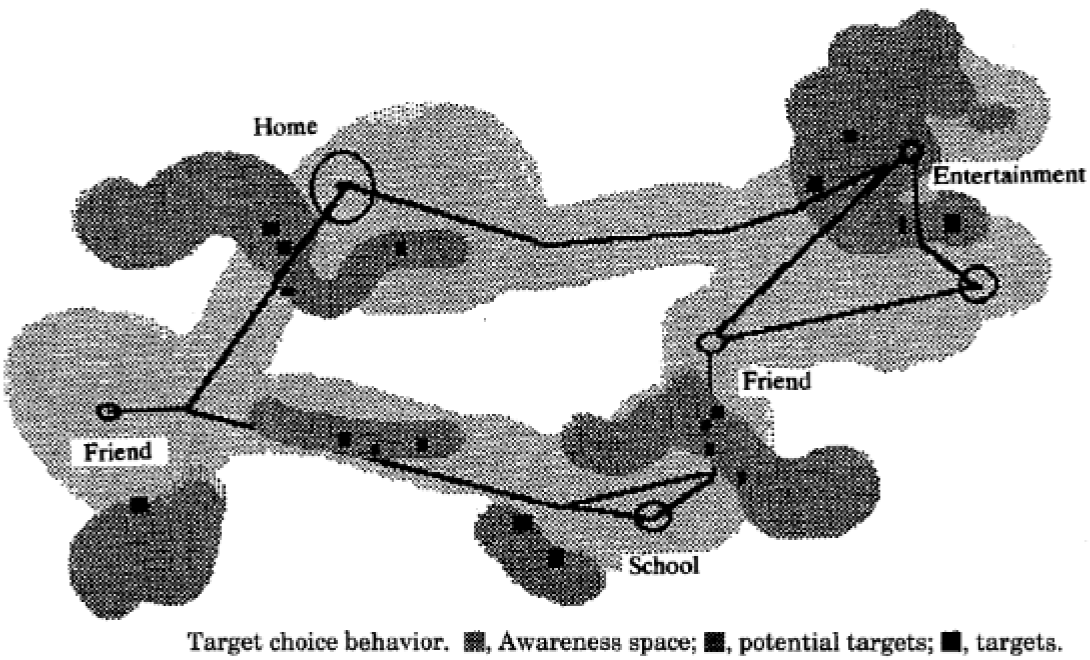


Figure 4.2: Visualization of the awareness space concept (Brantingham and Brantingham, 1993).

Offender neighborhood

For the most accurate spatial analysis of burglaries, the place of residence of convicted offenders is important data, especially from the viewpoint of awareness space theory as the place of residence of an offender is a crucial node. Unfortunately, these data are very privacy-sensitive and are therefore hard to obtain. The theory of offender neighborhoods can be used as an alternative by providing a characterization of the living area of potential offenders.

The theory of offender neighborhoods has its roots in the social-ecological approach on crime, based on the Chicago school. Pauwels et al. (2012) describe the social-ecological approach to crime as *"the idea that the environment where people live plays an independent role in the arise and continuation of crime in that location"* (p. 289). This idea suggests the analysis of neighborhoods as the place of residence of potential offenders, which can be used as a node from the awareness space theory. That is why the offender neighborhood theory analyses the neighborhood characteristics that can contribute to crime.

4.2.2 Offender risk factors

The potential offender risk factors that can be deduced from theory are discussed below. These factors are grouped by the theory that they are based on (for a summary of the risk sectors see table 4.2 at the end of this section).

Awareness space

From the viewpoint of the awareness space theory, the **place of residence of an offender** is an obvious and important node. This factor is mentioned in different scientific publications as an important explanatory factor for burglaries (Bernasco and Luykx, 2003; Bernasco and Nieuwbeerta, 2005; Bernasco, 2006; Brantingham and Brantingham, 1993; Groff and La Vigne, 2001; Nubani and Wineman, 2005).

The **city center** can be considered as an area that is very likely to be within the awareness space of many people, including those of potential offenders, because of the many facilities that can be found there. Based on this assumption, residential areas within or near city centers run a higher risk of being burgled (Bernasco and Luykx, 2003; Bernasco and Nieuwbeerta, 2005; Bernasco, 2006; Brantingham and Brantingham, 1993).

Also **public facilities**, like libraries or hospitals, and foodservice outlets, like **restaurants and bars**, are likely to fall within the awareness space of many people and are therefore a possible activity node (Bernasco, 2006; Brantingham and Brantingham, 1993).

Pawn shops are potential locations where burglars can sell their stolen goods and are therefore considered a potential activity node too (Caplan and Kennedy, 2011; Moreto et al., 2013).

Public transport stops are also likely activity nodes. Likewise, the **public transportation routes**, for example bus routes or railways, can be considered activity paths. The areas in the vicinity of these nodes and paths are within the awareness space of offenders (Moreto et al., 2013; Brantingham and Brantingham, 1993).

The **road network** is the most common way of creating paths between nodes. Certain types of roads can therefore be considered as risk factors, i.e. the main roads or the most 'integrated' roads in a network, because these roads are most likely to be within the awareness space of potential offenders (Brantingham and Brantingham, 1993; Groff and La Vigne, 2001; Moreto et al., 2013).

Offender neighborhood

Offender neighborhoods are often characterized by a higher percentage of low income households (Malczewski and Poetz, 2005; Moreto et al., 2013; Wilsem et al., 2006). It is assumed that people with lower incomes have relatively much to win by committing property crimes (Ehrlich, 1975). Therefore, the **household income** is not only relevant as a target risk factor, but also as offender risk factor. Where high income areas are considered as potential target, low income areas can be considered as the potential place of residence of offenders.

Just like low incomes, high levels of **unemployment** and lower levels of **education** are characteristics of offender neighborhoods (Malczewski and Poetz, 2005; Deadman, 2003). In addition, students with higher rates of **truancy** and lower levels of academic achievement are more likely to commit crimes (Weisburd et al., 2009).

Another research in north-west England showed that the most disadvantaged neighborhoods were characterized by a higher than average number of **lone-parent households** and high **ethnic heterogeneity** (Bowers, 1999; Malczewski and Poetz, 2005). That is why these characteristics are also linked to offender neighborhoods and considered as potential risk factors for burglaries.

Research also shows that burglars often share the same demographic characteristics. This **demographic risk group** consists of males with ages between 15 and 24 (Deadman, 2003; Wilsem et al., 2006). Areas with high percentages of this demographic group can be considered as areas where potential offenders live.

Offender risk factors
Awareness space
Place of residence
City center
Public facilities and restaurants/bars
Pawn shops
Public transport stops and routes
Road network
Offender neighborhood
Household income
Unemployment
Level of education
Truancy
Demographic risk group
Lone-parent households
Ethnic heterogeneity

Table 4.2: List summarizing the offender risk factors of residential burglaries based on a study of literature.

4.3 Setting

Setting is the final component of crime. Again, multiple spatial criminology theories are discussed that can help in finding relevant setting-related potential risk factors for burglaries. The theories discussed here are the 'social disorganization' theory and 'environmental design' theory. After this, the resulting potential setting risk factors are introduced.

4.3.1 Setting theories

Social disorganization

Just like the offender neighborhood theory, the social disorganization theory has its roots in the social-ecological approach based on the Chicago School. Sampson and Groves (1989) describe social disorganization as: *"the inability of a neighborhood to achieve the common goals of its residents and maintain effective social controls."* They continue that: *"Empirically, the structural dimensions of community social disorganization can be measured in terms of the prevalence and interdependence of social networks in a community, both informal (e.g. friendship ties) and formal (e.g. organizational participation), and in the span of collective supervision that the community directs toward local problems"* (p. 777).

With respect to crime, social disorganization refers to the ability of a community, or the lack of it, to supervise and control its own members (Kawachi et al., 1999). Social disorganization results in a lack of social cohesion and social control in a neighborhood or community, thus forming a risk factor for crime and burglaries (Bernasco and Nieuwbeerta, 2005; Bernasco, 2006; Caplan and Kennedy, 2011). Inhabitants experience less of a connection with the environment they live in and with their neighbors. This causes inhabitants to feel less responsible for their living environment.

The 'advantages' of social disorganization to potential burglars are that inhabitants are less likely to recognize strangers in their neighborhood and to be alarmed by suspicious activities. And if they are alarmed, they are less likely to act on it by for example calling the police. Thus, social disorganization can potentially result in serious setting-related risk factors for burglaries.

Environmental design

Environmental design theories assume that the spatial layout and design of the living environment contribute to the risk of crime (Bernasco, 2006; Brown and Altman, 1983; Cozens et al., 2005; Nubani and Wineman, 2005). These theories can be traced back to the iconic work of Jane Jacobs, who emphasized that the design of the public space influences the social control of and responsibility over the public space by citizens (Jacobs, 1961).

Closely related to this work is the design of 'defensible spaces' promoted by the architect and city planner Oscar Newman. Defensible spaces allow residents to survey their territory and allow clear articulation of the boundaries between public and private regions. This should ensure the residents' latent territoriality and sense of community (Brown and Altman, 1983). Therefore, environmental design theories are closely related to social disorganization theories, emphasizing the physical environment as a factor influencing the social structure of a neighborhood.

Many of the theories related to environmental design can be brought together under the concept of 'Crime Prevention Through Environmental Design' (CPTED). CPTED is based on the notion that *"the proper design and effective use of the built environment can lead to a reduction in the fear and incidence of crime, and an improvement in the quality of life"* (Crowe and Zahm, 1994).

CPTED is characterized by a total of six components: territoriality, surveillance, access control, activity support, image/maintenance and target hardening (Cozens et al., 2005) (see figure 4.3). The most relevant components are introduced below: territoriality, surveillance and target hardening.

- *Territoriality*

  Territoriality is a design concept for public space where the goal is to evoke a sense of responsibility with the legitimate users of a space, while preventing illegitimate users from committing offenses. The main idea is to use physical attributes to separate public, public-private and private space, to define ownership and to define acceptable patterns of usage.

  Different forms of territoriality include 'symbolic barriers' and 'real barriers' (Cozens et al., 2005). Symbolic barriers are aimed at conveying a sense of ownership and territoriality, separating private from public space in more implicit ways. Examples are gardens, hedges and even 'welcome doormats'. Real barriers limit access to a space through more explicit ways by using for example fences or differences in elevation.

- *Surveillance*
  Surveillance points to the physical design of public space to promote the informal or natural surveillance opportunities for residents and to stimulate guardianship over a space. When potential offenders feel like they are being watched, even when this is not the case, it is less likely that they will commit crimes.

  Three different types of surveillance are discerned: 'natural surveillance', for example self-surveillance opportunities provided by windows; 'formal surveillance', for example police patrols; and 'mechanical surveillance', for example by CCTV and street lighting (Cozens et al., 2005).

  The level of natural surveillance, the number of 'eyes on the street', is largely determined by the design of the public space and in particular the road network. The level of integration of street segments in the road network of an area can be an important factor (Cozens et al., 2005; Nubani and Wineman, 2005).

  One method for analyzing road networks is called 'space syntax'. Space syntax is a collection of theories about the social use of space developed in the late sixties by Hillier and Hanson (Nubani and Wineman, 2005). The space syntax variable 'integration' is used to define the level of accessibility of street segments from all other street segments within a spatial system. Integration measures how many turns have to be made from a street segment to reach all other street segments in the network, using the shortest paths (Jiang and Claramunt, 2002). Higher levels of integration will increase activity in an area, which will increase the opportunities for natural surveillance, ultimately leading to less burglaries (Cozens et al., 2005). On the other side however, integrated roads are more likely to be in the awareness space of offenders, potentially increasing the risk of burglaries.

- *Target hardening*
  The last relevant component of CPTED is target hardening. Target hardening refers to increasing the efforts that offenders must expend in the commission of a crime (Cozens et al., 2005; Nubani and Wineman, 2005). This can be seen as the most traditional approach towards crime prevention (Cozens et al., 2005). Examples of target hardening are reinforced locks, windows and doors.

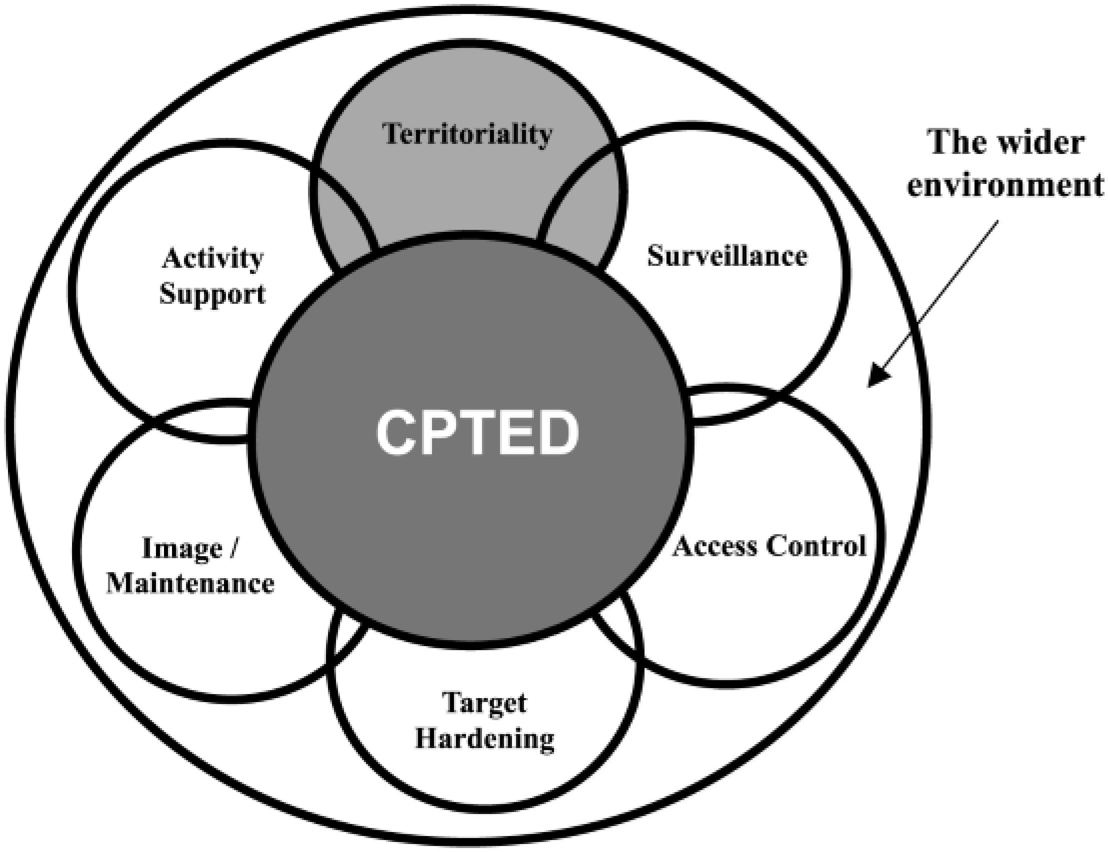


Figure 4.3: 'Crime Prevention Through Environmental Design' and its components (Cozens et al., 2005).

4.3.2 Setting risk factors

Based on the theories of social disorganization and environmental design, multiple potential setting risk factors can be identified. These are discussed below (for a summary of the risk sectors see table 4.3 at the end of this section).

Social disorganization factors

One of the variables that is often mentioned with respect to social disorganization is **ethnic heterogeneity** (Bernasco and Luykx, 2003; Bernasco and Nieuwbeerta, 2005; Bernasco, 2006; Caplan and Kennedy, 2011; Malczewski and Poetz, 2005; Weisburd et al., 2009; Wilsem et al., 2006). The main assumption is that a higher level of ethnic heterogeneity in a neighborhood increases the number of inhabitants who are not socially integrated. This can increase the level of anonymity, which can decrease social cohesion and social control and thus ultimately lead to more favorable settings for offenders.

Another essential characteristic of social disorganization is **residential mobility** (Bernasco and Luykx, 2003; Bernasco and Nieuwbeerta, 2005; Bernasco, 2006; Caplan and Kennedy, 2011; Malczewski and Poetz, 2005; Wilsem et al., 2006). It is assumed that a high residential mobility hinders the social cohesion in and the emotional connection with the neighborhood. Therefore, inhabitants feel less responsible for their environment and do not have a sense of guardianship.

Likewise, inhabitants of neighborhoods with lower percentages of **homeownership** also have less ties with their local environment and express less guardianship. The underlying assumption is that homeowners are more likely to invest in their living environment because of the direct relationship with their property value (Groff and La Vigne, 2001; Malczewski and Poetz, 2005; Weisburd et al., 2009; Wilsem et al., 2006).

Election turnout can be seen as a further indication of the local involvement of residents (Weisburd et al., 2009). Also higher **truancy** rates are considered a characteristic of socially disorganized areas (Weisburd et al., 2009). And finally, a higher number of reports of **nuisance** and high **crime rates** can be an indication of lower levels of social control and local guardianship (Groff and La Vigne, 2001; Perry, 2013; Bernasco, 2006).

Environmental design factors

Based on the territoriality design concept, **symbolic barriers** are used to express ownership and to separate private from public space (Cozens et al., 2005). The presence of these barriers might decrease the risk of burglaries. Furthermore, the presence of **physical barriers**, like locks, fences or gates, and **security**, like CCTV or security personnel, also reduce the possibilities for offenders (Brown and Altman, 1983; Johnson, 2001). These physical barriers can be related to the territoriality, surveillance and target hardening elements of CPTED.

Many potential risk factors are associated with the possibilities for surveillance in an area based on the environmental design. For example, **back alleys** are an element of the built environment that are favorable to potential offenders. Back alleys, especially those with dead ends, are likely to be poorly surveyed and moreover, they increase the number of getaway options for offenders (Bernasco, 2006; Chih-Feng et al., 2000).

Likewise, lack of **street lighting** can decrease the risk of being spotted (Bernasco, 2006; Groff and La Vigne, 2001; Nubani and Wineman, 2005). Surveillance possibilities are also lower in areas with **mixed land uses**, because of a lower residential population. This again can lead to higher risks of crime (Hillier and Sahbaz, 2007).

Dwellings that are less visible from the street run a higher risk of being burgled. The visibility of dwellings can be affected by objects such as hedges, fences or trees (Bernasco, 2006; Coupe and Blake, 2006). Therefore, the presence and size of these **objects affecting visibility** and their distance to potential targets could be considered.

Burglary rates vary over different **dwelling types**. It matters whether a dwelling is on a corner lot, is a semi-detached unit or is a detached unit. For example, corner lots are more susceptible to burglaries, because offenders believe that the chance of being observed by neighbors is smaller (Groff and La Vigne, 2001; Brantingham and Brantingham, 1993). Hillier and Sahbaz (2007) show with their space syntax based research that the risk of burglaries increases with the number of sides of a dwelling exposed.

Furthermore, areas that are more integrated within the road network, and thus have a higher **spatial integration**, can experience a lower risk of burglaries because of the higher level of surveillance over the area.

Areas around the **edges of homogeneous residential areas** run a higher risk of burglaries. This is mainly because offenders get less comfortable towards the centers of neighborhoods, as they are more likely to be considered strangers there (Brantingham and Brantingham, 1993). In addition, accessibility of edge areas is generally better than in the center of neighborhoods.

And finally, **police** surveillance activities can influence the spatiotemporal patterns of residential burglaries. The presence of and distance to police stations is a factor that can potentially reduce burglaries. The assumption is that less burglaries occur nearer to police stations. Offenders take the presence of police units and their response time into consideration (Caplan and Kennedy, 2011; Chainey and Ratcliffe, 2005; Deadman, 2003; Weisburd et al., 2009).

Setting risk factors
Social disorganization
Ethnic heterogeneity
Residential mobility
Homeownership
Election turnout
Truancy
Nuisance
Crime
Environmental design
Symbolic barriers
Physical barriers
Security
Back alleys
Street lighting
Mixed land use
Objects affecting visibility
Dwelling type
Spatial integration
Edges of homogeneous residential areas
Police

Table 4.3: List summarizing the setting risk factors of residential burglaries based on a study of literature.

4.4 Conclusion

This chapter introduced the potential risk factors for residential burglaries. These risk factors are categorized by the three components of crime: target, offender and setting. For each of these components the most relevant theories were discussed to provide a solid basis for the resulting risk factors.

It must be noted that not all risk factors that are identified can be operationalized. The necessary data is not always available, for example because data is only available on a higher level of aggregation, or the data could not be obtained, for example due to privacy concerns (addresses of convicted burglars).

Furthermore, distance related risk factors are operationalized based on euclidean distances, the distance 'as the crow flies', or on network distances, the distance following the road network. The calculation of network distances is much harder, as the distance of each cell to any other cell has to be calculated which is very time consuming. That is why network distance is only calculated for risk factors that include a limited number of end or starting nodes, like police stations.

Most of the required data is available on the spatial level of individual addresses. These data have to be aggregated to the grid cells of the study area. But in some cases, data are only available on a higher spatial level, i.e. neighborhoods. This means that the neighborhood data have to be assigned to the underlying grid cells, smoothing out spatial patterns between cells of the same neighborhood. This issue is referred to as 'ecological fallacy'. Heywood et al. (2006) states that the problem of ecological fallacy occurs *"when it is inferred that data for areas under study can be applied to the individuals within those areas"* (p. 193). While the problem of ecological fallacy is acknowledged here, it is decided to still include neighborhood data, as the focus of this research is more on finding a suitable method for the spatiotemporal analysis of burglaries than on the results: using these data is considered to be a better option than to not have any data at all.

Table 4.4 provides an overview of the risk factors that are operationalized based on the outcomes of the literature study and are used as potential explanatory variables in the regression analysis. The overview includes the most essential characteristics: name, description, type (target, offender or setting) and source of the data. More detailed information about for example the operationalization process and the underlying assumptions can be found in appendix A. More information about the data sources is included in appendix B.

Table 4.4: The input variables.

Name	*Description*	*Type*	*Source*
Dependent variable			
Burglaries	The number of recorded residential burglaries per 1000 residential addresses.	Dependent	MOR
Independent variables			
Household income	Average household income in euros.	Target	Neighborhood statistics
Welfare benefits	Population between 18 and 65 years old receiving social welfare as percentage of the complete population.	Target	Neighborhood statistics
Cars per household	Average number of cars per household.	Target	Neighborhood statistics
Property value	Average property value in euros.	Target	WOZ
Risky properties	Detached, semi-detached or corner properties as percentage of the complete housing stock.	Target and setting	BAG

Table 4.4: The input variables.

Name	*Description*	*Type*	*Source*
Rental properties	Rented properties as percentage of the complete housing stock.	Target and setting	WOZ
Building density	Number of residential addresses.	Target	BAG
Distance city center	Average euclidean distance in meters from dwellings to the city center.	Offender	BAG
Distance public facilities	Average euclidean distance in meters from dwellings to the nearest public facility.	Offender	BAG
Distance retail and catering	Average euclidean distance in meters from dwellings to the nearest cell with high number of retail and catering establishments.	Offender	BAG
Distance public transport node	Average euclidean distance in meters from dwellings to the nearest public transport nodes.	Offender	9292
Distance highway entry	Network distance in meters from cells to the nearest highway entry.	Offender	NWB
Accessibility	Average integration value based on space syntax analysis of road network.	Offender and setting	NWB
Distance low incomes	Average euclidean distance in meters from dwellings to the nearest cell with a high number of low incomes.	Offender	Neighborhood statistics
Distance welfare benefits	Average euclidean distance in meters from dwellings to the nearest cell with a high percentage of people receiving welfare benefits	Offender	Neighborhood statistics

Table 4.4: The input variables.

Name	*Description*	*Type*	*Source*
Distance demographic risk group	Average euclidean distance in meters from addresses to the nearest cell with a high number of males from 15 to 24 years old.	Offender	GBA
Distance ethnic heterogeneity	Average euclidean distance in meters from addresses to the nearest cell with high ethnic heterogeneity.	Offender	GBA
Distance rental properties	Average euclidean distance in meters from addresses to the nearest cell with a high number of rental properties.	Offender	WOZ
Ethnic heterogeneity	Population not born in The Netherlands as a percentage of the total population.	Setting	GBA
Residential mobility	Average number of years of residence at same address.	Setting	Neighborhood statistics
Election turnout	Voters for municipal elections of 2010 as percentage of eligible voters.	Setting	Neighborhood statistics
Nuisance	Number of reports of nuisance as percentage of total number of addresses.	Setting	MOR
Crime	Number of reports of crime (not burglaries) as percentage of total number of addresses.	Setting	MOR
Construction year	Average year of construction of residential objects.	Setting	BAG
Distance street lighting	Average euclidean distance of dwellings to nearest street light in meters.	Setting	BOR
Mixed land use	Residential addresses as percentage of total number of addresses.	Setting	BAG
Distance shrubbery	Average euclidean distance in meters from dwellings to the nearest patch of shrubbery.	Setting	BOR

Table 4.4: The input variables.

Name	*Description*	*Type*	*Source*
Distance to street	Average euclidean distance in meters from dwellings to the nearest street.	Setting	NWB
Edge dwellings	Addresses of dwellings at the edge of homogeneous residential areas as percentage of total number of addresses.	Setting	BAG
Distance police station	Network distance in meters from cells to the nearest police station.	Setting	BRT

Chapter 5

Burglary analysis results

This chapter focuses on the process and results of the spatiotemporal analysis of burglaries, answering the last research subquestion: *what are the results of the spatiotemporal analysis of residential burglaries?* The results of the spatiotemporal analysis of burglaries in Haarlem are presented, of which the proposed method was outlined in chapter 3 and the included potential risk factors in chapter 4.

The first section demonstrates the steps of the analysis in practice, making a distinction between the seasons. The aim is to find the significant risk factors explaining the spatial patterns, and based on these significant risk factors and their beta coefficients, to create a risk terrain surface.

The second section focuses on the interpretation of the results found in the first section. Are the significant relationships found between the independent variables and the dependent variable as was expected based on theory? The differences between the seasons are discussed as well.

5.1 Spatiotemporal analysis of burglaries

The spatiotemporal analysis of burglaries is performed following the steps outlined in chapter 3. For each of the seasons three main subjects are covered.

First, a quick overview of the situation is presented via maps showing the burglary rates and the hot and cold spots in the study area.

Second, the model building process is presented, covering potential issues like multicollinearity and spatial autocorrelation. To ensure that there are enough outcome events (i.e. burglaries) to prevent overfitting, the data is aggregated per season over the years 2010, 2011, 2012 and 2013. The independent variables are also aggregated over these four years and then averaged. As a result the final model is presented, showing for each season which risk factors significantly affect burglary counts, how much and in which direction. As the modeling steps are the same for all the individual seasons, these steps are only discussed in high detail for the first season: winter. For the following seasons, the modeling steps are discussed in less detail and the focus is more shifted towards the outcomes.

Third, the results are visually presented by creating a risk terrain surface. It shows the most risky areas per season based on the significant risk factors for that season. A hot spot map of the risk terrain surface is then created and compared to the initial hot spot map of actual counts of burglaries to assess the general accuracy of the risk terrain surface and to identify potential areas where burglary rates might increase in the future.

5.1.1 Burglaries during winter

Situation

878 burglaries occurred during the winter, which is 30% of all burglaries from 2010 to 2013. These burglaries occurred most often in the south-east districts of Haarlem: in the *Schalkwijk* district (see figure 5.1). The concentration of burglaries in the south-east is also clearly visible from the results of a hot spot analysis (see figure 5.2). Cold spots can be found in the North, East and Center districts.

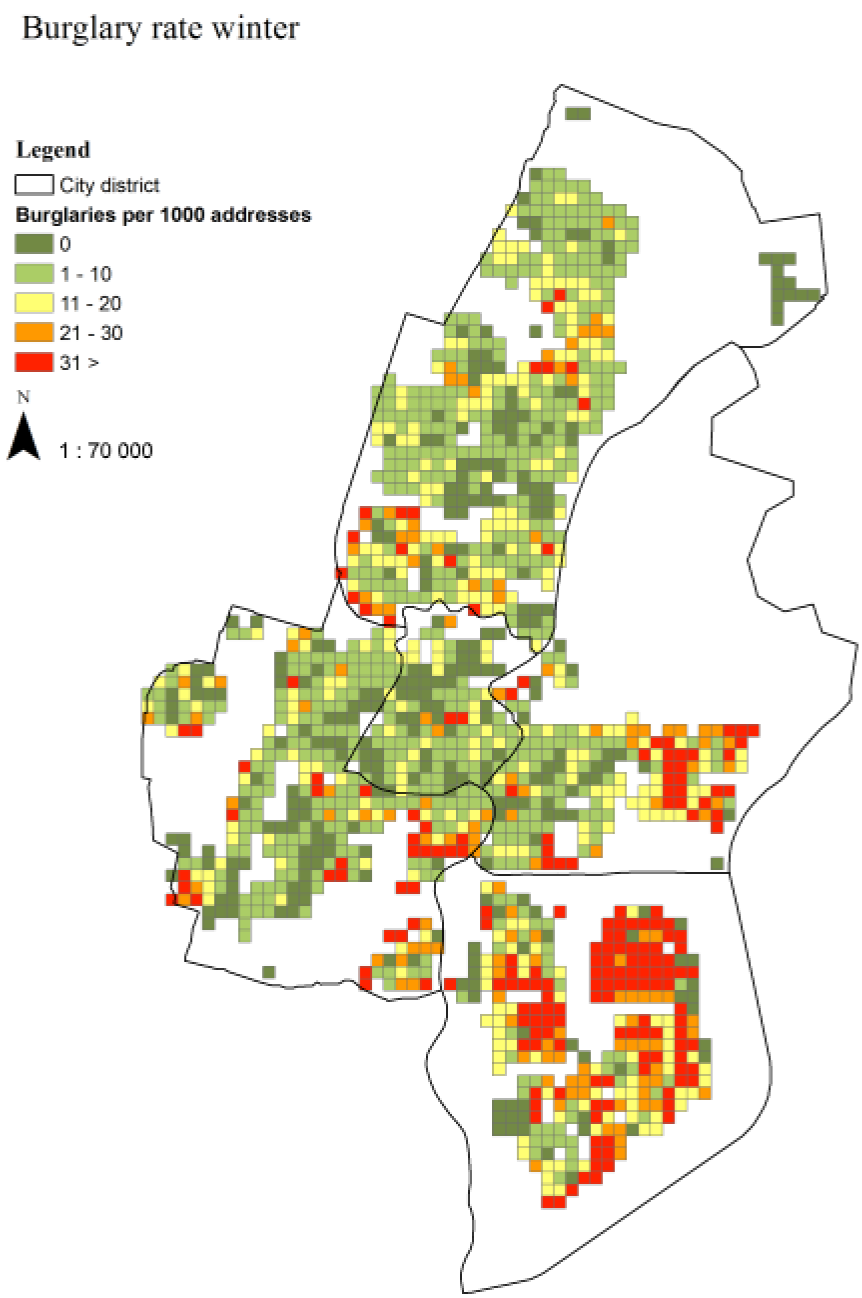


Figure 5.1: Map showing the burglary rates per cell during winter.

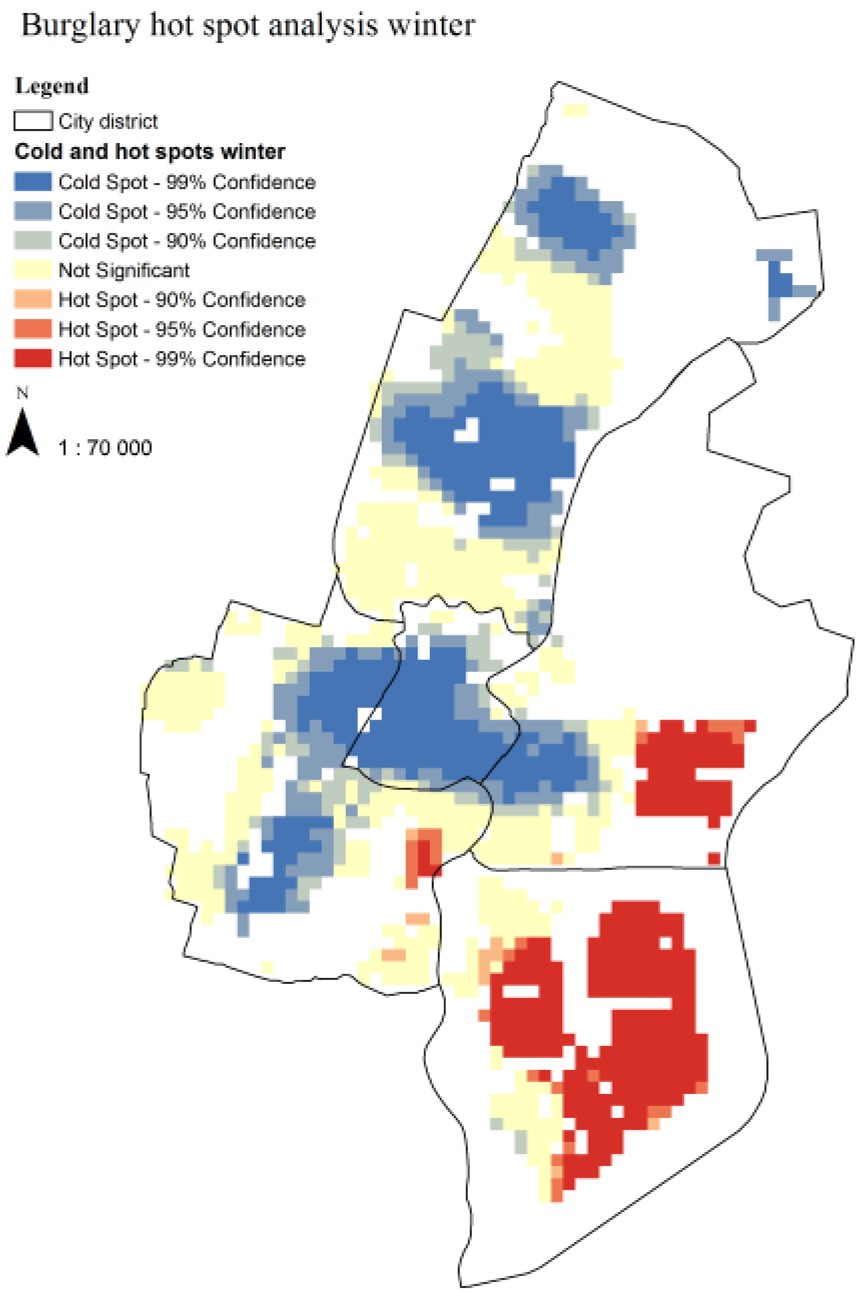


Figure 5.2: Map showing the significant cold and hot spots during winter. A fixed euclidean distance band of 400 meters is used for the analysis.

Model building

The first step is to check for any multicollinearity between the independent variables. For all pairs of independent variables a bivariate correlation coefficient is calculated, Spearman's r. When there are bivariate coefficients higher than 0.9, the cutoff value defined earlier, one of the involved independent variables has to be removed from the model.

A check for multicollinearity showed that there are no independent variables with coefficients higher than 0.9, meaning there is no multicollinearity. However, the analysis did show relatively strong relationships ($r \geq 0.7$) between some of the independent variables. These are summarized in table 5.1. As the independent variables for all seasons are the same, only the dependent variable changes, there is no need to check for multicollinearity for the other seasons.

Most of the high correlations found between independent variables are not unexpected. For example, the relationship between 'household income' and 'property value' is obvious, as both can be related to the general wealth of the population in an area. But for example the relationship between 'distance highway entry' and 'distance low incomes' is less obvious. A visual inspection of the data shows that areas with high percentages of people with low incomes are also areas that are relatively close to highway entries, possibly due to lower property values typically associated with higher traffic volumes in an area. This can explain the relatively high correlations between some of the independent variables.

The next step is to define an initial model and check the residuals for spatial autocorrelation. The spatial relationship is conceptualized as a fixed distance band of 400 meters, as the literature study suggested this is the distance burglaries influence each other (Johnson, 2008). Based on a calculation of Moran's Index, the residuals from the initial fitted regression model can be described as clustered (see figure 5.3). This means that spatial autocorrelation exists in the model's residuals. Based on the z-score, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

To deal with the spatial autocorrelation, a spatial lag variable is added to the regression model. The spatial lag variable is based on a spatial weights matrix where neighbors are again defined as the grid cells within 400 meters of the base cell. After including the spatial lag variable in the initial model, it is again tested for spatial autocorrelation. Unfortunately, the spatial autocorrelation could not be completely removed from the model, but adding the spatial lag variable caused Moran's I to drop from 0.12 to 0.04, a decrease of more than 67%.

The final model is summarized in figure 5.4 (only including the significant independent variables and the spatial lag variable) and the full model is specified in appendix C. The top three risk factors that explain most of the variance in burglaries during the winter are in descending order:

1. distance to welfare benefits,

2. distance to retail and catering, and
3. risky properties.

Independent variable 1	Independent variable 2	Spearman's r
Distance highway entry	Distance low incomes	0.89
Household income	Election turnout	0.81
Household income	Welfare benefits	-0.79
Household income	Propery value	0.74
Residential mobility	Cars per household	0.73
Welfare benefits	Distance highway entry	-0.72
Welfare benefits	Distance welfare benefits	-0.72
Rental properties	Distance rental properties	-0.72

Table 5.1: Correlation coefficients where $r \geq 0.7$ sorted from high to low.

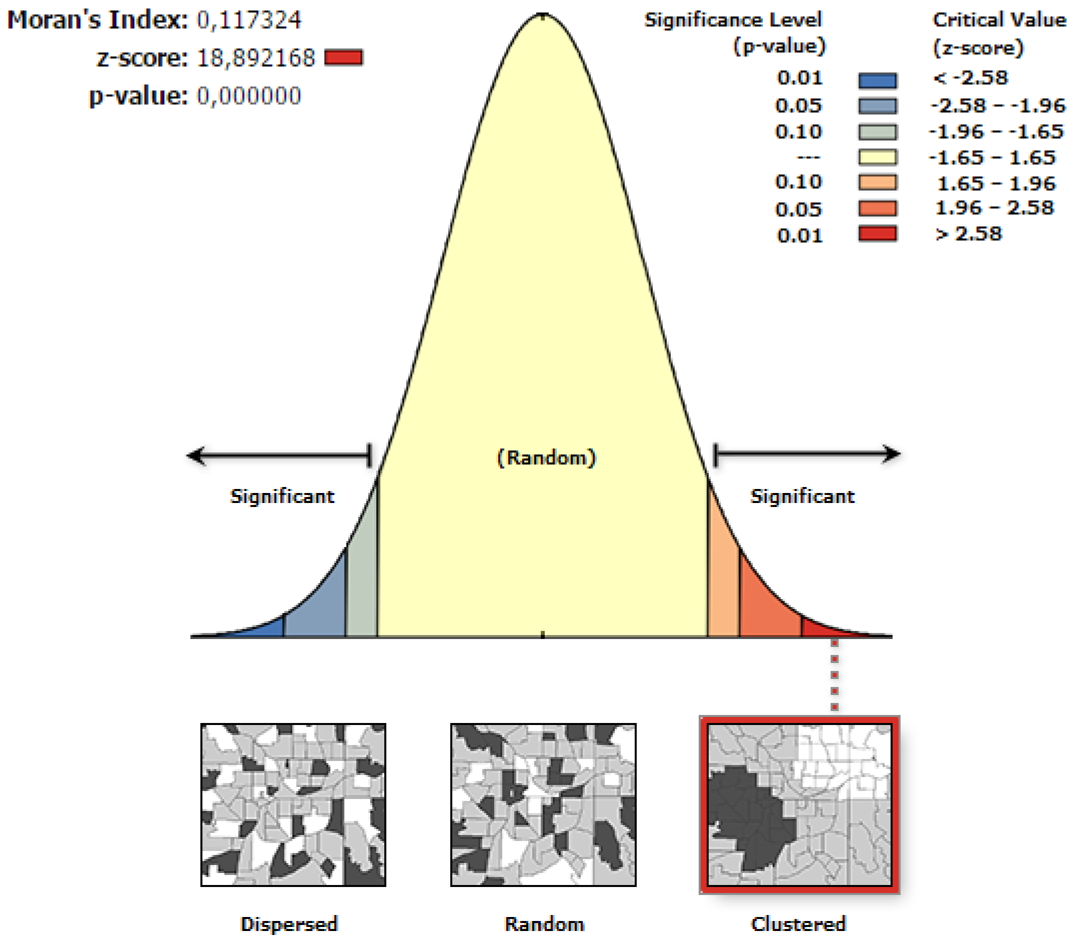


Figure 5.3: Results of the check for spatial autocorrelation among the residuals of the initial model for burglaries during winter. A fixed distance band of 400 meters is used.

Parameter	Beta coefficient	Standardized beta coefficient	Significance	Interpretation
Spatial lag	0,038	0,024	0,000	
Distance welfare benefits	0,000	-0,014	0,000	An increase of the distance to the nearest area with high percentages of welfare benefits with 100 meters, causes a 3% decrease of the fitted mean burglary rate.
Distance retail and catering	-0,001	-0,011	0,000	An increase of the distance to the nearest concentration of retail and catering establishments with 100 meters, causes a 8% decrease of the fitted mean burglary rate.
Risky properties	0,920	0,008	0,000	An increase of the risky properties of 10 percentage point, causes a 10% increase of the fitted mean burglary rate.
Distance city center	0,000	0,008	0,019	An increase of the distance to the city center with 100 meters, causes a 1% increase of the fitted mean burglary rate.
Building density	-0,005	-0,007	0,000	An increase of 10 addresses per grid cell (100 by 100 meters), causes a 5% decrease of the fitted mean burglary rate.
Distance ethnic heterogeneity	-0,001	-0,007	0,006	An increase of the distance to the nearest area with high ethnic heterogeneity with 100 meters, causes a 5% decrease of the fitted mean burglary rate.
Welfare benefits	-0,064	-0,006	0,029	An increase of the population percentage receiving welfare benefits with 1 percentage point, causes a 6% decrease of the fitted mean burglary rate.
Distance demographic risk group	-0,001	-0,006	0,000	An increase of the distance to the nearest concentration of residents in the demographic risk group with 100 meters, causes a 13% decrease of the fitted mean burglary rate.
Distance rental properties	-0,001	-0,005	0,038	An increase of the distance to the nearest area with a high percentage of rental properties with 100 meters, causes a 7% decrease of the fitted mean burglary rate.
Crime	0,230	0,004	0,040	An increase of the number of crime reports per address of 1, causes a 26% increase of the fitted mean burglary rate.
Edge dwellings	0,165	0,003	0,030	An increase of the percentage of edge dwellings with 10 percentage point, causes a 2% increase of the fitted mean burglary rate.

Figure 5.4: The final model for the winter including only the significant parameters with 95% confidence and sorted by standardized beta coefficient.

Risk terrain surface

A risk layer with value 1 for low risk, value 2 for medium risk and value 4 for high risk is created for each of the 11 significant risk factors during winter. Figure 5.5 shows an example risk layer for the risk factor *risky properties*. All of the risk layers are then weighted by their respective standardized beta coefficients and combined into one risk terrain surface. Figure 5.6 shows the final burglary risk terrain surface for the winter. Risk values are scaled relative to the lowest risk value. This means that the dark red areas with a value of close to 3, have about a 3 times higher risk of burglaries based on the included risk factors when compared to the areas with the lowest risk.

Next, a hot spot map is created based on the risk terrain surface. This hot spot map can then be compared to the initial hot spot map, based on the actual burglary count during the winter (see figure 5.7). This comparison can give an indication of the general accuracy of the created risk terrain surface. In addition, potential future hot spots can be identified.

The comparison shows a rather similar image between the actual burglary rates and the modeled burglary rates. The highest concentration of burglaries is found in the southeast district of *Schalkwijk*. Based on the individual risk layers it can be concluded that this concentration can be explained based on the presence of multiple offender-related risk factors: many characteristics of an offender neighborhood are found in *Schalkwijk*. *Schalkwijk* has many people receiving welfare benefits, has a high ethnic heterogeneity, has many people in the demographic risk group (males aged 15 to 24) and has a high percentage of rental properties.

In addition, the model-based hot spot map also shows some potential areas where burglaries might increase in the future. These areas can be found in the northern district *Noord* and the eastern district *Oost* (numbered 1 to 3 in figure 5.7).

The first potential hot spot is in the northwest of the study area, in the neighborhood *Delftwijk* (area 1 on the map). It can be explained by a culmination of offender risk factors. The area is close to a cluster of retail and catering establishments, close to areas with high ethnic heterogeneity, close to areas with many rental properties and close to areas with a high percentage of the population in the demographic risk group for offenders. So it can be concluded that this area is mostly at risk because of the high potential as an offender neighborhood.

The second potential hot spot is in the northeast of Haarlem, within the neighborhood *Spaarndam* (area 2 on the map). This area is mostly at risk because of the many risky properties (detached, semi-detached and corner properties), the high building density, the high amount of welfare benefits and the close distance to areas with a high population in the demographic risk group. Therefore, *Spaarndam* is at increased risk because of a combination of all types of risk factors: target, offender and setting.

The third potential hot spot is in the eastern district of Haarlem, in the neighborhood *Slachthuisbuurt* (area 3 on the map). This neighborhood is characterized by its proximity to areas with many people receiving welfare benefits, high ethnic heterogeneity, high percentages of rental properties and many people within the demographic risk group. It can be concluded that the *Slachthuisbuurt* is mostly at increased risk because of its close proximity to *Schalkwijk*, a neighborhood where potential offenders live.

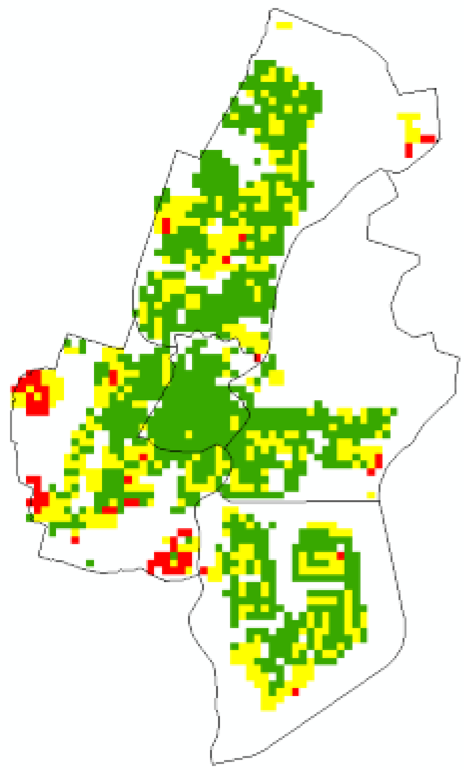


Figure 5.5: The intermediate risk layer for *risky properties* during winter. The green areas are low risk areas (risk value 1), the yellow areas are medium risk layers (risk value 2) and the red areas are high risk areas (risk value 4).

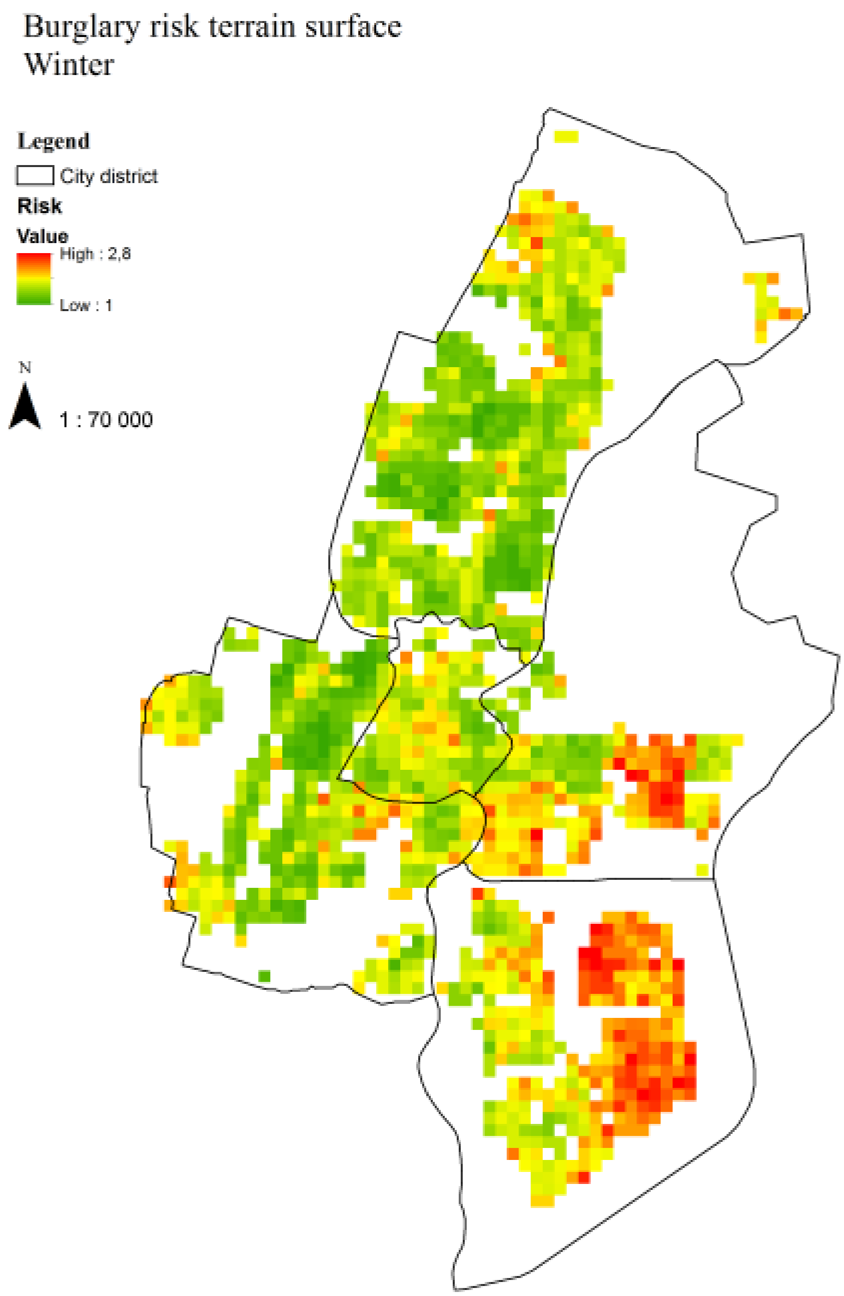


Figure 5.6: Burglary risk terrain surface for the winter.

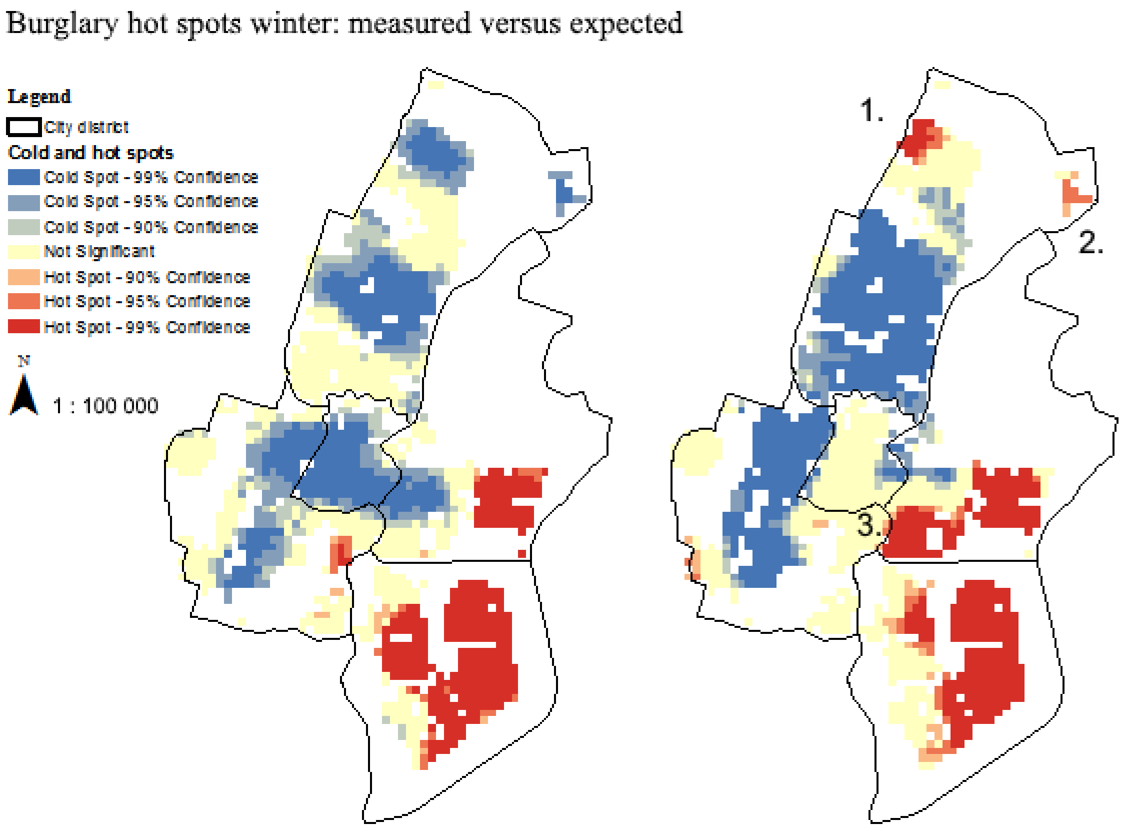


Figure 5.7: The measured burglary hot spots in the winter (left) versus the expected burglary hot spots based on the significant risk factors in the fitted model (right).

5.1.2 Burglaries during spring

Situation

642 burglaries occurred during the spring, which is 22% of the total number of burglaries from 2010 to 2013. Again, these burglaries occur most often in the south-east districts of Haarlem (see figure 5.8), which is also clearly visible from the results of the hot spot analysis (see figure 5.9). Cold spots are still found in the North, East and Center districts.

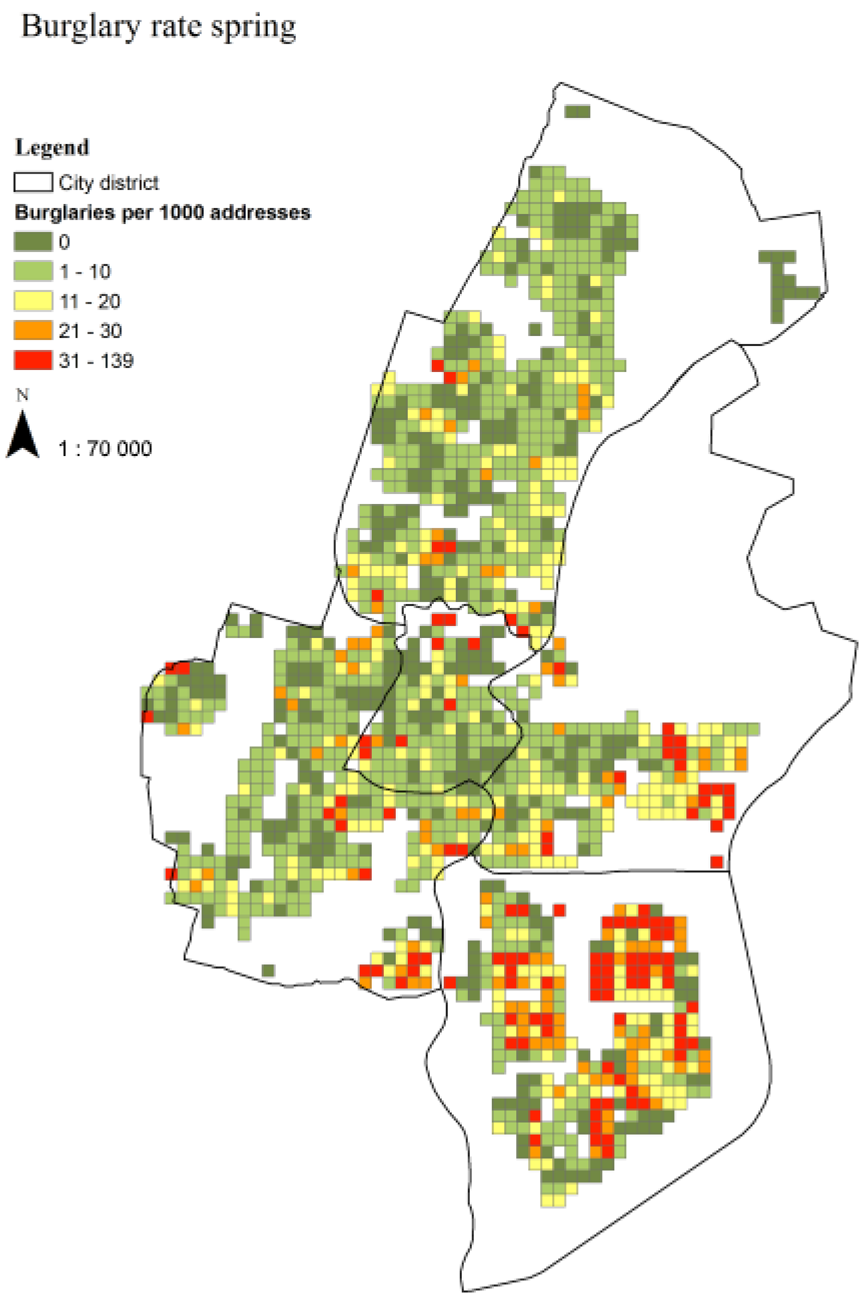


Figure 5.8: Map showing the burglary rates per cell.

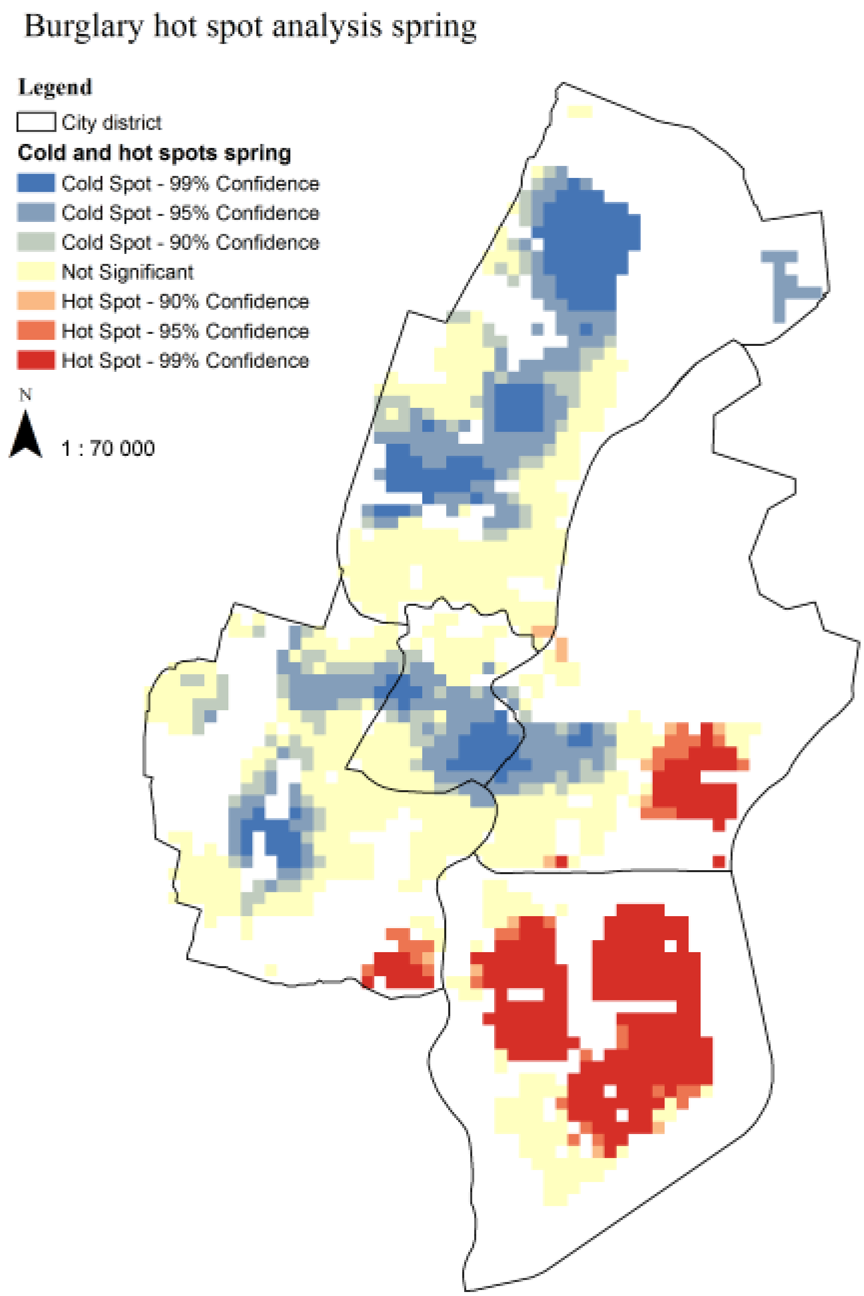


Figure 5.9: Map showing the significant cold and hot spots. A fixed euclidean distance band of 400 meters is used for the analysis.

Model building

The check for spatial autocorrelation shows there is significant clustering of the residuals of the initial model. After adding a spatial lag variable, all spatial autocorrelation is removed from the model and the residuals show a random spatial pattern.

The final model is summarized in figure 5.10 and the full model is specified in appendix C. The risk factors that explain most of the variance in burglaries during spring are in descending order:

1. risky properties,
2. building density, and
3. election turnout.

Parameter	Beta coefficient	Standardized beta coefficient	Significance	Interpretation
Spatial lag	0,050	0,026	0,000	
Risky properties	1,233	0,014	0,000	An increase of the risky properties of 10 percentage point, causes a 13% increase of the fitted mean burglary rate.
Building density	-0,006	-0,010	0,000	An increase of 10 addresses per grid cell (100 by 100 meters), causes a 6% decrease of the fitted mean burglary rate.
Election turnout	-0,015	-0,010	0,008	An increase of the election turnout of 10 percentage point, causes a 14% decrease of the fitted mean burglary rate.
Accessibility	0,379	0,007	0,002	An increase in the accessibility score of 0.1, causes a 4% increase of the fitted mean burglary rate.
Distance ethnic heterogeneity	0,000	-0,007	0,024	An increase of the distance to the nearest area with high ethnic heterogeneity with 100 meters, causes a 4% decrease of the fitted mean burglary rate.
Distance to street	0,017	0,006	0,047	An increase of the distance of dwellings to the street of 10 meters, causes a 18% increase of the fitted mean burglary rate.

Figure 5.10: The final model for the spring including only the significant parameters with 95% confidence.

Risk terrain surface

Based on the 6 significant risk factors for spring a risk terrain surface is created (see figure 5.11). Based on this risk terrain surface a hot spot map is created and compared to the initial hot spot map based on the actual burglary counts during spring (figure 5.12).

Again, the two hot spot maps paint a similar picture: the major hot spot is found in the southeast district of *Schalkwijk*. Based on the individual risk layers, it seems that the high ethnic heterogeneity in *Schalkwijk* is again a significant risk factor. But also the average distance of the dwellings to the nearest street, which is a setting risk factor, seems to be a relevant factor.

But besides the major hot spot in *Schalkwijk*, three other hot spots can be identified based on the risk terrain surface of the spring. These hot spots did not appear in the hot spot analysis of the actual burglary rates, thus they can signal areas of potential increases of burglaries in the future. These areas are labeled 1, 2 and 3 in figure 5.12.

The potential hot spot in *Spaarndam* (labeled 1) was identified during the winter, too. It can be explained by the high building density and the high amount of risky properties in that area.

The second potential hot spot is in the neighborhood *Duinwijk*. This neighborhood runs a higher risk of burglaries during spring based on the high percentage of risky properties, i.e. a high percentage of detached, semi-detached and corner properties.

The third potential hot spot can be found in the neighborhood *Slachthuisbuurt*, a potential hot spot also identified by the risk terrain surface for the winter. During the winter it was mainly the proximity to the potential offender neighborhood *Schalkwijk* that caused the increased risk of burglaries. For the spring, there also seems to be a relationship between a higher risk and the low election turnout in this neighborhood. Election turnout is a setting risk factor based on the social disorganization theory.

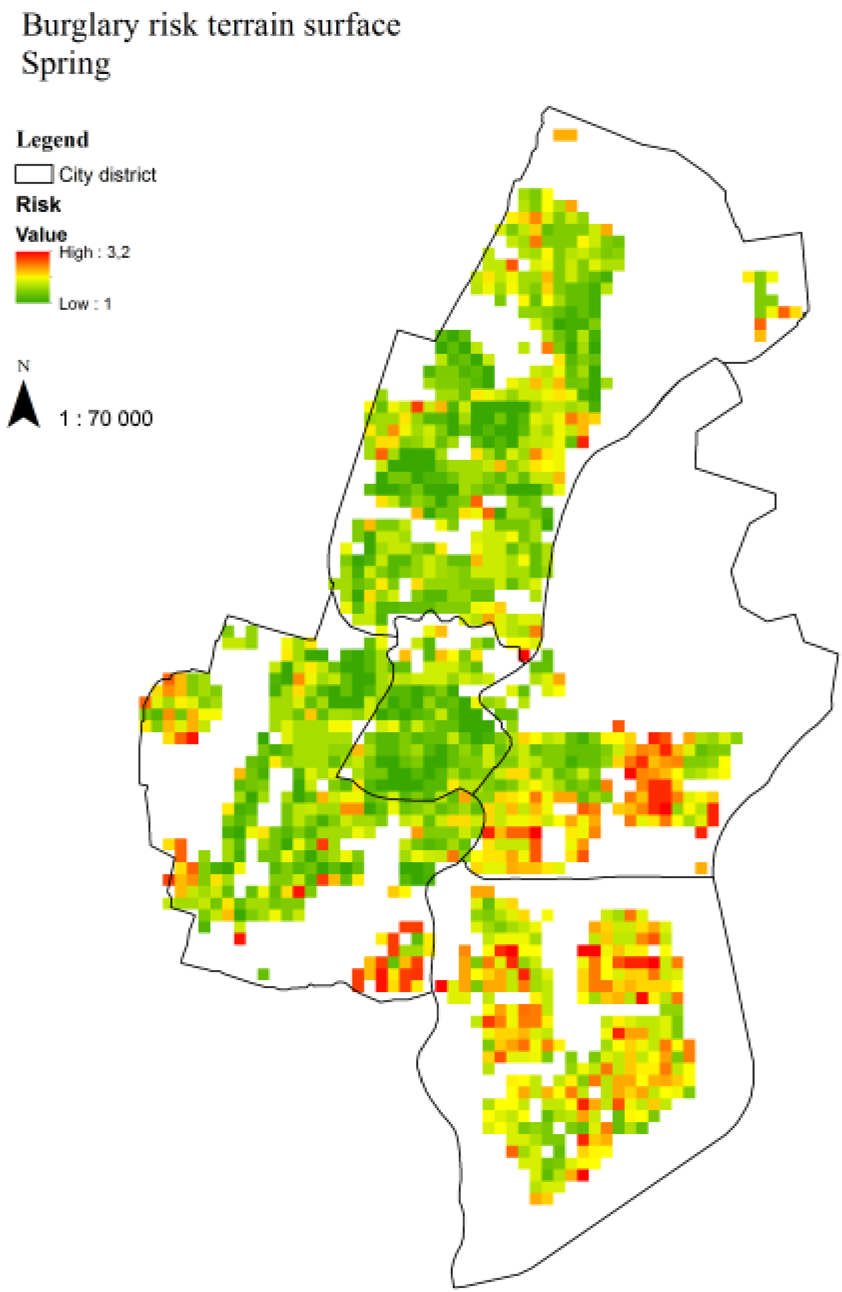


Figure 5.11: Burglary risk terrain surface for the spring.

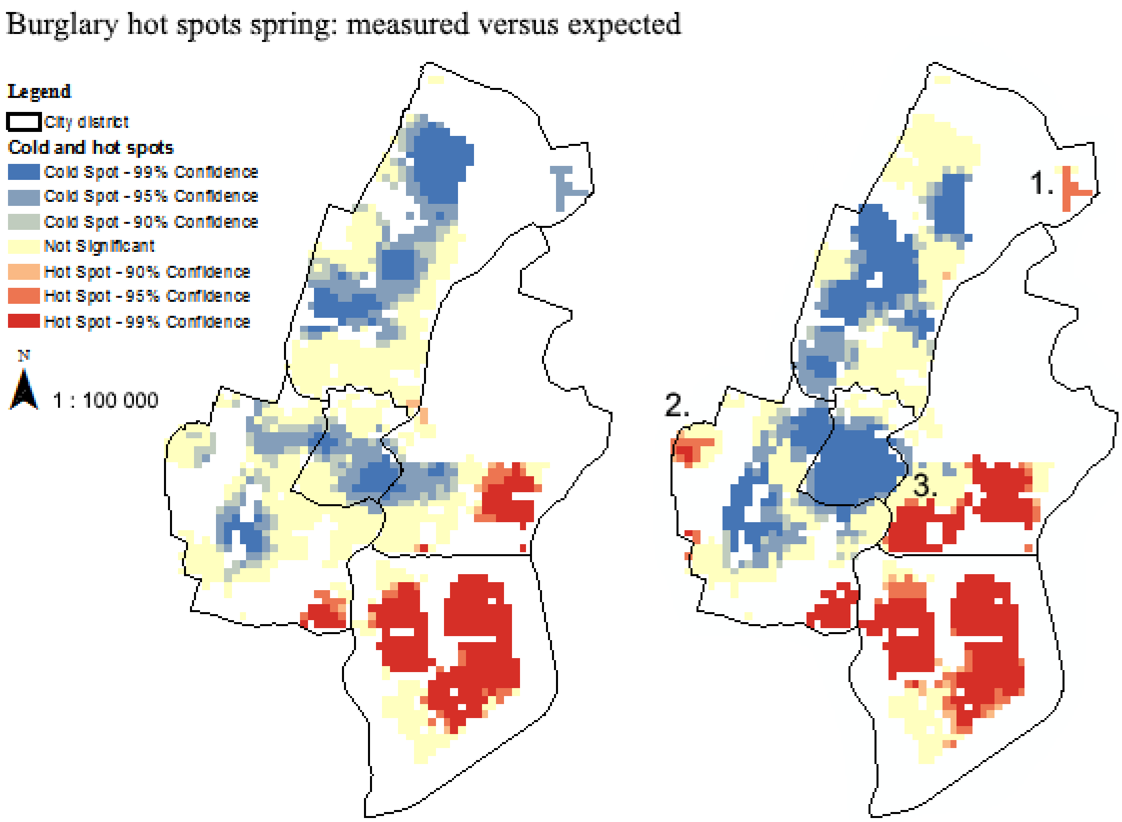


Figure 5.12: The measured burglary hot spots in the spring (left) versus the expected burglary hot spots based on the significant risk factors in the fitted model (right).

5.1.3 Burglaries during summer

Situation

687 burglaries occurred during summer, which amounts to 24% of total burglaries from 2010 to 2013. The general spatial pattern found in the previous analyses is not much different for the summer. The burglaries occur most often in the south-east districts of Haarlem (see figure 5.13), which is also clearly visible from the results of the hot spot analysis (see figure 5.14). Cold spots are still found in the North, East and Center districts, but it stands out that there is also a small cold spot discernible in the south-west of *Schalkwijk*.

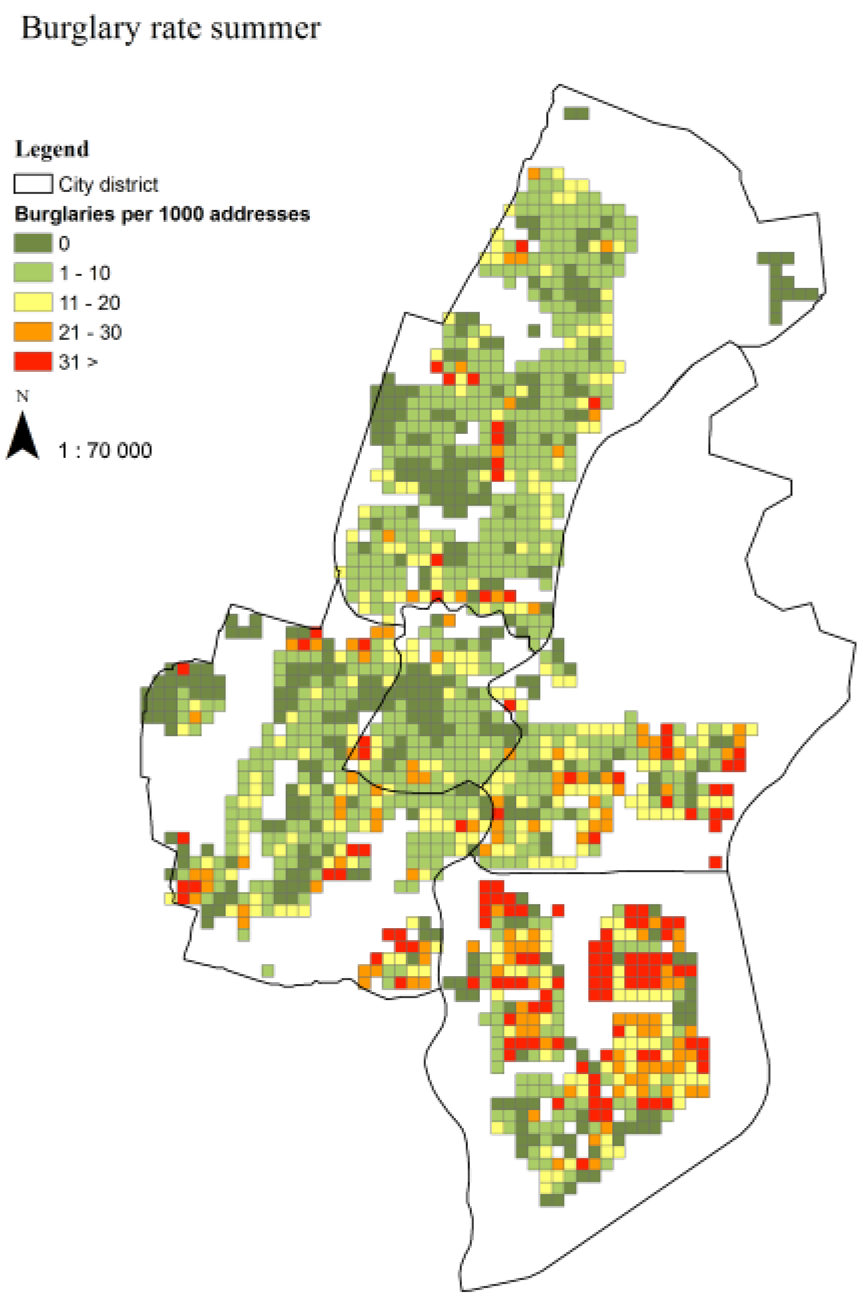


Figure 5.13: Map showing the burglary rates per cell.

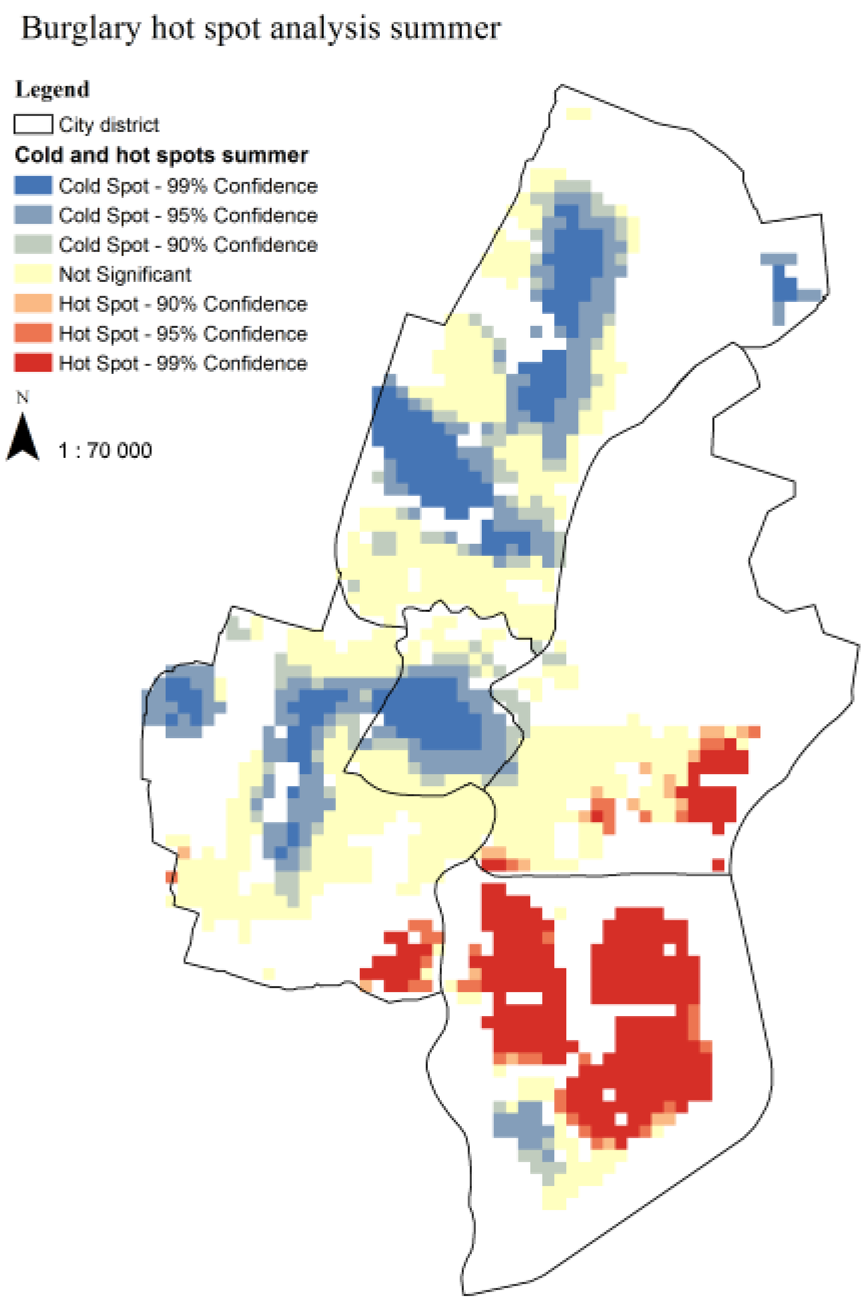


Figure 5.14: Map showing the significant cold and hot spots. A fixed euclidean distance band of 400 meters is used for the analysis.

Model building

The test of the residuals of the initial model signals the presence of spatial autocorrelation. A spatial lag variable is added to the model, which corrects for the spatial autocorrelation. Now the model residuals are not spatially clustered anymore, but show a random spatial pattern.

The final model is summarized in figure 5.15 and the full model is specified in appendix C. The risk factors that explain most of the variance in burglaries during summer are in descending order:

1. distance to welfare benefits,
2. property value, and
3. building density.

Parameter	Beta coefficient	Standardized beta coefficient	Significance	Interpretation
Spatial lag	0,050	0,026	0,000	
Distance welfare benefits	0,000	-0,010	0,043	An increase of the distance to the nearest area with high percentages of welfare benefits with 100 meters, causes a 2% decrease of the fitted mean burglary rate.
Property value	0,000	0,009	0,012	An increase of the property value of €10000, causes a 1% increase of the fitted mean burglary rate.
Building density	-0,005	-0,009	0,000	An increase of 10 addresses per grid cell (100 by 100 meters), causes a 5% decrease of the fitted mean burglary rate.
Distance ethnic heterogeneity	-0,001	-0,009	0,005	An increase of the distance to the nearest area with high ethnic heterogeneity with 100 meters, causes a 5% decrease of the fitted mean burglary rate.
Rental properties	0,316	0,008	0,004	An increase of rental properties of 10 percentage point, causes a 3% increase of the fitted mean burglary rate.
Distance rental properties	0,001	0,007	0,016	An increase of the distance to the nearest area with a high percentage of rental properties with 100 meters, causes a 8% increase of the fitted mean burglary rate.

Figure 5.15: The final model for the summer including only the significant parameters with 95% confidence.

Risk terrain surface

Based on the 6 significant risk factors for summer a risk terrain surface is created (see figure 5.16). From this, a hot spot map is created and compared to the initial hot spot map based on actual burglary counts (see figure 5.17).

The spatial trend of burglaries is very similar between the modeled and the actual hot spots. The main hot spot is again in *Schalkwijk*, in the southeast of the study area. Again, offender neighborhood risk factors are responsible.

Two potential hot spots can be identified: one area within the neighborhoods *Koninginnebuurt* and *Den Hout* (labeled 1 on the map in figure 5.17) and one in the *Slachthuisbuurt* (labeled 2).

The first potential hot spot can mainly be explained by the high property values in that area. Therefore, this potential hot spot is due to the presence of attractive targets in the area.

The second potential hot spot in the *Slachthuisbuurt* is caused by offender risk factors: it has many rental properties, it is close to areas with many people receiving welfare benefits and it is close to areas with high ethnic heterogeneity.

Burglary risk terrain surface
Summer

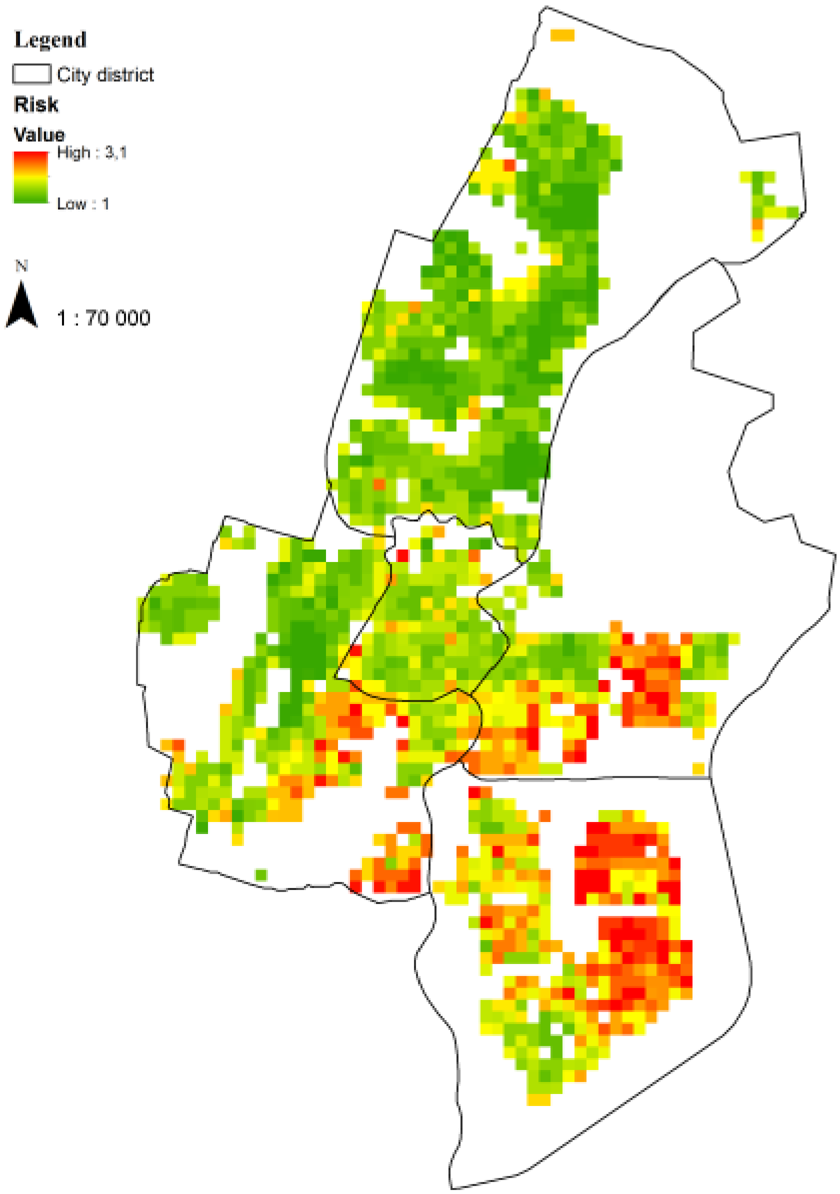


Figure 5.16: Burglary risk terrain surface for the summer.

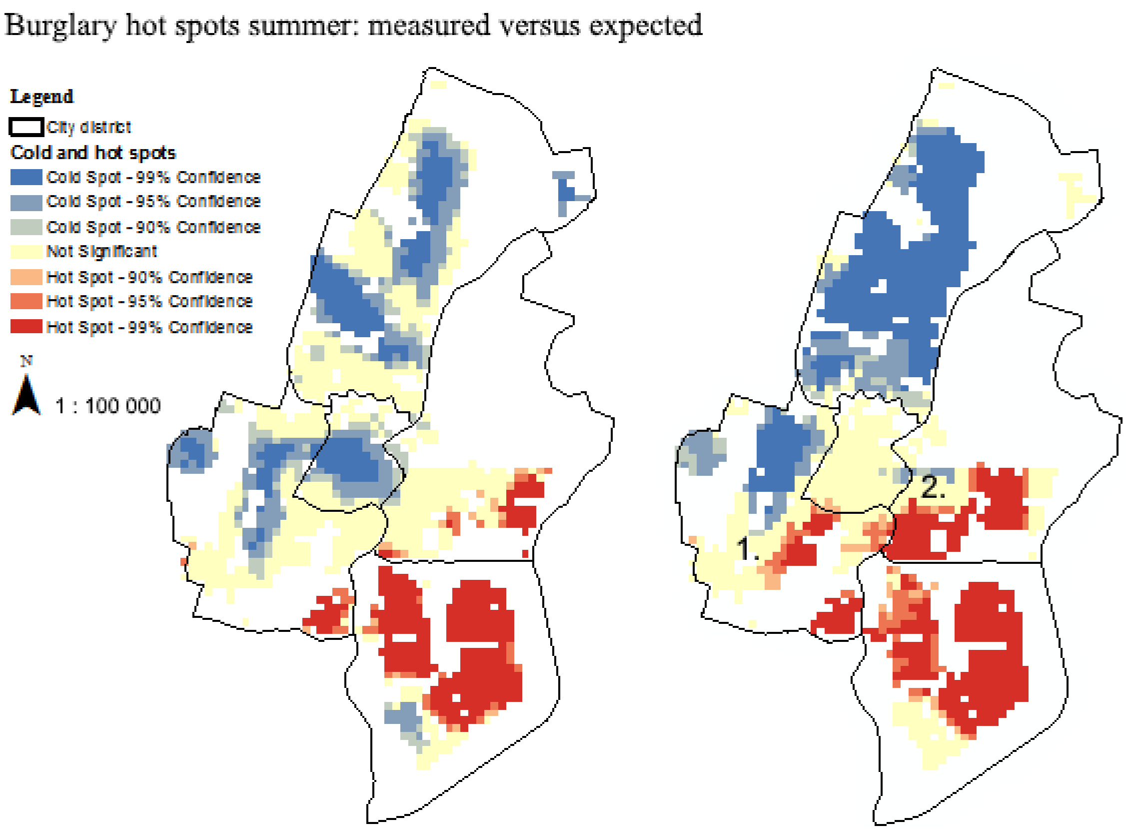


Figure 5.17: The measured burglary hot spots in the summer (left) versus the expected burglary hot spots based on the significant risk factors in the fitted model (right).

5.1.4 Burglaries during autumn

Situation

701 burglaries occurred during autumn, which comes down to 24% of the total number of burglaries from 2010 to 2013. Just like in the other seasons, burglaries occur most often in the south-east districts of Haarlem (see figure 5.18), which is illustrated by the results of the hot spot analysis (see figure 5.19). Cold spots are again found in the North, East and Center districts.

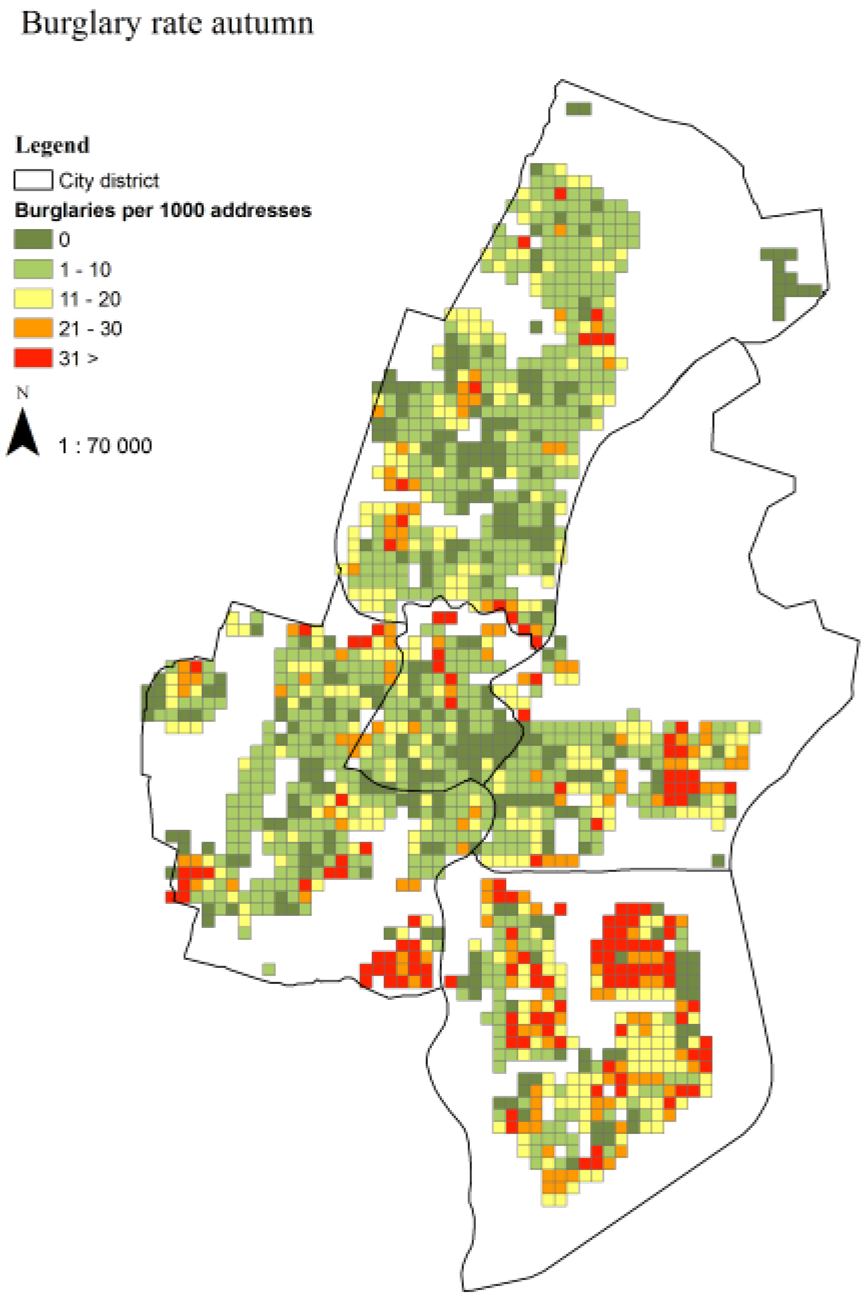


Figure 5.18: Map showing the burglary rates per cell.

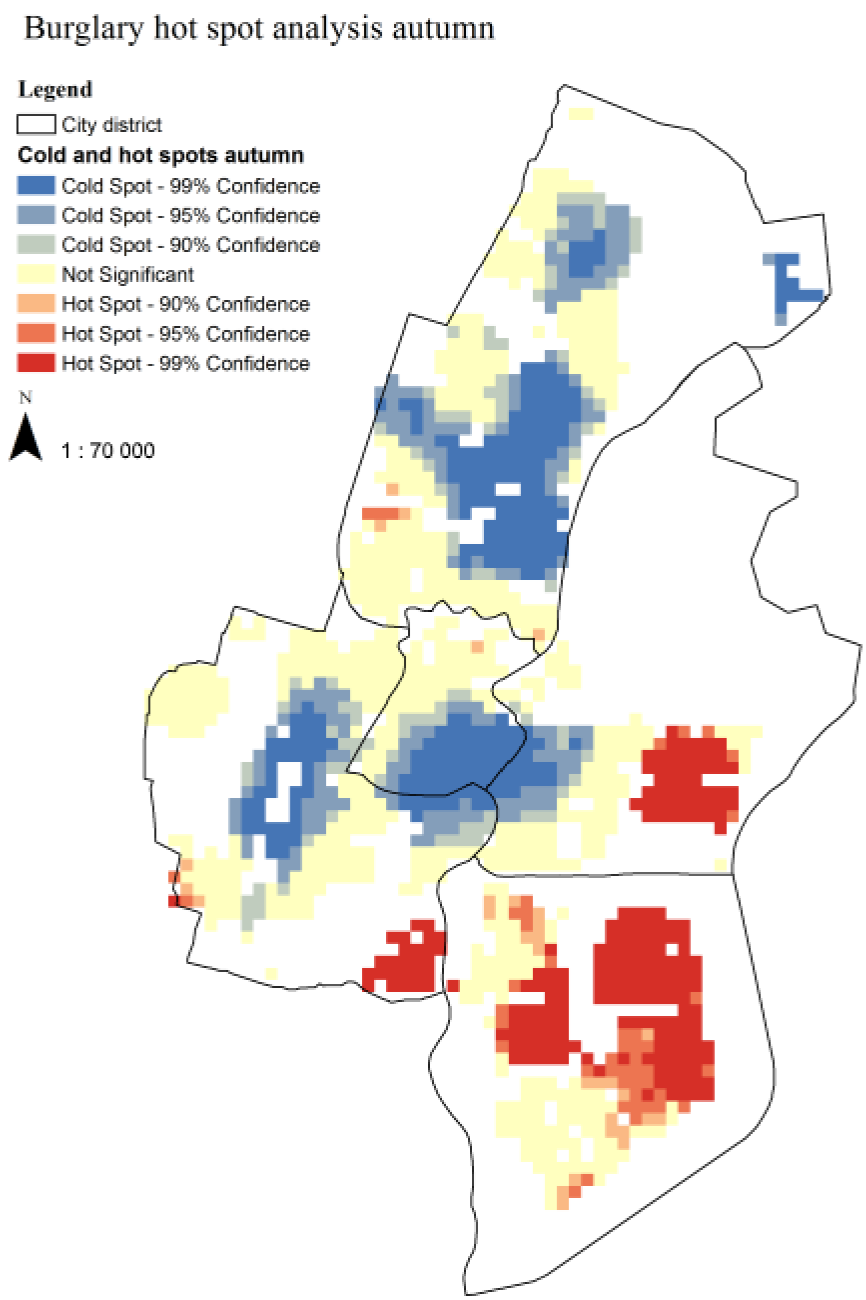


Figure 5.19: Map showing the significant cold and hot spots. A fixed euclidean distance band of 400 meters is used for the analysis.

Model building

Again, analysis of the residuals of the initial model shows that they are spatially autocorrelated and display a spatially clustered pattern. After adding a spatial lag variable, the residuals are no longer spatially autocorrelated.

The final model is summarized in figure 5.20 and the full model is specified in appendix C. The risk factors that explain most of the variance in burglaries during autumn are in descending order:

1. distance to ethnic heterogeneity,
2. risky properties, and
3. property value.

Parameter	Beta coefficient	Standardized beta coefficient	Significance	Interpretation
Spatial lag	0,052	0,027	0,000	
Distance ethnic heterogeneity	-0,001	-0,012	0,000	An increase of the distance to the nearest area with high ethnic heterogeneity with 100 meters, causes a 7% decrease of the fitted mean burglary rate.
Risky properties	0,707	0,008	0,007	An increase of the risky properties of 10 percentage point, causes a 7% increase of the fitted mean burglary rate.
Property value	0,000	0,008	0,048	An increase of the property value of €10000, causes a 1% increase of the fitted mean burglary rate.
Building density	-0,004	-0,008	0,002	An increase of 10 addresses per grid cell (100 by 100 meters), causes a 4% decrease of the fitted mean burglary rate.
Distance demographic risk group	-0,001	-0,004	0,049	An increase of the distance to the nearest concentration of residents in the demographic risk group with 100 meters, causes a 8% decrease of the fitted mean burglary rate.

Figure 5.20: The final model for the autumn including only the significant parameters with 95% confidence.

Risk terrain surface

Based on the 6 significant risk factors for autumn a risk terrain surface is created (see figure 5.21). Again, a hot spot map showing potential future hot spots is derived from the risk terrain surface and compared to the hot spots based on actual burglary counts (see figure 5.22).

Both hot spot analyses identify the southeast district of *Schalkwijk* again as the most risky area based on offender risk factors like the proximity to areas with high ethnic heterogeneity and to areas with a high population in the demographic risk group.

Only one potential hot spot can be identified and it is in the *Koninginnebuurt* (labeled 1 on the map in figure 5.22). This is caused by the relatively high property values in this area, making it an attractive target area to burglars.

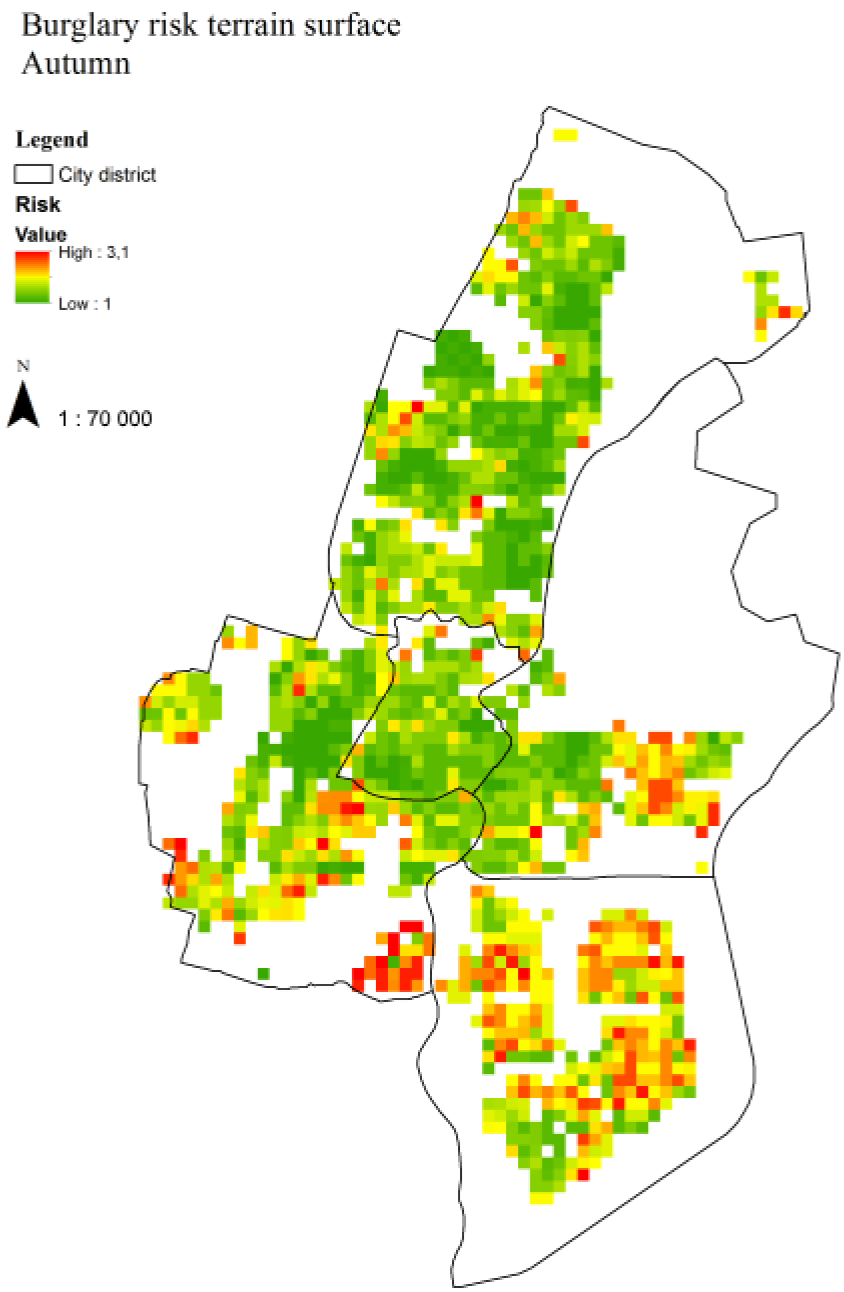


Figure 5.21: Burglary risk terrain surface for the autumn.

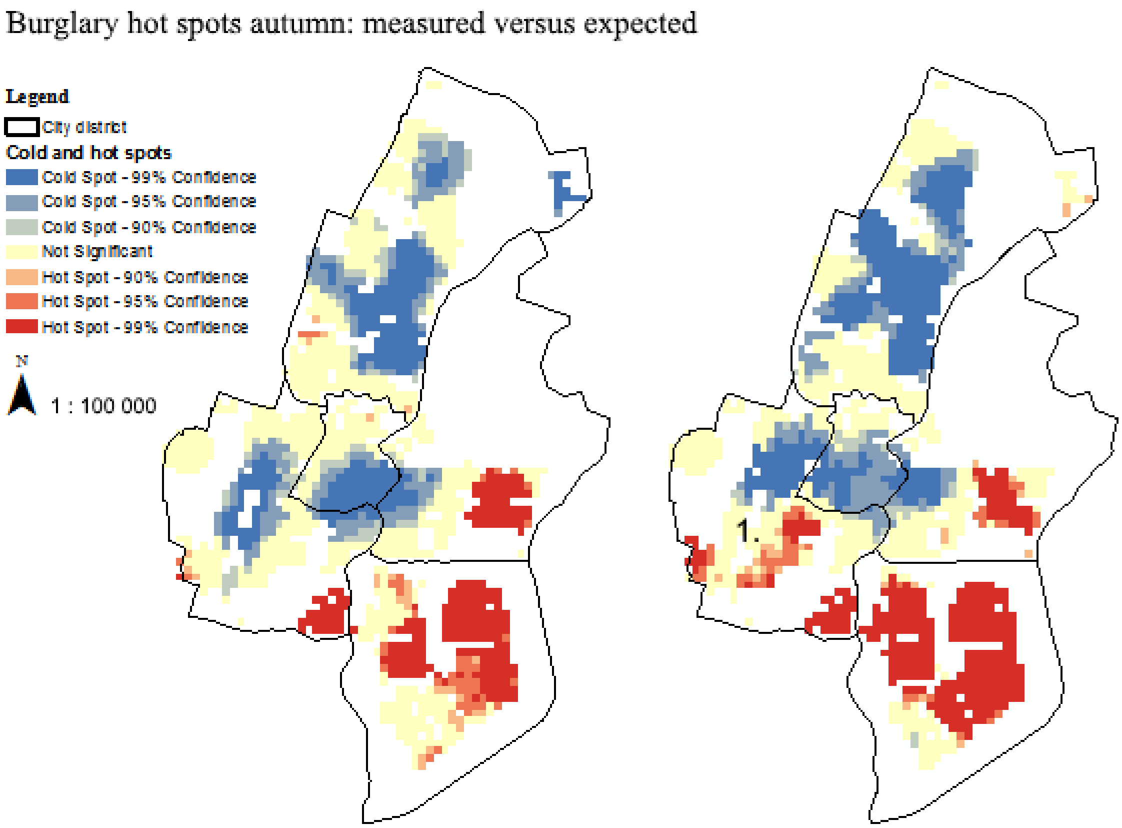


Figure 5.22: The measured burglary hot spots in the autumn (left) versus the expected burglary hot spots based on the significant risk factors in the fitted model (right).

5.2 Interpretation

This section focuses on the interpretation of the results found in the previous section. Are the significant relationships found between the independent variables and the dependent variable as expected based on theory? How are the different risk factor types (target, offender and setting) represented in the seasonal models? And what are the main differences between the seasons? These questions, and more, are answered here.

5.2.1 Risk factor hypotheses

The risk factors used in this research are derived from literature. Based on different theories, these risk factors have a potential influence on the burglary rate in an area. There are also hypotheses made describing the expected direction of a relationship: does an increase of independent variable x cause an increase or decrease of the dependent variable y? This first subsection reflects on these hypotheses related to the risk factors of burglaries.

The table in figure 5.23 summarizes all risk factors that turned out to be significant explanatory variables in explaining the burglary rates in at least one of the seasons. The most consistent risk factors in explaining burglary rates in Haarlem are:

1. *distance ethnic heterogeneity*,
2. *building density* and
3. *risky properties*.

These risk factors are significant in at least three seasons and all have relatively high average standardized beta coefficients.

Figure 5.23 also shows the relationship direction hypothesis and the actual measured relationship direction per season. Almost all hypotheses concerning the direction of a relationship are correct based on the findings from the Haarlem data, but there are some deviations: the most notable being *building density*.

The initial assumption was that a higher building density would cause a higher burglary rate, thus a positive relationship, based on the theories of rational choice and optimal foraging. One could reason that a higher density of potential targets makes an area more attractive to burglars. Nonetheless, all seasons consequently show a negative relationship between building density and burglary rates in Haarlem. How can this be explained? Of course this could be a data pattern only observable in Haarlem, making it a local deviation. But it could also be that the underlying theories of rational choice and optimal foraging are not applicable here. Perhaps the observed pattern could better be explained by the environmental design theory, making *building density* a setting risk factor instead of a target risk factor. A higher building density potentially increases the 'number of eyes' on the street, making an area less attractive to burglars.

Another dissonant is *distance rental properties*. The hypothesis was that there is a negative relationship between the distance to rental properties and the burglary rate. This hypothesis can be confirmed based on data from the winter but disproved based on data from the summer. Being close to areas with many rental properties is a risk factor in the winter, but the opposite is true for the summer. Therefore, there seems to be a seasonal effect on the influence of distance to rental properties on the burglary rate. More research is needed to find a conclusive explanation for this finding.

Finally, there were some risk factors that were initially not classified as one type of risk factor (see figure 5.23). It can be concluded based on the observed relationship directions that *rental properties* is more a setting risk factor than a target risk factor and that *accessibility* is rather an offender than a setting risk factor. More rental properties in an area cause a higher burglary rate, confirming the assumptions from the social disorganization theory. Likewise, a higher accessibility value in an area causes a higher burglary rate, confirming the assumptions based on the awareness space theory.

For fourteen of the initial risk factors derived from literature no evidence for a significant relationship with burglaries is found in the Haarlem data (see figure 5.24). Many of these are offender risk factors or setting risk factors based on the 'environmental design' theory.

There can be several explanations why no causal relationship is found for these risk factors. It can simply mean that a certain risk factor is not playing a part in explaining spatial patterns of burglaries in Haarlem. It could also be very well possible that the spatial unit used here, namely grid cells of 100 by 100 meters, has its effect on the outcomes. For example, the distance to street lighting could be a risk factor varying from dwelling to dwelling, but these local effects are smoothed out in spatial patterns on a smaller scale i.e. the average distance from a dwelling to street lighting is generally the same when measurements are aggregated over a larger area. A recommendation for future research is to try to quantify the effect that the size of spatial units can have on the modeling results by repeating the modeling process with larger as well as with smaller spatial units.

Risk factor	Type and theory	Relationship direction hypothesis	Significant in	Relationship direction	Average standardized beta coefficient
Distance ethnic heterogeneity	Offender: Offender neighborhood	Negative	Winter	Negative	-0,009
			Spring	Negative	
			Summer	Negative	
			Autumn	Negative	
Building density	Target: Rational choice / Optimal foraging	Positive	Winter	Negative	-0,008
			Spring	Negative	
			Summer	Negative	
			Autumn	Negative	
Risky properties	Target: Rational choice / Optimal foraging Setting: Environmental design	Positive	Winter	Positive	0,010
			Spring	Positive	
			Autumn	Positive	
Distance welfare benefits	Offender: Offender neighborhood	Negative	Winter	Negative	-0,012
			Summer	Negative	
Property value	Target: Rational choice / Optimal foraging	Positive	Summer	Positive	0,008
			Autumn	Positive	
Distance rental properties	Offender: Offender neighborhood	Negative	Winter	Negative	0,006
			Summer	Positive	
Distance demographic risk group	Offender: Offender neighborhood	Negative	Winter	Negative	-0,005
			Autumn	Negative	
Distance retail and catering	Offender: Awareness space	Negative	Winter	Negative	-0,011
Election turnout	Setting: Social disorganization	Negative	Spring	Negative	-0,010
Distance city center	Offender: Awareness space	Negative	Winter	Positive	0,008
Rental properties	Target: Rational choice / Optimal foraging	Negative	Summer	Positive	0,008
	Setting: Social disorganization	Positive			
Accessibility	Offender: Awareness space	Positive	Spring	Positive	0,007
	Setting: Environmental design	Negative			
Welfare benefits	Target: Rational choice / Optimal foraging	Negative	Winter	Negative	-0,006
Distance to street	Setting: Environmental design	Positive	Spring	Positive	0,006
Crime	Setting: Social disorganization	Positive	Winter	Positive	0,004
Edge dwellings	Setting: Environmental design	Positive	Winter	Positive	0,003

Figure 5.23: Summary of the significant risk factors for burglaries ordered by the number of seasons where they are significant and their average standardized beta coefficients.

Risk factor	Type	Theory
Household income	Target	Rational choice / optimal foraging
Cars per household	Target	Rational choice / optimal foraging
Distance public facilities	Offender	Awareness space
Distance public transport node	Offender	Awareness space
Distance highway entry	Offender	Awareness space
Distance low incomes	Offender	Offender neighborhood
Ethnic heterogeneity	Setting	Social disorganization
Residential mobility	Setting	Social disorganization
Nuisance	Setting	Social disorganization
Construction year	Setting	Environmental design
Distance street lighting	Setting	Environmental design
Mixed land use	Setting	Environmental design
Distance shrubbery	Setting	Environmental design
Distance police station	Setting	Environmental design

Figure 5.24: An overview of the risk factors that were not significant variables in explaining the spatial patterns of burglary patterns in Haarlem.

5.2.2 Risk factor types

Three types of risk factor are defined in this research: target, offender and setting risk factors. From all risk factors that are significant in at least one season, 3 are target risk factors (excluding *building density*), 7 are offender risk factors and 7 are setting risk factors (including *building density*). From this it seems that the target risk factors are somewhat underrepresented. But when looking at the more consistent risk factors across all seasons, i.e. the risk factors that are significant in at least two different seasons, a more nuanced picture appears. From those risk factors, 2 are target, 4 are offender and 2 are setting risk factors. So to conclude, the assumption that there are three types of risk factor responsible for burglaries seems to hold true, although there seem to be some seasonal effects in play as well.

5.2.3 Seasonal effects

As mentioned in the previous sections, there are seasonal differences in the risk factors and types of risk factor that explain the burglary rates. Figure 5.25 presents an overview of the significant risk factors per season. Figure 5.26 shows the risk factor types per season. In general, most burglaries occur in winter (30% of the total number of burglaries in all seasons). As mentioned earlier, this could be because of less hours of daylight and lower temperatures, reducing visibility and surveillance of the public space. Moreover, holidays like Christmas and New Year's Eve create opportunities for burglars. As this research cannot give a definitive answer, further (qualitative) research is needed to test these hypotheses.

Of the three most consistent risk factors, *risky properties* and *building density* seem to follow a similar pattern throughout the seasons. They both have the most explanatory power during spring. *Distance ethnic heterogeneity* however seems to follow a different pattern. During winter and spring it is the weakest explanatory variable of the three, but during summer it explains almost as much of the burglaries as *building density* and during autumn it is even the strongest explanatory variable of burglaries. Why living nearer to areas with high ethnic heterogeneity is particularly risky during autumn is unclear and could be subject to further research.

There are plenty examples where risk factors are only significant explanatory variables of burglaries during one or two seasons (see figure 5.25). The number of significant risk factors is highest in the winter. Of the 11 risk factors 2 are target, 6 are offender and 4 are setting risk factors. Especially during the winter, when most burglaries occur, there are relatively many 'unique' risk factors, of which many are offender-related. Living in or close to offender neighborhoods seems to be especially risky during winter.

There are 6 significant risk factors in the spring: 1 target, 2 offender and 4 setting risk factors. It stands out that there are some setting-related risk factors which are only significant during this season. These are *accessibility*, *distance to street* and *election turnout*. It seems that factors influencing the perceived risks for offenders are more important explanatory factors during spring than during the other seasons.

There are also 6 significant risk factors for the summer: 1 target, 3 offender and 2 setting risk factors. All offender risk factors are related to offender neighborhoods, implying that the risk of burglary in the summer is greatly influenced by living in or near offender neighborhoods.

For the autumn 5 significant risk factors are found: 2 target, 2 offender and 2 setting risk factors. During autumn, it stands out that target risk factors play a relatively large role in explaining burglary patterns.

The differences in risk factor types per season can also hint towards different types of offenders which are active during different seasons. Earlier in this research, a distinction was made between the rational and irrational offender where the rational offender has a predefined goal and the irrational offender acts on opportunity and instinct (see section 4.1.1). Target risk factors are usually more associated with rational offenders, where offender and setting risk factors are more associated with irrational offenders.

For example, it seems that due to the larger share of target risk factors that are significant explanatory variables for burglaries during autumn, this season is more popular with rational offenders and for 'planned burglaries'. It should be emphasized though that this is simply a hypothesis based on the findings from the analysis of the Haarlem data and further research is needed to prove or disprove this hypothesis in a broader context. The same can be said of the influence of offender neighborhoods on the spatial patterns of burglary rates in both the winter and the summer.

Winter	Type	Spring	Type	Summer	Type	Autumn	Type
Distance ethnic heterogeneity	Offender	Distance ethnic heterogeneity	Offender	Distance ethnic heterogeneity	Offender	Distance ethnic heterogeneity	Offender
Building density	Setting	Building density	Setting	Building density	Setting	Building density	Setting
Risky properties	Target Setting	Risky properties	Target Setting	Distance welfare benefits	Offender	Risky properties	Target Setting
Distance welfare benefits	Offender	Election turnout	Setting	Property value	Target	Property value	Target
Distance rental properties	Offender	Accessibility	Offender	Distance rental properties	Offender	Distance demographic risk group	Offender
Distance demographic risk group	Offender	Distance to street	Setting	Rental properties	Setting		
Distance retail and catering	Offender						
Distance city center	Offender						
Welfare benefits	Target						
Crime	Setting						
Edge dwellings	Setting						

Figure 5.25: An overview of the significant risk factors and their underlying theories per season.

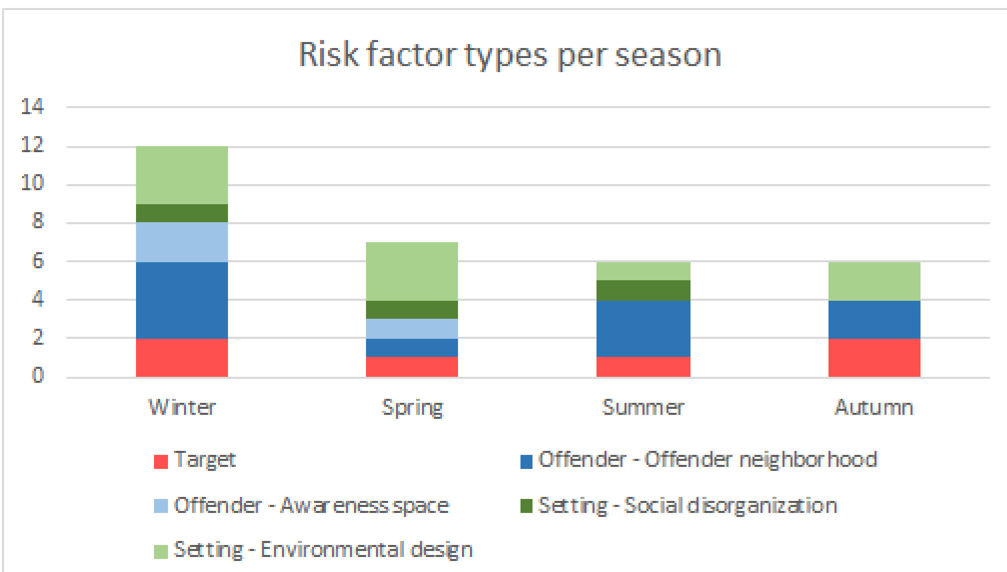


Figure 5.26: The significant risk factors grouped by risk factor type and underlying theory per season.

5.3 Conclusion

This chapter focused on answering the research subquestion: *what are the results of the spatiotemporal analysis of residential burglaries?*

The first section presented the situation regarding the spatial patterns of burglaries by showing burglary rates and hot spot maps. After that, the model building process was discussed aiming at modeling the causal relationship between different risk factors and burglaries. Meanwhile, issues like multicollinearity and spatial autocorrelation were assessed and were accounted for where needed. Multicollinearity did not turn out to be an issue as the correlation between the explanatory variables never exceeded the critical value of 0.9. Spatial autocorrelation was however present in the residuals of the initial model and were corrected for by adding a spatial lag variable to come to a final model per season. By adding this variable, the spatial autocorrelation was successfully removed in the models for most seasons, except for the winter model.

The first section concluded with a risk terrain surface, which mapped the risk of burglaries per grid cell based on the significant risk factors from the seasonal models. The persistent hot spot in the southeastern district called *Schalkwijk* can largely be explained by the presence of multiple offender risk factors. Many characteristics of an offender neighborhood are found in *Schalkwijk*: many people receiving welfare benefits, a high ethnic heterogeneity, many people in the demographic risk group (males aged 15 to 24) and a high percentage of rental properties.

The risk terrain models were subsequently compared with the hot spot maps of the actual burglaries to identify areas where burglary rates might increase in the future due to favorable conditions for burglaries. Potential future hot spots, based on a combination of several risk factors, can be found in the neighborhoods *Slachthuisbuurt*, *Spaarndam* and *Koninginnebuurt*.

The second section looked at the results of the spatiotemporal analysis and compared these with the hypotheses that resulted from the prior literature study. The three most consistent explanatory variables of burglaries turned out to be *distance ethnic heterogeneity*, *building density* and *risky properties*. From the 30 potential risk factors for burglaries, 16 are significant risk factors in at least one season and for 14 risk factor no significant causal relationships were found with the burglary rate. All three risk factor types (target, offender and setting) are important in explaining the spatial patterns of burglary rates. But there are also some variations between the different seasons, for example the more prominent role of target risk factors in the autumn and winter that might hint at a higher activity of rational offenders during these months when compared to the spring and summer.

Ultimately, the kind of insights acquired from the seasonal breakdown of burglary patterns and their explanatory variables can help decision makers develop better measures and policies against burglaries. It becomes clear which risk factors are responsible for certain spatial patterns in certain seasons and which specific areas in Haarlem are at increased risk of (future) burglaries. The role of this type of analysis in the context of smart cities is discussed in the concluding chapter: conclusion and discussion.

Chapter 6

Conclusion and discussion

At the beginning of this research a goal was formulated: to assess the practical usefulness of smart city concepts by finding a suitable method for explaining the structural spatiotemporal patterns of burglaries and applying this method in practice.

To help achieve this goal a main research question and four subquestions were formulated. The first section of this concluding chapter looks back at these subquestions and summarizes the most important findings and outcomes. Based on these findings and outcomes the main research question is answered.

The second section of this chapter involves a discussion of the findings and outcomes of this research. The goal is to put this research into a broader perspective by looking at the research process, possible improvements, generalizability of the results and potential areas, subjects and recommendations for further research.

6.1 Conclusion

6.1.1 Smart cities

The first research subquestion was: *what is a smart city and what are the underlying theoretical concepts?* It turned out that it is very difficult to pinpoint exactly what a smart city is. For this research the definition by Caragliu et al. (2011) was used:

> "We believe a city to be smart when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance" (p. 70).

The smart city is not a new concept on its own, but is better described as an evolution of existing theories for managing sustainable urban growth with a central role for ICT. It takes ideas from 'creative cities', 'business-oriented cities', 'green cities', 'wired cities' and more and combines them under the banner of 'smart cities'. Therefore, the term 'smart city' can be misleading as it does not necessarily represent something entirely new. It is likely that the term caught on as cities and local governments used the term for promotion and city marketing purposes. They like to be seen as modern and innovative in their competition to attract businesses, and social and human capital.

But simply because the term 'smart city' is used as a buzzword, does not automatically mean that there is no value in the concepts it represents. Two elements of the smart city that were further explored are smart governance and smart living.

Smart governance is aimed at using ICT as a means of increasing accountability, transparency and ultimately efficiency of governmental organizations. Smart governance includes collaborating and communicating efficiently both internally between departments and externally with citizens and stakeholders. Practical examples of smart governance can be found by looking at the increasing amount of data published as open data, horizontal and vertical integration of governmental departments through the concept of e-government and the development of a system of key registers in the Netherlands: a crucial part of the national spatial data infrastructure.

Smart living is about maximizing the quality of life of citizens by utilizing ICT. Smart living covers a broad area of subjects that can be related to quality of life, like environment and pollution, housing costs and access, health care and public health, education provision and levels, and art and cultural diversity. The focus here is on the aspect of crime and public safety. Two relevant concepts were introduced related to smart living and public safety: problem-oriented policing and predictive policing. In short, problem-oriented policing advocates a pro-active strategy against crime by focusing on longterm analysis of enduring problems causing crime. Predictive policing applies mainly quantitative techniques to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions.

There is also critique on the smart city and its concepts. These mainly focus on the term 'smart city' being used for place marketing and promotion, making it a hollow term, and on violations of privacy as smart cities generally involve gathering, storing and combining large sets of data, including personal data.

6.1.2 Method

The second research question was: *what methods are available for the spatiotemporal analysis of crime data?* Three broad categories were investigated: hot spot and near repeat methods, grid and raster methods, and univariate and multivariate regression methods. Within each of these categories several specific methods were discussed including their pros and cons. As a result, for this research multivariate regression in combination with grid and raster methods is used to analyze spatiotemporal patterns of burglaries. These methods are suitable for finding the structural causes of crime and include significance tests to assess and quantify the explanatory power of multiple independent variables. Moreover, grid and raster methods are strong in the interpretation and communication of the results.

As there are multiple multivariate regression methods available, three of the most common in crime analysis were discussed: linear regression, poisson regression and negative binomial regression. Negative binomial is selected because this method respects the specific nature of crime data, it allows for dependencies among individual crime events and it is supported by good results from other researches involving the analysis of crime data.

To help the interpretation and presentation of the results, the raster method of 'risk terrain modeling' is used. Risk terrain maps assist in strategic decision making and tactical action by showing where conditions are ideal for events to occur in the future. Separate map layers representing the presence, absence, or intensity of each significant risk factor at every place throughout a terrain is created, and then all map layers are combined using weights determined by the regression analysis to produce a composite 'risk terrain' map with attribute values that account for all risk factors at every place throughout the geography.

In short, the method used for the spatiotemporal analysis of crime data involves the following steps: identifying the potential risk factors; defining the study area and operationalizing the risk factors; building the negative binomial regression model while keeping in mind issues like multicollinearity, spatial autocorrelation and seasonality; and finally creating a seasonal risk terrain surface.

6.1.3 Risk factors

The third research question was: *what are the possible risk factors for residential burglaries?* To identify the risk factors, the assumption is made that crime events can be traced back to the combination of three components: a target, an offender and the setting. For each of these components, theories from spatial criminology were consulted.

For the target component this is the theory of 'rational choice' or 'optimal foraging'. These theories assume that a target is chosen based on the goal of maximization of the profit. These theories can be related to the 'rational offender': burglars who prepare their offenses with a predefined goal in mind.

For the offender component two theories are discussed: 'awareness space' and 'offender neighborhood'. The awareness space theory assumes that burglars tend to commit their offenses in areas that are familiar to them: areas within their 'routine activity space'. The theory about offender neighborhoods is used to identify the likely living areas of offenders based on the characteristics of these neighborhoods.

Finally, for the setting component theories about 'social disorganization' and 'environmental design' are considered. The theory about social disorganization assumes that crime and burglaries are more likely to be committed in areas where there is a lack of social cohesion and social control, resulting in residents who do not feel connected to their neighborhood and feel less responsible for their living environment. The theory about environmental design assumes that the spatial layout and design of the living environment contributes to the risk of crime and burglaries.

Based on the three components of crime and the associated theories an extensive list of potential risk factors is created based on the available data sources (see table 4.4).

6.1.4 Results

The fourth and final research subquestion is: *what are the results of the spatiotemporal analysis of residential burglaries?* Four negative binomial regression models are created: one for each season. Multicollinearity did not turn out to be above the predefined critical level, but there are some variables with relatively high bivariate correlation values. Spatial autocorrelation is corrected for in the models by adding a spatial lag variable. This removed all spatial autocorrelation from the models, with the exception of the winter.

Based on the resulting regression models an overview of the significant risk factors per season can be created, including the direction of the causal relationship and the standardized beta coefficients (see for a summary of the significant risk factors figure 6.1). From the 30 risk factors that entered the model building process, 16 risk factors turned out to be significant in explaining variations in burglary rates in at least one season. The most consistent risk factors, meaning that these risk factors are significant in at least three seasons, are: the distance to areas with high ethnic heterogeneity, the building density within an area and the share of risky properties (detached, semi-detached and corner properties) in an area. The found directions of the relationships were in most cases in correspondence with the hypotheses.

The assumption that there are three types of risk factor (target, offender and setting) responsible for burglaries seems to hold true, although there seem to be some seasonal effects in play as well (see figure 6.2). Most burglaries occur during the winter. This could be because of less hours of daylight and lower temperatures, reducing visibility and surveillance of the public space. Moreover, holidays like Christmas and New Year's Eve create opportunities for burglars. The significant risk factors and the underlying theories differ from season to season (see again figure 6.2). These differences in risk factor types per season can hint towards different types of offenders which are active during different seasons. Target risk factors are usually more associated with rational offenders, where offender and setting risk factors are more associated with irrational offenders. For example, it seems that due to the larger share of target risk factors that are significant explanatory variables for burglaries during autumn, this season is more popular with rational offenders and for 'planned burglaries'. It should be emphasized though that this is simply a hypothesis based on the findings from the analysis of the Haarlem data and further research is needed to prove or disprove this hypothesis in a broader context. The same can be said about the influence of offender neighborhoods on the spatial patterns of burglary rates in both the winter and the summer.

Risk terrain maps were created for each season by combining the risk layers of the significant risk factors and weighing each layer based on its standardized beta coefficient. The persistent hot spot in the southeastern district called *Schalkwijk* can largely be explained by the presence of multiple offender risk factors. Many characteristics of an offender neighborhood are found in *Schalkwijk*: many people receiving welfare benefits, a high ethnic heterogeneity, many people in the demographic risk group (males aged 15 to 24) and a high percentage of rental properties. Potential future hot spots, based on a combination of several risk factors, can be found in the neighborhoods *Slachthuisbuurt*, *Spaarndam* and *Koninginnebuurt*.

Risk factor	Type and theory	Relationship direction hypothesis	Significant in	Relationship direction	Average standardized beta coefficient
Distance ethnic heterogeneity	Offender: Offender neighborhood	Negative	Winter	Negative	-0,009
			Spring	Negative	
			Summer	Negative	
			Autumn	Negative	
Building density	Target: Rational choice / Optimal foraging	Positive	Winter	Negative	-0,008
			Spring	Negative	
			Summer	Negative	
			Autumn	Negative	
Risky properties	Target: Rational choice / Optimal foraging Setting: Environmental design	Positive	Winter	Positive	0,010
			Spring	Positive	
			Autumn	Positive	
Distance welfare benefits	Offender: Offender neighborhood	Negative	Winter	Negative	-0,012
			Summer	Negative	
Property value	Target: Rational choice / Optimal foraging	Positive	Summer	Positive	0,008
			Autumn	Positive	
Distance rental properties	Offender: Offender neighborhood	Negative	Winter	Negative	0,006
			Summer	Positive	
Distance demographic risk group	Offender: Offender neighborhood	Negative	Winter	Negative	-0,005
			Autumn	Negative	
Distance retail and catering	Offender: Awareness space	Negative	Winter	Negative	-0,011
Election turnout	Setting: Social disorganization	Negative	Spring	Negative	-0,010
Distance city center	Offender: Awareness space	Negative	Winter	Positive	0,008
Rental properties	Target: Rational choice / Optimal foraging	Negative	Summer	Positive	0,008
	Setting: Social disorganization	Positive			
Accessibility	Offender: Awareness space	Positive	Spring	Positive	0,007
	Setting: Environmental design	Negative			
Welfare benefits	Target: Rational choice / Optimal foraging	Negative	Winter	Negative	-0,006
Distance to street	Setting: Environmental design	Positive	Spring	Positive	0,006
Crime	Setting: Social disorganization	Positive	Winter	Positive	0,004
Edge dwellings	Setting: Environmental design	Positive	Winter	Positive	0,003

Figure 6.1: Summary of the significant risk factors for burglaries ordered by the number of seasons where they are significant and their average standardized beta coefficients.

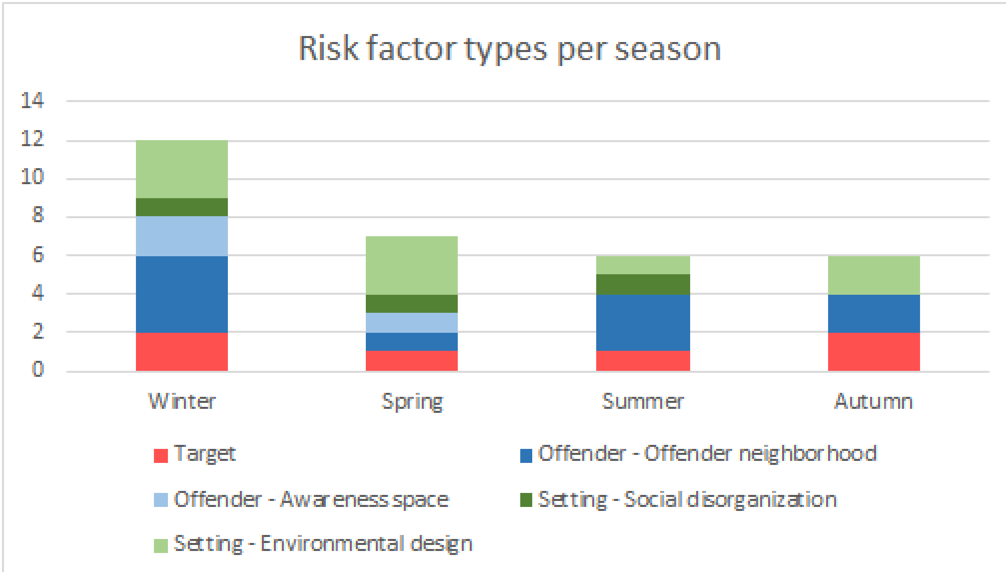


Figure 6.2: The significant risk factors grouped by risk factor type and underlying theory per season.

6.1.5 Main research question

The main research question was: *how can the application of smart city concepts help improve public safety by decreasing burglary rates?* The answer to this question is threefold. The application of smart city concepts can help decrease burglary rates by:

1. facilitating the efficient spatiotemporal analysis of burglaries,
2. motivating city and local governments to implement advanced analysis methods to support decision-making, and
3. promoting smart cities and related concepts by generating attention.

Smart city as facilitator

Especially the characteristics of smart governance, like the ambition to vertically and horizontally integrate datasets through spatial data infrastructures and to publish this data as open data, provide the necessary data-related conditions for a successful analysis of spatiotemporal patterns of burglaries.

The vertical and horizontal integration of datasets about demographics, socioeconomic statistics, topographical objects and more, allows researchers to combine datasets and to possibly find causal relationships that could not be found otherwise. The case study in Haarlem performed for this research certainly demonstrated how crime analysis can benefit from a spatial data infrastructure. Such a data infrastructure makes it easy to link different datasets to each other: for example relating property values to addresses and addresses to residents to allow for more complex relationships to be modeled. It facilitates the standardized geoprocessing of spatial data.

Moreover, the integration of datasets across different levels and between different departments of government, forces these datasets to be standardized, structured, complete and up-to-date together with the storage of valuable metadata. In short, smarter governance and the development of SDIs can lead to higher quality data and it is obvious how higher quality data can benefit all types of spatiotemporal analysis, including those of burglaries and other crime events.

And because these integrated and high quality datasets are increasingly published as open data, anyone has access to it. In this way value can be added to the data and it could potentially spur economic development as opportunities arise for the development of new analysis methods and reporting tools. Even individual enthusiasts have the opportunity to tinker around with the datasets and find innovative ways to apply the data. All in all, opening up datasets to the public can improve analysis methods, find innovative new analyses, and increase the general involvement of businesses and individuals in the improvement of the quality of life. As there is power in numbers, giving access to open datasets to the public can only benefit analysis methods and provide new insights.

To conclude, the case study in Haarlem demonstrated how a smarter governance can benefit the spatioanalysis of burglaries by providing integrated and high quality data sets and make these accessible as open data through spatial data infrastructures. The case study would not have been possible without these aspects of smart cities. In this light, the smart city can be seen as facilitator.

Smart city as motivator

Smart cities can motivate city and local governments to look at cities and their management differently and to discover new possibilities to support their decision making processes. The concept of smart living shows how information and communication technology can play an important role in solving or mitigating issues related to for example health, education, environment and of course public safety. Concepts like problem-oriented policing and predictive policing fit very well within this idea, with their comprehensive data collection and analyses and their aim on structural causes of crime and making forecasts of future incidents. An example of such an analysis is provided in this research, but there are also other examples like the *Criminaliteits Anticipatie Systeem* (CAS; Crime Anticipation System) developed in Amsterdam. Knowing when and where crime risk factors coincide can help focus the allocation of sparse police resources and better inform decision making processes of city and local governments.

Therefore, the smart city concept is seen as a motivator for cities and local governments to look at existing urban issues differently by demonstrating the power of modern information and communication technologies in for example massive and complex calculations, data mining, and analysis; which helps in the automatic discovery of patterns, rules and knowledge, and provides remote monitoring, control and feedback to the real world for intelligent city management and informed decision-making.

Smart city as promoter

A much heard critique is that 'smart city' is just a buzzword without any significance or real meaning. Although this might be true in some occasions and to some extent, the fact that the term is seen as a buzzword implies that smart cities apparently created a buzz: it drew attention and created publicity, and notably not only in the academic world. And that is exactly why the smart city as a buzzword might not be such a bad thing after all. It generates awareness among city and local governments, as well as among businesses and citizens. Therefore, the 'smart city' has the potential to reach a larger public by informing people about the potential benefits, draw in funds and subsidies and to actually realize smart city concepts. Obviously, it is important to acknowledge that the term smart city can evoke negative reactions, for example concerning the privacy of citizens, so it is important to address how the privacy of citizens can be respected while focusing on the intention of improving the quality of life.

That is why the smart city can be seen as a promoter and generator of attention and publicity which, when used 'smartly', can help to draw positive publicity and funds to promote smart city concepts from the drawing board to reality.

To conclude, the application of smart cities can ultimately help to decrease burglary rates by looking at the smart city as facilitator, motivator and promoter. Meanwhile, the case study of the spatiotemporal analysis of burglaries in Haarlem also demonstrated the added value of specific smart city concepts by increasing accountability, transparency and effectiveness of city and local governments. This also shows that being a smart city should never be the ultimate goal, but that being a smart city can ultimately result into more efficient and effective management of a city or local government.

6.2 Discussion

This final section looks back at the research process to identify potential improvements and to assess to what extent the findings and results are generalizable in other study areas. It also looks at the future by discussing the potential areas, subjects and recommendations for further research.

One of the goals of this research was to find a suitable method for explaining the structural spatiotemporal patterns of burglaries and to apply this method in practice. Several lessons can be learned from this.

One issue that is common in spatial analysis is spatial autocorrelation. This was dealt with in this research by adding a spatial lag variable to the regression model. While the results were mainly positive, there was still some spatial autocorrelation present in the final model for the winter. An alternative method to correct for spatial autocorrelation was also discussed and is called 'eigenvector spatial filtering'. As other researches show promising results with this method, it would have been interesting to see if it would outperform the spatial lag method in the Haarlem case study. Unfortunately, eigenvector spatial filtering is topology-based, meaning that all of the spatial units should be adjacent to each other. This was not the case for Haarlem, as cells without address locations were removed from the study area. It does not make sense to include spatial units where no outcome events, i.e. burglaries, could ever be recorded. Including these cells would have resulted in a strongly skewed data set, hampering the spatiotemporal analysis.

A possible solution to this type of problem would be to develop a standard dataset of spatial units specifically designed for social research purposes, i.e. for studies that involve data that is related to persons.

This should be a topology-based set of spatial units ideally covering a large administrative area, like a municipality, country or possibly an even larger administrative area. The spatial units are ideally as small as possible to enable the mapping of detailed spatial patterns. To ensure privacy, the borders of the spatial units can be drawn in such a way that it includes a predefined minimum number of inhabitants. This prevents data to be related to individuals directly, but also prevents gaps in the study area because there can by definition be no areas without any inhabitants. Spatial units would be smaller in populated areas and larger in unpopulated areas to ensure the same level of detail between spatial units.

Such a standard set of spatial research units can allow for easier comparison of data patterns through time or between different studies. Ideally, these spatial research units are developed and maintained by a governmental organization that publishes it as open data to allow anyone to use it. An example of how such a standard set of spatial units for social research purposes might look like is included in figure 6.3.

Another learned lesson is related to the near repeat theory. The assumption of this theory is that burglaries occur closer to each other in both space and time than can be expected based on chance. To include this effect, a spatial lag variable was calculated based on a spatial weights matrix where the area of influence of a burglary was set to 400 meters.

This parameter is an approximation based on a study of literature. It could be that the area of near repeats is actually larger or smaller in the specific case of Haarlem. That is why it is recommended for future research to calculate the area of near repeats prior to the modeling process to verify the area of influence and to potentially calculate a more accurate spatial lag variable.

The spatiotemporal analysis of burglaries performed in this research used data from the municipality of Haarlem. But to what extent are the findings and results generalizable to for example other municipalities in The Netherlands or to other countries?

The significant risk factors for burglaries found in Haarlem cannot automatically be generalized to other municipalities or cities. However, it can be hypothesized that the most consistent risk factors for explaining spatiotemporal patterns of burglaries in Haarlem, also play their role in explaining burglary patterns in comparable Dutch municipalities. Comparable here refers to municipalities with the same urban character as Haarlem and about the same number of inhabitants. But in general, the significant risk factors and their relationship directions and explanatory power, are likely to vary considerably between different Dutch municipalities, most certainly when more rural and less populated Dutch municipalities are considered. The risk factors for explaining burglaries in other countries are even more likely to deviate from the significant risk factors found in Haarlem.

But although the findings and results cannot easily be generalized to other areas, the method used for the spatiotemporal analysis can. This method can be seen as a form of standardized geoprocessing, described by Kiehle et al. (2006) as the next step of SDIs. The input data that was used was mostly derived from data within the Dutch system of key registers. As this system of key registers is maintained and accessible nation wide, all Dutch municipalities can use the same input data as Haarlem. And because of the structured and uniform character of the input data, meaning that the input data always comes in the same format, it is possible to fully automate the method used in this research for the spatiotemporal analysis of burglaries. This could make the analysis of burglaries, and potentially other types of crime, very accessible to all municipalities in the Netherlands.

For other countries, the method is still relevant but automation of the analysis process is only possible if the input data, meaning data about possible risk factors, is accessible as standardized datasets through a national spatial data infrastructure. The concepts from the smart city could help stimulate the development of these national spatial data infrastructures, by using the smart city as motivator and promoter (see section 6.1.5), to facilitate the standardized geoprocessing of spatial data.

Finally, some suggestions can be made for areas or subjects for future research.

The spatiotemporal analyis of burglaries conducted for this research focused on finding the structural causes of burglaries including seasonal variations. This was used to identify areas of elevated risk in the municipality of Haarlem per season. The analysis of spatial patterns of burglaries based on smaller time intervals, for example months, weeks, days or even hours, would be a valuable extension to this research and for decision-makers. Knowing which risk factors cause the temporal variations from hour to hour or from day to day, could allow for more detailed predictions. Combining structural risk factors with short-term risk factors like weather conditions or holidays, can provide hourly predictions of burglaries or other types of crime. This type of (almost) live monitoring of crime allows the police to know where and when to best deploy their scarce resources.

For this type of live monitoring of crime it is essential that the time of a crime event is recorded with as much accuracy as possible, as opposed to only registering the time a crime is reported to the police. For burglaries this can be difficult, as it is often hard to pinpoint the exact time a burglary occurred. When the exact time of a crime event is uncertain, recording a likely time interval is a good alternative. Without this data it is very hard to accurately model crime patterns with small time intervals.

Finally, a standardized method for geoprocessing crime data, like the method applied in Haarlem, can potentially be used in the field of urban planning as 'Planning Support System' (PSS). Geertman and Stillwell (2003) state that: "*Planning Support Systems involve a wide diversity of geo-technology tools (geographical information and spatial modelling systems) that have been developed to support public or private planning processes (or parts thereof) at any defined spatial scale and within any specific planning context*" (p. 5). PSS are computer-based tools that planners can use to enhance their analytical, problem-solving and decision making capabilities.

The spatiotemporal analysis of burglaries fits very well with the characteristics of PSS which include: data collection, spatial and trend analysis, data modelling, prediction and prescription, visualisation and display, and the transformation of basic data into information. Furthermore, PSS is specific and customized to focus on the task it is designed for. It can be described as a task-specific system. Finally, PSS pays specific attention to long-range problems and strategic issues (Geertman and Stillwell, 2003).

It would therefore be interesting to see if the method for the spatiotemporal analysis of burglaries presented in this research, can be further developed to be used as a planning support system. To be usable to decision makers the method applied in Haarlem should be extended to include report preparation, collaborative decision-making and scenario building. In the future smart city, multiple scenarios for new urban development could be assessed by a standardized analysis of the risk of burglaries: a spatiotemporal analysis identifies the relevant risk factors for burglaries in a certain locality and provides areas of elevated risk based on a culmination of risk factors. The outcomes are summarized in a clear report that can be used directly by decision makers to score different scenarios or to adjust spatial plans.

But although the current method used in Haarlem is far from ready to be used as PSS, it certainly has the potential to better support decision makers in their activities. It is just another example of how modern information and technology can continue to make cities smarter, by allowing the integration of and access to datasets and complex analysis methods, to better understand and predict patterns in social phenomena like crime, to support decision making and make governance more efficient and to ultimately improve the quality of life of citizens.

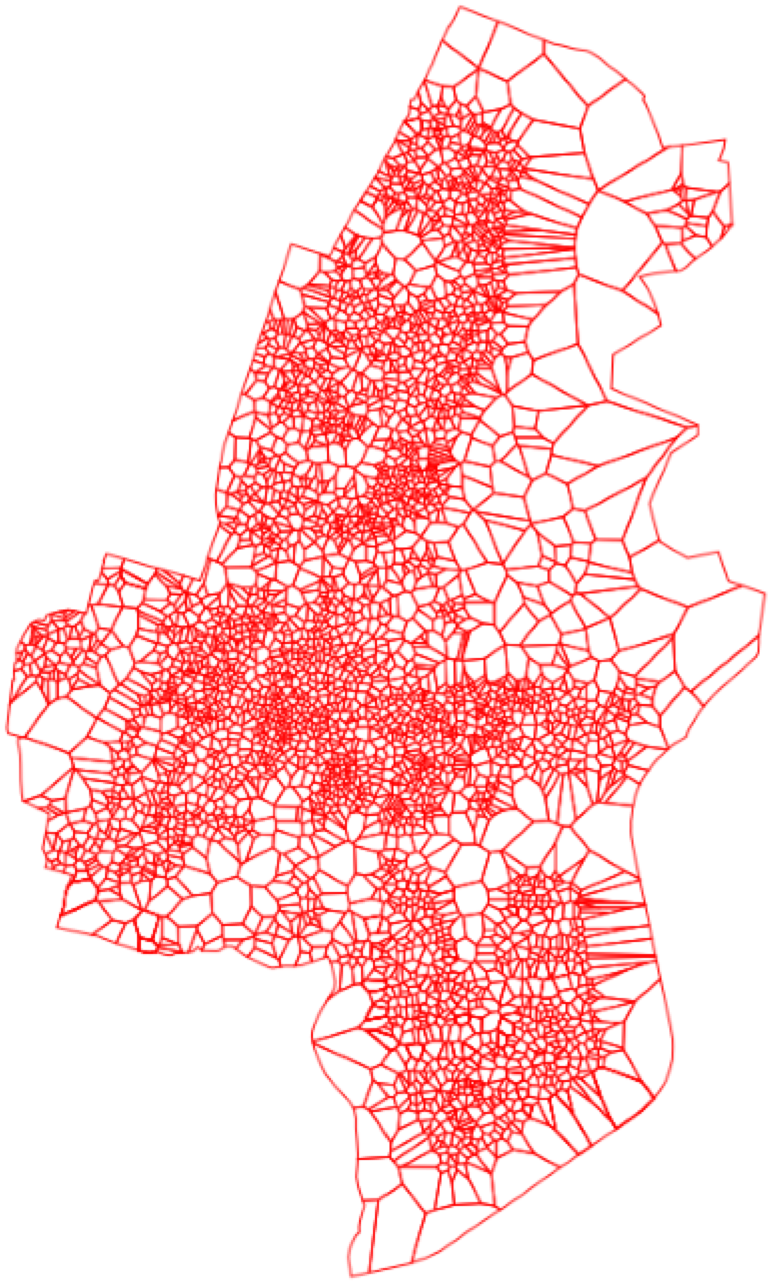


Figure 6.3: An example map of topology-based spatial research units of Haarlem. The units are based on postcode areas. Areas with more inhabitants have more postcodes and are therefore smaller. This allows more detailed mapping of spatial patterns in populated areas.

Bibliography

Akkermans, M., R. Kloosterman, K. Knoops, G. Linden, and E. Moons (2015). Veiligheidsmonitor 2014. Technical report, Centraal Bureau voor de Statistiek.

Anselin, L. and A. K. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs 155*, 237–290.

ASC (2015). About asc. `http://amsterdamsmartcity.com/about-asc`.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine 66*(3), 411–421.

Beine, M., F. Docquier, and H. Rapoport (2001). Brain drain and economic growth: theory and evidence. *Journal of development economics 64*(1), 275–289.

Bernasco, W. (2006). Co-offending and the choice of target areas in burglary. *Journal of Investigative Psychology and Offender Profiling 3*(3), 139–155.

Bernasco, W. (2010). Modeling micro-level crime location choice: Application of the discrete choice framework to crime at places. *Journal of Quantitative Criminology 26*(1), 113–138.

Bernasco, W. and F. Luykx (2003). Effects of attractiveness, opportunity and accessibility to burglars on residential burglary rates of urban neighborhoods. *Criminology 41*(3), 981–1002.

Bernasco, W. and P. Nieuwbeerta (2005). How do residential burglars select target areas? a new approach to the analysis of criminal location choice. *British Journal of Criminology 45*(3), 296–315.

Bettencourt, L. M., J. Lobo, D. Helbing, C. Kühnert, and G. B. West (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences 104*(17), 7301–7306.

Big Brother Awards (2015, October). Korpschef nationale politie genomineerd voor expertprijs vanwege 'predictive policing'. https://bigbrotherawards.nl/nl_NL/korpschef-van-de-nationale-politie-genomineerd-voor-expertprijs-vanwege-predictive-policing/.

Boba, R. (2003). *Problem analysis in policing*. Police Foundation Washington, DC.

Bowers, K. (1999). Exploring links between crime and disadvantage in north-west england: an analysis using geographical information systems. *International Journal of Geographical Information Science 13*(2), 159–184.

Bowers, K. J., S. D. Johnson, and K. Pease (2004). Prospective hot-spotting the future of crime mapping? *British Journal of Criminology 44*(5), 641–658.

Brantingham, P. and P. Brantingham (1995). Criminality of place. *European Journal on Criminal Policy and Research 3*(3), 5–26.

Brantingham, P. L. and P. J. Brantingham (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology 13*(1), 3–28.

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician 48*(3), 209–213.

Brown, B. B. and I. Altman (1983). Territoriality, defensible space and residential burglary: An environmental analysis. *Journal of environmental psychology 3*(3), 203–220.

Budhathoki, N. R., Z. Nedovic-Budic, et al. (2008). Reconceptualizing the role of the user of spatial data infrastructure. *GeoJournal 72*(3-4), 149–160.

Caplan, J. M. and L. W. Kennedy (2010). *Risk Terrain Modeling Manual: Theoretical Framework And Technical Steps Of Spatial Risk Assessment For Crime Analysis*. Rutgers Center on Public Security.

Caplan, J. M. and L. W. Kennedy (2011). Risk terrain modeling compendium. *Rutgers Center on Public Security, Newark*.

Caplan, J. M. and L. W. Kennedy (2014). Forecasting of shootings using risk terrain modeling. *Encyclopedia of Criminology and Criminal Justice*, 1682–1692.

Caplan, J. M., L. W. Kennedy, and E. L. Piza (2013). *Risk Terrain Modeling Diagnostics Utility User Manual (Version 1.0)*. Rutgers Center on Public Security.

Caragliu, A., C. Del Bo, and P. Nijkamp (2011). Smart cities in europe. *Journal of urban technology 18*(2), 65–82.

Chainey, S. and J. Ratcliffe (2005). *GIS and crime mapping.* John Wiley & Sons.

Chih-Feng, S. S. et al. (2000). Housing layout and crime vulnerability. *Urban Design International 5*(3-4), 3–4.

Chourabi, H., T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo, and H. J. Scholl (2012). Understanding smart cities: An integrative framework. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, pp. 2289–2297. IEEE.

Coe, A., G. Paquet, and J. Roy (2001). E-governance and smart communities a social learning challenge. *Social Science Computer Review 19*(1), 80–93.

Coupe, T. and L. Blake (2006). Daylight and darkness targeting strategies and the risks of being seen at residential burglaries*. *Criminology 44*(2), 431–464.

Cozens, P. M., G. Saville, and D. Hillier (2005). Crime prevention through environmental design (cpted): a review and modern bibliography. *Property management 23*(5), 328–356.

Crowe, T. D. and D. L. Zahm (1994). Crime prevention through environmental design. *Land Development magazine*, 22–27.

Datoo, S. (2014, April). Smart cities: are you willing to trade privacy for efficiency? `http://www.theguardian.com/news/2014/apr/04/if-smart-cities-dont-think-about-privacy-citizens-will-refuse-to-accept-change-says-cisco-chief`.

de Vocht, A. (2008). *Basishandboek SPSS 16.* Bijleveld Press.

de Vocht, A. (2009). *Syllabus Statistiek: Sociale Geografie en Planologie.* Faculteit Geowetenschappen, Universiteit Utrecht.

Deadman, D. (2003). Forecasting residential burglary. *International Journal of Forecasting 19*(4), 567–578.

Diniz-Filho, J. A. F. and L. M. Bini (2005). Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography 14*(2), 177–185.

Dodgson, M. and D. Gann (2011). Technological innovation and complex systems in cities. *Journal of Urban Technology 18*(3), 101–113.

Dormann, C. F., J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. D. Kissling, et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography 30*(5), 609–628.

E-overheid (2013, November). Haarlem zet overlast op de kaart. http://www.e-overheid.nl/actueel/i-nup-toppers/intItem/haarlem-zet-overlast-op-de-kaart/2164.

E-overheid (2014, November). Stelsel van basisregistraties. http://www.e-overheid.nl/onderwerpen/stelselinformatiepunt/stelsel-van-basisregistraties.

Eck, J., S. Chainey, J. Cameron, and R. Wilson (2005). Mapping crime: Understanding hotspots.

Ehrlich, I. (1975). On the relation between education and crime. In *Education, income, and human behavior*, pp. 313–338. NBER.

Esri (2013). How spatial autocorrelation (global moran's i) works. http://resources.arcgis.com/en/help/main/10.1/index.html#/How_Spatial_Autocorrelation_Global_Moran_s_I_works/005p0000000t000000/.

Farrar, D. E. and R. R. Glauber (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92–107.

Farrell, G. and K. Pease (1994). Crime seasonality: Domestic disputes and residential burglary in merseyside 1988–90. *British Journal of Criminology 34* (4), 487–498.

Flom, P. L. and D. L. Cassell (2007). Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. In *NorthEast SAS Users Group Inc 20th Annual Conference: 11-14th November 2007; Baltimore, Maryland*.

Florida, R. (2004). *The rise of the creative class and how it's transforming work, leisure, community and everyday life (Paperback Ed.)*. New York: Basic Books.

Gale, R. (2013). An application of risk terrain modeling to residential burglary. *TCNJ Journal of Student Scholarship 15*, 1–9.

Gardner, W., E. P. Mulvey, and E. C. Shaw (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological bulletin 118* (3), 392.

Geertman, S. and J. Stillwell (2003). Planning support systems: an introduction. In *Planning support systems in practice*, pp. 3–22. Springer.

Giffinger, R., C. Fertner, H. Kramar, R. Kalasek, N. Pichler-Milanovic, and E. Meijers (2007). Smart cities-ranking of european medium-sized cities. Technical report, Vienna University of Technology.

Gorr, W. and A. Olligschlaeger (2001). *Crime Hot Spot Forecasting: Modeling and Comparative Evaluation, Final Project Report*. US Department of Justice.

Greenfield, A. (2013). Against the smart city (the city is here for you to use). *Do Projects, New York City*.

Griffiths, D. N. (2013). City cynic: 'against the smart city' by adam greenfield (review). `http://www.forbes.com/sites/danielnyegriffiths/2013/12/02/city-cynic-against-the-smart-city-by-adam--review/`.

Groff, E. R. and N. G. La Vigne (2001). Mapping an opportunity surface of residential burglary. *Journal of Research in Crime and Delinquency 38*(3), 257–278.

Groff, E. R. and N. G. La Vigne (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies 13*, 29–58.

Gruen, A. (2013). Smart cities: The need for spatial intelligence. *Geo-spatial Information Science 16*(1), 3–6.

Guthery, F. S., L. A. Brennan, M. J. Peterson, and J. J. Lusk (2005). Invited paper: Information theory in wildlife science: Critique and viewpoint. *Journal of Wildlife Management 69*(2), 457–465.

Helbich, M. and J. Jokar Arsanjani (2014). Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartography and Geographic Information Science* (ahead-of-print), 1–15.

Heywood, I., S. Cornelius, and S. Carver (2006). *An Introduction to Geographical Information Systems*. Pearson Education.

Hielkema, H. and P. Hongisto (2013). Developing the helsinki smart city: the role of competitions for open data applications. *Journal of the Knowledge Economy 4*(2), 190–204.

Hillier, B. and O. Sahbaz (2007). Beyond hot spots; using space syntax to understand dispersed patterns of crime risk in the built environment. *(Proceedings) Conference on crime analysis at the Institute of Pure and Applied Mathematics*.

Hollands, R. G. (2008). Will the real smart city please stand up? intelligent, progressive or entrepreneurial? *City 12*(3), 303–320.

Huang, F. L. and D. G. Cornell (2012). Pick your poisson: A tutorial on analyzing counts of student victimization data. *Journal of School Violence 11*(3), 187–206.

Irvin, R. A. and J. Stansbury (2004). Citizen participation in decision making: is it worth the effort? *Public administration review 64*(1), 55–65.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. Vintage Books ed. Vintage Books.

Jenkins, M. J. (2014). Problem-oriented policing. *The Encyclopedia of Theoretical Criminology*.

Jiang, B. and C. Claramunt (2002). Integration of space syntax into gis: new perspectives for urban morphology. *Transactions in GIS 6*(3), 295–309.

Johnson, D. (2008). The near-repeat burglary phenomenon. In S. Chainey and L. Tompson (Eds.), *Crime Mapping Case Studies: Practice and Research*, Chapter 8, pp. 123–132. Wiley Online Library.

Johnson, M. P. (2001). Environmental impacts of urban sprawl: a survey of the literature and proposed research agenda. *Environment and Planning A 33*(4), 717–735.

Johnson, S. D., W. Bernasco, K. J. Bowers, H. Elffers, J. Ratcliffe, G. Rengert, and M. Townsley (2007). Space–time patterns of risk: a cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology 23*(3), 201–219.

Kawachi, I., B. P. Kennedy, and R. G. Wilkinson (1999). Crime: social disorganization and relative deprivation. *Social science & medicine 48*(6), 719–731.

Kiehle, C., K. Greve, and C. Heier (2006). Standardized geoprocessing–taking spatial data infrastructures one step further. In *Proceedings of the 9th AGILE International Conference on Geographic Information Science. Visegrád, Hungary.*

Klein Haneveld, R., S. Boes, and N. Kop (2012). Woninginbraken: Een onderzoek naar het fenomeen woninginbraken en mogelijke aanpak hiertegen. Technical report, Politieacademie.

Lam, W. (2005). Barriers to e-government integration. *Journal of Enterprise Information Management 18*(5), 511–530.

Layne, K. and J. Lee (2001). Developing fully functional e-government: A four stage model. *Government information quarterly 18*(2), 122–136.

Li, D., J. Shan, Z. Shao, X. Zhou, and Y. Yao (2013). Geomatics for smart cities-concept, key techniques, and applications. *Geo-spatial Information Science 16*(1), 13–24.

Liu, H. and D. E. Brown (2003). Criminal incident prediction using a point-pattern-based density model. *International journal of forecasting 19*(4), 603–622.

Lynn, L. E., C. J. Heinrich, and C. J. Hill (2000). Studying governance and public management: Challenges and prospects. *Journal of Public Administration Research and Theory 10*(2), 233–262.

Maguire, M. (1980). Impact of burglary upon victims, the. *Brit. J. Criminology 20*, 261.

Malczewski, J. and A. Poetz (2005). Residential burglaries and neighborhood socioeconomic context in london, ontario: Global and local regression analysis. *The Professional Geographer 57*(4), 516–529.

Marche, S. and J. D. McNiven (2003). E-government and e-governance: The future isn't what it used to be. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration 20*(1), 74–86.

Martinez-Balleste, A., P. A. Pérez-Martínez, and A. Solanas (2013). The pursuit of citizens' privacy: a privacy-aware smart city is possible. *Communications Magazine, IEEE 51*(6).

Ministerie van Infrastructuur en Milieu (2014). Smart cities. `http://www.platform31.nl/uploads/media_item/media_item/36/22/Smart_Cities_krant-1422533241.pdf`.

Mooij, J. E. (2003). *Smart Governance?: Politics in the Policy Process in Andhra Pradesh, India.* Overseas development institute (ODI).

Moreto, W. D., E. L. Piza, and J. M. Caplan (2013). a plague on both your houses?: Risks, repeats and reconsiderations of urban residential burglary. *Justice Quarterly* (ahead-of-print), 1–25.

Moss Kanter, R. and S. S. Litow (2009). Informed and interconnected: A manifesto for smarter cities. *Harvard Business School General Management Unit Working Paper* (09-141).

Nam, T. and T. A. Pardo (2011). Conceptualizing smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, pp. 282–291. ACM.

NOS (2013, November). Burgers in actie tegen inbraken. `http://nos.nl/artikel/578125-burgers-in-actie-tegen-inbraken.html`.

NOS (2014, October). Vinexwijk trekt inbrekers. `http://nos.nl/artikel/707492-vinexwijk-trekt-inbrekers.html`.

Nubani, L. and J. Wineman (2005). The role of space syntax in identifying the relationship between space and crime. In *Proceedings of the Fifth International Space Syntax Symposium.*

NU.nl (2015, October). Plasterk en nationale politie winnen big brother awards. `http://www.nu.nl/internet/4153994/plasterk-en-nationale-politie-winnen-big-brother-awards.html`.

Odendaal, N. (2003). Information and communication technology and local governance: understanding the difference between cities in developed and emerging economies. *Computers, Environment and Urban Systems 27*(6), 585–607.

Olligschlaeger, A. M. (1997). Artificial neural networks and crime mapping. In D. Weisburd and T. McEwen (Eds.), *Crime Mapping and Crime Prevention*. Criminal Justice Press.

Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of quantitative criminology 16*(1), 21–43.

Pauwels, L., F. M. Weerman, W. Bernasco, and B. Völker (2012). Ruimtelijke criminologie: van woonbuurt tot cyberspace en van politiestatistiek tot space-time budgets. *Tijdschrift voor Criminologie 54*(4), 291–305.

Perry, W. L. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation.

Politie (2013, October). Misdaad in kaart. `http://www.politie.nl/nieuws/2013/oktober/28/00-wat-doet-u-tegen-woninginbraak.html`.

Pyke, G. H., H. R. Pulliam, and E. L. Charnov (1977). Optimal foraging: a selective review of theory and tests. *Quarterly Review of Biology*, 137–154.

Rajabifard, A. and I. P. Williamson (2001). Spatial data infrastructures: concept, sdi hierarchy and future directions.

Ratcliffe, J. (2010). Crime mapping: spatial and temporal challenges. In *Handbook of quantitative criminology*, pp. 5–24. Springer.

Rogerson, R. J. (1999). Quality of life and city competitiveness. *Urban studies 36*(5-6), 969–985.

RTLNieuws (2015, August). Aantal inbraken daalt, maar niet overal. `http://www.rtlnieuws.nl/nieuws/binnenland/aantal-inbraken-daalt-maar-niet-overal`.

Sampson, R. J. and W. B. Groves (1989). Community structure and crime: Testing social-disorganization theory. *American journal of sociology*, 774–802.

Schaffers, H., N. Komninos, M. Pallot, M. Aguas, E. Almirall, T. Bakici, J. Barroca, D. Carter, M. Corriou, J. Fernadez, et al. (2012). Fireball white paper on smart cities as innovation ecosystems sustained by the future internet.

Scott, M. S. (2004). Burglary of single-family houses in savannah, georgia: A final report to the u.s. department of justice, office of community oriented policing services on the field applications of the problem-oriented guides for police project. `http://www.popcenter.org/library/researcherprojects/BurglarySingleHouses.pdf`.

Scripter, M. W. (1970). Nested-means map classes for statistical maps. *Annals of the Association of American Geographers 60*(2), 385–392.

Sorensen, D. W. (2004). Temporal patterns of danish residential burglary. *Copenhagen, Justits Ministeriet*.

Spelman, W. and J. E. Eck (1987). *Problem-oriented policing.* US Department of Justice, National Institute of Justice.

Taylor, R. B. (2003). Crime prevention through environmental. *Handbook of environmental psychology*, 413.

The University of Texas (2012, April). Common misteaks mistakes in using statistics: Spotting and avoiding them. `http://www.ma.utexas.edu/users/mks/statmistakes/StatisticsMistakes.html`.

Tilley, N. (2003). Problem-oriented policing, intelligence-led policing and the national intelligence model. *London: Jill Dando Institute of Crime Science, University College London*.

UN (2014). World urbanization prospects: The 2014 revision. Technical report, United Nations, Department of Economic and Social Affairs.

van der Werff, M. (2013). Handhaving, veiligheid en overlast. Technical report, Onderzoek en Statistiek, Gemeente Haarlem.

van Dijk, B. (2015, February). Big data: nieuwe software voorspelt diefstal en straatroof. `http://fd.nl/economie-politiek/1091395/politie-laat-computer-criminaliteit-voorspellen`.

van Noort, W. (2015, October). Hoe de slimme stad een dom idee kan worden. `http://www.nrc.nl/handelsblad/2015/10/17/de-slimme-stad-kan-een-dom-idee-worden-1546062`.

Walsh, D. (1986). Victim selection procedures among economic criminals: The rational choice perspective. *The reasoning criminal: Rational choice perspectives on offending*, 39–52.

Weisburd, D., E. Groff, and S.-M. Yang (2009). Understanding developmental crime trajectories at places: social disorganization and opportunity perspectives at micro units of geography. *National Institute of Justice, Washington, DC*.

Weisburd, D., C. W. Telep, J. C. Hinkle, and J. E. Eck (2010). Is problem-oriented policing effective in reducing crime and disorder? *Criminology & Public Policy 9*(1), 139–172.

Wetenschappelijke Raad voor het Regeringsbeleid (2011). *IOverheid*, Volume 86. Amsterdam University Press.

Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of animal ecology 75*(5), 1182–1189.

Willems, D. and R. Doeleman (2014). Predictive policing –wens of werkelijkheid? *Het Tijdschrift voor de Politie 4*(5), 39–42.

Wilsem, J. v., K. Wittebrood, and N. D. De Graaf (2006). Socioeconomic dynamics of neighborhoods and the risk of crime victimization: a multilevel study of improving, declining, and stable areas in the netherlands.

Yan, Y. Y. (2004). Seasonality of property crime in hong kong. *British Journal of Criminology 44*(2), 276–283.

Appendices

Appendix A

Input data

This appendix provides detailed information about the input data that are used in the regression model. The focus is on the steps that were followed to operationalize the risk factors. But also the underlying assumptions of how independent variables relate to the dependent variable are discussed including references to the theory and the source of the data. More information about the data sources can be found in appendix B.

A.1 Residential addresses

Theory and assumptions

The residential addresses are the main unit of analysis for this study. Both the dependent variable and the independent variables are related to residential addresses before being aggregated to grid cells. That is why the process of identifying residential addresses is discussed first.

Source data

The data is extracted from the *Basisadministratie Adressen en Gebouwen* (key registry for addresses and buildings) or BAG.

Operationalization

The extraction of the residential addresses is basically based on a selection from the BAG dataset.

1. All addresses which are designated as addresses with a residential purpose are selected. Note that this excludes 'special' residential addresses, for example addresses of residential units for elderly or handicapped people. These are excluded because these residential addresses are often located in large building complexes. The assumption here is that these building complexes are intrinsically different from regular housing, making it

unlikely that the general theories for spatial patterns of burglaries apply here. Addresses belonging to houseboats and mobile homes are excluded for the same reason.

2. Only residential addresses with a construction year before 2010 are selected. This is done to ensure that only addresses that existed for all four years of study are selected to allow for a comparison between these years.

The map in figure A.1 shows the resulting residential addresses that are used in the analysis.

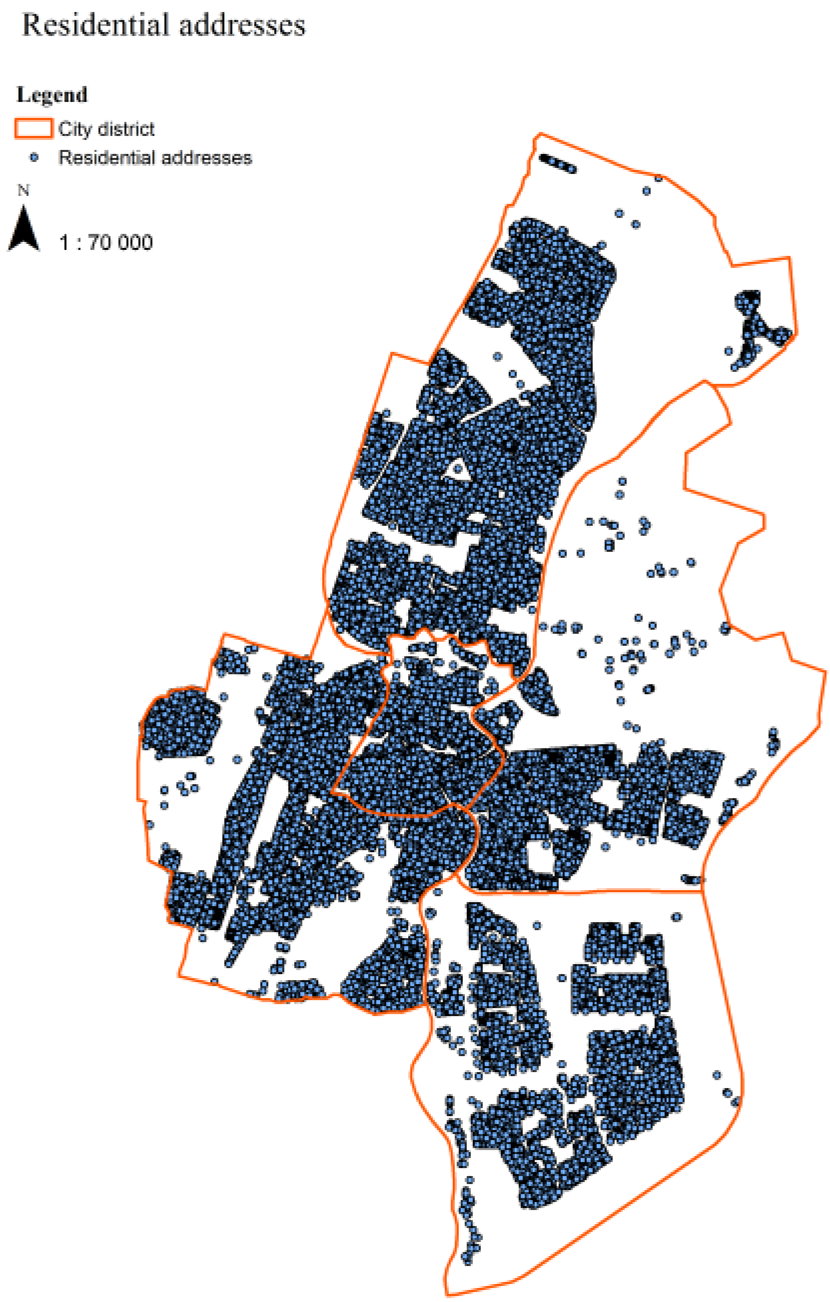


Figure A.1: The residential addresses in Haarlem used for the spatiotemporal analysis in this study.

A.2 Burglaries

Theory and assumptions

As the residential burglaries are the dependent variable, there are no relevant assumptions or references to the theory to mention here.

Source data

The data about burglaries is derived from the *Meldingen Openbare Ruimte* (reports concerning public space). In this registration, the data on burglary events is provided by the police. An important data limitation is that, due to privacy concerns, burglary events are related only to streets. As the data is aggregated to 100 meter square cells, this does not pose a problem in most cases. Only when streets are exceptionally long, for example several hundreds of meters, some inaccuracies can be expected in the data. The operationalization section covers how is dealt with this issue.

Operationalization

1. The first step in the operationalization process is to aggregate the different subcategories of burglaries that exist within the police database. These include (in Dutch):

   - inbraak woning;
   - gekwalificeerde diefstal in/uit woning;
   - diefstal in/uit woning (geen braak);
   - diefstal in/uit woning (niet gekwalificeerd);
   - inbraak woning met geweld;
   - gekwalificeerde diefstal met geweld in/uit woning;
   - diefstal in/uit woning met geweld (geen braak); and
   - diefstal met geweld in/uit woning (niet gekwalificeerd).

   There are several motives for aggregating the events in these categories. First, the police also aggregates the events in these categories when discussing or presenting figures about residential burglaries. It makes sense to follow their conventions. Second, it is assumed that the explanations for these different events do not differ significantly. The assumption is made that these categories represent the same phenomenon. Finally, aggregating the data also ensures that there is sufficient data to perform statistical analyses on and to draw viable conclusions.

2. A data selection is made based on the year or season at hand. Data is selected for the years 2010, 2011, 2012 and 2013. Seasonal data is aggregated over these four years.

3. The next step adds a geographic component to the burglary events by relating the street codes in the burglary dataset to street codes in the BAG. A data field is added with the number of burglaries per street.

4. The number of burglaries per street is assigned to all residential addresses. A datafield is added containing the number of addresses per street.

5. To get an approximation for the number of recorded burglaries per residential address, the number of burglaries per street is divided by the number of residential addresses per street.

6. The final step is to convert the number of burglaries per residential address to grid cell values for the study area. This is done by aggregating the burglaries per address inside each grid cell. Then, based on the number of addresses per cell, the number of burglaries per 1000 addresses is calculated. This value is rounded to create integer values.

A.3 Household income

Theory and assumptions

Household income is defined as a target risk factor. Household income is believed to be an indicator for the expected gains of a burglary. This corresponds with rational choice and optimal foraging theories. The assumption is that there is a positive relationship with burglaries:

> *a higher household income results in a higher number of burglaries.*

Source data

The data on household income comes from the Central Bureau for Statistics. It is only available on a neighborhood level, which unfortunately causes local spatial patterns to be smoothed. The data is extracted for the years 2010, 2011 and 2012. Data for the year 2013 was unavailable. For 2013, the data of 2012 is used.

Operationalization

1. The household income figure is assigned to the residential addresses of each neighborhood.

2. The household income is then averaged for the addresses within each grid cell and this value is assigned to the cell.

A.4 Welfare benefits

Theory and assumptions

Welfare benefits is a target risk factor based on rational choice and optimal foraging theories. The percentage of the population receiving welfare benefits can be linked to unemployment and are considered an indicator for the expected gains of a burglary. The assumption is that there is a negative relationship with burglaries:

> *a higher percentage of people receiving welfare benefits causes a lower number of burglaries.*

Source data

The data on welfare benefits comes from the municipality of Haarlem. It is only available on a neighborhood level, which unfortunately causes local spatial patterns to be smoothed. The data is extracted for the years 2010, 2011, 2012 and 2013.

Operationalization

1. The welfare benefit percentage is assigned to the residential addresses of each neighborhood.
2. The welfare benefit percentage is then averaged for the addresses within each grid cell and this value is assigned to the cell.

A.5 Cars per household

Theory and assumptions

Cars per household is a target risk factor based on rational choice and optimal foraging theories. Cars per household are considered an indicator for the expected gains of a burglary. The assumption is that there is a positive relationship with burglaries:

> *a higher number of cars per household causes a higher number of burglaries.*

Source data

The data on cars per household comes from the Central Bureau for Statistics. It is only available on a neighborhood level, which unfortunately causes local spatial patterns to be smoothed. The data is extracted for the years 2010, 2011, 2012 and 2013.

Operationalization

1. The cars per household figure is assigned to the residential addresses of each neighborhood.

2. The cars per household are then averaged for the addresses within each grid cell and this value is assigned to the cell.

A.6 Property value

Theory and assumptions

Property value is a target risk factor based on rational choice and optimal foraging theories. Property values are considered an indicator for the expected gains of a burglary. The assumption is that there is a positive relationship with burglaries:

> *a higher property value causes a higher number of burglaries.*

Source data

The data on property values comes from the *Waardering Onroerende Zaken* (real estate valuation) registration or WOZ. Only the latest data (2014) is available. This does not pose problems, as property values are averaged and property values generally do not fluctuate heavily.

Operationalization

1. The property values are assigned to the residential addresses.

2. The property values are averaged for the addresses within each grid cell and this value is assigned to each grid cell.

A.7 Risky properties

Theory and assumptions

Risky properties is a target risk factor based on rational choice and optimal foraging theories. Risky properties include properties that have a high number of exposed sides, including detached, semi-detached or corner properties. These properties are often valued higher. That is why risky properties are considered an indicator for the expected gains of a burglary. The assumption is that there is a positive relationship with burglaries: a higher percentage of risky properties causes a higher number of burglaries.

Risky properties are considered a setting risk factor too, based on environmental design theory. As risky properties have less neighbors, the chance of offenders being observed is smaller. The assumption of a positive relationship with burglaries stays the same:

a higher percentage of risky properties causes a higher number of burglaries.

Source data

The data on risky properties is derived from the *Basisadministratie Adressen en Gebouwen* (key registry for addresses and buildings).

Operationalization

1. All addresses that are associated with detached, semi-detached or corner properties are selected.

2. The number of detached, semi-detached or corner addresses are summed per grid cell and labeled as 'risky properties'.

3. The risky properties count is divided by the total number of residential addresses within each grid cell. This percentage is assigned to each grid cell.

A.8 Rental properties

Theory and assumptions

Rental properties is a target risk factor based on rational choice and optimal foraging theories. The general idea is that people who rent a home are generally less wealthy, making these properties less attractive for offenders. This is why the percentage of rental properties is considered an indicator for the expected gains of a burglary. The assumption is that there is a negative relationship with burglaries:

a higher percentage of rental properties causes a lower number of burglaries.

Rental properties are also considered a setting risk factor, based on social disorganization theory. Residents in rental properties generally have less ties with their local environment and express less guardianship. This means that at the same time the assumption is that there is a positive relationship with burglaries:

a higher percentage of rental properties causes a higher number of burglaries.

Source data

The source data is from the *Waardering Onroerende Zaken* (real estate valuation) registration or WOZ. This registration includes the ownership of properties.

Operationalization

1. All properties without a natural person in the WOZ-ownership field are selected. This means that there is no A-number present, which is a number associated with natural persons. Although there may be some exceptions, in most cases this query returns the rental properties. Note that no distinction is made between social renting and private renting.

2. The next step is to link the rental properties to the addresses to include the geometry.

3. Finally the number of rental property addresses is counted per grid cell. This value is divided by the total number of addresses to get a percentage of rental properties per grid cell.

A.9 Building density

Theory and assumptions

Building density is a target risk factor based on rational choice and optimal foraging theories. The idea is that offenders have a higher chance of finding a suitable target within an area with a higher building density. This is why the building density is considered an indicator for the expected gains of a burglary. The assumption is that there is a positive relationship with burglaries:

> *a higher building density causes a higher number of burglaries.*

Source data

The source data is from the *Basisadministratie Adressen en Gebouwen* (key registry for addresses and buildings).

Operationalization

The residential addresses are counted per grid cell. This value is also the density (number of residential addresses per hectare) because all grid cells are the same size.

A.10 Distance city center

Theory and assumptions

Distance to the city center is an offender risk factor based on the awareness space theory. The general idea is that the city center is very likely to be in the awareness space of offenders. As offenders prefer targets within their awareness space, addresses within or close to the city center are considered to be at a higher risk. The assumption is that there is a negative relationship with burglaries:

a larger distance to the city center causes a lower number of burglaries.

Source data

The source data is from the municipality of Haarlem, specifically a data set with all city districts.

Operationalization

1. The grid cells within the city district *Centrum* (Center) are selected.
2. The euclidean distance of all residential addresses to the nearest city center grid cell is calculated in meters.
3. The distance from the residential addresses to the city center is averaged per grid cell. This value is assigned to the grid cell. Grid cells with value 0 are within the city center district.

A.11 Distance public facilities

Theory and assumptions

Distance to public facilities is an offender risk factor based on the awareness space theory. The general idea is that public facilities are very likely to be in the awareness space of offenders. As offenders prefer targets within their awareness space, addresses close to public facilities are considered to be at a higher risk. The assumption is that there is a negative relationship with burglaries:

a larger distance to public facilities causes a lower number of burglaries.

Source data

The source data is from the *Basisadministratie Adressen en Gebouwen* (key registry for addresses and buildings).

Operationalization

1. A selection is made of all addresses which have the purpose of *maatschappelijke doeleinden* (societal purpose), *overheidsvoorzieningen* (governmental services), *onderwijsvoorzieningen* (educational services) and *medische voorzieningen* (medical services). These kinds of purposes are all public in nature.
2. The euclidean distance of all residential addresses to the nearest public facility is calculated in meters.
3. The distance from the residential addresses to the nearest public facility is averaged per grid cell. This value is assigned to the grid cell.

A.12 Distance retail and catering

Theory and assumptions

Distance to retail and catering is an offender risk factor based on the awareness space theory. The general idea is that retail and catering establishments are very likely to be in the awareness space of offenders. As offenders prefer targets within their awareness space, addresses close to a high concentration of retail and catering establishments are considered to be at a higher risk. The focus is on percentages because a single shop, bar or restaurant is not likely to be within the awareness space of many offenders. But a high concentration of shops or bars is much more likely to be in the awareness space of offenders. The assumption is that there is a negative relationship with burglaries:

> *a larger distance to a high quantity of retail and catering establishments causes a lower number of burglaries.*

Source data

The source data is from the *Basisadministratie Adressen en Gebouwen* (key registry for addresses and buildings).

Operationalization

1. A selection is made of all addresses which have the purpose of *detailhandel* (retail), *horeca minder hinderlijke typen* (catering establishments that cause less nuisance) and *horeca hinderlijke typen* (catering establishments that cause more nuisance).

2. The number of retail and catering addresses are counted per grid cell. This number is divided by the total number of addresses within a grid cell. This results in a percentage of retail and catering establishments per grid cell.

3. The mean and standard deviation of these percentages is calculated.

4. The value of 1 standard deviation is added to the mean value. All grid cells with percentages of retail and catering establishments equal to or above this value are considered as grid cells with a high concentration of retail and catering establishments.

5. The euclidean distance of all residential addresses to the nearest grid cell with a high concentration of retail and catering establishments is calculated in meters.

6. The distance from the residential addresses to the nearest grid cell with a high concentration of retail and catering establishments is averaged per grid cell. This value is assigned to the grid cell. Grid cells with value 0 are cells with a high concentration of retail and catering establishments.

A.13 Distance public transport node

Theory and assumptions

Distance to a public transport node is an offender risk factor based on the awareness space theory. The general idea is that public transport nodes are very likely to be in the awareness space of offenders. As offenders prefer targets within their awareness space, addresses close to transport nodes are considered to be at a higher risk. The assumption is that there is a negative relationship with burglaries:

> *a larger distance to a transport node causes a lower number of burglaries.*

Source data

The source data is from the 9292 open data repository. The transport nodes, i.e. bus stops or train stations, are based on the situation at the beginning of 2014.

Operationalization

1. The coordinates that are available in the source data are converted to points.
2. The euclidean distance of all residential addresses to the nearest public transport node is calculated in meters.
3. The distance from the residential addresses to the nearest public transport node is averaged per grid cell. This value is assigned to the grid cell.

A.14 Distance highway entry

Theory and assumptions

Distance to a highway entry is an offender risk factor based on the awareness space theory. The general idea is that highway entries are very likely to be in the awareness space of offenders. As offenders prefer targets within their awareness space, addresses close to highway entries are considered to be at a higher risk. The assumption is that there is a negative relationship with burglaries:

> *a larger distance to a highway entry causes a lower number of burglaries.*

Source data

The source data is from the *Nationaal Wegen Bestand* (national road repository). Roads of the years 2010, 2011, 2012 and 2013 are available.

Operationalization

1. The highway entries are based on a visual inspection of the road network. Three highway entries in or near the study area are identified (see figure A.2).

2. The network distance of all grid cell center points to the nearest highway entry is calculated in meters. This value is assigned to the grid cell.

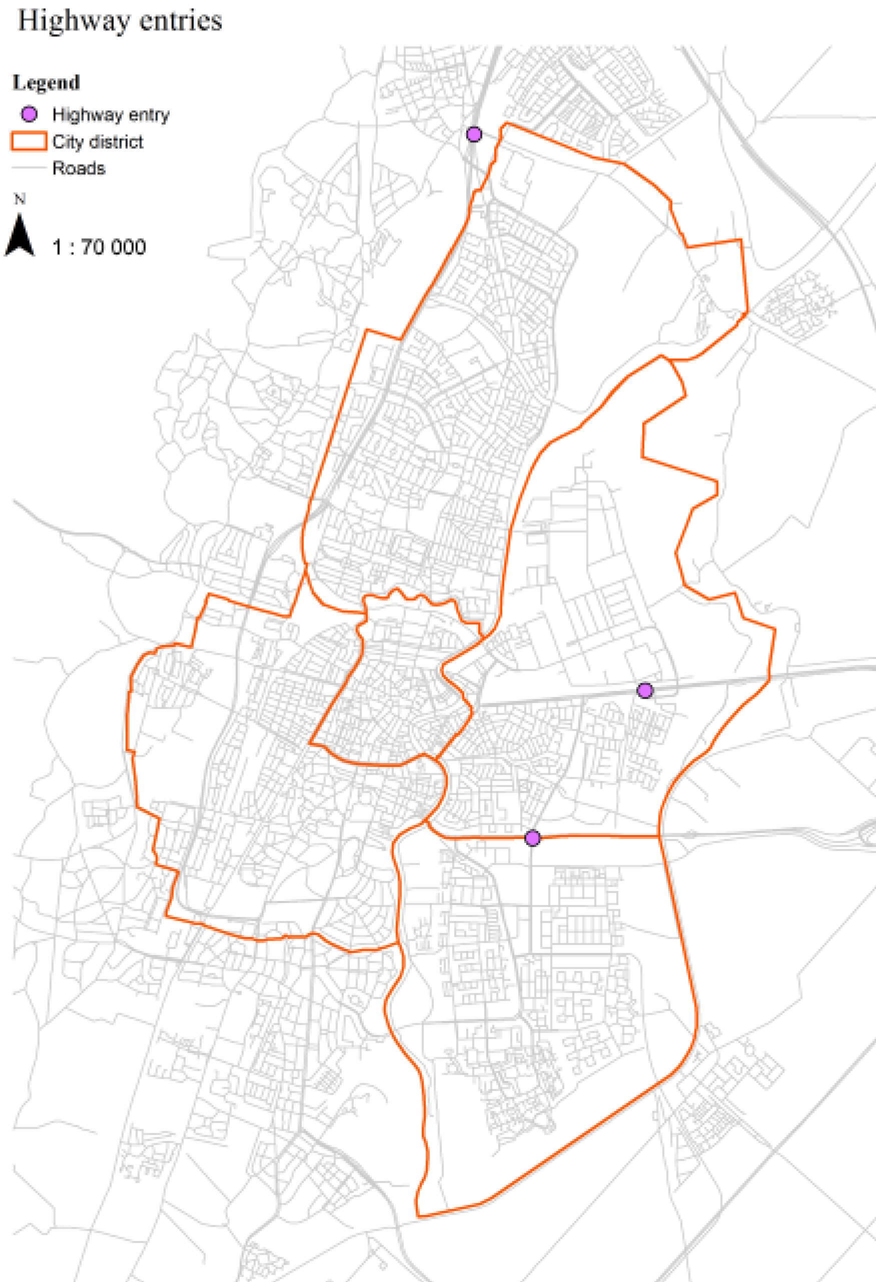


Figure A.2: The highway entries.

A.15 Accessibility

Theory and assumptions

Accessibility is an offender risk factor based on the awareness space theory. The general idea is that more integrated roads are more likely to be in the awareness space of offenders. As offenders prefer targets within their awareness space, addresses close to more integrated roads are considered to be at a higher risk. The assumption is that there is a positive relationship with burglaries:

a more integrated road causes a higher number of burglaries.

Accessibility is also a setting risk factor based on environmental design theory. The general idea is that more integrated road segments are under greater surveillance. The assumption is that there is a negative relationship with burglaries:

a more integrated road causes a lower number of burglaries.

Source data

The source road network data is from the *Nationaal Wegen Bestand* (national road repository). Roads of the years 2010, 2011, 2012 and 2013 are available.

Operationalization

1. Space Syntax analysis of the road network integration is performed. The resulting integration values are assigned to each road segment representing how integrated a road segment is. Integration measures how many turns have to be made from a street segment to reach all other street segments in the network, using shortest paths. An example of the results based on the road network of 2010 is displayed in figure A.3.

2. The residential addresses inherit the integration value of the road segment that is closest to them.

3. The integration values of the residential addresses are averaged per grid cell and this value is assigned to the grid cells.

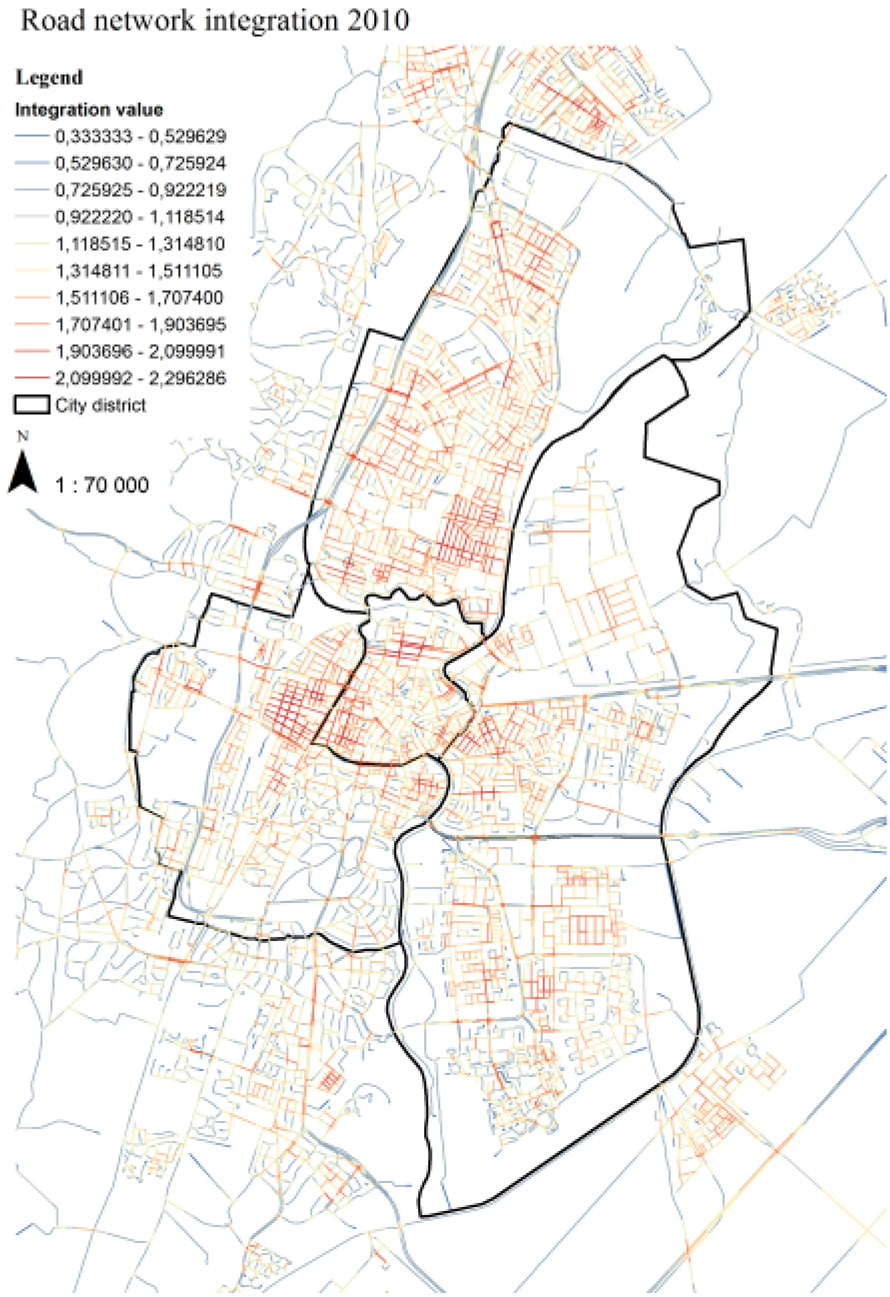


Figure A.3: The road network integration of the road segments in the study area in 2010.

A.16 Distance low incomes

Theory and assumptions

Distance to areas with the lowest incomes is an offender risk factor based on the offender neighborhood theory. The general idea is that offenders typically live in areas with lower incomes and that offenders choose their targets near their place of residence. Therefore areas with the lowest incomes, and areas close by, run a higher risk of burglaries. The assumption is that there is a negative relationship with burglaries:

> *a larger distance to areas with the lowest incomes causes a lower number of burglaries.*

Source data

The data on household income comes from the Central Bureau for Statistics. It is only available on a neighborhood level, which unfortunately causes local spatial patterns to be smoothed. The data is extracted for the years 2010, 2011, 2012. Data for 2013 was unavailable. For 2013, the data for 2012 is used.

Operationalization

1. The household income figure is assigned to the residential addresses of each neighborhood.
2. The household income is then averaged for the addresses within each grid cell and this value is assigned to the cell.
3. The mean and standard deviation of these income values is calculated.
4. The value of 1 standard deviation is subtracted from the mean value. All grid cells with household incomes equal to or below this value are considered as grid cells with the lowest household incomes.
5. The euclidean distance of all residential addresses to the nearest grid cell with a low income is calculated in meters.
6. The distances from the residential addresses to the nearest grid cell with a low income are averaged per grid cell. This value is assigned to the grid cell. Grid cells with value 0 are cells with the lowest incomes.

A.17 Distance welfare benefits

Theory and assumptions

Distance to areas with high percentages of people receiving welfare benefits is an offender risk factor based on the offender neighborhood theory. The general idea is that offenders typically live in areas with higher percentages of people receiving welfare benefits and that offenders choose their targets near their place of residence. Therefore areas with the highest percentages of welfare benefits, and areas close by, run a higher risk of burglaries. The assumption is that there is a negative relationship with burglaries:

> *a larger distance to areas with the highest percentages of people receiving welfare benefits causes a lower number of burglaries.*

Source data

The data on welfare benefits comes from the municipality of Haarlem. It is only available on a neighborhood level, which unfortunately causes local spatial patterns to be smoothed. The data is extracted for the years 2010, 2011, 2012 and 2013.

Operationalization

1. The percentage of people receiving welfare benefits is assigned to the residential addresses of each neighborhood.
2. The percentage is then averaged for the addresses within each grid cell and this value is assigned to the cell.
3. The mean and standard deviation of these income values is calculated.
4. The value of 1 standard deviation is added to the mean value. All grid cells with household incomes equal to or above this value are considered as grid cells with high percentages of people receiving welfare benefits.
5. The euclidean distance of all residential addresses to the nearest grid cell with a high percentage is calculated in meters.
6. The distances from the residential addresses to the nearest grid cell with a high percentage on welfare are averaged per grid cell. This value is assigned to the grid cell. Grid cells with value 0 are cells with the highest percentages.

A.18 Distance demographic risk group

Theory and assumptions

Distance to areas with high percentages of people within the demographic risk group (males aged 15 to 24) is an offender risk factor based on the offender neighborhood theory. The general idea is that offenders are more likely to be in this demographic group and that offenders prefer targets near their place of residence. Therefore areas with the highest percentages of people within the demographic risk group, and areas close by, run a higher risk of burglaries. The assumption is that there is a negative relationship with burglaries:

a larger distance to areas with the highest percentages of people within the demographic risk group causes a lower number of burglaries.

Source data

The source data is derived from the *Gemeentelijke Basis Administratie Persoonsgegevens* (municipal basic registration of personal data).

Operationalization

1. The individuals that are both male and aged from 15 until 24 (risk group) are selected.
2. These individuals are geo-referenced based on the address field.
3. The number of people within the risk group are counted per grid cell and divided by the total population of that grid cell to get a risk group percentage.
4. The mean and standard deviation is calculated from the risk group percentage.
5. A value of 1 standard deviation is added to the mean value. All grid cells with risk group percentages equal to or above this value are considered as grid cells with high risk group percentages.
6. The euclidean distance of all residential addresses to the nearest grid cell with a high percentage is calculated in meters.
7. The distances from the residential addresses to the nearest grid cell with a high risk group percentage are averaged per grid cell. This value is assigned to the grid cell. Grid cells with value 0 are cells with the highest percentages.

A.19 Distance ethnic heterogeneity

Theory and assumptions

Distance to areas with high ethnic heterogeneity is an offender risk factor based on the offender neighborhood theory. The general idea is that offenders are more likely to live in areas with high ethnic heterogeneity and that offenders prefer targets near their place of residence. Therefore areas with the highest ethnic heterogeneity, and areas close by, run a higher risk of burglaries. The assumption is that there is a negative relationship with burglaries:

a larger distance to areas with the highest ethnic heterogeneity causes a lower number of burglaries.

Source data

The source data is derived from the *Gemeentelijke Basis Administratie Persoonsgegevens* (municipal basic registration of personal data).

Operationalization

1. The individuals that are not born in The Netherlands are selected.
2. These individuals are geo-referenced based on the address field.
3. The number of people born abroad are counted per grid cell and divided by the total population of that grid cell to get a percentage of people born abroad.
4. The mean and standard deviation is calculated from the percentage of people born abroad.
5. A value of 1 standard deviation is added to the mean value. All grid cells with percentages of people born abroad equal to or above this value are considered as grid cells with high ethnic heterogeneity.
6. The euclidean distance of all residential addresses to the nearest grid cell with a high percentage is calculated in meters.
7. The distances from the residential addresses to the nearest grid cell with high ethnic heterogeneity are averaged per grid cell. This value is assigned to the grid cell. Grid cells with value 0 are cells with the highest ethnic heterogeneity.

A.20 Distance rental properties

Theory and assumptions

Distance to areas with high percentages of rental properties is an offender risk factor based on the offender neighborhood theory. The general idea is that offenders are more likely to live in low income areas. The assumption is made that people with lower incomes are more likely to rent their homes than to buy them. Furthermore, offenders prefer targets near their place of residence. Therefore areas with the highest percentage of rental properties, and areas close by, run a higher risk of burglaries. The assumption is that there is a negative relationship with burglaries:

> *a larger distance to areas with the highest percentage of rental properties causes a lower number of burglaries.*

Source data

The source data is from the *Waardering Onroerende Zaken* (real estate valuation) registration or WOZ. This registration includes the ownership of properties.

Operationalization

1. All properties without a natural person in the WOZ-ownership field are selected. This means that there is no A-number present, which is a number associated with natural persons. Although there may be some exceptions, in most cases this query returns the rental properties. Note that no distinction is made between social renting and private renting.
2. The next step is to link the rental properties to the addresses to include the geometry.
3. Finally the number of rental property addresses is counted per grid cell. This value is divided by the total number of addresses to get a percentage of rental properties per grid cell.
4. The mean and standard deviation is calculated from the percentage of rental properties born abroad.
5. A value of 1 standard deviation is added to the mean value. All grid cells with percentages of rental properties equal to or above this value are considered as grid cells with a high percentage of residential properties.
6. The euclidean distance of all residential addresses to the nearest grid cell with a high percentage of rental properties is calculated in meters.
7. The distances from the residential addresses to the nearest grid cell with a high percentage of rental properties are averaged per grid cell. This value is assigned to the grid cell. Grid cells with value 0 are cells with the highest percentage of rental properties.

A.21 Ethnic heterogeneity

Theory and assumptions

Areas with high ethnic heterogeneity are a setting risk factor based on the social disorganization theory. The general idea is that areas with higher ethnic heterogeneity are often characterized by residents who are not socially integrated. Therefore areas with higher ethnic heterogeneity run a higher risk of burglaries. The assumption is that there is a positive relationship with burglaries:

> *a higher ethnic heterogeneity causes a higher number of burglaries.*

Source data

The source data is derived from the *Gemeentelijke Basis Administratie Persoonsgegevens* (municipal basic registration of personal data).

Operationalization

1. The individuals that are not born in The Netherlands are selected.
2. These individuals are geo-referenced based on the address field.
3. The number of people born abroad are counted per grid cell and divided by the total population of that grid cell to get a percentage of people born abroad. This value is assigned to the grid cell.

A.22 Residential mobility

Theory and assumptions

Residential mobility is a setting risk factor based on the social disorganization theory. The general idea is that people who move more often, feel less responsibility for their neighborhood and have a lesser sense of guardianship. The assumption is that there is a negative relationship with burglaries:

> *a higher number of years living on one address causes a lower number of burglaries.*

Source data

The source data is from the municipality of Haarlem. It is only available on a neighborhood level, which unfortunately causes local spatial patterns to be smoothed. The data is extracted for the years 2010, 2011, 2012 and 2013.

Operationalization

1. The average number of years of residence at one address is assigned to the residential addresses of each neighborhood.
2. The number of years are then averaged for the addresses within each grid cell and this value is assigned to the cell.

A.23 Election turnout

Theory and assumptions

Election turnout is a setting risk factor based on the social disorganization theory. The general idea is that people who do not vote in local elections, feel less responsibility for their local environment and have a lesser sense of guardianship. The assumption is that there is a negative relationship with burglaries:

a higher election turnout causes a lower number of burglaries.

Source data

The source data is from the municipality of Haarlem and contains the people who voted as a percentage of the eligible voters. It is only available on a neighborhood level, which unfortunately causes local spatial patterns to be smoothed. Moreover, the data is only available for 2010 as the local elections were in this year. The data for 2010 is extrapolated for the subsequent years.

Operationalization

1. The percentage of voters is assigned to the residential addresses of each neighborhood.
2. The percentage of voters is then averaged for the addresses within each grid cell and this value is assigned to the cell.

A.24 Nuisance

Theory and assumptions

Nuisance is a setting risk factor based on the social disorganization theory. The general idea is that areas with more nuisance experience less social control. The assumption is that there is a positive relationship with burglaries:

more nuisance causes a higher number of burglaries.

Source data

The data about burglaries is derived from the *Meldingen Openbare Ruimte* (reports concerning public space). The data on nuisance events are provided by the municipality of Haarlem.

Operationalization

1. The following types of nuisance are selected for each year from the dataset:
   - *dieren (met eigenaar)* (animals (with owner));
   - *hangjongeren, verslaafden, zwervers* (loitering, addicts, homeless people);
   - *illegaal, verkeerd aangeboden afval* (illegal, wrong placement of waste);
   - *vernieling, vandalisme* (destruction of property, vandalism); and
   - *vervuiling, lozing* (pollution, waste).

2. The reported nuisance events are counted per grid cell.

3. The number of nuisance events is divided by the number of residential addresses per grid cell. This 'nuisance rate' is assigned to each grid cell.

A.25 Crime

Theory and assumptions

Crime is a setting risk factor based on the social disorganization theory. The general idea is that areas with more crime experience less social control. The assumption is that there is a positive relationship with burglaries:

> *more crime causes a higher number of burglaries.*

Source data

The crime data is derived from the *Meldingen Openbare Ruimte* (reports concerning public space). In this registration, the data on crimes is provided by the police. An important data limitation is that, due to privacy concerns, crime events are related only to streets. As the data is aggregated to 100 meter square cells, this does not pose a problem in most cases. Only when streets are exceptionally long, for example several hundreds of meters, some inaccuracies can be expected in the data. The operationalization section covers how is dealt with this issue.

Operationalization

1. The first step in the operationalization process is to select the crime events of each analysis year, excluding residential burglaries. This is done because burglaries are already a dependent variable.

2. The next step adds a geographic component to the crime events by relating the street codes in the burglary dataset to street codes in the BAG. A data field is added with the number of crimes (excluding burglaries) per street.

3. The number of crimes per street is assigned to all residential addresses. A datafield is added containing the number of addresses per street.

4. To get an approximation for the number of recorded crimes per residential address, the number of crimes per street is divided by the number of residential addresses per street.

5. The final step is to convert the number of crimes per residential address to grid cell values for the study area. This is done by aggregating the crime events per address inside each grid cell. This value is assigned to the grid cell.

A.26 Construction year

Theory and assumptions

Construction year is a setting risk factor based on the environmental design theory. The general idea is that older properties generally have less target hardening present, i.e. worse window and door locks. The assumption is that there is a negative relationship with burglaries:

> *a more recent construction year causes a lower number of burglaries.*

Source data

The source data is from the *Basisadministratie Adressen en Gebouwen* (key registry for addresses and buildings).

Operationalization

1. The construction year of all residential addresses is averaged per grid cell. The average construction year is assigned as the cell's value.

A.27 Distance street lighting

Theory and assumptions

Distance to the nearest street light is a setting risk factor based on the environmental design theory. The general idea is that properties closer to a street light are less attractive to offenders, because the risk of being spotted is higher. The assumption is that there is a positive relationship with burglaries:

> *a larger distance to street lighting causes a higher number of burglaries.*

Source data

The source data is from the *Beheer Openbare Ruimte* (management of the public space) register of the municipality of Haarlem.

Operationalization

1. The euclidean distance from all residential addresses to the nearest street light is calculated.

2. The distances are averaged per grid cell and this average distance to the nearest streetlight is assigned to the grid cell.

A.28 Mixed land use

Theory and assumptions

Mixed land use is a setting risk factor based on the environmental design theory. The general idea is that areas with more mixed land use, have less residents who can survey the area. The assumption is that there is a positive relationship with burglaries:

> *more mixed land use causes a higher number of burglaries.*

Source data

The data on risky properties is derived from the *Basisadministratie Adressen en Gebouwen* (key registry for addresses and buildings).

Operationalization

1. Addresses with purposes other than 'residential' are selected and counted per grid cell.
2. The selected addresses are divided by the total number of addresses in each grid cell. This percentage is assigned to the cell.

A.29 Distance shrubbery

Theory and assumptions

Distance to the nearest shrubbery is a setting risk factor based on the environmental design theory. Shrubbery refers here to patches of land with high densities of medium to high bushes (in Dutch: *bosplantsoen*, see figure A.4). The general idea is that properties closer to shrubbery are more attractive to offenders, because the risk of being spotted is lower. Close by shrubbery could also aid in getting away from the scene, providing cover for offenders. The assumption is that there is a negative relationship with burglaries:

> *a larger distance to shrubbery causes a lower number of burglaries.*

Figure A.4: Example of shrubbery or a *bosplantsoen* (source: `http://tinyurl.com/m7u9apv`)

Source data

The source data is from the *Beheer Openbare Ruimte* (management of the public space) register of the municipality of Haarlem.

Operationalization

1. The euclidean distance from all residential addresses to the nearest shrubbery is calculated.
2. The distances are averaged per grid cell and this average distance to the nearest shrubbery is assigned to the grid cell.

A.30 Distance to street

Theory and assumptions

Distance to the street is a setting risk factor based on the environmental design theory. The general idea is that properties closer to the street are less attractive to offenders, because the risk of being spotted by for example passersby is higher. The assumption is that there is a positive relationship with burglaries:

> *a larger distance to the street causes a higher number of burglaries.*

Source data

The road network source data is from the *Nationaal Wegen Bestand* (national road repository). Roads of the years 2010, 2011, 2012 and 2013 are available.

Operationalization

1. The euclidean distance from all residential addresses to the nearest street is calculated.

2. The distances are averaged per grid cell and this average distance to the nearest street is assigned to the grid cell.

A.31 Edge dwellings

Theory and assumptions

Edge dwellings are a setting risk factor based on the environmental design theory. The general idea is that properties closer to the edge of a neighborhood are more attractive to offenders, because the risk of standing out is lower and there are more getaway options. The assumption is that there is a positive relationship with burglaries:

> *a higher percentage of edge dwellings causes a higher number of burglaries.*

Source data

The source data is from the *Basisadministratie Adressen en Gebouwen* (key registry for addresses and buildings).

Operationalization

1. A concave hull polygon is created based on the residential address and the neighborhood they belong to.

2. An inner buffer of 50 meters is created. This is an arbitrary value. This buffer is defined as the edge area of the neighborhood (see figure A.5).

3. All residential addresses within the buffer are considered ‘edge dwellings’. These edge dwellings are counted per grid cell.
4. The number of edge dwellings per cell is divided by the total number of residential addresses per cell. This percentage of edge dwellings is assigned to each grid cell.

Figure A.5: A concave hull polygon that is created for the residential addresses of a neighborhood including the 50 meter buffer. This process is repeated for all neighborhoods.

A.32 Distance police station

Theory and assumptions

Distance to the nearest police station is a setting risk factor based on the environmental design theory. The general idea is that properties closer to a police station are less attractive to offenders. The short distance to police stations also shortens the response time of police officers. The assumption is that there is a positive relationship with burglaries:

> *a larger distance to a police station causes a higher number of burglaries.*

Source data

The location of police stations is derived from the BRT (see figure A.6).

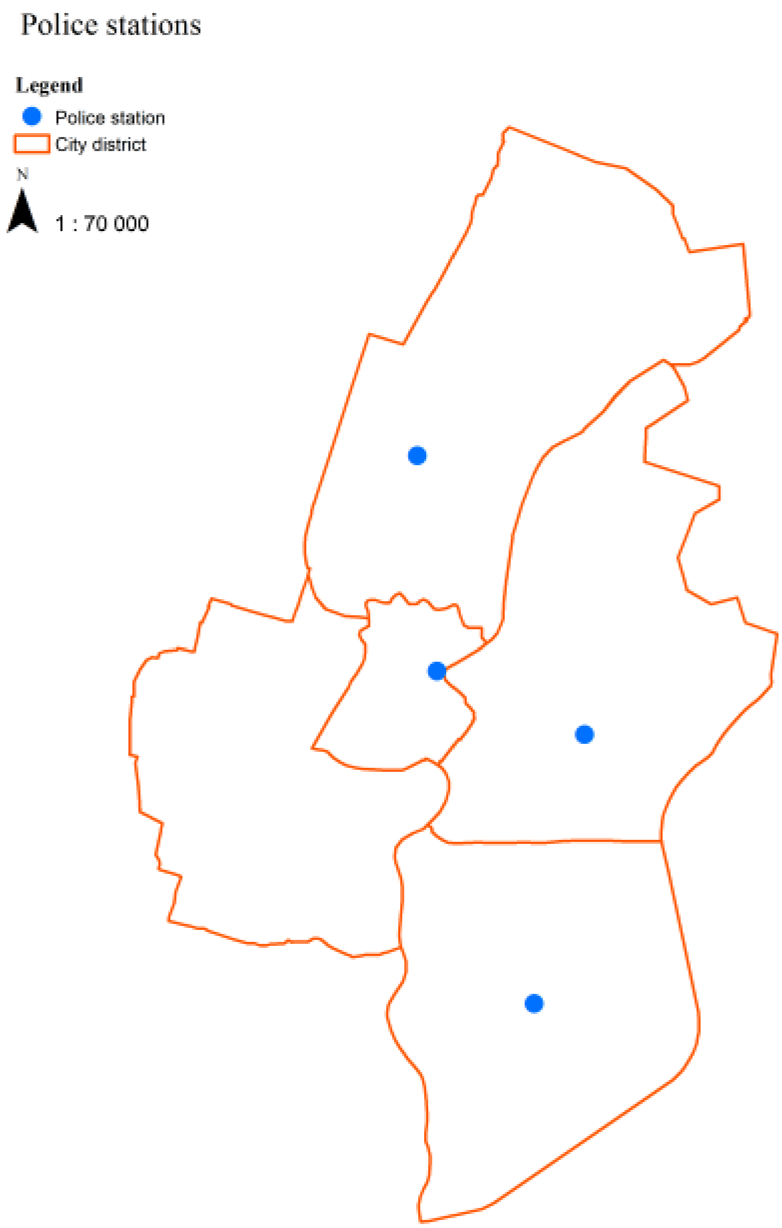


Figure A.6: The police stations in the study area based on BRT.

Operationalization

1. The network distance from all the grid cell center points to the nearest police station is calculated.

2. The distances are averaged per grid cell and this average distance to the nearest police station is assigned to the grid cell.

Appendix B

Data sources

This appendix aims at giving more in-depth information about the data sets and data registrations that are used as the source for the data in the analysis.

B.1 Basisadministratie Adressen en Gebouwen

Description

The *Basisadministatie Adressen en Gebouwen* (key register addresses and buildings), or BAG, is the most used data source for this study. It is a part of the *stelsel van basisregistraties* (system of key registers) (see figure B.1). This system is still in development, some registers, links and geometries are missing at this moment.

The BAG stores information about all addresses and buildings in The Netherlands, together with the geometry. The BAG can be divided in two parts: one part containing mandatory elements and one part containing optional elements. Only the mandatory elements are synchronized with the *landelijke voorziening*, the national data repository that collects the data of key registers from all source holders. Much information can be derived from the mandatory elements, such as purpose of an object, construction year of a building and most importantly the geometry.

The data that is used in this study is obtained via the municipality of Haarlem. An advantage is that also the optional elements are available, contrary to the public version. The optional elements provide for example more detailed information about the purpose of an object and the type of dwelling, for example: is it an apartment or a detached property?

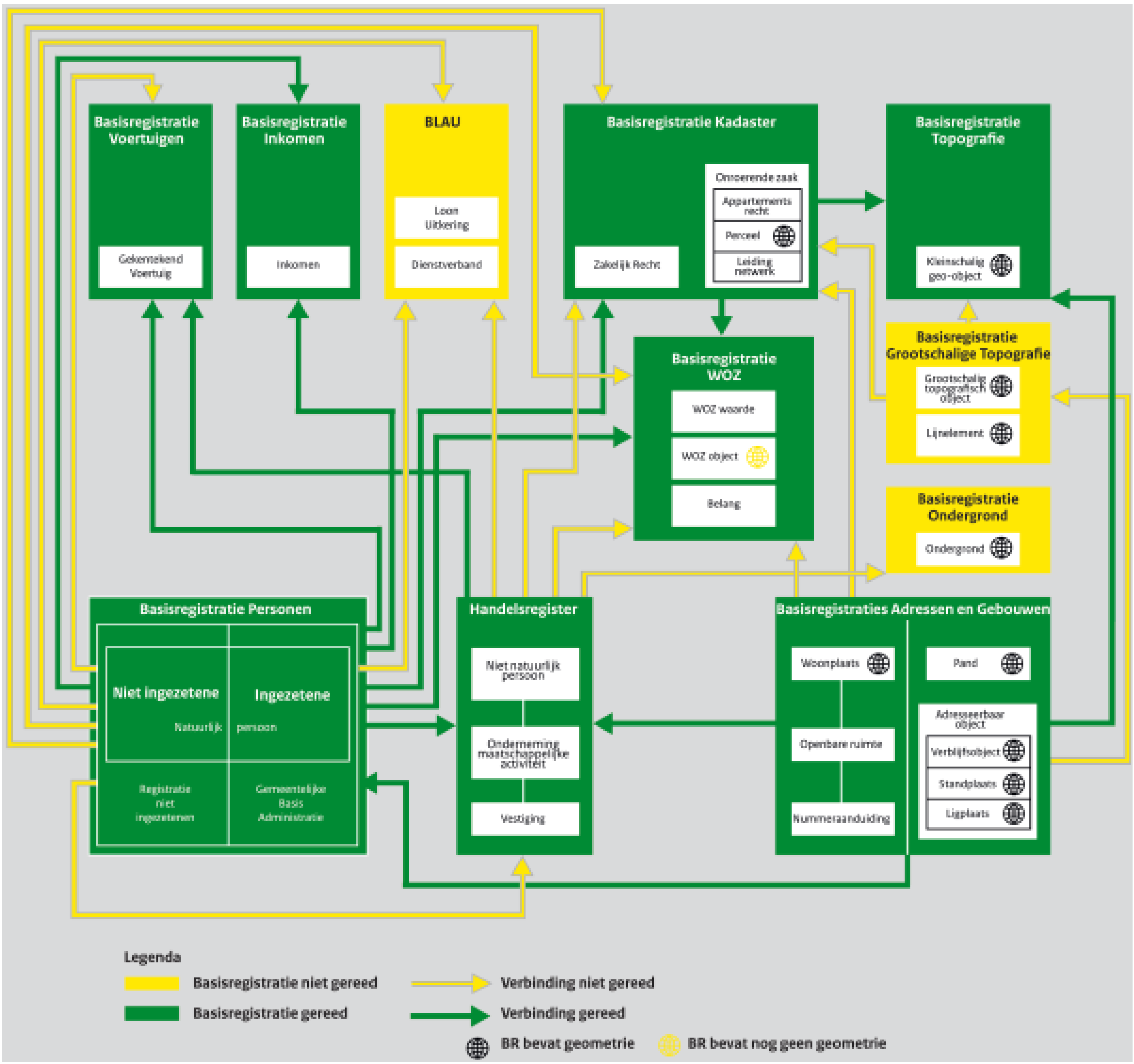


Figure B.1: Overview in Dutch of the *stelsel van basisregistraties* (system of key registers) (source: `http://tinyurl.com/aocuz58`).

Source holder

The BAG is administered on a municipality level.

Availability

The mandatory information that is sent to the national data repository is publicly available[1].

Geometry

This dataset contains geometry: addresses are represented by points and buildings by polygons.

B.2 Gemeentelijke Basis Administratie Persoonsgegevens

Description

The *Gemeentelijke Basis Administratie Persoonsgegevens* (municipal key register personal data) or GBA stores all personal information, such as name, date of birth and address. In the near future it will be a part of the system of key registers, integrated in the *Basisregistratie Personen* (key register persons) and forming connections with for example the BAG and WOZ key registers (see figure B.1).

Source holder

The municipalities are the source holders of the GBA.

Availability

The GBA is not publicly available. Therefore, data for this study could only be obtained at an aggregated level directly from the municipality of Haarlem.

Geometry

The GBA does not contain geometry but persons can be related to an address.

B.3 Waardering Onroerende Zaken

Description

The *Waardering Onroerende Zaken* (real estate valuation) or WOZ provides information about the value and ownership of a building and is for example used to calculate taxes for homeowners. The WOZ is also a key register and thus part of the system of key registers (see figure B.1).

[1] for example via `http://bagviewer.pdok.nl/`

Source holder

Municipalities are the share holders of the WOZ.

Availability

The WOZ is not publicly available on the level of individual addresses. For this study, data is directly derived from the municipality of Haarlem on an aggregated level.

Geometry

The WOZ does not contain geometry. WOZ-objects can be related to BAG-objects by comparing addresses. In the future, geometry will be added to the WOZ-objects.

B.4 Basis Registratie Topografie

Description

The *Basis Registratie Topografie* (key register topography) or BRT contains small-scale topographic information. It is a part of the system of key registers (see figure B.1). Governmental organizations are obliged to use the BRT-maps as the basis for their work.

Source holder

The Kadaster is the source holder of the BRT.

Availability

The BRT is freely and publicly available.

Geometry

The BRT contains geographic information about many different objects in the form of points, lines and polygons.

B.5 Nationaal Wegen Bestand

Description

The *Nationaal Wegen Bestand* (national road repository) or NWB includes road segments and hectometer points. The NWB is not a key register on its own, but the road data is included in the BRT. The NWB is used here because it includes earlier versions of the dataset, allowing the extraction of data of different years of analysis.

Source holder

Rijkswaterstaat maintains the NWB.

Availability

The NWB is available as open data.

Geometry

The road segments are represented as lines and hectometer points as points.

B.6 Beheer Openbare Ruimte

Description

The *Beheer Openbare Ruimte* (management of the public space) or BOR dataset includes data that is relevant for the management and maintenance of the public space by municipalities. It includes for example information about street furniture, like benches and street lights, public greenery, roads and sewers. The BOR does not only include the location of these objects, but also their state. The BOR is not a key registry, but data from the BOR is included in the BRT, the key register topography, and in the future also the BGT, the key register for large-scale topography.

Source holder

The municipalities are share holder for the BOR-data.

Availability

The BOR is not publicly available, but some information can be accessed through the key register for topography.

Geometry

The BOR includes all types of geometry: points, lines and polygons.

B.7 Meldingen Openbare Ruimte

Description

The *Meldingen Openbare Ruimte* (reports concerning public space) or MOR collects all reports that are filed by residents concerning the public space. These are for example reports of broken street lights, parking-related nuisances or reports of damaged trees after a storm.

The municipality of Haarlem has extended the MOR to also include all sorts of crime event data provided by the police. They did this because they wanted to get a more comprehensive image of the public safety in the municipality. Reports of burglaries, that are at the basis of this study, are derived from this 'extended MOR'.

Source holder

The municipalities are the source holder for the reports filed by residents concerning the public space. The police is source holder for the data concerning crime events.

Availability

The MOR register of Haarlem is not publicly available. Especially, the data concerning crime events is for internal use by authorized personnel only. For example the municipality of Amsterdam did make the MOR available[2].

Geometry

The MOR uses point features to represent reports by residents and crime reports.

B.8 Neighborhood statistics

Description

The *Gemeente Haarlem* was an important source of information for neighborhood statistics. These statistics are gathered by the department of *Onderzoek en Statistiek* (research and statistics).

Source holder

The municipality of Haarlem is the source holder.

Availability

Neighborhood statistics are published via the *Buurtmonitor*[3] (neighborhood monitor). Unfortunately, data is not available on a lower spatial scale than on the neighborhood level. This is due to privacy concerns.

Geometry

The neighborhood statistics data has no geometry. The data can be georeferenced by using the neighborhood code that is included.

[2] `http://www.opdekaart.amsterdam.nl/mor`
[3] http://haarlem.buurtmonitor.nl/

B.9 9292 Open Data

Description

9292 Open Data is an open data repository offering all sorts of data related to public transport. This includes for example locations of stops, time schedules for buses and trains and trip fares. For this study, the General Transit Feed Specification (GTFS) data is used. This includes, among other data, the data of bus and train stop locations.

Source holder

The public transport agencies providing information for this dataset are the source holders.

Availability

The data is publicly available[4].

Geometry

The public transport stops are represented as points.

[4] http://9292opendata.org/datacollecties

Appendix C

Model results

Parameter Estimates Winter

Parameter	B	Std. Error	Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
Spatiallag	,038	,0045	,029	,047	70,727	1	0,000
Distanceretailandcatering	-,001	,0002	-,001	,000	22,377	1	,000
Buildingdensity	-,005	,0014	-,008	-,003	15,349	1	,000
Distancedemographicriskgroup	-,001	,0004	-,002	-,001	13,554	1	,000
Riskyproperties	,920	,2551	,420	1,420	13,010	1	,000
Distancewelfarebenefits	,000	9,5576E-05	-,001	,000	12,330	1	,000
Distanceethnicheterogeneity	-,001	,0002	-,001	,000	7,539	1	,006
Distancecitycenter	,000	6,3392E-05	2,421E-05	,000	5,485	1	,019
Welfarebenefits	-,064	,0294	-,122	-,007	4,760	1	,029
Edgedwellings	,165	,0758	,016	,313	4,729	1	,030
Distancerentalproperties	-,001	,0003	-,001	-3,890E-05	4,302	1	,038
Crime	,230	,1121	,011	,450	4,218	1	,040
Distancepublicfacilities	,001	,0005	,000	,002	2,261	1	,133
Distanceshrubbery	,000	,0003	-,001	,000	1,721	1	,190
Distancetostreet	,011	,0086	-,006	,028	1,644	1	,200
Distancepublictransportnode	,000	,0004	,000	,001	1,627	1	,202
Residentialmobility	-,035	,0295	-,093	,023	1,432	1	,231
Ethnicheterogeneity	-,395	,4204	-1,219	,429	,881	1	,348
Householdincome	8,711E-06	9,4099E-06	-9,732E-06	2,715E-05	,857	1	,355
Distancestreetlighting	,009	,0100	-,010	,029	,838	1	,360
Mixedlanduse	,185	,2283	-,263	,632	,654	1	,419
Constructionyear	-,001	,0014	-,004	,002	,461	1	,497
Distancehighwayentry	5,676E-05	9,3843E-05	,000	,000	,366	1	,545
Accessibility	,039	,1226	-,201	,279	,102	1	,750
Rentalproperties	-,030	,1090	-,243	,184	,074	1	,785
Distancelowincomes	1,968E-05	,0001	,000	,000	,031	1	,861
Nuisance	-,192	1,2547	-2,651	2,268	,023	1	,879
Carsperhousehold	-,065	,4921	-1,029	,900	,017	1	,895
Electionturnout	-,001	,0059	-,012	,011	,016	1	,899
Propertyvalue	4,996E-08	4,2524E-07	-7,835E-07	8,834E-07	,014	1	,906
Distancepolicestation	-6,206E-06	5,3174E-05	,000	9,801E-05	,014	1	,907
(Intercept)	4,119	2,7790	-1,328	9,565	2,197	1	,138
(Scale)	1[a]						
(Negative binomial)	1[a]						

a. Fixed at the displayed value.

Figure C.1: The final full model for the winter including all parameters and sorted by significance. This is the direct output from the SPSS software.

Parameter Estimates Spring							
			Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
Spatiallag	,050	,0065	,037	,063	59,467	1	,000
Riskyproperties	1,233	,2631	,717	1,749	21,952	1	,000
Buildingdensity	-,006	,0014	-,009	-,003	18,836	1	,000
Accessibility	,379	,1249	,134	,624	9,222	1	,002
Electionturnout	-,015	,0057	-,026	-,004	6,971	1	,008
Distanceethnicheterogeneity	,000	,0002	-,001	-5,627E-05	5,084	1	,024
Distancetostreet	,017	,0084	,000	,033	3,960	1	,047
Householdincome	1,708E-05	9,2528E-06	-1,058E-06	3,521E-05	3,406	1	,065
Distancewelfarebenefits	,000	9,4436E-05	,000	2,799E-05	2,767	1	,096
Propertyvalue	-6,751E-07	4,4946E-07	-1,556E-06	2,058E-07	2,256	1	,133
Distanceshrubbery	,000	,0003	,000	,001	2,064	1	,151
Distancehighwayentry	,000	9,7060E-05	-6,651E-05	,000	1,625	1	,202
Distancecitycenter	-7,099E-05	6,1528E-05	,000	4,960E-05	1,331	1	,249
Distancedemographicriskgroup	,000	,0004	,000	,001	1,138	1	,286
Distancepolicestation	-5,427E-05	5,2952E-05	,000	4,952E-05	1,050	1	,305
Distancerentalproperties	,000	,0003	-,001	,000	,904	1	,342
Rentalproperties	-,101	,1109	-,318	,117	,827	1	,363
Residentialmobility	-,027	,0310	-,088	,034	,753	1	,386
Distanceretailandcatering	,000	,0002	,000	,000	,672	1	,412
Constructionyear	,001	,0014	-,002	,004	,651	1	,420
Ethnicheterogeneity	,334	,4247	-,499	1,166	,618	1	,432
Mixedlanduse	-,191	,2477	-,676	,295	,593	1	,441
Distancepublicfacilities	,000	,0005	-,001	,001	,575	1	,448
Nuisance	-,897	1,3667	-3,576	1,781	,431	1	,511
Welfarebenefits	,016	,0297	-,042	,074	,285	1	,594
Crime	,065	,1366	-,203	,333	,225	1	,635
Distancestreetlighting	,004	,0103	-,016	,024	,181	1	,671
Edgedwellings	-,021	,0774	-,172	,131	,070	1	,791
Carsperhousehold	,056	,4987	-,921	1,034	,013	1	,910
Distancelowincomes	-9,465E-06	,0001	,000	,000	,007	1	,934
Distancepublictransportnode	-1,087E-05	,0004	-,001	,001	,001	1	,976
(Intercept)	-,522	2,8439	-6,096	5,052	,034	1	,854
(Scale)	1[a]						
(Negative binomial)	1[a]						

a. Fixed at the displayed value.

Figure C.2: The final full model for the spring including all parameters and sorted by significance. This is the direct output from the SPSS software.

Parameter Estimates Summer							
			Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
Spatiallag	,050	,0067	,037	,063	56,199	1	,000
Buildingdensity	-,005	,0014	-,008	-,003	14,161	1	,000
Rentalproperties	,316	,1112	,098	,534	8,083	1	,004
Distanceethnicheterogeneity	-,001	,0002	-,001	,000	7,773	1	,005
Propertyvalue	1,133E-06	4,4919E-07	2,523E-07	2,013E-06	6,358	1	,012
Distancerentalproperties	,001	,0003	,000	,001	5,841	1	,016
Distancewelfarebenefits	,000	9,2862E-05	,000	-6,139E-06	4,105	1	,043
Distanceretailandcatering	,000	,0002	-,001	1,059E-05	3,604	1	,058
Distanceshrubbery	,000	,0003	-,001	7,979E-05	2,850	1	,091
Constructionyear	,002	,0014	-,001	,005	2,456	1	,117
Householdincome	-1,325E-05	8,8937E-06	-3,068E-05	4,183E-06	2,219	1	,136
Distancepublicfacilities	,001	,0005	,000	,002	2,178	1	,140
Distancetostreet	-,012	,0088	-,029	,006	1,742	1	,187
Accessibility	,140	,1194	-,094	,374	1,366	1	,242
Riskyproperties	,277	,2535	-,220	,774	1,191	1	,275
Electionturnout	,006	,0057	-,005	,017	1,074	1	,300
Distancelowincomes	,000	,0001	,000	,000	,902	1	,342
Distancestreetlighting	,008	,0099	-,011	,028	,683	1	,409
Nuisance	,649	1,1179	-1,542	2,840	,337	1	,562
Ethnicheterogeneity	-,200	,4113	-1,006	,607	,235	1	,628
Residentialmobility	,013	,0307	-,048	,073	,170	1	,680
Mixedlanduse	-,081	,2416	-,555	,393	,112	1	,737
Distancedemographicriskgroup	,000	,0004	-,001	,001	,094	1	,759
Edgedwellings	-,021	,0755	-,169	,127	,076	1	,783
Distancehighwayentry	2,112E-05	9,1633E-05	,000	,000	,053	1	,818
Distancecitycenter	-1,282E-05	5,9772E-05	,000	,000	,046	1	,830
Distancepolicestation	1,040E-05	5,2014E-05	-9,154E-05	,000	,040	1	,841
Crime	-,008	,1234	-,249	,234	,004	1	,951
Welfarebenefits	,000	,0308	-,060	,061	,000	1	,991
Distancepublictransportnode	3,765E-06	,0004	-,001	,001	,000	1	,992
Carsperhousehold	-,001	,5053	-,991	,990	,000	1	,999
(Intercept)	-2,296	2,6912	-7,571	2,978	,728	1	,394
(Scale)	1[a]						
(Negative binomial)	1[a]						

a. Fixed at the displayed value.

Figure C.3: The final full model for the summer including all parameters and sorted by significance. This is the direct output from the SPSS software.

Parameter Estimates Autumn							
			Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
Spatiallag	,052	,0066	,039	,065	62,186	1	,000
Distanceethnicheterogeneity	-,001	,0002	-,001	,000	12,847	1	,000
Buildingdensity	-,004	,0014	-,007	-,002	9,629	1	,002
Riskyproperties	,707	,2614	,195	1,219	7,314	1	,007
Propertyvalue	8,912E-07	4,5096E-07	7,304E-09	1,775E-06	3,905	1	,048
Distancedemographicriskgroup	-,001	,0004	-,002	-3,149E-06	3,872	1	,049
Distanceretailandcatering	,000	,0002	-,001	6,117E-05	2,516	1	,113
Distancestreetlighting	,014	,0097	-,005	,033	2,186	1	,139
Distancepolicestation	-7,390E-05	5,2811E-05	,000	2,961E-05	1,958	1	,162
Carsperhousehold	-,685	,5059	-1,677	,306	1,834	1	,176
Distancecitycenter	7,532E-05	5,9235E-05	-4,078E-05	,000	1,617	1	,204
Accessibility	-,148	,1205	-,384	,089	1,501	1	,220
Mixedlanduse	,255	,2321	-,200	,710	1,210	1	,271
Nuisance	1,330	1,2652	-1,150	3,810	1,105	1	,293
Distancewelfarebenefits	-8,289E-05	8,8460E-05	,000	9,049E-05	,878	1	,349
Constructionyear	,001	,0014	-,002	,004	,734	1	,392
Crime	,102	,1258	-,145	,348	,653	1	,419
Electionturnout	,005	,0056	-,006	,016	,645	1	,422
Rentalproperties	,083	,1068	-,126	,292	,603	1	,437
Welfarebenefits	,024	,0314	-,037	,086	,591	1	,442
Ethnicheterogeneity	-,313	,4144	-1,125	,499	,571	1	,450
Distancelowincomes	7,228E-05	,0001	,000	,000	,435	1	,510
Residentialmobility	,018	,0303	-,042	,077	,345	1	,557
Distancehighwayentry	4,307E-05	9,5751E-05	,000	,000	,202	1	,653
Distanceshrubbery	,000	,0003	,000	,001	,178	1	,673
Distancepublicfacilities	,000	,0005	-,001	,001	,139	1	,710
Distancerentalproperties	,000	,0004	-,001	,001	,124	1	,725
Distancepublictransportnode	,000	,0004	-,001	,001	,110	1	,741
Householdincome	2,476E-06	9,5670E-06	-1,628E-05	2,123E-05	,067	1	,796
Edgedwellings	-,018	,0765	-,168	,132	,058	1	,810
Distancetostreet	,001	,0091	-,016	,019	,026	1	,873
(Intercept)	-,906	2,8457	-6,483	4,671	,101	1	,750
(Scale)	1[a]						
(Negative binomial)	1[a]						

a. Fixed at the displayed value.

Figure C.4: The final full model for the autumn including all parameters and sorted by significance. This is the direct output from the SPSS software.