Forecasting Crime Events Applying Risk Terrain Modeling –

An Implementation and Evaluation of the
Risk Terrain Modeling-Approach
Using the Example of the City of Salzburg, Austria

by

Milena Kocher

2nd Bachelor Thesis

Submitted in partial fulfillment of the requirement of the degree
Bachelor of Science in Engineering

Carinthia University of Applied Science
School of Geoinformation and Environmental Technologies

Supervisor: Dr. Michael Leitner
Department of Geography and Anthropology, Louisiana State University,
Baton Rouge, USA

Baton Rouge, June 2014

Science Pledge

By my signature below, I certify that my thesis is entirely the result of my own work. I have cited all sources I have used in my thesis and I have always indicated their origin.

Villach, 06/01/2014 Milena Koder

Place, Date Signature

Acknowledgments

This thesis has benefited greatly from the support of many people, some of whom I would sincerely like to thank here. It was this research project which contributed to my increasing and deep interest into the field of crime analysis and prediction.

First of all, my deepest gratitude is to my supervisor Dr. Michael Leitner, Department of Geography and Anthropology, Louisiana State University in Baton Rouge, Louisiana, USA, who not only gave me the opportunity to complete my internship at the Louisiana State University but also to write my thesis within the context of this interesting project. Your detailed and constructive comments and your excellent advice guided me in all stages of this thesis.

My special thanks goes to Dr. Joel Caplan, Rutgers School of Criminal Justice, Newark, USA, whose support regarding specific questions on the Risk Terrain Modeling technique was an enormous help for me. Furthermore, I would like to thank the Rutgers School of Criminal Justice for providing me with the educational version of the RTM Diagnostics Utility software.

This thesis would not have been possible without the necessary data, which were provided by the Austrian Federal Criminal Police Office and the State Police Headquarters of the city of Salzburg. I would like to take this opportunity to thank you.

I would also like to express my appreciation to the Austrian Marshall Plan Foundation for their generous financial support, which enabled me to do my research project for my internship at the Louisiana State University.

Finally, but first in my heart, my family is due my deep gratitude for their endless support throughout my studies and my whole life. None of this would have been possible without your love and patience – to you all my thesis is dedicated.

Zusammenfassung

Diese Bachelorarbeit befasst sich mit der Implementierung und Evaluierung der Risk Terrain Modeling (RTM) Methode, mit der es möglich ist, Vorhersagen für Straftaten zu modellieren. Durch diese Methode können Gebiete lokalisiert werden, die eine erhöhte Wahrscheinlichkeit aufweisen, dass eine Straftat begangen wird. Die zu Grunde liegenden Annahmen fokussieren sich nicht auf vorhergegangene Straftaten, sondern auf Risikofaktoren, die einen Einfluss auf ihre Umgebung haben und die Wahrscheinlichkeit erhöhen können, dass eine Straftat begangen wird. Bei RTM handelt es sich um einen neueren Ansatz, der in Österreich bis jetzt noch nicht getestet wurde. Am Beispiel der Stadt Salzburg werden Vorhersagen für die vier ausgewählten Deliktarten Autodiebstahl, Einbruch, Körperverletzung und Raub für 2013 sowie 2014 erstellt. Zusätzlich werden die Ergebnisse des Jahres 2013 evaluiert und miteinander verglichen. Zuerst müssen die Risikofaktoren für die Deliktarten beschaffen bzw. selbst erfasst und geokodiert werden. Mit der Software RTMDx Utility können in einem ersten Schritt die Risikofaktoren, die mit der Deliktart korrelieren sowie deren Einfluss identifiziert werden. Die in der Software notwendige durchschnittliche Blocklänge wird basierend auf dem digitalisierten Straßennetz berechnet. Für die Risikofaktoren kann ein maximaler räumlicher Einfluss von einer bis vier Blocklängen angegeben werden. Aufgrund des fehlenden Fachwissens werden die Modelle für alle vier Optionen berechnet. Auf Basis der Ergebnisse der Software können die Risikofaktoren zu Risikolayern operationalisiert werden, was mithilfe zweier selbst entwickelter Modelle in ArcGIS geschieht. Danach folgt die Kombination der Risikolayer zu einer finalen Risikokarte, die entsprechend klassifiziert und unter Einhaltung kartographischer Aspekte fertiggestellt wird. Es werden vier Risikoklassen erstellt, die von geringem bis hin zu höchstem Risiko reichen. Die Evaluierung der Vorhersagen für 2013 geschieht mithilfe des Predictive Accuracy Indexes (PAI), wofür ein weiteres Modell in ArcGIS entwickelt wird. Ein erstelltes Diagramm zeigt die Ergebniswerte der Evaluierung, das zweite Diagramm stellt den Anteil der richtig vorhergesagten Straftaten gegenüber der Größe der vorhergesagten Gebiete dar. Diese Information ist besonders für die Polizei und andere Entscheidungsträger von Bedeutung. Insgesamt wurden 27 Modelle berechnet und Vorhersagen erstellt, da die Deliktarten teilweise noch in Jahreszeiten oder in Unterkategorien unterteilt wurden. Die Vorhersagen für Körperverletzungen, die Ergebnisse für die Jahreszeiten Frühling und Sommer umfassen, konnten die höchsten PAI Werte mit 31 und 23 erzielen. An zweiter Stelle kommt Raub mit einem PAI Wert von 18. Im Vergleich zu diesen Ergebnissen schnitten die Vorhersagen für Einbruchsdelikte allgemein und Einbruchsdelikte in Gebäude mit Werten von 4 schlecht ab. Die schlechteste Vorhersage wurde für Autodiebstähle erreicht, mit einem PAI Wert von 2. Abschließend lässt sich sagen, dass sich die Methode RTM auch für österreichische Städte, in diesem Fall für die Stadt Salzburg, anwenden lässt. Den größten Einflussfaktor für die Genauigkeit der Vorhersagen bilden die Risikofaktoren, die für das jeweilige Untersuchungsgebiet angepasst werden müssen.

Abstract

This research project deals with the implementation and evaluation of the Risk Terrain Modeling (RTM) technique, which allows localizing places, where the probability is high that a crime event will take place. RTM does not focus on previous events that happened, but on risk factors which have an influence on the environment and can increase the probability of the risk that a crime will be committed. RTM is a recently developed approach, that has not yet been tested in Austria. Using the example of the city of Salzburg, predictions are made for the four different crime events assault, auto theft, burglary, and robbery for 2013 as well as for 2014. In addition, the results of 2013 are evaluated and compared.

In a first step, the risk factor data for the crime events have to be obtained or rather self-captured and geocoded. With the RTMDx Utility software the risk factors which correlated with a crime event and their influence can be identified. The required average block length is calculated based on the digitized street network. For the risk factors, a maximum spatial influence can be set from one to four block lengths. Due to a lack of expertise, the models are calculated for all four options. Based on the results of the software, the risk factors can be operationalized to risk map layers, which is done using two developed models within ArcGIS. After that, the risk map layers are combined to a final risk terrain map, which is classified and finalized according to cartographic aspects. Four risk classes are created, which range from low risk to highest risk. The evaluation for the predictions of 2013 is done using the Predictive Accuracy Index (PAI) based on a developed model in ArcGIS. One diagram shows the results of the evaluation, and a second diagram shows the percentage of correctly predicted crime events in respect to the size of the predicted areas. This information is particularly interesting for the police and other decision-makers. In sum, 27 models were calculated and predictions made, because the crime events partly have been separated into seasons or sub-types. The predictions for assaults, which include results for the seasons spring and summer, resulted in the highest PAI values with 31 and 23. These results are followed by robbery with a PAI value of 18. Compared to these results the predictions for burglaries in general and burglaries into buildings with PAI values of four perform rather poorly. The worst prediction was achieved for auto thefts with a PAI value of 2. To sum up, the RTM technique can also be applied to Austrian cities, in this case to the city of Salzburg. The biggest influencing factor for the accuracy of the predictions is the risk factor data, which have to be particularly adapted for the project area.

Table of Contents

List of Abbreviations

BIC	*Bayesian Information Criterion*
BKA	*Austrian Federal Criminal Police Office*
CriPA	*Criminal Predictive Analytics*
CS	*Coordinate System*
CUAS	*Carinthia University of Applied Sciences*
GIS	*Geographic Information System*
GUI	*Graphical User Interface*
ICTF	*Intelligence Crime Task Office*
LPD	*State Police Headquarters of the city of Salzburg*
LTPD	*Lawrence Township Police Department*
LSU	*Louisiana State University*
PAI	*Predictive Accuracy Index*
RTM	*Risk Terrain Modeling*
RTMDx	*Risk Terrain Modeling Diagnostics Utility*
SAGIS	*Geographic Information System of the city of Salzburg*
SD	*Standard deviation*
SIMO	*Security Monitor*
URL	*Uniform Resource Locator*
USA	*United States of America*
WMS	*Web Map Service*

1. Introduction

This research project deals with the forecasting of crime events applying the recently developed technique Risk Terrain Modeling. The goal is to implement and evaluate the method for selected crime events for 2013 and to make predictions for 2014 that can be used by the Salzburg Police. This chapter describes the motivation and problem definition of the project as well as the method of solution. Furthermore, the expected results and the structure of the thesis are given.

1.1 Motivation

In 2012, there were about 548,000 reported crimes committed in Austria (Bundesministerium für Inneres, 2013). In the city of Salzburg the reported crimes amounted to 15,201 – that are more than 40 reported crimes every day – whereby property crimes had the largest proportion with about two-thirds of all reported criminal offenses (Salzburger Nachrichten, 2013). Among the varied tasks of the police which not only include recording criminal offenses and their investigations, also crime prevention plays a big role. Appropriate measures have to be taken into account in order to protect potential victims and prevent future crimes. One part of the preventive work is the aversion of crimes, this means that preparations and implementations of measures are made so that crimes are prevented and the population is protected (Jura Forum, 2014). For this purpose there exist several techniques in Geoinformation, "(…) such as Hotspot Mapping, Near Repeat Analysis, and Risk Terrain Modeling (RTM)" (Kennedy et al., 2011, p1), among others. While hotspot mapping is a widely available method which is also used in the Federal Criminal Police Office (Bundeskriminalamt Wien) in Austria (Kampitsch et al., 2008), RTM is a more recent approach and has not yet been tested in Austria. Results of this research will show, if the approach is helpful for the Austrian police.
The main reason why this research project is done in the USA at the Louisiana State University (LSU) in Baton Rouge is, because the supervisor, Professor Dr. Michael Leitner is an expert in crime modeling and mapping in general, and in the prediction of crime, in particular. On the other hand, criminal predictive analytics is still in its infancy in Austria. However, the Austrian police is very keen in the RTM and other predictive crime modeling approaches and would like to implement those approaches into their proactive decision-making processes.

1.2 Problem Definition

Motivated by the number of reported crimes and the fact that Risk Terrain Modeling can aid in strategic-planning and prevention-based operations (Caplan & Kennedy, 2010), the primary goal of this research project is to implement and evaluate the RTM technique using the example of the city of Salzburg and different crime types. The results are evaluated regarding the impact that different spatial influences of risk factors have on the final risk terrain map.

1.3 Method of Solution

The project will begin by getting familiar with the RTM process. Also, the required data have to be obtained or self-captured, respectively. Based on the available data and number of the reported crimes of the different crime types, specific crime events can be selected. The next step is the implementation of the risk terrain models using the RTM Diagnostics Utility software (RTMDx) based on different spatial influences. After that the models are operationalized and finalized within ArcGIS using self-developed models. Followed by a comparison and evaluation of the predictions for 2013 the best model can be identified and used for the prediction for 2014.

1.4 Expected Results

The expected results from this research project are the following:

- Predictions of locations where it is most likely that crime offenses will be committed in the city of Salzburg for 2013 and 2014 for different crime events.
- Development of models for the operationalization types Proximity (higher risk to be near a risk factor) and Density (higher risk thru a concentration of risk factors at a particular location) as well as for the evaluation in order to enable a semi-automatically process of the operationalization and evaluation.
- Evaluation and validation of the analysis results, including the comparison of risk terrain maps based on a different spatial influence.

1.5 Structure of the Thesis

At the beginning of this thesis different approaches how to predict future crime locations are described, the advantages of the RTM are given as well as its process described in detail. The section methodology deals with the project area and the required data. Furthermore, it contains the implementation process which involves the selection of particular crime events, the implementation of the risk terrain models and the visualization and evaluation. The chapter ends with a summary reviewing the implementation. Chapter four shows the results and the interpretation. In particular, the predictions and evaluations for 2013 are presented as well as the prediction for 2014. In chapter five the used methods are discussed and the work is reflected critically. It is followed with a final conclusion and a future outlook. At the end the references, the list of figures and the list of tables are given.

2. Theoretical Background

This chapter includes the relevant theory regarding approaches how to predict locations with a high risk that a crime will be committed in the future. The assumptions behind criminogenic factors as well as the Risk Terrain Modeling are described in detail. Also the used evaluation method is described.

2.1 Crime Analysis and Prediction

Crime analysis methods, in particular predictive crime analytics gained in importance over the last years. Instead of reacting to committed crimes the developed approaches show a trend towards predictive policing, which allows to predict where and when a crime event is likely to happen in the future (Perry et al., 2013). "Predictive policing refers to any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention", stated John Morgan (Uchido, 2009). This includes the usage of Geographic Information Systems (GIS) and statistical methods which allow the forecasting of future crime locations based on crime event data from the past or other data, such as environmental data (Ferguson, 2012).

For the prediction of future crime events there are different methods which are divided into two main approaches by Caplan et al. (2010 and 2012). One group includes so called reactive or retrospective methods, the other group represents the proactive approach. While methods of the retrospective approach, such as hotspot mapping or near repeat analysis are based on past criminal events, the proactive approach of Risk Terrain Modeling focuses on the interaction between environmental factors and the commitment of a crime event (Caplan et al., 2010; Caplan et al., 2012). An alternative classification is suggested by Perry et al. (2013), separating the methods into three groups, based on their context and benefits. The first group has the goal to determine where a crime event will be committed and includes hot spot analysis, regression models, data-mining, and near-repeat methods. The second group involves temporal and spatiotemporal analysis methods in order to predict when a crime event will occur. Risk terrain analyses, which include RTM, are the third group and focus on geospatial factors that influence the likelihood that a crime event will occur. While many used methods are retrospective, it is more important to "(…) identify early warning signals across time and space" (Groff & La Vigne, 2002, p.30), which is made possible using proactive methods.

2.2 Criminogenic Factors

That some areas have a high concentration of crime, also called hotspots, was found out through years of research and experience of police agencies (Caplan & Kennedy, 2010). The identification of the relationship between the geography and the act of crime events is an essential step when analyzing and mapping crime, state Murray & Grubesic (2013). That above all, the unit of analysis plays an important role regarding the areas where crime takes place is discussed by Weisburd et al. (2009; 2012).

Brantingham & Brantingham (2008) state that crime does not take place randomly, but it is dependent on the context of the given area, which they name the "urban backcloth". This backcloth involves the location of the city in a larger region, the existing road network, land use and also the "(…) socio-economic status of residents" (Brantingham & Brantingham, 2008, p.87). A city is structured through the above mentioned basic elements and these lead to locations that have a higher probability that crime concentrates there (Brantingham & Brantingham, 2008). The Brantinghams define a further concept how crime and place are related to each other, through crime generators and crime attractors, which are places that represent crime hot spots. Crime generators are areas with a high concentration of people, regardless of any criminal activity, which might be shopping malls, sport stadia, or entertainment venues. There is a higher chance that crime takes place in these areas, because of the combination of many people and the given settings which may attract offenders. However, crime attractors are locations where the opportunities to commit a crime are provided and attract offenders. These can involve areas with a high concentration of bars or prostitution, drug markets, insecure parking lots, and shopping malls. There also exist so called crime-neutral areas and often an area is a mixture of crime generators and crime attractors (Brantingham & Brantingham, 2008). RTM is based on so called criminogenic factors, or risk factors, that correlate with crime events and could be identified through years of research. The RTM Compendium lists the criminogenic factors for 17 different crime events. These factors do not produce the crime, but influence the environment and can point out locations which are at higher risk that a crime event will take place because the conditions are more attracting for offenders (Caplan & Kennedy, 2011), because the offender is influenced by "(…) situational and environmental features" (Groff & La Vigne, 2002, p.32). The risk factors, which are called "leading indicators" by Groff & La Vigne (2002) enable to predict future crime events and even show the reasons why future hotspots will develop.

2.3 Risk Terrain Modeling

The following chapters describe the theory behind the Risk Terrain Modeling technique, the advantages to retrospective analysis methods and show best-practice projects.

2.3.1 State of the Art

Geoinformation technologies are an essential part in fighting against crime in the USA since the 1990s (Kampitsch et al., 2008). Predictive policing, which includes methods and technologies for the prediction of future crime events, has been implemented in the USA for quite some time (Wilson et al., 2009; Ferguson, 2012; Friend, 2013). The RTM technique was developed by Joel M. Caplan and Leslie W. Kennedy at Rutgers University School of Criminal Justice, USA and has been tested and implemented with success several times, which is shown by best-practice projects (Caplan & Kennedy, 2010). While hotspot mapping is a widely available method which is also used in the Federal Criminal Police Offices (BKA) in Austria (Kampitsch et al., 2008), RTM is a more recent approach and has not yet been tested in Austria. Criminal predictive analytics is still in its infancy in Austria. However, the Austrian police is very keen in the RTM and other predictive crime modeling approaches and would like to implement those approaches into their proactive decision-making process.

2.3.2 Concept

RTM is an approach to risk analysis and was developed at the School of Criminal Justice at Rutgers University, USA, by Joel M. Caplan and Leslie W. Kennedy (Caplan & Kennedy, 2010). Instead of focusing on previous events that happened, as retrospective methods do, RTM focuses on the "(…) dynamic interaction between social, physical, and behavioral factors (…)" (Piza et al., 2010, p.1), which means that these risk factors have a bearing on the environment. These variables, also called criminogenic factors, can increase the probability of the risk that a crime will be committed, because the conditions at a particular place become more attractive for offenders. A list of criminogenic factors for 17 different crime types including auto theft, residential burglary, or gun shootings can be found in the RTM Compendium (Caplan & Kennedy, 2011). For each criminogenic factor an individual risk map layer is created and at the end all layers are combined to a final risk terrain map, partially with different weights. This final map shows locations where the risk is high for a future criminal offense. This information can be used by the police and decision-makers for strategic planning and preventive operations (Caplan & Kennedy, 2010). The benefits of using the RTM technique are varied. Costs can be saved because of more efficient resource allocation, and the results are useful for the police regarding tactical planning. The police agencies can act in a more proactive way and suppress crime before it happens (Caplan et al., 2011). Studies have shown that this approach has advantages compared to retrospective methods (see the following chapter 2.3.3 "Advantages to Retrospective Approaches").

2.3.3 Advantages to Retrospective Approaches

The reasons why the RTM technique is used are varied and described below. While traditional density maps, such as hotspot maps, show locations where crime events occurred, the RTM technique can find out where future crime events are most likely to occur. This proactive approach allows the police to potentially prevent and interrupt crime (McCue, 2011). Caplan and Kennedy (2010) state that retrospective hotspot maps, which are one alternative to the prospective RTM approach can correctly forecast locations of some future crime events, as well. McCue (2011) states that some areas "(…) associated with an increased likelihood for future incidents (…)" (McCue, 2011, p.5) would be missed with retrospective hotspot maps, but are found through RTM (Caplan & Kennedy, 2010).

A research study which was implemented by Caplan et al. (2011) compared the correctness of prediction of the RTM technique and the retrospective hotspot mapping method. The research was built upon two time periods of six months each for the crime event of shootings. The Risk Terrain models as well as the hotspot maps were produced in the same way to ensure the correctness of the comparison. Supposing that the retrospective method has a significant predictive validity too, the assumption was made that the correctness of the predictions of the risk terrain model would be higher (Caplan et al., 2011). After the model was created the predictions of both models were compared to the real incidents of shootings. The result showed that "(…) 21% more shootings occurred in high-risk cells predicted by the risk terrain map compared to the retrospective map. The top 10% of high-risk cells in the Period 1 risk terrain map correctly predicted 42% of future shootings, compared to 21% that were identified by the Period 1 retrospective map" (Caplan et al., 2011, p. 374). The retrospective map predicted future incidents too, but the risk terrain model was more exact (Caplan et al., 2011).

Another study, implemented by McCue, also compares hotspot mapping with the RTM technique, which is shown in Figure 1. The black dots symbolize known auto parts thefts, the blue ellipses show the result of traditional density mapping. The other dark gray areas are the result of the RTM technique, and the gray triangles show the crime incidents which occurred after creating the model. The figure shows that all five future auto parts thefts took place in an area predicted by the RTM method. This is because several environmental factors in those areas identified by the RTM method increased the conditions which attract crime offenders (McCue, 2011).

Figure 1 - Traditional density map compared to the RTM approach
(taken from Figure 2 in McCue, 2011)

If RTM did not show any improvements compared to retrospective hotspot maps, it would not make sense for the police to use it. Retrospective hotspot mapping takes "(…) locations of past events to forecast locations of future similar events (…)", state Caplan and Kennedy (Caplan & Kennedy, 2010, p.30). However, this would mean that the environment is static and that the location of crime events would not change over time. RTM uses the approach that the environment is dynamic and influenced by different factors. But hotspot mapping is useful, as well. For example, it can aid police in tactical processes where an immediate intervention has to take place in order to suppress crime by shifting resources to general areas with high crime. Hence, RTM is more useful for strategic planning and the development of preventive operations (Caplan & Kennedy, 2010).

2.3.4 Process of RTM

The entire process of implementing RTM includes ten steps. These steps have to be performed when the model is created manually within the GIS software, such as ArcGIS (Caplan & Kennedy, 2010). Recently, the Risk Terrain Modeling Diagnostics Utility software was released, which automates most steps of RTM (Caplan et al., 2013a). However, the underlying concept of the ten steps is still the same, although some of the used statistics might vary and are described in chapter 2.3.6. The ten steps of the process how to create a risk terrain model are described below (Caplan & Kennedy, 2010).

1) Selection of an outcome event
In the first step it has to be decided for which crime event the map should be produced, e.g. gun shootings or residential burglary. If RTM should be applied to several crime events, then several risk terrain models have to be produced, implementing the process for each crime event, respectively. The reason for that is that the criminogenic factors differ for each outcome event.

2) Selection of a study area
The study area can be of any extent, including large and small areas. The only limitation is the availability of data, which have to cover the entire study area and have to be obtained from a reliable source. The generated results need to be meaningful for further actions and all obtained base maps as well as the criminogenic factors should cover the whole study area. Also the unit of analysis should be considered, so that the study area does not get too large for the outcome event, which could lead to distortion in the results.

3) Selection of a time period to create the map for
The decision of the duration of the time period is an important aspect because it influences the way how the information can be used in short- and long-term decision-making. Also the seasonal circumstances have to be considered, so that the duration of the first time period is also relevant for the duration of the modeled time period. Depending on the chosen time period, the results have to be interpreted differently, so that the extracted information is useful. A one year time period is reasonable for a general predictive model, because the criminogenic factors of the environment are representative, stated Caplan (2014a).

4) Base maps for the study area have to be obtained
At least a shapefile which includes a polygon of the whole study area is necessary to clip the generated raster map to the boundaries of the study area. If statistical tests are made to validate the model, the cells outside the study area would lead to distorted results. Other useful base maps are road networks, waterways, parks or forested areas. These layers can then be used for the final map in order to give the user a better geographic context. It is important that base maps are used which are representative for the modeled time period.

5) Identification of aggravating and mitigating risk factors

All risk factors which are related to the chosen outcome event have to be identified, through literature review, review of reports, documents from research centers or interviewing experts. A list of correlated variables for different outcome events can be found in the RTM Compendium (Caplan & Kennedy, 2011). Aggravating risk factors increase the probability that a crime event will take place, while mitigating decrease the probability that a crime will be committed.

6) Selection of particular risk factors

The usage of all identified risk factors does not always mean that the produced models are better compared to when only a portion of all identified risk factors is selected. The selected data have to be validated and obtained from a reliable source, otherwise they could bias the results. RTM does always depend on the availability of data. One approach is to include only the ones that are most correlated with the outcome event. This concept is known as "less-can-be-more" and research studies were done by Piza et al. (2010). The following two approaches can be applied in order to decide which risk factors to include and which to exclude:

- Ad-hoc method:
  This method assumes that research results which were found from different sources are valid and adequate for use in a new application. All identified risk factors can be used, or only a few, depending on research or other knowledge. Another method would be to produce different risk terrain maps and test their validity to find out which model is the best. It is important that it can be explained why the chosen factors were used.

- Empirical method:
  As the name suggests this method requires further statistical analysis and an extra dataset of outcome event locations. The place-based connection on the chosen crime event of every risk factor is tested and only the most significantly correlated are used. Using this method needs more time and available data, but it maximizes the credibility and validity of the model. The statistical tests that are used, to find out which risk factors correlate most with the outcome event, are recommended to be Chi-squared tests. The workflow how to implement this step is described below.
  - First, all risk factors have to be operationalized to separate risk map layers. A map for the outcome event has to be generated too, for the selected time period of step 3.

- In a next step a blank vector grid has to be created which covers the whole study area. The cell size of this grid layer has to be the same that will be used for the produced risk terrain maps.
- Within the blank grid layer it will be calculated if a cell intersects with any of the separated risk map layers. This step can be done with "Spatial Joins" or "Select by Location" functions in ArcGIS.
- The blank grid layer has a new attribute field for each risk factor as well as for the outcome event. The attribute table has to be exported and saved in the .dbf-format to use it for the final statistical test.
- For the statistical test, a software is required which can produce cross tabulation tables and calculate Chi-squared tests, for example SPSS or Microsoft Excel. The dbf-file is imported and the Chi-squared tests performed. Therefore each risk factor column is calculated with the outcome event column.
- Through the cross tabulation the percentage of cells can be find out for the risk factor and the crime outcome event. Based on this result, only risk factors which are significantly correlated with the crime outcome event are selected. The RTM Manual recommend to use a p-value of 0.05 or less for this purpose. A further criteria is to only choose factors where the "(…) percentage of cells with both the outcome event and the risk factor was greater than the percentage of cells with a risk factor but no outcome event" (Caplan & Kennedy, 2010, p.84).

7) Operationalization of risk factors to risk map layers

For each risk factor an individual raster map layer is created, which shows its "(…) presence, absence or intensity (…)" (Caplan & Kennedy, 2010, p.84) throughout the study area. Several steps have to be performed and are described below.

- Determination of the cell size:

  Because the operationalized risk map layer has to be in raster format, the cell size has to be defined, which is based on the desired precision. The cell size defines how smooth the raster is and how precise the results are shown. The decision about the cell size is subjective but should be meaningful for the study area. This means that the cell size should be useful for the interpretation and further actions.

- Operationalization:
  Through the operationalization the influence of a risk factor can be calculated. Therefore it is necessary to find out in which way a risk factor influences the environment. As an example, the risk factor "bars" might have the following types of influence on the environment. The concentration of bars whereas a density map would be calculated or to be near a bar, whereby near has to be defined, e.g. 100 meters.

- Classification:
  Once all risk factors are operationalized, it is necessary to assign a classification scheme. This should be meaningful, either through binary values (cell has a risk or cell has no risk) or the cells are classified in a more accurate way. Within the RTMDx Utility software, the risk map layer are classified into binary values, because the more accurate classification method adds nuances, but did not produce better models (Caplan, 2014b). If the risk map layers are classified more accurately, Caplan & Kennedy (2010) recommend the following classification scheme: All cells with values greater than +2 standard deviations (SD) can be assigned as "high risk", those cells with values between the mean value and +2 SD as "medium risk" and all other cells with values below the mean as "low" or "no risk".

- Indexing:
  In this step weights are assigned to risk map layers. All values within each layer have to get an index, or weight, indicating its relationship to the risk factor and its influence on the outcome event. The process of indexing or the so-called standardization, has to be done in the same way for all risk map layers, otherwise the values cannot be compared. It does not matter how the risk factors were operationalized, but the indexing scheme has to be equal for each layer.

- Risk values:
  The last step includes the assignment of risk values to all layers using reclassification. Through this step each cell gets a new fixed value representing the risk. While the symbology only has an impact on the visualization, the reclassifying process modifies the attribute values of the data. Therefore the "Reclassify" tool can be used within ArcGIS.

8) Inter risk map layer weighting

There are two types of weighting which are described below.

- Un-weighted model:
  For a so called un-weighted model, the influence is weighed about the geography, which means that every risk factor is operationalized to a risk map layer and then indexed as described (see step 7).

- Weighted model:
  A weighted model is formed when the layers are weighed relative to each other. This method should be used if there are risk factors which have a bigger influence and are more important.
  - To calculate the weights it is recommended to apply a logistic regression analysis to find out how the risk factors correlate with the outcome event. This process is similar to the one done in step 6.
  - First, a blank vector grid has to be created whereby each risk map layer has to be in vector format.
  - In the blank vector grid layer new attribute columns are added. For each risk map layer a new column is created, which can be done with the function "Spatial Join". The new columns present the risk value of each layer for the cell. Furthermore a column which represents the number of incidents of the outcome event in each cell is defined. A last new column is created which includes a binary value showing if there is an incident in a cell or not.
  - In a next step, the table has to be exported and saved in the .dbf-format. After that it can be imported into a software product that can calculate logistic regressions, e.g. SPSS.
  - Within the regression model each risk map layer is shown through its variable, the dependent variable is the created column which shows if any incidents happened or not.
  - The result represents the influence that each risk factor has on the outcome event in those cells where the risk factor is present. The weight for layers is given through the Beta Coefficient (B). This value has to be multiplied with each cell in the relevant layer. In a last step, the "Reclassify" tool can be used to assign the new weighted risk values to the raster risk map layers.

9) Combination of all layers to a final map

All risk map layers are combined in order to create a composite map using the "Raster Calculator". As a result a new raster is created whereby the cells of the risk map layers are added through map algebra and generate a new risk value. This map represents the final risk terrain map.

10) Finalization of the risk terrain map

The last and final step includes measures to present the risk terrain map in a meaningful way. This might involve the clipping process of redundant areas as well as symbolization and design by reverting to cartographic concepts. The obtained base maps can be added to the map so that it can be interpreted easier.

The classification of the final risk terrain map is not dependent on the classification of the separate risk map layers. It depends on what one wants to show with the final map. That can be highlighted areas which have a risk value greater than the mean value or using stretched symbology. Caplan (2014b) recommends creating three or four classes which are based on standard deviations, because that allows for consistency across the separate risk map layers. The other methods are based on algorithms that are unique for each dataset and lead to difficulties regarding comparisons of different risk terrain models. Using the classification based on standard deviations enables easy comparisons. The classes are separated by values less than the mean, greater than the mean but less than one SD, greater than one SD but less than two SDs and all values greater than two SDs. In addition, all areas with values greater than two SDs are statistically approximately the top five percent of the risk areas (Caplan, 2014b). It is important to note that the chosen classification scheme can involve areas that would be defined as hotspots. These areas can be presented differently by using different classification methods. To produce statistically significant clusters a further tool, such as "Hotspot Analysis" or "Cluster and Outlier Analysis" can be used. With a hotspot analysis method, the final risk terrain map can show places with a high risk and low risk which is not only visually significant, but also statistically significant (Kennedy et al., 2012).

2.3.5 Best Practice Projects

In the following, two best practice projects showing an application of the RTM technique for burglaries are described.

Application of RTM to Residential Burglaries

The RTM technique was implemented to identify the most risky areas for residential burglaries in Lawrence Township (New Jersey). Based on research the following three risk factors were determined for this study area: bus stops, calls for suspicious vehicles and calls for suspicious persons. Data were provided from the Lawrence Township Police Department (LTPD) as well as from the USA Census Bureau. As a first step the addresses needed to be geocoded which was done in ArcGIS using the geocoding toolbox. In a next step, a Moran's I statistic for spatial autocorrelation was used to find out if the burglary calls are clustered, which did not. Then the data were converted from a geographic coordinate system to a projected coordinate system and used to create a kernel density image of the burglary calls. The next main step was the operationalization of each of the chosen risk factors. Through the kernel density images for each risk factor, the concentration of the risk factors could be visualized. With the "Raster Calculator" in ArcGIS the final risk terrain map was generated. This map showed risk indices from zero (no risk factor is present) to three (all risk factors are present) based on 100 square foot cells. In order to use the data for statistical analysis, the result layer representing each risk factor were joined. The statistical test showed that all risk factors are statistically significant associated with locations that show a high concentration of burglary crime events. The final risk terrain map shows locations where it is most likely that a burglary crime event will be committed in the future.

Application of RTM to Burglaries in Morris County

In 2007, the Morris County Prosecutor's Office Intelligence Crime Task Force (ICTF) was founded, in order to implement predictive policing methods. One big goal was to reduce the crime countywide without using further resources. Through an analysis of crime event data, problematic areas could be identified. These were classified in a next step and police departments could focus on these areas to prevent further crime. For this process a GIS was used and the RTM technique was adapted for residential burglaries. For the model the following five criminogenic factors were determined: Past burglaries, residential location of individuals that were arrested for burglary between 2009 and 2011, proximity to main highways, locations with a high concentration of males between 16 and 24, and locations of apartment complexes and hotels. Through geographic profiling the investigations could be supported. After implementing the RTM technique in Morris County (New Jersey), there was a significant reduction regarding crime offenses, the total crime index decreased by about 11% since 2007, and violent crime even by 21% (Paul & Joiner, 2011).

2.3.6 RTM Diagnostics Utility

The RTM Diagnostics Utility, short RTMDx, is a software product which mostly automates the steps of the Risk Terrain Modeling technique, especially steps six through nine, which were described in the RTM process above. The software developed by Rutgers School of Criminal Justice and was released recently (Rutgers, 2014). Each risk factor that is included in the model is tested on its spatial influence on the outcome event. That step allows finding out only the most correlated risk factors which then are included in the calculation in order to produce the "best" risk terrain model (Caplan et al., 2013a). The outcome event data are only used for the calculation of the correlation, the location itself is not included, as it is done by retrospective methods (Caplan, 2014b). The involved risk factors are reported and weighted regarding their influence on the outcome event. For the analyses which risk factors should be included, the empirical method (which is described in chapter 2.3.4) is implemented. Further, there exist two versions of the RTMDx Utility, a professional version and an educational one. For this research project the free educational version is used which does not produce a final map, compared to the professional version where the final risk terrain map is provided in the form of a GeoTIFF image. This task has to be implemented manually by operationalizing the risk factors using the "best" model as described in the report (Caplan et al., 2013a). Figure 2 shows the Graphical User Interface (GUI) of the RTMDx software.

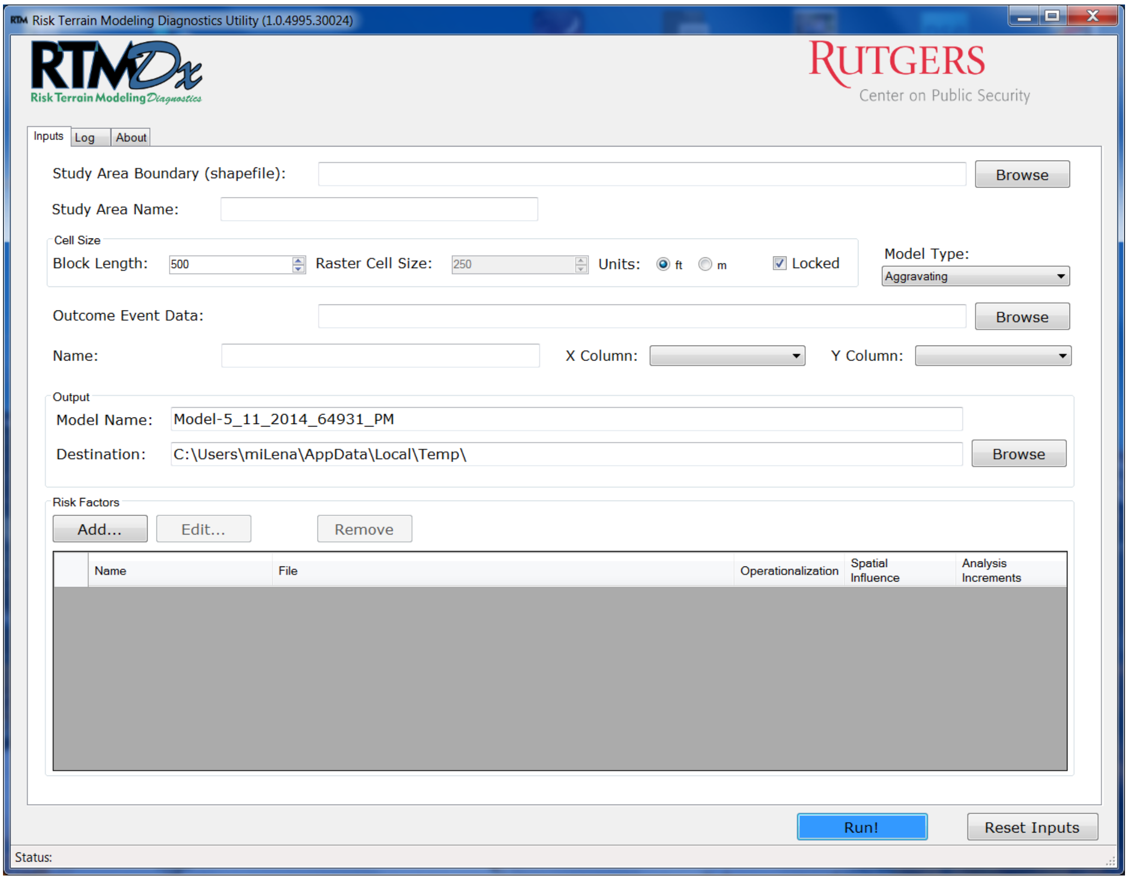

Figure 2 – The GUI of the RTM Diagnostics Utility software

The software includes the three tabs "Inputs", "Log", and "About", whereby the tab "Inputs" is the most important one for the analysis. In a first step the shapefile representing the study area as well as its name has to be specified. The next part includes the information regarding the cell size. The “Block Length” and “Raster Cell Size” are set to 500 and 250 by default, but can be changed manually. For the unit Feet and Meters are selectable. The user manual recommends setting the value for the block length “(…) to the mean length of a block face in the study area” (Caplan et al., 2013a, p.17) and the value for the raster cell size to the half of it. The produced risk terrain model can either be an aggravating or a protective model type. While aggravating means that the risk factors correlate with the locations of the outcome event, protective would assume that the risk factors correlate with the absence of the outcome event. Because the latter one has not been researched quite long, most risk terrain models have been implemented using the aggravating model type. In the next step the outcome event data have to be included, which is necessary for analyzing the most correlated risk factors. For the model an output destination and a name have to be declared (Caplan et al., 2013a). The next section refers to the risk factors that are included. For each added risk factor further parameters have to be set, as represented in Figure3.

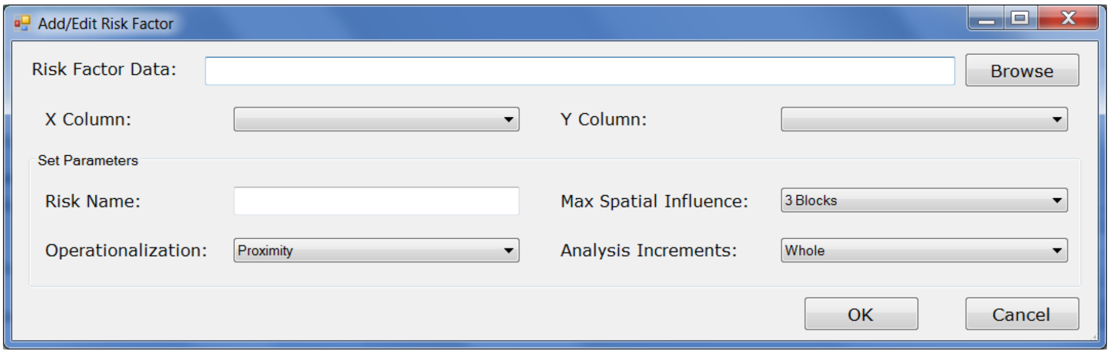

Figure 3 – The parameters for a risk factor

First, the dataset of a risk factor and its name have to be specified, whereby the data has to be in the form of point features either in .shp or .csv format. The “Max Spatial Influence” parameter determines within which distance the risk factor has an influence on the environment. This distance is then included in the statistical analysis for the risk factor. It can be set up to four blocks, because empirical research showed that this is the maximum spatial influence. “Operationalization” includes the type how the risk factor should be operationalized and can be set to “Proximity”, “Density”, or “Both Proximity and Density”. While density defines that a high concentration of the risk factor increases the likelihood that a crime event will take place, proximity “(…) assumes that being within a certain distance from a risky feature increases the likelihood of (…) crime event locations” (Caplan et al., 2013a, p21). If the operationalization type is not obvious, the option “Both Proximity and Density" can be used, wherefore the type is empirically tested and set by the RTMDx software or it can be calculated manually with a nearest neighbor analysis. For the last parameter "Analysis Increments" the options "Half" and "Whole" are selectable. While "Half" takes increments for half- and whole-blocks, "Whole" only takes increments for the whole length of a block. All mentioned parameters have to be specified for each risk factor (Caplan et al., 2013a).

After all parameters for the model are determined and all risk factors included, the model can be started. When the calculations are finished, the result is summarized in a report. For the professional version a risk terrain map is created in the form of a GeoTIFF image. For the educational version, the information in the report can be used to operationalize and finalize the risk terrain model using GIS software (Caplan et al., 2013a).

The advantages by using the RTMDx Utility software compared to manually produced risk terrain models are the following. It is time saving, because the statistical tests are done automatically by the software. Using the professional, but commercial version, which is combined with an online training, neither the operationalization nor the finalization of the risk terrain model has to be done manually. Further, the statistical methods used in the software were improved with the collaboration of Jeremy Heffner from Azavea (Caplan et al., 2013a).

The statistics used in the RTMDx Utility identify which risk factors are statistically significantly correlated with the outcome event and calculate their spatial influence. While the input has to be in vector format, the calculations and the output are in raster format. For the operationalization type "proximity", it is measured if a raster cell is within a particular distance to the risk factor, which is then expressed as binary values. For "density", it is tested if a raster cell is within a high concentration of the risk factor. Also for that operationalization type, binary values are calculated. Cells with values greater than two standard deviations are set to one, and all other cells are set to zero. For the operationalization process, a series of variables is tested, and both types are based on a kernel stamping density method. In order that no bias in the correlations occur, which could happen because a large number of variables is calculated, a penalized Poisson regression model is calculated, using cross-validation. In a next step, the model is simplified using a bidirectional stepwise regression process, where the Bayesian Information Criterion (BIC) is measured. Within this process, the BIC value is measured at the beginning. Then, each variable is added and the BIC is measured again to find the model with the best (in this case the lowest) BIC value. Through this process the optimal model is identified which only includes the most correlated risk factors and their spatial influences. Because the crime events might be related to each other, which is the contrary assumption of the Poisson process, a negative binomial distribution is calculated. This includes a variable to show "(…) over-dispersion of counts, which can help to represent dependency between the crime events" (Heffner, 2013, p.37). At the end, the model with the best BIC value between the two distributions is selected as best model and included in the RTMDx report as result (Heffner, 2013).

2.4 Evaluation

Testing the predictive validity of the produced model is an important task, and built upon a simple principle: A predictive model is implemented and then it is tested how many crimes indeed happened in the predicted areas. It is an optional step and requires an extra dataset of outcome events of the modeled time period, but through the evaluation it can be shown with which certainty the results can be used and how good a method is to predict future crime events (Chainey et al., 2008). It is also possible to find out, which time period or which risk factors are most suitable for the prediction. In order to evaluate the model, additional data of the outcome events of the following time period are required (Caplan & Kennedy, 2010).

A widely used approach is the so called hit rate method, whereby it is calculated how many crimes happened indeed in the predicted areas. But the disadvantage of this method is that the size of the area is not considered, although it plays an important role. Another technique which includes the size is the Search Efficiency Rate by calculating the number of crime events based on square kilometers, but it makes comparisons more difficult. In this research project, the evaluation of the produced risk terrain models is done using the Prediction Accuracy Index (PAI). Through the PAI the size of the whole study area is included as well as the size of the predicted areas with a high risk of future crime events. The PAI was invented in order to "(…) consider the hit rate against the areas where crimes are predicted to occur with respect to the size of the study area" (Chainey et al., 2008, p.12). To calculate the Prediction Accuracy Index, the hit rate is divided by the percentage of the predicted areas in relationship to the whole study area. The hit rate is defined as the number of crime events which reside in the predicted areas divided by the whole number of crime events. Figure 4 shows the formula to calculate the PAI:

$$PAI = \frac{Hit\ Rate}{Area\ Percentage} = \frac{\frac{n}{N}}{\frac{a}{A}}$$

Formula 1

The “n” shows the number of crime events that happened in the predicted areas, the “N” is the total number of crime events in the whole area. The “a” represents the area of locations considered to be high risk (i.e., the predicted areas) and parameter “A” is the size of the whole study area. The advantage of the PAI is that it is not difficult to calculate and can be used for study areas of any size and for any crime events and techniques (Chainey et al., 2008).

3. Methodology

This chapter presents the methodology of the thesis. In the first two sections the problem definition and the used method of solution are given in more detail. Then, the project area and the required data are described. The section implementation includes and discusses the selection of particular crime events, the necessary data capture and geocoding, the way how the risk terrain models are implemented and the visualization as well as the evaluation. The last subchapter summarizes the whole section.

3.1 Problem Definition

This bachelor project is done within the context of the research project Criminal Predictive Analytics (CriPA), which focuses on predictions of future criminal activity. Using different approaches and techniques, it will further investigate which methods are appropriate to predict crime trends (Joanneum Research, 2014). For this bachelor project the proactive risk assessment technique Risk Terrain Modeling is implemented and evaluated for the first time for an Austrian city. Risk terrain models are implemented for the four different crime types Assault, Auto Theft, Burglary and Robbery, and varying spatial influences of selected risk factors for 2013 and 2014. The predictions are based on a one-year or seasonal time period, respectively. These results then can be compared and evaluated, showing differences between the selected spatial influences. Furthermore, the impact of the available risk factor data on the final risk terrain map is evaluated. The research results should show if and for which crime types the method is appropriate or otherwise outline possible reasons. In addition, the information provided by the predictions for 2014 can be used by Salzburg Police.

3.2 Method of Solution

In a first step, research on the concept and process of RTM as well as on the RTMDx Utility software has to be done. Crime events have to be selected based on relevant crime types in the city of Salzburg and their risk factor data have to be obtained or self-captured, respectively.

Then, the risk terrain models can be implemented using the RTMDx Utility software and different spatial influences. Using the software the correlated risk factors and their weights can be identified. After that, within ArcGIS the models for 2013 are operationalized and finalized using self-developed models for the operationalization types "proximity and density", based on data from 2012. In this way, the models can be implemented semi-automatically. Using a unified visualization, the models can be interpreted and compared to find out differences. Followed by the evaluation with real crime data from 2013, the results can be compared numerically, showing the most appropriate model and the percentage of correctly predicted crime offenses in 2013. This information can then be used to make the predictions for 2014, based on data from 2013 using the same process. Figure 5 presents an overview about the schematic procedure.

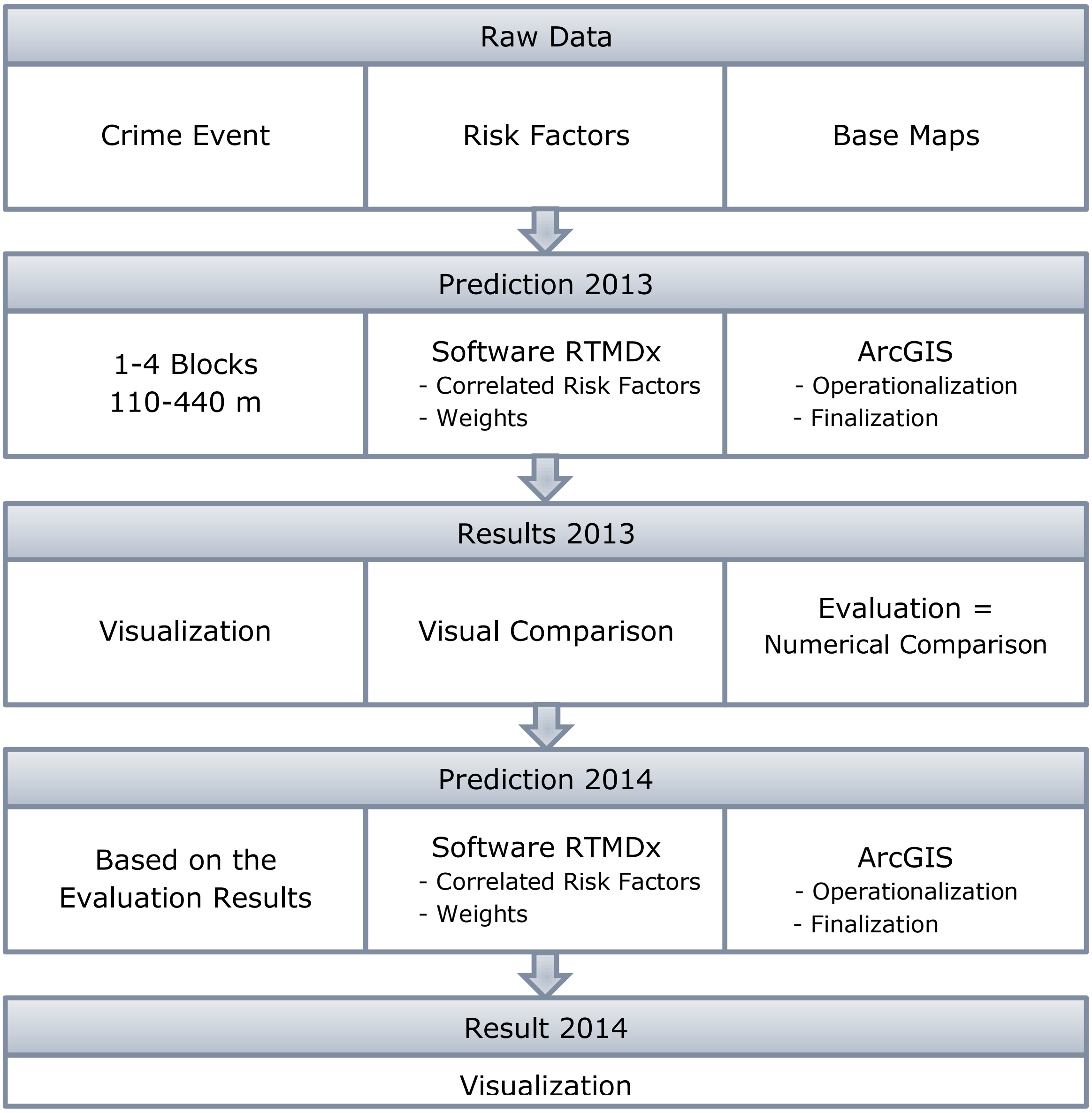

Figure 4 – Schematic procedure of the method of solution

3.3 Project Area

As can be seen in Figure 6, the project area for this research is the city of Salzburg, which is located in the western part of Austria. The city of Salzburg is the capitol of the state of Salzburg. It is characterized by the river Salzach, several forested hills, and an old town. It has about 148,000 inhabitants (as of January 2014) and an area of about 65.68 km² (Stadt Salzburg, 2014). In the map, the streets, rivers and lakes as well as the forests are visible. The gray areas represent residences.

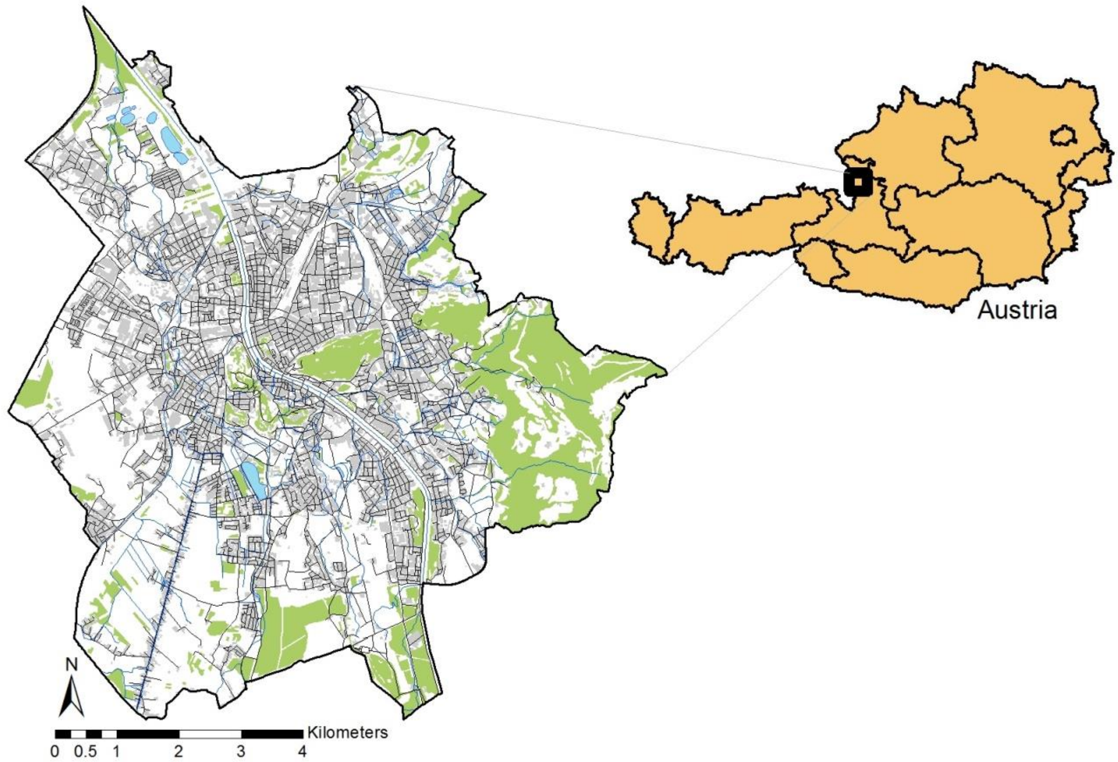

Figure 5 – On the right is the location of the study area within Austria. On the left the city of Salzburg is shown.

3.4 Geodata

The following subchapters explain the required data for the analysis and visualization and can be separated into the three categories crime data, risk factor data, and base map data. The data is presented in tables, including six columns. The first column represents the name of the dataset. In the second column the file format and in the third column the extension is given. Column four represents the year where the data were captured and in column five the format of the data is shown, which can either be raster or vector, whereby vector data can be in point, polyline, or polygon format. The last column represents the data source, which can be the Austrian Federal Criminal Police Office (BKA), the State Police Headquarters of the city of Salzburg (LPD), the Geographic Information System of the city of Salzburg (SAGIS), or self-captured.

3.4.1 Crime Data

Crime data are necessary to find out if there is a correlation between the outcome event and the risk factors and it is also used for the evaluation. The data was provided by the Austrian Federal Criminal Police Office, where the data are stored in the Security Monitor (SIMO) database. The data were available in form of address-level crime locations for the selected crime events and are shown in Table 1.

Table 1 – Crime Data

Data Name	File Format	Extension	Year	Data Format	Data Source
Assault	Comma Separated Value	.csv	2012 2013	Vector Point	BKA
Auto Theft	Comma Separated Value	.csv	2012 2013	Vector Point	BKA
Burglary	Comma Separated Value	.csv	2012 2013	Vector Point	BKA
Robbery	Comma Separated Value	.csv	2012 2013	Vector Point	BKA

3.4.2 Risk Factor Data

The risk factor data were necessary for the calculation of the risk terrain models. In total, 15 risk factor datasets were provided or could rather be captured, as seen in Table 2. All risk factors are aggravating except police departments, which represent a mitigating risk factor.

Table 2 – Risk Factor Data

Data Name	File Format	Extension	Year	Data Format	Data Source
Banks	Shapefile	.shp	n/a	Vector Point	Self-captured
Bars & Pubs	Shapefile	.shp	n/a	Vector Point	Self-captured
Buildings	Shapefile	.shp	Date of survey	Vector Polygon	SAGIS
Bus Stops	Shapefile	.shp	2009	Vector Point	SAGIS
Cash Points	Shapefile	.shp	n/a	Vector Point	LPD
Clubs & Discos	Shapefile	.shp	n/a	Vector Point	Self-captured
Entertainment Venues	Shapefile	.shp	n/a	Vector Point	Self-captured
Leisure & Fast-food Outlets	Shapefile	.shp	n/a	Vector Point	Self-captured
Nightclubs	Shapefile	.shp	n/a	Vector Point	Self-captured
Official Buildings	Shapefile	.shp	2013	Vector Point	SAGIS
Pawn Shops	Shapefile	.shp	n/a	Vector	Self-captured

				Point	
Police Departments	Shapefile	.shp	n/a	Vector Point	LPD
Post Offices	Shapefile	.shp	2014	Vector Point	Self-captured
Railway Stops	Shapefile	.shp	2010	Vector Point	SAGIS
Schools	Shapefile	.shp	2011	Vector Point	SAGIS

3.4.3 Base Map Data

The data representing basically layer are important for the visualization purpose, so that the information can be presented and interpreted easier. The shape of the city of Salzburg was also used for the calculation and based on the street network the average block length could be calculated. Table 3 represents the base map data.

Table 3 – Base Map Data

Data Name	File Format	Extension	Year	Data Format	Data Source
City of Salzburg	Shapefile	.shp	n/a	Vector Polygon	SAGIS
Forests	Shapefile	.shp	Sheet line date	Vector Polygon	SAGIS
Lakes	Shapefile	.shp	n/a	Vector Polyline	SAGIS
Rivers	Shapefile	.shp	n/a	Vector Polyline	SAGIS
Street Network	Shapefile	.shp	n/a	Vector Polyline	Self-captured

3.5 Implementation

The chapter implementation deals with the necessary steps to solve the problem, defined in the beginning (see chapter 3.1 "Problem Definition"). The workflow of the main aspects of the implementation, including data, the RTMDx Utility software, as well as ArcGIS and the results, which include the prediction and evaluation, is shown in Figure 7.

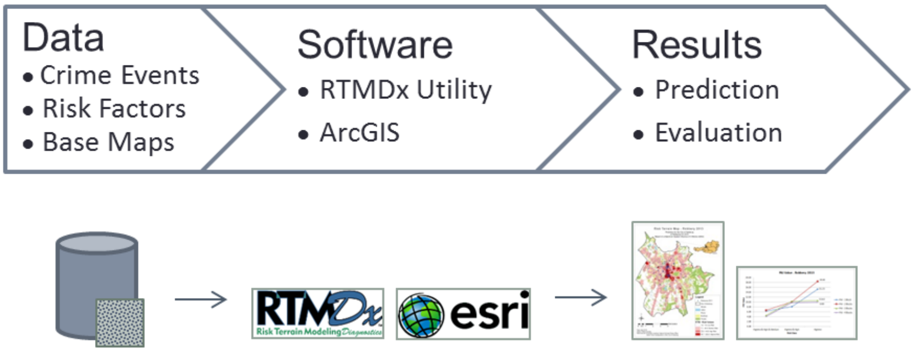

Figure 6 – Main implementation steps

3.5.1 Selection of Crime Events

To decide for which crime events the risk terrain models should be implemented and the predictions evaluated, three criteria have been defined. First, the risk factors for the crime event have to be known, based on a reliable source such as the RTM Compendium, which lists the risk factors for 17 different outcome events. Second, widespread criminal offenses should be analyzed. The third aspect is that the crime event has to be relevant in the city of Salzburg. These aspects were discussed with a team of the State Police Headquarters of the city of Salzburg at a meeting before the start of this thesis research. Based on these three criteria the following four crime events for this research project were selected: Assault, Auto Theft, Burglary, as well as Burglary into buildings, and Robbery.

3.5.2 Data Capture and Geocoding

Some of the risk factor data could not be obtained but are very important data for the whole analysis process. Because of that the data were self-captured and geocoded. In sum, eight different risk factors were captured, which are included in the table of the chapter 3.4 "Geodata". The data capture process was based on as much reliable sources as possible. The name of the data and the name as well as the Uniform Resource Locator (URL) of the data source and the number of captured features are presented in Table 4.

Table 4 – Self-captured risk factor data

Data name	Data Source Name	Data Source URL	Number of Features
Banks	Bankkonditionen	www.bankkonditionen.at	129
Bars & Pubs	Salzburg-Night	www.salzburg-night.at	123
Clubs & Discos	Salzburg-Night	www.salzburg-night.at	25

Entertainment Venues	Salzburg News	http://www.salzburg.com/	18
Leisure & Fast-food Outlets	Tourism Salzburg City of Salzburg	http://www.salzburg.info/ http://www.stadt-salzburg.at/	32
Nightclubs	Salzburg-Night	www.salzburg-night.at	25
Pawn Shops	Herold	http://www.herold.at/gelbe-seiten/branchen-az/pfandleihe	6
Post Offices	Austrian Post	www.post.at	32

For each layer the name of the feature and its address, including the street, house number, zip code, and city were recorded. The next task was to geocode these addresses in order to assign points with an x- and y-coordinate to the addresses. For the geocoding process a table is necessary which includes the addresses which should be geocoded as well as a reference dataset represented by a map that includes the addresses of the research area (Cote, Harvard University). The Carinthia University of Applied Sciences (CUAS) has the possibility to use ArcGIS Online for geocoding processes, where a reference dataset is provided already. For the self-captured data an account was unlocked in order to process the geocoding task. After the login the Geocoding Service is provided in "Ready-To-Use Services – Geocoding". The "Geocoding" toolbar was added and the tool Geocode Addresses chosen, whereby the "World Geocode Service (ArcGIS Online)" of the "Ready-To-Use Services" was selected. The geocoding process resulted in 99% matched addresses, shown in Figure 8, and the two data which were tied. If data are identified as tied, there are several addresses with the best match score, but located at different places (ArcGIS Help, 2012). The tied data could be matched manually by assigning the right address.

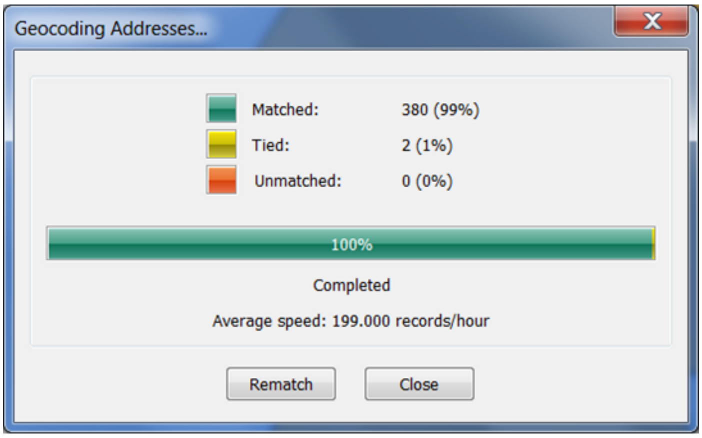

Figure 7 – Geocoding process

A layer representing the street network of the city of Salzburg was necessary, too. On the one hand it represented an important base map layer and on the other hand it was necessary for the calculation of the average block length, which was required for the software. Table 5 shows the data name and data source information as well as the number of captured street segments, which were in fact 4,635.

Table 5 – Self-captured base map data

Data name	Data Source Name	Data Source URL	Number of Features
Street Network	SAGIS	http://www.salzburg.gv.at/themen/se/sagis/download.htm	4,635

The Web Map Service (WMS) "Straßengraph_WMS_Land_Salzburg" could be downloaded for free from SAGIS and connected in ArcGIS using the following URL of the WMS Server: http://service.salzburg.gv.at/ArcGIS/services/Strassengraph_WMS_Land_Salzburg/MapServer/WMSServer?version=1.1.1. A new polyline shapefile was created and the street segments digitized using the Editor. The digitized street network of the city of Salzburg is represented in Figure 9.

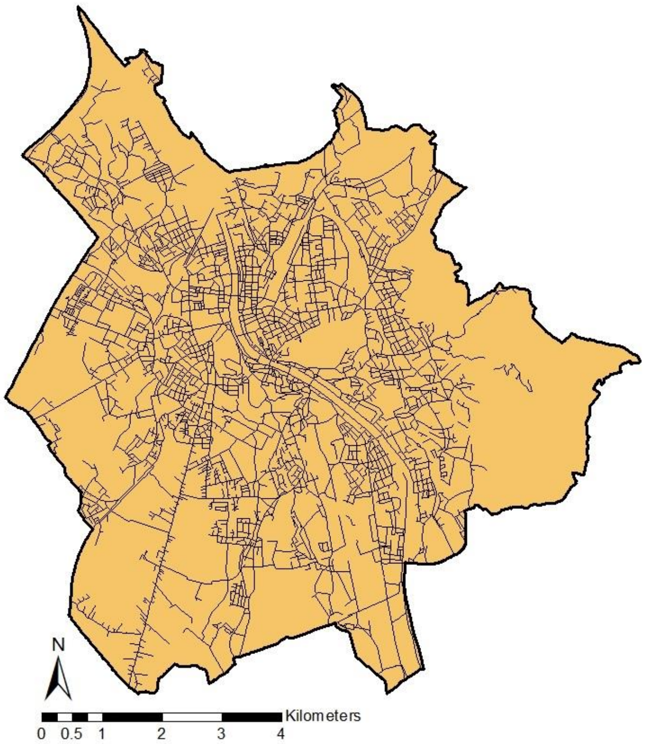

Figure 8 – Digitized street network of the city of Salzburg

3.5.3 Data Evaluation and Preparation

Before the analyses were started, the data were evaluated. For the RTMDx Utility software the data had to be in vector format, represented by points. The two layers "Cash Points" and "Police Departments" included features for the whole state of Salzburg. Because only features within the city of Salzburg were required, these have been extracted using the "Select by Location" function in ArcGIS.

The crime event data were provided in .csv tables and had to be converted in order to visualize and use the data for the software. This task could be solved using the function "Create Feature Class" – "From XY Table…" within ArcGIS. The parameters for the x and y field have to be specified as well as the coordinate system (CS). Because the data are projected in the MGI Austria Lambert CS, but the used CS was MGI Austria GK M31, the latter one had to be specified within the "Advanced Geometry Options". Otherwise the data would not be at the correct spatial location.

The last feature class that had to be modified was buildings, which were represented by polygons. Because the RTMDx Utility software only takes point features as risk factors, the polygons had to be converted to points. For this task a model which automates several steps was created by Rutgers University and is provided for free. It can be used within ArcGIS and converts the polygons to points in an appropriate way for the RTM process, so that both the perimeter and interior of a building are represented (Caplan et al., 2013b).

Using the tool, the spatial influence distance has to be specified, which is the same as specified in the software.
Although the coordinate system was specified when the data were added to ArcGIS, it was not always updated within the feature class. Hence the feature classes had to be projected again, using the function "Project" which can be found in the "Projections and Transformations" toolbox.
In the RTMDx Utility software the average value of the block length has to be specified. A block is defined as consisting of several adjacent properties and surrounded by a street (Wikipedia, 2014). In other words, the block length is the distance between two intersecting streets. In order to calculate this parameter for the city of Salzburg, the length of the digitized street network was calculated. This could be done by creating a new column of the type double. Within the attribute table the option "Calculate Geometry" has to be used to calculate the length. Then, the option "Statistics" shows the mean length of the street network, presented in Figure 10 which was rounded up to 110m.

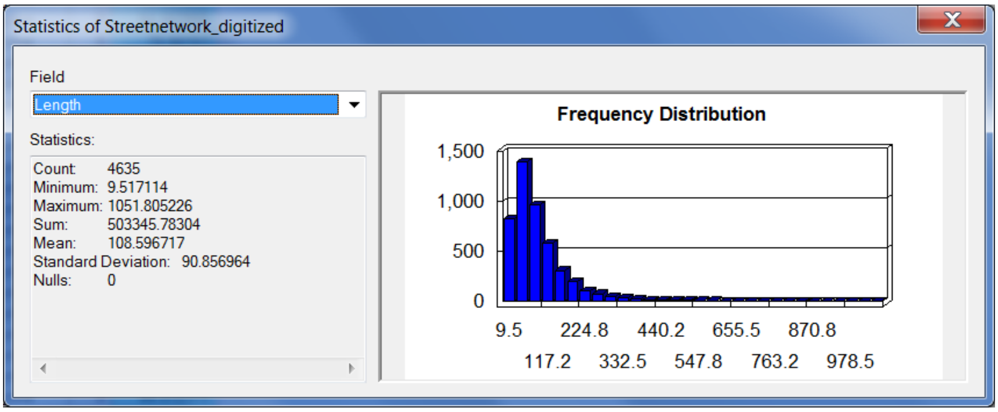

Figure 9 – Statistics of the field Length of the street network

3.5.4 Requirements Concerning the Models

For the implementation of the risk terrain models several requirements were determined. The time period was set to one year, because that is reasonable for a general predictive model, as Caplan stated (2014a). If no results could have been calculated, the risk terrain models were implemented for seasons. For each risk factor included in the model, three further parameters had to be specified. The maximum spatial influence represents the influence of a risk factor on the environment and can be set to one, two, three, or four blocks. For the analysis the model was calculated four times so that each option was selected once. The operationalization type can be set to proximity, density or both and was set according to the description of the risk factor in the RTM Compendium. If the option is not obvious, the option both was selected and the software calculated which option fits better. The last parameter which has to be specified is analysis increments, which can be set to half or whole. For the calculation of the models the option half was used, in order to get more accurate results, although the running time was doubled.

3.5.5 Implementation of the Risk Terrain Models

The implementation of the risk terrain models is structured by the following workflow. At first, the steps one to five of the RTM process described in chapter 2.3.4 have to be considered. After that each model is calculated four times using the RTMDx Utility software. Based on the best presented model in the report, the model has to be operationalized and combined manually within ArcGIS.

For the calculation with the software, the required parameters have to be set. The study area, the block length, and the cell size have to be specified as well as the model type, the outcome event data, and the output location defined. A name for the model and its output location has to be determined. Then, the identified risk factors can be added, defining the data itself, the risk factor name, the operationalization type, the maximum spatial influence, and analysis increments. After all risk factors are added and the calculation done, the result is summarized in the form of a report. This includes the best model and lists the correlated risk factors, the spatial influence on the outcome event, and their weight. The formula to combine the separate risk map layers is given too.

The operationalization was done within ArcGIS, based on the information of the RTMDx Utility report. A model was developed for the operationalization type proximity and density, respectively, using the "Model Builder". Through the "Model Builder" several geoprocessing tools can be connected, whereby the output of one tool is used as input for another tool (ArcGIS Help, 2012). Because often the same steps have to be implemented, the models enable an automatization of the operationalization. Both models operationalize a risk factor showing its influence on the environment and result in binary values showing cells which have an influence or which have no influence.

The model for the operationalization of a risk factor with the type proximity is shown in Figure 11. At first, a buffer is created based on the input risk factor and a given distance by the user. This distance represents the spatial influence and has to be set to the value listed in the report. In a next step, the created buffer features are joined with an empty vector grid of the city of Salzburg. This task is necessary, so that the whole study area is covered. The blank vector grid has only to be calculated once, which could be done using the tool "Create Blank Vector Grid of Study Area". This tool is included in the toolbox "Risk Terrain Tools" and is provided for free at the website of RTM (http://rutgerscps.weebly.com/rtm.html). After the join, the feature is converted to a raster and reclassified into binary values using the raster calculator. The reclassification is based on the field which counts how many buffer features overlay a cell. If a cell has a value greater than one, the final value is set to one, otherwise the cell gets a value of zero.

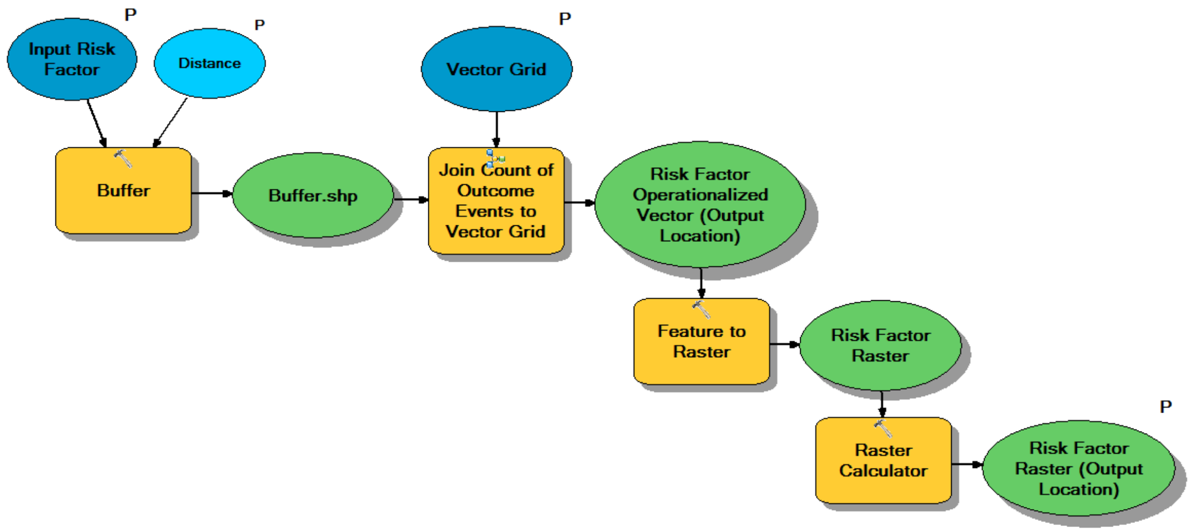

Figure 10 – Model for the operationalization type proximity

Figure 12 represents the Graphical User Interface of the tool. The user has to specify an output location for the operationalized risk factor, the blank vector grid, and the input risk factor as well as the distance representing the influence.

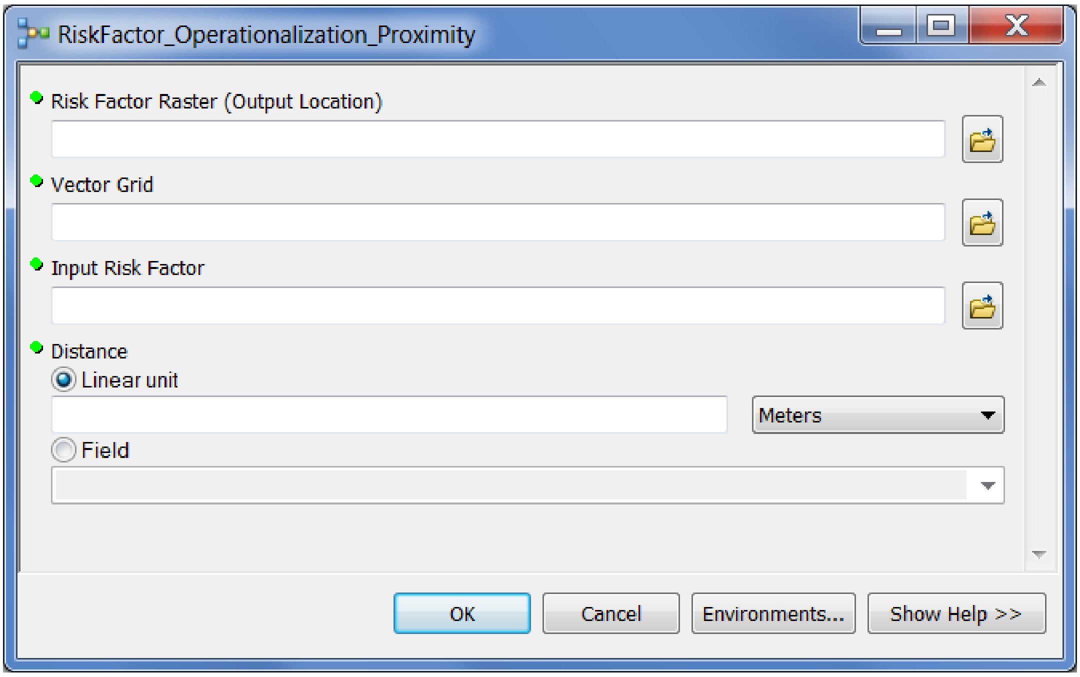

Figure 11 – GUI for the operationalization type proximity

The model to operationalize a risk factor with the type density is shown in Figure 13. In a first step a kernel density is calculated, based on the input risk factor and a search radius which is equivalent to the spatial influence. With the tool "Get Raster Properties" the standard deviation value of the raster can be read out and is used to reclassify the raster. Within the raster calculator all values that are greater than two standard deviations are assigned the new value one and all other cells get the value zero.

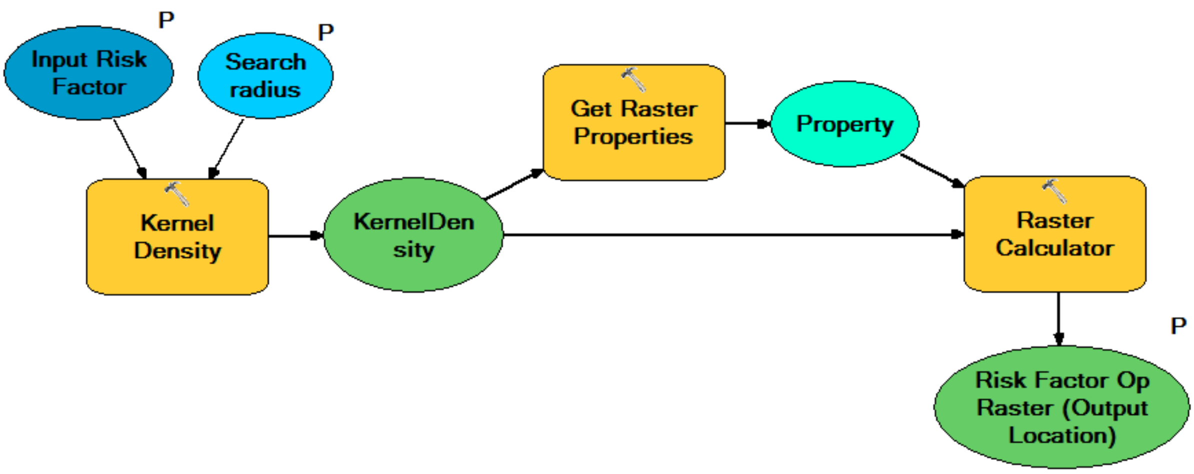

Figure 12 – Model for the operationalization type density

The GUI of the operationalization for density is presented in Figure 14. The input risk factor, the search radius, and the output location for the operationalized risk factor have to be specified.

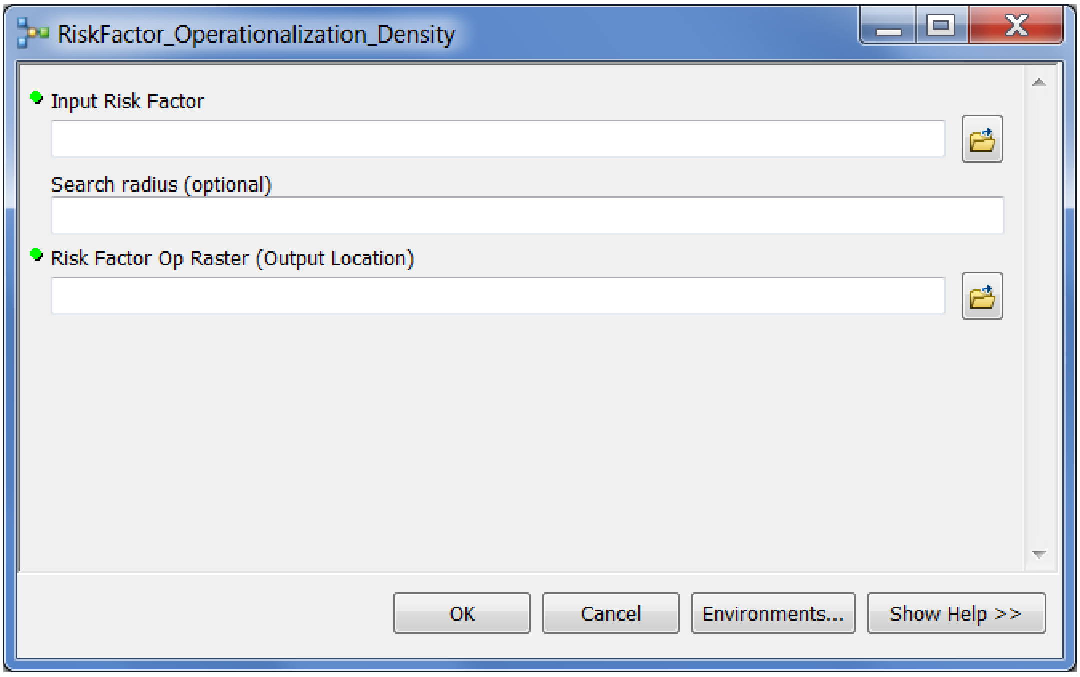

Figure 13 – GUI for the operationalization type density

After all risk factors are operationalized, the separate risk map layers have to be combined. This can be done using the "Raster Calculator" based on the provided formula of the report. The risk values of the risk terrain map are then shown by a raster through stretched values by default.

3.5.6 Finalization and Visualization of the Results

The last step includes the finalization and visualization of the map. In order to compare the risk terrain maps visually, a standardized classification scheme has to be defined. For all produced risk terrain models four classes were created, which can be seen in Table 6. The classification was based on standard deviations whereby the break points were set manually. The first class, "Low Risk" includes all cells with values below the average value. The second class "Medium Risk" includes all cells with values above the mean and below one standard deviation. The third class is named "High Risk" and represents values above one standard deviation to values below two standard deviations. All values that are greater than two standard deviations make up the fourth class, "Highest Risk". The classes are represented by different red tones to show the risk. The class with the lowest risk is not assigned a color, in order to not overload the map, because a large portion of the map would be colored in a light red tone. The other classes are given red tones from light red to dark red.

Table 6 – Classification scheme for the risk values

Class Name	Range of Values	Color
Low Risk	< mean	no color
Medium Risk	> mean to < +1SD	light red (Rose Quartz)
High Risk	> +1SD to < +2SD	medium red (Medium Coral Light)
Highest Risk	> +2SD	dark red (Poinsettia Red)

After the classification, the map can be created using the different base maps, which makes the interpretation for the user easier. All important cartographic elements such as scale bar, north arrow, legend, title, creation date, and data sources are added to the map as well.

3.5.7 Evaluation and Comparison of the Results

The goal of the concluding comparison and evaluation is to show differences between the models and to find out which of the four calculated models for a crime event is the most appropriate one.

First, a visual comparison and evaluation of the predictions for 2013 can be made. Therefore the outcome event data for 2013 are added to the map to see if and how many crime events fall into predicted areas.

In a next step, the numerical comparison, or, more explicitly, the evaluation is done. As described in chapter 2.4, for the evaluation of the results the Predictive Accuracy Index is used. Each produced model is evaluated three times, for the class "Highest Risk", "Highest and High Risk" and for "Highest, High, and Medium Risk" and the results are summarized by two diagrams.

For the calculation of the PAI another model was produced within ArcGIS and thus made it possible to calculate the PAI value automatically. The model was implemented three times, each for the different risk classes, which are defined above. The steps are basically the same, but there are a few differences regarding the extraction of the risk classes. The model for the class "Highest Risk" is shown in Figure 15.

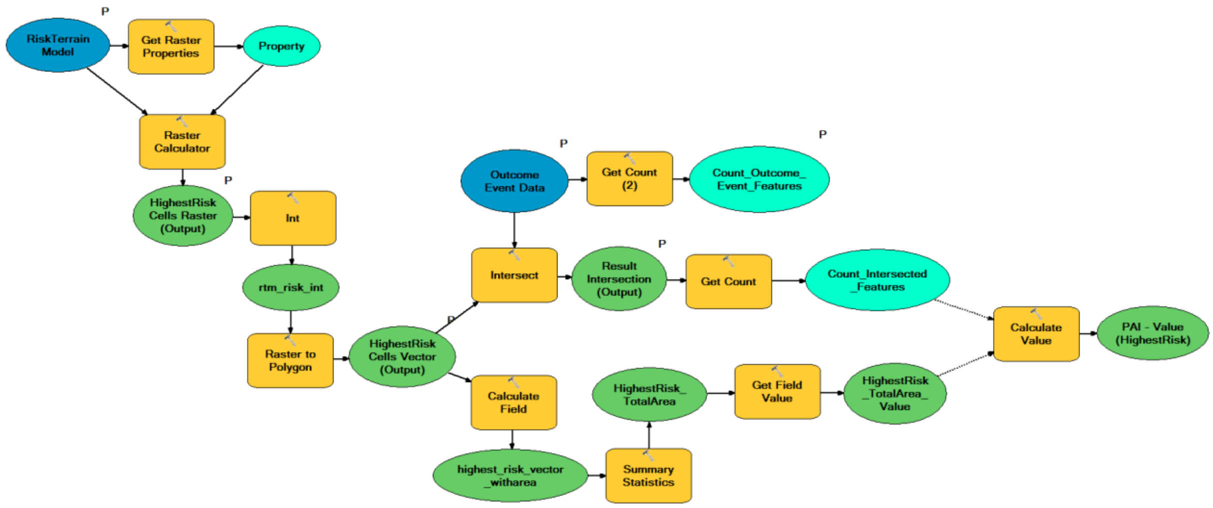

Figure 14 – Model for the evaluation

In a first step, the specified risk terrain model is used to calculate the standard deviation or rather the mean value of the raster. Then this information is used to extract only cells which fall into a specific risk class using the "Raster Calculator". Are the values above the mean, one standard deviation, or two standard deviations, respectively, these cells are assigned the value one, all other cells are set to zero. Then, the raster is converted into integer values so that it can be converted into vector format afterwards. This layer represents the risk cells in vector format which is required for further calculations. For the outcome event data, which are set by the user, the number of features is calculated using the tool "Get Count". After that, the layer with the risk cells is intersected with the outcome events in order to get the number of features which fall into the predicted areas. A further tool has to be used to calculate the size of the predicted areas. Therefore, for a field the size of the area is calculated (see Figure16), based on the Python expression "!shape.area!".

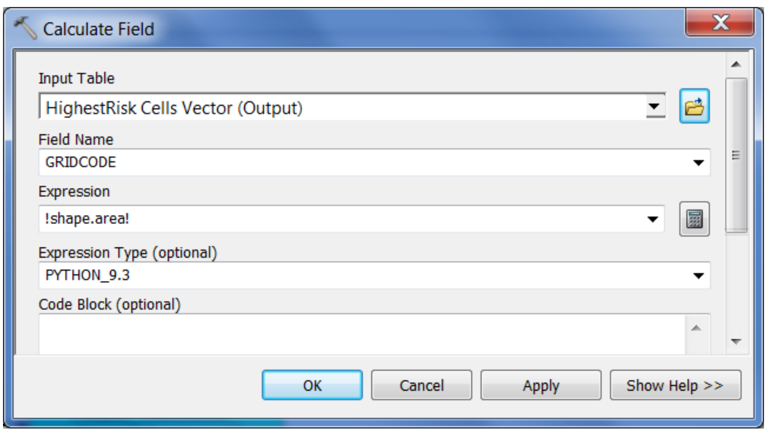

Figure 15 – The tool "Calculate Field"

The tool “Summary Statistics” is used to sum up the field values to calculate the total size. This value is then read out by the tool “Get Field Value”. After that, all required parameters for the formula are available and the PAI value can be calculated. This task is done with the tool “Calculate Value”. Figure 17 presents the final calculation. The expression shows the formula and the separate parameters are read out within the code block. It has to be mentioned that within the formula the multiplication with 100 has to be at the beginning and 100 has to be specified as a double (100.00), otherwise the calculation is not based on doubles and ends in an error.

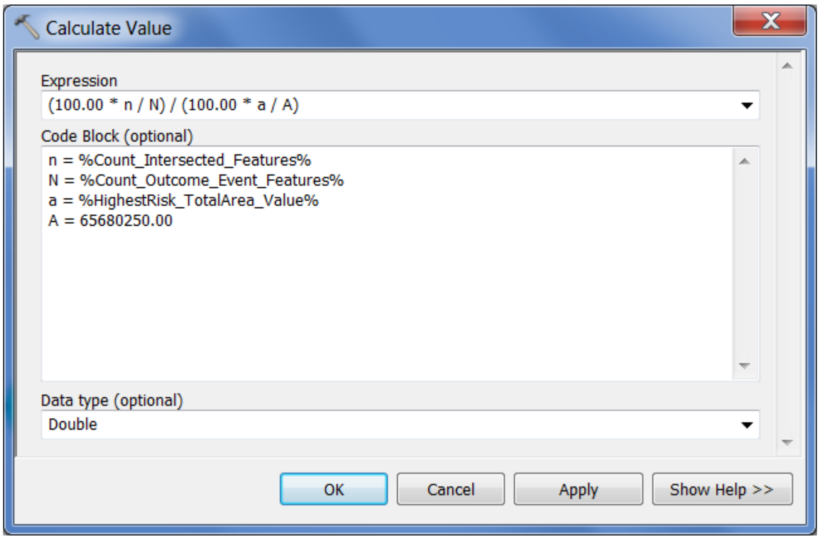

Figure 16 – The tool “Calculate Value”

To calculate the PAI values for a model, the user has to specify the risk terrain model, the outcome event data, and the output locations for the raster and vector representing the risk cells and for the intersection of the risk cells and the outcome event, see Figure 18.

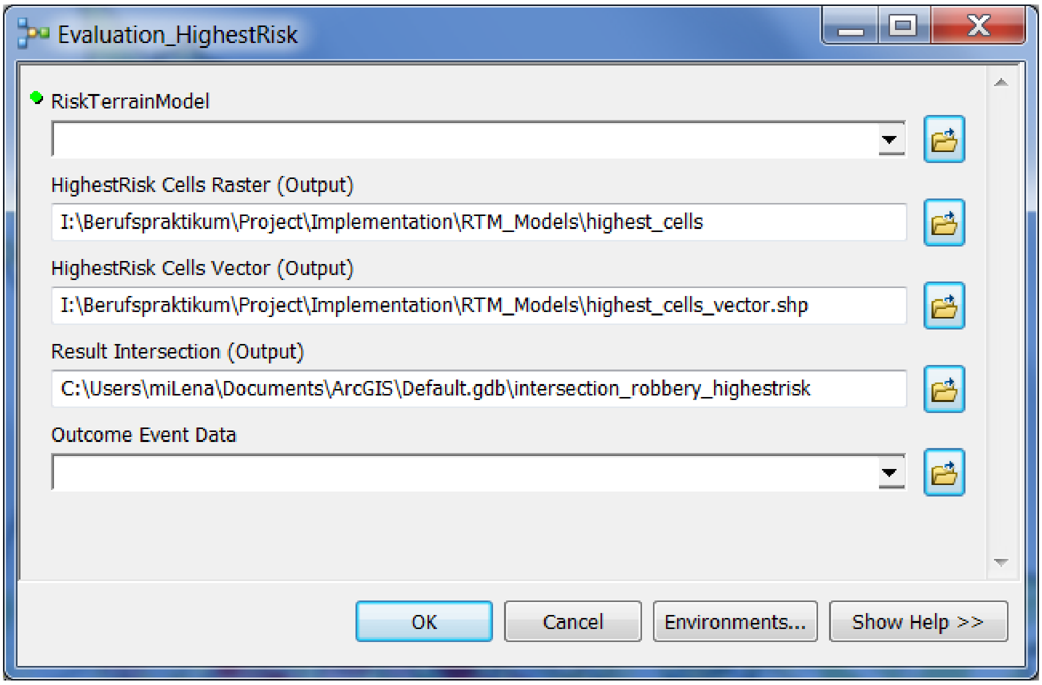

Figure 17 – GUI for the evaluation

In order to compare the results and because of the huge amount of information, two tables and the corresponding charts were created within Microsoft Excel. The first table and chart show the different risk classes and the PAI values for the different block lengths. The second table and chart represent the percentage of correctly predicted crime events in respect to the size of the predicted areas, which is given in square kilometers. This information is useful for the police and other decision-makers.

3.6 Summary

The chapter "Methodology" is divided into several subchapters. At the beginning, the problem definition is given and shows the challenges of this research project. The subchapter "Method of Solution" presents the workflow to solve the defined problems. Subsequently, the project area and the necessary data for the project are discussed. The implementation is described in more detail. This not only includes the selection of crime events and the data capture as well as the data evaluation, but also the detailed process how the risk terrain models were calculated and operationalized. Furthermore, the finalization and visualization of the predictions is described and the way how the results were evaluated and compared is given. The main steps of this chapter are summarized and presented in Figure 19.

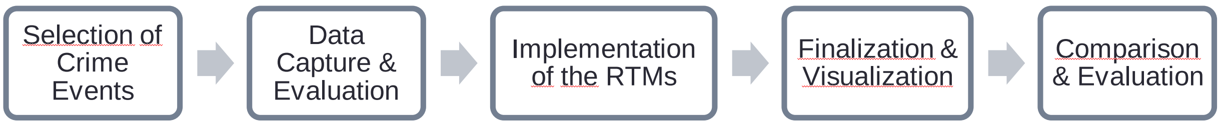

Figure 18 – The main steps of the chapter "Methodology"

4. Results and Interpretation

In this chapter the results and findings of the research project are presented and interpreted. At first, the risk factors for the separate crime events and the identified correlated factors are given. A further main focus lies on the predictions and the evaluation of the risk terrain models.

4.1 Risk Factors

Risk factors for the crime events were identified using the RTM Compendium, which lists and describes the risk factors for 17 different outcome events. Although the risk factors are related to the outcome event, based on the specific settings of the project area, only some of the risk factors might correlate with the outcome event, which is tested within the calculation of the RTMDx Utility. For the four crime events the risk factors regarding offense locations are listed in a table which also shows which data could be obtained. In a map the spatial distribution of the risk factors within the city of Salzburg is presented.

4.1.1 Assault

Table 7 below shows risk factors which are related to the outcome event “Assault”. Assaults have the highest percentage of all crime events reported in the city of Salzburg. More than 1,500, exactly 1,644 assaults were reported in 2012 and 2013 on average. All relevant risk factors cannot be included, because gang activities and drug dealing areas are not that prevalent in the city of Salzburg.

Table 7 – Risk factors of assault

Risk Factors (Compendium)	Obtained data
Bars and Nightclubs	Bars and Pubs Nightclubs
Drug Trade	--
Entertainment Venues	Entertainment Venues
Gang Activity	--
Nightclubs	Nightclubs
Schools and School Property	Schools

Figure 20 shows the spatial distributions of the risk factors. Because the calculation based on a one year prediction did not result in any correlated risk factors, the prediction was made for the seasons, including all available risk factors. Additionally to the risk factors mentioned above further risk factors could be identified, which are also shown in the following map.

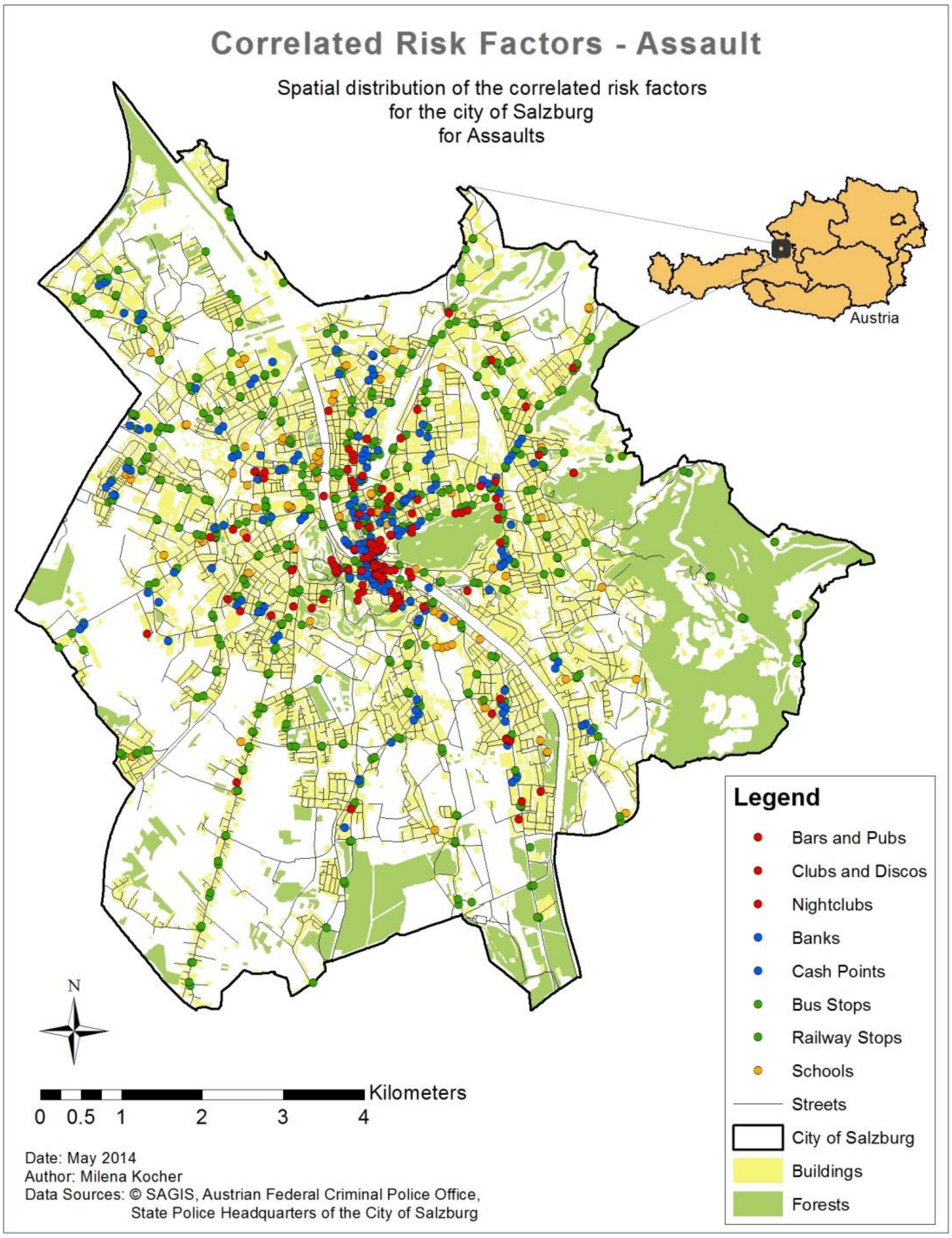

Figure 19 – Correlated risk factors of assault

4.1.2 Auto Theft

The risk factors for the crime event "Auto Theft" are listed in Table 8. 32 auto thefts were reported every year per average in 2012 and 2013. Because the required data was very specific, some datasets could not be obtained and only three of twelve risk factors were available.

Table 8 – Risk factors of auto theft

Risk Factors (Compendium)	Obtained data
Household Income	--
Land-Use Type	Buildings (Residences)

	Official Buildings
Locations of Most Frequently Stolen	--
Locations of Older Vehicles Parked	--
Nighttime Entertainment Venues	Clubs and Discos Nightclubs
Parking Lots	--
Property Value	--
Proximity to Bars	Bars and Pubs
Proximity to High Schools	Schools
Single-family Households	--
Vehicle Availability	--
Vehicles Parked	--

Only the risk factor “Schools” was significantly correlated and its spatial distribution is presented in Figure 21.

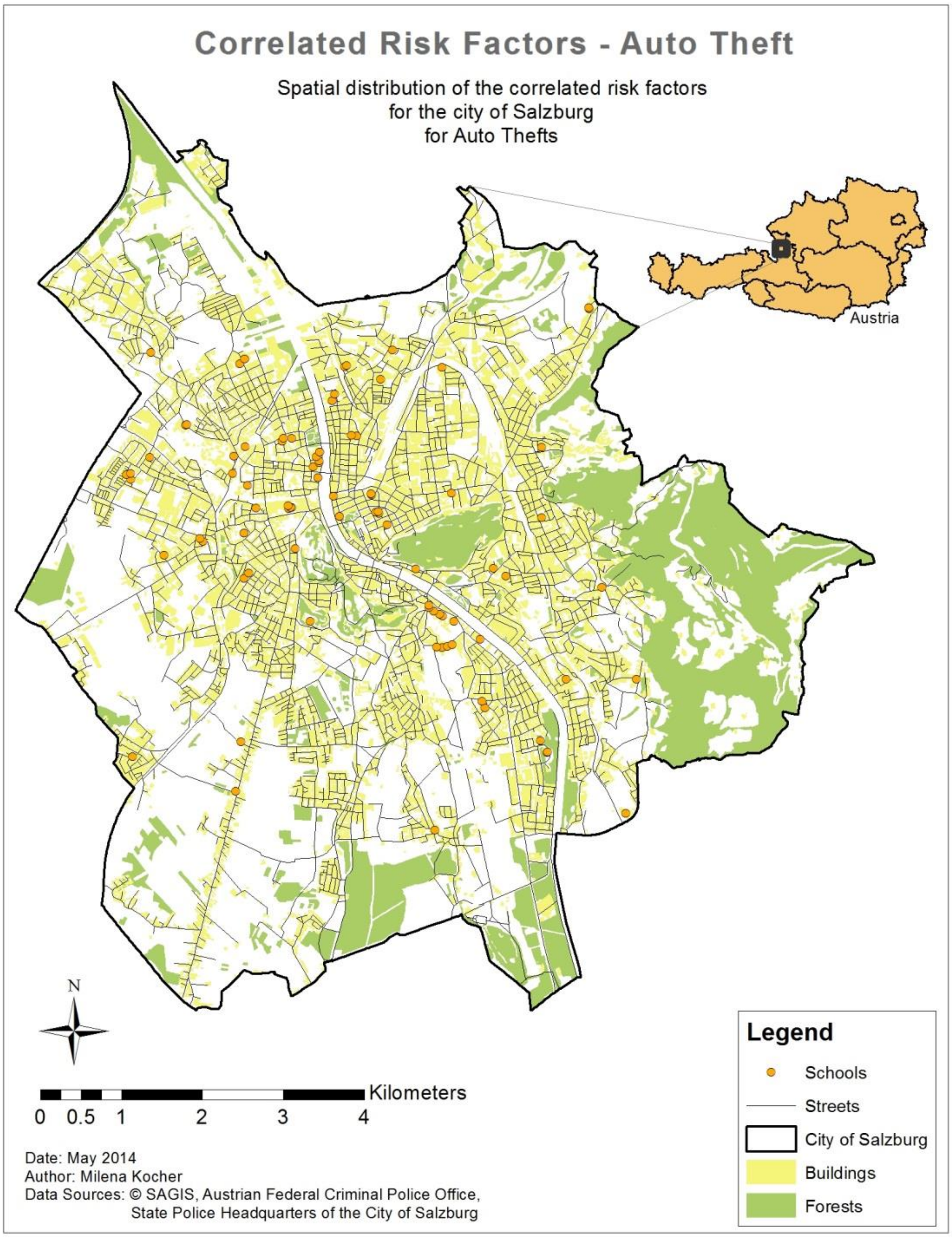

Figure 20 – Correlated risk factors of auto theft

4.1.3 Burglary

Table 9 lists the risk factors which have an influence on "(Residential) Burglary". Because the risk factors are not given for general burglaries, the analyses were made twice. Once it is made for burglaries and the second time only for residential burglaries. In the city of Salzburg, 2,333 burglaries were committed per year in 2012 and 2013, for burglaries into buildings per average 814 burglaries were reported.

Table 9 – Risk Factors of burglary

Risk Factors (Compendium)	Obtained data
Land Use Type (residential)	Buildings (Residences) Official Buildings
Measures of Social Disorganization	--
Proximity to Pawn Shops	Pawn Shops
Proximity to Police Stations	Police Departments
Proximity to Public Transportation	Bus Stops Railway Stops

The spatial distribution of burglaries is shown in Figure 22. Very dominant are the residences. All areas without residences can be excluded, at least for the analysis done for "Residential Burglary".

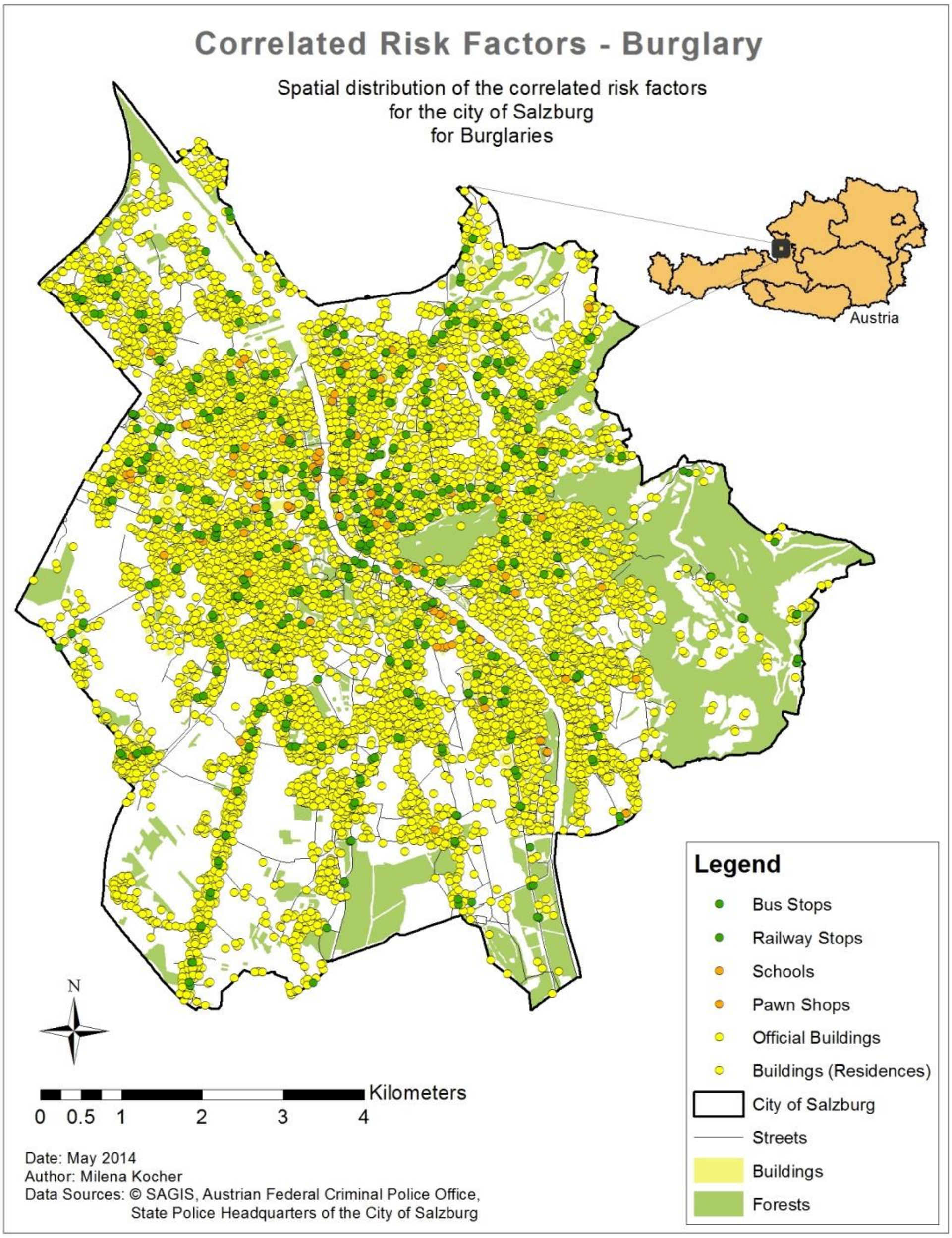

Figure 21 – Correlated risk factors of burglary

4.1.4 Robbery

The risk factors for the crime event "Robbery" are different depending on the fact whether the victim has already committed a crime offense or not. The first option "Risks in Relation to Criminal Victim Selection" assumes that persons with a criminal background are good victims because the chance that they report the crime offense is not that high. For that type the two risk factors "Proximity to Drug Dealing Areas" and "Proximity to Areas with High Prostitution" are listed. Because none of these risk factors can be identified for the city of Salzburg, the analysis is only done for the second option. This is named "Risks in Relation to Non-Criminal Victim Selection" and is based on the assumption that law-abiding persons are good victims because they might not be as dangerous and aggressive as persons with a criminal background. All six risk factors, presented in Table 10 could be obtained. On average, 79 robberies per year were committed in the city of Salzburg in 2012 and 2013.

Table 10 – Risk factors of robbery

Risk Factors (Compendium)	Obtained data
Proximity to Banks and Cash Points	Banks Cash Points
Proximity to Bars, Pubs and Exotic Clubs	Bars and Pubs
Proximity to Leisure and Fast-Food Outlets	Leisure and Fast-Food Outlets
Proximity to Post Offices	Post Offices
Proximity to Public Transport	Bus Stops Railway Stops
Proximity to Schools	Schools

The correlated risk factors of the outcome event "Robbery" are presented in Figure 23.

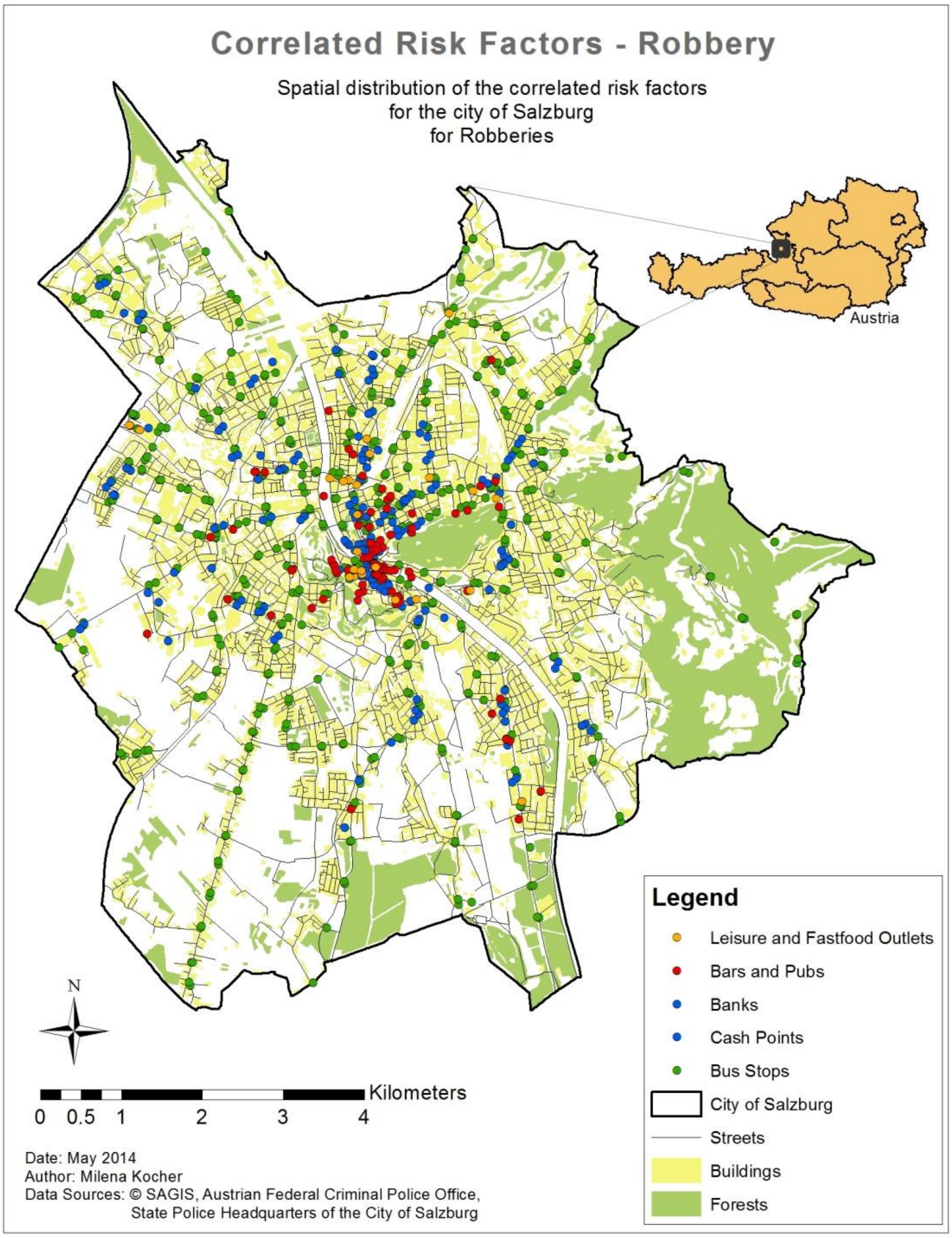

Figure 22 – Correlated risk factors for robbery

4.2 Predictions and Evaluation of the Results

In this subchapter, the predictions of the calculated risk terrain models and the evaluation are presented. First, the result of the RTMDx Utility report and the prediction for 2013 are discussed. Because of the relative high number of calculated models and in order that the structure did not change, only the best out of the four models is described. The other prediction results can be found in the appendix. Furthermore, the evaluation for 2013 which contains two tables and two charts is shown and interpreted. Based on the result of the evaluation the prediction for 2014 can be done and is presented, as well.

4.2.1 Assault

First, the models for "Assault" were calculated based on a one year time period, but it resulted in an error message, see Figure 24. No risk factors could be found which significantly correlated with the outcome event data. Also the calculation with all available risk factors was not successful.

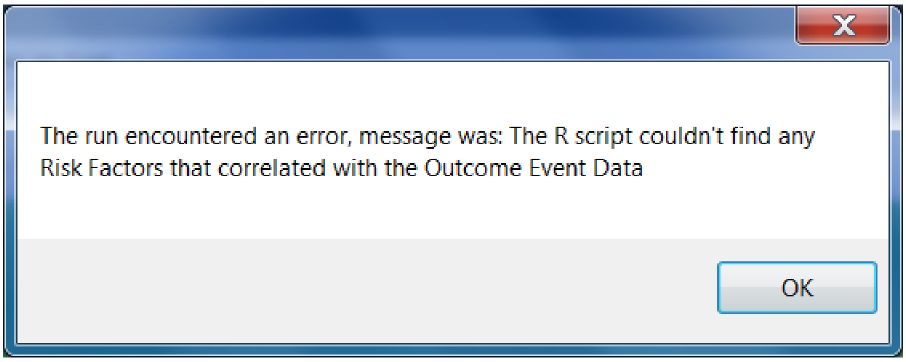

Figure 23 – RTMDx Utility error message

In a next step, models were calculated for all four seasons, four times each, to include the different spatial influences. The models were calculated using all eligible risk factors. In fact, the following were added: Banks, Cash Points, Clubs and Discos, Leisure and Fast-Food Outlets, Bus Stops, Railway Stops and Pawn Shops. For spring and summer, results for all four different spatial influences could be calculated. For fall and winter, no risk factors were identified that correlated with the outcome event, and the same error message as shown in Figure 24 was shown.

The best model for the prediction of assaults for spring 2013 could be calculated with a spatial influence of one block (110 meters). The result of the report shows the correlated risk factors, their operationalization type, the spatial influence, and the weight which is equivalent to the relative influence of each risk factor. Five risk factors were found which are presented in Table 11.

Table 11 - RTMDx report for assault for spring 2013

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bars and Pubs	Density	110	1.4810	4.3973
Rate	Railway Stops	Proximity	110	1.4170	4.1247
Rate	Cash Points	Proximity	110	1.2585	3.5201
Rate	Bus Stops	Proximity	110	1.1922	3.2943
Rate	Nightclubs	Proximity	110	1.0705	2.9168
Rate	Intercept	--	--	-4.7359	--
Overdispersion	Intercept	--	--	0.6018	--

The finalized risk terrain map for the prediction of assaults for spring 2013 is presented in Figure 25. The map shows the city of Salzburg with the most important base layers including forests, rivers, lakes, buildings, and residences. The risk values define the value how high the risk is and are visualized in different red tones. When adding the actual assaults for spring 2013, the resulting map can be evaluated visually. It can be seen that almost all crime offenses fall into a risk class, many of them even in the class "Highest Risk". The other block lengths resulted in similar predictions, but with a few differences.

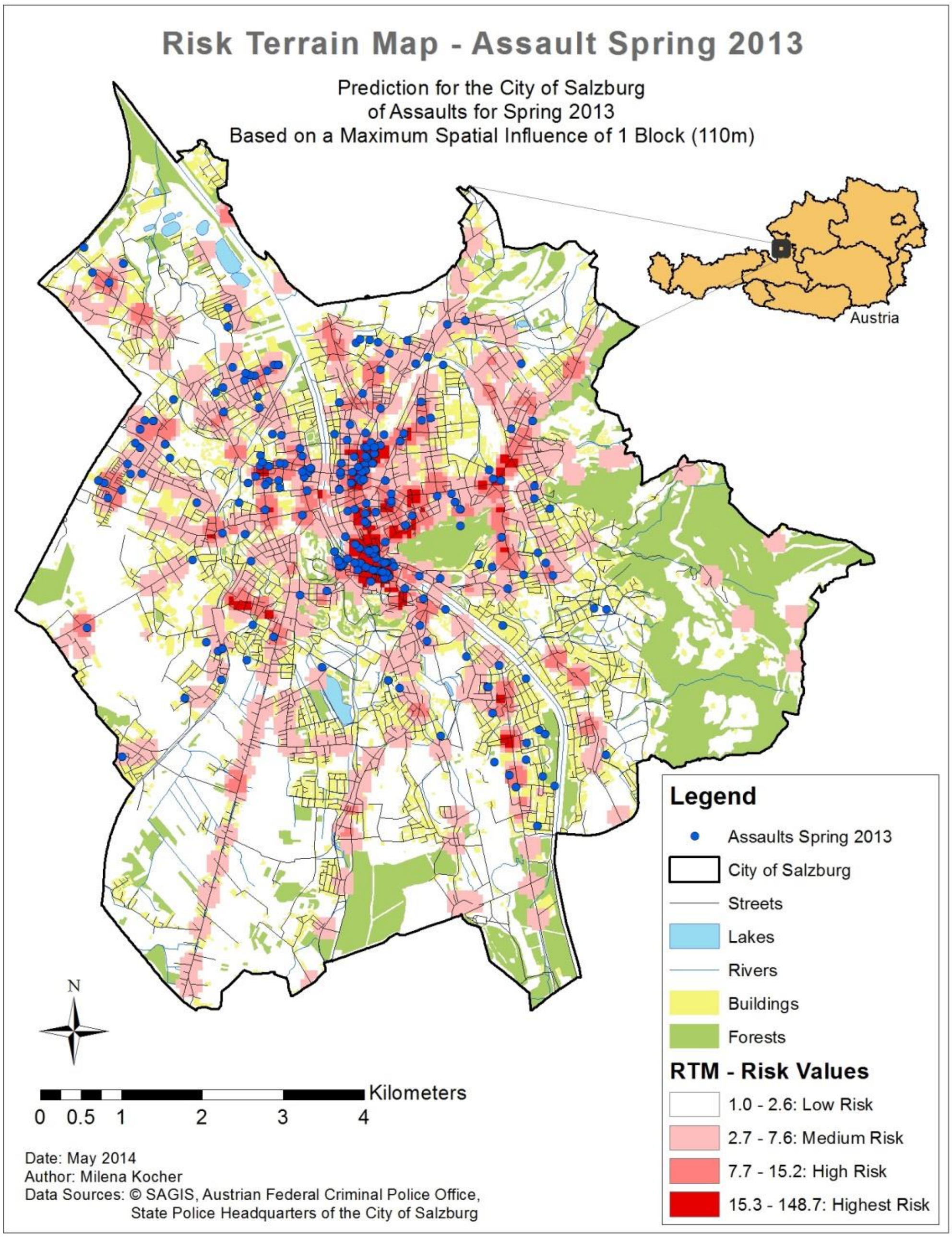

Figure 24 – Prediction of assaults for spring 2013 (1 Block)

After the prediction, the results could be evaluated. The first evaluation shows the different risk classes and the calculated PAI value for the different block lengths, which can be seen in Table 12 and Figure 26. The PAI values range from 2.99 up to 31.37, which is the highest PAI value. Except the highest value, all other values are relatively similar. Based on this information, the model calculated for one block represents the best model.

Table 12 – Evaluation for assault spring 2013

Assault Spring 2013

	Risk Class		
	Highest & High &	Highest & High	Highest
PAI - 1 Block	2.99	9.27	31.37
PAI - 2 Blocks	4.21	6.11	16.08
PAI - 3 Blocks	3.73	9.69	16.26
PAI - 4 Blocks	3.30	9.32	16.37

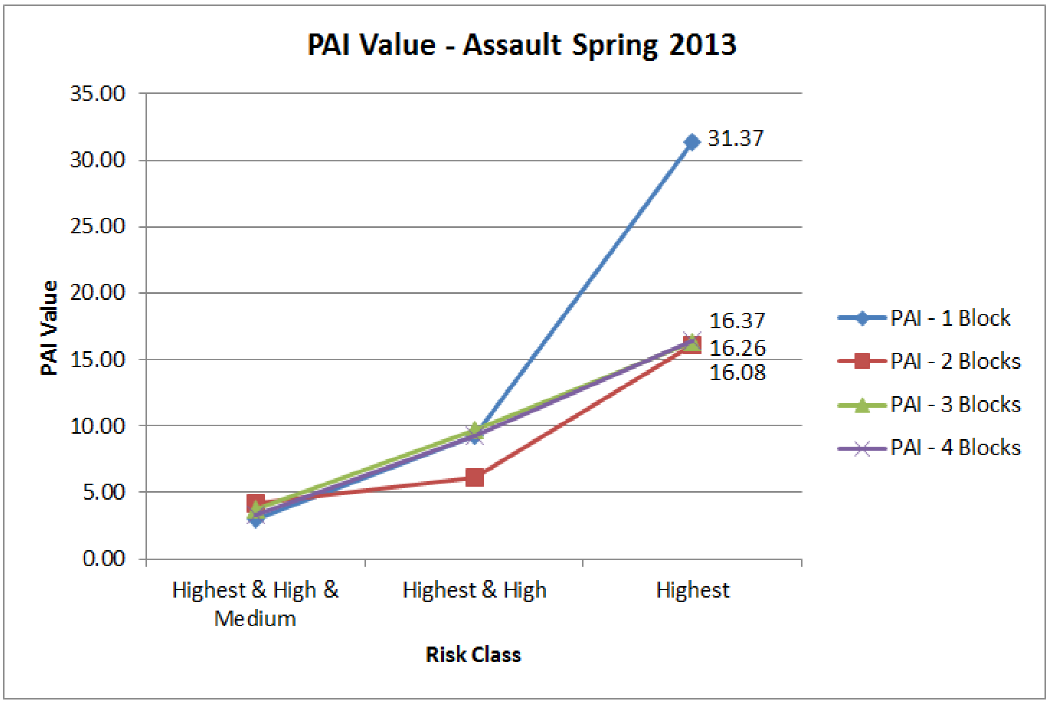

Figure 25 – Evaluation for assault spring 2013

The next evaluation shows the correctly predicted assaults related to the size of the predicted areas (given in square kilometers). The results are presented for the risk class "Highest Risk". The model for one block correctly predicted 37% of the actual assaults in 2013 for an area of 0.78km². A similar result regarding the percentage could be calculated for four blocks; in fact 39% were predicted correctly, but for an area of 1.5 km². The models for two and three blocks correctly predicted 47% and 48%, respectively, for an area of 1.91 and 1.96 km². This result shows that the higher the percentage of correctly predicted crime events, the bigger the area gets. Based on how much area can be covered by the police or other decision-makers, this result presents useful information and can be seen in Table 13 and Figure 27.

Table 13 – Correctly predicted assaults for spring 2013

Assault Spring 2013

	Correctly Predicted Assaults Spring 2013 (Highest Risk)		Area in km²
1 Block	153 of 412	37.14%	0.78
2 Blocks	198 of 412	48.06%	1.96
3 Blocks	195 of 412	47.33%	1.91
4 Blocks	159 of 412	38.59%	1.55

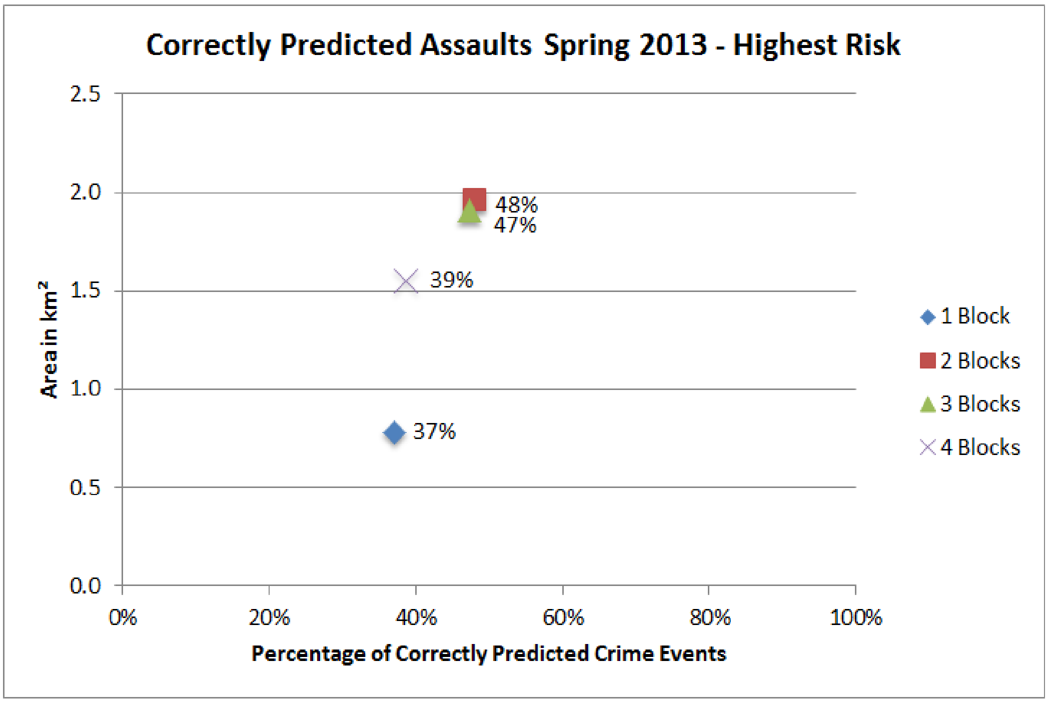

Figure 26 – Correctly predicted assaults for spring 2013

The prediction for 2014 was based on a spatial influence of the risk factors of one block. For the best model, which is shown in Table 14, some different risk factors were identified which is because the outcome event data were from 2013 and not from 2012 anymore. That supports the assumption that the environment is dynamic and changes over time.

Table 14 – RTMDx report for assaults for spring 2014

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Clubs and Discos	Density	110	1.3150	3.7248
Rate	Cash Points rightCS	Proximity	110	1.2556	3.5099
Rate	Nightclubs	Proximity	110	1.2366	3.4439
Rate	Bars and Pubs	Proximity	110	1.2339	3.4346
Rate	Bus Stops	Proximity	110	0.8656	2.3764
Rate	Intercept	--	--	-4.6333	--
Overdispersion	Intercept	--	--	0.6593	--

The following map, see Figure 28, shows the prediction of assaults for spring 2014. The size for the class "Highest Risk" is 1.2km², for "Highest and High Risk" it is 1.5km², and when including the medium risk class too, it is 17.6km².

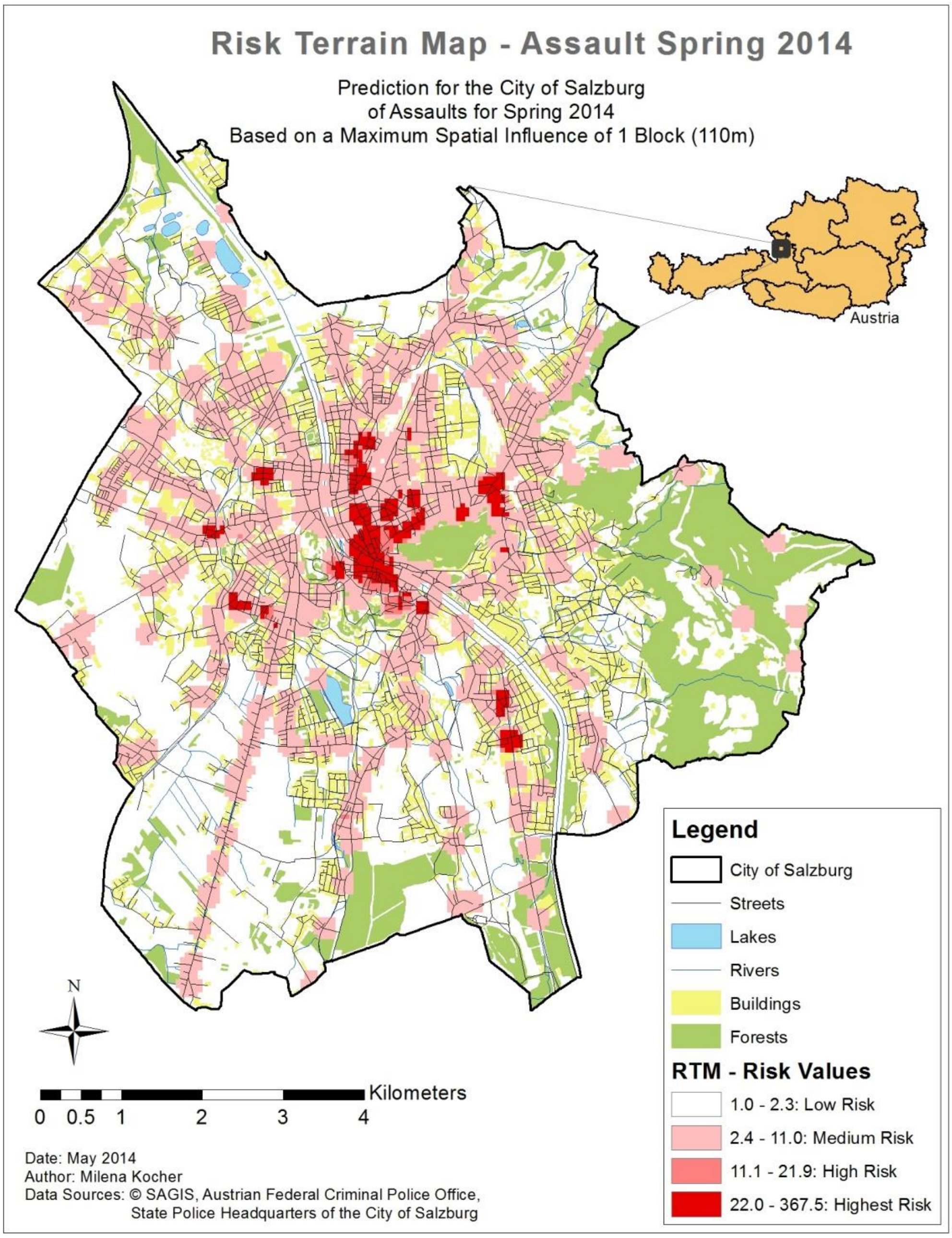

Figure 27 – Prediction of assaults for spring 2014

Furthermore, the model was calculated for the season summer, whereby the best model could be calculated with a spatial influence of one block (110 meters). Table 15 shows the best model, which includes five risk factors.

Table 15 – RTMDx report for assault for summer 2013

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bars and Pubs	Proximity	110	1.5125	4.5381
Rate	Bus Stops	Proximity	110	1.1798	3.2537
Rate	Cash Points	Proximity	110	0.9925	2.6979
Rate	Clubs and Discos	Proximity	110	0.9841	2.6753
Rate	Nightclubs	Proximity	110	0.8574	2.3570
Rate	Intercept	--	--	-4.7364	--
Overdispersion	Intercept	--	--	0.7739	--

The map which shows the prediction as well as the actual assaults for 2013 is presented in Figure 29. Many of the actual assaults in 2013 are within predicted areas. The predictions based on a spatial influence of two, three, and four blocks are given in the appendix.

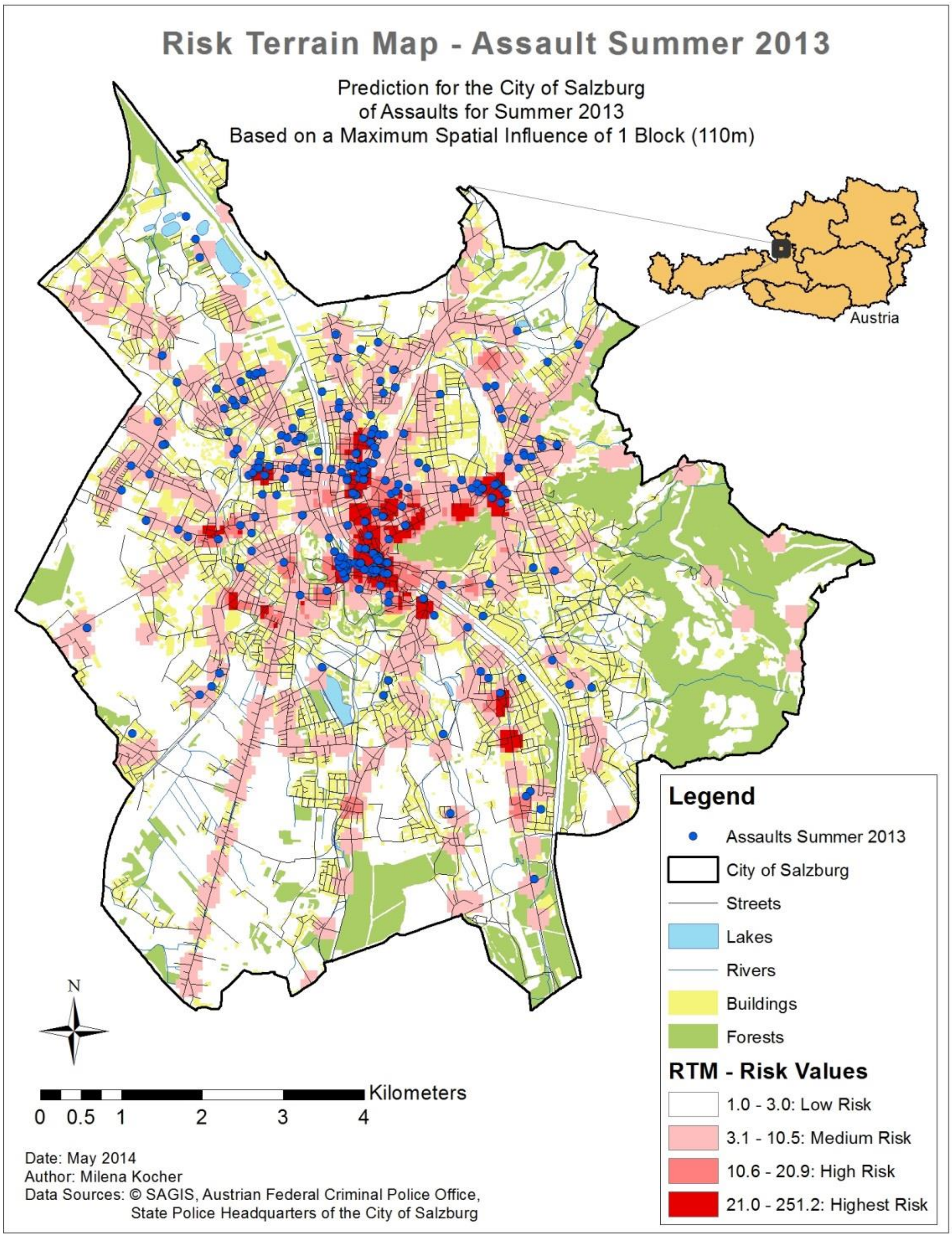

Figure 28 – Prediction of assaults for summer 2013 (1 Block)

The highest PAI value is 23.4 and could be calculated for the model based on one block. The results for three and four blocks are identical and therefore, are only once operationalized and finalized. This information is seen in Table 16 and Figure 30.

Table 16 – Evaluation for assault for summer 2013

Assault Summer 2013

	Risk Class		
	Highest & High &	Highest & High	Highest
PAI - 1 Block	3.34	13.54	23.40
PAI - 2 Blocks	4.25	6.52	22.12
PAI - 3 Blocks	4.21	4.38	18.44
PAI - 4 Blocks	4.21	4.38	18.44

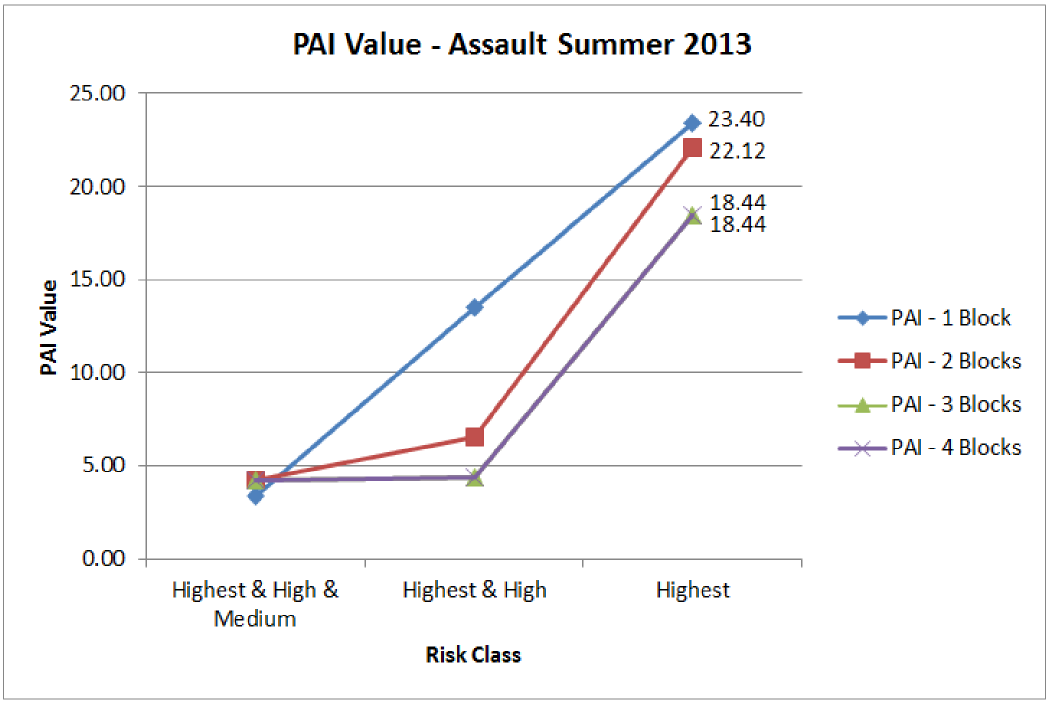

Figure 29 – Evaluation for assault for summer 2013

The correctly predicted crime events are presented in Table 17 and Figure 31. What is striking is the fact that the percentage of correctly predicted crime events and the area is similar for all four block lengths. All values are between 44% and 1.25km² and 49% and 1.73km².

Table 17 – Correctly predicted assaults for summer 2013

Assault Summer 2013

	Correctly Predicted Assaults Summer 2013 (Highest Risk)		Area in km²
1 Block	198 of 446	44.39%	1.25
2 Blocks	209 of 446	46.86%	1.39
3 Blocks	217 of 446	48.65%	1.73
4 Blocks	217 of 446	48.65%	1.73

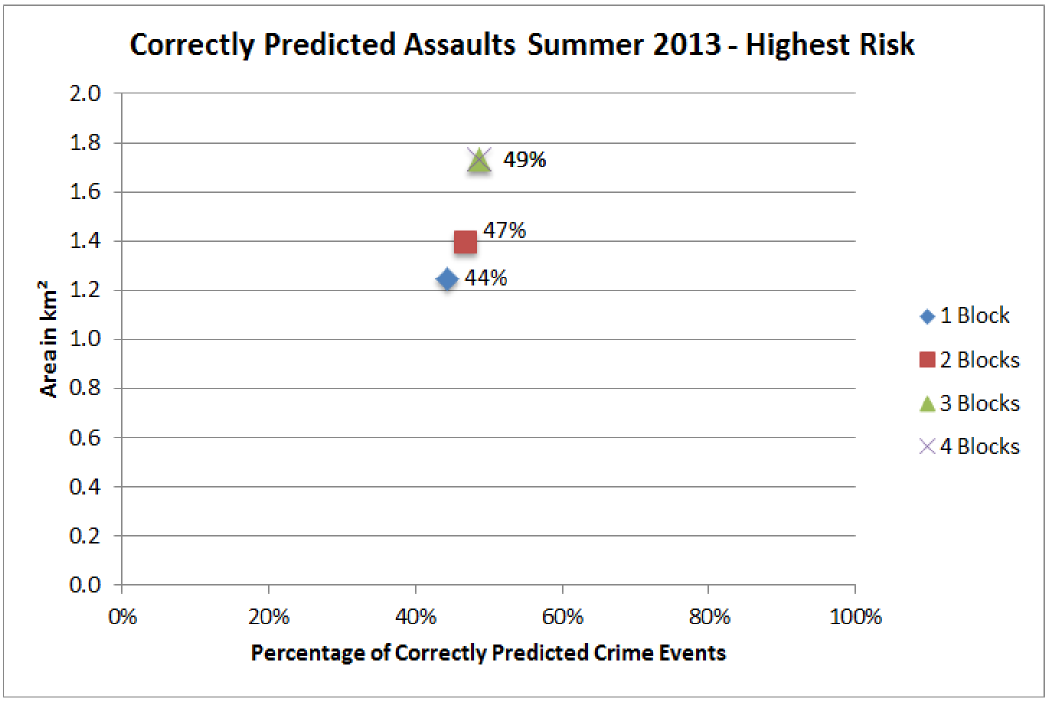

Figure 30 – Correctly predicted assaults for summer 2013

Hence, the prediction for assault for summer 2014 was based on a spatial influence of the risk factors of one block. As Table 18 shows, the best model includes two additional risk factors.

Table 18 – RTMDx report for Assault for summer 2014

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bars and Pubs	Proximity	110	1.6663	5.2925
Rate	Railway Stops	Proximity	110	1.6033	4.9694
Rate	Clubs and Discos	Proximity	110	1.3402	3.8198
Rate	Nightclubs	Density	110	1.1275	3.0879
Rate	Bus Stops	Proximity	110	1.0265	2.7913
Rate	Cash Points	Proximity	110	0.9347	2.5464
Rate	Schools	Proximity	110	0.8131	2.2548
Rate	Intercept	--	--	-4.7683	--
Overdispersion	Intercept	--	--	0.8886	--

The prediction for summer 2014 of assaults shows relatively small areas (see Figure 32). The class "Highest Risk" has a size of 1.5km², the highest and the high risk class have a size of 1.7km² and with the addition of the medium risk class the size is 7.4km².

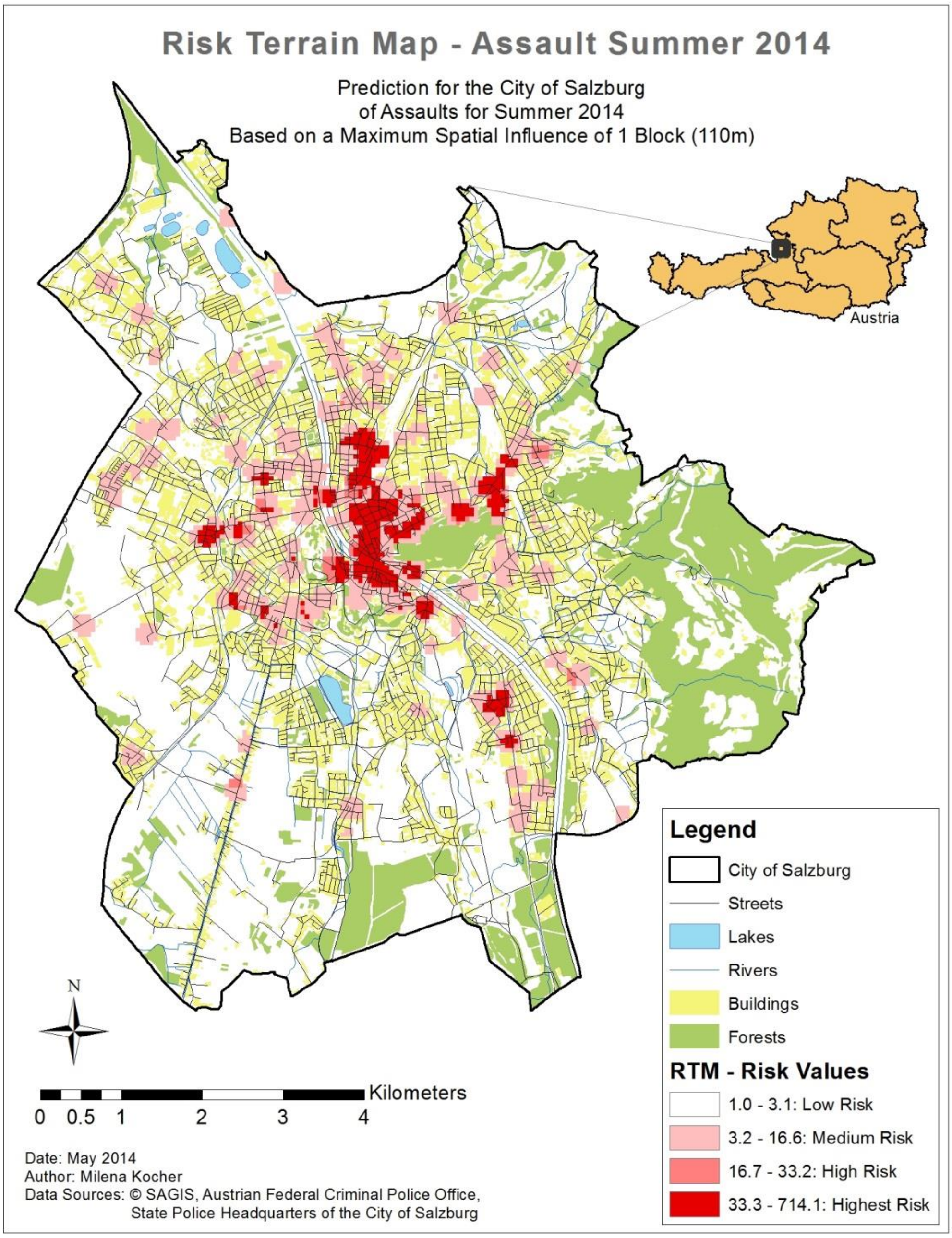

Figure 31 – Prediction for assaults for summer 2014

4.2.2 Auto Theft

The calculated model for auto theft, based on a spatial influence of one block resulted in the error message that no risk factors correlated. For the other block lengths, models could be calculated, but included only one correlated risk factor, due to the fact that only three of twelve risk factors could be obtained (see chapter 4.1.2). Furthermore, the model for three and four blocks was the same.

The prediction for auto thefts for 2013 is presented in Table 19, based on a spatial influence of two blocks. Only the risk factor “Schools” was identified to be significantly correlated.

Table 19 – RTMDx report for auto theft for 2013

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Schools	Proximity	220	1.6713	5.3191
Rate	Intercept	--	--	-7.0471	--
Overdispersion	Intercept	--	--	-1.9058	--

Because only one risk factor was included, it was not possible but also not necessary to make the combination, because only one risk map layer was created. Because of that, the risk values are presented in binary format, showing its presence or absence (see Figure 33). It is apparent that the size of the predicted areas is very big compared to the crime offenses and not many of the outcome events are within the predicted areas.

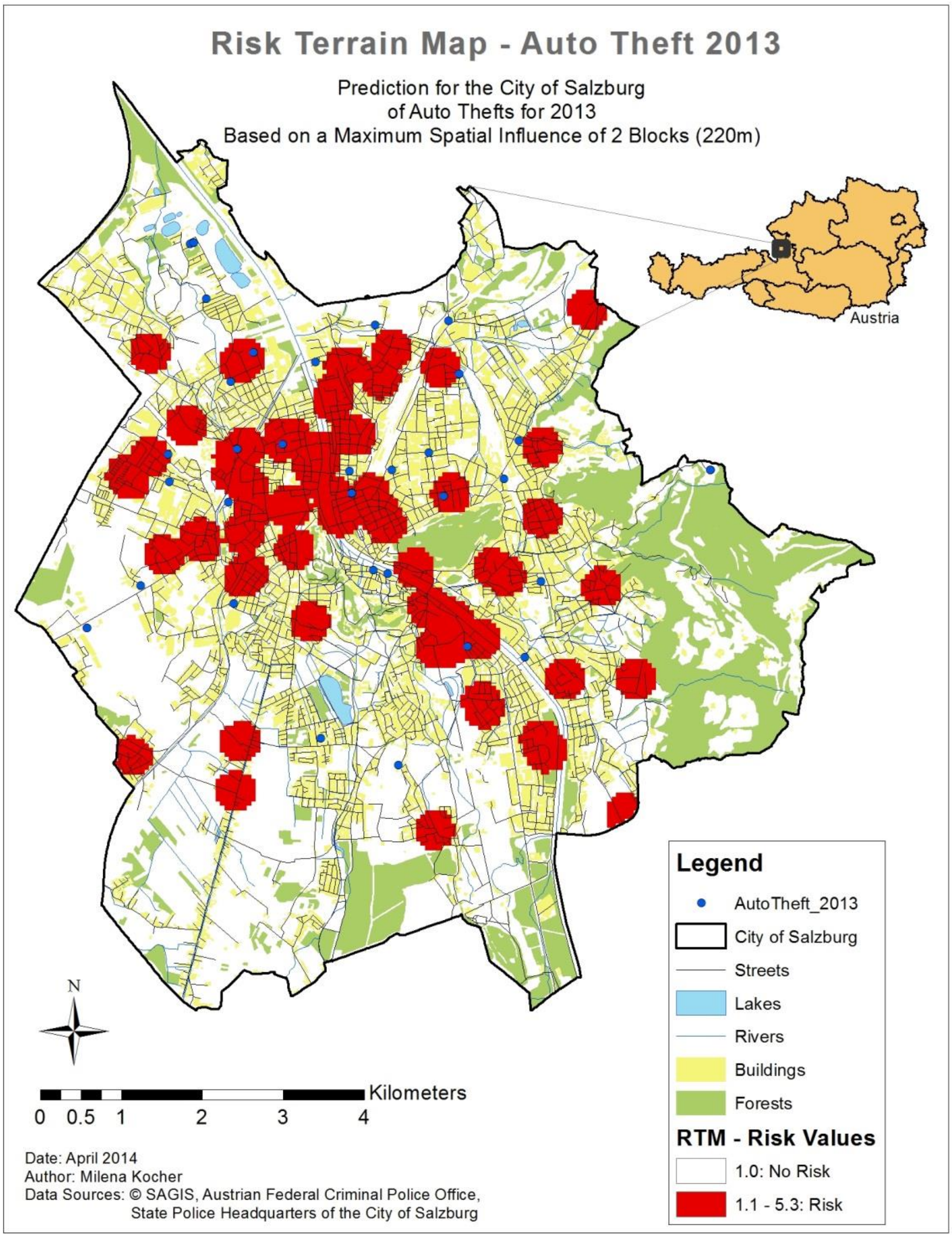

Figure 32 – Prediction of auto thefts for 2013 (2 Blocks)

The evaluation could also only be done for the one existing risk class. The highest PAI value was 1.71, calculated for two blocks (see Table 20 and Figure 34).

Table 20 – Evaluation for auto theft for 2013

Auto Theft 2013

	Risk
PAI - 1 Block	
PAI - 2 Blocks	1.71
PAI - 3 Blocks	1.40
PAI - 4 Blocks	1.40

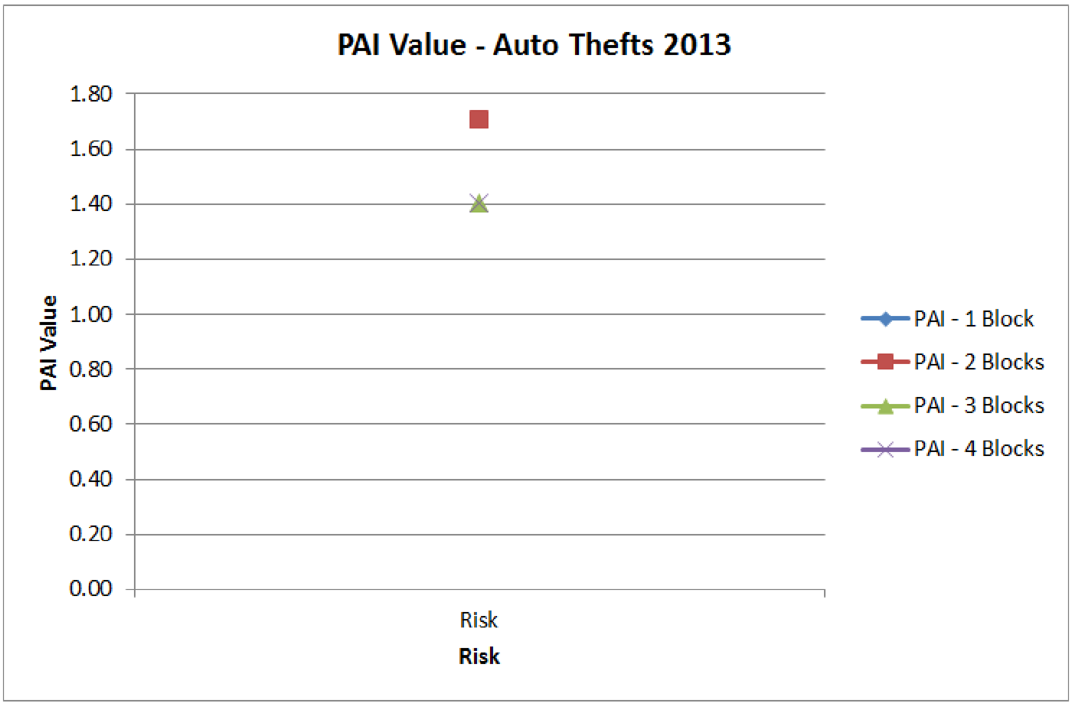

Figure 33 – Evaluation for auto theft for 2013

In the following table and chart (Table 21 and Figure 35) it is obvious that the prediction for auto thefts is not very useful. Less than 30% of the actually happened auto thefts could be predicted correctly, and the area is nearly 10km² or larger.

Table 21 – Correctly predicted auto thefts for 2013

Auto Theft 2013

	Correctly Predicted Auto Thefts 2013 (Risk)		Area in km²
1 Block			
2 Blocks	9 of 35	25.71%	9.90
3 Blocks	10 of 35	28.57%	13.37
4 Blocks	10 of 35	28.57%	13.37

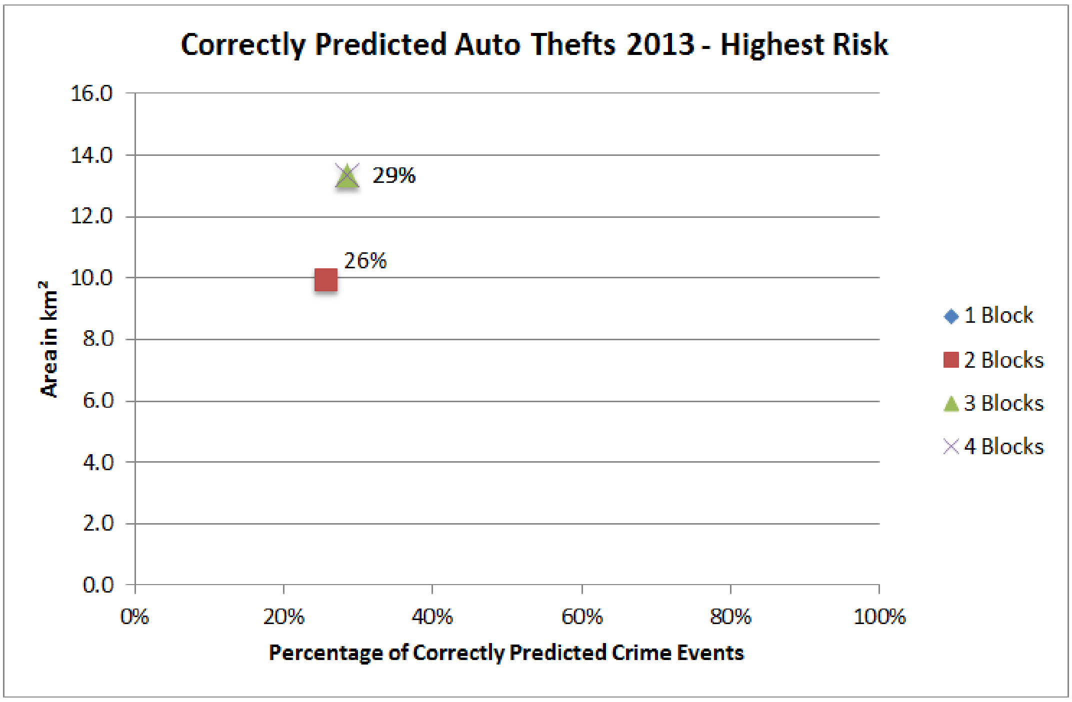

Figure 34 – Correctly predicted auto thefts for 2013

The prediction for 2014 was calculated with a spatial influence of two blocks, but no result could be calculated. That shows again, that it is not possible to make a precise prediction with few available risk factors. A further reason can be that the obtained risk factors would have a small influence on the outcome event in any case and the important risk factors were not included.

4.2.3 Burglary

For the crime event burglary, the four different models were calculated twice. At first, all crime subtypes of the dataset "Burglaries" were included and in a second step only selected ones, which are related to burglaries into buildings, were used for the calculations.

Table 22 shows the result for the calculation of burglaries based on one block. Four different risk factors were identified to be correlated.

Table 22 – RTMDx report for burglary (all) for 2013

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Pawn Shops	Proximity	110	1.4526	4.2742
Rate	Bus Stops	Proximity	110	1.2807	3.5992
Rate	Official Buildings	Proximity	110	1.0231	2.7818
Rate	Schools	Proximity	110	0.7873	2.1975
Rate	Intercept	--	--	-2.7854	--
Overdispersion	Intercept	--	--	0.4129	--

In the map shown in Figure 36, the predicted areas as well as the actual burglaries are presented. Many crime offenses are visible and although the predicted areas seem to be bigger, they are covered by the crime event data.

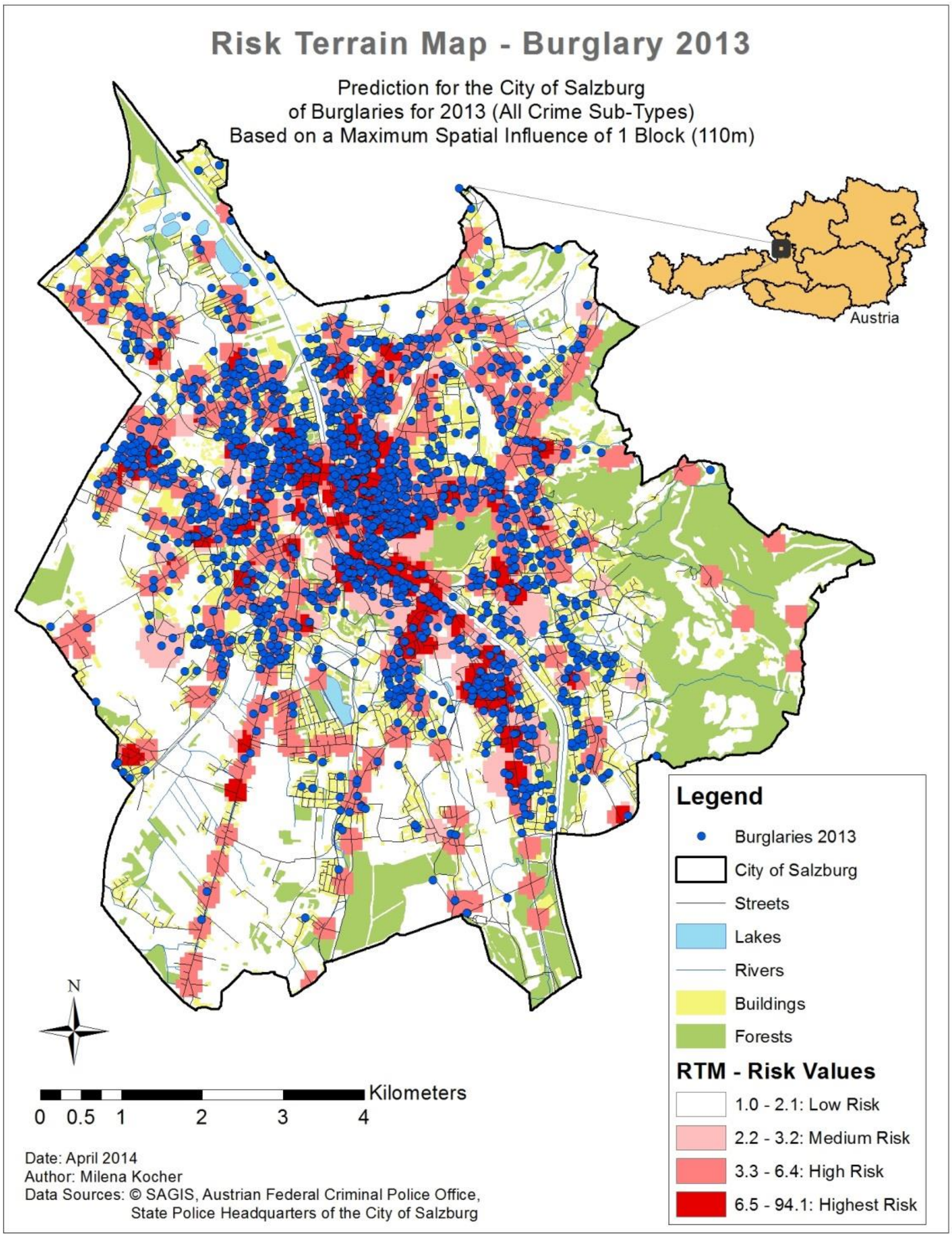

Figure 35 – Prediction of all burglaries for 2013 (1 Block)

The evaluation shows that the highest PAI value is 4.46 and was calculated with a maximum spatial influence of one block (see Table 23). In Figure 37 it is visible, that the PAI values are all quite similar.

Table 23 – Evaluation for all Burglaries for 2013

Burglary 2013 (All Crime Sub-Types)

	Risk Class		
	Highest & High & Medium	Highest & High	Highest
PAI - 1 Block	2.39	2.52	4.46
PAI - 2 Blocks	2.16	2.27	3.40
PAI - 3 Blocks	2.02	2.02	4.35
PAI - 4 Blocks	2.24	2.91	4.11

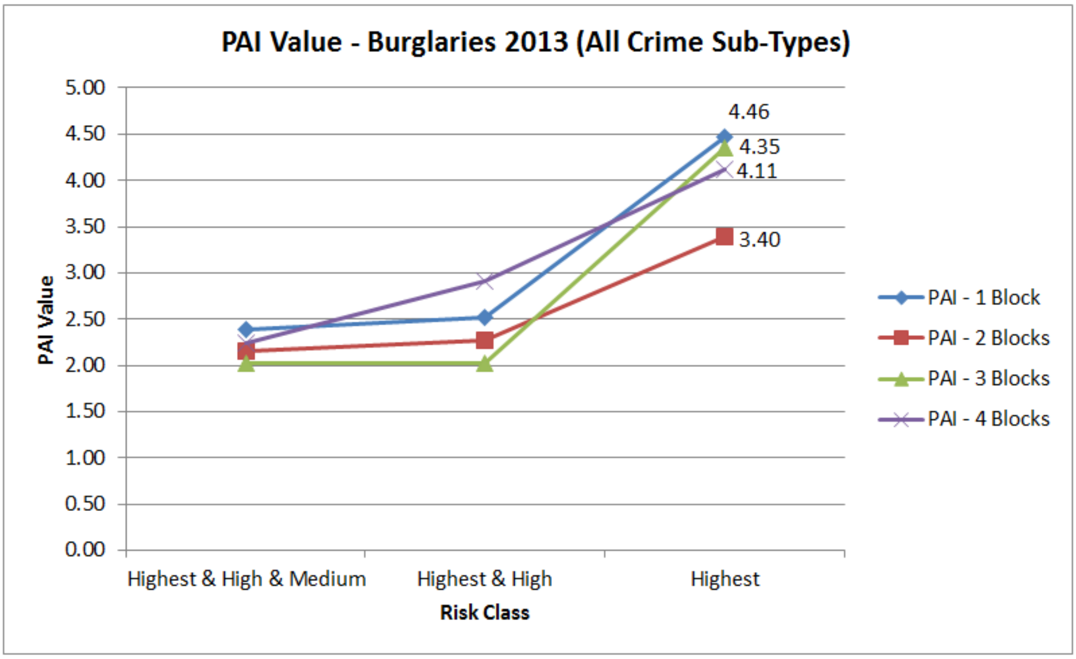

Figure 36 – Evaluation for all burglaries for 2013

Table 24 and Figure 38 show that the correctly predicted burglaries, including all crime sub-types, are between 25% and 44% with areas between 3.7 and 8.6 km².

Table 24 – Correctly predicted burglaries (all) for 2013

Burglary 2013 (All Crime Sub-Types)

	Correctly Predicted Burglaries 2013 (all) (Highest Risk)		Area in km²
1 Block	629 of 2473	25.43%	3.74
2 Blocks	1101 of 2473	44.52%	8.61
3 Blocks	855 of 2473	34.57%	5.22
4 Blocks	1035 of 2473	41.85%	6.69

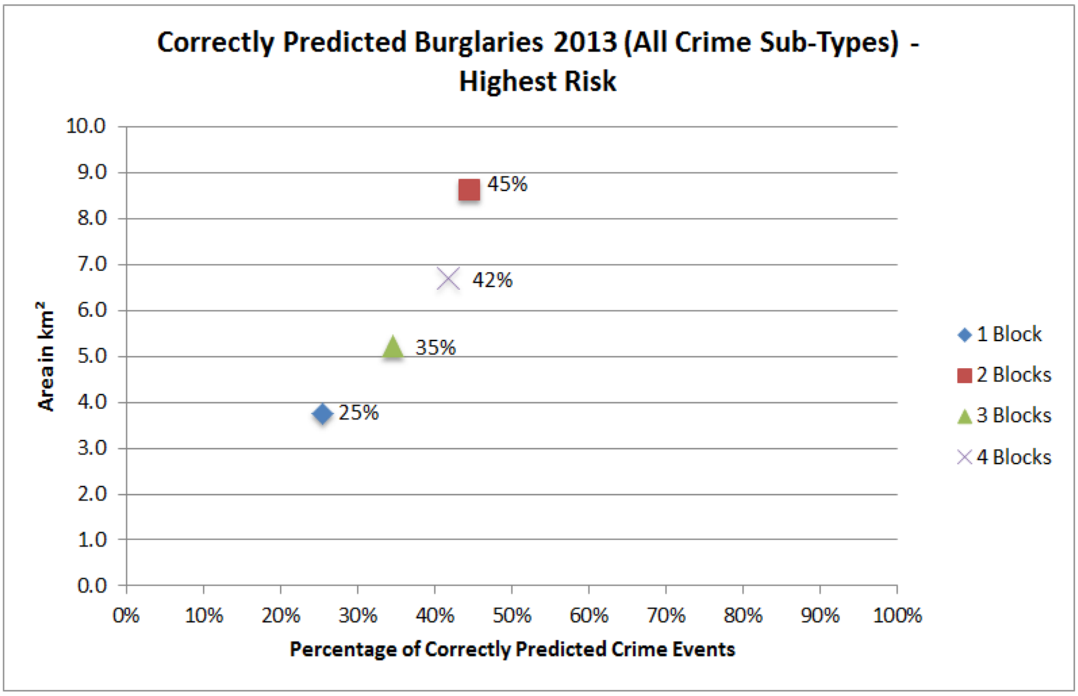

Figure 37 – Correctly predicted burglaries (all) for 2013

The prediction for 2014 was based on a maximum spatial influence of one block. As Table 25 shows, six different risk factors have an influence on burglaries for 2014.

Table 25 – RTMDx report for all burglaries for 2014

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Pawn Shops	Proximity	110	1.2993	3.6667
Rate	Official Buildings	Proximity	110	1.1895	3.2854
Rate	Bus Stops	Proximity	110	1.0991	3.0015
Rate	Schools	Proximity	110	0.7204	2.0553
Rate	Railway Stops	Proximity	110	0.6844	1.9825
Rate	Buildings Residences 1Block	Density	110	0.6270	1.8719
Rate	Intercept	--	--	-2.7016	--
Overdispersion	Intercept	--	--	0.4573	--

The prediction for 2014 is presented in Figure 39. It is particularly interesting to note, that there are a lot of small predicted areas, and many areas are predicted as being at the highest risk. The size of the risk class "Highest Risk" amounts to 6.32km², for the highest and high risk class the size is 16.79km², and for the highest, high and medium risk class it is 18.12km².

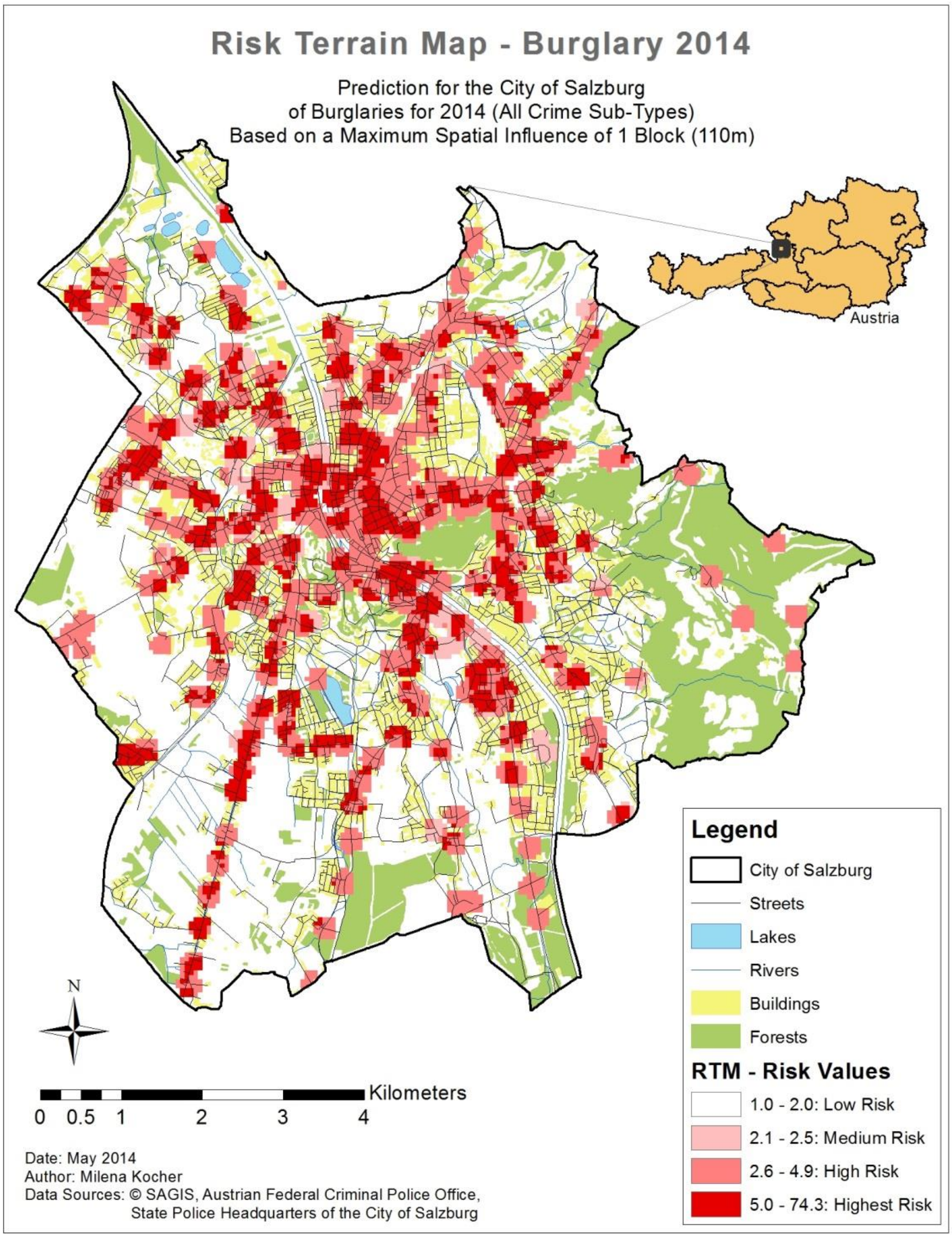

Figure 38 – Prediction of all burglaries for 2014

In a next step, the analyses were run again for burglaries, but only with selected crime subtypes. These subtypes were chosen with respect of burglaries and included burglaries in apartments, in houses, in cellars, in companies, and in stores. Excluded were bicycle and cell phone theft, burglaries into cars and all other categories. The best model for the prediction of 2013 had a maximum spatial influence of four blocks and included five risk factors, as Table 26 shows.

Table 26 – RTMDx report for selected burglaries for 2013

"Best" Model Specification

The RTMDx Utility determined that the best risk terrain model was a **Negative Binomial type II** model with **5** risk factors and a **BIC** score of **5918.7**. The model also includes an intercept term that represents the background rate of events and an intercept term that represents overdispersion of the event counts:

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Buildings Residences 4Blocks	Proximity	55	4.4635	86.7907
Rate	Bus Stops	Proximity	165	0.8251	2.2822
Rate	Pawn Shops	Proximity	440	0.7907	2.2049
Rate	Schools	Proximity	385	0.4658	1.5933
Rate	Official Buildings	Proximity	440	0.3481	1.4164
Rate	Intercept	--	--	-8.1428	--
Overdispersion	Intercept	--	--	-0.3014	--

The visualized result is presented in Figure 40. It can be seen that the size of the predicted areas is very big, related to all other previous calculations, and there are a lot of areas which are indicated as highest risk.

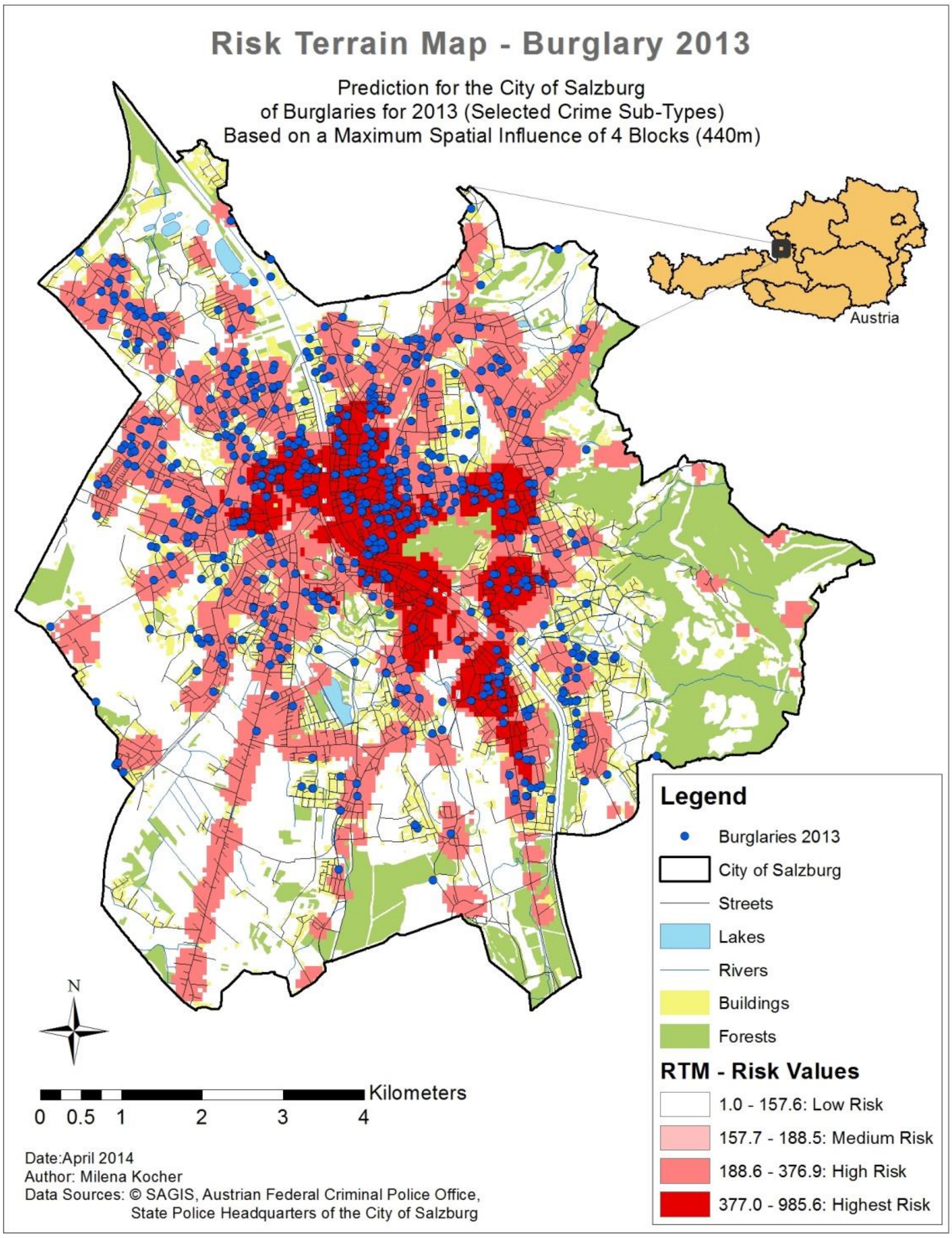

Figure 39 – Prediction of selected burglaries for 2013 (4 Blocks)

The PAI values range from 2.41 to 4.08. The calculation based on four blocks has the best value (see Table 27). It is interesting to note, that the PAI values are all very similar within the different risk classes, except for the highest value, which is explained in Figure 41.

Table 27 – Evaluation for selected Burglaries in 2013

Burglary 2013 (Selected Crime Sub-Types)

	Risk Class		
	Highest & High & Medium	Highest & High	Highest
PAI - 1 Block	2.41	2.41	2.77
PAI - 2 Blocks	2.24	2.24	3.22
PAI - 3 Blocks	2.26	2.26	3.05
PAI - 4 Blocks	2.20	2.20	4.08

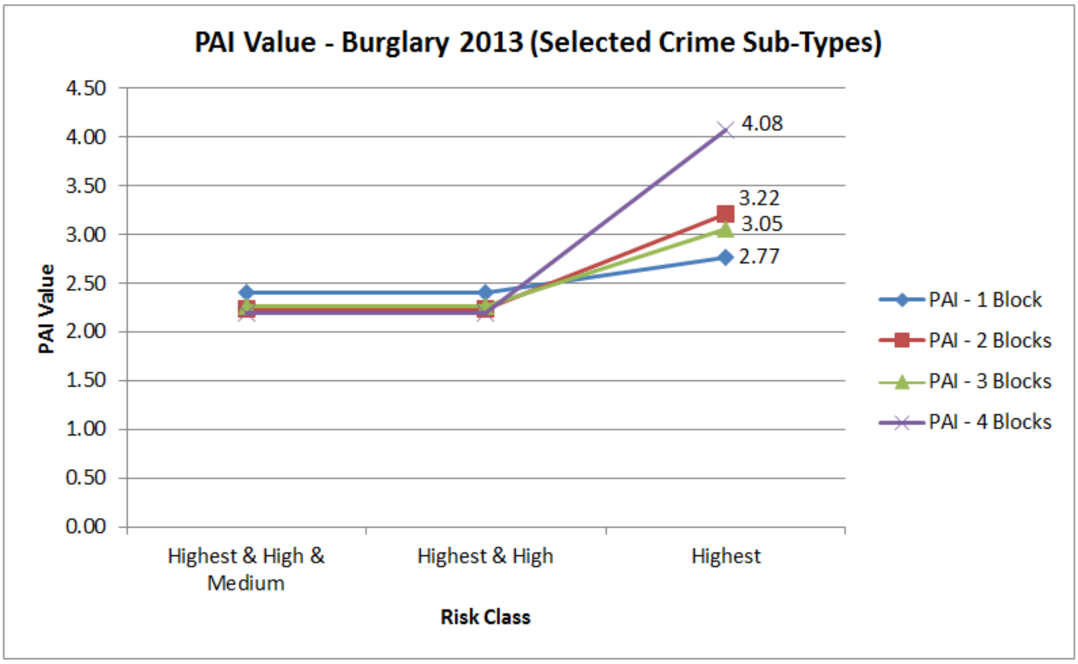

Figure 40 – Evaluation for selected burglaries in 2013

Regarding the correctly predicted burglaries, between 23% and 50% could be predicted correctly, for areas between 5.3 and 11 square kilometers. The result for four blocks predicted 33% correctly for an area of about 5.3km² (see Table 28 and Figure 42).

Table 28 – Correctly predicted selected burglaries in 2013

Burglary 2013 (Selected Crime Sub-Types)

	Correctly Predicted Burglaries 2013 (selected) (Highest Risk)		Area in km²
1 Block	182 of 765	23.79%	5.57
2 Blocks	287 of 765	37.52%	7.66
3 Blocks	388 of 765	50.72%	10.91
4 Blocks	252 of 765	32.94%	5.31

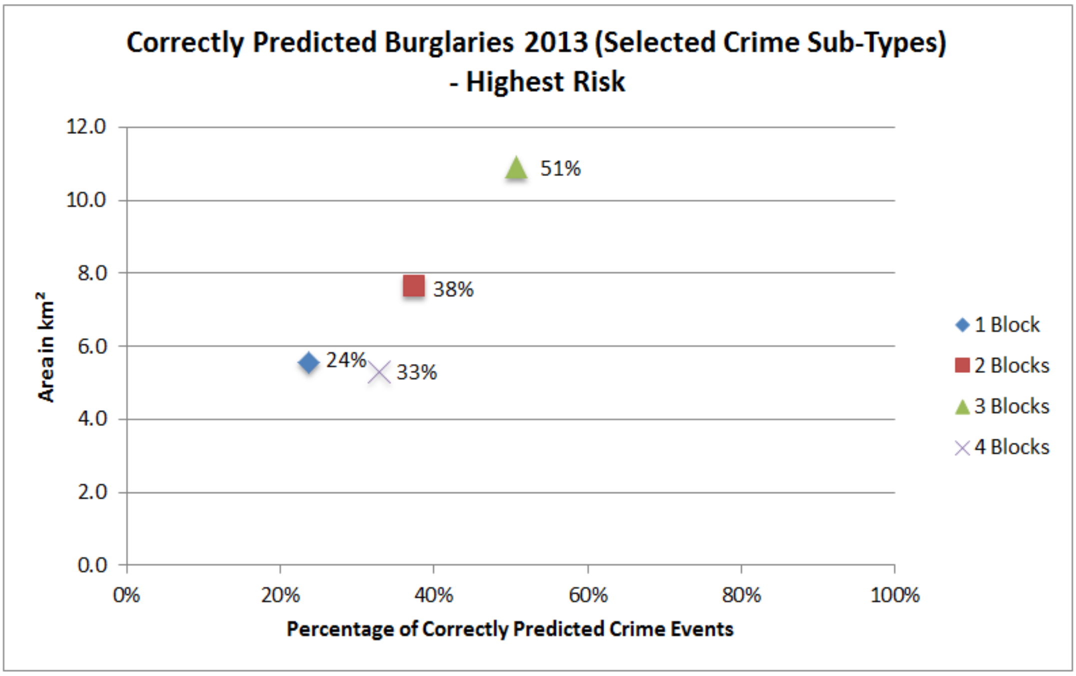

Figure 41 – Correctly predicted selected burglaries in 2013

The prediction for burglaries was based on a maximum spatial influence of the risk factors for four blocks. As Table 29 shows, five risk factors were identified to be significantly correlated with burglaries into buildings.

Table 29 – RTMDx report for selected burglaries in 2014

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Buildings Residences 4Blocks	Proximity	55	3.6935	40.1852
Rate	Bus Stops	Proximity	330	1.4031	4.0678
Rate	Pawn Shops	Proximity	440	0.6046	1.8305
Rate	Schools	Proximity	440	0.5915	1.8066
Rate	Railway Stops	Proximity	440	0.4779	1.6127
Rate	Intercept	--	--	-8.3605	--
Overdispersion	Intercept	--	--	-0.2232	--

Figure 43 presents the prediction of burglaries for 2014, including the selected crime subtypes. The predicted areas are very big and cover quite a lot of the project area. In fact, the highest risk areas have a size of 6.69km². Because there are no areas included in the medium risk class, "Highest and High Risk" as well as "Highest, High, and Medium Risk" both have an area of 21.22km².

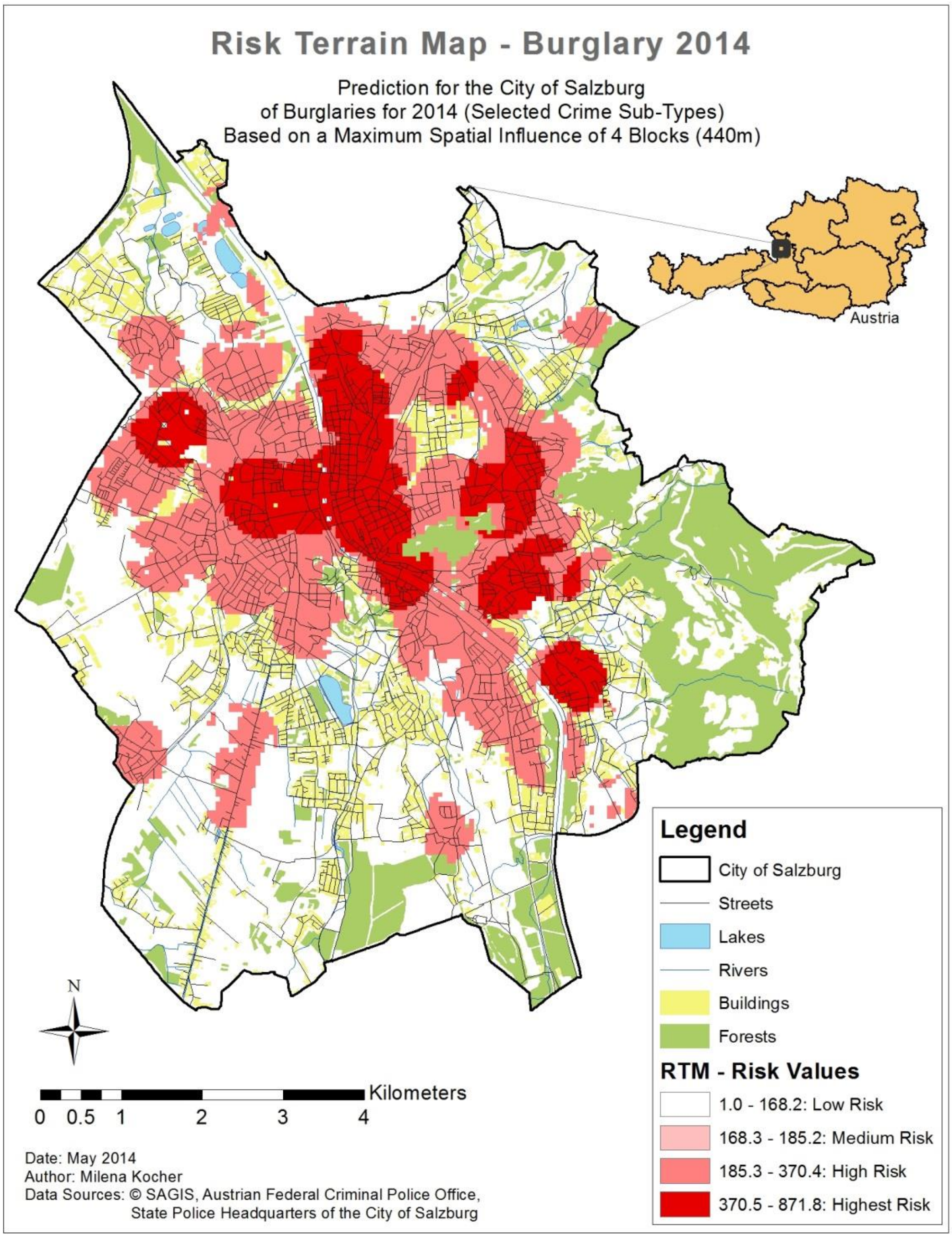

Figure 42 – Prediction of selected burglaries for 2014

4.2.4 Robbery

For the prediction of "Robbery" for 2013, the model based on a spatial influence of two blocks included four risk factors, which can be seen in Table 30.

Table 30 – RTMDx report for robberies for 2013

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bus Stops	Proximity	165	1.9629	7.1199
Rate	Leisure and Fastfood Outlets	Proximity	220	1.2723	3.5691
Rate	Bars and Pubs	Density	165	1.1798	3.2537
Rate	Banks	Proximity	110	1.1166	3.0545
Rate	Intercept	--	--	-7.4491	--
Overdispersion	Intercept	--	--	-0.4264	--

In Figure 44, the finalized map is presented, which shows the predicted areas and the robberies of 2013. It is visible that many of the actual crime events are within predicted areas, many even in the highest risk area, which is concentrated in the center of the city of Salzburg.

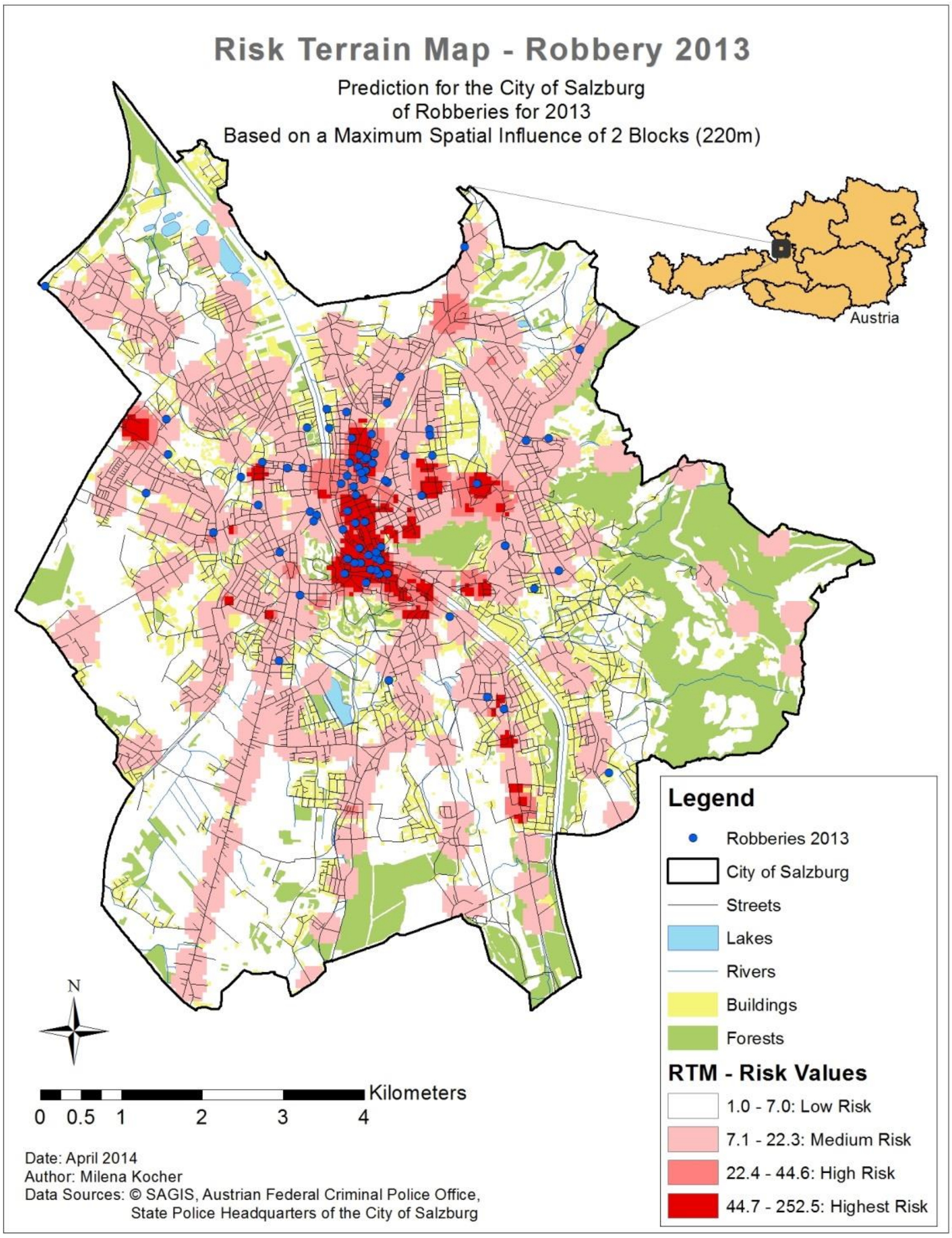

Figure 43 – Prediction of robberies for 2013 (2 Blocks)

The evaluation shows, that the highest PAI value was calculated for the model with two blocks and was 18.46. This result can be seen in Table 31 and Figure 45.

Table 31 – Evaluation for Robberies for 2013

Robbery 2013

	Risk Class		
	Highest & High &	Highest & High	Highest
PAI - 1 Block	6.26	8.09	15.23
PAI - 2 Blocks	6.63	10.10	18.46
PAI - 3 Blocks	4.54	10.28	10.63
PAI - 4 Blocks	4.09	9.75	9.89

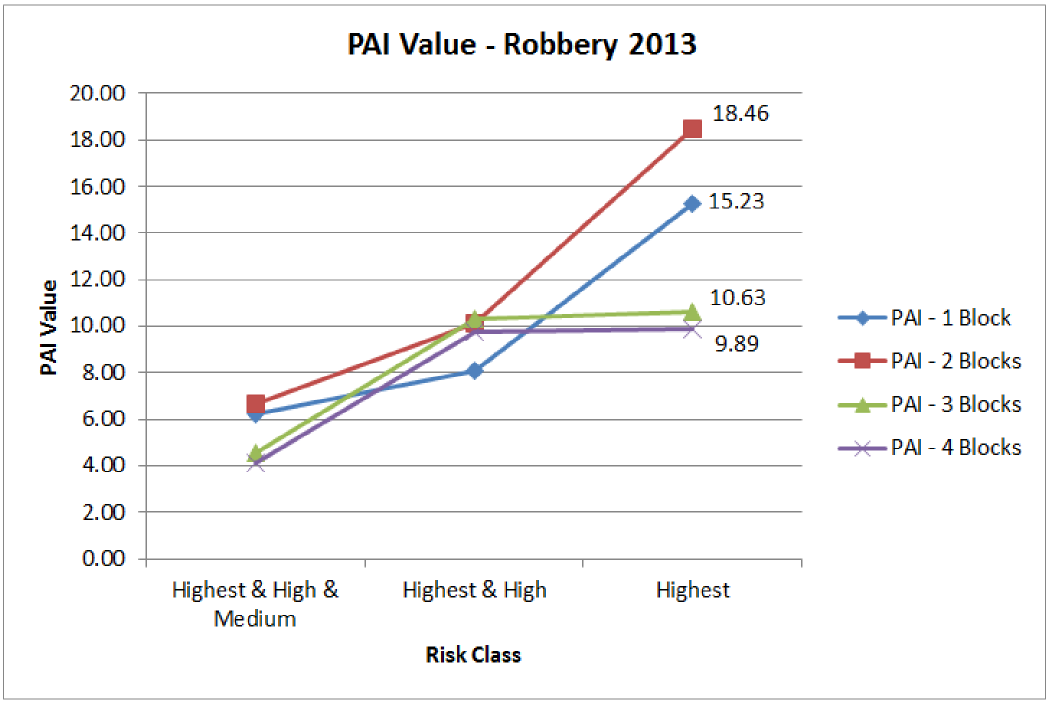

Figure 44 – Evaluation for robberies for 2013

Table 32 and Figure 46 present the correctly predicted robberies and the predicted areas for the block lengths. For a spatial influence of one and two blocks, 31% and 43% of the robberies were correctly predicted. The areas covered are 1.3 and 1.5 square kilometers, respectively. For three and four blocks the prediction was correct for 51%, but this includes an area of more than 3km².

Table 32 – Correctly predicted robberies for 2013

Robbery 2013

	Correctly Predicted Robberies 2013 (Highest Risk)		Area in km²
1 Block	26 of 85	30.59%	1.32
2 Blocks	37 of 85	43.53%	1.55
3 Blocks	43 of 85	50.59%	3.12
4 Blocks	43 of 85	50.59%	3.36

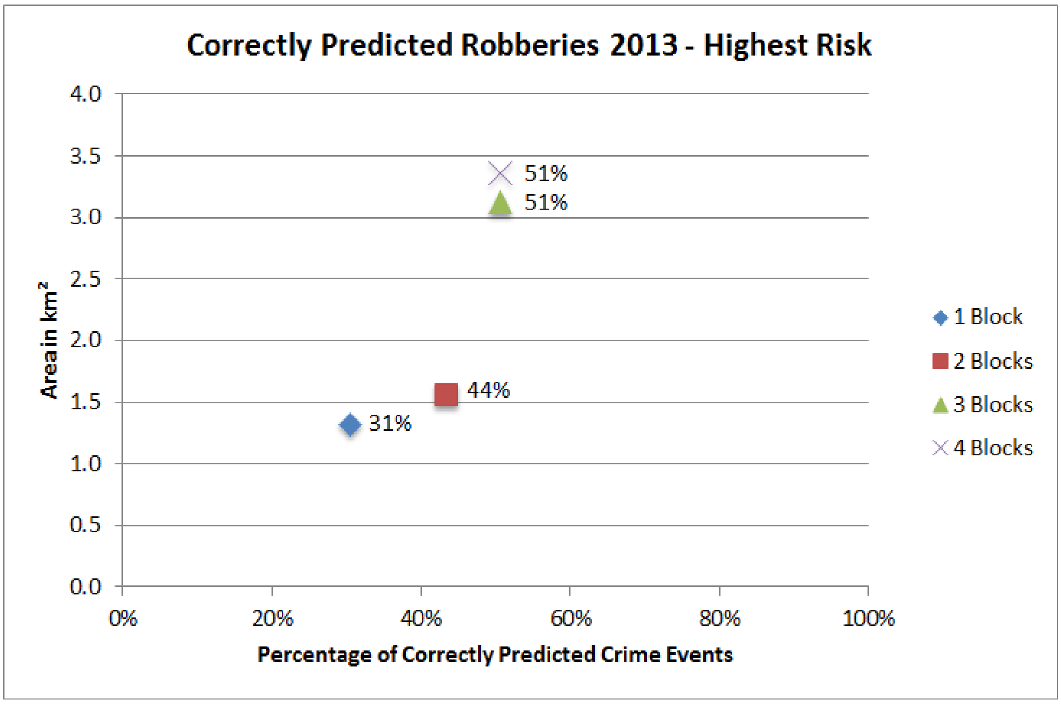

Figure 45 – Correctly predicted robberies for 2013

Based on the evaluation result, the prediction for 2014 was based on a maximum spatial influence of two blocks. The best model (see Table 33) includes also four risk factors, but one different. Instead of bars the railway stops were identified to be correlated significantly.

Table 33 – RTMDx report for robberies for 2014

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Banks	Proximity	165	1.8559	6.3975
Rate	Leisure and Fastfood Outlets	Proximity	220	1.6443	5.1774
Rate	Railway Stops	Proximity	220	1.2267	3.4100
Rate	Bus Stops	Proximity	110	1.0244	2.7854
Rate	Intercept	--	--	-6.9039	--
Overdispersion	Intercept	--	--	-1.7717	--

In the final map (see Figure 47) areas with a high risk that a robbery will take place in 2014 are shown. The "Highest Risk" class predicts areas with a size of 1.67 km², "Highest Risk" and "High Risk" classes, an area of 2.33 km², and for "Highest, High, and Medium Risk" classes almost 10 km² are predicted.

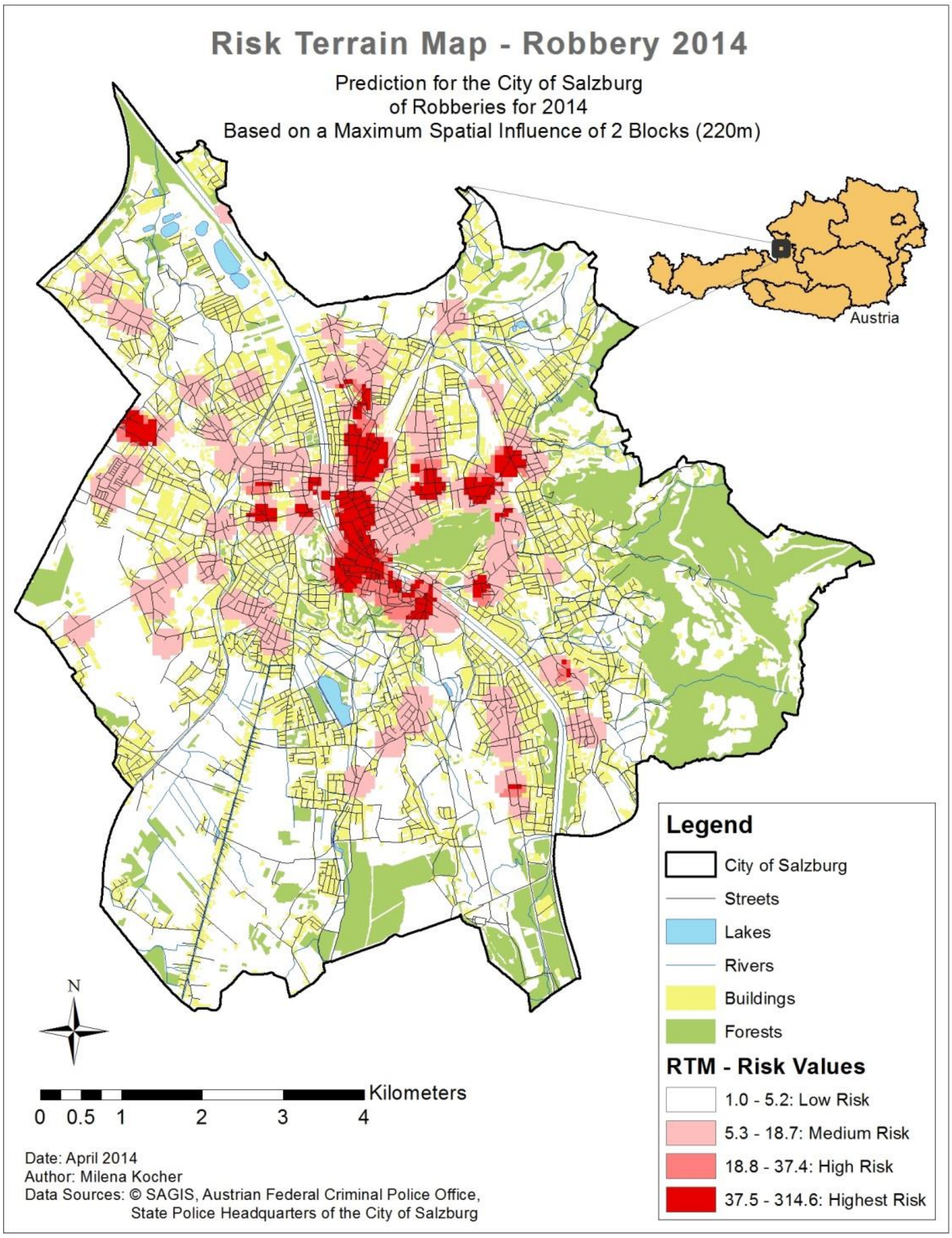

Figure 46 – Prediction for robberies in 2014

4.3 Comparison of Results

In this subchapter, the results of the predictions and evaluations are summarized and discussed.

In sum, 42 different models were calculated with the RTMDx Utility software, whereby 27 of these models identified at least one correlated risk factor and provided a result. For the other 15 models which resulted in an error, no risk factors could be found which correlated with the crime event. The 27 successfully calculated models were further operationalized and finalized to a risk terrain map, showing the prediction for a particular outcome event for 2013 or 2014.

For 2013, always four models were calculated and finalized to a risk terrain map, which could then be evaluated and the best model be identified. Based on the information of the evaluation, the prediction for 2014 was done. These analyses were done for the four different crime events, but because of differences in the modeled time period or the data selection, six final results could be created. This included four models and predictions for 2013, two evaluations, and the prediction for 2014.

For assaults, the calculation for one year was not successful, and so the models were also created for each of the four seasons. For spring and summer predictions could be made, but for fall and winter no risk factors correlated with assaults. This can be an indicator that the spatial distribution of assaults in fall and winter might be different from the distribution of assaults in spring and summer. For auto thefts, only the models based on two, three, and four blocks were successful for 2013. For 2014, no model could be calculated. The crime event burglary was separated into all burglaries and into burglaries into buildings. For both categories all models could be calculated. For robberies, all models for 2013 as well as the model for 2014 could be calculated.

The PAI values for the above described crime events and their subdivisions, are given in Table 34 and Figure 48. The highest PAI value was reached for assaults in spring, which amounted to 31.37. Also, the model for assault in summer showed a high value of 23.4. The model for auto theft with a value of 1.71 and for burglaries with values of 4.08 and 4.46 show the lowest PAI values. The PAI value for robberies amounts to 18.46. It can be seen that the models for assaults as well as for robberies have high PAI values compared with the values for auto thefts and burglaries.

Table 34 – Comparison of PAI values for different crime events

Crime Event 2013	PAI Value
Assault Spring	31.37
Assault Summer	23.40
Auto Theft	1.71
Burglary All	4.46
Burglary Selected	4.08
Robbery	18.46

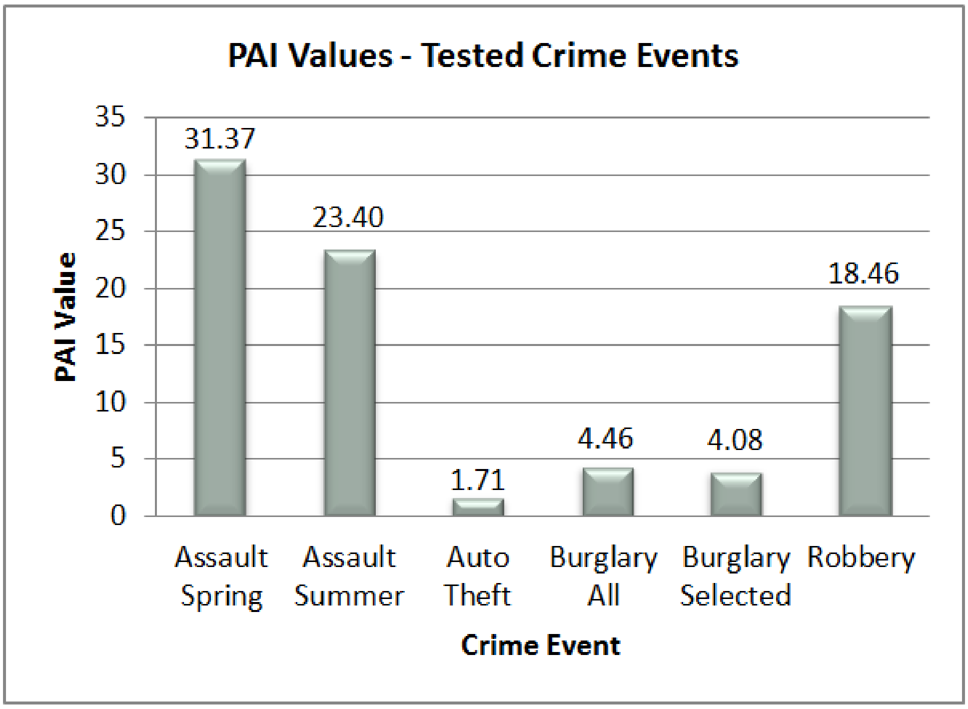

Figure 47 - Comparison of PAI values for different crime events

Table 35 summarizes the correctly predicted crime events and the area for the six predictions of 2013. In general, between 25% and 44% could be predicted correctly. The modeled areas range from 0.78km² to 9.9km².

Table 35 – Comparison of correctly predicted crime events

Crime Event	Correctly Predicted Crimes		Area in km²
Assault Spring	153 of 412	37.14%	0.78
Assault Summer	198 of 446	44.39%	1.25
Auto Theft	9 of 35	25.71%	9.90
Burglary all	629 of 2473	25.43%	3.74
Burglary selected	252 of 765	32.94%	5.31
Robbery	37 of 85	43.53%	1.32

The same information is shown in Figure 49. It is visible that the result for auto thefts, shown through the triangle, is not quite successful, because only 25% were predicted correctly and the area amounts to nearly 10km². The results for burglary (all burglaries and selected burglaries) are relatively good, compared to the results of the other crime types. Between 25% and 32% were predicted correctly for an area of 3.7 and 5.3km², respectively. The best results were achieved for assaults (spring and summer) as well as for robbery. The result for assault in spring correctly predicted 37% with the smallest area of only 0.78km². The results for assaults for summer and robbery correctly predicted 44% and 43% for an area of 1.25 and 1.32km².

The reason for the poor prediction of auto thefts might be mainly due to the fewer obtained risk factors, as described in chapter 4.1.2.
Although there was no result for assaults based on a one year time period, the results for spring and summer are quite good. It is also interesting to note, that the results for burglaries including all crime sub-types was better than the result of including only selected crime sub-types. For robberies, all risk factors were available and also the high data quality of robberies might lead to such a good result. The high data quality can be explained by the fact, that the time and location of a robbery can be reported exactly.

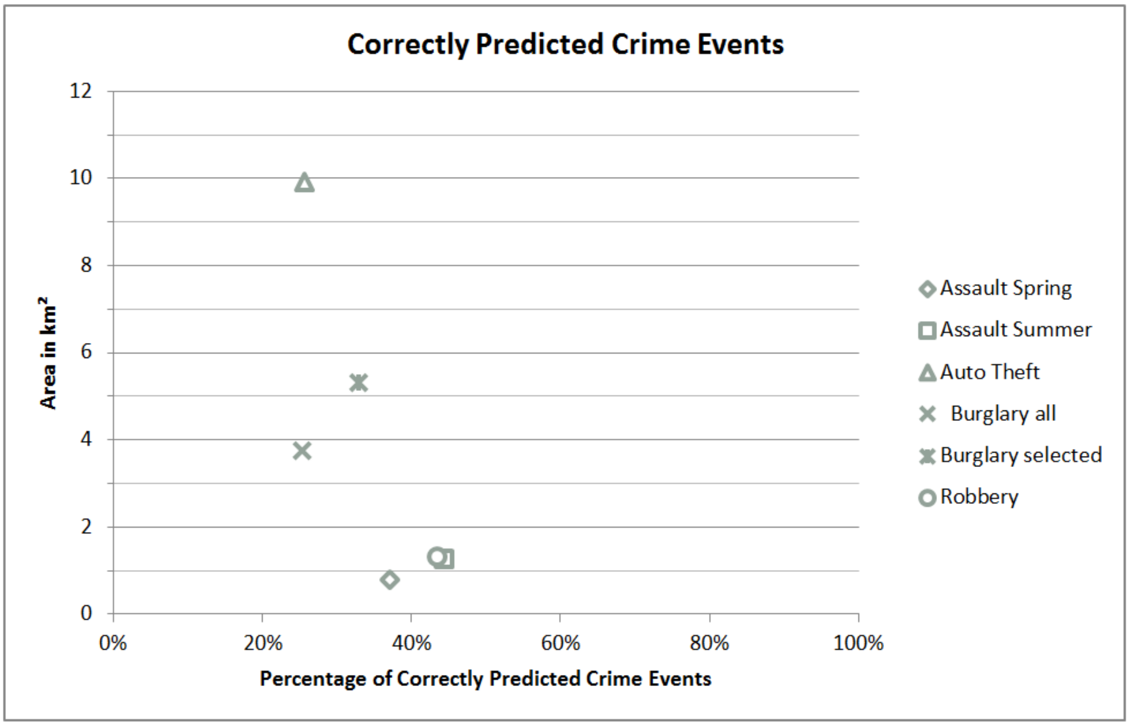

Figure 48 – Comparison of correctly predicted crime events

5. Discussion

In this chapter a critical reflection of the project is included and it is further discussed, if the expected results and goals were reached.

5.1 Critical Reflection

How good the RTM technique works for a specific project area is affected by different factors. Above all, the availability and quality of the risk factor data is crucial. Since this technique was developed and implemented in the USA, the selected risk factors listed in the RTM Compendium are influenced by research done in the USA. It might be possible that for the project area of this research, the city of Salzburg, further risk factors exist that were not included in this thesis. In addition, not all risk factor data listed in the RTM Compendium were available or could be obtained because of a lack of expertise, as is true for the factor "social disorganization". For other risk factor data it was difficult, if not impossible, within the context of this work to obtain them. A further aspect which has to be considered is that both, the obtained as well as the self-captured risk factor data, were partly out of date. The self-captured data were based on research using the internet, where the most reliable sources were used. However, this cannot necessarily guarantee that the data are up to date.

5.2 Are the Applied Methods Appropriate?

Although the data quality regarding correctness of the self-captured data may not be considered as high, it was the only possibility to obtain these data. Through the geocoding process the captured features could be assigned exact x and y coordinates. Using the RTMDx Utility software, a standardized workflow for the calculation was enabled. Although it required further information, such as the average block length or the spatial influence, the process was more appropriate compared to a fully manually developed model, because the statistics have been improved. The average block length for the city of Salzburg was calculated based on the digitized street network and can be considered valid. Because the maximum spatial influence of the risk factors could not be estimated, it was a logic workflow to test four different options. The developed models for the operationalization make it possible to operationalize risk factors in the same way and to accelerate the process. To sum up, the RTM technique and the workflow used in this research project show that RTM is an appropriate method to predict crime events also for Austrian cities, although it works better for some crime events than for others. The availability of up to date risk factor data is a problem regarding RTM, because these data are often captured by administrative bodies, and are not updated that often.

The evaluation was done using the Predictive Accuracy Index. It has to be mentioned that this method does not consider how many crime events are committed. From the author’s point of view, it would make sense that the more crime events happen the bigger the predicted area can be. As an example, although if both assaults and auto thefts predict 50% of all happened crime events correctly, it would be reasonable that for robberies with more than 1,500 crime offenses per year a bigger area can be predicted than for auto thefts with only 32 crime offenses per year. However, using this method, a standardized comparison between the results was possible, which was the main purpose.

5.3 Have the Expected Results and Goals of the Thesis been reached?

The expected results and goals of this research project could have been reached. Risk terrain models for the four different crime events assaults, auto thefts, burglaries, and robberies were implemented and their results evaluated. Furthermore, based on the evaluation results, predictions for 2014 were done, which can be used by Salzburg Police. For the operationalization and the evaluation processes, models could be implemented to enable a semi-automatically process.

6. Conclusion and Future Work

In this chapter, the research that has been done and the results that have been found are summarized. Furthermore, possible aspects that can be considered in the future are given.

6.1 Summary

Within this research project it was possible to make predictions for four different crime events, including assaults, auto thefts, burglaries, and robberies, for the city of Salzburg. For this purpose, the technique RTM was used. This technique is primarily based on the concept that there exist risk factors that have an influence on crime events. For some risk factors, the data had to be self-captured and geocoded. In sum, eight datasets were captured, which are useful risk factors that correlate with many different outcome events and can be used for further projects, too. In a first step, the implementation was done with the RTMDx Utility software, in a second step the operationalization, combination, and finalization was done within ArcGIS. Therefore, two models for the operationalization types "proximity" and "density" were developed, as well as a model for the evaluation. These models enable a semi-automatically and thus a faster calculation compared to a manual operationalization process. For the evaluation the Predictive Accuracy Index (PAI) was used, wherefore the model automatically calculates the PAI value. The results of the risk terrain models could be compared with two charts. First, the different PAI values for the models based on different maximum spatial influences were compared. Based on these results, the prediction for 2014 could have been made, based on the maximum spatial influence of the best model of 2013. Second, the percentage of correctly predicted crime events was presented, in respect of the size of the predicted areas. Especially these results could show how useful the predictions are for the police or other decision-makers. The developed standardized visualization, which includes the most important base map layers as well as the predicted areas, enables an easier interpretation for the user. The final comparison of the risk terrain models showed that there are differences between the accuracy of the predictions which is due to the fact that not all risk factor data could be obtained for the different crime events.

6.2 Conclusion and Future Work

To sum up, this research project showed that the RTM technique can be applied to Austrian cities as well, although there are differences regarding the accuracy of the predictions. For future projects, it has to be considered that above all the availability and quality of the risk factor data are crucial for the accuracy of the predictions. Another factor is the spatial distribution of the crime event, which can vary during the year, and might lead to the conclusion that for some crime events predictions should be made on a seasonal basis rather than on a one year period. Risk factors which are specific for the study area as well as socio-economic factors could improve the risk terrain models, too. Different evaluation methods, which could include the number of crime incidents per year, could be tested to see if the results differ regarding the method.

References

ArcGIS Help, 2012. What is ModelBuilder? [online]. Available from: http://resources.arcgis.com/en/help/main/10.1/index.html#//002w00000001000000 [Accessed 14 May 2014]

Bundesministerium für Inneres (BMI), 2013. Pressegespräch "Sicherheit in Österreich: Kriminalitätsentwicklung und Maßnahmenplan 2013". Available from: http://www.bmi.gv.at/cms/BK/presse/files/522013_KrimEntwicklung.pdf [Accessed 11 May 2014]

Brantingham, P. J., Brantingham, P. L., 2008. Crime pattern theory. In Wortley, R., Mazerolle, L. (Eds.), Environmental Criminology and Crime Analysis, 2008, p.78-93.

Caplan, J. M., Kennedy, L. W., 2010. Risk Terrain Modeling Manual.

Caplan, J. M., Kennedy, L. W. (Eds.), 2011. Risk Terrain Modeling Compendium.

Caplan, J. M., Kennedy, L. W., Miller, J., 2011. Risk Terrain Modeling: Brokering Criminological Therory and GIS Methods for Crime Forecasting. Justice Quarterly, 28 (2), p.360-281. Download via LSU Libraries Homepage.

Caplan, J. M., Kennedy, L. W., Piza, E. L., 2012. Joint Utility of Event-Dependent and Environmental Crime Analysis Techniques for Violent Crime Forecasting. Crime & Delinquency, 2013, 59 (2), p.243-270. Download via LSU Libraries Homepage.

Caplan, J. M., Kennedy, L. W., Piza, E. L., 2013a. Risk Terrain Modeling Diagnostics Utility User Manual (Version 1.0). Available from: http://www.rutgerscps.org/docs/RTMDxUserManual_FinalDraft_CaplanKennedyPiza.pdf [Accessed 17 March 2014]

Caplan, J. M., Kennedy, L. W., Piza, E. L., 2013b. Risk Terrain Modeling Diagonstics Utility User Manal Appendix (Rev. 1). Available from: http://www.rutgerscps.org/software [Accessed 13 May 2014]

Caplan, J. M. (jcaplan@andromeda.rutgers.edu), 2014a, 19 February. RE: Research RTM. Email to M. Kocher (Milena.Kocher@edu.fh-kaernten.ac.at).

Caplan, J. M. (jcaplan@andromeda.rutgers.edu), 2014b, 5 April. RE: Research RTM. Email to M. Kocher (Milena.Kocher@edu.fh-kaernten.ac.at).

Chainey, S., Tompson, L., Uhlig, S., 2008. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. Security Journal, 2008, 21, p.4-28.

Available from: http://www.palgrave-journals.com/sj/journal/v21/n1/full/8350066a.html [Accessed 8 April 2014]

Cote, P., Harvard University. Address Geocoding with ArcGIS. Available from: http://www.gsd.harvard.edu/gis/manual/geocoding/ [Accessed 13 May 2014]

Ferguson, A. G., 2012. Predictive Policing and Reasonable Suspicion. Emory Law Journal, 62 (2), p259-325. Download via LSU Libraries Homepage.

Friend, Z., 2013. Predictive Policing: Using Technology to Reduce Crime. Available from: http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin/2013/April/predictive-policing-using-technology-to-reduce-crime [Accessed 21 February 2014]

Gale, R., 2013. An Application of Risk Terrain Modeling to Residential Burglary. TCNJ Journal of Student Scholarship, 9, p1-9. Available from: http://joss.pages.tcnj.edu/files/2013/04/2013-Gale.pdf [Accessed 21 February 2014]

Groff, E. R., La Vigne, N. G., 2002. Forecasting the Future of Predictive Crime Mapping. Crime Prevention Studies, 3, p.29-57. Available from: http://www.popcenter.org/library/crimeprevention/volume_13/03-groff.pdf [Accessed 20 May 2014]

Heffner, J., 2013. Statistics of the RTMDx Utility. In J. Caplan, L. Kennedy, E. Piza, Risk Terrain Modeling Diagnostics Utility User Manual (Version 1.0), p35-39. Available from: http://www.rutgerscps.org/docs/RTMDxUserManual_FinalDraft_CaplanKennedyPiza.pdf [Accessed 09 April 2014]

Joanneum Research, 2014. CRIPA Crime Predictive Analytics [online]. Available from: http://www.joanneum.at/policies/referenzprojekte/cripa.html [Accessed 13 May 2014]

Jura Forum, 2014. Gefahrenabwehr [online]. Available from: http://www.juraforum.de/lexikon/gefahrenabwehr [Accessed 11 May 2014]

Kampitsch, K., Leitner, M., Schabauer, H., 2008. Der Einsatz von Geografischen Informationssystemen im österreichischen Bundeskriminalamt. Available from: http://www.agit.at/php_files/myagit/papers/2008/6834.pdf [Accessed 11 May 2014]

Kennedy, L. W., Caplan, J. M., 2011-2013. Rutgers Center on Public Security [online]. http://www.rutgerscps.org [Accessed 26 February 2014]

Kennedy, L. W., Yasemin, G., Caplan, J. M., 2012. Analyzing and Visualizing Worldwide Spatial Data: An Application of Risk Terrain Modeling.

Kennedy, L.W., Caplan, J. M., Piza, E. L., 2011. Hotspot mapping, near repeat analysis, and risk terrain modeling. Presentation to UCL Department of Security and Crime Science. Available from: http://www.ucl.ac.uk/secret/secret-student-seminars-11/hotspot-mapping [Accessed 11 May 2014]

Levine, N., 2010. CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations. Chapter 4 – Spatial Distribution. Available from: http://www.icpsr.umich.edu/CrimeStat/files/CrimeStatChapter.4.pdf [Accessed 17 March 2014]

McCue, C., 2011. Proactive Policing: Using Geographic Analysis to Fight Crime. Geography & Public Safety, 2 (4), p3-5. Available from: http://www.nij.gov/topics/technology/maps/Documents/gps-bulletin-v2i4.pdf?Redirected=true [Accessed 25 February 2014]

Murray, A. T., Grubesic, T. H., 2013. Exploring Spatial Patterns of Crime Using Non-hierarchical Cluster Analysis. In Leitner, M. (eds), Crime Modeling and Mapping Using Geospatial Technologies, 2013, p.105-124

Uchida, C. D., 2009. A National Discussion on Predictive Policing: Defining Our Terms and Mapping Successful Implementation Strategies. Available from: https://www.ncjrs.gov/pdffiles1/nij/grants/230404.pdf [Accessed 21 February 2014]

Paul, J. S., Joiner, T. M., 2011. Integration of Centralized Intelligence with Geographic Information Systems: A Countywide Initiative. Geography & Public Safety, 3 (1), p5-7. Available from: http://www.cops.usdoj.gov/files/RIC/Publications/e081117392_GPS_Vol3_Iss1-508.pdf [Accessed 21 February 2014]

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., Hollywood, J. S., 2013. Predictive Policing – The Role of Crime Forecasting in Law Enforcement Operations. Available from: http://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.sum.pdf [Accessed 21 February 2014]

Piza, E. L., Kennedy, L. W., Caplan, J. M., 2010. Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies. Available from: http://www.rutgerscps.org/docs/NewarkRTM_CaseStudy_Brief.pdf [Accessed 25 February 2014]

Rutgers, State University of New Jersey Office of the Vice President for Research and Economic Devel., 2014. Predicting crime with big data, affordably [online]. ScienceDaily. Available from: www.sciencedaily.com/releases/2014/01/140110142212.htm [Accessed 18 March 2014]

Salzburger Nachrichten (SN), 2013. Kriminalstatistik Salzburg: Leichte Anzeigen-Zunahme. Available from: http://www.salzburg.com/nachrichten/salzburg/chronik/sn/artikel/kriminalstatistik-salzburg-leichte-anzeigen-zunahme-46249 [Accessed 11 May 2014]

Stadt Salzburg, 2014. Salzburg in Zahlen. Available from: http://www.stadt-salzburg.at/internet/politik_verwaltung/salzburg_in_zahlen.htm [Accessed 11 May 2014]

Weisburd, D., Bernasco, W., Bruinsma, G. J. N., 2009. Putting Crime in its Place – Units of Analysis in Geographic Criminology. Download via LSU Libraries Homepage

Weisburd, D., Groff, E. R., Yang, S., 2012. The Criminology of Place. Street Segments and Our Understanding of the Crime Problem. Download via LSU Libraries Homepage

Wikipedia, 2014. City block. Available from: http://en.wikipedia.org/wiki/City_block [Accessed 13 May 2014]

Wilson, R. E., Smith, S. C., Markovic, J. D., LeBeau, J. L., 2009. Geospatial Technology Working Group – Meeting Report on Predictive Policing. Available from: http://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Documents/geospatial-twg-predictive-policing-meeting-summary.pdf [Accessed 21 February 2014]

List of Figures

List of Tables

Appendix

The appendix shows the results for assaults, auto thefts, burglaries, and robberies, the RTMDx reports, and the predictions for 2013 for the other three maximum spatial influences.

For "Assault" for spring 2013 the best model was calculated for one block. Table.A 1 and Figure.A 1 show the result for the calculation based on a maximum spatial influence of two blocks, where seven risk factors were included.

Table.A 1 – RTMDx report for assault for spring 2013 (2 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bus Stops	Proximity	220	1.8300	6.2339
Rate	Railway Stops	Proximity	110	1.4720	4.3579
Rate	Clubs and Discos	Proximity	165	1.0297	2.8002
Rate	Bars and Pubs	Density	220	0.9743	2.6492
Rate	Nightclubs	Density	220	0.8188	2.2678
Rate	Banks	Proximity	220	0.7062	2.0263
Rate	Cash Points	Proximity	165	0.6732	1.9606
Rate	Intercept	--	--	-5.9273	--
Overdispersion	Intercept	--	--	0.5349	--

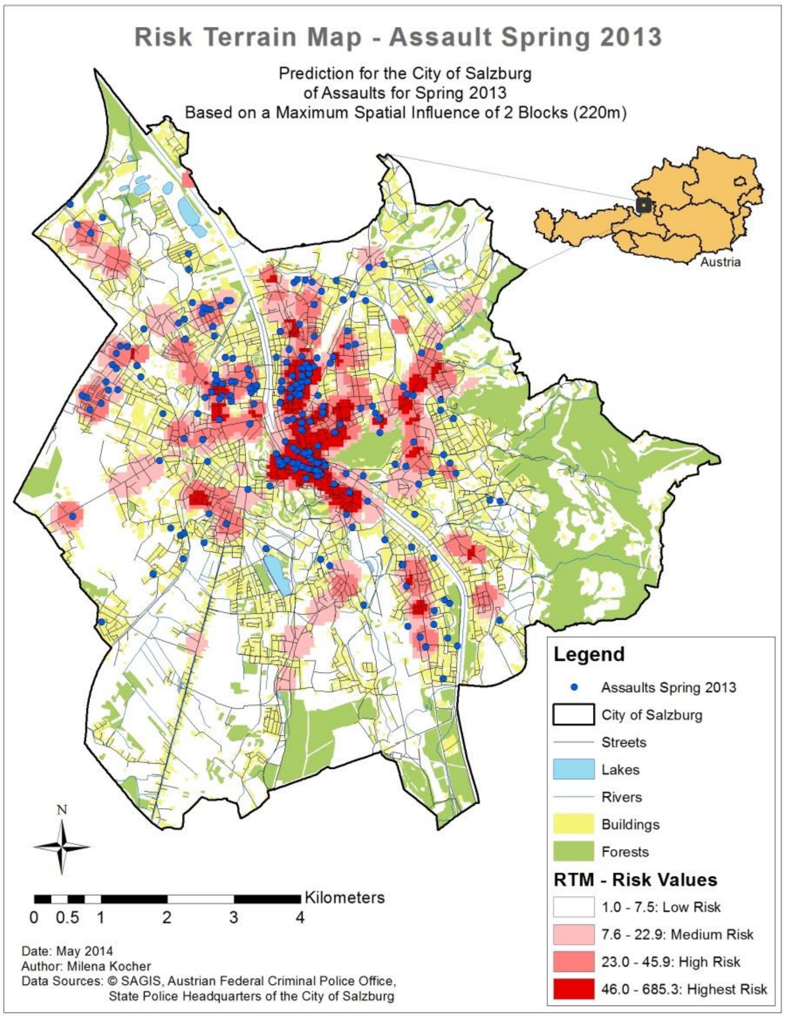

Figure.A 1 - Prediction for assault for spring 2013 (2 Blocks)

In Table.A 2 and Figure.A 2 the calculation for assaults for spring 2013, based on three blocks (330m), can be seen, which included five risk factors.

Table.A 2 - RTMDx report for assault for spring 2013 (3 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bus Stops	Proximity	220	1.7913	5.9972
Rate	Bars and Pubs	Density	220	1.0812	2.9482
Rate	Cash Points	Proximity	330	1.0711	2.9186
Rate	Clubs and Discos	Proximity	165	1.0280	2.7955
Rate	Nightclubs	Proximity	275	0.7043	2.0224
Rate	Intercept	--	--	-5.9896	--
Overdispersion	Intercept	--	--	0.5634	--

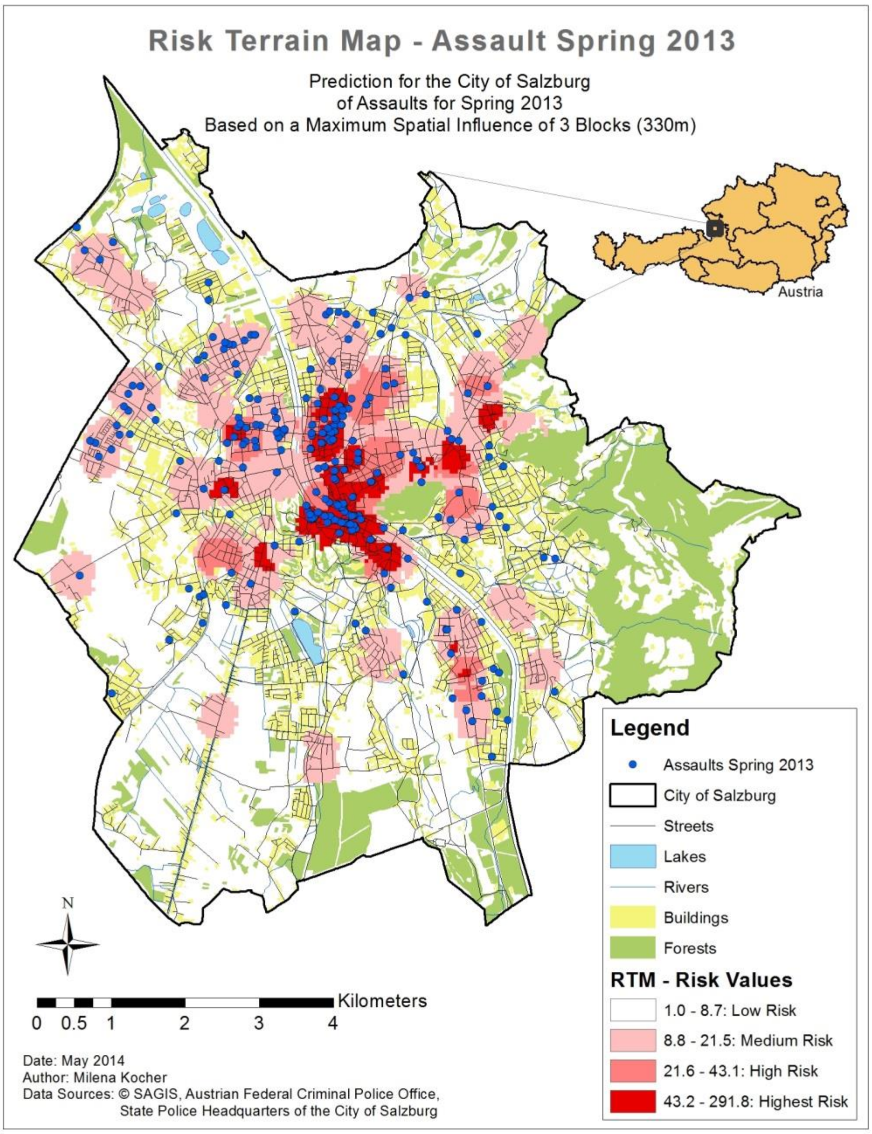

Figure.A 2 – Prediction for assault for spring 2013 (3 Blocks)

The result for the calculation for assaults for spring 2013 with a maximum spatial influence of four blocks (440m) can be seen in Table.A 3 and Figure.A 3.

Table.A 3 – RTMDx report for assault for spring 2013 (4 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bus Stops	Proximity	220	1.6609	5.2640
Rate	Cash Points	Proximity	440	1.3609	3.8997
Rate	Bars and Pubs	Density	220	1.1394	3.1249
Rate	Clubs and Discos	Density	165	1.0337	2.8114
Rate	Nightclubs	Proximity	275	0.7211	2.0567
Rate	Intercept	--	--	-6.2113	--
Overdispersion	Intercept	--	--	0.5648	--

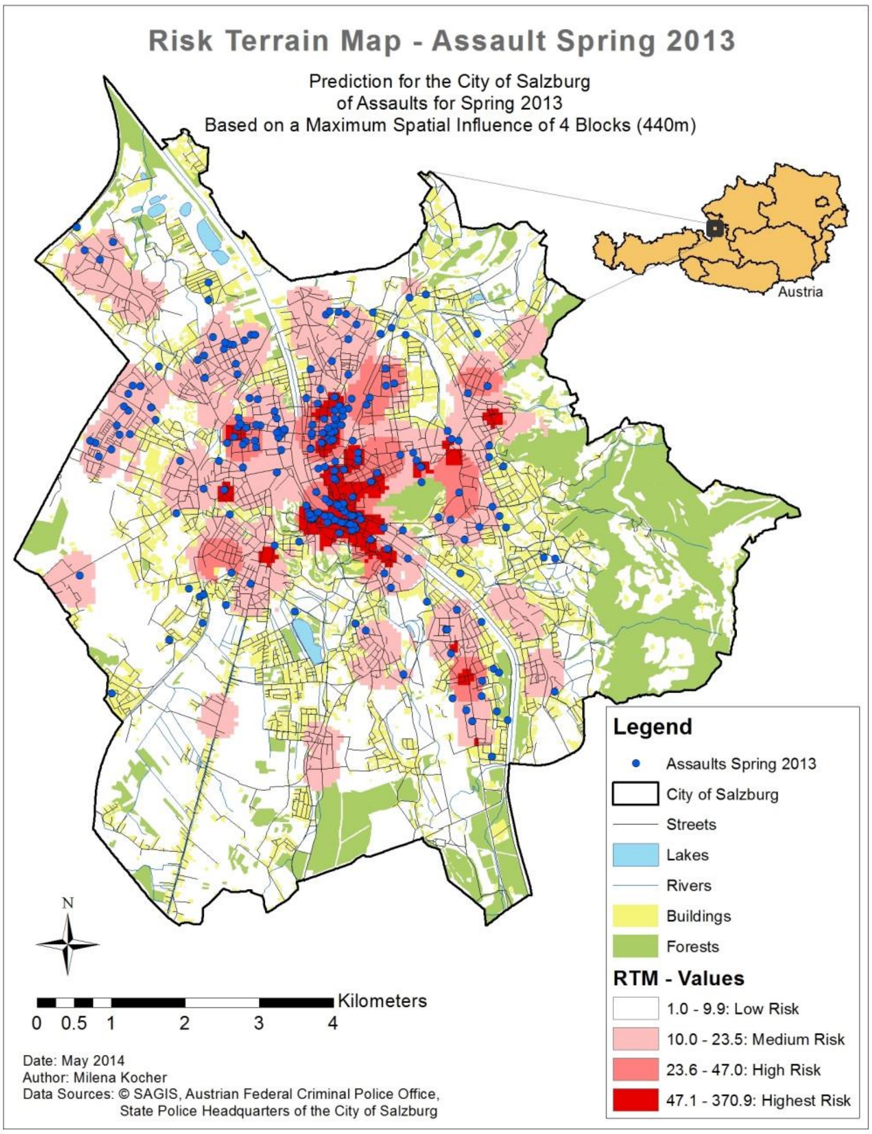

Figure.A 3 - Prediction for assault for spring 2013 (4 Blocks)

The next results are presented for assaults for summer 2013, for two, three, and four blocks. First, the result for the calculation based on a two block lengths is given (see Table.A 4 and Figure.A 4).

Table.A 4 – RTMDx report for assault for summer 2013 (2 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bus Stops	Proximity	220	1.6389	5.1495
Rate	Clubs and Discos	Density	165	1.2083	3.3478
Rate	Bars and Pubs	Density	110	1.0154	2.7605
Rate	Banks	Proximity	220	0.9157	2.4985
Rate	Nightclubs	Density	165	0.7472	2.1111
Rate	Cash Points	Proximity	220	0.6654	1.9452
Rate	Intercept	--	--	-5.8742	--
Overdispersion	Intercept	--	--	0.7210	--

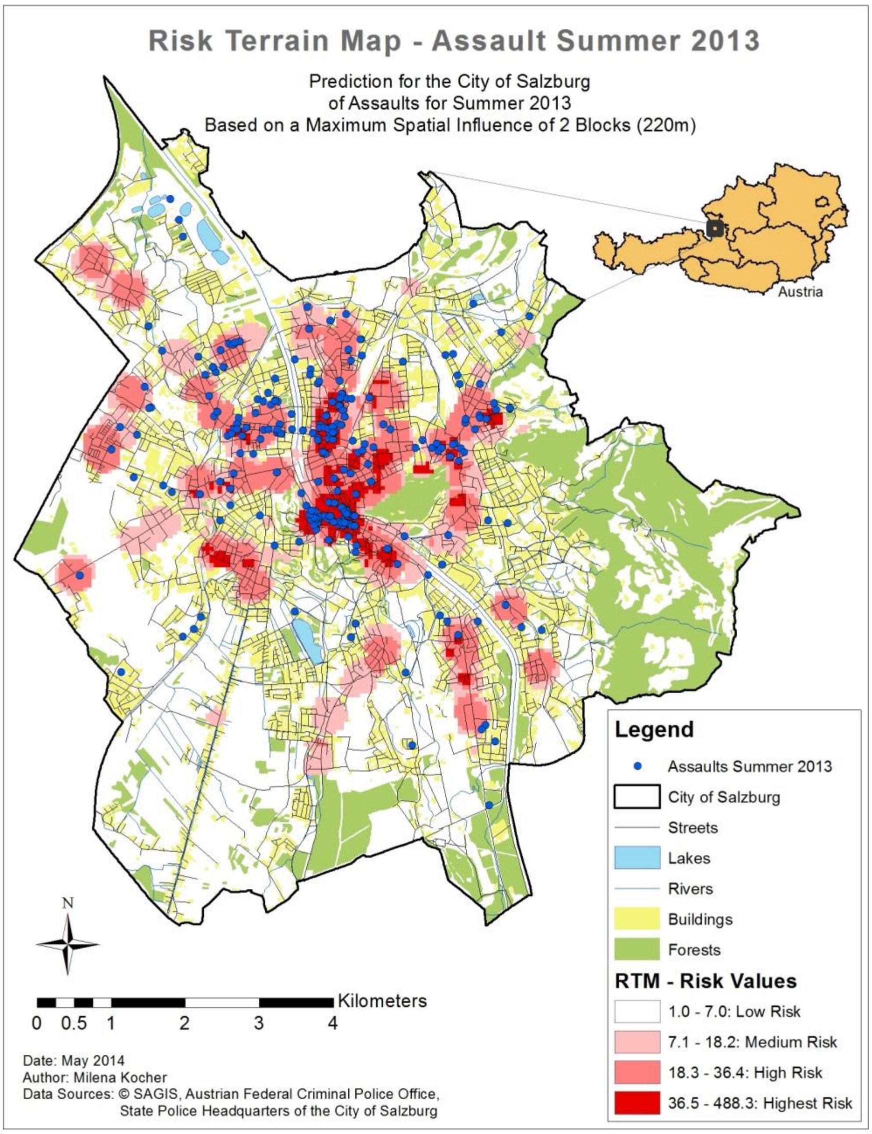

Figure.A 4 – Prediction for assault for summer 2013 (2 Blocks)

The calculation based on a maximum spatial influence of three and four blocks showed the same risk factors and spatial influences. This leads to same results. Thus, the result of the report (see Table.A 5) and the prediction (see Figure.A 5) are given only once.

Table.A 5 – RTMDx report for assault for summer 2013 (3, 4 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bus Stops	Proximity	220	1.5208	4.5759
Rate	Cash Points	Proximity	330	1.5134	4.5421
Rate	Clubs and Discos	Density	165	1.2691	3.5576
Rate	Bars and Pubs	Density	110	1.1322	3.1025
Rate	Nightclubs	Density	165	0.8097	2.2473
Rate	Intercept	--	--	-5.9840	--
Overdispersion	Intercept	--	--	0.7249	--

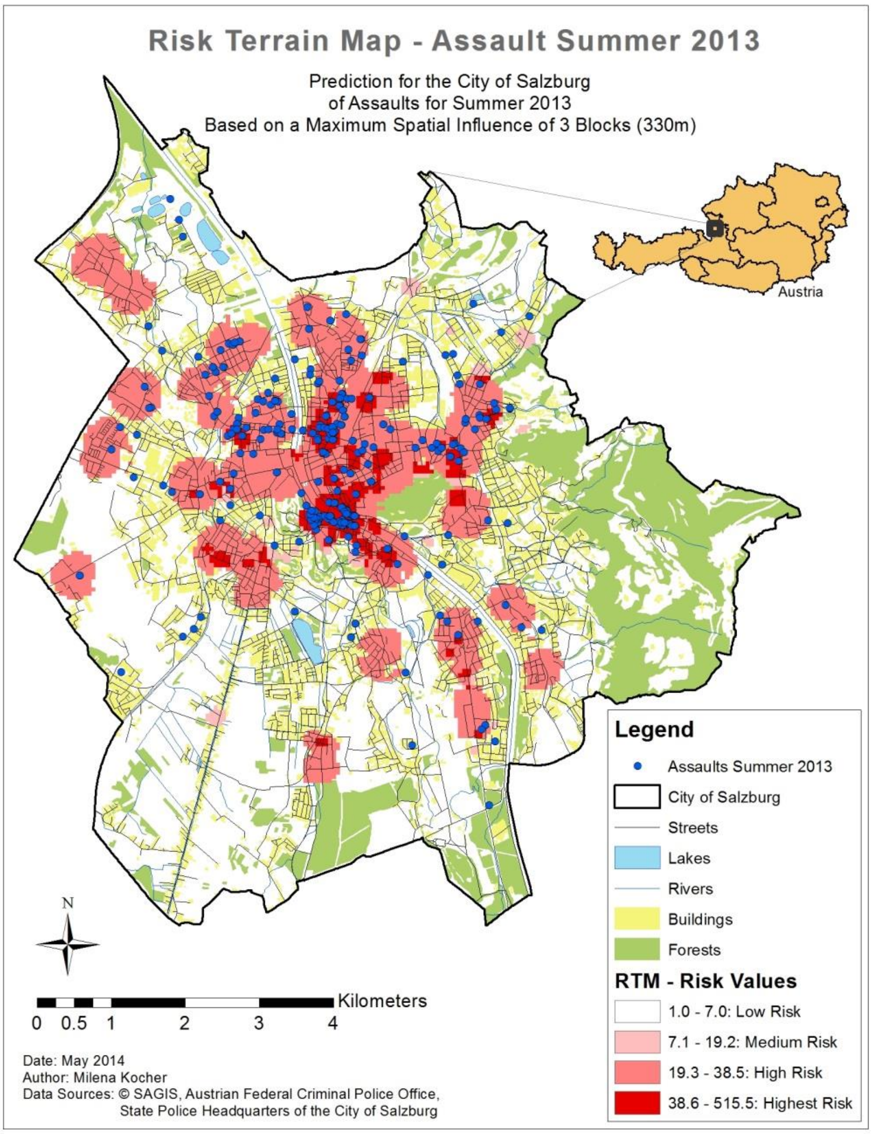

Figure.A 5 – Prediction for assault for summer 2013 (3, 4 Blocks)

The next section presents the results for auto thefts. No model could be calculated for one block and the best model was calculated for two blocks. The results for three and four blocks are the same and shown in Table.A 6 and Figure.A 6.

Table.A 6 – RTMDx report for auto theft for 2013 (3, 4 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Schools	Proximity	275	1.8025	6.0648
Rate	Intercept	--	--	-7.2695	--
Overdispersion	Intercept	--	--	-1.9169	--

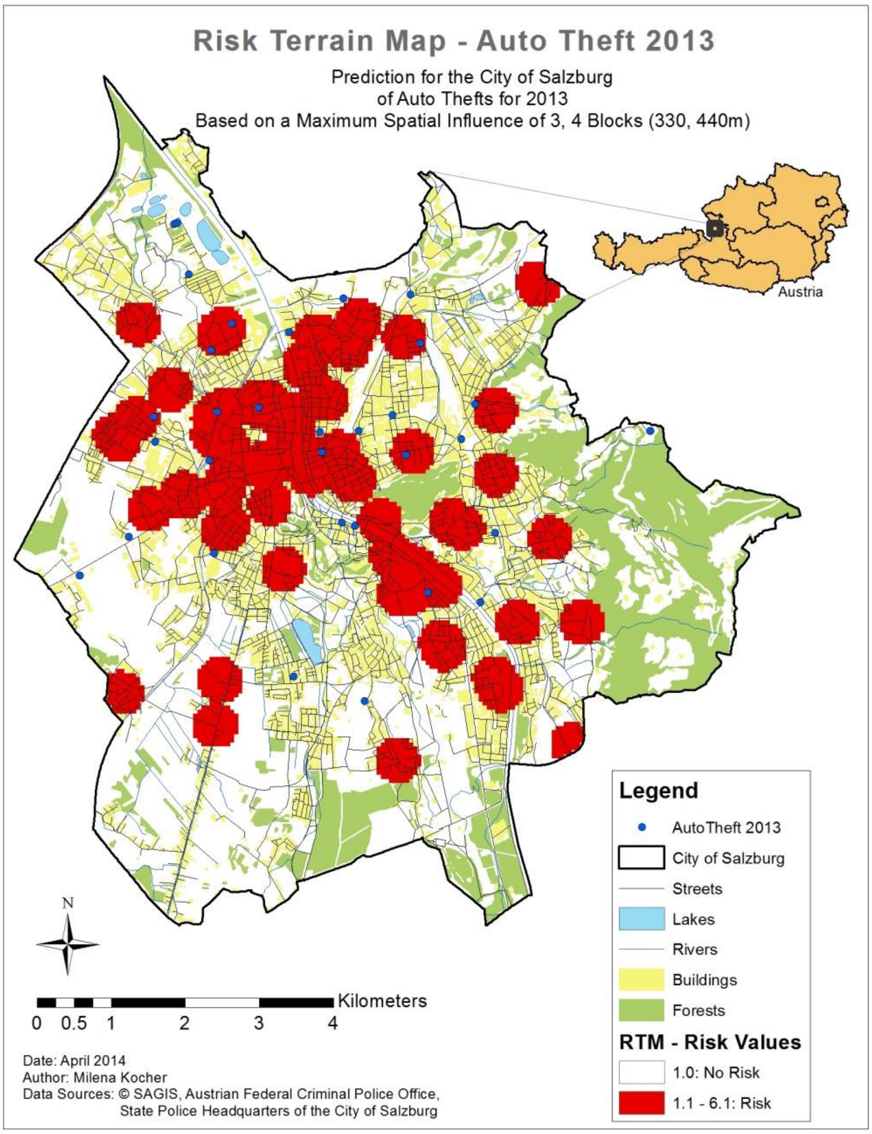

Figure.A 6 – Prediction for auto theft for 2013 (3, 4 Blocks)

For burglaries, first the results for the calculations with all crime subtypes are presented. The best result could be calculated based on a maximum spatial influence of one block. Table.A 7 and Figure.A 7 show the result based on two block lengths.

Table.A 7 – RTMDx report for Burglary (all) 2013 (2 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Buildings Residences 2Blocks	Proximity	55	4.3678	78.8699
Rate	Bus Stops	Proximity	165	0.8764	2.4024
Rate	Pawn Shops	Proximity	165	0.8491	2.3377
Rate	Official Buildings	Proximity	165	0.8030	2.2322
Rate	Schools	Proximity	220	0.5419	1.7192
Rate	Railway Stops	Proximity	220	0.5288	1.6969
Rate	Intercept	--	--	-6.9698	--
Overdispersion	Intercept	--	--	0.3733	--

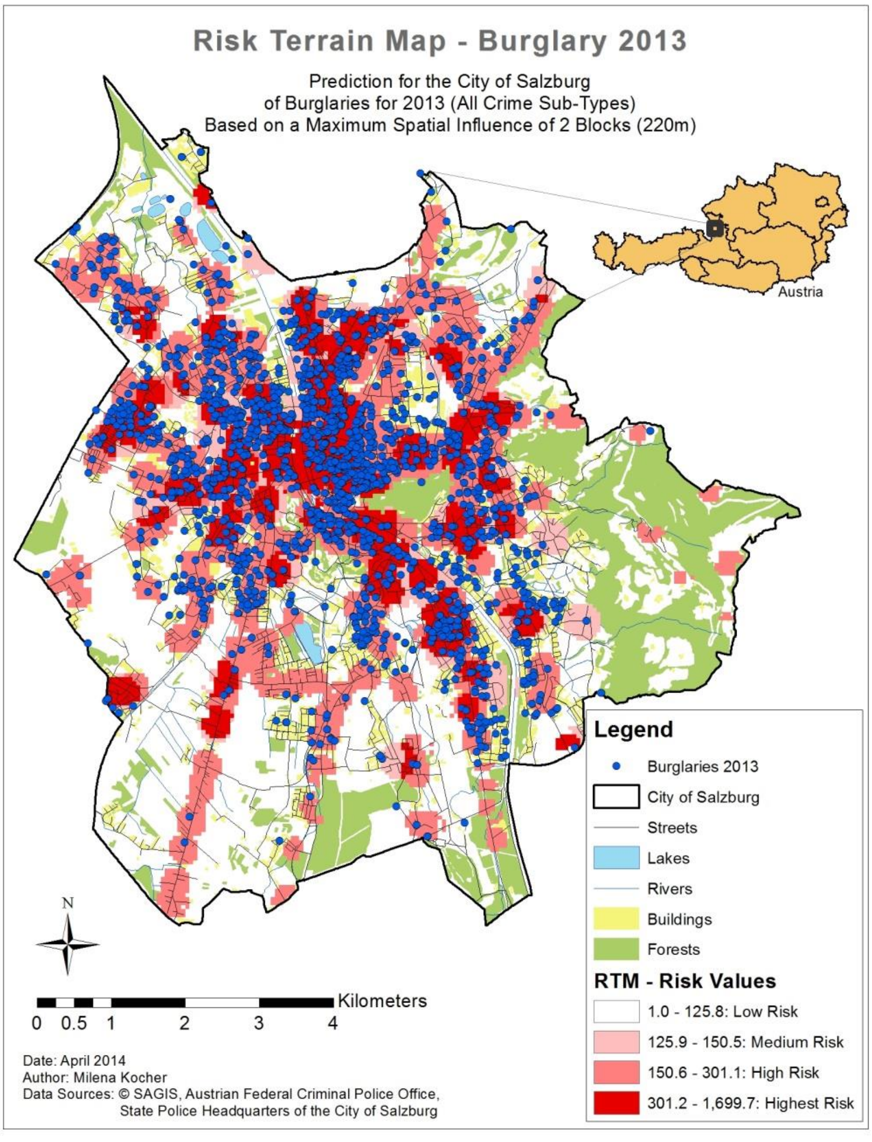

Figure.A 7 – Prediction for all burglaries for 2013 (2 Blocks)

In the next table and figure (see Table.A 8 and Figure.A 8), the result for three blocks is shown.

Table.A 8 - RTMDx report for all burglaries for 2013 (3 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Buildings Residences 3Blocks	Proximity	55	5.2288	186.5688
Rate	Bus Stops	Proximity	220	0.9920	2.6967
Rate	Official Buildings	Proximity	330	0.5930	1.8094
Rate	Pawn Shops	Proximity	330	0.4936	1.6382
Rate	Schools	Proximity	330	0.4795	1.6152
Rate	Railway Stops	Proximity	330	0.3045	1.3559
Rate	Intercept	--	--	-8.0424	--
Overdispersion	Intercept	--	--	0.3863	--

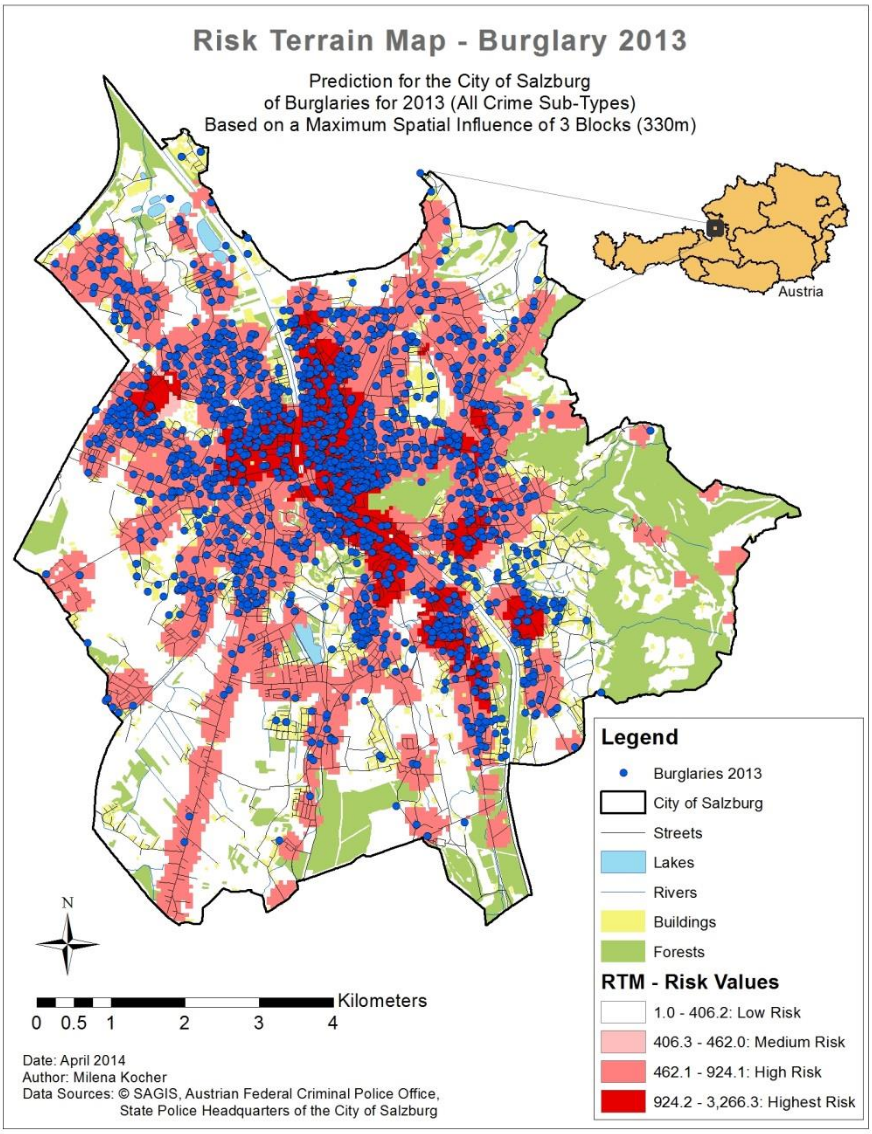

Figure.A 8 - Prediction for all burglaries for 2013 (3 Blocks)

The result for four block lengths can be seen in Table.A 9 and Figure.A 9.

Table.A 9 - RTMDx report for all burglaries for 2013 (4 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Buildings Residences 4Blocks	Proximity	55	4.0229	55.8629
Rate	Bus Stops	Proximity	165	0.7568	2.1315
Rate	Pawn Shops	Proximity	440	0.6085	1.8377
Rate	Schools	Proximity	385	0.5676	1.7640
Rate	Official Buildings	Proximity	440	0.5068	1.6599
Rate	Railway Stops	Proximity	440	0.4006	1.4927
Rate	Intercept	--	--	-6.8830	--
Overdispersion	Intercept	--	--	0.3547	--

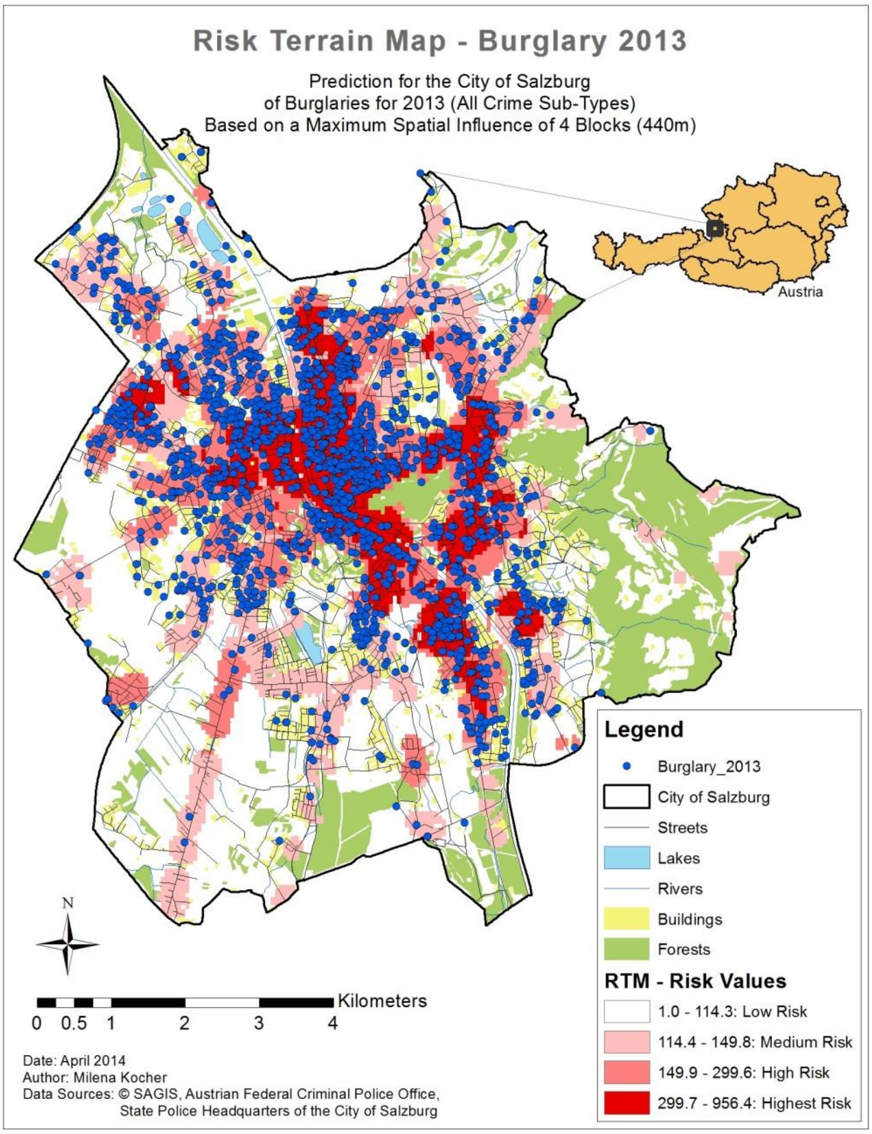

Figure.A 9 - Prediction for all burglaries for 2013 (4 Blocks)

The best result for burglaries, where only selected crime subtypes were included, was achieved with four blocks. The result based on one block is given in Table.A 10 and Figure.A 10.

Table.A 10 – RTMDx report for selected burglaries for 2013 (1 Block)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Pawn Shops	Proximity	110	1.7260	5.6181
Rate	Bus Stops	Proximity	110	1.3085	3.7006
Rate	Official Buildings	Proximity	55	1.0362	2.8185
Rate	Schools	Proximity	110	0.6146	1.8489
Rate	Buildings Residences 1Block	Density	110	0.4937	1.6383
Rate	Intercept	--	--	-3.7813	--
Overdispersion	Intercept	--	--	-0.2432	--

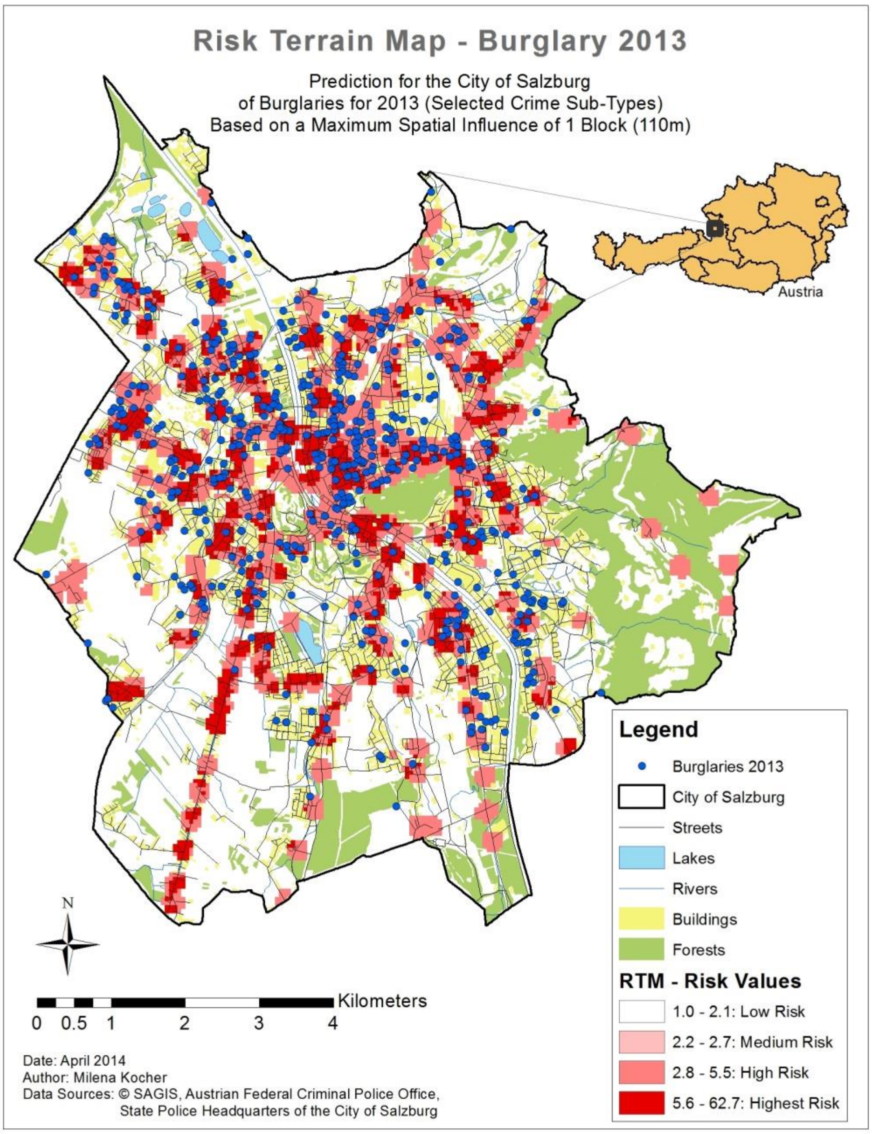

Figure.A 10 – Prediction for selected burglaries for 2013 (1 Block)

The results for selected burglaries, based on two block lengths are shown in Table.A 11 and Figure.A 11.

Table.A 11 - RTMDx report for selected burglaries for 2013 (2 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Buildings Residences 2Blocks	Proximity	55	4.5824	97.7487
Rate	Pawn Shops	Proximity	165	1.0735	2.9256
Rate	Bus Stops	Proximity	165	0.9503	2.5866
Rate	Official Buildings	Proximity	165	0.6014	1.8247
Rate	Schools	Proximity	220	0.3470	1.4148
Rate	Intercept	--	--	-8.0691	--
Overdispersion	Intercept	--	--	-0.2693	--

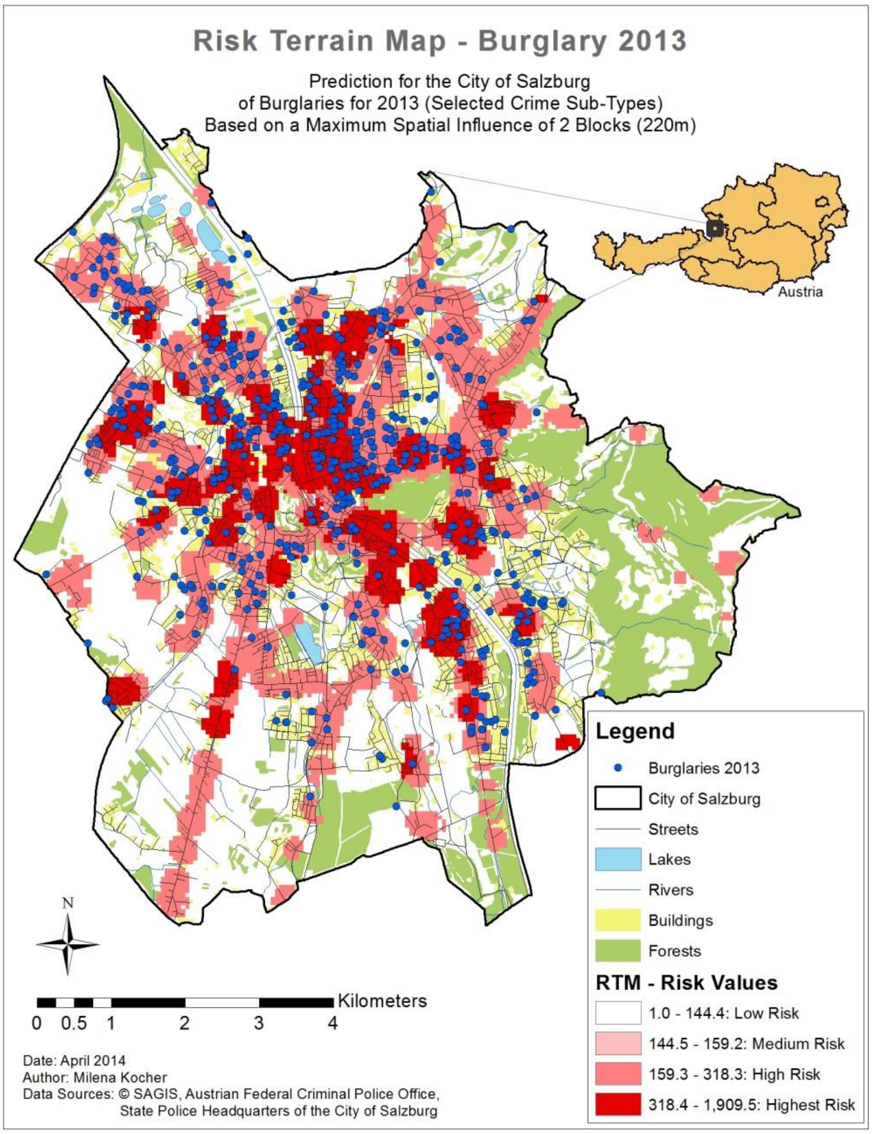

Figure.A 11 – Prediction for selected burglaries for 2013 (2 Blocks)

Table.A 12 and Figure.A 12 present the results for selected burglaries based on a block length of 330m.

Table.A 12 - RTMDx report for selected burglaries for 2013 (3 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Buildings Residences 3Blocks	Proximity	55	5.2140	183.8279
Rate	Bus Stops	Proximity	165	0.8791	2.4086
Rate	Pawn Shops	Proximity	330	0.8650	2.3750
Rate	Schools	Proximity	330	0.4876	1.6285
Rate	Intercept	--	--	-8.7905	--
Overdispersion	Intercept	--	--	-0.2787	--

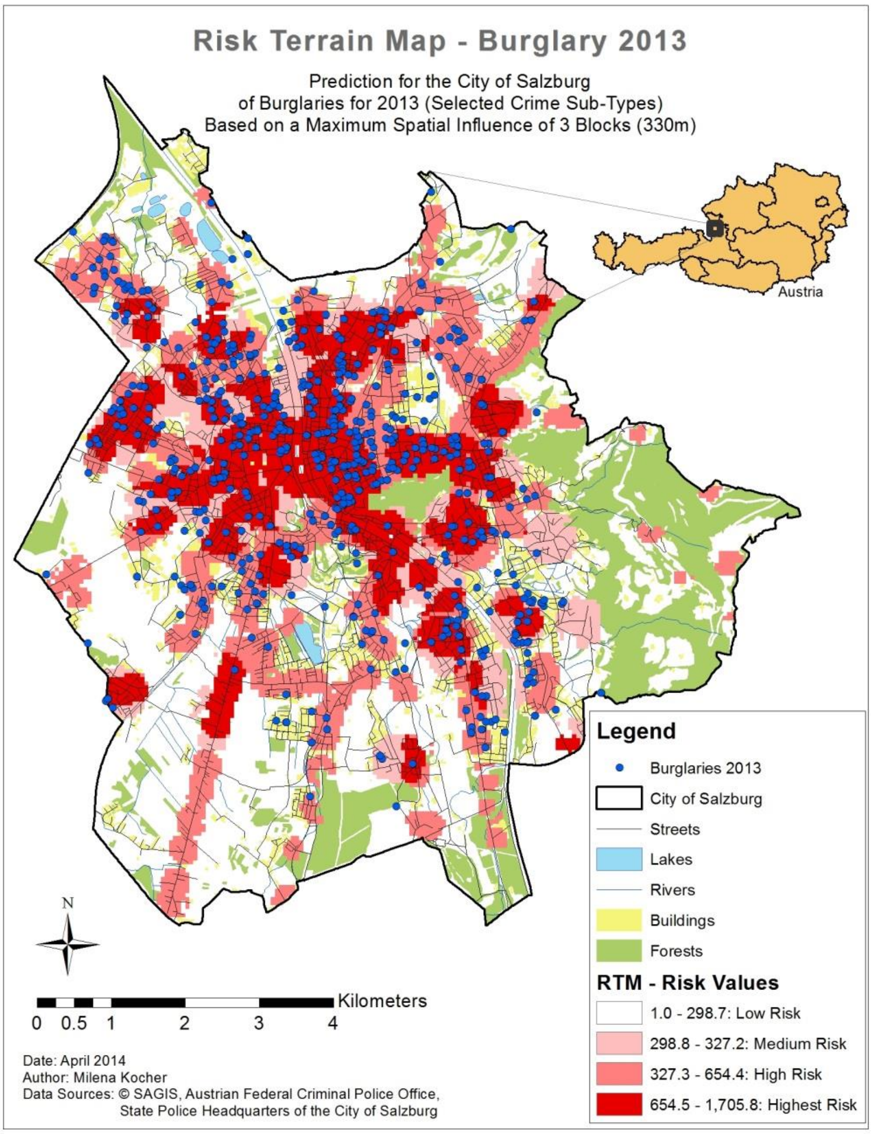

Figure.A 12 – Prediction for selected burglaries for 2013 (2 Blocks)

The next section shows the results for robberies. The best model was calculated with a maximum spatial influence of two blocks. At first, the result based on a maximum spatial influence of one block is presented (see Table.A 13 and Figure.A 13).

Table.A 13 – RTMDx report for robberies for 2013 (1 Block)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bars and Pubs	Proximity	110	1.7646	5.8392
Rate	Banks	Proximity	110	1.4838	4.4097
Rate	Bus Stops	Proximity	110	1.2753	3.5798
Rate	Intercept	--	--	-6.7084	--
Overdispersion	Intercept	--	--	-0.3823	--

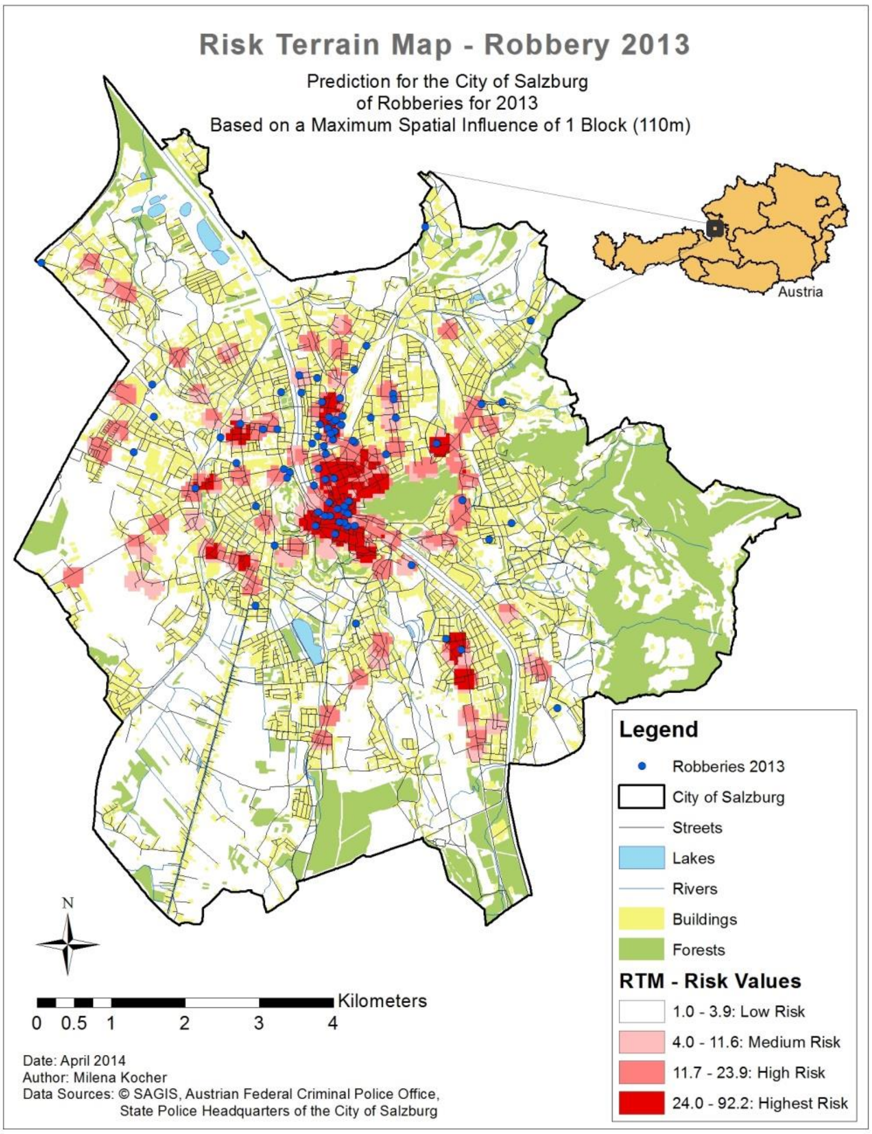

Figure.A 13 – Prediction for Robberies for 2013 (1 Block)

Table.A 14 and Figure.A 14 show the result for the RTMDx report and the prediction based on a maximum spatial influence of three blocks.

Table.A 14 - RTMDx report for robberies for 2013 (3 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Bus Stops	Proximity	165	1.6082	4.9938
Rate	Leisure and Fastfood Outlets	Proximity	275	1.3412	3.8236
Rate	Cash Points	Proximity	330	1.2432	3.4667
Rate	Bars and Pubs	Density	165	1.2319	3.4277
Rate	Intercept	--	--	-7.7132	--
Overdispersion	Intercept	--	--	-0.4217	--

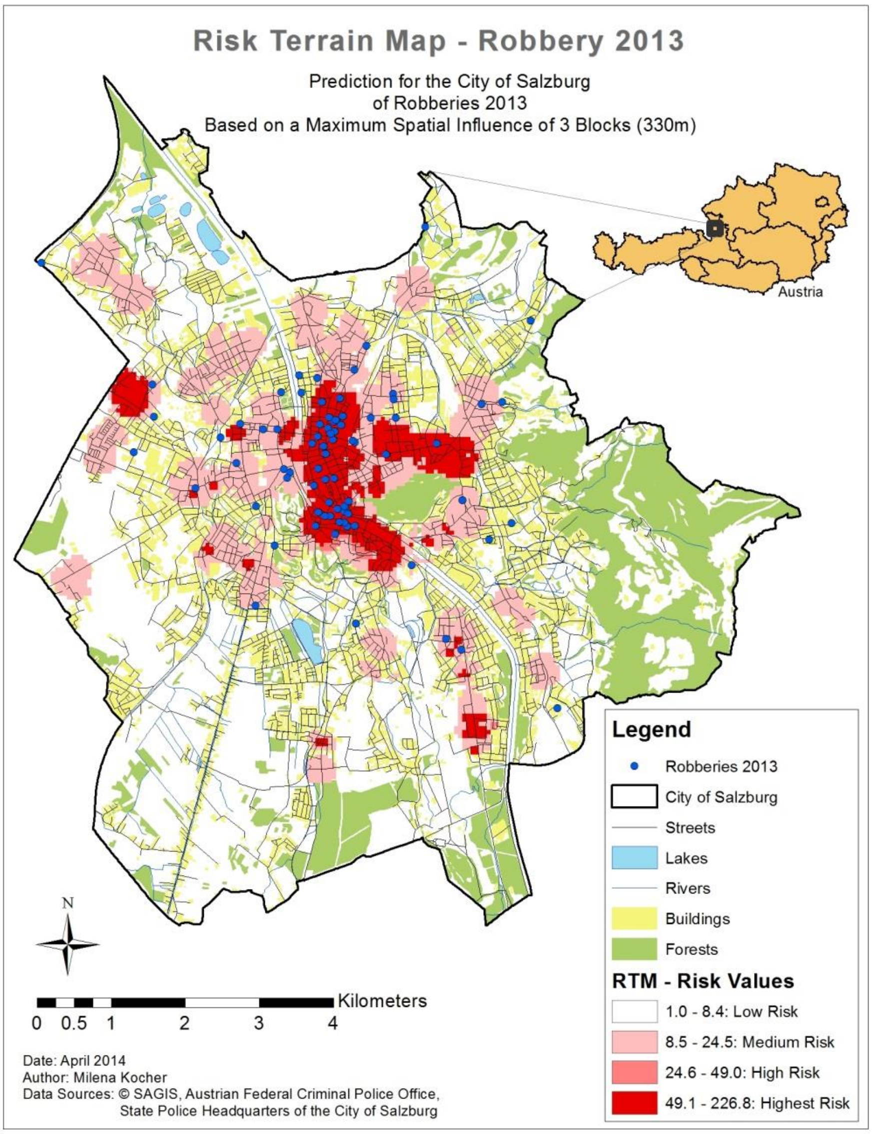

Figure.A 14 - Prediction for robberies for 2013 (3 Blocks)

In Table.A 15 and Figure.A 15, the results for robberies based on a maximum spatial influence of 4 blocks are presented.

Table.A 15 - RTMDx report for robberies for 2013 (4 Blocks)

Type	Name	Operationalization	Spatial Influence	Coefficient	Relative Risk Value
Rate	Cash Points	Proximity	385	1.7221	5.5963
Rate	Bus Stops	Proximity	165	1.4420	4.2291
Rate	Leisure and Fastfood Outlets	Proximity	275	1.2534	3.5022
Rate	Bars and Pubs	Density	165	1.1901	3.2874
Rate	Intercept	--	--	-8.0030	--
Overdispersion	Intercept	--	--	-0.4254	--

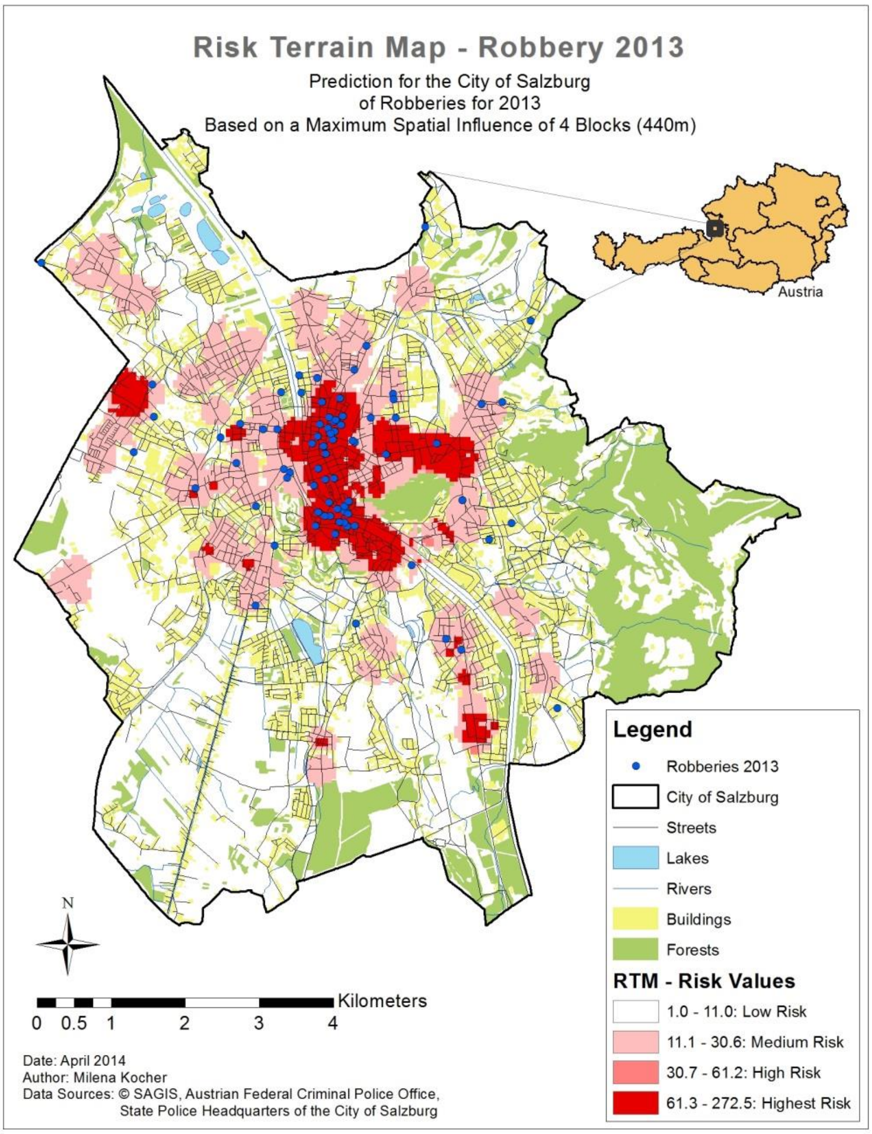

Figure.A 15 - Prediction for robberies for 2013 (4 Blocks)