University of Groningen - Artificial Intelligence
Sentient Information Systems BV

Master thesis October 2, 2007

Evaluating agent-based modelling as prediction tool for crime

Author:
Jeroen van Dijk
Student no. 1290126

Supervisors:
Internal: Bart Verheij
External: Rob van der Veer

Abstract

Crime forecasting is an instrument of growing importance for police enforcement around the globe. Police departments in the Netherlands have started first trials with forecasting techniques. Sentient Information Systems BV, the company at which I conducted my research, provide crime analysis software for the police and were interested in alternative techniques to forecast crime. Because of this partnership, anonymous data on crime incidents and criminal individuals was available at Sentient for development purposes.

Existing crime forecasting methods use data on crime incidents and related variables. This research studied the possibilities of the use of data on criminal individuals for the forecasting of crime. A natural choice for the modelling of crime on an individual level is the agent-based modelling (ABM) methodology. Previously developed crime ABM models have been investigated to find useful theories, techniques and ideas. Because the previous models were not meant for crime forecasting we used criminological literature to find additional ideas and techniques.

This research gives an overview of an effort to use the ABM methodology to simulate crime based on data on individual criminals. The most important results are an overview of useful crime theories and techniques for the ABM methodology, a first-effort implementation of an ABM model to predict crime with individual data, a discussion of the limitations of this approach and suggestions for future work.

Table of Contents

1 Introduction

In the late 1980s until the early 1990s many developed countries realised that the traditional mode of policing was not suitable to fight the rapidly increasing crime rates. In the UK this came together with a significant financial constraint (Chainey and Ratcliffe 2005). This problem caused the police departments to review their strategy towards fighting crime for more efficiency. Recently, automated forecasting techniques have been developed to make predictions on crime numbers for areas in order to better position and schedule police manpower. With these techniques crime can be better predicted in time and place, allowing police departments to plan their use of manpower more efficiently. Police departments in the Netherlands have therefore shown interest in crime forecasting.

This research, a collaboration between the University of Groningen and Sentient Information Systems BV, is interested to find alternatives to existing crime forecasting techniques. Sentient is a company that applies data mining solutions for several different kinds of customers. Sentient has developed DataDetective, a software suite for sophisticated data analysis. The Dutch police use DataDetective to match and mine their criminal databases. DataDetective enables the police organisation to rapidly find complex relations between suspect data, incident data, government data, weather records and socio-demographic data. Future versions of the DataDetective will encapsulate a novel technique for the forecasting of crime trends in time and space. Sentient is interested in new alternatives to their current technique in order to improve their service to police organisations. In this research we will propose and built a model based on a different methodology.

In this thesis we will propose a model that simulates future crime from an individual perspective. The inspiration of the proposed model is based on a common police practice to prevent crime. This police practise is to keep an eye on the individuals that have been convicted or suspected of crimes in the past. The experience is that ex-offenders are more likely to commit new crime incidents than persons that are not known by the police. Police officers thus keep an eye on offenders when they return in society after a conviction, because they might lead them to new crime incidents. Ex-offenders are thus used as a predictor of future crime. This police practice has inspired us to predict crime from an individual perspective instead of the crime incident perspective that is used by Sentient's current technique and other crime forecasting techniques. In crime forecasting techniques that use a crime incident perspective the number of crimes or crime incident data is used to forecast future crime trends. Sentient's current method and some other methods also use other correlated variables, but we have not found an example where data on individuals has been used.

One of the hypothesized advantages of the individual perspective to the macro perspective is that crime trends caused by a specific offender will cease or start again when this offender is respectively constrained by a prison sentence or set free again. The idea is that by using an individual perspective, one can use specific information that is available about (criminal) individuals. This information is to our knowledge currently not used in other forecasting methods. For example, one motivation to use information on offenders is that they can be kept prisoner or constrained in other ways. During this period of restriction these criminals are not able to commit crime (at least we may hope so). The possible crime trends that were caused by these criminals will probably not be proceeded or proceeded by different individuals. The crime trends will at the least be influenced by the arrest or conviction of these persons. This is in

contrast with other methods to forecast crime that ignore data on individual criminals and are based on criminal incident data and other (cor-)related variables. These methods can, possibly, wrongly forecast that a certain crime trend will prolong in the future while the offender responsible for this specific trend has been captured yesterday. Hence, this research is thus based on the belief that the focus on the individual perspective with the use of individual data has an additional value to current crime analysis methods.

Sentient has built up close relations with the police departments Amsterdam-Amstelland and Midden-West Brabant in The Netherlands. The police district in Midden-West Brabant has already used the forecast technique developed at Sentient for the strategic positioning of road blocks. Both districts are interested in new techniques to improve their efficiency of police enforcement. Therefore, nameless crime data from Midden-West Brabant has been made available for this research at Sentient. This data contains information on crime incidents and the involved individuals[1]. The most important properties of this data that are used in this research is the time of occurrence, the place of occurrence and the crime type of a crime incident.

A first glance at literature for individual based crime modelling provided a motivation for the use of the Agent-Based Modelling (ABM) methodology. Gunderson and Brown (2000) described a multi-agent methodology to predict physical and cyber crime. This methodology describes how data mining techniques such as clustering can be used to discover criminal agents from crime data. These criminal agents can then be allowed to interact in a synthetic environment that is constructed from data of the environment. The output of Gunderson and Brown's proposed model is a threat surface, in which regions with high threat values represent regions with high likelihood for future criminal events. A complete implementation of the proposed methodology of Gunderson and Brown (2000) has until this moment not been published yet[2]. The PhD-thesis of Gunderson (2003) shows the implementation of a part of the methodology described in Gunderson and Brown (2000). Gunderson (2003) showed how preferences of criminals can be extracted from crime incidents with the use of a clustering algorithm. Furthermore, Gunderson showed how these preference structures could be used to make a prediction about future crime. The work of Gunderson and Brown (2000) provided the main direction to our research. We will, however, not derive the criminal agents from the crime incident data as in Gunderson (2003), but we base our criminal agents on suspect data and derived some of their preferences from the coupled crime incident data. In section 2.4 we will discuss how the work of Gunderson and Brown (2000) has contributed to our research.

With the ideas of Gunderson and Brown (2000) in mind we have formulated research questions discussed in the next section. After we have discussed the research questions we will summarize why and when Agent-Based Modelling (ABM), our research methodology, should be used. Finally, we will continue to describe the resources we have available in this research.

In the subsequent chapters we will work out the research questions formulated in this chapter. In chapter 2 we will further investigate literature on ABM models that simulate crime to find more theories and aspects valuable for our purpose. At the end of this chapter we describe the results of this search: a summary of the most relevant agent-based crime models and a

[1]The data available at Sentient is nameless meaning that no connections can be made between the data on

[2] Donald E. Brown wrote in an email conversation that the work of Gunderson and Brown (2000) has been continued, but no papers were published on it. He also confirmed that a simulation for predicting attacks has been built, but has not been tested yet.

description of concepts that can be reused in this research. We will make a choice for a subset of these concepts. Chapter 3 discusses insights and techniques from crime literature that are used in our model. The theories useful for our model will be summarized at the end of this chapter. The properties of the ABM methodology are discussed in chapter 4. The design and implementation of our model is discussed in 5 as well as the results of experimentation with the model. Finally, in chapter 6 we will discuss our research and do suggestions for future research.

1.1 Research Questions

The central claim of this thesis is that an individual perspective and the Agent-Based Modelling methodology are fruitful for crime prediction. A fully operational agent-based model that accurately predicts future crimes (compared to other methods) would obviously be the best proof for this claim. Unfortunately, my research does not extend that far. This is because, if such a model is at all possible, the complexity of such a model will be high and therefore it takes an enormous amount of time needed to design, implement and test this kind of model, and by probably more than one person. For this reason we will design and build a model based on a limited set of principles and a limited complexity to show the value of our approach.

We will defend the central claim: "an individual modelling perspective and the Agent-Based Modelling methodology are fruitful for crime prediction" by answering the research questions below throughout this thesis. The first research questions try to answer if there are theoretical grounds and techniques for an ABM model on crime simulation:

1. **What theories are useful for the modelling of crime in an ABM?**
2. **What type of crime is suitable to simulate with an ABM?**
3. **What techniques are useful for the modelling of crime in an ABM?**
4. **What data is available and is useful as input for our ABM?**

The first two questions will be answered by the discussion of previous ABM models that simulated crime in chapter 2. The third question is answered in parts in the discussion of previous ABM crime simulations, crime theory and the discussion of the ABM methodology. Question 4 is already partly answered in this chapter in the last section. However, we are also interested in additional data that are not available just by the existence of this project. In the discussion of this thesis, chapter 6, we will also outline why the availability of certain data is currently one of the main obstacles for some fundamental insights such as the effect of police enforcement on crime. Additionally to the theoretical research questions we have also defined practical research questions which will be answered when we build the model:

5. **What are the possible uses of ABM for crime analysis?**
6. **How can we evaluate our ABM?**
7. **Can we build an ABM based on individual data that provides crime results similar to real crime numbers?**

Question 5 will be answered at the end of 2 where we outline the possible uses of ABM for crime analysis. This summary of possible uses is reviewed in chapter 6. Question number 5 will be answered throughout this thesis and an evaluation method will be used when we compare our model predictions with real observations. Question number 7 will be answered in chapter 5 where we build and evaluate our model. When all these questions are answered we can defend or reject our central claim.

1.2 Agent-Based Modelling: Why and When?

In the previous parts we have outlined why we want an individual perspective to simulate crime and we have formulated research questions. Here we will summarize the reasons why and when ABM should be used in general. An additional discussion of the methodology can be found in chapter 4.

Bonabeau (2002) explains the advantages of ABM over other modelling techniques. Bonabeau writes that ABM is just a mindset, a synonym of ABM could be microscopic modelling and an alternative is macroscopic modelling. Bonabeau (2002) captures the advantages of ABM over other modelling techniques in three statements:

- ABM captures emergent phenomena
  Emergent phenomena result from the interaction between individual agents. The whole is more than the sum of the parts. For example, a traffic jam is caused by the interaction between the individual drivers. Emergent phenomena can be counterintuitive, e.g. the traffic jam possibly moves in the direction opposite to the one of the cars.
- ABM provides a natural description of a system
  For many cases ABM is most natural for describing and simulating a system. Bonabeau (2002) gives the following example:

  [...]It is more natural to describe how shoppers move in a supermarket than to come up with the equations that govern the dynamics of the density of shoppers. Because the density equations result from the behaviour of shoppers, the ABM approach will also enable the user to study aggregate properties. ABM also makes it possible to realize the full potential of the data a company may have about its customers: panel data and customer surveys provide information about what real people actually do. Knowing the actual shopping basket of a customer makes it possible to create a virtual agent with that shopping basket rather than a density of people with a synthetic shopping basket computed from averaging over shopping data.

- ABM is flexible in several ways.
  It is, for example, easy to add more agents to the model. Also adding more (complex) behaviour and interaction to the models agent is a natural operation. Another way of flexibility is the ability to change levels of description and aggregation: one can choose to play with the model at a system level or at an individual level.

Bonabeau (2002) summarizes when it is best to use ABM, these reasons and more are captured in more recent articles of Macal and North (2005; 2006) where they sum up the following reasons (quoted):

- When there is a natural representation as agents
- When there are decisions and behaviours that can be defined discretely (with boundaries)
- When it is important that agents adapt and change their behaviours
- When it is important that agents learn and engage in dynamic strategic behaviours
- When it is important that agents have a dynamic relationships with other agents, and agent relationships form and dissolve

- When it is important that agents form organizations, and adaptation and learning are important at the organization level
- When it is important that agents have a spatial component to their behaviours and interactions
- When the past is no predictor of the future
- When scaling-up to arbitrary levels is important
- When process structural change needs to be a result of the model, rather than a model input

Many examples are given in the articles of Macal and North (2005; 2006) and Bonabeau (2002). There are a few reasons that are relevant but have not been (explicitly) mentioned above. In economics, to model markets with traditional techniques one had to make assumptions about the homogeneity of agents making perfect decisions and long-run equilibriums making these problems analytically and computationally tractable. Therefore an additional reason to use ABM is when the population modelled is heterogeneous, each individual is (potentially) different (Bonabeau 2002). The second additional reason why ABM is becoming more useful is because increasingly more micro-data is obtained. For example, big supermarket concerns have enormous amounts of data of their customers. They know the favourite peanut butter and the favourite shopping hours of a specific customer. This data can be used to more realistically model the behaviour of their customers (Macal and North 2005; 2006). Third, the ever growing computer power allows us now to run simulations with many sophisticated agents. The computational complexity that comes with ABM is therefore becoming a smaller problem (Bonabeau 2002). Finally, to make the list of ingredients for successful ABM complete several modelling platform for ABM have emerged. In the past only experienced programmers were able to create an ABM model. ABM platforms become easier to use with every new release. In the future these platforms are expected to have the same complexity as the software now used for traditional methods (Samuelson and Macal 2006).

Above several reasons are outlined when to use ABM. All except for the last reason are applicable to the domain of crime modelling.

- Criminals, victims and police units are all natural representations of agents.
- Criminals, victims and police units can adapt their behaviour therefore for a model to be based on reality agents should be able to adapt and change their behaviours.
- The behaviours of criminals, victims and police units can be defined discretely (see Groff 2006, in section 2.2).
- Especially criminals can learn and engage in dynamic strategic behaviour when offending.
- Criminals can form dynamic relations with other criminals that can form and dissolve. Co-offending is an example of such a dynamic relation.
- For the simulation of crime the spatial component is an important aspect to the behaviour and interactions of criminals, victims and police units.
- The past is no direct predictor of the future. Criminals' behaviour is especially sensitive to police enforcement.
- Scaling up is important because the police is interested in the effect on police enforcement on crime numbers.

In summary, ABM is a methodology that can be used for crime modelling for several reasons.

1.3 Data resources

This research is a collaboration between the University of Groningen and Sentient Information Systems BV. The latter has close relations with the police departments in The Netherlands, Amsterdam-Amstelland and Midden-West Brabant. Together with our research there is another running research at Sentient that compares the forecasting technique at Sentient with other techniques. Both projects fall under another project at Sentient called GeoPredict. The goal of these projects is to research new crime forecasting techniques. Crime data has been made available for the GeoPredict project at Sentient.

The database contains anonymous[3] data about arrested suspects and incidents of the past 5 years of Midden-West Brabant (June 2001 to June 2006). Some incidents are coupled to suspects and suspects are coupled to one or more incidents. For the analysis of the crime data and the construction of the data sets we can use DataDetective, the software suite for sophisticated data analysis developed by Sentient. The advantage of DataDetective is that it is very user friendly for our purpose. Only for the coupling of antecedents of criminals we had to create a program to put the antecedents of each criminal in an xml file. In the Appendix is explained how DataDetective is used.

Additionally to the crime data we have Geographic Information System (GIS)[4] maps available to create a more realistic environment. GIS maps contain data about an environment that is spatially referenced to the earth. The street vector file we have available has no 6-digit zip codes, which is necessary to provide a small enough level of detail. Therefore, instead of streets, a file with just the 6-digits zip codes as points will be used. Unfortunately, this makes movement between zip codes less realistic because the agents have to 'jump' from one zip code to another. More on this is said in chapter 5.

There is no data available on convicted criminals. This data is hard to get to. Therefore we have to assume that arrested suspects are the criminals that have committed the crimes where they are suspected of. Because many are suspected of several antecedents this seems a reasonable assumption. From now one we will call these suspects criminals for simplicity.[5] In chapter 5 we will explain how we will use the above described data in our model.

1.4 Summary

In this chapter we have introduced the background of this research and we have defined a set of research questions to defend the central claim of this thesis: "an individual modelling

---

[3] Meaning that names of victims and criminals and house numbers are removed, and noise is added to zip codes.

[4] A more complete definition of GIS can be found on http://en.wikipedia.org/wiki/Gis: *A geographic information system (GIS) is a system for capturing, storing, analyzing and managing data and associated attributes which are spatially referenced to the earth. In the strictest sense, it is a computer system capable of integrating, storing, editing, analyzing, sharing, and displaying geographically-referenced information. In a more generic sense, GIS is a tool that allows users to create interactive queries (user created searches), analyze the spatial information, edit data, maps, and present the results of all these operations. Geographic information science is the science underlying the geographic concepts, applications and systems, taught in degree and GIS Certificate programs at many universities.*

[5] Despite this assumption we do support the Presumption of innocence (http://en.wikipedia.org/wiki/Presumption_of_innocence).

perspective and the Agent-Based Modelling methodology are fruitful for crime prediction". Furthermore, we have discussed the reasons when and why to use ABM as methodology. In the last section the data that is available was described. In the next chapter we will continue to discuss more ABM models on crime to answer our research questions from earlier work.

2 Previous work on crime ABM

In the introduction we have set some goals for this research. In this chapter we will discuss previous work on crime modelling with ABM. First we will discuss previous work globally. Next, we will continue to highlight the most relevant work and how their work can be reused. This chapter ends with a summary and some conclusions regarding the direction of this research. Some answers are found on research questions as well.

One of the first modern agent models related to crime is from Epstein (2002). In his article an ABM of civil violence is presented. His model comes in two variants. In the first a central authority seeks to suppress decentralized rebellion. In the second variant a central authority seeks to suppress communal violence between two warring ethnic groups. Epstein's model is not intended to replicate a particular case, but it is intended to generate certain characteristic phenomena and core dynamics. The conclusion of Epstein (2002) is that agent-based methods "offer a novel and promising approach to understanding the complex dynamics of decentralized rebellion and interethnic civil violence, and, in turn, to fashioning more effective and efficient policies to anticipate and deal with them".

Van Baal (2004) discusses the use of computer simulations for studies into criminal deterrence. Van Baal outlines his computer program which is a modelling environment designed to investigate the effect social networks and crime deterrence policies have on a population of potential offenders. The work of van Baal (2004) is not only relevant to use because it models crime, it is also relevant because it provides good examples of how to statistically evaluate an ABM.

Bosse, Gerritsen et al. (2007) emphasise the lack of modelling of physical and mental aspects in other crime models and therefore present an ABM with agents that have a complex internal model by extending the general BDI-agent model (Georgeff and Lansky 1987; Rao and Georgeff 1991, adopted from Bosse, Gerritsen et al. 2007). Brantingham and Brantingham (2004) describe a model that is based on criminological theory.

In the next sections we will discuss four models, that are relevant to this research according to our research questions. These models provide concrete examples of model elements and used theories.

2.1 Liu et al.'s model: predicting crime patterns with ABM

The work of Liang (2001) is the first work found on crime simulation with cellular automata. His work is continued in Liu, Wang et al. (2005). In this work the possibility of simulating individual crime events and generating plausible crime patterns is explored. In Liu, Wang et al. (2005) the Routine Activity Theory (RAT, see section 3.1) is applied to explicitly model crime processes. The model is calibrated to closely fit real crime patterns. Also the potential use of a cellular automata simulation model[6] as a virtual laboratory for testing new crime theories is evaluated. Street robbery is used as an example to illustrate the characteristics of the model.

[6] Because there is overlap between cellular automata models and ABM's it is not clear what the exact differences are between these types of models. However, cellular automata model's can be considered to be a subset of ABM's in which the agents are spatially-explicit, homogeneous and dense. ABM's are thus less constricted. See Amblard (2002) for a small overview and also http://www.red3d.com/cwr/ibm.html for some extra information.

Liu, Wang et al.'s motivation for using simulation to test the RAT is because the theory is non-linear. The interaction between offender, target and place is complex. Data about the status of the offender (motivation), target (desirability and guardian capability) and place (Management efficiency) is usually unavailable.

Crime incidents, both successful and failed, cause anxiety, fear, depression and hostility (Norris 1997; Hollway and Jefferson 2000, adopted from Liu, Wang et al. 2005). Liang (2001) and Liu, Wang et al. (2005) use *tension* in place agents and target agents as a surrogate concept to represent the overall psychological reaction to a crime event. When a crime occurs this adds tension to the crime location and target. Following the work of Liang (2001) and Liu, Wang et al. (2005) tension decreases in space (not for target agents) and time. The model of Liu, Wang et al. (2005) only considers the spatial propagation of tension when a neighbours' tension is higher, because people tend to pay more attention to bad news or high tensions.

Liu, Wang et al.'s model holds three entities: offenders, targets and crime places. An offender agent has two properties: location and motivation. The location on a given day and time is related to the offenders routine activities. Based on empirical evidence of Wright and Decker (1997, adopted from; Liu, Wang et al. 2005) that offenders do not travel great distances to commit robbery. Liu, Wang et al. (2005) make the assumption that the probability of an offender going to a place is inversely related to the distance from the offenders' home to the place when the distance exceeds a small threshold value (Brantingham and Brantingham 1993; Block and Block 2000, the latter adopted from Liu, Wang et al. 2005). A random process is then used to assign the location of an offender. Experienced criminals are more motivated than less experienced offenders. This motivation is increased by a successful robbery or decreased after being discouraged by a failed robbery. A novice offender tends to change its motivation at a faster rate than an experienced offender. A target agent has four properties: location, tension, desirability, guardian capability and has also four corresponding behaviours that update the value of these properties. Again a random process determines the placement of targets on the streets when the routine activities are unknown. More targets are placed on streets that are more accessible. Target tension only exists for targets that have been attacked by street robbers. Liu, Wang et al. (2005) assume that target agents decrease their desirability and increase their guardian capability after being robbed to avoid future attacks. It is not clear from the text what exactly the influence is of tension on the behaviour of the target agents. A place agent, that represent a potential crime place such as gas station or cafe, has the following properties: the accessibility of the place, place tension and management effectiveness. Occurrences of crime increase the tension of a place, including the neighbour place agents. A decrease or increase in tension also causes a decrease or increase in management effectiveness. The accessibility of a place depends on its connectivity to the streets and the capacity of the streets. Equation 2.1 shows how the properties of these agents are used to determine the likelihood of crime. δ stands for Desirability, μ for Motivation, $\propto$ for Accessibility, γ for Capability and ε for Effectiveness.

$$L = \frac{\delta\mu\alpha}{(0.1 + \varepsilon)(0.1 + \gamma)}$$

Equation 2.1 Formula for the likelihood of crime in the model of Liu et al. (2005)

The model is calibrated using a real crime data set. The model gave the following plausible results:

- Repeat location: simulated street robberies are located in a few locations being consistent with earlier findings of Eck and Weisburd (1995, adopted from Liu, Wang et al. 2005).
- Repeat victimization: A small group of victims is victimized relatively often.
- Repeat offending: a small number of offenders are responsible for a disproportionately large amount of crime (Spelman 1994, adopted from Liu, Wang et al. 2005).
- Increasing risks and difficulties and reducing rewards of crime reduces the opportunity of crime (Clarke 1992).

Liu, Wang et al. (2005) conclude that Routine Activity Theory cellular automata models have the potential to become a tool for improving understanding and control of crime patterns. In the long run these models could be used for bench testing policies prior to field experimentation and implementation. One of the limitations of the RAT cellular automata model is the parameter calibration. According to Liu, Wang et al. the best calibration may never be achieved due to the computational complexity of a cellular automata based simulation with a large number of parameters. Moreover, the calibration relies on the experience and expertise of the user.

In summary, Liu, Wang et al. (2005) show that a cellular automata, a special kind of ABM model, after calibration, can closely fit real crime patterns using concepts of the routine activity theory.

2.2 Groff's model: using GIS and ABM to test crime theory

The work of Groff (2006) demonstrates how formalizing theory in a computational laboratory can provide a better understanding of how spatio-temporal aspects of human activity influences the incidence and distribution of street robbery events. The point of this research is to operationalize the assumptions of routine activity theory in an artificial society and test whether the model outcomes matches the predicted outcomes of the theory. In this research the routine activity is formalized in a GIS ABM. Groff (2006) demonstrates several ways to improve the realistic value of the environment of her ABM. She uses block group level population figures to describe the distribution of residences across Seattle. Employment data is used to describe the number of employees per zip code area. The model has 18,024 points that are potential activity locations that identified through the use of retail and service establishments (e.g. groceries stores, convenience stores, dry cleaners, gyms and so forth). A street network database file is used to structure the movement of agents.

The model has two types of agents, named civilians and cops. Civilians have activity spaces and can have three kind of possible roles (offender, victim and guardian) depending on the particular situation. Cops are agents that just have formal guardianship (no crime occurs when there is a cop on a street). Civilians with criminal propensity can take up all three roles, those without can only be victim or guardian. Furthermore, each civilian has a unique set of characteristics including wealth and employment status. Factors such as guardianship, caused by a civilian that has the role of guardian or by a formal guardian (a cop), and the presence of a suitable target (the wealth of the potential victim) are considered by the civilian that has criminal propensity. Agents in the models have four places they visit each day: a home, a main node (e.g. work or school), and at least two other frequently visited places (such as a gym, grocery store, etc.). The paths taken between these places are structured and constrained by the street network of Seattle. The size of these activity spaces is influenced by the distribution of residential housing, jobs, schools, retail and services.

Groff (2006) has translated the guardianship of a police officer to be absolute. Thus no crime occurs when a police officer is around. The decision behaviour of the agents is based on the rational choice theory. Groff has several conditions and 12 variables to experiment with. Examples of parameters are: "the number of police", "time to wait before able to re-offend", "initial wealth distribution", "perception of target suitability" and "the perception of guardianship". Groff uses wealth and a utility value for street robbery to determine whether or not a crime should be committed. Second, a line of sight is used for the different perceptions of the main elements of routine activity theory. The perceptions are, although based on theory, not empirically based.

Groff (2006) has used her ABM to validate the theoretical framework of routine activity in crime. She used the crime of street robbery in Seattle as a basis to test hypothesis created from routine activity theory (originally developed by Cohen and Felson 1979, see section 3.1). The focus of her research was to use the ABM as a virtual laboratory to vary different variables to see what happens with the crime rates. She used several techniques to verify her result: Ripley-K to see if the spatial clustering of crime events in her model is not random and an ANOVA (analysis of variance) test to determine whether the differences between the conditions were significant. Furthermore, a visual inspection to analyse the spatial pattern was done using kernel density. In summary, this research shows how environment data can be used to make a model more realistic and how an ABM can be analysed.

In summary, Groff has formalised the routine activity and rational choice theory in an ABM model to test the routine activity theory for a real environment.

2.3 Melo et al.'s model: adaptive police planning

A more practical use of ABM for policing can be found in Melo, Belchior et al. (2005). This article describes a tool for assisting the investigation of different strategies of agent physical reorganisations. It is used in the public safety domain for helping in the study of strategies of preventive policing. The aim of the model is to analyze and compare the effect of different police routes on the reduction of crime rates. Melo, Belchior et al.'s work is extended in Reis, Melo et al. (2006) with a genetic algorithm to find the most optimal police routes and crime hotspots automatically.

There are several entities in the model: notable points, an emergency central (CIOPS), police units and criminals. There are two objects that are part of the simulation that are not characterized as agents: police stations, the starting/end point of police routes and the criminal's residence, the point where the criminals are during the period that they are not committing crimes. Notable points are establishments that are potential targets for a criminal, such as gas stations, lottery houses, squares and shopping centres. Their main properties are financial value available at the moment, public illumination of the surroundings, demographic density and tension point. The financial value and demography vary according to the time of the day. Tension point is a representation of the state of the victim after the occurrence of a crime. This concept of tension is similar as in the models of Liang (2001) and Liu, Wang et al. (2005). The Emergency Central (CIOPS)'s function is to receive SOS calls from notable points and send the police team that is closest to the place of occurrence. Each police unit has at least one route, accomplishing the preventive policing of the area that they occupy. The police team only leaves the route when a call is received from the CIOPS agent.

A criminal is the one that causes crime occurrences in the model. They have two state variables: ideal satisfaction and current satisfaction. The ideal satisfaction represents the necessary value of the criminal to be satisfied and it determines that the criminal will not make any criminal analysis on the environment until being unsatisfied again. The current satisfaction represents the satisfaction value that the criminal possesses at the moment. All criminals possess a vision that allows them to see cells (the environment is represented by a grid of cells). The criminal has a personality, novice, intermediate or dangerous) that determines the experience. Depending on this experience criminals have a different ideal satisfaction and a different preference to certain targets.

For the decision to commit a crime the following factors are analyzed: the existence of police within the area of the criminal's vision and the level of public illumination of the notable points at the moment of the analysis. The personality of the criminal also interferes in this decision. For example, some personalities prefer notable points that are badly illuminated. If the criminal commits a crime this will increase his current satisfaction, the notable point will have zero financial value and tension is spread in the notable place and neighbouring areas. The criminal makes a comparison between his current satisfaction and the ideal satisfaction. If the current satisfaction is greater than the ideal satisfaction then he does not intend to commit crimes and returns to his residence where he can stay until unsatisfied again (this satisfaction decays every tick). No criminal analysis is done if not unsatisfied.

The above described model is designed to be used by police agencies to get insight in the different strategies for patrolling. The model of Reis, Melo et al. (2006) goes even further and finds the optimal patrolling routes for police teams given the parameters of the model. In summary, this discussed work shows a practical example in which an ABM is used for the planning of police patrols.

2.4 Gunderson's work: deriving preferences of criminals

The research of Gunderson and Brown (2000) and Gunderson (2003) was already mentioned in the introduction and has been an inspiration of this research.. The first research presents a method to forecast crime by simulating the behaviour of criminals that are derived from the data of crime incidents. The criminals are derived by the grouping criminals on preferences from crime data. Every incident has unique properties, time of the day, weather, etc.. These properties define the preferences of the derived groups of criminals. When the criminals are constructed they interact with the environment that is divided into three main surfaces; an opportunity surface, a guardian surface and a distance surface. The concepts opportunity surface and guardianship surface come from the concepts used in the routine activity theory and rational choice theory. The opportunity surface contains all features that influence the perception of opportunity e.g. the median income of an area. The guardianship surface contains all features that influence the perception of being prevented from carrying out a crime. The resulting behaviour of the criminal agents is the forecast for crime and is compared with actual crime data. This research almost defends the central claim of this thesis. However, the PhD-research of Gunderson (2003) does not exactly implement the model we had expected based on Gunderson and Brown (2000). Furthermore, we are interested in building an ABM model that uses data on individual criminals. We will discuss the research of Gunderson (2003) below because it does present an interesting concept that could possibly be integrated in an ABM.

Gunderson (2003)'s research is interesting because it demonstrates how preferences of criminals in crime can be obtained from crime data. Agents are constructed from preference structures extracted from crimes committed in the environment. This means that an agent can represent a group of criminals, one criminal or just a part of a criminal. This can be explained by the following example. Many robbers can have the preference for weather, time of day and place to choose their targets so one agent would represent this group. However, one of these robbers could do some weekend burglary as well. This part of the robber would be represented by another agent. (Gunderson) uses her model to make a prediction of crime by creating a regression model for each discovered agent. In summary, the model of Gunderson shows how to extract different types of criminal preferences from criminal incident data. These preferences can possibly be coupled to existing criminals in order to simulate their behaviour.

2.5 Summary and conclusions

In this chapter we have discussed ABM examples that have simulated crime in one way or the other. From this review we have learnt what the current applications are of ABM for crime modelling. The first conclusion is that what we will do is quite new to the field of ABM as well as crime modelling, this research is the first known research that tries to predict future crime with an ABM with the use of criminal data on individuals. In section 1.1 we have stated our research questions. We can already answer some of these partly. The first research question "What theories are useful for the modelling of crime in an ABM?" is partly answered by the work of previous crime ABM models (Liu, Wang et al. 2005; Groff 2006; Groff 2007). These models use the routine activity theory and the rational choice theory in their models. We will explore these theories in the next chapter in section 3.1 and 3.2.

We can also answer the second research question: "What type of crime is suitable to simulate with an ABM?". Street robbery is simulated in two of the discussed models (Liu, Wang et al. 2005; Groff 2006; Groff 2007). Groff mentions four advantages of street robbery:

- *[...] it is an instrumental crime and thus more likely than expressive crimes to involve a rational decision process (Clarke and Cornish 1985; Cornish and Clarke 1986; Walsh 1986).*
- *[...] street robbery is by definition restricted to the street or some other exposed area rather than in a residence or business and thus involves the public intersection of offender and target in space and time.*
- *[...] police presence is assumed to be more effective against street level crime then crimes that take place indoors (e.g. domestic violence).*
- *[...] street robbery elicits a high level of fear among residents because of its suddenness and potential for serious injury and thus is of considerable interest to both law enforcement and the public (Feeney, 1986).*

In the next chapter in section 3.2 properties of decisions in street robbery are presented. In section 3.5 the offender profile for street robbery is discussed.

The next research question we can partly answer is: "What are the possible uses of ABM for crime analysis?". Groff (2006) showed that an ABM can be used to test a crime theory; the routine activity theory. Melo, Belchior et al. (2005) showed that an ABM can be used to find

optimal police patrolling routes. In Liu, Wang et al. (2005) was shown that an ABM can be used to replicate crime patterns.

The final research question we can discuss is: "How can we evaluate our ABM?". Groff has shown three ways of evaluation: Ripley-K, Kernel Density and the ANOVA test. We will repeat these methods later when we know more about the domain of crime.

The four discussed models above provide elements that can be integrated into our model. We will summarize these elements below as requirements for our work:

- The use of routine activity theory, see section 2.1 (Liu, Wang et al. 2005), and section 2.2 (Groff 2006; Groff 2007).
- The use of the rational choice theory, see section 2.2 (Groff 2006; Groff 2007).
- The use of the concept of Tension, see section 2.1 (Liu, Wang et al. 2005), and section 2.3 (Melo, Belchior et al. 2005).
- The use of environmental data to create a realistic environment, see section 2.2 (Groff 2006; Groff 2007).
- The use of activity spaces in which agents are active, see section 2.2 (Groff 2006; Groff 2007).
- The use of the crime type street robbery, see section 2.1 (Liu, Wang et al. 2005) and section 2.2 (Groff 2006; Groff 2007).
- The use of the concept satisfaction and ideal satisfaction for criminals, see section 2.3 (Melo, Belchior et al. 2005).
- The use of genetic algorithms to optimize police enforcement, see section 2.1 (Liu, Wang et al. 2005).
- The use of an opportunity surface and guardianship surface to, respectively, promote and inhibit criminal behaviour, see 2.4 (Gunderson and Brown 2000).
- The use of preferences of agents derived from crime data, see 2.4 (Gunderson and Brown 2000; Gunderson 2003).
- The visual inspection of the predictions of the model, with or without kernel density, see section 2.1 (Liu, Wang et al. 2005), and section 2.2 (Groff 2006; Groff 2007).
- The use of Ripley's K to see if the spatial distribution of crime is not random section 2.2 (Groff 2006; Groff 2007)..
- The use of One-way ANOVA to test significant aggregated macro variables section 2.2 (Groff 2006; Groff 2007).

The list with requirements is too long to integrate completely into this research. Therefore is chosen to continue with only some of these concepts. The routine activity theory is common to all except for one of the discussed works. We will therefore discuss the concepts of routine activity theory in the next chapter. The same holds for the rational choice theory. Furthermore, Groff has created activity spaces for the criminal agents in her model based on environmental data that is not available to us. For this reason a method from theory to derive the activity spaces from the data we have on criminals is discussed in the next chapter. Evaluation techniques are investigated further when we know more about our own model. In the next chapter we will discuss crime literature to obtain more theoretical insights. In chapter 6 we will say more about the concepts we have not used in suggestions for future work.

3 Discussion of Crime Theory

In this chapter we will discuss criminological theories and techniques for the understanding of crime. Since we have chosen to model street robbery, we will where possible focus on this specific crime type. Otherwise we will discuss the general principles of crime. We will start off by describing two important crime theories, the routine activity theory and the rational choice theory. These theories have been used to explain macro behaviour of crimes as well as behaviour on an individual level. The discussion on the rational choice theory also contains a part on the theory on the Residual Career Length, a study on the remaining career length of criminals. In the next section the theory on Repeat Victimisation is introduced. This theory explains why some places and some victims are more victimised than others. This theory has inspired the idea of how previously discussed opportunity and guardianship surface can be created in an ABM model. The subsequent section describes theory and techniques to estimate the geographic profile of an offender. Geographic profiling provides a technique to estimate the activity spaces discussed in the previous chapter. In the following section the profile of the typical offender and victim in street robbery will be described to give us background information. The next section introduces predictive crime mapping. Finally, we will summarize our findings and answer some research questions. Furthermore, here we will define the concepts that will be used in our model.

3.1 Routine activity theory

Routine activity theory was initiated by Cohen and Felson (1979) to explain predatory crime on a macro-level. Since then it has been developed to become a useful mechanism in the examination of criminal opportunities and crime prevention (Chainey and Ratcliffe 2005). Clarke and Felson (2004) have continued their work on the subject. The theory is based on the assumption that criminal behaviour is directed by opportunities in the routine activities of the potential offender. The theory is summarised by Farrell (2006) as follows:

*A crime occurs when a **suitable target** and a **potential offender** meet at a suitable time and place lacking **capable guardianship** [emphasis from the cited author].*

Crime opportunity is defined in Equation 3.1 by Chainey and Ratcliffe (2005).

$$\text{crime opportunity} = \text{potential offender} + \text{suitable target} - \text{capable guardian}$$

Equation 3.1 The definition of crime opportunity by Chainey and Ratcliffe (2005)

Targets can be persons, businesses or other grounds, vehicles or particular consumer products. Suitable targets, however, are a subclass of targets where they are perceived in certain ways. In case of a street robbery the 'suitable target' is a person perceived to carry valuable items, be unarmed and is unlikely to fight back. In case of a burglary of a house, a 'suitable target' could be a house, perceived to contain things of value and is unguarded.

The terms 'suitable' and 'perceived' are important here. One offender can perceive an object as a 'suitable target', while a second offender does not share this perception[7]. The suitability varies between criminals, types of crime, by site, by situation and with variations in the settings (Brantingham and Brantingham 1993). The risk of crime can be reduced when the perceived

[7] This has strong relations with the rational choice theory which we will discuss in section 3.2.

target suitability is decreased. This could be done by increasing the security of the target (Farrell 2006).

A potential offender can be anyone around us. The idea is that under the right circumstances anyone can commit a crime and thus is a potential offender (Walsh and Ellis 2003, adopted from Farrell 2006). However, there is a small group of career criminals that are responsible for a disproportionate amount of crime (Blumstein, Cohen et al. 1986; Townsley and Pease 2002).

As with the previous mentioned terms, capable guardianship is also adjustable in the sense that it depends on the circumstances and on the perception of others. A guardian is a broad term in this sense that it can be anything from a dog to a CCTV camera. A capable guardian can be a store manager or just another customer that is thought to be alert. Circumstances determine whether or not the guardian is capable. For instance, two colleagues walking together can be each other's guardians. The same holds for parents that company their children. The capability of the guardian depends also on whether they are perceived of calling the police or interfere directly. Note that again it is important that the guardian is perceived as being capable, and not whether the guardian is really capable. Another important premise of crime is that the potential offender and suitable target have to interact in time and space for the occurrence of a crime. Police uses this fact, for instance, in a soccer stadium where different supporter groups are physically separated and where not possible capable guardians are positioned (in the form of stewards).

3.2 Rational choice theory

In this section we will describe relevant chapters from the book of Cornish and Clarke (1986). This book was the outcome of a conference by the Home Office at Christ's College, Cambridge, England, in July 1985. The conference was designed to provide a forum for exploring and elaborating a decision-making approach of the explanation of criminal behaviour. Although newer articles have been written about the rational choice theory this book is still a recommended reading according to the Centre for Problem-Oriented Policing[8].

The cited text below (Cornish and Clarke 1986, p. 1) was an important starting assumption of this conference:

[...]offenders seek to benefit themselves by their criminal behaviour; [...] this involves the making of decisions and of choices, however rudimentary on occasion the processes might be; and [...] the processes exhibit a measure of rationality, albeit constrained by limits of time and ability and the availability of relevant information. [Assumed is that... JvD]

Even though this assumption was recognised to be fitting some offences better than others, it was felt that the rational components were also present in crimes that seemed pathologically motivated or impulsively executed. In what follows the relevant parts of Cornish and Clarke (1986) are discussed.

[8] See http://www.popcenter.org/library-recommended_readings_2.htm. Last visited on 6-5-2007.

3.2.1 Rationality and risk

Walsh (1986) has interviewed offenders to assemble data on commercial burglary and robbery. According to Walsh (1986) the behaviour of burglars and robbers is rationally bounded. This conclusion originally comes from the work of Bennett and Wright (1984) who describe the choice to offend and the usual planning of the offence as conscious, only this rationality is limited to what seems reasonable to the offender given the condition at the given time he or she is in. Bennett and Wright (1984) prefer the idea of limited rationality because it is not presumed that the offender is taking all relevant variables in consideration each time a crime is considered. Other seemingly unrelated factors often take over when deciding to commit a crime. The conclusion of Bennett and Wright (1984, p. 152) is that offenders at the time of offending see their behaviour as being rational, although this can be completely different at another time when the offender is, for example, in a different state of mind.

Walsh (1986) noted that most criminals that get caught are often seen as irrational because of the risks taken. This is not correct because even with a high amount of rationality crime still involves risks. Crime therefore does not imply irrationality.

Walsh (1986) states that for rationality to be total the amount of information is rather infinite than finite. The example is given of a military special service situation where planners beyond a certain point stop trying to acquire more information realising that the aggressor always has the advantage, and justifying the outcome by the gain only. This also holds for the economic criminal, the criminal that has a financial motivation. According to Walsh's study offenders accept that things may turn out differently. They do not see this as being in their control or due to lack of foresight, but as a part of their 'job'. Within the group of robbers 52% had planned their robberies and of these individuals 25% had planned for months or years . [9]

The commitment of the robbers compared to the burglars is higher. 11% of the burglars said that nothing could stop them from offending a particular crime once planned, compared to 54% of the robbers. The typical way for all robbers to choose their victim was by knowledge acquired from employment, residence, observation or gossip (47%) (Walsh 1986).

Walsh (1986) has two alternative explanations to irrationality for the seemingly small profits for economical criminals. The first is the difficulty to predict exact gains in advance. The second emphasises the point of view of the criminal; although the gain can be small, it can be adequate for the offender's immediate requirements[10] and therefore subjectively much larger than they appear.

3.2.2 The decision to rob

Feeney (1986) has interviewed 113 California offenders charged with robbery and convicted of robbery or related offences. More than half of them said they did no planning at all, and over 60% said they had not even thought about being caught before the robbery. Over 50% of the money motivated offenders were using it for drugs or food. When first-time offenders were interviewed they indicated that they felt fear when approaching their victims. They felt sympathy with their victims and even sometimes left money when the victims would say they

[9] It should be noted that the distinction between bank robbery and street robbery is not made in the article of. Street robbery is not a crime that is very suited for planning since the target is a person and thus dynamic.

[10] The immediate requirements have overlap with the concept of the current satisfaction of a criminal (Melo, Belchior, et al. 2005, see section 2.3).

really needed it (for rent, for example). More experienced robbers tended to view their victims more as objects rather than persons and were more hardened and less fearful. This attitude was generally already present after a few robberies (Feeney 1986). An example is given by Feeney (1986) of a criminal that began robbing using heroin at an age of 13. By his 26th he had committed over a thousand robberies without a conviction until his present sentence. He had been arrested five times, but each time the charges were dropped. This illustrates why some robbers are fearless for getting caught. This fearlessness of getting caught is supported by Dutch research (see Ferwerda, Jonkmans et al. 1998 in section 3.5).

According to Feeney (1986) although some decisions of robbers do not seem rational, most of them are clearly rational according to the definition used by Clarke and Cornish (1985). The offenders choose robbery to satisfy their desires and needs. Also, whether or not they committed other types of crime, robbery was a carefully considered part of their repertoire (Feeney 1986). Feeney (1986) argues that these decisions would seem more rational if more planning and more concern about the possibility of arrest was identified, however, it is not that different from what normal people do in their daily lives. Experienced robbers say not to plan much. Their experience, however, compensates this.

3.2.3 The decision to give up crime

The decision to give up crime is often caused by a shock of some sort, by a delayed deterrence process, or both (a figure is shown in Cusson and Pinsonneault 1986, p. 74). This was already remarked by Conwell (in Sutherland 1937, p. 182). Most offenders suffered from such a shock during the last crime. Cusson and Pinsonneault (1986) provide the following definition: "Delayed deterrence is the gradual wearing down of the criminal drive caused by the accumulation of punishments." The successions of arrest and imprisonments have their effect on the long run. Offenders engender a pervasive fear which becomes extremely great over the years (Cusson and Pinsonneault 1986). Cusson and Pinsonneault (1986) mention the four components of delayed deterrence:

- A higher estimate of the cumulative probability of punishment.
  With age criminals raise the estimates of the certainty of punishment. Young criminals generally do not realize that each new crime increases the cumulative probability of getting caught. According to one of the by Cusson and Pinsonneault (1986) interviewed criminals: "Every time you commit one, you risk being arrested. The law of averages is against you; the prisons are there to prove it."
- Increasing difficulty in "doing time".
  When offenders get older they feel more that they are wasting time and ruining their lives.
- An awareness of the weight of previous convictions on the severity of the sentences.
  Criminals are aware that they get longer sentences with more crimes.
- Spreading of fear.
  The criminal is getting more paranoid, because he is always nervous of getting caught.

After some time when delayed deterrence has its effect or a shock has occurred, the will to pursue their criminal career becomes weaker. The offenders then have a period of crisis and conclusions are that theft does not pay enough and the criminal way of life becomes a problem (Cormier, Kennedy et al. 1959; Shover 1983; Cusson and Pinsonneault 1986).

Dying in prison is seen as the ultimate failure (Braly 1976). Thus giving up crime is not a positive decision, the wish to go straight, but a negative decision, avoid another imprisonment (Cusson and Pinsonneault 1986).

Hirschi and Gottfredson (1983) have made a convincing argument that there is a direct link between age and crime (Cusson and Pinsonneault 1986). Individuals that quit crime at the end of adolescence have a normal maturation. The ones that stop during their thirties have a late maturation (Cormier, Kennedy et al. 1965; Glueck and Glueck 1974, adopted from Cusson and Pinsonneault 1986). Experiences that accelerated the process of maturation are the discovery of reading, studying, learning a trade etc. Reading broadened their perspective according to the majority of the interviewed respondents of Cusson and Pinsonneault (1986). In some cases, the ex-offender is tempted to commit new thefts often because of money problems (Cusson and Pinsonneault 1986). Some of them lose their job, others have a bad regulation of their expenditures resulting in debts. Others committed crimes when they were idle, bored, hopeless. Meetings with former inmates made it easier to recommit.

3.2.4 Residual career length of an offender

In the previous section we have discussed reasons to give up crime. An interesting research direction are studies over the so-called Residual Career Length (RCL) and Residual Number of Offenses (RNO), meaning the remaining time and number of offenses in criminal careers up to the point of termination. These studies try to discover what exactly causes one criminal to stop after an imprisonment and the other goes on without any difference. Furthermore, very recent research has been working on a method to predict the RCL of criminals (Kazemian, Blanc et al.).

Recent research has shown the importance of the distribution of RCL. (Kazemian and Farrington 2006, adopted from Kazemian, Blanc et al.) discuss the potential theoretical and policy relevance of RCL. From a theoretical viewpoint, RCL reflects the age-crime distributions of active offenders. (Kazemian, Blanc et al.) discuss the predictive potential of measure of past criminal behaviour (i.e., age of onset, past number of offences, and the time since the last conviction) on future offending.

Currently official police records are the only source for estimating RCL, since the use of self-report, although possibly more accurate, are mostly not available at the sentencing stage. To investigate this (Kazemian, Blanc et al.) computed risk scores to assess the ability to predict the Official Residual Career Length (ORCL) and the Self-report Residual Career Length (SRCL) based on the four most influential variables (age at offence, conviction number, time since the last conviction, and age of onset). (Kazemian, Blanc et al.) were able to predict the ORCL and the SRCL better than chance. (Kazemian, Blanc et al.) suggest that predictions of RCL based on information available in official records may be more accurate when using samples of high-rate offenders. An important finding of (Kazemian, Blanc et al.) is that the distributions of SRCL and ORCL were often highly similar for high-rate offenders. (Kazemian, Blanc et al.) emphasize however that the high-rate offenders deviate greatly from the norm. They also note that even in the scenarios where the prediction is accurate, there is no guarantee that incapacitation of the offender in question will prevent the occurrence of the offence; especially being true in group crimes, where the offender may easily be replaced.

The main policy implications of RCL are related to sentencing and incapacitation decisions. This becomes clear when offenders are arrested and convicted, sentencers must decide whether the

offender should be confined and if so, the length of this confinement. One of the main purposes of confinement is to prevent future crimes from these individuals. Therefore, errors on this length of confinement are made when people have residual criminal careers that extend beyond their release from prison, or when they remain confined after the end of their criminal careers. Ideally, confinement should be applied to offenders during their years of criminal activity. A better estimation of the time remaining of a criminal career possibly leads to a more effective use of limited prison space, and avoid long prison terms for individuals that are unlikely to re-offend (Spelman 1994, adopted from Kazemian, Blanc et al.).

Although related, there is an important difference between estimates of the total career length and the residual criminal career length. The first might be interesting from a theoretical perspective, the latter, however, has significant more practical potential. As noted earlier, it might be useful for sentence determination, but it might also be useful for crime prediction. When offenders are correctly modelled as active offenders during their RCL, the final result will be more reliable.

3.3 Repeat victimisation

Crime and disorder problems are concentrated among a relatively few offenders, victims and places. Repeated studies have shown that a small number of victims reports a relative great amount of crimes (Chainey and Ratcliffe 2005). For street robbery, this has been noticed in several studies (Farrell 2006). The risk of a repeated event is the greatest just after the event (Polvi, Looman et al. 1991). For burglary, this risk decreases with time to obtain the background risk after a few months (Ratcliffe and McCullagh 1998). This brings us to the notion of Repeat Victimisation (RV, Miller 2005), an important principle in crime. This concept states that a target (e.g. people, vehicles or buildings) that have suffered a crime once have a disproportionally high probability to be a victim of the same type of crime again (Sparks, Genn et al. 1977; Hindelang, Gottfredson et al. 1978). Several studies show that this elevated risk is communicable to properties within an area around the victimised property. The elevated risk does not remain indefinitely but restores itself to the mean in the following period. The length of this period depends on the type of crime.

Farrell (2005) describes the current findings on RV. Farrell shows that 10 percent of the victims experience 50 percent of the crimes and presents 17 reasons why the prevention of RV is an attractive crime prevention strategy.

Brunsdon, Higgs et al. (2005) state that the same observation holds for the time dimension. Crime peaks at specific hours of the day, days or months. Some researchers have refined the notion of RV. Here we will list the most important:

- Virtual RV, this is a form of RV where types of targets are re victimised. For example, burglars that target the same type of house, because they know the layout, the risks and rewards may be similar to a previous target of the same type.
- Near RV. In the case of near RV the targets are chosen based on their location. This means that other targets that are near a previously victimised target share an elevated risk.
- Tactical RV is defined as RV where the offender picks the targets based on the tactics or skills of the offender. An example is a burglar that only picks locks, or always climbs through open windows, etc.

Eck, Chainey et al. (2005) splits Repeat Victimization into 4 different categories: repeat places, repeat streets, repeat areas and repeat victims. Each of these categories have different causes and call for different analysis.

- Places can be revictimised, or can be a location of reoccurring crime, because the behaviour regulation by the place managers fails. An example is a bar where the staff is instructed to minimise the chance of an assault, e.g. by limiting the number of drinks one can consume. This attracts people who desire a well regulated environment and repels people that do not (and are more likely to cause problems). Repeat places tend to be stable over time and a list of the most serious places can provide very useful police targets.
- Repeat streets can have a high degree of revictimisation for several reasons. It can be a common pathway of potential offenders and/or victims between their routine activities. Another reason is that some streets contain many places that are potential targets, like stores. A different reason is that some offences concentrate along streets or street segments. Prostituting and drug dealing are examples of such offences.
- Areas that are more frequently revictimised can be explained by several factors. One important aspect is the level of social control. Social networks are important in keeping up social control. The forming and stability of these networks in an area are affected by high levels of migration, poverty and racism. The willingness of the population in an area to prevent crime and keep up social norms also affects the crime levels of an area. A final possibility is the existence of concentrations of crime opportunities.
- Repeat victims are people that are victimised multiple times. Explanations for revictimisations might be found in the lifestyle and occupation of the victims. A prostitute or drug dealer for example has an increased chance of becoming a victim. Another explanation might be that the pathways between routine activities of a victim intersect with high risk (or repeat) places, streets or areas.

3.4 Geographic Profiling

In the previous sections prominent crime theories have been discussed. In this section we will discuss geographic profiling because this will be applicable to our simulated criminal agents in later stages. We need to identify the activity spaces of criminals geographically, and this is exactly what geographic profiling tries to do. Here is a definition for geographic profiling (Rossmo 2000):

Geographic profiling is a criminal investigative methodology that analyses the locations of a connected series of crimes to determine the most probable area of offender residence. It is applied in cases of serial murder, rape, arson, robbery and bombing, though it can be used in single crimes that involve multiple scenes or other significant geographic characteristics.

Geographic profiling can be used for two types of situations. The first situation is the one in which offenders for a series of coupled incidents are unknown, here geographic profiling is used to estimate the home location. The second situation is when the police have a criminal of which the home location and incidents are known. In this case the police can predict from the geographic profile where the offender can possibly strike again or if the offender has a connection with unsolved cases. There exists a range of strategies for the prediction of the home location of an offender. These are clearly described in (Levine 2004). Snook, Zito et al. (2005) have tested if the results of more complex geographic profiling techniques give better results for

more complex tasks[11] and tasks in general. They have judged the different strategies on complexity in terms of the number of mathematical operations that were necessary to complete the calculations necessary. Subsequently, Snook, Zito et al. (2005) compared the different techniques on accuracy. This comparison showed that the simple techniques were as good as the more complex tasks. We will therefore only describe a simple technique.

An article from a Dutch Police Journal Lopez (2005) also reported that the use of simple strategies are as accurate as more complex strategies, they argue that the circle technique[12] is a good practise for the police because it is simple and can be used without a computer. And with this technique the search area is not too big for the police. When an estimation of the area of which the next crime will take place is made the accuracy is good 80% for known offenders and 52% for unknown offenders.

Another relative simple strategy is by calculating the centroid whose coordinates are the mean of the x-coordinates and y-coordinates. The equation for deriving the coordinates is given in Equation 3.2, where x_i and y_i are the coordinates of crime locations and n is the total number of crime locations. This strategy is computational inexpensive, but can compete on accuracy with the more complex techniques (more information on these and other techniques can be found in Snook, Zito et al. (2005) and Levine (2004, chapter 10).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

Equation 3.2 Computation of the centroid coordinates

There are various ways measurements of accuracy for Geographic Profiling strategies. Two measures of accuracy commonly used are search cost (Canter, Coffey et al. 2000), or hit percentage (Rossmo 2000), and error distance (Levine and Associates 2000; Snook, Canter et al. 2002; Snook, Taylor et al. 2004). The search cost and hit percentage are equivalent strategies that are measured as the percentage of cells, in an overlaid grid, that need to be searched. The error distance measure was used for the evaluation of the strategies in Snook, Zito et al. (2005).

Stangeland (2005) repeat a critic from Canter (2003) saying that geographic profiling "presents at best, a way to assign priorities to time consuming police investigation, and recommends the investigative paths that are most likely to give results"'. According to Homant and Kennedy (1998) geographic profiling also has a qualitative component, this is based on the reconstruction and interpretation of the offenders' mental map. Factors such as the hunting style of the offender, the density of the potential victims, the location of major roads and highways, physical and psychological boundaries, and zoning and land use can be used to improve the geographic profile after the quantitative prediction has been made.

In section 5.2.1 we will describe how we will use geographic profiling in our model.

[11] Tasks are called to get more complex when the number of crimes increases.

[12] One draws a circle through the outer points; the centre of the circle is the estimated home location. More information is found in Snook (2005).

3.5 Offender profile

For every crime there exists a different kind of profile of the criminal, with different kind of preferences. For street robbery, the crime that we will use for our research, exists also a specific offender profile. Ferwerda, Jonkmans et al. (1998) have investigated street robbery in the region "Midden- en West-Brabant" in the Netherlands. This investigation is based on 249 researched street robbery incidents (of which 152 'real' street robberies and 97 bag robberies). Within street robbery they make the distinction between bag robbery and 'real' street robberies. They make this distinction because there are differences in the profile of both the offender and the victims the modus operandi and the value of the catch. More than half of all street robberies are committed in the weekends. The relative small city centre is the most unsafe location. Connections between the city centre and the outer districts have also a higher risk for being victimised by street robbery. Most bag robberies are between 18.00 and 01.00 hour. For 'real' street robbery this lies between 18.00 and 05.00. Between 22.00 and 23.00 is the highest risk for street robbery. So there seems to be a relation between street robbery and going out (to a bar).

Property	bag robber victim	real' street robbers victim
Gender and nationality	Dutch women	almost always Dutch men
Age	Average 51 years	Average is 25 years
Company	Mostly alone	Mostly alone
Travel by foot or by bike	Mostly by foot or by bike	Mostly by foot or by bike

Table 3.1 The properties of two types of victims involved in street robbery

Property	bag robber	real' street robbers
age	young < 20 years	average 21 years
number	alone or with companion	mostly with companion
vehicle	motorbike or scooter for quick transport	travel by foot
weapons	seldom	often verbally or 'with fists'
used violence	minimal	often unnecessary violence
gender	male	male

Table 3.2 The properties of two types of street robbers involved in street robbery

As implied earlier, there are two types of street robbers. The first are bag robbers and the second are 'real' street robbers. The properties of the individuals involved in street robbery are summed up in Table 3.1 and Table 3.2. The difference lies mostly in the use of violence and the way of operating. Bag robbery is seen as a relative 'clean' and quick crime. Bag robbers are not as well known as 'real' street robbers. This can be explained by the small chance of getting caught for bag robbery. Because of this Ferwerda, Jonkmans et al. (1998) continue to write about street robbers only including 'real' street robbers. From now on we will also neglect bag robbery and consider street robbery as being 'real' street robbery exclusively. Street robbers have on average 18 antecedents. Most of them come from poorer neighbourhoods. Street robbery is no start delict and it seems that street robbery is a crime of which the chance of getting caught is small, while the reward is quite high on average, therefore the change of repetition is high.

Now we have gained insight in the main crime theories and the profile of a street robber we will continue to discuss techniques that are used to predict crime.

3.6 Predictive crime mapping

In the previous sections we have presented important crime theories. In this section techniques other than the ABM technique that are used to predict crime are briefly discussed. The field of crime prediction and forecasting is still in its infancy but it gets increasing attention in recent years. First, we will discuss the hot spot mapping technique shortly. Subsequently, predictive hot spot techniques and other predictive techniques to predict crime patterns are discussed briefly. Finally, the evaluation of the prediction of these techniques is discussed.

3.6.1 Crime hot spot mapping

One of these techniques is the use of hot spot mapping as forecasting method. This is a common technique in the crime mapping field. A hot spot is an area of high crime density (Eck, Chainey et al. 2005). Hot spot mapping has the advantage of being flexible in time scale. One can use the technique to determine where crimes will most likely take place in the next 24 hours for targeting police patrols. For strategic crime prevention planning purposes the technique is used to predict the next 12 months (Chainey and Ratcliffe 2005). According to Chainey and Ratcliffe (2005) animation provides an opportunity to inspect spatio-temporal patterns. The current way that this animation is created is by creating a sequence of single snapshots of a hot spot map. Chainey and Ratcliffe (2005) make an important note however that a snapshot does not provide a detailed behaviour of the changing crime patterns, it only tells the analyst what has changed not how this transition has taken place. This finding provides us with an advantage of ABM that naturally can show spatio-temporal patterns.

3.6.2 Predictive crime mapping techniques

The predictive hot spot mapping technique can be divided into two categories: retrospective and prospective. Retrospective mapping uses crime data of the same month of the previous data or the data of a complete past year to forecast future crime events. If, for example, the data of 12 months is used in a retrospective method this means that the crimes of yesterday have the same weight as crime that occurred a year ago. This is conflicting with (recent) knowledge about crime. Evaluations of crime prediction techniques suggest that forecasting methods that use retrospective data, especially those using data for the same month of the previous year should not be used, because they are not accurate (Gorr and Olligschlaeger 2002; Groff and Vigne 2002; Chainey and Ratcliffe 2005). More sophisticated is the leading indicators method (Gorr and Olligschlaeger 2002). This technique uses past and current values of independent variables that are associated with a certain crime variable to predict the future value of the crime variable. Leading indicators are, for example, shots fired, calls for service, disorderly conduct offences etc. (Groff and Vigne 2002). According to Groff and Vigne (2002) the leading indicator method requires a fair amount of understanding of environmental criminology theory and multivariate spatial modelling. Bowers, Johnson et al. (2004) introduce a prospective mapping method that is based on research on repeat victimization (see section 3.3) and has some similarities with the leading indicator method, using each crime event as a leading indicator. It appears to be that crime is communicable, meaning that properties within 400 meters of burgled residents are at significant greater risk for being victimised for up to two months after the initial incident. Crimes in the past are therefore weighted to their communicable risk, the events that have happened more recently and those that happened close to other similar events get a greater weight. Bowers, Johnson et al. (2004) claim that their method is 30% more accurate than existing hot spot mapping techniques such as kernel density estimation and thematic mapping of boundary areas.

Additionally to predictive hot spot techniques there are also other types of predictive techniques. To start with the simplest technique, univariate methods use a previous value of one variable to predict the future. An example of a common data-based, non-model police method uses a month data of a year ago as the forecast of the same month this year. Gorr and Olligschlaeger (2002) call this method "naive lag 12" and show that this is the 'very worst forecast and comparison method' to assess police activities. Univariate methods, although straightforward, can be the least accurate if only simple methods are applied (Gorr and Olligschlaeger 2002; Gorr and Harries 2003). Point process modelling Liu and Brown (2003; 2004) is based upon the preferences shown by offenders in how they offend in past incidents (Groff and Vigne 2002). In this method multivariate models that attempt to explain behaviour are combined with components from kernel density estimation and a geostatistical interpolation technique called kriging. This technique has distinct advantages over other methods because it has a theoretical base (rational choice theory, see section 3.2) and it can point out explanatory variables. Another group of techniques are artificial neural networks (Olligschlaeger 1997). These techniques are trained on past crime data and iteratively learns the conditions that influence the occurrence of crime in time and space, to predict where and when future crime will occur. This technique holds promise, however, it cannot show which input (independent variables) are providing predictive power and it is atheoretical (Groff and Vigne 2002). Groff and Vigne (2002) write that the current state of knowledge seems to indicate that the straightforward exponential smoothing techniques are as effective as more advanced ones. However, they also note that their review of the current used methods is premature, because some methods are still in development. A more recent article with comparisons of methods such as that of Groff2002 has not been found. In the next section is a common way police departments anticipate on crime is discussed.

3.6.3 The use of prediction techniques by police departments

In previous text we have discussed current forecasting techniques briefly, here the common police method is discussed. A traditional way of the police to forecast crime is to use the data of the previous year as indication of what can be expected for the current year (Gorr and Harries 2003).

Currently, the way in which the Region Police of Amsterdam-Amstelland (RPAA) predicts crime can be divided into two parts. Firstly, there is a prediction for three months ahead with a capacity planning report. This report shows recent crime trends by crime numbers per police precinct of the last three months. The long term trend, i.e. what is normal for the predicted month, is shown by crime numbers of this month of the last five years. This report is used to determine how many persons are needed per district during the different shifts. Secondly, to determine the patrolling routes of police units within the precincts, hotspot maps of the last four weeks are compared against the four weeks before those weeks, and against the same period in the previous year. The comparisons are made by showing difference hotspot maps. The maps used here are not meant to show trends in numbers, but are meant to show the points of attention. Both these approaches are not called predictive by the police.

The current state of the art in crime forecasting at the RPAA is based on time series. The number of crimes per district per crime is forecasted by the autoregressive integrated moving average (ARIMA) technique. This technique gives the RPAA a forecast of six months. This technique is useful for capacity planning of manpower but does not help the planner in positioning manpower beneath the district level.

We have not found more information on police behaviour and the effect of enforcement on crime levels. In our model police behaviour is therefore simulated as if they get information about crime levels once a day.

3.6.4 Evaluation of crime predictions

In this section we will discuss the research of Bowers, Johnson et al. (2004) and how they evaluate their forecast. Bowers, Johnson et al. (2004) talk about prospective hot-spotting with an accent on residential burglary. Residential burglary has probably a different focus than street robbery. Incidents from 2 months earlier are still relevant to burglary whereas street robbery incidents might have a different scope. However, Bowers, Johnson et al. (2004) provide a good example to evaluate crime forecasting methods. Since crime forecasting is used specifically for the effective deployment of police manpower this technique requires a special evaluation. A perfect prediction of crime is impossible therefore one has to do concessions. So instead of saying, if it is not perfect it's wrong, one can use a less specific approximate and compare that with other methods. Thus instead of comparing a method with a perfect answer it should be compared with other methods, with the prediction of random walk method[13] as a minimal performance goal.

Bowers, Johnson et al. (2004) compare their prospective hotspotting with retrospective hotspotting and a traditional method. The discussed prospective hot spot technique makes use of recent data up to two months from the current data. Different from other methods, it is event-based rather than area-based. This is not arbitrary but based on evidence from literature (Johnson and Bowers 2004). Retrospective techniques use data of the same period but a year earlier. The traditional method, is based on the number of incidents in police beats. The techniques are first compared by visual inspection of the created hot spots. Second, a more objective approach is proposed based on the following criteria (quoted from Bowers, Johnson et al. 2004):

- Hit rate - the number of new crimes that are captured the predicted hot spots areas.
- Hot-spot area - the extent of the hot-spot areas. It is expressed both in terms of total area across all hot spots and average size of hot-spot area.
- Search efficiency rate - the number of crimes successfully predicted per kilometre squared. Using a standardised index allows different procedures and different hot spots to be meaningfully compared.
- Number of hot spots - account of the number of different hot spot areas produced by the technique.
- Area-to-perimeter ratios - These are calculated by simply dividing the area of the hot spot by its perimeter. The larger this is, the more compact the area is; in other words, the more area is covered in the shape per length of perimeter. This could be seen as a measure of efficiency of the hot spot in terms of how practical it is to cover the area.

With these criteria in mind, the prospective hot-spotting method is always better than the retrospective method, and according to the search efficiency rate also better than the traditional method. There is one exception, if focused on the Area-to-perimeter ratio, the traditional method outperforms the prospective hot-spotting method. The authors suggest that there might

[13] The random walk method uses crime numbers from the previous month or year as forecast for the future. An example of such a method is the "naïve lag 12" method (see section 3.6 in Gorr and Olligschlaeger 2002).

by a trade off between the search efficiency rate and the area-to-perimeter ratio. The authors also conclude that the Hit rate criteria is misleading since it does not constrain the size of hot spots, there could be one hot spot that covers the whole map resulting in a hit rate of 100%. Another conclusion, is that the simple prospective method gives better results in terms of search efficiency rate than, the more complex, retrospective methods.

In summary, the article of Bowers, Johnson et al. (2004) is very useful for our domain. It describes a useful method to evaluate forecasting methods that are based on different techniques and output different maps. When looking at the evaluation of ABM forecasts this article also implicitly suggests that a generic evaluation method of an ABM forecast is hard and instead should be domain specific to be usable in practise e.g. in our case a method that takes into account the practical value of a crime forecast for efficient policing (see Bowers, Johnson et al. 2004 for an example).

3.7 Summary and conclusions

In this chapter we have discussed several relevant theories. As already suggested by the previously discussed ABM's in the previous chapter, routine activity theory and the rational choice theory are the only well-founded theories we have found that could be useful for modelling crime on an individual level. This confirms the answer we found in the previous chapter on the research question: "What theories are useful for the modelling of crime in an ABM?". Furthermore, we found possible useful theories and techniques that answer this question and the question: "What techniques are useful for the modelling of crime in an ABM?". The theories we found and the belonging techniques are found in: the theory on Repeat Victimisation, Geographic Profiling and the theory on Residual Career Length:

- The theory on Repeat Victimisation can be used to estimate the opportunity surface. In section 2.4 the concept of opportunity and guardianship surface was presented (Gunderson and Brown 2000). The theory of Repeat Victimization (see section 3.3) gives us evidence to use our data to construct an opportunity surface for our model. The risk of a repeated event is the greatest just after the crime event (Polvi, Looman et al. 1991). This risk decreases with time to obtain the background risk (Ratcliffe and McCullagh 1998). In our model we will use this risk to define opportunity. By using crime frequencies for locations and recent incidents, we can obtain the background risk and recent crime trends. This risk surface is translated as a surface that indicates opportunity.
- Geographic profiling theory described in section 3.4 provides us with techniques to determine the approximate location of criminals based on data of previous crimes. These techniques can also be used the other way around, to approximate the activity space of a criminal. In our model presented in chapter 5 the centroid technique (Equation 3.2) will be used to estimate the activity spaces.
- The Residual Career Length is a theory that studies the remaining time a criminal will continue to commit offences. Kazemian, Blanc et al. have very recently worked on a method to predict the remaining activity of offenders. This method performed better than chance in estimating the career length. This method could be interesting in the future for estimating the life span of simulated criminals in an ABM.

We also found an additional answer to the question: "What are the possible uses of ABM for crime analysis?". According to Chainey and Ratcliffe (2005) animation provides an opportunity to inspect spatio-temporal patterns, this is something an ABM naturally shows.

In section 3.6.4 we discussed methods for the evaluation of crime forecasting methods in order to answer the research question: "How can we evaluate our ABM?". One important point we learnt was that the minimal performance goal of an crime forecast is that of the random walk or "the naïve lag 12" method. The other discussed criteria might be useful later. In our evaluation of our model we will first look at the correlation between the predicted crime and the observed crime and compare this correlation with the naïve lag 12 prediction and observed crime correlation.

We have seen proof that crime is in a large extent a rational act. However, this rationality should be regarded from the offender's point of view. Another important observation is that with each crime the risk of getting caught increases, and the criminal in question is aware of this. Although, criminals mostly behave according to known patterns, described in the theories discussed in this chapter, these patterns are influenced by individual factors and preferences that are hard to predict. Feeney (1986) also states that robbers know a lot about themselves and about robberies that no one else knows. These specific preferences can possibly be derived from crime incident data using a preference extraction method (see e.g. Gunderson and Brown 2000; Gunderson 2003) in section 2.4).

To summarise, we have gained insight in the behaviour of street robbers and criminal patterns in general. We found techniques and some general rules that can be applied in an ABM. We extracted the following requirements for our ABM:

- The use of the concepts routine activity theory: guardianship, potential offender and suitable target (see section 3.1).
- The use of the concept of delayed deterrence to illustrate the increasing fear of criminals (see section 3.2.3).
- The use of the idea that when criminals get older it is more likely that they quit crime (see section 3.2.3).
- The use of the idea that under the right circumstances ex-offenders tend to recommit crime (see 3.2.3).
- The use of the concept of repeat victimization (see 3.3).
- The use of a Geographic profiling technique to estimate the activity space of a criminal (see section 3.4).
- The use of the defined strategy of the police (see section 3.6.3).
- The use of the minimal requirement for crime prediction (see3.6.4).

As in the previous chapter only a subset of these requirements can be used in our model. The first requirement is a logical choice to implement because Groff (2006) already created a model based on these concepts, which we can borrow. The next three requirements, starting with delayed deterrence, are a bit abstract and therefore hard to formalise in our model. For this reason and because our model will not run for a long period in the future, because this makes a forecast less reliable, we will neglect these three requirements. The theories of Repeat Victimisation and Geographic Profiling are used as described in the beginning of this section. In

section 5.2.3 the behaviour police enforcement model is described, this is roughly based on the strategy described in 3.6.3. In the next chapter the ABM methodology is discussed.

4 Agent-Based Modelling as methodology

In this chapter the Agent Based Modelling (ABM) methodology is explored for our purpose. An overview will be given on the field of simulation in general and ABM. We will discuss the used concepts and the properties of the ABM methodology. Furthermore, some critical notes on the methodology are discussed. Requirements for our ABM model will be extracted from this exploration.

4.1 Simulation

4.1.1 Simulation in the social sciences

Social scientists are concerned with the understanding and explanation of social phenomena. The prediction of social phenomena is viewed with scepticism, because of both the difficulty of making one and also the possibility that the forecast itself will influence the outcome (Gilbert and Troitzsch 2005). For example, in our case we could construct an ABM that predicts crime rates and is used to determine where to place police forces without incorporating this changed variable into the model. This model would ignore the influence of police forces on crime occurrences and thus the use of the forecast would change the context on which the model is based. Therefore we reason, if the goal is to forecast a certain variable to anticipate with a controlled variable this controlled variable should be incorporated as well. So in our context, when we predict crime to help police manpower scheduling, the adjusted scheduling should be included as a variable in the model as well, because it can mean that there will be a different forecast (Gorr and Harries 2003).

A researcher is interested in a certain phenomenon and the goal of an ABM is to create a target that is simpler than the target itself (Gilbert and Troitzsch 2005, originally from Zeigler 1985; Doran and Gilbert 1994). The hope is that conclusions are drawn from the model that can be applied to the target because the target and model are sufficiently similar (Gilbert and Troitzsch 2005). In social sciences targets are always dynamic entities which change over time and react to their environment, encapsulating both structure and behaviour. For this reason the model must also be dynamic. In ABM this model is represented as a computer program. Alternatives to ABM are given in Gilbert and Troitzsch (2005), such as microsimulation and system dynamics.

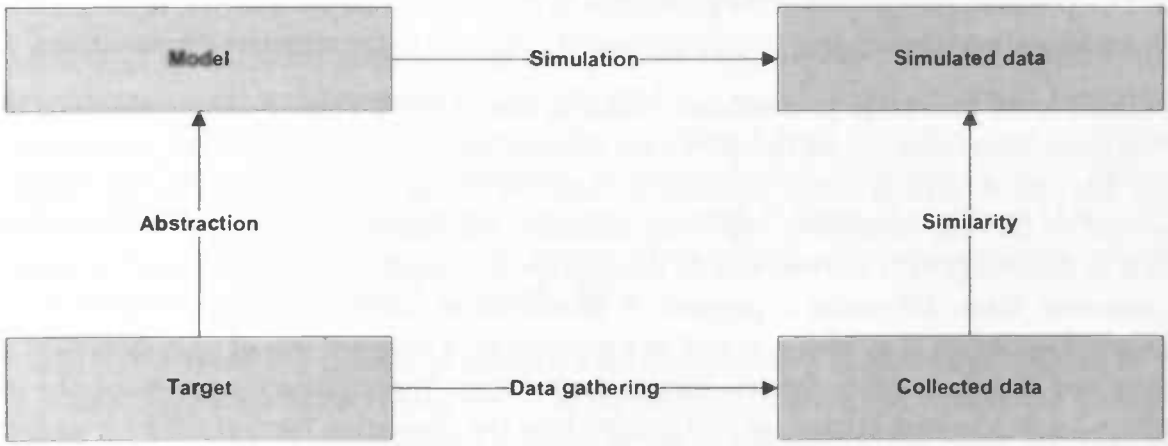


Figure 4.1 The logic of simulation as a method (adopted from Gilbert and Troitzsch 2005)

In Figure 4.1 the logic of a simulation as a method is shown (Gilbert and Troitzsch 2005). The researcher develops a model based on a target phenomenon. This model is used to produce data of the behaviour measured. This data can then be compared with the data collected to check whether the outcomes of the model are similar to the real data.

4.1.2 Stages of simulation-based research

Gilbert and Troitzsch (2005) describe the stages involved in simulation-based research. One starts by finding a question which answer is unknown. The next step is to define the target for modelling. Observations are then required to provide parameters and initial conditions.

Then assumptions have to be made and the model has to be designed. For this it is recommended to use a package instead of starting from scratch, because of the many issues that take time when writing a program. For example, writing code to show plots and charts is a difficult task and therefore time consuming. Most packages, however, provide facilities to display output variables. Additionally it is recommended to use a version control system so that it is easier to work with different versions of models. The last steps are verification, validation and sensitivity analysis. Verification will be the debugging of the model using unit tests. The validation of the model is to check whether the simulation is a good model of the target. Validity can be ascertained by comparing the output of the simulation with data collected from the target. Finally, if the model appears to be valid, a sensitivity analysis will be done to test how sensitive the model is for its initial conditions and parameter values.

There are several caveats to the above described approach of validation. Firstly, it is likely that both the model and the target processes are stochastic thus preventing exact correspondences. Whether the differences between the distributions are large enough to cast doubt depends on the expected distribution of the output measures. As in our case, unfortunately, this distribution is usually unknown and not easy to estimate. Secondly, most simulations are path-dependent, as a result of which outcomes can be very sensitive to the precise value of some assumptions. Gilbert and Troitzsch (2005) explain also that as results obtained from the simulation can match those from the target, certain aspects of the target, however, cannot be reproduced by the model. Gilbert and Troitzsch (2005) mention an example of a model that predicted the growth of the world population. In this model the predictions for the next 50 years in the future looked plausible. However, retrodiction of the population to the situation 20 years in the past, using the same model and parameters, was completely wrong compared to the actual world population.

Publication is the final stage in simulation research. Axelrod1997 poses several difficulties in writing about simulation. These categories are divided into four categories: ambiguities in the model description, gaps in model description, clear but wrong model descriptions and finally, difference in the way computers represent numbers. For example, in one of the replicated models by Axelrod (1997) 9/3 was exactly equal to 2 + 1; in another implementation it was not. To overcome these difficulties a protocol is developed to improve the ABM methodology (Richiardi, Leombruni et al. 2006), as well as a protocol for a standard way of describing ABM's (Grimm and Railsback 2005; Grimm, Berger et al. 2006). These protocols are discussed in section 4.2.4. In the next section we will discuss how the simulation method differs from the traditional research methods, deduction and induction.

4.1.3 Simulation as a third way of doing science

Axelrod (2005)'s goal is to advance simulation research as a mature field. His article is summarised here, because it gives a good overview of the field of simulation. When not cited otherwise the information from this section comes from Axelrod (2005).

Axelrod (2005) mentions many terms that are used to describe the field of simulation in the social sciences. Examples are artificial society, complex system, agent-based model, multi-agent model, individual-based model, bottom-up model, adaptive system, computational model. There are no strict distinctions between these terms. We will continue to use the term ABM as we will argue in section 4.2.1.

Particular strengths of simulation as research method is that publications come from virtually all of the social sciences (anthropology, business, economics, human evolution, environmental planning, law, information, organization theory, political science, public policy) given publications with the term 'simulation' in it). A Weakness of this method is that it has little identity as a field in its own right (given the fact that these publication are diverged of different journals). The definition for simulation used by Axelrod (2005) comes from (Bratley, Fox et al. 1987):

Simulation means driving a model of a system with suitable inputs and observing the corresponding outputs.

For the purpose of understanding, the assumptions underlying the agent-based model should be simple. The complexity should be in the simulated results. For the purpose of prediction and training the assumptions need to be quite complicated because accuracy is important and not simplicity.

Simulation starts with a set of explicit assumptions, like deduction, but it does not prove theorems. It generates data that can be analysed as with induction. However, the data comes from a specified set of rules rather than from real observations. Induction can be used to find patterns, deduction can be used to find consequences of assumptions and simulation modelling can be used as a sort of mental experiment. Therefore Axelrod (2005) calls simulation a third way of doing science.

Advantages of simulation are that it can help to think about concepts as mind experiments. An agent based model simulation can be implemented in a so-called computational laboratory to vary agents and the landscape systematically, providing a level of control that is difficult to achieve in traditional social science methods (Epstein and Axtell 1996; Dibble 2006). Deduction is possible when the assumption of rational choice is used. It allows the analysis of adaptive as well as rational agents. New data can be generated on request and does not contain omissions as is normal in empirical research (Axelrod 1997).

Analysing the result of a simulation model is hard and according to Axelrod (2005) can be done in at least three ways by describing details of the history of a given run. At first, the history of a model can be told in a chronological order in terms of key events that occur. For example, a simulation of the labour market can be described as the sequence of key events such as strikes.

The history can also be told from an individual actor's point of view. Another way is to describe the history from a global point of view. This can be done, for example, in a model where agents

have a certain amount of wealth. The distribution of wealth over time can be described to analyse the extend of inequality among the agents. The latter is the best way for seeing large-scale patterns, however, detailed histories are needed to determine the explanation for these large patterns.

The detailed history can be of a single (typical) run. Statistical analysis of a whole set of runs is necessary to determine whether the inferences being drawn from the illustrative history are really well-founded. This should be done by changing the random seed. The effect of changing the parameters systematically is also a way to analyse the model. A model should be built incrementally, i.e. one should start with a model with agents having random behaviour and slowly making the behaviour more complex. The first (random) version can be compared with the more complex versions to see if the results are caused by the new complexity. Typically if the changes are quantitative the statistical method should be regression, if qualitative this should be analysis of variance. Two questions should then be raised. First, are the differences statistically significant (not likely being caused by chance)? Second, are the differences substantially significant (large enough in magnitude to be important)?

The final step in research is sharing results. This is mostly done by publications. Unfortunately, simulation results are typically quite sensitive to the details of the model. The analysis of the results often includes some narrative description of histories of one or more runs, and such a narrative description often takes a good deal of space. Since simulation results often address an interdisciplinary audience one has to take into account that not everyone understands the jargon of simulation. Therefore and because computer simulations are still very new it is required to explain the power and the limitations of the methodology. Some researchers refer to an external source to give a complete description of their model.

One of the things that has almost never been done so far, but is important nevertheless, is the replication of published models. Replication is important for science, because it is a precaution against cumulative errors and gives more faith in the claimed results. Also robustness of inferences from models can be tested with replication. For example, Axtell, Axelrod et al. (1996) replicated eight models and found many replication problems. Galan and Izquierdo (2005) replicated a model of Axelrod (1986) and found that the results of the original model were not as reliable as one would hope. The described methodology for ABM in section 4.2.4 is, among other things, meant to simplify replication. In the next section we will focus on ABM in particular.

4.2 Agent-Based Modelling

In this section we will describe the properties of ABM, when and how it should be used. The essential idea of ABM is, that phenomena, even very complex ones, are best modelled as autonomous agents that are relatively simple and are equipped with simple rules for interaction (Samuelson and Macal 2006). The range of applications is wide, from the modelling of households to simulate the impact of the growing rural population on the forests and panda habitat (An, Linderman et al. 2005) and from the behaviour of customers (Baxter, Collings et al. 2003), to the modelling of the space marketplace (Charania and DePasquale 2005) to the simulation of the impact of a biological attack on the populations behaviour (Carley, Altman et al. 2006). In the following some definitions for ABM will be given, subsequently we will continue with practical notes and we will end with a methodology description.

4.2.1 Definitions

As with simulation many names are used for Agent-Based Modelling. ABM is widely used just as ABS (Agent-Based Systems), and IBM (Individual-Based Modelling). Others (see for example Macal and North 2005; Samuelson and Macal 2006) use Agent-Based Modelling and Simulation (ABMS). Interesting here is that even though Samuelson and Macal (2006) explicitly state they will use the term ABMS to prevent confusion with the use of agents in computer science, in their articles they still use the term Agent-Based Modelling now and then. Because even the researchers who claim to be consistent are not consistent we will continue to use the term Agent-Based Modelling or the acronym ABM. Researchers such as Axelrod (2005) and Richiardi, Leombruni et al. (2006) mention the importance of one term for the future of the field of ABM. We agree with this statement, but it is not enough just to mention that this must be done. Prominent researchers in the field of ABM should come to a consensus on the term and argue why this specific term has to be used. We have not found such a discussion and we will therefore use the most common term found in literature.

There seems to be a common agreement, albeit with different names, on what ABM is. However, for the term "agent" different concepts are used. For the sake of clarity we will give the definition of the term "agent" as will be used throughout this thesis and we will give some examples.

A clear definition of an agent, specific to computer science, is given by Wooldridge (2002):

*An **agent** is a computer system that is situated in some environment and that is capable of **autonomous action** in this environment in order to meet its design objectives.*

Implicitly in this definition is that the agent has some design objective. In contrast to the definition of Wooldridge (2002), in ABM an agent does not have to meet a specific design objective in the sense that it has certain goals to accomplish. In ABM an agent acts upon its (artificial) environment given the behavioural rules it embodies. It is not important if a certain problem will be solved by the agents, it is the how that is important. In simple ABM models this idea of how should be translated as how interaction between simple agents emerges to complex macro behaviour. The definition of ABM in Bonabeau (2002) also contains a definition of agents:

In agent-based modelling (ABM), a system is modelled as a collection of autonomous decision-making entities called agents. Each agent individually assesses its situation and makes decisions on the basis of a set of rules. Agents may execute various behaviours appropriate for the system they represent.

We will use this definition because it is flexible enough for our purpose. However, it does not say anything about the characteristics an agent can have. Macal2005 give an overview of the possible characteristics an agent can have:

- An agent is identifiable, a discrete individual with a set of characteristics and rules governing its behaviour and decision-making capability.
- An agent is situated, living in an environment within which it interacts with other agents.
- An agent is goal-directed, having goals to achieve (not necessarily objectives to maximise) with respect to its behaviours.
- An agent is autonomous and self-directed.

- An agent is flexible, and has the ability to learn and adapt its behaviours over time based on experience.

There are some characteristics that are universal in ABM, such as situatedness and autonomy, others are a choice of the researcher. Macal and North (2005) call the capability of agents to make independent decisions a fundamental property, requiring agents to be active rather than passive. Apart from this property, other properties of agents can differ strongly from one model to another, or even within models, making the agents diverse, heterogeneous and dynamic in their attributes and behavioural rules. Behavioural rules can vary in their sophistication, the amount of information is used in the agent decisions, the agent's internal model of the environment, the memory of past events and the use in its decisions (Macal and North 2005). As mentioned by Axelrod (2005) the decision for these properties has to depend on the purpose of the model. This diversity is one of the reasons what makes ABM interesting (Macal and North 2005).

4.2.2 ABM and GIS

Agent-based models for simulation can be grouped into different categories. There are models that simulate networks e.g. networks of friends, this allows an agent's neighbourhood to be defined more generally and flexibly (Macal and North 2005), there are models that simulate certain phenomena without specific spatial properties using a grid and there are models that have an important spatial aspect. Examples of the latter are agent-based models that have a Geographic Information Systems (GIS) surface. This surface is normally based on a real environment and may include buildings, rivers etc. This information can be derived from companies such as ESRI (http://www.esri.com). The GIS topology is the most relevant to us, because we are interested in simulating crime for a real environment.

GIS is are systems to capture, manipulate and analyse spatial data. Geographic information science is the science that studies and has founded the geographic concepts, applications and systems. GIS has been used extensively in studying human-environment interactions. GIS alone however, is not able to capture functions or dynamic processes practically (An, Linderman et al. 2005, originally from Peuquet 1999).

While at the same time ABM has focussed on sophisticated representations of time and behaviour at the expense of sophisticated representations of space and spatial relationships (Brown, Riolo et al. 2005), GIS has been focussing on the representation of the spatial dimension, at the expense of the temporal dimension (Peuquet 1999). From a GIS point of view, ABM offers the possibility to model dynamic processes. From an ABM point of view, advantage of the use of GIS data in combination with ABM is that it makes models more realistic and makes them therefore more plausible, because the environment is better represented in the model. This idea is supported by Torrens (2003) who has criticised ABM models that tried to model urban phenomena for the lack of spatiality. Other researchers, that discuss the integration of agent-based models and GIS, conclude also that ABM should be used in combination with GIS for a better model of reality (Brown, Riolo et al. 2005).

Gimblett (2002) discusses several conceptual and technical questions that have been raised after attempts to integrate ABM and GIS techniques. Issues generally fall into questions of ontology and process, i.e. how are entities and processes represented. Bian (2003) concluded that the environment within an individual-based model can be represented either as object-

based, thus maintaining object orientation in both the data and the model, or field-based, wherein agents interact with a discretized environment of attributes. The cellular automata approach, where the individual organisms as well as the environment are represented in the same grid field, is popular in ecological modelling but has significant limitations. The grid forces arbitrary movement directions and distances what may cause unrealistic simulation results. Brown, Riolo et al. (2005) give examples of several implementations where ABM is coupled with GIS. They also discuss the different types of possible software implementations for the coupling between GIS and ABM which are discussed in the next section.

4.2.3 Modelling platforms

There are several simulation platforms for ABM that can be used, e.g.: Swarm, NetLogo, StarLogo, MASON, Ascape and Repast[14]. There are researchers (Tobias and Hofmann 2004; Macal and North 2005; Railsback, Lytinen et al. 2006) that have evaluated some of these tools for the use of ABM. RePast (North, Collier et al. 2006) seems to be the most promising, with possibilities to integrate GIS as well as implementations in different programming languages; Java (Repast J, Repast S), .NET (Repast .Net) and Python (Repast PY), each having its own (dis-)advantages.

Railsback, Lytinen et al. (2006) evaluated five modelling platforms; MASON, Repast, NetLogo, Java Swarm and Objective C Swarm. In their conclusions they strongly recommend NetLogo for simple, grid-based models. They also recommend NetLogo for fast prototyping that may later be implemented in other low-level platforms (such as RePast): this prototyping can be a quick and thorough way to explore design decisions. However, they mention that experienced programmers can feel uncomfortable with the simplified programming environment of NetLogo. MASON is a good choice for experienced programmers working on computationally extensive models. Java Swarm was meant to provide Swarm for Java users, however now better alternatives exist in Java. Objective-C Swarm was the first ABM framework, is stable and still has some advantages such as a clever design, well-organised structure and a fairly complete set of tools. Repast has been marked as the most complete Java platform, with additional functionality such as geographical and network functions. A general drawback of all platforms is, that the basic documentation is largely incomplete. Because Repast is generally regarded as the first choice for ABM development, Railsback, Lytinen et al. (2006) pose some critical notes on how Repast could be improved.

What all the papers discussed earlier have in common is that Repast is indicated as the leading ABM platform. Given the functionality, the relatively large user base, helpful mailing list and many examples in literature Repast is currently the most complete platform. Thus only one choice remains i.e. the choice for Repast J, Repast S Repast PY or Repast .Net. Recently, the new version of Repast, Repast Symphony (see Figure 4.2) has been distributed (North, M.J et al. 2005; North, Howe et al. 2005). This platform promises to be easier to use with additional functions to Repast J. However, at the moment not all the functionality of Repast J has been implemented, for example the important GIS functionality. Additionally, since this is an alpha version bugs are more likely to appear and there are only a few implemented examples available. Furthermore, the documentation is minimal compared to the documentation of the other Repast versions. Repast S is therefore not a good option yet.

[14] For a complete up-to-date overview of ABM platforms should be looked at SwarmWiki2006 or http://www.econ.iastate.edu/tesfatsi/acecode.htm\#ACE

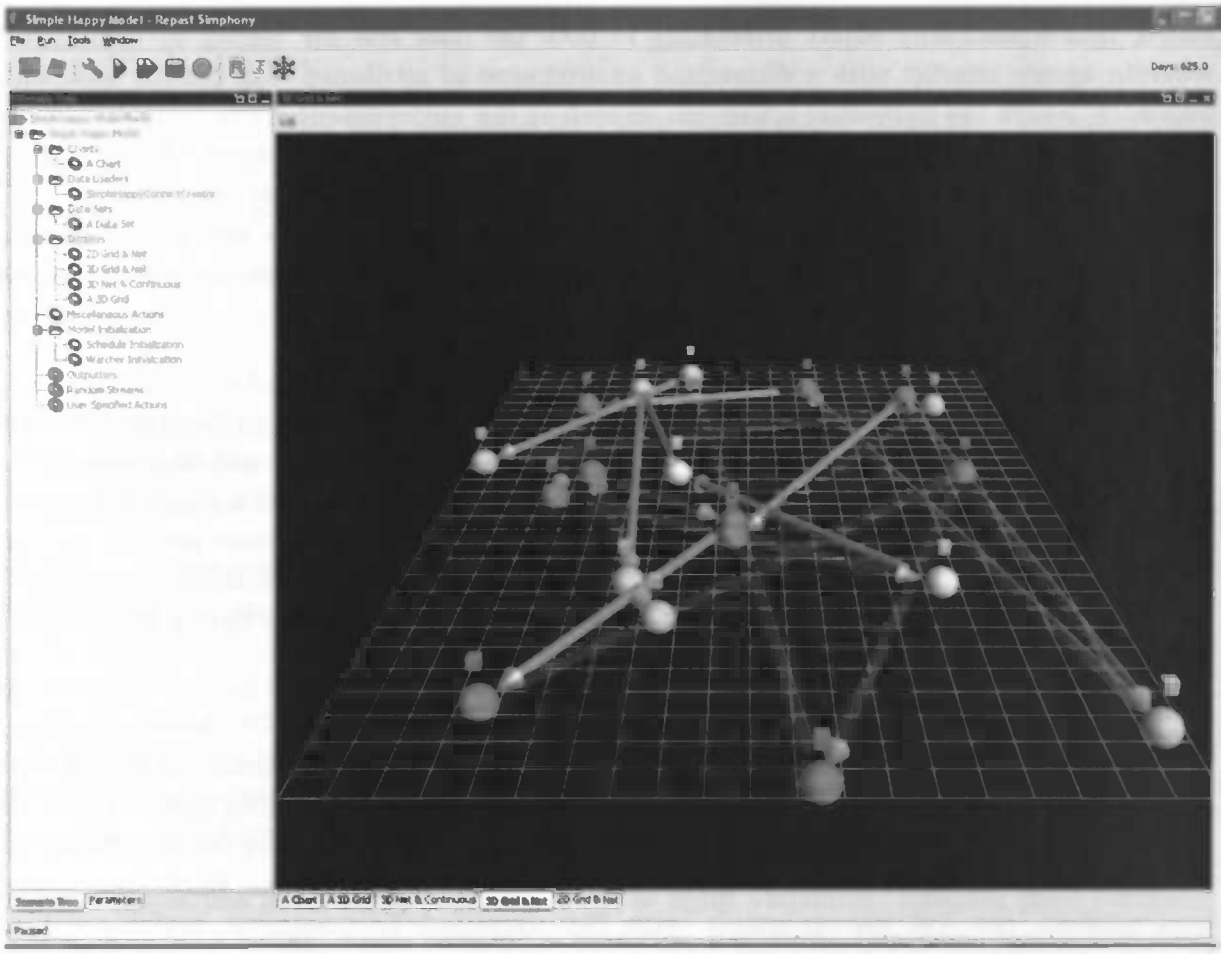


Figure 4.2 The Repast Symphony 'Simple Happy Model' displayed in a 3D grid and network display

(North, Collier et al. (2006) have discussed their experiences implementing models in the other three versions. In this discussion it turns out that Repast .Net has some advantages over Repast J due to properties of the .Net language, however some of these disadvantages of Java have been overcome with the introduction of Java 5 as the authors note. Likewise, disadvantages of the IDE of .NET are also expected to be solved with the introduction of the new version of Microsoft Visual Studio. One of the biggest advantages of the use of Repast J is the many possible extensions through a wealth of Third Party Java Libraries, e.g. Repast J has stronger GIS support thanks to the integration with the OpenMap library.

The comparison between Repast J and Repast PY is easier because Repast PY is based on Repast J and can therefore use the same libraries. The advantage of Repast PY is the point-and-click interface that is meant for novices and meant to provide fast prototyping. However, Railsback, Lytinen et al. (2006) found that Repast J in combination with the Eclipse IDE was more productive than Repast PY, even for beginners. The conclusion is, that Repast J seems to be the better choice for our research. Nevertheless, future research should consider using Repast Symphony.

Now that we have considered many platforms and their reviews it finally depends on the personal preferences for the choice of the platform. However, according to the general opinion it is better to use a common platform than to start from scratch (Axelrod 2005; Gilbert and Troitzsch 2005).

4.2.4 Methodology

In this section the work of Richiardi, Leombruni et al. (2006) is summarized. Richiardi, Leombruni et al. propose a common protocol for ABM in their work because computer-simulated models mostly lack a reference to an accepted methodological standard. ABM does not have a protocol such as the well-established protocol used in traditional modelling practise in the social sciences. This is one of the main reasons for scepticism among mainstream social scientists and economists, resulting in a low acceptance of papers with agent-based methodology in prominent journals (Leombruni and Richiardi 2005; Richiardi, Leombruni et al. 2006). The criticism of Richiardi, Leombruni et al. is that many articles ignore the basics of experimental design. The comparison between artificial and real data is often naive and parameters' values are chosen without a proper discussion. It is often not possible to understand the details of the implementation of an agent-based simulation, making replication a difficult or impossible task. To overcome these issues they discuss common pitfalls in agent-based modelling research.

Link with literature Each article should make references to the theoretical background of the phenomenon that is investigated. If a model is being innovated one should make references to that model. Additionally, references should be made to relevant non-simulation literature since this is what the audience that are not acquainted with ABM is used to.

Structure Basic features that characterise a simulation model are often not explicitly discussed. This includes the features: treatment of time (discrete or continuous), treatment of fate (stochastic or deterministic), the representation of space (topology), the population evolution (birth and death processes), the treatment of heterogeneity (how and what are the variables that differ across individuals), the interaction structure (localised or non-localised), the coordination structure (centralised or decentralised), the type of individual behaviour (optimising, satisfying, etc.).

Analysis After the specification of the model we come to the analysis. Since ABM simulations can generate enormous amounts of data there are many statistics to look at. In many cases it is not meaningful to do a full exploration of all the variables. Richiardi, Leombruni et al. (2006) give the example when parts of the model (e.g. the demand side for firms' output in a model of labour participation) are only sketched. Only a subset of the defined variables of the model should then be investigated. This partial exploration should be clearly stated and motivated.

The analysis of a model can be performed in equilibrium, out-of-equilibrium, or both. This should be clearly stated. There can be an equilibrium at a micro-level, which is defined as a state where individuals' strategies are constant. Or there can be an equilibrium at a macro-level as a state where some (aggregate) statistics of the system are stationary. In ABM usually the second definition of equilibrium is used since in ABM agents individual behaviour is relatively simple.

There are two ways of investigating a model: global investigation and local investigation. In global investigation it is important how the model behaves in broad regions of the parameters' space, i.e. for general values of the initial conditions and the parameters. This is generally the case when the model has been built with a theoretical perspective: the relationship between in- and output has to be understood without reference to the real data. In local investigation interest lies only in restricted regions of the parameters' space. Generally this is the case when the model is built with an empirical goal: one wants to replicate some empirical phenomenon of

interest and thus one only wants to explore the dynamics of the model around the estimated values of the parameters.

Parameter estimation can be preliminary to a local investigation. Richiardi, Leombruni et al. (2006) do not make a distinction between estimation and validation of a model. They refer to the estimation process as follows:

Estimation is the process of choosing the values of the parameters that maximise the accordance of the model's behaviour (somehow measured) with the real world system.

Not all parameters have to be treated the same. Some have natural real counterparts for which the value is known. Simulating a function in ABM that is not known, implies that estimating unknown parameters directly will not work. Structural estimation is, however, still possible by simulation-based techniques (Richiardi, Leombruni et al.).

Richiardi, Leombruni et al. (2006) define sensitivity analysis as a collection of tools and methods used for investigating how sensitive the output values of some models are to changes in the input values (see Brown and Robinson 2006 for an example). They use the following definition for a good model:

A "good" simulation model (or a "significant" result) is believed to occur when the output values of interest remain within an interval (which has to be defined), despite "significant" changes in the input value (which also have to be defined) [emphasis from the cited author].

Sensitivity analysis include the following kinds of input variability: Random Seed Variation, Noise Type and Noise Level Variation, Parameter Variation, Temporal Model Variation, Variation in the level of data aggregation, Variation in the decision process and capabilities of the agents, variation of sample size (see Richiardi, Leombruni et al. 2006 for more details).

(Richiardi, Leombruni et al. 2006) state that even a wrong model can be estimated, therefore it has to be validated as well. They use the definition of van Dierendonck (1975):

The term 'validatity can be formally defined as the degree of homomorphism between one system and a second system that it purportedly represents'.[15]

To asses overall validity other forms of validity have to be considered: theory validity, model validity, program validity, operational validity, empirical validity, (see Richiardi, Leombruni et al. 2006 for more details).

The degree of replicability of models is determined by several aspects of the simulation model, examples are programming language, tools, representation formalisms, development methodologies. For example, an open source license of the tools used is an important requirement. Furthermore, Richiardi, Leombruni et al. (2006) recommend the use of UML. [16]

---

[15] Homomorphism is used instead of isomorphism because the goal is to represent one system by a less complex system. Equality in complexity would mean that the systems are isomorphic (Richiardi, Leombruni, et al. 2006) .

[16] UML stands for Unified Modelling Language and is a general-purpose modelling language that includes a graphical notation used to create an abstract model of a system, referred to as a UML model (from http://en.wikipedia.org/wiki/Unified_Modeling_Language).

For describing individual- and agent-based models a new protocol has been developed, the ODD protocol (Grimm and Railsback 2005; Grimm, Berger et al. 2006). [17] This protocol consists of seven elements. The first three elements provide an overview, the fourth element explains general concepts underlying the model's design, and the remaining three elements provide details. The developers of this protocol believe standardization of the description will aid in the replicability and understanding of ABM's.

4.3 Summary

In this chapter we have discussed the simulation methodology and focussed on ABM. Several pitfalls have been presented by other researchers. We have chosen the Repast J as ABM platform as a framework for our model (see section 4.1.2 and 4.2.3). The following requirements of an ABM have been obtained:

- Choose the values of parameters with a proper discussion (see section 4.2.4).
- Use of a random seed to repeat random event if necessary (see section 4.1.3).
- Use retrodiction to forecast known data to test the model (see section 4.1.2).
- Use the ODD protocol to describe the model in a standard and clear way (see section 4.2.4).

These requirements will be fulfilled except for the last one, the use of the ODD protocol. We tried to use this protocol to describe our model but we got stuck with it. The protocol is formulated by ecologist who usually do not include human agents, the used concepts therefore do not match very well. Another fact is that there are currently little examples, especially for the modelling of human behaviour. The ODD protocol might be useful in the future when it has evolved and more examples are available for descriptions of models with human agents.

For the evaluation of the ABM we have obtained the following requirements:

- When analysing the results, use regression for quantitative differences in results and use analysis of variance if the differences are qualitative.
- When a subset of the defined variables of the model is investigated this should be clearly stated and motivated (see section 4.2.4).
- Usage of sensitivity analysis, including: random seed variation, noise type and noise level variation, parameter variation, temporal model variation, variation of sample size (see section 4.2.4).
- Discuss validity of the model, including theory validity, model validity, program validity, operation validity, empirical validity (see section 4.2.4).

We did not meet all these requirement when we analysed the model due to time constraints. However, we think we did the most important requirements. A sensitivity analysis was conducted and the correlation between the observed crimes and the predicted crimes was analysed. The theoretical validity of the model is discussed by presenting the underlying concepts of the components of the model. In the next chapter our model is presented.

[17] ODD stands for Overview, Design concepts, and Details the three building blocks of the protocol.

5 The model

In the previous chapters we have obtained requirements for the ABM model that is meant to illustrate the central claim of this thesis: "an individual modelling perspective and the Agent-Based Modelling methodology are fruitful for crime prediction". In this section we will describe the model that should defend this claim, it simulates the criminal behaviour of a group of criminals. The connection with the earlier described requirements will be mentioned when the concepts of the model are described. The criminals in the model are based on real offenders recorded in the police database of Midden-West Brabant, the Netherlands. The environment is a realistic map of the region of Tilburg. The possible crime locations in this environment are based on the zip codes in this area. The initial value of the state variables of the locations and criminals are based on criminal data from the police database. Furthermore, there are police agents that are introduced to illustrate the cat and mouse game that takes place between police and criminals. These police agents are not based on real police officers due to the lack of availability of data on police enforcement. In the following sections we will describe the model concepts without going into detail about the model's parameter settings. These will be discussed in the experimental setup in which we test the model on different parameter settings and scenario's. Finally, this chapter will end with a discussion on the model choices, the outcomes and the limitations.

5.1 Model environment - the region Tilburg

The crime events unfold on a digital map of the region of Tilburg, a city in the Netherlands. A screenshot of two maps are given in Figure 5.1.

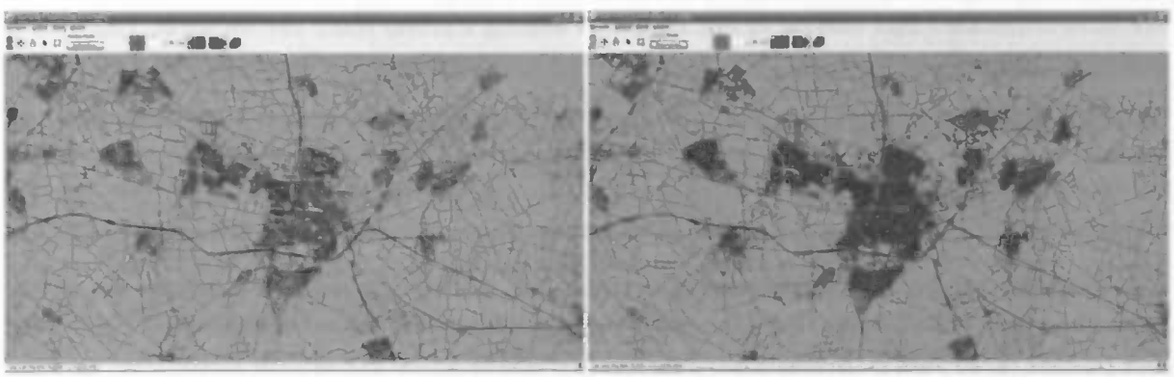


Figure 5.1 The model environment. The left image (A) shows the streets in the region of Tilburg. The right image (B) shows the same but with the Zipcodes projected as small yellow squares.

These maps represent the region of Tilburg, a region that is about 10 by 10 kilometres. [18] In the left map only streets are shown. Streets are only included to give context to the environment and do not influence the behaviour of the model. Since this model's use would be for crime analysis the idea is that the model's results are easier to interpret when the environment can be recognized from e.g. the streets.

Locations are based on the zip codes that are present in this region. We will call these locations Zipcodes from now on. These Zipcodes are represented by the small squares in the right image

[18] To be precise, the region we have taken into account has the following extents: latitude minimum, longitude minimum 4.90325, 51.5051; latitude maximum, longitude maximum 51.6147, 5.23167 (the latitude/longitude coordinate projection is based on the WGS84 system which is used by OpenMap). These extents are derived from the coordinates of the location of incidents recorded in the region Tilburg.

of Figure 5.1. Note that only Zipcodes are shown in a part of the right image. This is because we focussed on the area of Tilburg. The Zipcodes in the area around Tilburg are ignored, and therefore not shown, because this reduces the computational complexity. There are a total of 6447 Zipcodes present in the selected region. Before we discuss the agents of our model in the next section we will continue to explain more about the properties of the Zipcodes.

5.1.1 Zipcodes

Zipcodes, the objects that represent locations in the model, have a value for the opportunity of crime and a guardianship value. This is based on the idea of opportunity and guardianship surfaces (Gunderson and Brown 2000). The initial opportunity values of the Zipcodes are based on the crime frequencies of the previous years retrieved from the crime data made available by the police of Midden-West Brabant. We assume here that the crime frequency of four previous years[19] can be used to obtain the normal opportunity for crime on each location. In Equation 5.1 is shown how the normal opportunity is calculated. The number of crimes at a zip code is divided by the total number of crimes in the data. Some Zipcodes have an opportunity value of zero for crime because there has been no crime in the past on that location.

$$normal\ opportunity_x = \frac{number\ \ of\ crimes_x}{total\ \ number\ \ of\ crimes}$$

Equation 5.1 The calculation of the normal opportunity for Zipcode x

Additional to the normal opportunity value we also use the data of the last two months to include recent crime trends. This is based on research on repeat victimization. Bowers, Johnson et al. (2004) argue that crime until two months earlier indicate an increased chance on crime. We have translated this finding in our model as adding a crime trace per incident. This rule is described below[20]:

$$Opportunity_{x,t_0} = normal\ opportunity_x\ +\ (crimeTrace * number\ recent\ crime\ events)$$

Rule 5.1 Opportunity initialization rule: the opportunity at Zipcode x at t_0 (initialization)

Guardianship values are initially zero, because we did not have data available on police presence. After initialization, the opportunity and guardianship values are updated every step. The precise order will be discussed later. Here we will describe the update steps. First we have the opportunity increase rule for when a crime occurs:

<u>Opportunity increase rule:</u> add a crime trace to the opportunity value after a crime event.[21]

Rule 5.2 Opportunity increase rule

[19] We used crime data of the previous four years, from 2001-2004. The reason for this is that we only have data of January 2001 to June 2006. We used the data on the year 2005 to test the model results. We ignored the data of 2006 because it was only six months of data. Other options, e.g. testing the model results on the data of 2004, would mean that we had less previous crime data and would therefore be less reliable for the estimation of the normal crime frequencies for the Zipcodes.

[20] The crimeTrace is a model parameter and is set according to a sensitivity analysis discussed later.

[21] The rule can be stated formally. Call the current opportunity value o^t. Then the new opportunity value o^{t+1} is given by

$o^t = o^t + crimeTrace$.

'=' should be read as becomes.

The opportunity increase rule has the same theoretical grounds as the previous opportunity initialization rule; previous crimes indicate an increased chance on crime on that specific location. This theory also says something about the decay of the chance on crime, providing us with the opportunity decay rule:

> *Opportunity decay rule: at each Zipcode, opportunity decays with a factor alpha.*[22]

Rule 5.3 Opportunity decay rule

This rule translates a part of the theory of repeat victimization into our model. When a crime occurs the opportunity increases, as we have seen in rule number Rule 5.2. This rule shows the process after a crime event. Just after a crime the opportunity is the highest, then it slowly decays every step until a new crime is committed..

Like the opportunity values, guardianship values also change during a model run. Every time a Police agent visits a Zipcode, a guardian trace is left (hence Police agents are considered guardians). We have not found evidence for such a mechanism or other effects of police presence. However, since we wanted to create a model that accounts for the effect of police presence on crime we had to think of a plausible rule. We thought it is reasonable to assume that a Police agent leaves some trace of his presence (just like a criminal when he or she commits a crime as in Rule 5.2), therefore we have defined Rule 5.4. This rule states that the guardianship value at a Zipcode will increase when a guardian is present at that Zipcode.

> Guardianship increase rule: at each Zipcode, on each time interval if a guardian is present, increase the guardianship value with the value of theta.[23]

Rule 5.4 Guardianship increase rule

The idea of this rule is that when guardians such as police officers are present people feel safer and are therefore less frightened. Guardianship inhibits criminal behaviour, as we will see later in Rule 5.6. An increase in guardianship when a guardian is present can therefore also be translated as that criminals feel there is a higher chance of being caught. The feeling of safety for normal civilians and the feeling of getting caught both fade away slowly if no Police agent is present. This is described in Rule 5.5.

> Guardianship decay rule: at each Zipcode, guardianship decays with a factor beta every time interval.[24]

Rule 5.5 Guardianship decay rule

Every time step the value for guardianship decays with a certain factor (a model parameter). The consensus behind it is that we assume that people feel most safe just after they have seen a police agent, a feeling that slowly decays when no new police unit has been noticed. In the same

---

[22] The rule of opportunity decay can be stated formally. Call the current opportunity value o^t. Then the new opportunity value o^{t+1} is given by $o^{t+1} = o^t * (1 - \text{opportuntiyDecay})$.

[23] The rule of guardianship increase can be stated formally. Call the current guardianship value g^t. Then the new guardianship value g^{t+1} is given by $g^t = g^t + \text{guardianshipTrace}$. The '=' should be red here as becomes.

[24] The rule of guardianship decay can be stated formally. Call the current guardianship value g^t. Then the new guardianship value g^{t+1} is given by

$$g^{t+1} = g^t * (1 - \text{guardianshipDecay}).$$

way, criminals, whom are inhibited by guardianship according to the routine activity theory, will feel less frightened to commit a crime when they have not seen police around for some time.

5.2 The Agents

Just as there is an initial distribution of opportunity among the Zipcodes, there is also an initial population of our main agents, the Criminals[25]. These Criminals are based on the criminal records from the crime data we have available. In our model these criminals continue their criminal life and offend when they cross a Zipcode with a relative high opportunity and a low guardianship value. According to Ferwerda, Jonkmans et al. (1998) street robbery is a crime of which the chance of getting caught is small (see section 3.5). For this reason and because we had no ideas on how to model the arrest process, the Criminals in our model cannot be caught in the model no matter how many crimes they commit in the model.

In the following we will describe how the crime data is used to construct our criminal agents, but first we will explain the properties and behaviours of Criminals. Additionally to the Criminals there are also Police agents that represent, as the name implies, police units. These agents are less sophisticated than Criminals and are directed by a central process. This simplification is introduced, because we had no data or theory on the behaviour of police units. On initialization, Police agents are positioned based on the initial distribution of opportunity. We will go into more detail next.

5.2.1 Selection and initialization of criminal agents

Criminals have several properties; a crime frequency, a list with possible locations and distribution of activity per hour of the week. The first two are specific for each Criminal, the latter is based on one distribution thus equal for all Criminals. We will explain what these properties mean and how they are chosen. The behaviours of Criminals depend on the values of these properties. First we will explain the properties of Criminals, in the next section we will continue with the behaviours.

As we already mentioned in the preface of this section, Criminals are based on real offenders in a crime database. These offenders are based on real persons with different preferences and habits. For example, offender A could have been recorded for 3 crimes while offender B has been recorded for 12 crimes in the same month. A and B obviously differ in the frequency they commit crime. Likewise, A has committed his crimes in the centre of a city and in the suburbs, while B was only interested in the centre. Consequently, A and B also differ in their Geographic Profile (i.e. activity space).

The differences in crime frequency and activity space are captured respectively in the CrimeFrequency and in a list of Zipcodes. The CrimeFrequency is simply captured by the following equation:

$$CrimeFrequency = \frac{numberOfCrimes}{numberOfDaysActive}$$

Equation 5.2 Computation of the crime frequency of a criminal

[25] We use capitalization to indicate that we talk about agents in our model.

This crime frequency of criminals is corrected with a factor so that the average crime frequency of criminals is 1. This is done to prevent very small values of the crime frequencies that could cause zero crime occurrences in a model run.

The numberOfCrimes are the number of crimes recorded for an offender. The numberOfDaysActive is the number of days between the first crime and the start of the model. A property of this equation is that a Criminal representing an offender recorded for 9 incidents over 3 years has a lower crime frequency than one representing an offender recorded for 4 incidents since last year. This seems reasonable. It also means that a first offender recorded a month before the start of the model has a relative high crime frequency, because the number of active days is low. The estimation of the crime frequency therefore becomes more reliable when more crimes have been recorded.

According to previous work (Liu, Wang et al. 2005; Groff 2006; Groff 2007) on crime ABM in chapter 2, we found that street robbery is a crime suited to be modelled by an ABM. Therefore, the criminals in the model are constructed from the crime records on street robbery. Other crime types are completely neglected. A crime record of an offender in our used crime database consists of personal information and the coupling to incidents in which the offender is involved.[26] Each record of an incident contains, among other things, a location and a date. This is the information that will be used for the construction of the criminal agents. Each criminal agent in our model is characterized by a crime frequency and an area in which it is active, the so-called activity space.

In chapter 3 we have discussed the concept of Geographic profiling. This field of study tries to identify the geographic profile of criminals. Methods from Geographic Profiling have been used to search for the home location of a criminal, but also to identify the area of activity. As such, we have used a variation on one of the discussed techniques in chapter 3 to estimate the activity spaces of our criminals. It has been shown that currently the complex techniques do not outperform the simple techniques in accuracy. A variation on the centroid technique (discussed in 3.4) is used to create the activity spaces of our criminal agents. In Figure 5.2 are all activity spaces shown together in the model environment. An activity space of a criminal is represented by a red transparent circle. The centre of this circle is the centroid of the locations of the crimes from the offender's record[27]. The greatest distance between the centroid and one of the crime locations is used as radius. The criminal will be active on the Zipcodes that lie within the area of the circle. More specific initialization settings are discussed when we describe the experiments.

Criminals have a preference for offending on specific hours of the week. Criminals are homogenous in this property, because on average recorded offenders did not have enough recorded street robbery offences to derive a distribution of activity from. A distribution of the most popular hours in a week for offending is, therefore, used to give all Criminals preference to

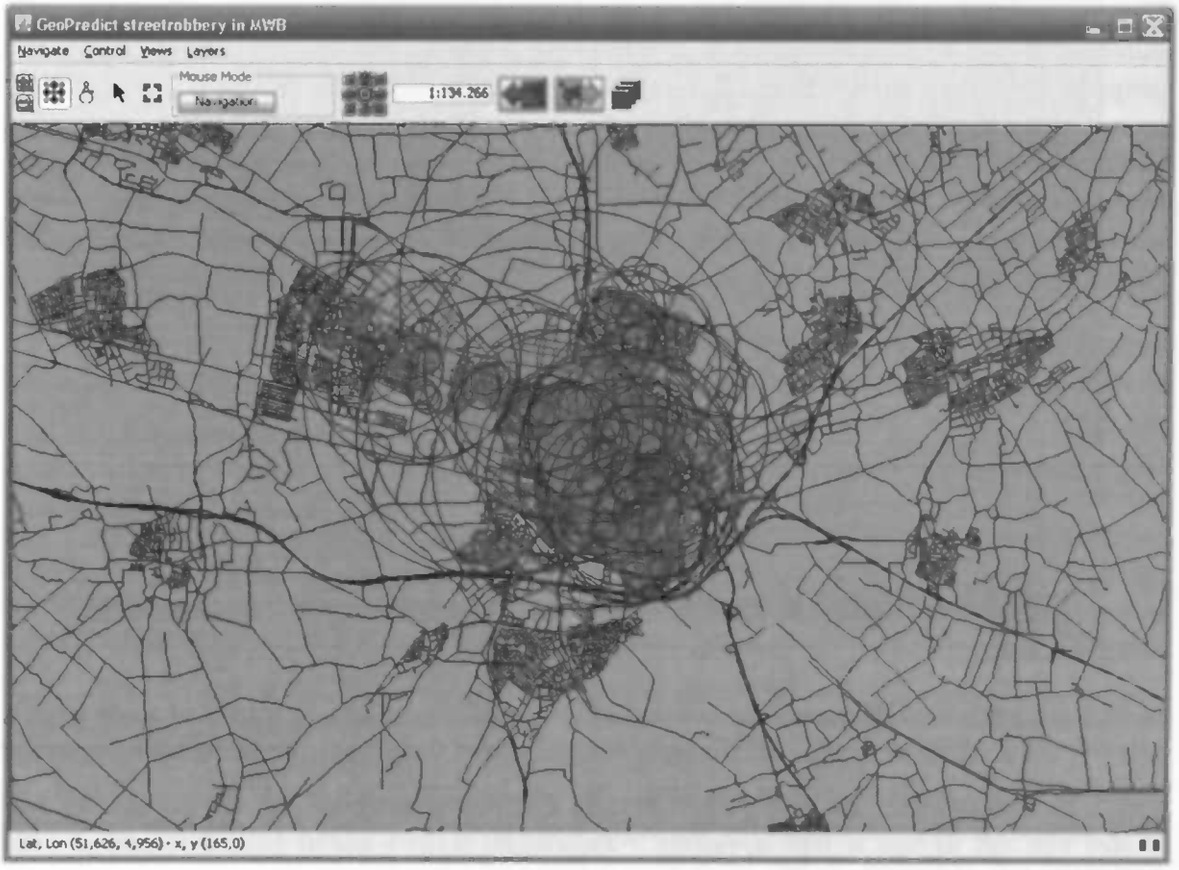
[26] The offender can either be involved as a criminal or as victim. We have only focused on the first to keep it simple.

[27] The centroid technique was already mentioned in section 3.4 and is given by two equations. The first is to find for the *x*-coordinate of the centroid, the second for the *y*-coordinate of the centroid:

$$Centroid_x = \frac{\sum_i^n location_{x_i}}{n}, Centroid_y = \frac{\sum_i^n location_{y_i}}{n}$$

offend on certain hours of the week. This distribution will be used in the criminal behaviour to offend. This value lies between 0 and 1. In Rule 5.6 this preference is captured by $Activity_t$

Figure 5.2 The activity spaces (transparent circles) shown in the model environment



5.2.2 Two behaviours of Criminals: movement and offending

The Criminal agents in our model exhibit two behaviours: moving and committing crime. A Criminal can only move within its activity space. A Criminal can either move randomly towards a new location or he can move based on the attractiveness of Zipcodes. The calculation for the latter is given by the following equation:

$$Attractiveness_{x,t} = Opportunity_{x,t} - Guardianship_{x,t}$$

Equation 5.3 Computation of the attractiveness to commit crime for Zipcode x on time step t

Thus the attractiveness of Zipcode X on time T is the opportunity of X on time T minus the Guardianship of X on time T. We assume here that a Criminal has absolute knowledge within its activity space and so the Criminal chooses the Zipcode with the highest Attractiveness in its activity space. Whether Criminals choose their next location randomly or based on the attractiveness is a model setting.

The other behaviour Criminals have is offending. When they are at a Zipcode they decide whether or not to offend by the following rule:

```
IF (NOT CapableGuardianship(Zipcode_x,t)
        AND (CrimeFrequency *Attractiveness_x,t * Activity_t ) ≥ RANDOM)
THEN [OFFEND]
```

Rule 5.6 Rule for a Criminal to offend on Zipcode x on time step t

This rule says that a criminal will offend when there is currently no capable guardianship and the attractiveness of the current Zipcode times the activity value of the current hour exceeds a random value between 0 and 1. For clarity the part of Capable Guardianship has been extracted in a separate function:

```
IF (Police_x,t) AND guardianSensitivity ≤ RANDOM)
        THEN TRUE
ELSE
        THEN FALSE
```

Rule 5.7 Rule for capable guardianship on Zipcode x on time step t

In short, there is capable guardianship when there is currently no police present and the guardianSensitivity exceeds a random value between 0 and 1.

5.2.3 Behaviour of Police agents

As mentioned earlier, police units are represented by Police agents. Each model run, a fixed number of police agents is included into the model. This is a parameter setting. The police agents are like Criminals in the sense that they are represented by circles and have a dynamic location, a Zipcode. However, these are the only similarities. Police agents do not choose their own location as Criminals and they do not offend. Police agents are not restricted to locations in their activity space, because they have none. Police agents are directed by a central process which sends Police agents to the top number of most attractive locations for crime. Other options were to let Police agents behave autonomously in the environment, but this also meant that we had to correctly model individual behaviour of police agents. Since we have found neither data nor theory on the behaviour of police officers this was no option to us. Furthermore, it is also realistic to assume that police agents are directed by a central process that monitors crime levels. Future models should, however, consider other options.

5.3 Implementation details

In this section we will describe implementation specific details. These details are important enough to include in the main text instead of the appendix. First, we will describe framework specific properties. Second, we will describe the details of the implemented order of the processes earlier described.

5.3.1 Framework

In chapter 4 we already mentioned that we found in literature that is better to use a framework than to start from scratch when applying ABM, because this prevents common errors to be introduced. The Repast J 3.1 framework is regarded as the best framework available for our needs. Repast J 3.1 provides a good coupling with the GIS library OpenMap allowing us to use GIS files we have available on the region of Tilburg. We already explained that the Zipcode objects in our model are based on the zip codes in the region of Tilburg. These Zipcodes are loaded from a so-called shapefile. The complete process is described in the Appendix. The streets in the background are loaded from a shapefile as well.

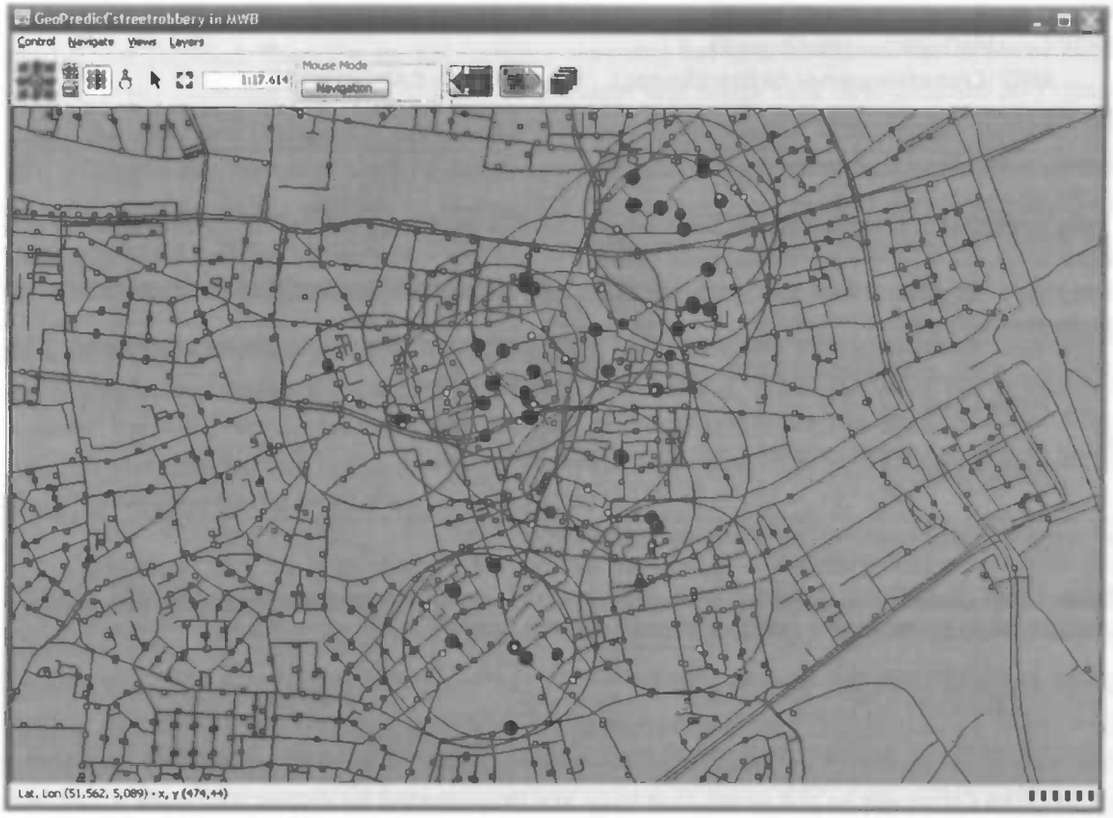


Figure 5.3 The model in action. The green circles are the Criminals and the blue circles are the Police agents. A model run with demonstration data is used so that the agent are easier spotted. The small squares represent the Zipcodes.

The Zipcode objects are implemented as Java objects. In fact, they are implemented as if they were agents according to the Repast framework[28]. However, this is due to the fact that it provides us with some functionality in the Repast framework that would otherwise be hard to achieve. Zipcodes, however, just as described earlier, have a static location and lack behaviour in the sense that they change the state variables of other agents or objects; they are passive.

Just as Zipcodes, Criminals and Police agents in the model are implemented as Java objects. Another option would have been to implement these agents as different Java threads, allowing 'real autonomy' and asynchronous behaviour. However, we have not found examples of such models using the Repast framework. Furthermore, this would make the development and analysis of the model a lot harder and the runtime a lot slower. As an alternative, asynchronous behaviour can be simulated by invoking the behaviour of agents in a random order, so that the order for every step is different. As such, we randomize the order of the list of Criminals before every step. In the next section we will say more about the order of the processes involved of the model.

[28] In Repast J all agents are implemented as Java objects. Our agents (and Zipcodes) implement the OpenMapAgent interface, which basically means that we can make use of some standard functionality of the Repast framework such as the displaying the agents on the map. For more information on this we refer to the website of Repast (repast.sourceforge.net).

5.3.2 Schedule

In the previous section we have described several processes: the initialization of the model, the update of Zipcodes and the behaviours of agents. We have not, however, explained in what order these processes are executed in the model. This will be done in this section.

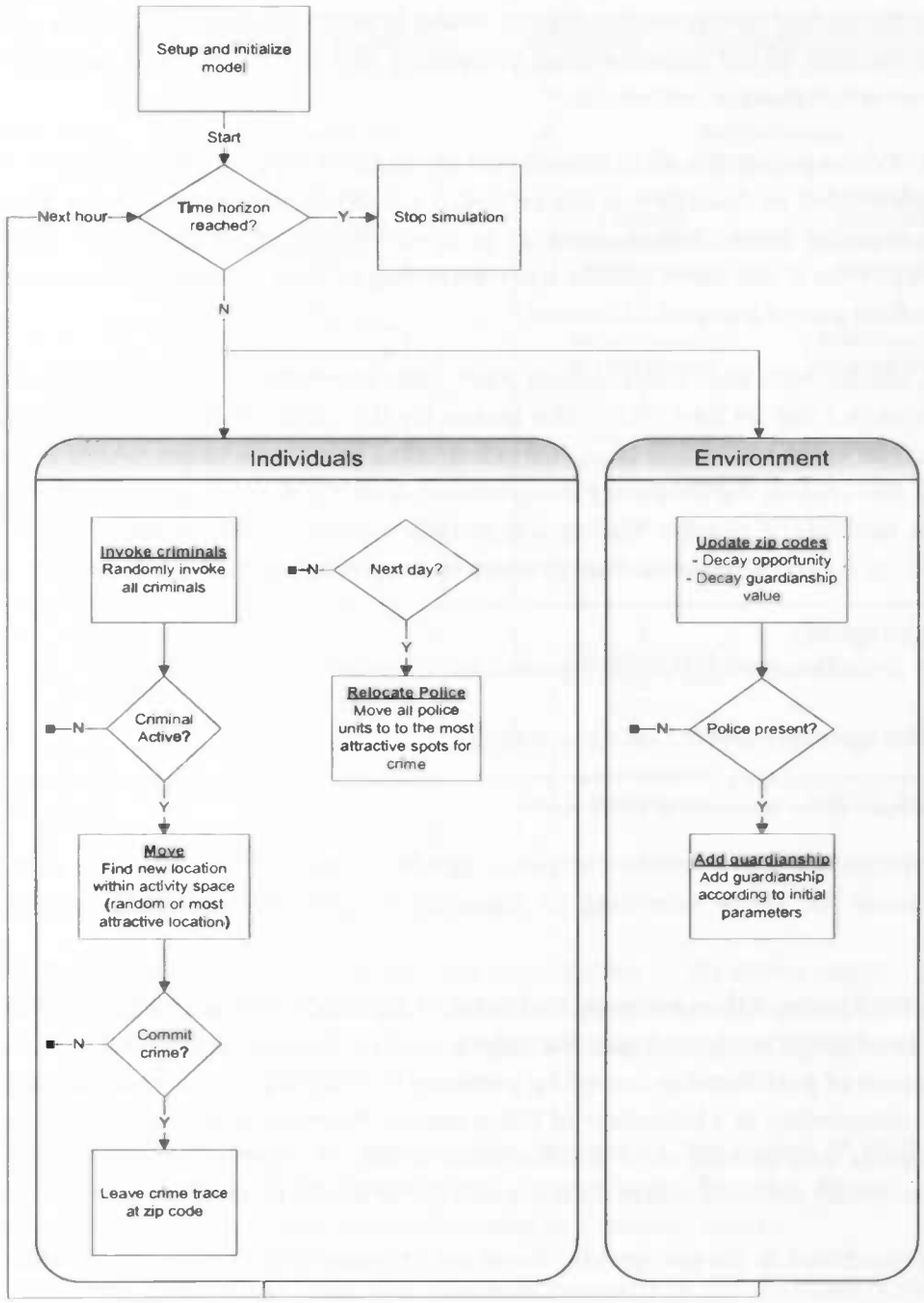


Figure 5.1 Schedule of the model processes

In Figure 5.1 a flowchart of the model is shown. First the model is setup and initialized as described in the Appendix in the section 'Initialization of the model'. The model then proceeds in steps that represent the simulation of one hour in real time. This process continues until a set time horizon is reached i.e. a set number of hours.

A step is divided in the update of agents, and the environment, the Zipcodes. First the list of Criminals is shuffled. Then the model iterates over this list in a for-loop. For every criminal is checked if the Criminal is active. If so, the Criminal will move to a new location and consider crime. As described in section 5.5 Criminals are homogenous in their temporal preferences. This is due to the fact that we had not enough data for each criminal to extract a complete temporal profile. As we will see in section 5.5 the model is set so that every Criminal will be active every step because of the computational complexity and it does not add much to the model results as we will discuss in section 5.5. [29]

Thus, given the previous paragraph all Criminals will move every step to a new location. This movement is implemented as described in the section 5.2.2. When the Criminal is on his new Location he will consider committing a crime as is described in the same section. When a Criminal commits a crime it will leave a crime trace according to Rule 5.2 Opportunity increase rule. This was the final part of a step of a Criminal.

In the flowchart can be seen that Police agents have only one behaviour, movement. Police agents move only once a day on time 00.00. The reason for this is that Police agents cannot be informed about crime numbers on real time. Instead, it takes some time to see where crime is concentrated. For this reason, the frequency of movement is once a day and midnight is chosen because this is the natural end of a day. Earlier was already mentioned that movement of Police agents is directed by a central process. This process is implemented as follows: the list with

```
N <- Number(PoliceAgents)
ListOfZipcodes <- OrderDescendingOnAttractiveness(ListOfZipcodes)
FOR(i=0 to N)
        ListOfPoliceAgents[i].moveTo(ListOfZipcodes[i])
```

Pseudocode 5.1 Pseudocode of the movement of Police agents

Zipcodes is ordered so that the *n* number of police agents is sent to the *n* Zipcodes with the highest attractiveness for crime according to Equation 5.3. The pseudocode is shown in Pseudocode 5.1.

We have come to the update of the environment, the list of Zipcodes. This update involves three updates. The decay of opportunity and guardianship according Rule 5.3 and Rule 5.5. The last update is the increase of guardianship caused by presence of Police agents, as in Rule 5.4. This could as well be implemented as a behaviour of Police agents. However, this was easier because this way Police agents only needed to be invoked once a day. This implementation is just an optimization of the model code and makes no difference in the model behaviour.

All the processes described in the paragraphs above are repeated until a set number of hours are simulated.

[29] We could have just ignored this model property in the model description, because we do not use it. However, it is part of the model design and the reason for neglecting this property might be important for future models.

5.4 Initialization of the model

In the previous section we have described concepts of the model. Here we will describe the initialization process of the model. This part of a model run is especially important because the model is initialized with empirical data, that of crime records.

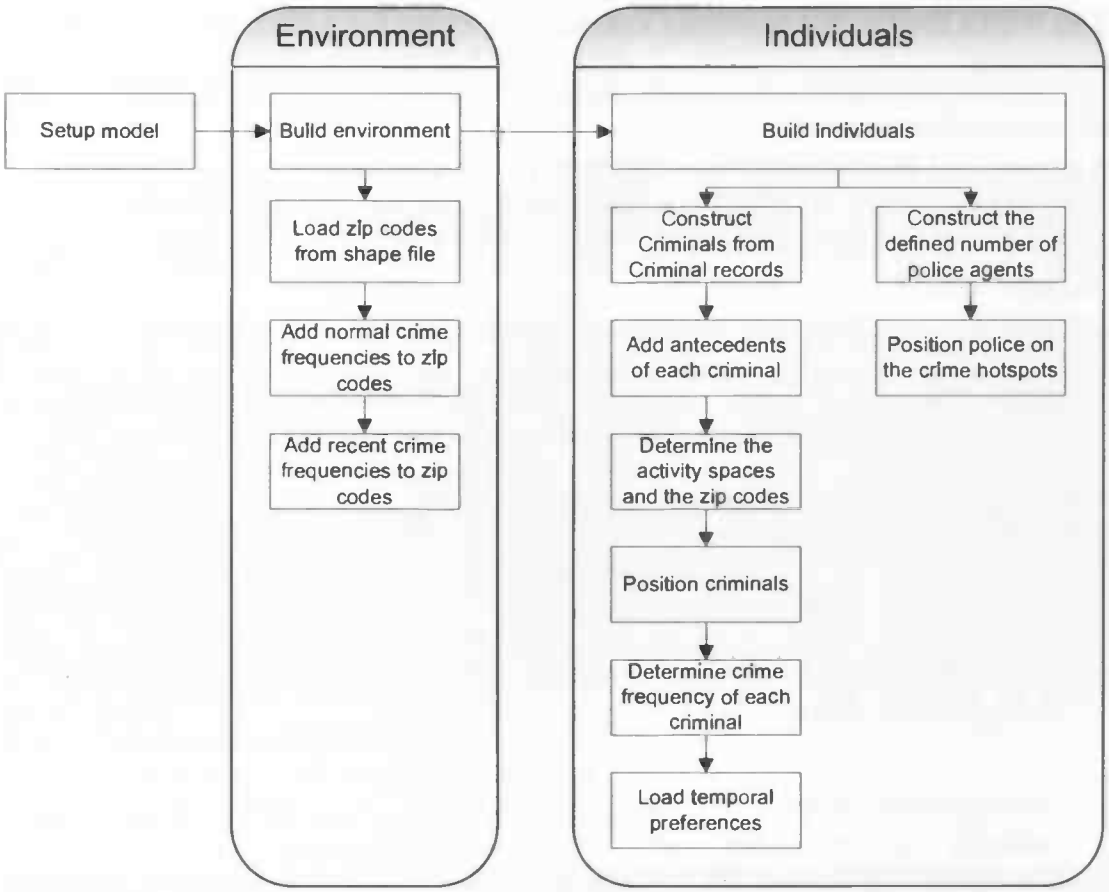


Figure 5.4 Flowchart of the initialization of the model

The setup of the model is divided into the construction of the environment and the creation of agents. The environment is built first, Zipcodes objects are constructed. After the properties of Zipcodes are defined; the agents, the Criminals and the Police agents are constructed and initialized.

Zipcode objects are based on the zip-codes present in the region of study, Tilburg. Crime numbers are derived from the database and coupled to the Zipcodes. Recent crimes, the crimes of two months ago, add crime trace to Zipcodes as described earlier.

Criminals are constructed from the criminals recorded in the database. Criminals that were recorded for street robbery in Tilburg before the start date of the model are selected. The past crimes of these criminals are used to compute the crime frequency and to determine the Zipcodes in their activity space (see Section 5.2.1).

Police agents are initialized based on the model parameter NumberPoliceAgents. This parameter determines how many agents will be active in the model. The initial locations of these Police agents are based on the initial distribution of crime attractiveness as computed by Equation 5.3 Computation of the attractiveness to commit crime for Zipcode.

5.5 Sensitivity analysis

In the previous part of this of the chapter we described the concepts of our model. Here we will describe the model parameters that have been used. Some of them were already shortly mentioned in the update rules of the Zipcodes and the behaviours of Criminals. After the description of parameters, we will test how different parameters setting influence the model results. This will be done with a sensitivity analysis which tests the model with many different parameter settings.

5.5.1 Model initialization data

Earlier in section 5.1.1 was described how Zipcodes are initialized and in section 5.2.1 was described how Criminals are initialized. Both processes rely on empirical data. For our experiments we used the data summarized in Table 5.1. The data set with criminals contains data on 262 offenders. [30]

Data set	
Criminals	File with offenders that have committed street robbery in the region Tilburg in the period from January 2001 until December 2004. The offences used to derive preferences of these criminals are also only street robberies from the region Tilburg in the period from January 2001 until December 2004.
Distribution of activity per hour of the week	The temporal profile of all criminals is based on the distribution of street robberies per hour of the week in the region Tilburg in the period of January 2001 until December 2004.
Crime frequencies per zip-code	The crime frequencies per zip-code are based on a file with the number of street robberies for each zip-code in the region Tilburg in the period of January 2001 until December 2004.
Recent street robberies	Recent street robberies from November and December 2004 in the region Tilburg are used to add a crime trace as described in Rule 5.1 Opportunity initialization rule.

Table 5.1 Summary of the empirical data used for the initialization for the model

5.5.2 Model parameters

A table with a summary of the model parameters is shown in Table 5.2. The parameters can be divided by the object or agent that depends on it. The parameter NumberPoliceAgents simply

[30] For the exact derivation of this data one can look into the Appendix 1 The data set with Criminals is based on the *incidentSelectionFilename*, personSelectionFilename and incidentPersonFilename. The data set with the distribution of activity per hour of the week is described in the appendix as *temporalProfileFilePath*. The file with the crime frequencies per zip-code is described as *frequencyFilePath*. The file with the recent street robberies is described as *incidentsLastTwoMonthsFilename*.

sets the number of Police agents that are present in the model, no more, no less. The parameters CrimeTrace, GuardianTrace, GuardianDecay and OpportunityDecay influence the way Zipcodes are updated. These parameters were already mentioned in section 5.1.1 where the update rules of Zipcodes were discussed. The parameters control the amount of opportunity and guardianship is added or subtracted each step or after a crime event.

Parameter	Description
NumberPoliceAgents	The number of Police agents in the model.
CrimeTrace	The value that is added to the opportunity value on a Zipcode when a crime is committed.
GuardianTrace	The value that is added to the guardianship value on a Zipcode when a police unit is present.
GuardianshipDecay	The proportion of the opportunity value of a Zipcode that is subtracted when updated.
OpportunityDecay	The proportion of the opportunity value of a Zipcode that is subtracted when updated.
GuardianSensitivity	The sensitivity (in %) of all Criminals to Police agents present on the same Zipcode in their choice to commit crime.
AgentChooseZipcodes	The Boolean that sets whether the criminals should choose their Zipcodes randomly or based on the attractiveness value.
MinimumRadiusActivitySpaceIn KM	The minimum radius of the activity space in kilometres for Criminals.
NumberActiveHoursCriminals	The number of hours all Criminals are active a day.
RandomSeed	The seed that feeds the random number generator.

Table 5.2 Description of the model parameters

The parameters GuardianSensitivity, AgentChooseZipcodes and MinimumRadiusActivitySpaceInKM influence the behaviour of criminals directly. GuardianSensitivity was already mentioned along with the criminal behaviour 'Offending'. GuardianSensitivity can be regarded as the amount of respect criminals have for Police presence. GuardianSensitivity controls the chance that a Police agent has influence on the criminal behaviour. For example, according to Rule 5.6 and Rule 5.7 when this value is 100 it means that no crime will occur when a Criminal meets a Police agent on a Zipcode. In the same way, a value of 50 means that a present Police agent will restrain a Criminal for 'thinking' of crime with a chance of 0.5. Thinking here means that when the criminal is not constrained by Police presence he does not necessarily offend (see Rule 5.6). The parameter AgentChooseZipcodes controls whether or not new locations from the activity space are chosen randomly or based on the attractiveness of Zipcodes. In the latter case, the Zipcodes with the highest attractiveness are chosen (see section 5.2.2).

The last parameter that influences behaviour of Criminals is the MinimumRadiusActivitySpaceInKM parameter. This parameter sets the minimum radius of the circle that represents the activity space. Ideally, this parameter should not have been necessary, however, in our crime data offenders with only one offense are registered. Without this parameter using the geographic profiling technique to calculate the radius for these criminals would result in radii with length zero, because that is the distance between a point and itself. Furthermore, there exist Criminals in the database that have committed offenses only very close to eachother, making the activity space very small as well. The

MinimumRadiusActivitySpaceInKM parameter was introduced to give Criminals a minimum size of their activity space.

Finally, the RandomSeed is the parameter that is the seed for the random number generator. Using a fixed seed for a random number generator makes it possible to regenerate a model run that is based on (partly) random events. Furthermore, by changing the seed one can check if the model behaviour is not caused just by the accidental choice a seed.

5.5.3 Parameter settings

Now we have described the model parameters, we will describe the parameter settings for the sensitivity analysis. We were forced to limit the number of values for each parameter because of the computational complexity of the model. A model run for the simulation of a year took up to one hour on a computer with a CPU of 3.19 GHZ and 2.0 GB of RAM. In Table 5.3 we have summarized the values over which the model was varied. These values are chosen with care and are discussed in the next paragraph. Given the possible combinations of parameter the number of possible runs is 5832. This means that the complete sensitivity analysis for a year would take about 5832 hours, or 243 days! We have tried to optimize the code, but we have not been able to make significant improvements in execution speed. Therefore, instead of running the model for one year we have done the sensitivity analysis for runs that simulated one month. Given that a run of one month still involves 744 steps[31] this should be enough to reveal possible trends in a model run and should also make the outcomes of the model significant enough to be compared with real crime data. One could say we should have varied the model parameters more, however every addition of a parameter value would increase the number of possible combinations dramatically. Therefore, we have chosen a number of variations that would be acceptable in execution time.

When looking at Table 5.3 we see that for the parameter NumberPoliceAgents three values are chosen: 1, 163 and 225. This number is not based on the number of police units present at a given moment in the region Tilburg. We could have asked the police how many police units are active in the region Tilburg at a specific time, however, in real life not all police units are occupied with the crime street robbery and we only model street robbers. A correct estimation of the right number of police is thus hard to make. Therefore, we looked at how crime levels in our model were influenced by different numbers of Police agents. A very high number of Police agents prevent any crime occurrence. We found that the chosen values give a nice correlation on average.

Parameters	Values
NumberPoliceAgents	1, 163, 225
CrimeTrace	0.01, 0.005, 0.001
GuardianTrace	0.001, 0.0005, 0.0001
GuardianshipDecay	0.001, 0.0005, 0.0001
OpportunityDecay	0.001, 0.0005, 0.0001
GuardianSensitivity	25, 75, 100
AgentChooseZipcodes	False, True
MinimumRadiusActivitySpaceInKM	0.3
NumberActiveHoursCriminals	24
RandomSeed	100, 300, 400, 500

Table 5.3 Values for the model parameters in the sensitivity analysis

[31] When January is used there are 24 times 31 hours i.e. 744 hours.

The next four settings; CrimeTrace, GuardianTrace, GuardianshipDecay and OpportunityDecay are closely related. Together they influence the increase and the decay of the attractiveness of Zipcodes for crime (see section 5.1.1). For this reason they should be discussed together. The last three parameters are varied over the same values, this is because the rules that use these parameters are used every step of the model and we had no reason to give one parameter higher values than another. The first parameter, CrimeTrace is different in that sense. This parameter's value is added to the opportunity value when a crime is committed (see Rule 5.2). Under normal conditions it is not expected that a crime will occur every hour or more on a zip-code. Rule 5.2 is consequently not executed for every Zipcode on every step like the other rules for the other three parameters. Therefore to give this parameter meaning, the value is varied over the values of OpportunityDecay, the other parameter that controls the amount of chance in the opportunity value of Zipcodes (see Rule 5.3), with a factor from 1 to 100.

The values of GuardianSensitivity were varied over 3 values 25, 75, 100 to see what the effect is when *all* Criminals are less or more feared for Police presence. The parameter AgentChooseZipcodes was either true of false, meaning that the Criminals' choice for a next location was either random or based on the attractiveness of Zipcodes. This parameter was introduced to see the effect of the distribution of opportunity value on the crime distribution.

The MinimumRadiusActivitySpaceInKM is only set on 0.3. We did not want this value to be too small, because only very few Zipcodes would then be available when a Criminal had few prior offenses. We also did not want the area to be too big, because that would deteriorate the geographical information we have on an offender. Therefore we have chosen for a radius with the value of 0.3 kilometres, making the minimum activity spaces covering an area of a small neighbourhood. We have chosen not to vary this in the sensitivity analysis, because this would increase the computation time and we thought it would not have a major influence (for small changes).

Finally, we have four different values for the RandomSeed to prevent having model results that can be contributed only to the accidental choice of this value. In other models such as that of Groff (2006) five values were used. Basically, when it looks like the model results are very strongly influenced by the random seed, this value should be varied until one is confident enough to have found a typical outcome.

5.5.4 Comparisons of the model output: correlation with crime data

In the previous part we have described what the parameters were and why we have chosen for certain parameter values for the sensitivity analysis. The sensitivity analysis is used to see what model results can be expected with a variety of parameter combinations. The model will output crime numbers per zip code. To get an idea how good certain parameter settings are we look at the correlation between the distribution of crimes predicted and the distribution of crimes observed. Before we actually execute the sensitivity analysis we will first describe what kind of correlations we should expect. In crime forecasting the random walk method is commonly used as a minimum requirement of a forecast (Gorr and Olligschlaeger 2002, see section 3.6). The random walk method takes past crime rates to predict crime rates in the future. For example, a random walk method with a lag of a year will predict the same crime numbers observed in December 2004 for December 2005. Since the random walk is regarded as a minimum requirement for a forecast, we have computed the correlations between different years of crime

data. This gives us the correlations we may minimally expect when we compare the results of a good forecast with observed crime rates.

Other evaluation methods suggested by previous research (Groff 2007, section 2.2); Ripley-K, Kernel Density and the ANOVA test. These tests do not tell how well a prediction is but give only certainty about properties of the model output. The Ripley-K measure can tell whether a spatial distribution of points is random by comparing the distribution with a random distribution. Due to the use of references to human-generated patterns (e.g. population, housing and in our case zip-codes of the region Tilburg) this is already not the case because human-generated patterns are never random (this was mentioned in Groff 2007). The Kernel Density method can create a hot spot map (section 3.6.1) of a distribution of points. This method is also hard to use for comparison because there is no standard way to create a map, therefore it requires insight into the theoretical grounds of this method and experimentation with the creation of maps. The ANOVA test can be used to test if the results found are significant and not based on random processes. In section 3.6.4 we discussed the evaluation technique of Bowers, Johnson et al. (2004), an evaluation technique based on the police practise. This method evaluates predictions on the ease with which they can be put into practise. For example, a crime can correctly be predicted for a certain area but if this area is too big it is hard to cover by a police patrol. Our model has not been designed to output a crime prediction that is usable by the police. Instead, the output is the number of crime events that have taken place on zip-codes. The correlation with real observed crime is therefore a lot easier to use.

The correlation measure we used is the Pearson product-moment correlation coefficient[32]. This coefficient was used because it shows how two variables increase and decrease together. The model might predict an increase correctly, while the exact number was wrong. A correctly predicted increase is valuable for the police organization. The Pearson product-moment correlation will value a correctly predicted increase or decrease in crime. Other methods such as the Mean Squared Error (MSE) will remain low only when the absolute error is low. Furthermore, when there are many zero values, and we will have many Zipcodes that have no crime events during a model run, methods such as the MSE will be mislead.

When we first compared the correlations between the crime numbers of zip-codes between years of crime data we noticed that the correlations were quite low. The cause for this low correlation is that there are, for example, only 307 street robberies in the region Tilburg in the year 2005. Since we have chosen to simulate only for one month this number becomes even smaller. For January 2005 the number of street robberies is only 20. Given that there are 6447 Zipcodes in the selected region, the chance to correctly predict the (relative) number of crimes for a zip-code becomes very small. For these reasons, we divided the region Tilburg in two grids of cells, one with 10 by 10 cells and another with 100 by 100 cells.

In chapter 3 we found that the minimal requirement of a crime forecasting technique is to be better than the "lag 12" method. Therefore we have looked at the correlations between different years of crime data in order to see what correlation should be expected for our model forecast. In Table 5.4 the correlations between different years of crime data are shown for the 10 by 10

[32] The Pearson product-moment correlation coefficient (PMCC) is obtained by dividing the covariance between the two variables by the product of their standard deviations. PMCC $= \frac{\sum Z_X Z_Y}{n-1}$, $Z = \frac{x-\mu}{\sigma}$ (from http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient).

grid and the 100 by 100 grid. The correlations between 10 by 10 grids are very high. This is not very strange, because most street robberies occur in the centre of the region of Tilburg and the cells are relatively big. The whole selected area of the region Tilburg is about 10 by 10 kilometres, thus one cell is one kilometre squared for the 10 by 10 grid. The average correlation is 0.90. This correlation will be used as a minimum requirement for our model prediction on the 10 by 10 grid. For the 100 by 100 grid cells are 100 by 100 meter. This smaller cell size makes it harder to correctly predict crime events. The lower correlations show this principle. We will use the average of around 0.25 as a minimum expectation for the prediction on the 100 by 100 grid.

Compared years for January	Grid 10 by 10	Grid 100 by 100
2005-2004	0.865635866	0.261725178
2005-2003	0.894598285	0.234043301
2005-2002	0.943618865	0.30382291
2005-2001	0.956800857	0.306128811
2004-2003	0.859139023	0.241889546
2004-2002	0.818953535	0.199005225
2004-2001	0.804028096	0.153759446
2003-2002	0.925441186	0.327316896
2003-2001	0.903996837	0.294361693
2002-2001	0.975458585	0.279864121

Table 5.4 Correlations between years of crime data

In summary, we have obtained the minimum requirements for a prediction of our model. We will continue to discuss the results of the sensitivity analysis in the next section.

5.5.5 Results

In the previous section we discussed the parameter settings and the data loaded on initialization. In this section we will discuss the results of the sensitivity analysis. We have compared the values of the model's predictions on crime with that of the real observations on correlation in the same way we compared crime data between years as described in the previous section. When we compare the found correlations, the correlations between crime data of previous years, discussed in the previous section, are used as a reference point of what can be expected.

The combinations of the values in Table 5.3 gives a total of 5832 runs. This number is too much to discuss in full here. We will therefore summarize the correlations over all runs in Table 5.5.

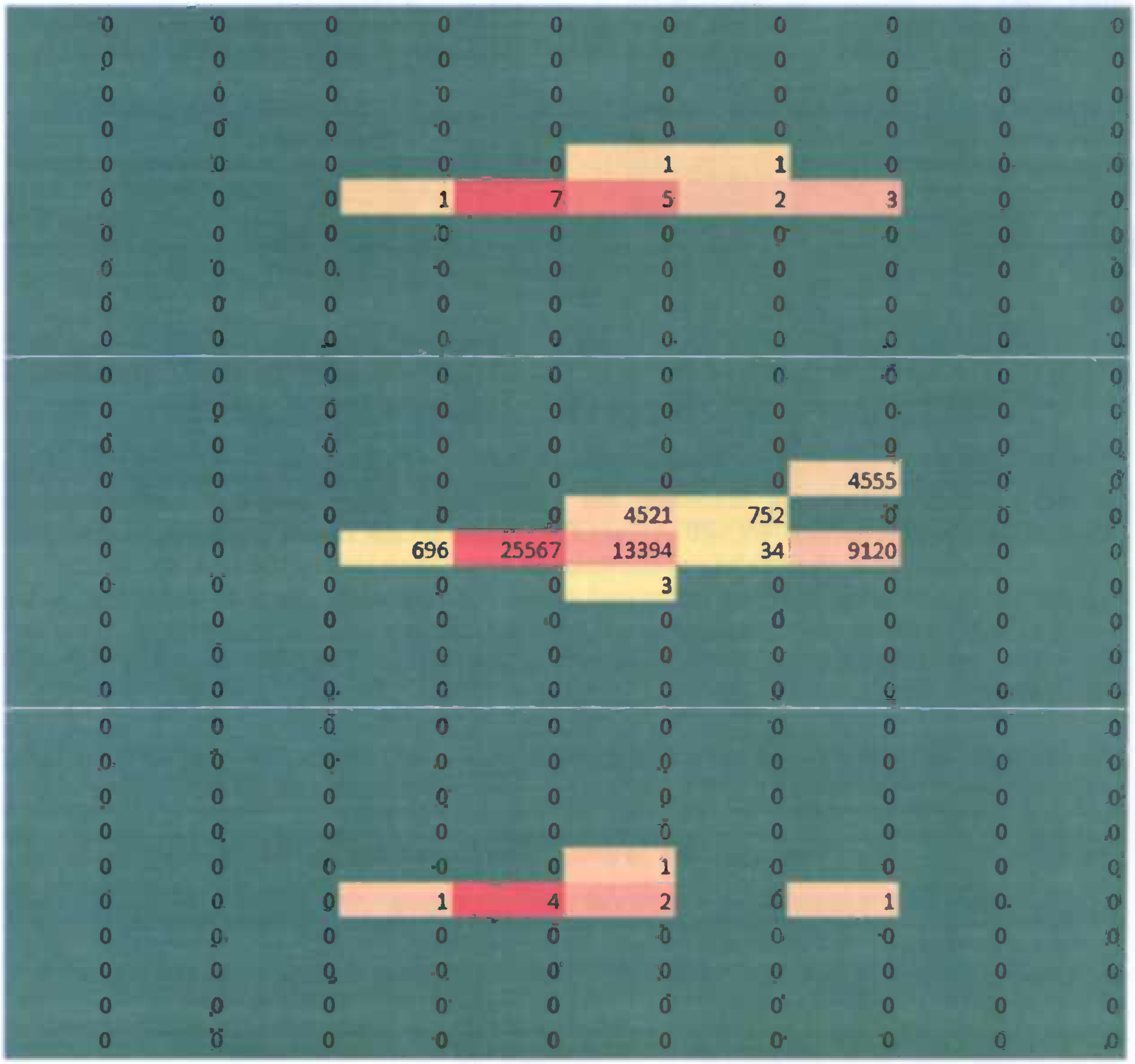
Value	Correlation 10 x 10 grid	Correlation 100 x 100 grid	Number of crimes
Mean	0.601309542	0.044823698	26736.42661
Standard deviation	0.349595352	0.06388215	31867.05558
Minimum	-0.020633108	-0.001414704	3
1st quartile	0.22281563	-0.000428797	44
Median	0.763954509	0.015642618	4430
3st quartile	0.881371277	0.066164824	49510
Maximum	0.949272602	0.42083416	106315

Table 5.5 Summary statistics for all model runs

When we look at the correlation of the 10 x 10 and the 100 x 100 grid the first thing to notice is that the maximum correlation values are quite high. Even the mean and median correlation are quite high at first sight. However, when we compare these correlations with the correlations between the recorded crimes of several years discussed in the previous section, we have to

conclude that only the correlations on the 10 x 10 grid are quite high, even if we look at the median and average. For example, when we look at the mean and median of the 100 by 100 grid it does not come above 0.05, while the minimum expected correlation is 0.25 according to Table 5.4.

When we look at the highest correlations for the 10 by 10 grid we see no significant parameter setting that causes these high correlations. There is no random seed that dominates and neither does any of the other parameters. However, striking is that the correlations that are high are mainly caused by the runs that have a high number of crimes. With a few exceptions the runs with more than 5000 number of crimes predicted have a high score on correlation (all above 0.838245275). The lowest correlations come from runs with relative few predicted crimes, there are some exceptions to this rule however. A run with a very high correlation (0.94312517845483) comes from a run with only 9 crimes. The distributions of crime



0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0
0	0	0	1	7	5	2	3	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	4555	0	0
0	0	0	0	0	4521	752	0	0	0
0	0	0	696	25567	13394	34	9120	0	0
0	0	0	0	0	3	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0	1	4	2	0	1	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Figure 5.5 Grids compared for the 10 by 10 grid. The top grid (A) shows the observed street robberies for January 2005 in the region Tilburg. The middle grid (B) shows the grid with the best correlation (0.949272602). The grid below (C) shows a grid with a very high correlation (0.94312517845483) , but with a low number of crimes (9).

predicted by these runs discussed are shown in Figure 5.5. The parameters of this run are presented in Table 5.6.

Figure 5.5 shows that although the correlation may be very high this still does not mean that the results are predicted very precise in location: In grid B two cells with crime are predicted where no crime has been observed, while in grid C two cells have been missed. This is a property of the correlation measure. Only when two variables, the observation of crime and the prediction of crime change together the correlation improves. The places without overlap do not influence the correlation much, only when it is a relative big part of the sum of all values. A further important point to notice is that it looks like all the zero values are predicted very well. To understand that this is not very surprising one must keep in mind that the geographical distribution of Zipcodes already provides some sort of bias; some cells cover areas that do not have Zipcodes in it and other cells cover areas that are highly populated and have therefore many Zipcodes. [33]

Parameters	Highest correlation	High correlation, low crime number
NumberPoliceAgents	225	225
CrimeTrace	0.01	0.005
GuardianTrace	0.0005	0.001
GuardianshipDecay	0.0005	0.001
OpportunityDecay	0.001	0.0001
GuardianSensitivity	75	100
AgentChooseZipcodes	True	False
MinimumRadiusActivitySpaceInKM	0.3	0.3
NumberActiveHoursCriminals	24	24
RandomSeed	300	500

Table 5.6 The parameter settings of runs with a high correlation on 10 by 10 grid.

[33] In the Netherlands zip-codes are based on the number of addresses, more populated areas have thus more zip-codes per square meter than less populated areas.

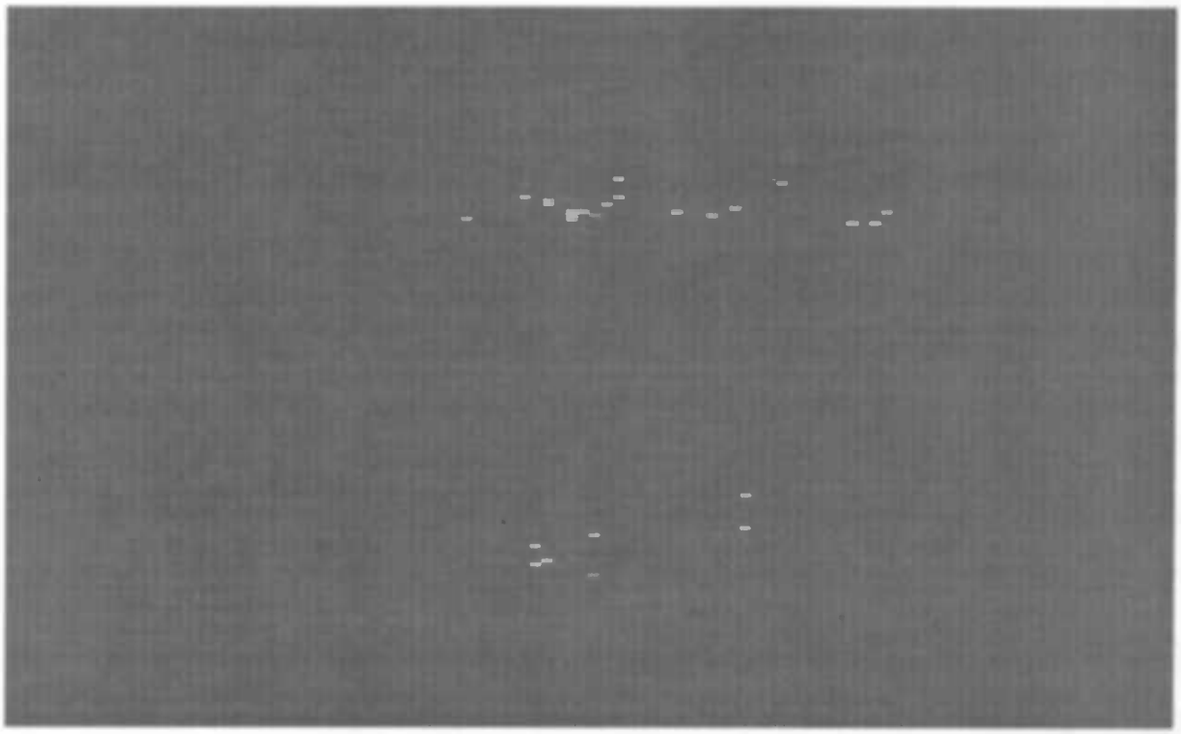


Figure 5.6 Two 100 by 100 grids, the top grid (A) shows the distribution of observed street robberies for January 2005, the grid below (B) shows the best prediction for this period. Cells with value 0 are in both grids green. Red is the highest value in both grids. The highest value of the grid A is 2 while the highest value of grid B is 703. Grid A shows a total of 20 crimes, while grid B shows a total of 58642.

In conclusion, the model performs quite well for the prediction of the 10 by 10 grid. We will now look at the correlations for the 100 by 100 grids. Striking is that a high correlation on the 10 by 10 grid does not necessarily mean a high correlation on the 100 by 100 grid, sometimes even the opposite. The run with the highest correlation (0.420834160462375), of which the grid is shown in Figure 5.6 together with the observed crimes, has a medium correlation (0.742056948) on the 10 by 10 grid. When we look at the grids in Figure 5.6 it becomes clear why the correlation on the 10 by 10 grid is lower; it seems to have only 2 values in common. It does not seem that the correlation of 100 by 100 alone tells us much.

From the previous discussed runs we saw that one correlation alone does not tell us much, therefore we will look now at the run that has the highest overall correlation with the grids of observed crimes. In Figure 5.7 is the run shown with the highest overall correlation. We will analyze the trend lines of this run and compare it with a run with few crimes and a run with an average crime number.

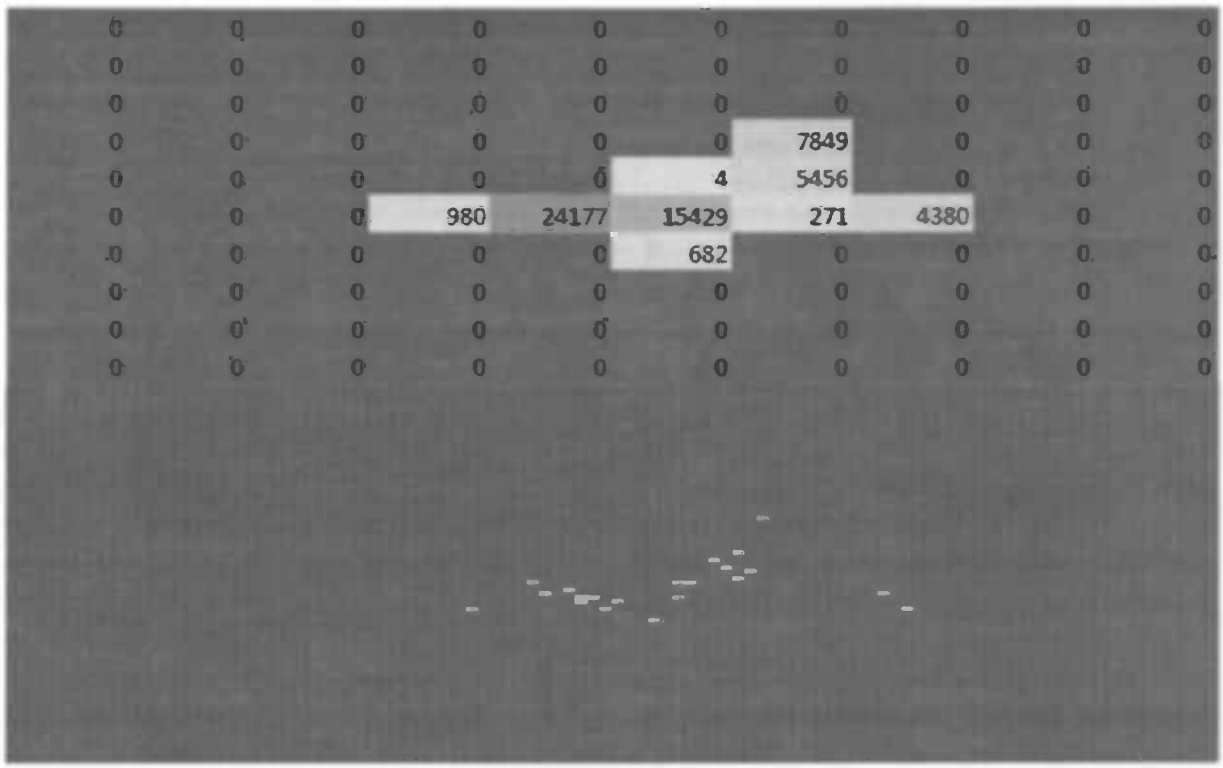


Figure 5.7 The grids of the run with the highest overall correlation, respectively 0.915581778 and 0.260418861 for the 10 by 10 grid and the 100 by 100 grid

The trend lines of the sum of all opportunity and guardianship values of the run with the best correlation are shown in Figure 5.8. The total opportunity grows exponentially in time. The total guardianship grows also exponentially but at a much lower rate. Given Equation 2.1 it is clear that the attractiveness value for crime grows exponentially as well. When the crime attractiveness value is high enough it will cause that a Criminal will offend every step when there is no capable guardianship. This can be seen in Rule 5.6; when the product of crime attractiveness and the activity is 1 or higher (the random value lies between 0 and 1) a Criminal will offend without capable guardianship.

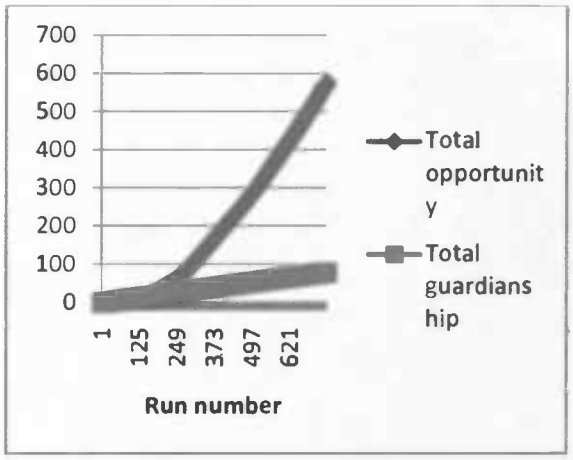
Parameters	Values
NumberPoliceAgents	225
CrimeTrace	0.01
GuardianTrace	0.0005
GuardianshipDecay	0.0001
OpportunityDecay	0.0001
GuardianSensitivity	75
AgentChooseZipcodes	True
MinimumRadiusActivitySpaceInKM	0.3
NumberActiveHoursCriminals	24
RandomSeed	400

Table 5.7 The parameter settings for the run with the overall highest correlation.

When looking at the parameters of this run in Table 5.7 it can be seen that GuardianSensitivity is 75, meaning that on average Criminals are insensitive to Police presence in 25% of the time steps. Thus even when there is a Police agent present at the Zipcode, when the crime attractiveness is high enough a Criminal will commit once every 4 steps on average. This explains the high number of crimes seen in Figure 5.7. Another possible explanation is the

relocation frequency of Police agents. Police agents move only once a day, in contrast to Criminals who can relocate every hour to a (more attractive) location.

Figure 5.8 The total opportunity and total guardianship shown during the run with the best overall correlation

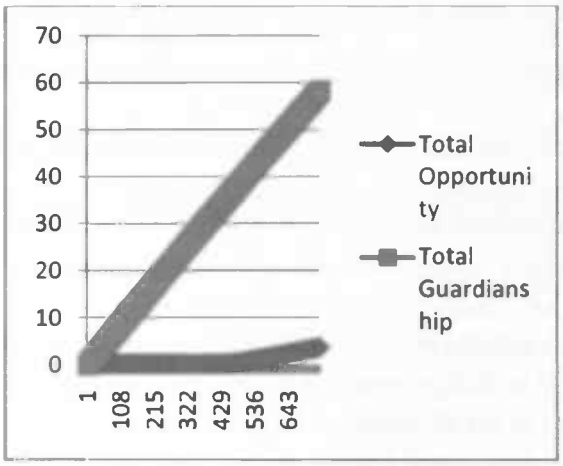


Parameter	Value
NumberPoliceAgents	225
CrimeTrace	0.01
GuardianTrace	0.0005
GuardianshipDecay	0.0001
OpportunityDecay	0.0001
GuardianSensitivity	75
AgentsChooseZipcodes	True
MinimumRadiusActivitySpaceInKM	0.3
NumberActiveHoursCriminals	24
RandomSeed	400

Table 5.8 Parameter value of the run with the highest overall correlation.

When we look at the trends lines in Figure 5.9 we see a different situation. The total guardianship is greater than the total opportunity. Without knowing the correlation and crime numbers one could think that the crime attractiveness during this run was too low for Criminals to offend. However, the blue line, the line of the total opportunity, sees a small increasing trend after 450 runs. Given Rule 5.2, this must be due to Criminals that have committed crime. When looking at the number of police agents, 163 and knowing that there are 262 Criminals (section 5.5.1), there are thus always criminals that are free to offend. Another explanation is that, although opportunity decays faster than the guardianship on average, there can be more than one Criminal present on a Zipcode at a time, while only one Police agent. This situation can cause the opportunity value to increase faster than the guardianship value when these Criminals start to commit crimes. The number of crimes of this model run, 773, indicates that this must have been the case. During the model run the number of crimes per step grew from 0 to 3.

Figure 5.9 Trend line of the total opportunity and total guardianship for a run with a higher total guardianship than total guardianship



Parameter	
NumberPoliceAgents	163
CrimeTrace	0.005
GuardianTrace	0.0005
GuardianshipDecay	0.0001
OpportunityDecay	0.0005
GuardianSensitivity	50
AgentsChooseZipcodes	False
MinimumRadiusActivitySpaceInKM	0.3
NumberActiveHoursCriminals	24
RandomSeed	500

Table 5.9 Parameters of model run with a higher total guardianship than total guardianship

This increase seems to be the consequence of the random choice for a location, set by AgentsChooseZipcodes seen in Table 5.9. The correlations for the 10 by 10 and 100 by 100 grid resemble this random choice, because they are respectively 0.0166835408107613 and 0.000431833041078861.

We have now shown two situations where either the total opportunity or the total guardianship dominated. We have not found a model run where the total opportunity or total guardianship grows equally or goes up and down. We seem to have made design choices that cause the dominance of either opportunity or guardianship. In the next section we will discuss scenario's in which we focus on specific parameters. The last scenario leaves out the initial distribution of crime, causing different trends of the opportunity and guardianship values.

5.6 Scenario tests

In the previous section was shown what effect different parameters settings can have on model results. We did not find very strange results here. No parameter really dominated the outcomes. In this section we will test the model on three scenarios to illustrate possible uses of the model. In the first scenario we will test the effect of different properties of Police enforcement: no enforcement, random and targeted enforcement. In the second scenario we will test the effect of differences in GuardianSensitivity settings to see what the effect is when Criminals have more or less respect for police presence. In the final scenario we will test what the effect is when the initial opportunity distribution is not based on previous crimes but on a uniform distribution. With the first two scenario's we hope to show the possible value of our model for tests on police policy. With the final scenario we want to test if the derivation of activity spaces tell us something about future crime. The parameters values obtained from the run with the best overall correlation are reused in the scenarios when possible.

5.6.1 Police enforcement

In this scenario experiment will be analysed what the effects is of no police and randomly directed police compared to targeted police. The goal of this experiment is to see what differences it makes for the model results if Police agents are not directed to the most attractive Zipcodes, but instead sent randomly to Zipcodes. It is expected that targeted police enforcement will be more efficient in preventing crime than randomly directed police enforcement. The 'null condition' is the setting in which no police is present.

The model parameters will have the values as shown previously in Table 5.8. Only the number of police agents is obviously adjusted when we test the effect of no police enforcement. The data used in this experiment is the same as in the previous experiment (see section 5.5.1). In Figure 5.10 the trend lines of the run with randomly directed Police agents and the run with no Police agents are shown. As expected there are more crime occurrences in the runs that have no (targeted) police enforcement. For the run with random Police agents we see 102568 crimes and for the run without Police agents this is 102441. Nothing different from what we have seen before. When we compare the run with random Police enforcement with the run with targeted Police enforcement on crime number and correlation, according to the settings of the run with the highest overall correlation described in section 5.5.5, we see nothing significant different on the 10 by 10 grid. The correlation for run with random police enforcement on the 10 by 10 grid is 0.9198465081309583 and on the 100 by 100 grid 0.04819238221132854. The lower correlation on the 100 by 100 grid could be due to random effects. In Figure 5.10 is the trend line shown of the total opportunity and guardianship during the model run with no Police

enforcement and the run with random Police enforcement. As can be seen in the figure the trend lines on total opportunity are pretty much the same. In summary, in our model random Police enforcement is as effective as no Police enforcement.

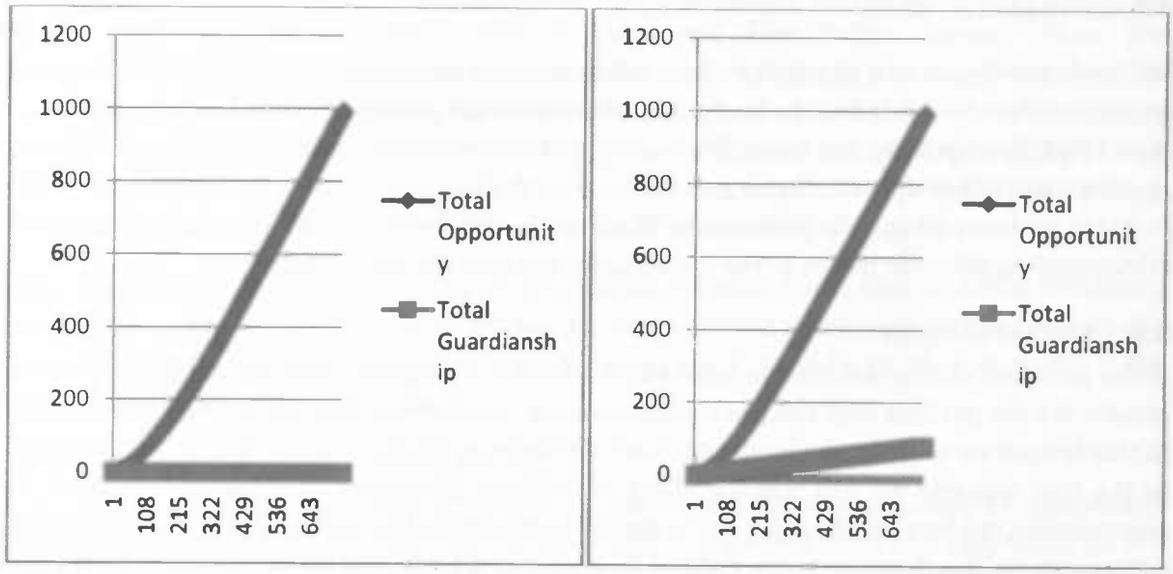


Figure 5.10 On the right (A) the total opportunity and total guardianship during the run with randomly directed Police agents . On the left (B) the total opportunity and total guardianship during the run with no Police agents

5.6.2 Guardian sensitivity

In the previous section we tested the effect of targeted Police enforcement compared to random Police enforcement. In the next experiment we will test the effect of different settings of guardianSensitivity. This parameter was earlier described as the amount of respect criminals have for police presence. In this experiment is thus tested what the effect is of a lower or higher amount of respect for Police presence on crime rates. It is expected that a higher amount of respect causes lower crime numbers, thus a higher GuardianSensitivity is expected to lower crime numbers.

The model parameters will have again the values as shown previously in Table 5.8. Only the GuardianSensitivity is adjusted. The data used in this experiment is the same as in the previous experiment (see section 5.5.1).

In Figure 5.11 trend lines are shown for runs with the different guardian sensitivity values. There were 6 runs in total, the GuardianSensitivity was 0, 20, 40, 60, 80, and 100. The trend of the total guardianship was the same for all these runs and is shown as one trend line. Apparently the Police agents are not influenced by the GuardianSensitivity parameter. We had expected that a higher GuardianSensitivity would results in less crime, however this is not supported by the trend lines in Figure 5.11. The lowest crime number is caused by a GuardianSensitivity of 60. The highest crime number is caused by a GuardianSensitivity of 20. Both these results are counterintuitive. One would expect to see the most opportunity because of crime events with a GuardianSensitivity of 0.

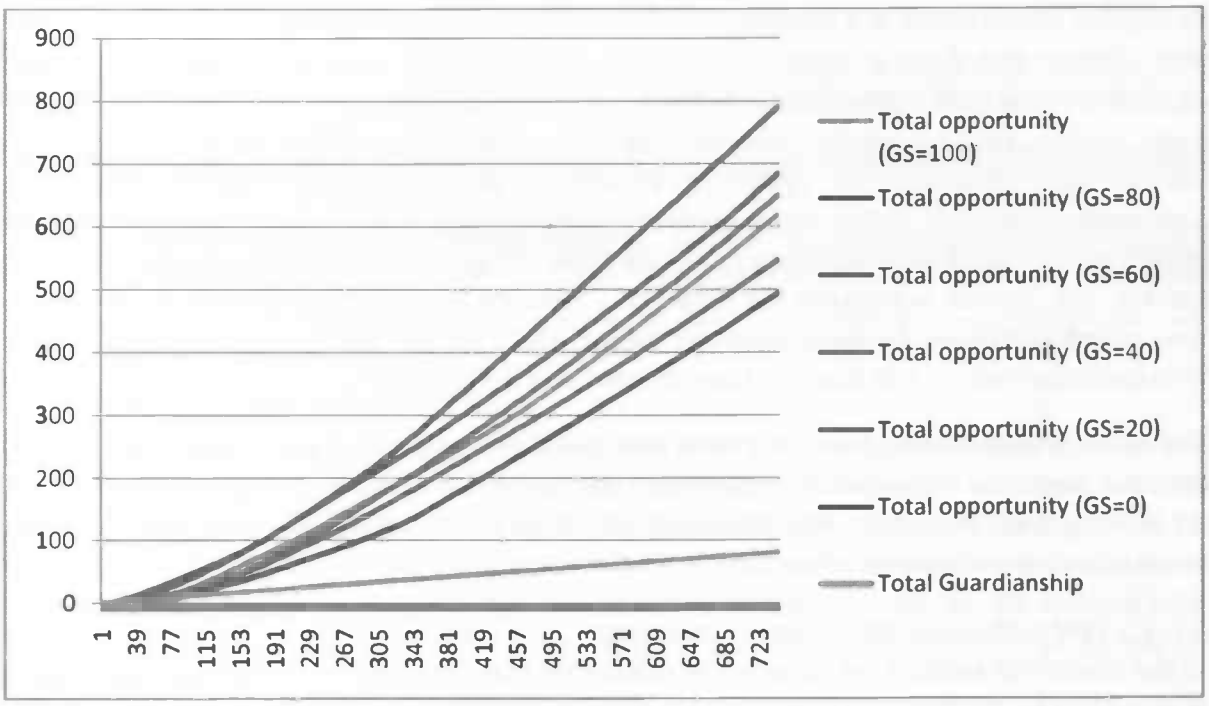


Figure 5.11 Total Opportunity for different guardian sensitivity settings . The total guardian sensitivity is the same for all runs.

When we look at the crime numbers in Table 5.10 one can see, as expected, the same situation. We can conclude higher GuardianSensitivity does not cause a higher crime number. Similarly, it can be concluded that a lower GuardianSensitivity does not cause a lower crime number. Another observation is that a GuardianSensitivity of 0 is not the same as the complete absence of Police agents. This becomes clear when one looks at the crime numbers of both settings. For a GuardianSensitivity of 0 this is 70440 and for the setting with no police this is 102441. This cannot be caused by other factors because we used the same settings over all scenarios. The reason that a GuardianSensitivity of 0 is not the same as no Police agents has to do with the attractiveness of Zipcodes for crime (see Equation 5.3). Rule 5.4 states that the guardianship is increased when a Police agent is present on a Zipcode. This means that the attractiveness is lowered. Thus although a zero GuardianSensitivity removes direct influence by Police agents (no capable guardianship) there is still an indirect influence on the decision to offend due to the guardianship value (Rule 5.6). This explains the difference between no guardians and no guardian sensitivity.

GuardianSensitivity	Correlation 10 by 10 grid	Correlation 100 by 100 grid	Crime number
100	0.6716384128392562	0.06419998860490131	63467
80	0.8569685612251631	0.16018578036199865	66694
60	0.88751468793194 3	0.17494910199904978	50552
40	0.9281732013628592	0.09461063045883507	56134
20	0.8818090356047086	0.06229963119035156	81408
0	0.938635561118479	0.04263366295793482	70440

Table 5.10 The correlations and crime numbers for different GuardianSensitivity values

The question remains why the highest GuardianSensitivity value does not cause the least crime and why the lowest values does not cause the most crime. Given Rule 5.6 Rule for a Criminal to offend on Zipcode x on time step t the chance on crime at a Zipcode should increase when there is

no capable guardianship at a Zipcode. Rule 5.7 Rule for capable guardianship on Zipcode x on time step t defines that there is capable guardianship when a Police agent is present at the current Zipcode and the GuardianSensitivity is smaller or equal to a random value. Since we used the same random seed for all the runs discussed here this cannot be the cause. The Criminals choose their locations on crime attractiveness just like Police agents. When the GuardianSensitivity is 100 Criminals will not commit when there is Police present thus the opportunity value slowly decays. In the mean while the guardianship of the Zipcode increases quickly. The Zipcode is thus not very attractive after a while and the Criminal will move to a new, more attractive location. This relocation must be the reason that the run with a GuardianSensitivity of 100 does not have the lowest crime number.

The same principle must also yield for a low GuardianSensitivity. A zero GuardianSensitivity does not mean the highest number of crimes due to a different movement pattern. Apparently 20 is an optimal setting for GuardianSensitivity from the Criminal's point of view. 60 is the worst situation for criminal behaviour.

5.6.3 Test with no initial opportunity values

In the sensitivity analysis we came to the conclusion that the initial distribution of opportunity is important for the final correlation. This is something that has not come as a surprise to us. We are more interested in what the effect is when we change the initial opportunity distribution to a uniform distribution and see if the initial distribution of the activity spaces of Criminals is something that gives us some predictive value.

After the sensitivity analysis we were interested in the influence of the initial opportunity values. Therefore we have initialized each Zipcode with an opportunity value of 1. This opportunity value thus does not have influence for the first choice for a Zipcode. However after a few runs the distribution of activity spaces is expected to have its influence on the opportunity distribution.

In Table 3.1 the correlations and crime numbers are shown for the four runs we tested with a uniform distribution of initial opportunity. The results are not very promising if the results are compared to what we found in the sensitivity analysis. For the runs shown in the table we used the parameter values found for the run with the best overall correlation in the sensitivity analysis. Only the random seed value was varied. The first run has actually the same random seed and has thus an equal parameter set. This first run shows a very low correlation on both the 10 by 10 grid and the 100 by 100 grid. The same holds for the other 3 random seeds. Only the correlations for seed 200 and 300 on the 10 by 10 grid look fair. These results are not convincing enough to conclude that the distribution of activity spaces of Criminals add something to the predictive value of the model. The cause can be that the activity spaces are not estimated well or the behaviour of the criminals within these activity spaces is too simple, or it is a combination of both.

Random seed	Correlation 10 by 10 grid	Correlation 100 by 100 grid	Crime number
400	0.39881417296374255	0.11513951787867942	23142
300	0.8983188246288093	0.026122915130092954	40380
200	0.8744776507400799	-0.0010013307340195452	61281
100	0.3368377482214786	-8.78302001774616E-4	31521

Table 5.11 Correlations and crime numbers for the runs with a uniform initial opportunity distribution

5.7 Summary and conclusions

In this chapter we have proposed a model of street robbery in the region Tilburg. The concepts of this model were derived from theory and the initialization of the model was done with real crime data. In the second half of this chapter we have tested the output of our model on the correlation with real crime data. In this investigation we found that this correlation was not very high on average, especially for the 100 by 100 grid. For the best run we showed what happened during a model run. Furthermore, we used the settings of this best run to test the model on scenarios.

To give an additional answer to the research question: "What are the possible uses of ABM for crime analysis?" we showed how the model can test three scenario's (section 5.6). The first scenario's tested the effect of the police. Our model suggested that random enforcement is just as effective as targeted police enforcement, however the results were suspicious because the correlations and crime numbers of two runs were equal. In the second scenario we tested the influence of the parameter GuardianSensitivity to see what the effect is of the amount of respect criminals have for the presence of police officers in their decision to commit crime. This scenario setting provided some interesting results. A lower GuardianSensitivity value did not necessarily mean a higher number of crime. Likewise, a higher value did not mean a lower crime number either. We think this is due to the relocation of Criminals. If this result can be translated back into the police practise is questionable, because of the low complexity of the model rules. It is nonetheless an interesting question for future models and the police practise.

In the last scenario we tested the influence of the initial distribution of opportunity by initializing the model with a uniform distribution of opportunity. We have not been able to show here that the distribution of activity spaces tells us something about the distribution of future crime. However, we have also not been able to show the opposite. Future models can try to replicate a similar method to ours and see what the relation is with the distribution of crime incidents. This does not have to be implemented in an ABM, but could probably be done in GIS software alone.

The scenario tests showed the principle of how an ABM could be used for the testing of scenarios and experimentation. The tests do not provide strong enough evidence for new theories, but they do hopefully inspire experiments with future models.

We answered the research question: "How can we evaluate our ABM?" in section 5.5.4 by using a correlation measure instead of the found techniques in chapter 2 and 3. The other evaluation techniques did not agree with the goal or did not fit with the output of the model. When we had shown that our ABM performed well on the prediction task of crime by showing high

correlations between the model's output and real crime data, we could have chosen to also use the other evaluation methods. The correlation method is certainly not ideal, because neighbouring cells do not count even though these 'near misses' could be marked as 'almost right'. Unfortunately, we are not acquainted with methods that can compute the correlation between two spatial distributions. An alternative is to create a grid with small cells and perform a smoothing operation over those cells and then compute the correlation. This, however, requires many arbitrary decisions. Besides the spatial results we have also looked at the temporal changes in the opportunity values and guardianship values.

Now we can answer the final research question: "Can we build an ABM that provides crime results similar to real crime numbers?". The answer is: not really. With some parameter settings a high correlation was found for the 10 x 10 grid, but the correlation was then a lot lower on the 100 x 100 grid. The correlations were not better than the lag 12 method, in section 3.6.2 described as "the very worst forecast and comparison method" (Gorr and Olligschlaeger 2002). Therefore, we can certainly not be positive about the predictive value of our ABM model. We should therefore consider this model as a first effort.

The central claim of this thesis is that an individual perspective and the Agent-Based Modelling methodology are fruitful for crime prediction. This claim was supported in chapter 2 by a discussion of literature on previous ABM models for crime. In this chapter we have made a contribution to this list of models. In chapter 2 and 3 we had found several concepts from crime literature that could be used in our ABM. We have chosen a subset of these concepts to keep the complete project attainable. These concepts obtained from the requirements in the previous chapters are met as follows:

- **The use of environmental data to create a realistic environment, see section 2.2 (Groff 2006; Groff 2007).**
  This requirement is met in 5.1 where is described how data of the region Tilburg is loaded.
- **The use of activity spaces in which agents are active , see section 2.2 (Groff 2006; Groff 2007).**
  This requirement is met in section 5.2.1 where the used Geographic Profiling technique to estimate activity spaces is described.
- **The use of the crime type street robbery, see section 2.1 (Liu, Wang et al. 2005) and section 2.2 (Groff 2006; Groff 2007).**
  This requirement is met in section 5.5.1 where we describe how the data of street robbery is loaded into the model.
- **The use of an opportunity surface and guardianship surface to, respectively, promote and inhibit criminal behaviour, see 2.4 (Gunderson and Brown 2000).**
  This requirement is with the use of Zipcodes, locations that have a value for opportunity and guardianship, see section 5.1.1.
- **The use of preferences of agents derived from crime data, see 2.4 (Gunderson and Brown 2000; Gunderson 2003).**
  This requirement is partly met by extracting the crime frequency and activity spaces of criminals in section 5.2.1.
- **The use of the concepts routine activity theory: guardianship, potential offender and suitable target (see section 3.1).**

- **The use of the concept of repeat victimization (see 3.3).**
  This requirement is met in 5.1.1. where a crime trace is added to a Zipcode when there was a recent crime before the start of the model.
- **The use of a Geographic profiling technique to estimate the activity space of a criminal (see section 3.4).**
- This requirement is met in section 5.2.1 where the used Geographic Profiling technique to estimate activity spaces is described.
- **The use of the defined strategy of the police (see section 3.6.3).**
  This requirement is roughly met in section 3.6.3. where the behaviour of Police agents is described.
- **The use of the minimal requirement of crime prediction (see 3.6.4).**
  This requirement is met in section 5.5.4. where the correlation between the predictions of the model and the observations of crimes is compared with the minimal expected correlation.

Despite the fact that we have not succeeded in building a proof of the claim that an individual-based modelling perspective and the use of ABM are fruitful from crime prediction. We do have shown how theories such a Geographic Profiling and Repeat Victimisation can be used in an ABM for crime. Furthermore, we have shown how data on individuals can be used to create criminals agents with different preferences. Future research can use this model as an example and might help to show that an ABM model is fruitful for crime prediction. In the next section we will discuss our research on a more global level and draw our final conclusions.

6 Final Conclusions

In this chapter we will first provide our answer to the research question formulated in the first chapter. In the second and last section suggestions for future work are discussed.

6.1 Answers on the research questions

The central claim of this thesis is that an individual perspective and the Agent-Based Modelling methodology are fruitful for crime prediction. We have tried to defend this claim by answering the research questions from section 1.1:

1. **What theories are useful for the modelling of crime in an ABM?**
   In chapter 2 and 3 we have found theories that can be used in an ABM for the modelling of crime: Routine activity theory, Rational choice theory, Geographic Profiling theory, Repeat Victimisation theory and theory on Residual Career Length. We used concepts of the routine activity theory and the rational choice theory in the design of our model. Furthermore, we used the theory on Repeat Victimisation to initialize and update the opportunity and guardianship surface. The technique to estimate the Residual Career Length was not mature enough yet to integrate into our model. Can we conclude that these theories were useful? The theories were useful in designing our model and the theories were implemented in a natural way. However, we have not been able to show that our model performed better than the minimal requirement of the lag 12 method. Therefore, we cannot conclude that the theories should be used in future models. For example, we used a simple Geographic Profiling technique because we had found in literature that more complex techniques were not better. In the future this might change and more complex techniques should be used.

2. **What type of crime is suitable to simulate with an ABM?**
   In chapter 2 we found several examples of previous models in which street robbery was modelled. Groff (2006) mentioned four reasons for the use of street robbery:

   - *[...] it is an instrumental crime and thus more likely than expressive crimes to involve a rational decision process (Clarke and Cornish 1985; Cornish and Clarke 1986; Walsh 1986).*
   - *[...] street robbery is by definition restricted to the street or some other exposed area rather than in a residence or business and thus involves the public intersection of offender and target in space and time.*
   - *[...] police presence is assumed to be more effective against street level crime then crimes that take place indoors (e.g. domestic violence).*
   - *[...] street robbery elicits a high level of fear among residents because of its suddenness and potential for serious injury and thus is of considerable interest to both law enforcement and the public (Feeney, 1986).*

   These reasons remain valid, however, we found that the number of incidents per offender was low. This resulted in a less reliable estimation of the activity spaces and crime frequencies. Most of the street robbers had committed others sort of crime as well. These other incidents could be used to make a better estimation of the activity

space and crime frequency. However, it is hard to compare a serious crime such as street robbery with a less serious crime such as shop lifting. This is a consideration that should be taken into account. Another option, probably a good one, is to derive the preferences of groups of criminals (see e.g. Gunderson and Brown 2000). Each offender found in the database can extract its preferences of the groups of preferences found. This does not mean that there have to exist groups of criminals with the same preferences. The total group of preferences could be divided in subsets of preferences to cluster on. Additionally, these preferences could be combined with the unique crime frequency and activity space when different type of crimes are used. Future research should take this consideration into account, but street robbery is still a good choice especially when there is an explicit interaction between the offender and victim, something we have not modelled (see e.g. Groff 2006). There exist however data on victims that could be modelled as well. An additional option is to model co-offending, a phenomenon that becomes clear when looking at the data; several criminals are arrested for the same crime incident. [34]

3. **What techniques are useful for the modelling of crime in an ABM?**
   As we already stated in the answer of the first research question we have not been able to show evidence of a model that performs well using certain techniques. However, the reason that we did not show that the model was successful might be an indication that, for example, the Geographic Profiling technique can be used smarter. There exist data about locations in a city, for example where groceries are. [35] We did not have this data available for our model. When this data is available, the Geographic Profile technique can be improved. Groff (2006) used this kind of data to randomly point activity spaces to criminal agents. When both techniques are combined this could lead to a better estimation of activity spaces. Furthermore the residual career length estimation (Kazemian, Blanc et al.) might be used in the future to estimate which criminals are still active and to estimate how active they are.

4. **What data is available and is useful as input for our ABM?**
   In our research we used data on crime incidents and criminal suspects to initialize the environment and to construct criminal agents in our model. Incidents were coupled to suspects to create a geographic profile and to extract the frequency for crime. Incident data on street robbery data was used to derive a distribution of activity per hour of the week. Incident data was used to determine the chance on crime geographically.

   The answer of this research question would be more useful if we change it to: "What data should be available as input for an ABM?". This changes the simple answer of 'not a lot' to a more interesting list of data sources. The list is sorted on the (expected) availability of the data:
   - Environmental data: data on the environment should be used to create a realistic model of the environment (see e.g. Groff 2006). There is no limit here what should be used and is only limited by the computational complexity. Clustering methods and criminologist can point the more interesting variables.

[34] The modeling of co-offending was suggested by Ken Pease and Shane Johnson in an email conversation.
[35] Environmental data can be obtained against payment via companies such as ESRI (www.esri.com).

- Crime incident data: crimes in the past can be used to find relations with other variables such as environmental data to find the preferences of criminals (Gunderson 2003).
- Data on victims can be used to include possible victims in the model. It is known that some people are victimised multiple times. Data on these individuals can be used to model (a part) of the victim population more realistically.
- Data on co-offending: co-offending exists in street robbery and we have actually found relations between criminals in our data set, however due to time constraints this was not implemented. It would be interesting to see in an ABM what happens when certain criminals fall out of a group. Would this change the behaviour of the whole group or would someone else take his/her place.
- Data on individual criminals: the data used in our model did not include information on possible imprisonment of criminals. We could thus have modelled criminals that were constrained by a sentence. This data does exist but is hard to get to.
- DNA data on unknown offenders: the database we had available consisted of offenders that were once arrested. For volume crime, however, this is only a small percentage of the total criminal population. There exist criminals that never get caught or too late and for only a small amount of their actual offenses. These criminals can be made visible when the DNA found at the crime scene is matched with previous found DNA on other crime scenes. Together this will give us the same information on offenders as the information we used in this research.
- Data on police enforcement: where were all the police units on time *t*. This will make empirical research possible on the effect of police enforcement, but more importantly, this can be used as input to an ABM model and can possibly be used to compare different strategies of police enforcement. The reason that this data was not available is because it does not exist. During a meeting with the Police of Amsterdam-Amstelland it became clear that the precise movement of police units is not registered. Hopefully this changes in the future now experiments take place with PDA's with GPS functionality in Groningen. These PDA's are connected with the police headquarters and can theoretically provide location information on police units.

As a final note on the use of data, we have defined several types of data that should be included into an ABM to forecast. This should be taken lightly in the sense that we do not except that next models will use all of the above described data. However, one should consider the options stated. We do believe that a model will be better when this data is used in a good way.

5. What are the possible uses of ABM for crime analysis?

In chapter 2 en 3 we found several possible uses of ABM in crime analysis:

- Melo, Belchior et al. (2005) showed that an ABM can be used to find optimal police patrolling routes (section 2.3).
- In Liu, Wang et al. (2005) was shown that an ABM can be used to replicate crime patterns (section 2.1).

- Groff (2006) showed that an ABM can be used to test a crime theory; the routine activity theory (section 2.2).
- Chainey and Ratcliffe (2005) make an important note however that a snapshot does not provide a detailed behaviour of the changing crime patterns, it only tells the analyst what has changed not how this transition has taken place. This finding provides us with an advantage of ABM that naturally can show spatio-temporal patterns (section 3.6.1).

In addition to the possible uses found in literature, we found that ABM presented in this research can be used to give insight in the criminal population. The display of our ABM shows the activity spaces of criminals, this gives an idea of where criminals are active and not only where crime is committed as with other methods. We believe that when the predictive value of the ABM model gets better – whether this is because of the use of additional data, a better modelling of behaviour, or the use of other theories or in the use of the current theories in another way – the ABM model will become more useful for several reasons. A good and robust prediction of crime and other variables give more certainty of model results when a model scenario is changed to see the effect. For example, in the future currently-not-available data on past police enforcement could be used in an ABM model that quite accurately predicts crime. If the modeller is certain that behaviour of the agents is still realistic when the scenario of the model changes this means that alternative police enforcement strategies could be tested on a real prediction, for example, to minimize victims of street robberies. The model of Melo, Belchior et al. (2005) does something similar, but in this model variable settings are based on the insight of the modeller not on how well they fit real data. Not only police enforcement strategies could be tested but also e.g. the placement of bars etcetera to see the effect on crime levels. A hypothesized possible use is thus scenario testing for real environments and real behaviour of people. Not only will this scenario testing be useful for the direct police practise also currently missing theories on police enforcement could be investigated.

In this research in section 5.6 we have shown example scenario experiments. Unfortunately, we were not able to prove something here, we did, however, illustrated the idea of such kind experiments. Future research might be able to show better examples of the scenario experiments.

Additionally to the above described usage, the individual-based crime forecasting approach might be used as an additional source of information to other crime forecasting techniques that only use incident data. Even if future individual-based crime models do not proof to be better forecasters than other methods, they might give additional insights in the origin of crime trends. For example, in Figure 5.2 is a distribution shown of geographic profiles of known offenders. Such kind of individual-based information might proof to be useful in addition to crime forecast methods that ignore data on individuals. Especially when crime needs to be forecasted for hypothetical situations as in the earlier mentioned examples, e.g. police enforcement planning but also the planning of new housing estate, individual-based modelling, and ABM in particular, is expected to be very useful in addition to other forecasting techniques.

6. **How can we evaluate our ABM model?**
   In chapter 2 and 3 we found theories to evaluate ABM for crime analysis: Groff (2007) has shown three ways of evaluation for her model: Ripley-K, Kernel Density and the ANOVA test. Furthermore, we found that the random walk or naïve lag 12 method should be used as minimal requirement for a crime prediction. Bowers, Johnson et al. (2004) showed a method to evaluate a crime prediction from the perspective of policing.

   In chapter 5 the model was evaluated with a correlation: the Pearson product-moment correlation coefficient. This coefficient was used because it shows how two variables increase and decrease together. The model might wrongly predict the exact crime number, but correctly predict an increase in crime. A correctly predicted increase is valuable for the police organization. The Pearson product-moment correlation will value a correctly predicted increase or decrease in crime. Other methods such as the Mean Squared Error (MSE) will remain low only when the absolute error is low. Furthermore, when there are many zero values, and we will have many Zipcodes that have no crime events during a model run, methods such as the MSE will be misled.

   In the conclusions of the previous chapter (section 5.7) we already concluded that there exist no good evaluation method for the comparison of two spatial distributions for our purpose.

7. **Can we build an ABM model based on individual data that provides crime results similar to real crime numbers?**
   In the previous chapter we already concluded that we have not been able to build an ABM model that provided similar crime results as real crime numbers. The predictions of our model were not convincing. However, this research has provided an example of how an ABM based on individual crime data could look like. Furthermore, we have shown how theories, especially of Geographic Profiling, can be used in an ABM.

The next question is whether or not these answers on the research questions defend the central claim of this thesis: "an individual modelling perspective and the Agent-Based Modelling methodology are fruitful for crime prediction". Based on the predictions of the model described in the previous chapter one would reject this claim. However, earlier research provided models that have proven to be useful (see chapter 2) for the testing of crime theories and the replication of crime patterns. It is possible that these kind of models should mature before one should think of using these models to predict crime. At the start of this research we searched for crime theories to build a model with the purpose of predicting crime. It might be wiser to first create a model that shows realistic crime patterns that agree with theory before it is tested on the prediction of real crime.

6.2 Suggestions for future work

In the previous section we discussed answers on the research questions defined in the first chapter. In this section suggestions for future work are made. To begin future work should look at the answers on the previously discussed research questions. Special attention should be given to the answer on what data is available and useful for our ABM. Furthermore, some possible uses of ABM for crime analysis are explained. The theories presented in this thesis (see the previous research questions) might also be useful to look at. The discussion of what crime

type to choose is also important, this should be considered carefully especially when one wants to use real crime data.

In addition to the research questions, we have also defined requirements in every chapter. Some of these are interesting to highlight for future work. For example in chapter 2 we saw several interesting concepts:

- The use of the concept of Tension, see section 2.1 (Liu, Wang et al. 2005), and section 2.3 (Melo, Belchior et al. 2005).

Tension accounts for the spreading of fear when a crime has taken place. Neighbouring areas are influenced by the event of a crime. This is an interesting model concept.

- The use of activity spaces in which agents are active , see section 2.2 (Groff 2006; Groff 2007).

Groff uses activity spaces that are based on environmental data such as locations of groceries, schools etc. This creation of activity spaces can be used with a geographic profiling techniques to better estimate routes of criminals and victims.

- The use of the concept satisfaction and ideal satisfaction for criminals, see section 2.3 (Melo, Belchior et al. 2005).

The concept of satisfaction by Melo, Belchior et al. provides a concept that could account for the temporal behaviour of criminals. With data on criminals this satisfaction can be estimated for each criminal in the manner we estimated the crime frequency of criminals (see section 5.2.1, Equation 5.2).

- The use of genetic algorithms to optimize police enforcement, see section 2.1 (Liu, Wang et al. 2005).

When a model is created with realistic behaviour of criminals, victims and police genetic algorithms can be used to optimize police enforcement in the model to see what enforcement strategies work and which do not.

- The use of preferences of agents derived from police data, see 2.4 (Gunderson and Brown 2000; Gunderson 2003).

The derivation of preferences from crime incident data with Gunderson's method can be used to model the criminals better.

In the discussed crime literature in chapter 3 we found that three concepts that have to do with the theory on Residual Career Length (see section 3.2.4):

- The use of the concept of delayed deterrence to illustrate the increasing fear of criminals (see section 3.2.3).
- The use of the idea that when criminals get older it is more likely that they quit crime (see section 3.2.3).
- The use of the idea that under the right circumstances ex-offenders tend to recommit crime (see 3.2.3).

In our model we did not account for the fact that criminals can quit crime or start again. Future models should consider modelling this for realistic behaviour of criminals. Furthermore, we did not model that criminals are caught by police agents. This is another design decision.

The last suggestion for future work is to review the use of a Geographic Profiling technique. Geographic Profiling is a new field with a lot of recent developments, this means that better techniques will be developed.

Bibliography

Amblard, F. (2002). Which Ties To Choose? A Survey Of Social Networks Models For Agent-Based Social Simulations. Proceedings of the 2002 SCS International Conference On Artificial Intelligence, Simulation and Planning in High Autonomy Systems, Lisbon, Portugal.

An, L., M. Linderman, et al. (2005). "Exploring Complexity in a Human–Environment System: An Agent-Based Spatial Model for Multidisciplinary and Multiscale Integration." Annals of the Association of American Geographers **95**(1): 54-79.

Axelrod, R. (1997). "Advancing the art of simulation in the social sciences." Complexity **3**(2): 16--22.

Axelrod, R. (2005). Handbook of Research on Nature Inspired Computing for Economy and Management. J.-P. Rennard, Hershey, PA: Idea Group Inc.

Axelrod, R. M. (1986). "An Evolutionary Approach to Norms." American Political Science Review **80**(4): 1095-1111.

Axtell, R., R. Axelrod, et al. (1996). "Aligning simulation models: A case study and results." Computational & Mathematical Organization Theory **V1**(2): 123--141.

Baxter, N., D. Collings, et al. (2003). "Agent-Based Modelling — Intelligent Customer Relationship Management." BT Technology Journal **21**(2): 126--132.

Bennett, T. and R. Wright (1984). Burglars on burglary: prevention and the offender. Aldershot, Hampshire, England; Brookfield, Vt., U.S.A., Gower.

Bian, L. (2003). "The representation of the environment in the context of individual-based modeling." Ecological Modelling **159**(2-3): 279--296.

Block, R. and C. R. Block (2000). Analyzing crime patterns: Frontiers of practice. V. Goldsmith, P. G. McGuire, J. H. Mollenkopf and T. Ross, Thousand Oaks, CA: Sage.: 137-152.

Blumstein, A., J. Cohen, et al. (1986). Criminal careers and career criminals, Washington , DC: National Academy Press.

Bonabeau, E. (2002). "Agent-based modeling: Methods and techniques for simulating human systems." PNAS **99**(3): 7280-7287.

Bosse, T., C. Gerritsen, et al. (2007). Cognitive and Social Simulation of Criminal Behaviour: the Intermittent Explosive Disorder Case. Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'07.

Bowers, K. J., S. D. Johnson, et al. (2004). "Prospective Hot-Spotting: The Future of Crime Mapping?" Br J Criminol **44**(5): 641-658.

Braly, M. (1976). Flase Starts: A Memoir of San Quentin and Other Prisons, Boston: Little, Brown.

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.

Brantingham, P. L. and P. J. Brantingham (2004). "Computer Simulation as a Tool for Environmental Criminologists." Security Journal **17**(1): 21-30.

Bratley, P., B. L. Fox, et al. (1987). A guide to simulation (2nd ed.). New York, NY, USA, Springer-Verlag New York, Inc.

Brown, D. G., R. Riolo, et al. (2005). "Spatial process and data models: Toward integration of agent-based models and GIS." Journal of Geographical Systems **7**(1): 25--47.

Brown, D. G. and D. T. Robinson (2006). "Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl." Ecology and Society **11**(1).

Brunsdon, C., G. Higgs, et al. (2005). Using Isosurfaces To Explore Geo-Temporal Patterns Of Crime.

Canter, D. (2003). Mapping murder. The secrets of geographical profiling. London, Virgin Books.

Canter, D., T. Coffey, et al. (2000). "Predicting Serial Killers' Home Base Using a Decision Support System." Journal of Quantitative Criminology **16**(4): 457--478.

Carley, K., N. Altman, et al. (2006). "Biowar: Scalable Agent-Based Model Of Bioattacks." IEEE Trans. on Systems, Man, and Cybernetics **36**: 252-265.

Chainey, S. and J. Ratcliffe (2005). GIS and Crime Mapping, West Sussex, England: Wiley.
Charania, A. and J. DePasquale (2005). Simulating the Dynamic Marketplace: An Introduction to the Nodal Economic Space Commerce (NESC) Model.
Clarke, R. and D. Cornish (1985). Modelling offenders' decisions: a framework for research and policy. M. Tonry and N. Morris. Chicago, University of Chicago Press. **6:** 225--256.
Clarke, R. V. (1992). Situational crime prevention: Successful case studies. R. V. Clarke, New York: Harrow and Heston: 3-36.
Clarke, R. V. and M. Felson, Eds. (2004). Routine Activity and Rational Choice (Advances in Criminological Theory), Transaction Publishers, New Brunswick (U.S.A).
Cohen, L. E. and M. Felson (1979). "Social Change and Crime Rate Trends: A Routine Activity Approach." American Sociological Review **44**(4): 588--608.
Cormier, B. M., M. Kennedy, et al. (1959). "The natural history of criminality and some tentative hypotheses on its abatement." The Canadian Journal of Correction **1**(4): 35-49.
Cormier, B. M., M. Kennedy, et al. (1965). Forensic Psychiatry and Child Psychiatry. E. Cameron. Boston, Little, Brown: 3-41.
Cornish, D. and R. V. Clarke, Eds. (1986). The Reasoning Criminal, New York: Springer-Verlag.
Cusson, M. and P. Pinsonneault (1986). The Reasoning Criminal. D. B. Cornish and R. V. Clarke, New York: Springer-Verlag: 72-82.
Dibble, C. (2006). Computational Laboratories for Spatial Agent-Based Models. Handbook of Computational Economics. L. Tesfatsion and K. L. Judd. The Netherlands, Amsterdam, Elsevier. **2:** 1511--1548.
Doran, J. E. and N. Gilbert (1994). Simulating Societies: The Computer Simulation of Social Phenomena. N. Gilbert and J. E. Doran. London, UCL press: 1-18.
Eck, J. E., S. Chainey, et al. (2005). Mapping Crime: Understanding Hot Spots, National Institute of Justice.
Eck, J. E. and D. Weisburd (1995). Crime and place. J. E. Eck and D. Weisburd, Monsey, NY: Criminal Justice Press: 1-33.
Epstein, J. M. (2002). "Modeling civil violence: An agent-based computational approach." PNAS **99**(90003): 7243-7250.
Epstein, J. M. and R. Axtell (1996). Growing Artificial Societies - Social Science from the Bottom, MIT Press, Cambridge, MA.
Farrell, G. (2005). The Handbook of Crime Prevention and Community Safety. N. Tilley. Cullumpton, Devon, UK, Willan Publishing: 145-172.
Farrell, G. (2006). The Handbook of Security. M. Gill, Palgrave Macmillan, London (U.K.): 179-199.
Feeney, F. (1986). The Reasoning Criminal. D. B. Cornish and R. V. Clarke, New York: Springer-Verlag: 53-71.
Ferwerda, H., H. Jonkmans, et al. (1998). "Het fenomeen straatroof onder de loep." Algemeen politieblad van het Koninkrijk der Nederlanden(20): 14-16.
Galan, J. M. and L. R. Izquierdo (2005). "Appearances Can Be Deceiving: Lessons Learned Re-Implementing Axelrod's 'Evolutionary Approach to Norms'." Journal of Artificial Societies and Social Simulation **8**(3).
Georgeff, M. P. and A. L. Lansky (1987). Reactive Reasoning and Planning. Proceedings of the Sixth National Conference on Artificial Intelligence, AAAI'87., AAAI Press.
Gilbert, N. and K. G. Troitzsch (2005). Simulation for the social scientist, Open university press, London (U.K.).
Gimblett, H. R., Ed. (2002). Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes, Oxford University Press, USA.
Glueck, S. and E. Glueck (1974). Of Delinquency and Crime: A Panorama of Years of Search and Research. Springfield, IL, Springfield, Ill: Charles C Thomas.
Gorr, W. and R. Harries (2003). "Introduction to crime forecasting." International Journal of Forecasting **19**: 551-555.

Gorr, W. and A. Olligschlaeger (2002). Crime Hot Spot Forecasting: Modeling and Comparative Evaluation, Summary, National Institute of Justice.

Grimm, V., U. Berger, et al. (2006). "A standard protocol for describing individual-based and agent-based models." Ecological Modelling **198**(1-2): 115--126.

Grimm, V. and S. F. Railsback (2005). Individual-based modeling and ecology, Princeton University Press, Princeton, N.J.

Groff, E. R. (2006). Exploring the Geography of Routine Activity Theory: A Spatio-Temporal Test Using Street Robbery, University of Maryland.

Groff, E. R. (2007). "Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery." Journal of Quantitative Criminology **23**(2): 75-103.

Groff, E. R. and N. G. L. Vigne (2002). Analysis for Crime Prevention (Crime Prevention Studies, Volume 13). N. Tilly, Criminal Justice Press, Monsey, New York, U.S.A.: 29-57.

Gunderson, L. and D. Brown (2000). Using a multi-agent model to predict both physical and cybercriminal activity.

Gunderson, L. and D. Brown (2000). Using a multi-agent model to predict both physical and cybercriminal activity. Systems, Man and Cybernetics, 2002 IEEE International Conference on, Nashville, TN, USA.

Gunderson, L. F. (2003). Using Data-Mining and Multi-Agent Simulation to Predict Criminal Behavior, the faculty of the School of Engineering and Applied Science, University of Virginia.

Hindelang, M. J., M. R. Gottfredson, et al. (1978). Victims of Personal Crime: An Empirical Foundation for a Theory of Personal Victimization, American Sociological Association.

Hirschi, T. and M. Gottfredson (1983). "Age and the explanation of crime." American Journal of Sociology **89**(3): 552-584.

Hollway, W. and T. Jefferson (2000). Crime risk and insecurity. T. Hope and R. Sparks, New York: Routledge: 31-49.

Homant, R. J. and D. B. Kennedy (1998). "Psychological Aspects of Crime Scene Profiling: Validity Research." Criminal Justice and Behavior **25**(3): 319--343.

Johnson, S. D. and K. J. Bowers (2004). "The Burglary as Clue to the Future: The Beginnings of Prospective Hot-Spotting." European Journal of Criminology **1**(2): 237-255.

Kazemian, L., M. L. Blanc, et al. Patterns of Residual Criminal Careers among a Sample of Adjudicated French-Canadian Males.

Kazemian, L. and D. P. Farrington (2006). "Exploring Residual Career Length and Residual Number of Offenses for Two Generations of Repeat Offenders." Journal of Research in Crime and Delinquency **43**(1): 89-113.

Leombruni, R. and M. Richiardi (2005). "Why are economists sceptical about agent-based simulations?" Physica A: Statistical Mechanics and its Applications **355**(1): 103--109.

Levine, N. (2004). CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.0), Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC.

Levine, N. and Associates (2000). Crimestat: A Spatial Statistics Program for the Analysis of Crime Incident Locations. Washington, DC, National Institute of Justice.

Liang, J. (2001). Simulating Crimes and Crime Patterns Using Cellular Automata and GIS, University of Cincinnati.

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.

Liu, H. and D. E. Brown (2004). "A New Point Process Transition Density Model for Space-Time Event Prediction." IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **34**(3): 310-324.

Liu, L., X. Wang, et al. (2005). Geographic Information Systems and Crime Analysis. F. Wang, Singapore: Idea Group: 197-213.

Lopez, M. J. J. (2005). "Geografische daderprofilering." Het tijdschrift voor de politie **66**(5): 31-33.

Macal, C. M. and M. J. North (2005). Tutorial on agent-based modeling and simulation. WSC '05: Proceedings of the 37th conference on Winter simulation, Winter Simulation Conference.

Macal, C. M. and M. J. North (2006). Tutorial on agent-based modeling and simulation part 2: how to model with agents. WSC '06: Proceedings of the 37th conference on Winter simulation, Winter Simulation Conference.

Melo, A., M. Belchior, et al. (2005). Analyzing Police Patrol Routes with the Simulation of the Physical Reorganization of Agents. Multi-Agent Simulation, 2005, Utrecht. 5th Multi-agent Based Simulation (MABS), Berlin : Springer Verlag.

Miller, H. J. (2005). "A Measurement Theory for Time Geography." Geographical Analysis **37**(1): 17-45.

Norris, F. H. (1997). Victims of crime. R. C. Davis, A. J. Lurigio and W. G. Skogan, Thousands Oaks, CA: Sage Publicationsm Inc.: 146-166.

North, M.J, et al. (2005). The Repast Simphony Runtime System. Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms, Chicago, USA.

North, M. J., N. T. Collier, et al. (2006). "Experiences creating three implementations of the repast agent modeling toolkit." ACM Trans. Model. Comput. Simul. **16**(1): 1--25.

North, M. J., T. R. Howe, et al. (2005). The Repast Simphony Development Environment. Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms.

Olligschlaeger, A. M. (1997). Artificial Neural Networks and Crime Mapping. Crime Mapping and Crime Prevention. D. W. a. T. McEwen. Monsey, NY, Criminal Justice Press: 313-348.

Peuquet, D. J. (1999). Geographical information systems (2nd ed.). P. A. Longley, M. F. Goodchild, D. J. Maquire and D. W. Rhind, New York: Wiley: 91-103.

Polvi, N., T. Looman, et al. (1991). "The Time Course Of Repeat Burglary Victimization." Br J Criminol **31**(4): 411-414.

Railsback, S. F., S. L. Lytinen, et al. (2006). "Agent-based Simulation Platforms: Review and Development Recommendations." Simulation **82**(9): 609-623.

Rao, A. S. and M. P. Georgeff (1991). Modelling Rational Agents within a BDI-architecture. Proc. 2nd Int. Conf. on Principles of Knowledge Representation and Reasoning, (KR'91). Morgan Kaufmann.

Ratcliffe, J. H. and M. J. McCullagh (1998). "Identifying Repeat Victimization With GIS." Br J Criminol **38**(4): 651-662.

Reis, D., A. Melo, et al. (2006). Towards Optimal Police Patrol Routes with Genetic Algorithms. International Conference on Security Informatics, 2006, San Diego. Proc. of the 3rd ISI 2006, Springer Verlag, Berlin (Germany).

Richiardi, M., R. Leombruni, et al. (2006). "A Common Protocol for Agent-Based Social Simulation." Journal of Artificial Societies and Social Simulation **9**(1): (online journal).

Rossmo, D. K. (2000). Geographic profiling, Boca Raton, Fla, CRC Press.

Rpaa (2007). A meeting with police (GIS-) experts and local intelligence officers from Regiopolitie Amsterdam-Amstelland (the police organisation for the region Amsterdam-Amstelland).

Samuelson, D. A. and C. M. Macal (2006). "Agent-Based Simulation Comes of Age." OR/MS Today **33**(4): 34-38.

Shover, N. (1983). "The Later Stages of Ordinary Property Offender Careers." Social Problems **31**(2): 208-218.

Snook, B., D. Canter, et al. (2002). "Predicting the home location of serial offenders: a preliminary comparison of the accuracy of human judges with a geographic profiling system." Behavioral Sciences & the Law **20**(1-2): 109--118.

Snook, B., P. J. Taylor, et al. (2004). "Geographic profiling: the fast, frugal, and accurate way." Applied Cognitive Psychology **18**(1): 105--121.

Snook, B., M. Zito, et al. (2005). "On the Complexity and Accuracy of Geographic Profiling Strategies." Journal of Quantitative Criminology **21**(1): 1--26.

Sparks, Genn, et al. (1977). Surveying crime, Cardiff University.

Spelman, W. (1994). Criminal incapacitation, New York: Plenum.
Stangeland, P. (2005). "Catching a Serial Rapist: Hits and Misses in Criminal Profiling." Police Practice and Research **6**(5): 453--469.
Sutherland, E. H., Ed. (1937). The Professional Thief, University of Chicago Press, Chicago.
Tobias, R. and C. Hofmann (2004). "Evaluation of free Java-libraries for social-scientific agent based simulation." Journal of Artificial Societies and Social Simulation **7**(1).
Torrens, P. M. (2003). Advanced Spatial Analysis. M. Batty, Redlands, CA: ESRI Press.
Townsley, M. and K. Pease (2002). "How efficiently can we target prolific offenders?" International Journal of Police Science & Management **4**: 323-331.
van Baal, P. (2004). Computer Simulations of Criminal Deterrence: from Public Policy to Local Interaction to Individual Behaviour, BJU Boom Juridische uitgevers: Den Haag, Netherlands.
van Dierendonck, A. (1975). "Inferential simulation: hypothesis-testing by computer simulation." Nederlands Tijdschrift voor de Psychologie **30**: 677--700.
Walsh, A. and L. Ellis, Eds. (2003). Biosocial Criminology, New York: Nova.
Walsh, D. (1986). The Reasoning Criminal. D. B. Cornish and R. V. Clarke, New York: Springer-Verlag: 39-52.
Wooldridge, M. (2002). An Introduction to MultiAgent Systems, John Wiley & Sons, Inc., New York (U.S.A).
Wright, R. T. and S. H. Decker (1997). Armed robbers in action : stickups and street culture, Boston, Mass : Northeastern University Press.
Zeigler, B. P. (1985). Theory of Modelling and Simulation. Malabar, Krieger.

Appendix: Manual Agent Model

Contents

Introduction

This document is meant as a manual for the agent model developed in the GeoPredict project at Sentient. The purpose of this model is to simulate the behaviour of real offenders to give a prediction for future crime rates. This is the first version of the model. The theoretical concepts behind this model can be found in the thesis that belongs to this model. In the thesis is explained how the behaviour of the agents work. Furthermore, experiments with scenario's are described in the thesis to show the value of the model.

The purpose of this manual is to show how to get the model up and running. We will therefore first discuss what the requirements are and how to start a first demo (section 'Requirements' and 'Quick start'). A more advanced usage is discussed in 'Installation steps for advanced usage'. In this section will be discussed how to create the different settings files and what all the parameters mean. The data export of GIS files from MapInfo and the crime data export from DataDetective are also discussed. In the section 'Experimentation' is explained how the model can be used for experimentation. Section 'Further Development' explains how the model can be extended. There are some auxiliary programs that have been developed during the project. They are described in section 'Auxiliary programs'. The libraries used in the model are shortly described in the sections 'Links' with links to the websites.

Requirements

Java 1.5 or 1.6 should be installed. Java 1.6 is more efficient than 1.5. This can be downloaded from the Sun Website (http://java.sun.com).

Quick start

A demo of the model, without using crime data, can be started by unzipping the archive into a random directory and clicking on the model.bat file. This only works under Windows. Under Linux (and in Windows too) the model can be executed by the following command:

Open a terminal and point to the installation directory. Type the following command:

```
java -jar model.jar -pf=parameters.txt
```

The first argument is the jar file of the model and the second is argument is the parameter file. More about this file will be said later. Go to the directory which contains the data settings file. To start load the demo_single_agents.xml file in the installation directory.

Installation steps for advanced usage

1. Unzip the zip archive to a location of choice.
2. Create the settings files.
3. Export crime data from DataDetective into the correct directory.
4. Creating GIS files.
5. Create a parameter file.
6. Run the model.

Step 1 is self explanatory. The other steps will be explained in the next sections. Step 6 was already explained in the previous step.

Create the settings files

The model reads several files in before starting the model. Two files are in the root directory of the model: ModelSettings.xml en AgentSettings.xml. Only the settings read from AgentSettings.xml can be adjusted on run time too. These settings files contain non-model specific settings. The one exception is the *startDate*, which is set in ModelSettings.xml instead of the parameters file.

Additionally to the parameter, model and agent settings, there are also settings files needed for the GIS and crime data to be loaded. The setting file for the crime data is requested on the start up of the model by an 'Open File dialog', the path to the settings file of the GIS file is a field in the crime data settings file. All these setting files will be discussed below. Examples of these files can be found in the archive belonging to this manual.

The parameter file set the initial value of the model parameters. In the case of a batch run this is the only way to set the parameters. The parameters are clearly described in my thesis; I will only mention them briefly below. Examples of parameter files are given in the model directory. To edit these look at the tutorial of Repast on http://repast.sourceforge.net/how-to/params.html .

Parameters:

Variable Name	Description
OpportunityDecay	The decay of opportunity per step (formula in thesis).
GuardianshipDecay	The decay of guardianship per step (formula in thesis).
GuardianTrace	The trace of guardianship a police agent leaves each step.
GuardianSensitivity	The sensitivity of the presence of a police agent on the choice of crime (0-100).
AgentsChooseZipcodes	Boolean value for whether or not to commit crime.
NumberPoliceAgents	The number of police agents.
NumberActiveHoursCriminals	The number of hours a day a criminal is active.
CrimeFrequencyConstant	Used to multiply the crime frequency of a criminal.
MinimumCrimeAttractiveness	The minimal attractiveness.
RngSeed	The seed for the random number generator.
minimumRadiusActivitySpaceInKM	The minimum radius of the activity space of a criminal.
NumberOfHours	The length of a model run in the simulated number of hours.
IntervalResultsInHours	The interval of writing results in model hours. (-1 means none).

ModelSettings.xml:

Variable name	Description
mapCenterLat	The latitude coordinate on which the OpenMap GIS display should be centred.
mapCenterLon	The longitude coordinate on which the OpenMap GIS display should be centred.
mapScale	The zoom factor (1: scale) of the OpenMap GIS display.
modelName	The name of the model.
delayPerStep	A delay in milliseconds can be set when running the model in visual mode.

millisecondsPerStep	The number of milliseconds that pass by in one step. Default is one hour (3600000 milliseconds).
startDate	The start date of the model (model parameter).
nameZipcodeLayer	The name of the zip code layer.
namePoliceLayer	The name of the police layer.
nameAgentLayer	The name of the agent layer.
crimeType	The description of the crime type (should be the same as in DataDetective).
nameActivitySpaceLayer	The name of the activity space layer.
nameAntecedentLayer	The name of the antecedents layer.
dataDir	The directory that contains the setting files. This can be relative to the models root directory or absolute.

AgentSettings.xml:

Variable name	Description
radiusAgentGraphicInKM	The radius in Kilometres of the graphic that represents a criminal.
agentTransparancy	The transparency of the colour of a criminal (0-255).
antecedentsShown	Boolean of whether or not to show the past crimes of a criminal.

Crime data xml file (no obligatory name):

Variable Name	Description
personSelectionFilename	Input xml file with the selected criminals (in input dir).
incidentSelectionFilename	Input xml file with the selected incidents (in input dir).
incidentPersonFilename	Input xml file with the relation between incidents and persons (in input dir).
incidentsLastTwoMonthsFilename	Input xml file that holds the incidents of the last two months (in input dir). Is only not directly loaded in the model.
recentIncidentsPerZipcodeFilename	The xml file that contains the recent incidents per zip code.
basePath	Base path of all data.
inputDir	Input dir, relative to selectionSpecificPath.
xlsDir	Directory with the XSL that converts the original xml files from DataDetective.
outputDir	Output dir, containing all the model results. Will be combined with a timestamp. Can be absolute or relative.
selectionSpecificPath	relative dir to basePath of this specific selection.
targetFilename	Path to the xml file on which the criminals are based (indirectly derived from DataDetective).
description	Description of the dataset.
shapefilePath	The location of the xml file that contains path information about the shapefiles.
frequencyFilePath	The XLS-file that contains the number of crimes in the past per zip code (derived from DataDetective).
temporalProfileFilePath	The XLS- file with the temporal profile of the robbers (derived from DataDetective).

temporalProfileStartRow	The row number on which the data starts.

The GIS settings file (no obligatory name) is best described in its original XML format:

```
<gisfiles dir="data/gis/">
 <types>
  <type dir="SHP">shp</type>
 </types>
 <files>
   <zipcodeFile type="shp "  filename="zipcodes.shp">
    <name>Zipcodes</name>
    <description>Zipcodes </description>
   </zipcodeFile>
   <file type="shp" filename="additional" order="0">
    <name> additional </name>
    <description/>
   </file>
    </files>
</gisfiles>
```

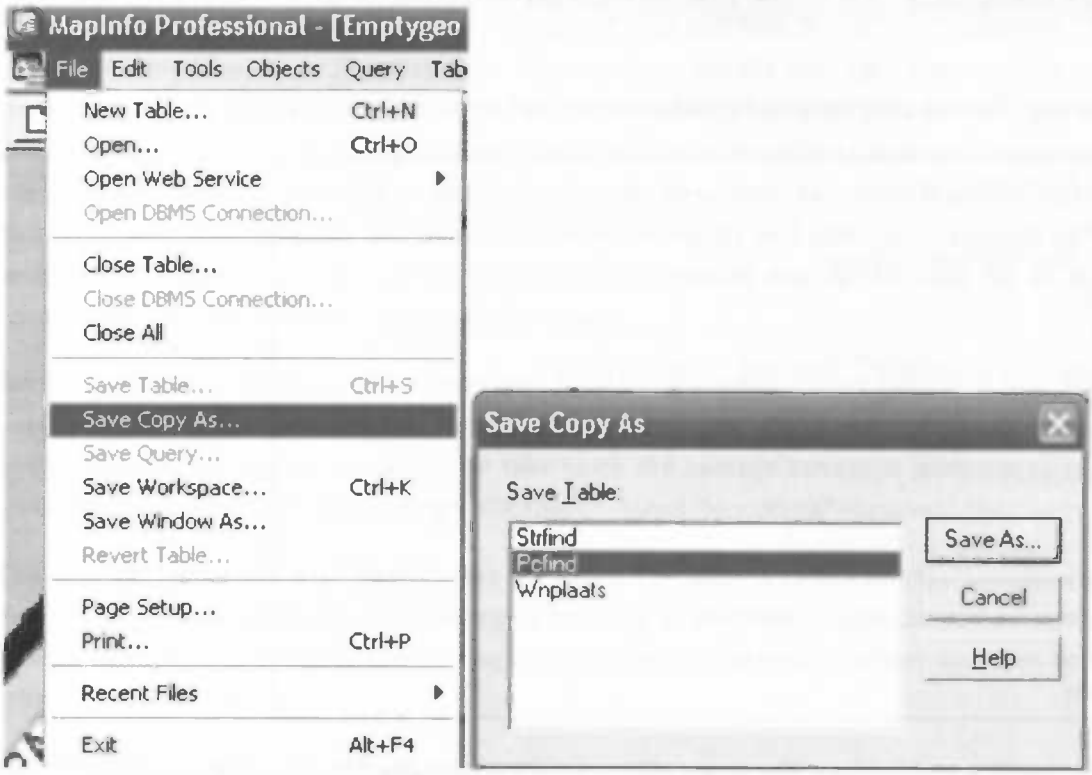
The red elements are obligatory, the green elements are optionally. The dir attribute should be the directory that contains the GIS files (either relative from model root or absolute). The attribute dir in the type node sets the subdirectory for specific types. In our case we only use shapefiles (shp) thus the files are located in "./data/gis/SHP/..." The zipcodeFile node is obligatory and contains the shapefile with zip code locations. The Zipcode agents in the model are based on this file. An arbitrary number of additional shapefile files can be added to make the GIS display more attractive. The integer value of order should start from 0 and increment with 1 if a shapefile is added. If there are five additional shapefiles the values 0, 1, 2, 3, 4 should be distributed among these nodes. Note that the more and the greater the shapefiles the slower the model will be.

GIS File export from MapInfo and GeoStreets

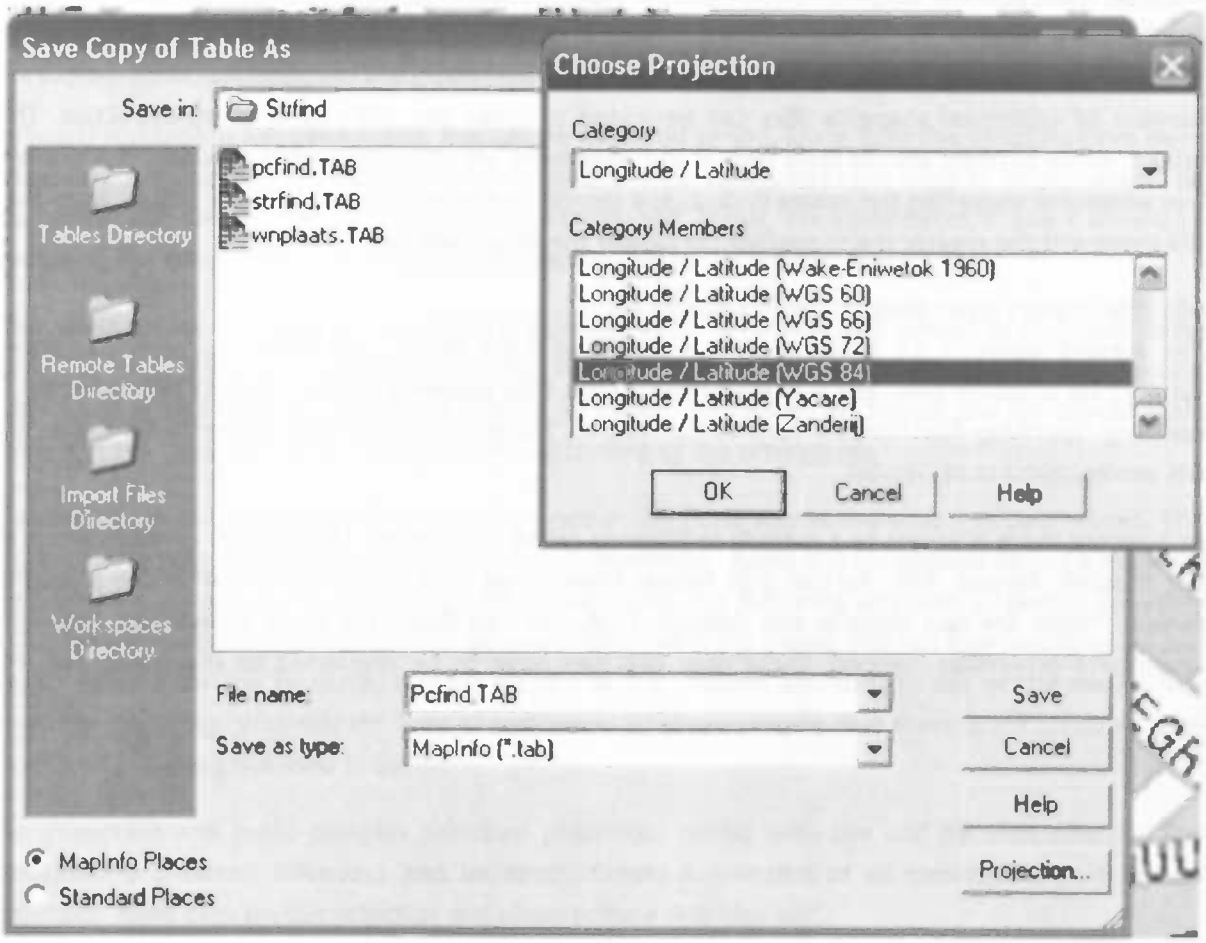
The current Repast J 3.1 library uses the OpenMap 4.6.3 library. The latter currently has some support for MapInfo files, but it is easier to use the ESRI shapefile format. Repast only provides a coupling OpenMap functionality for this file format. Therefore, I used ESRI shapefile to create the GIS environment in my model.

At Sentient there is access to a module in MapInfo called GeoStreets. This module contains maps in the MapInfo format .TAB. To use it in Repast these maps have to be converted to the shapefile format. There are two steps in this process. First, the TAB files have to be saved with the right coordinate projection. Second, these new TAB files have to be converted to shapefiles. For the

required shapefile, the one with zip codes the layer PCFind should be selected:

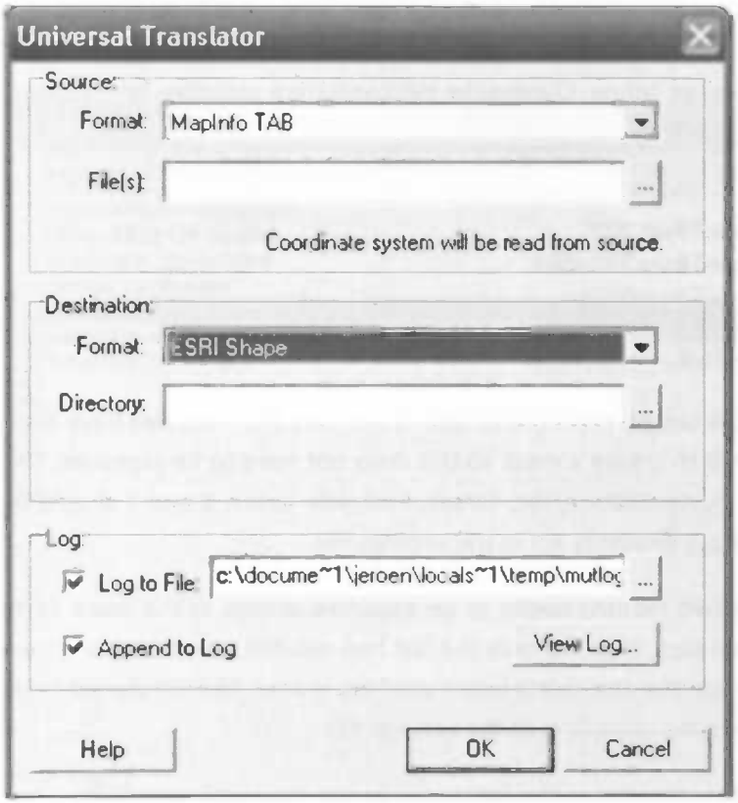


The right projection (Longitude/Latitude (WGS 84) should be chosen before the file is saved:



When the TAB file with Longitude/Latitude (WGS 84) projection is created, the file can be translated to an ESRI shapefile with the universal translator.

Choose the just created TAB file and export it to the ESRI shape format into the correct directory.



If desired, these steps can be repeated for other types of shapes that can serve as background such as streets etc..

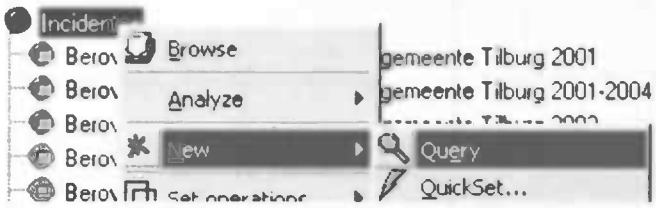
Additionally, for efficiency reasons and also for modelling reasons, the exported shapefiles have to be limited to the area of Tilburg. The coordinates for Tilburg have been extracted from the crime data itself. I have used the FindCoordinateExtentsFromIncidents.java program (described later on) to find extreme values for the coordinates. Next, I have used the GISFeatures class to create a new shapefile with these coordinate extents.

Data export from DataDetective

For a model to run on crime data we need to put four xml files in the input directory. These files need to be exported from DataDetective. We need an xml file that contains the data set of past crimes (previously described as *incidentSelectionFilename*), an xml file that contains the data set of crimes of the last two months (previously described as *incidentsLastTwoMonthsFilename*), an xml file that contains the selected criminals (previously described as *personSelectionFilename*), and finally the xml file that holds the relation between incidents and persons (previously described as *incidentPersonFilename*). This last file is not selection specific and can be used for all models if it contains all possible relations found in the database.

Additionally to the xml files, some other files have to be exported from DataDetective. We need to export a file that contains the frequency of crime per zip code (previously described as *frequencyFilePath*) and we need the file that holds the average temporal preferences of criminals (previously described as *temporalProfileFilePath*).

The xml file for selection of past crimes (of which the path is called *incidentSelectionFilename* is derived in the following way. First we need to make a selection in data detective: right-click and choose the option as in the following image below. Choose the right selection variables here (region, crime type, period etc.).

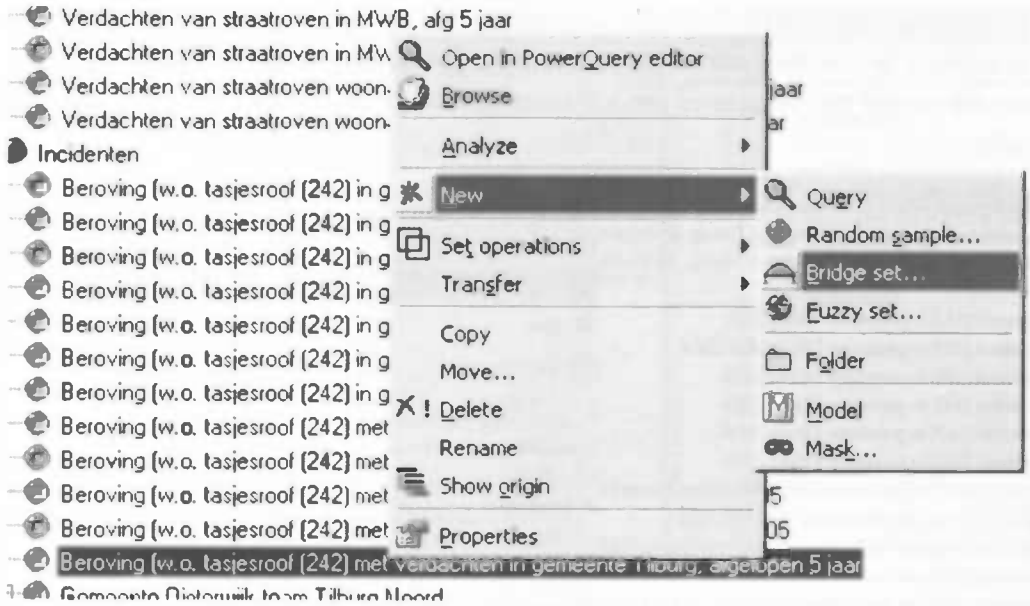


When the selection has been made, this can be exported to xml. Make sure the variables have been exported. A convenient way to do this is to create a mask so this does not have to be repeated. The variables GIDS-sleutel, Incidentsoort, Incidentdatum/tijd, Straat, Postcode-beter, X and Y should be included. Put the exported file in the input directory set in the settings file.

The selection for incidents of the last two months needs to be exported almost in the exact same way. The only difference is the selection step, in which only the last two months need to be selected. This could as well as be the last three months, but this is how I used my model. The file should again be put in the input directory with a filename according to the settings file.

The xml file with the selected criminals is created in the same way as the export of past crimes. The criminals are selected from a different datamart 'Personen'. Again it is convenient to use a mask for the selected variables to be exported. The model needs the following selection of person variables GIDS-sleutel, Geslacht, Leeftijdsklasse, Geboorteland and Woonplaats. The last four variables are not really needed for the research, but are helpful in the analysis and further use of the model. For example, by using 'Woonplaats' (city of residence) I became aware that there were criminals that operated in Tilburg but lived in Breda.

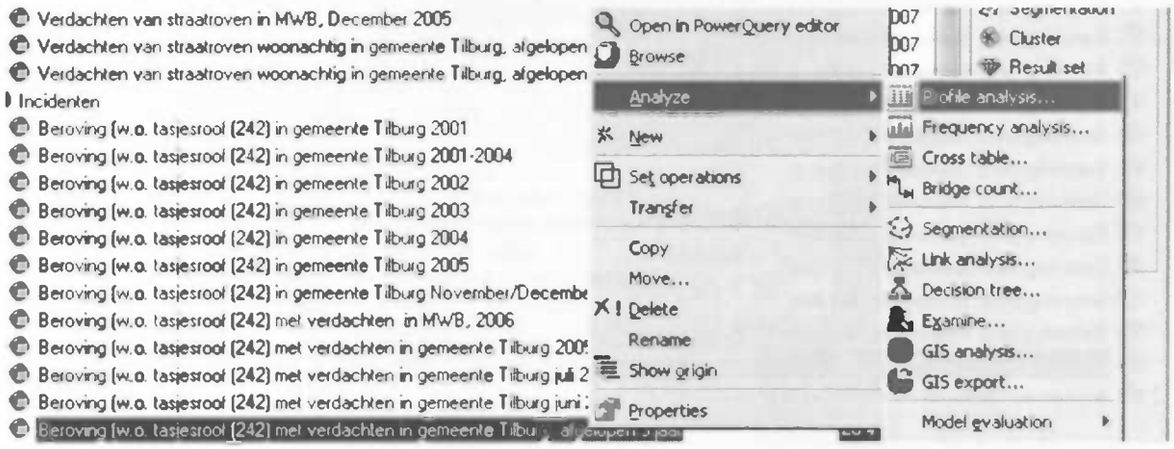
An important and more complex selection procedure comes with the xml file that contains the relationship between offenders and incidents. Create a selection of all incidents with a known offender. Right-click on this selection and choose "New -> Bridge set"



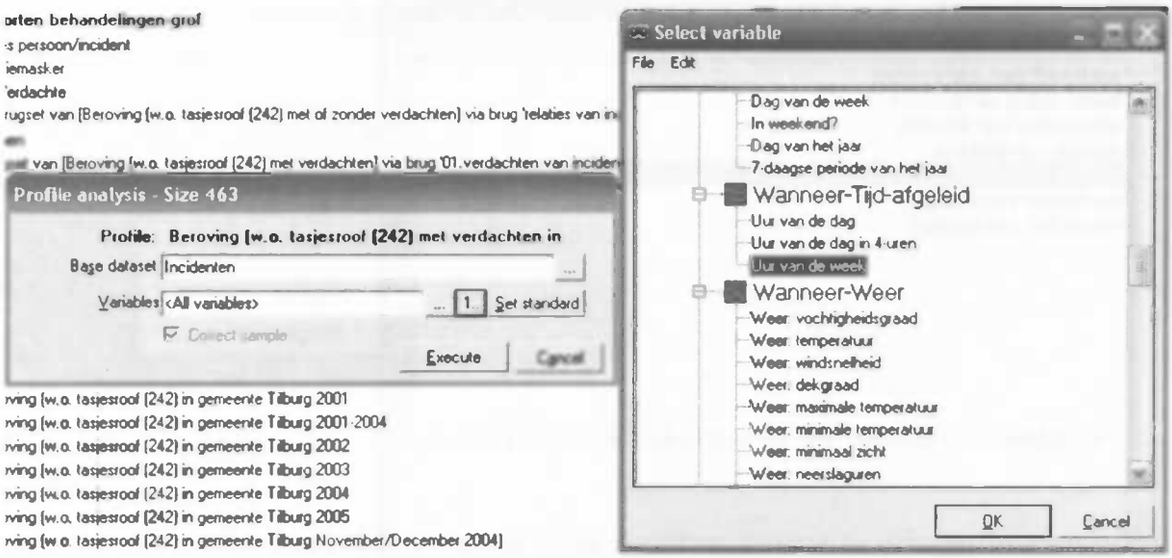
Choose relations of incident:

This bridge set will be saved under a specified name under the datamart 'Relaties persoon/incident'. This data selection should then be exported in the same way this was done with the incident selections. Export this selection to the location as specified in the data settings file.

The file with the temporal preferences of criminals (previously described by *temporalProfileFilePath*) is obtained by creating a profile analysis on the hour of the week. This could also be done in the same way as we will create the frequency per zip code file in the next step. However, this was implemented earlier.

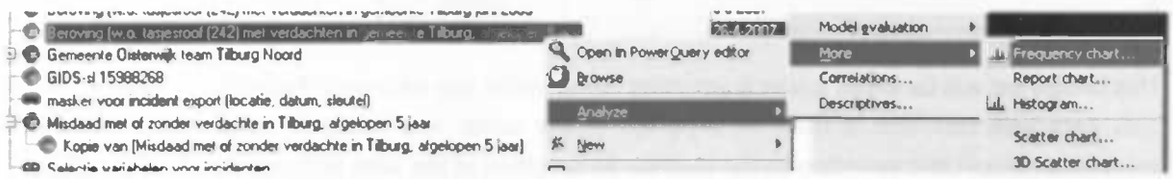


The selected variable should be '<uur van de week>' selected as follows:

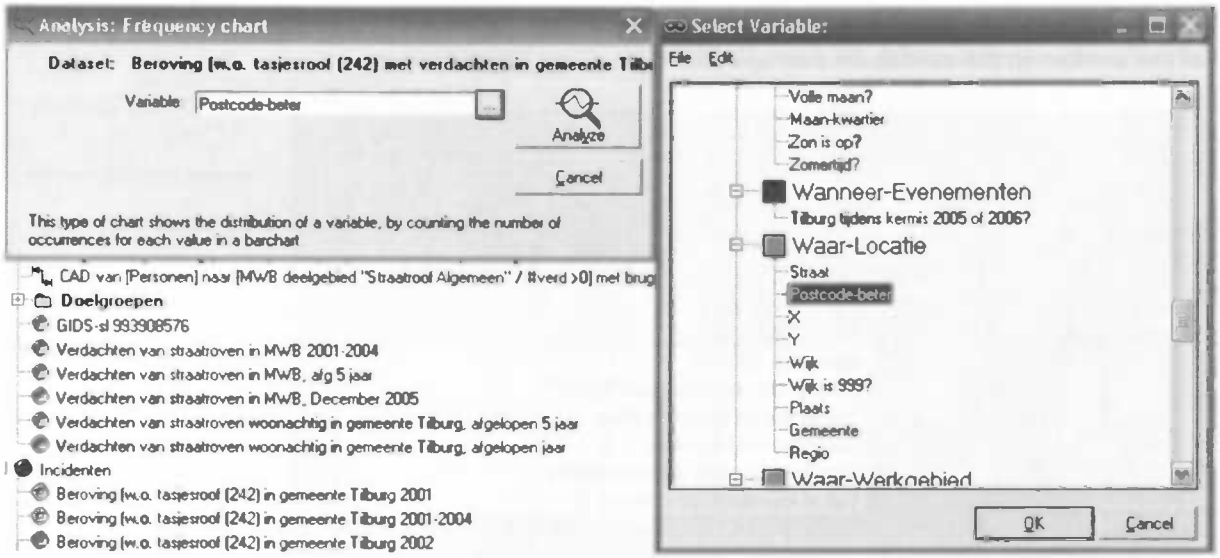


When the Button execute is pressed this will cause a window to show up with the requested values. The shown values should then be copied into Excel and saved as xls in the correct directory as set in the settings file.

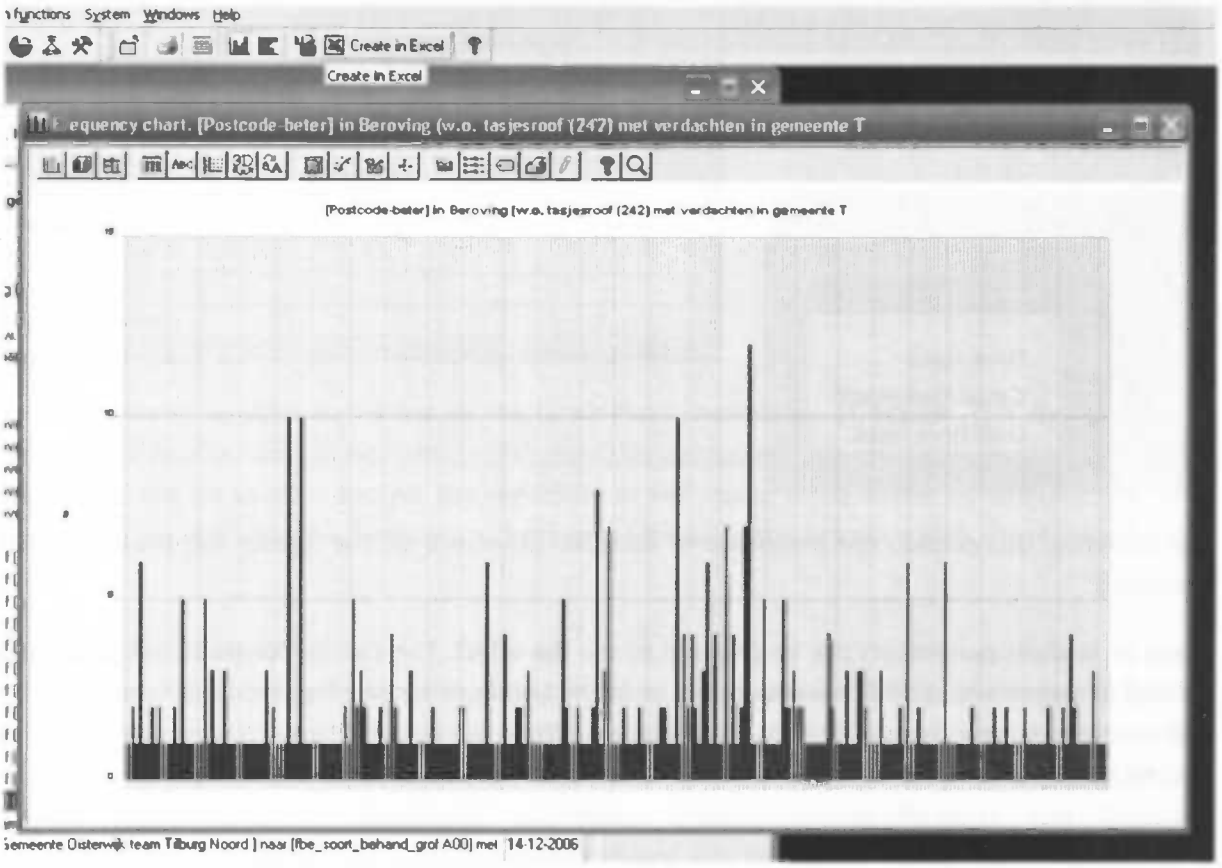
The frequency file of crime per zip code (previously described by *frequencyFilePath*) can be obtained by right-clicking on the selected incidents, then choose 'Analyze -> More -> Frequency chart':



The variable 'Postcode-beter' should be chosen in the following window:



The following window then shows up:

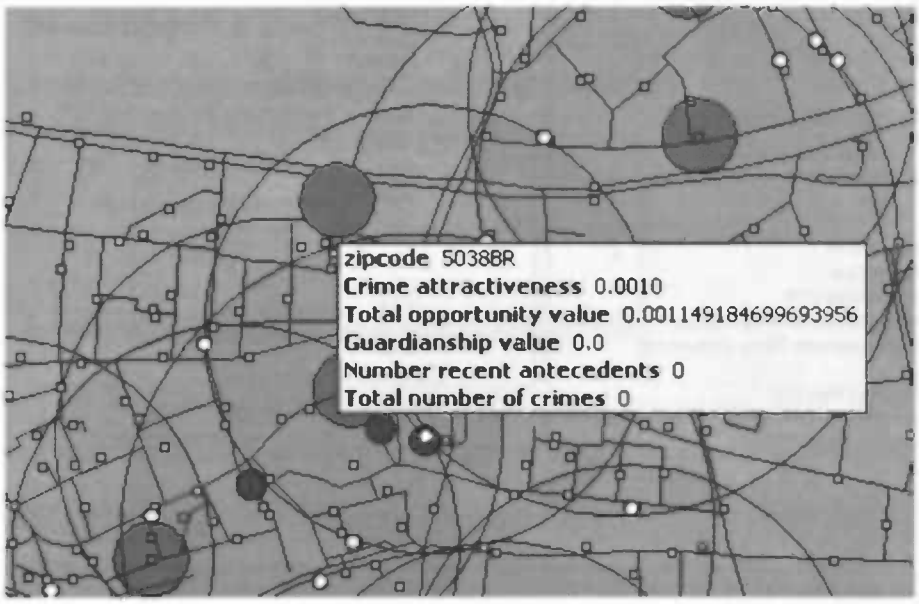


Click on 'Create in Excel' to create the excel file that contains the frequencies per zip code and save as xls file to the location as defined in the settings file. The first row is empty, this row should be deleted.

Experimentation

Experimentation can be done in two ways. One can visually experiment with the model and Monte Carlo like simulation runs can be done. We will explain the first option here first. Graphs show the crime rates and the GIS display shows the behaviour of the criminals and the police. The properties

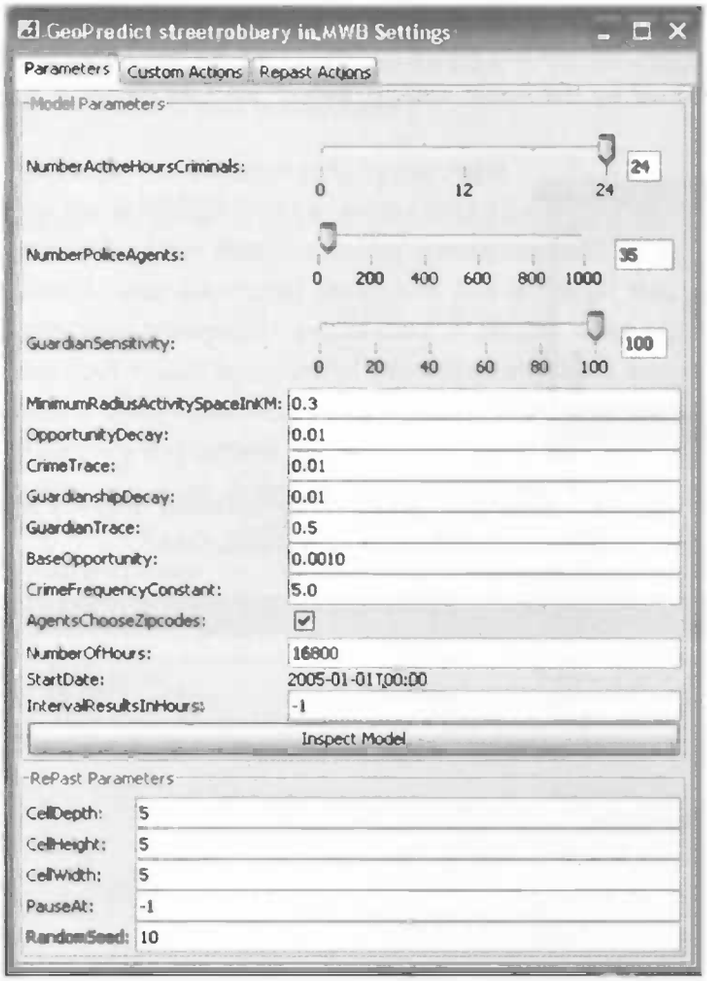
of Criminals, Police Agents and Locations can be probed. This is done by moving the mouse over one of the entities in the model. An example is shown in the figure below:



Another way to see the same properties is to right-click on one of the entities and click on the properties option.

By analysing the graphs, the environment and the behaviour of the agents the model can be analysed.

Base on analysis parameters can be changed to see the effect. This can be done before starting the model in the parameter file 'parameter.txt' or in the GUI itself by pausing the model and changing the model parameters as shown in the figure below. When the model is restarted these new settings will be used.



Another way of analysis is the use of the batch run functionality of Repast. This functionality is similar to a Monte Carlo simulation (http://en.wikipedia.org/wiki/Monte_Carlo_method). This way of analysis can be used to analyse the sensitivity of the model to its model parameters. We have used this analysis to find the parameters that caused the highest correlation with the recorded crime data.

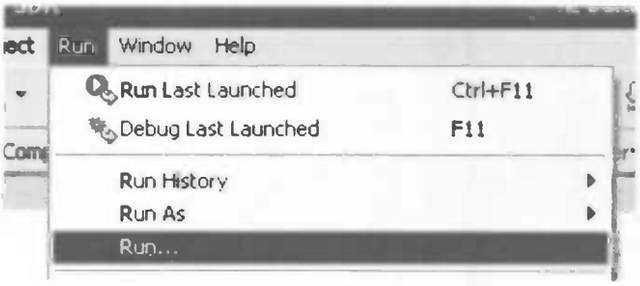
Further Development

In my thesis many possible improvements to the model are mentioned. Before one begins to further develop this model it is recommended to thoroughly study the thesis for details on the implemented concepts. Furthermore, the model is implemented using several libraries. The most important are Repast (http://repast.sourceforge.net), OpenMap (http://openmap.bbn.com) and Geotools (http://geotools.codehaus.org). Theses libraries should also be studied to see the limitations. It is very likely that improved versions of these or other libraries exist in the near future.

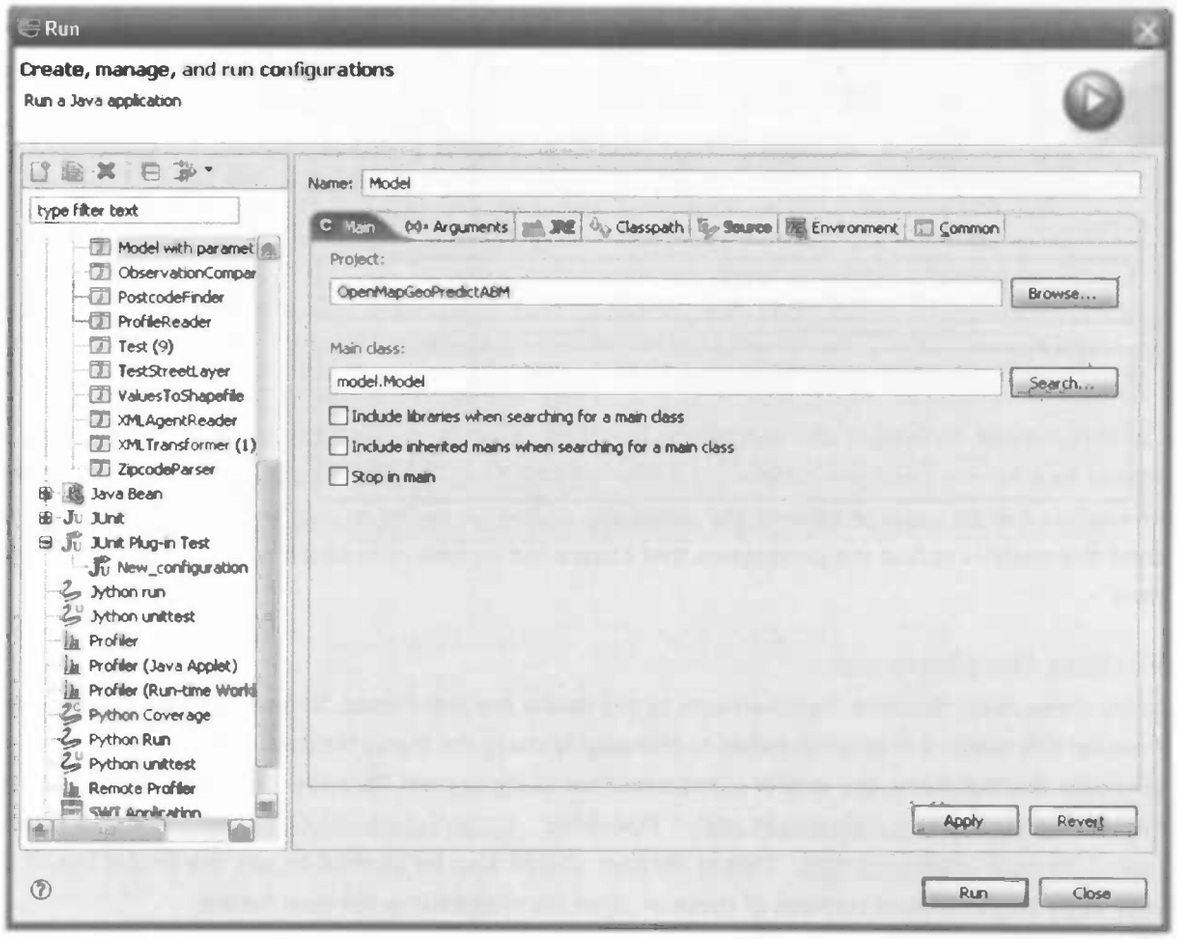
If one comes to the conclusion that the current model is useful, either as example or as basis, for a future model, here are some recommendations for further development.

The most convenient way for further development is by loading the source code into the IDE Eclipse (http://www.eclipse.org/). I used this program to develop the model.

The libraries needed can be found in library directory 'lib' in the model directory (check if newer editions of these libraries exist). When these libraries are added to the Eclipse Project[36] the model can be run from within Eclipse. This is done by the menu option "Run -> Run" as shown in the following figure.



The class Model contains the main method that should be called. In the following image a run configuration is shown. Note that our Eclipse project is called 'OpenMapGeoPredictABM'



When all things have gone well, clicking on run will cause the model to run.

[36] How this can be done is described here: https://eclipse-tutorial.dev.java.net/eclipse-tutorial/part1.html. It is convenient to add the source code and Javadoc to the libraries as well'.

Auxiliary programs

For the analysis and data preparation I have created several auxiliary programs. They are summarized in next two sections.

Extensible stylesheet (XSL) programs

For the preparation of the model data I have created several XSL programs. DataDetective has an export function that can export a data selection to XML. This XML data was not suited to load directly into the model because it had a format that was hard to read and more importantly additional information was needed. In section 'Data export from DataDetective' we have described the used export functions of DataDetective; these export functions are limited to the selection of data within a datamart[37]. For my model I had to couple incidents to criminals to estimate their crime frequency and activity space (see my thesis for more information on these concepts). In the table below the used programs are summarized. These programs should be run in this order (also represent in the number of their filenames). The model class 'ConfigurationLoader' runs the XSL programs when the required data is not available. Another option is to run these programs with msxsl.exe and a .bat file.

XSL program	Description
0_transform_incidents.xsl	Transforms xml data export from the incident datamart.
0_transform_persons.xsl	Transforms xml data export from the person datamart.
1_map_incidents_to_zipcodes.xsl	Sorts incidents per zip code.
1_map_persons_to_incidents.xsl	Couples criminals Id's to incident Id's.
2_couple_persons_to_incidents.xsl	Adds the id's of criminal's past crimes to a criminal.
3_extend_incidents_of_persons.xsl	Extends the incident information created in the previous step with extra information.
4_extend_with_cooffenders_of_persons.xsl	Couples the co-offenders id's to the offenders data created in the previous step.

Note that we have not used the data created by the last program. Although co-offending is an interesting aspect to model it was too much work to include this in my research. See the discussion in the thesis for more details.

Java programs

For the analysis of my model I have created several java classes that helped me to analyse the output of my model. The easiest way to run these classes is from within a created Eclipse project, because the classes use the same libraries as the model.

A summary of these classes is given in the table below.

Class name	Description
CreateFilesObservedCrimes.java	Creates a summary XLS files and grid files of crime per day, week, month and year.
FindCoordinateExtentsFromIncidents.java	Finds the minimum and maximum coordinates of an xml

[37] Explanation can be found here: http://en.wikipedia.org/wiki/Data_mart.

	file with incidents.
GridComparator.java	Compares two equally sized grids from two text files.
ModelRunComparator.java	Compares grids from model runs with grids of observed crimes and computes e.g. the Pearson correlation.
ObservationComparator.java	Compares grids between observed crimes and computes e.g. the Pearson correlation.
ValuesToShapefile	Creates a shapefile from an XLS file with crimes per zip code.
GISFeatures	Program to crop out the required region from a shapefile. This class has also several other functionalities used by other classes.

Links

Eclipse	Eclipse is an open-source software framework written primarily in Java. In its default form it is a Java Integrated Development Environment (IDE),
Repast	The Recursive Porous Agent Simulation Toolkit (Repast) is an agent-based modeling toolkit. Repast has multiple implementations in several languages and built-in adaptive features such as genetic algorithms and regression. Repast is free and open source.
OpenMap	BBN Technologies' OpenMap TM package is an Open Source JavaBeans TM based programmer's toolkit. Using OpenMap, you can quickly build applications and applets that access data from legacy databases and applications. OpenMap provides the means to allow users to see and manipulate geospatial information.
Geotools	GeoTools is an open source (LGPL) Java code library which provides standards compliant methods for the manipulation of geospatial data, for example to implement Geographic Information Systems (GIS).
CrimeStat	*CrimeStat* is a spatial statistics program for the analysis of crime incident locations.
Msxsl	The msxsl.exe command line utility enables you to perform command line Extensible Stylesheet Language (XSL) transformations using the Microsoft® XSL processor.