Comparison of selected crime prediction methods for an U.S. and an Austrian urban area

by

Heidrun Gudrun Dalmatiner

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

Supervisor

Internal & External Supervisor: Dr. Michael Leitner

Department of Geography and Anthropology, Louisiana State University, Baton Rouge, U.S.

Baton Rouge, May 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 theses and I have always indicated their origin.

Baton Rouge, 26.05.2014 Heidrun Gudrun Dalmatiner

Heidrun Gudrun Dalmatiner

Acknowledgments

Ich möchte mich vor allem bei meinem Betreuer Professor Michael Leitner für seine Unterstützung und Anregungen während meiner Zeit in den Vereinigten Staaten von Amerika bedanken und dafür, dass er mir die Möglichkeit geboten hat, mein Berufspraktikum an der Louisiana State University in Baton Rouge zu machen. Außerdem möchte ich der Marshall Plan Foundation und der Fachhochschule Villach für ihre finanzielle Unterstützung danken, ohne die es mir nicht möglich gewesen wäre, mein Berufspraktikum in den USA zu realisieren. Diese Zeit wird mir immer sehr gut in Erinnerung bleiben.
Weiters danke ich besonders meinem Partner für seine Hilfe und Unterstützung während meines gesamten Studiums.

Zusammenfassung

Das Ziel dieser Forschungsarbeit und der daraus resultierenden Bachelorarbeit ist es, Unterschiede in den Kriminalitätsmustern der Vereinigten Staaten von Amerika und Österreich herauszufinden. Es sollen von den bereits in den USA verwendeten Vorhersagemethoden von Delikten die am besten für Österreich geeigneten gefunden werden. Dabei ist eine Fragestellung, ob die in den USA verwendeten statistischen Methoden für die Vorhersage von Delikten in Österreich adaptiert werden müssen.

Vorhersagemethoden in der Kriminalanalyse sind besonders von Bedeutung, da sie dabei helfen können Verbrechen zu verhindern bevor sie passieren. Es gibt drei wichtige Methodenansätze für Vorhersagemethoden:

1. Cluster- oder Hot Spot - Analyse
2. Near repeat - Analyse
3. Risk Terrain Modeling (RTM)

Der Fokus in dieser Bachelorarbeit liegt auf den ersten beiden Methoden, die auch als retrospektive Vorhersagemethoden bekannt sind, da die Vorhersage zukünftiger Verbrechen auf bereits begangenen Verbrechen aufbaut. Der Risk Terrain Modeling – Ansatz wird von Milena Kocher in ihrer Bachelorarbeit behandelt und darum wird hier nicht näher darauf eingegangen. Das Ziel dieser Forschungsarbeit soll durch die Anwendung verschiedener statistischer Vorhersagemethoden an Datensätzen einer US-amerikanischen Stadt (Houston, Texas) und einer österreichischen Stadt (Wien) erreicht werden. Die am besten geeigneten statistischen Techniken werden angewendet und miteinander verglichen. Gewählte räumliche Clustermethoden sind die Kerndichteschätzung, Nächste Nachbar Hierarchische Clusteranalyse, local Moran's I und Getis-Ord Gi*. Für den zweiten Ansatz der Vorhersagetechniken, die Near-Repeat-Analyse, wurde die Software Near Repeat Calculator verwendet. Ergebnisse der räumlichen Clustermethoden werden evaluiert und mittels Prediction Accuracy Index (PAI) und Hit Rate interpretiert. Außerdem wird eine visuelle Interpretation der Ergebnisse durchgeführt.

Abstract
The goal of this research project is to explore differences in crime patterns between the U.S. and Austria and to discover which crime prediction method, which has already been used in the U.S., could also be applicable to crime prediction in Austria. One particular research question that needs to be answered is whether or not crime prediction methods developed in the U.S. need to be adapted before they can be used in Austria. This is important because crime prediction may help to prevent crime before it happens. Today there exist three important sets of methods for crime prediction: first, crime density mapping, crime cluster or hot spot mapping; second, near-repeat analysis; and third, Risk Terrain Modeling (RTM). The focus in this bachelor thesis will be on the first two methods, which are also referred to as retrospective prediction methods. The RTM approach, which is a prospective crime prediction method will not be discussed, since this approach will be researched in the bachelor thesis by Milena Kocher. The goal of this research will be achieved by carrying out statistical prediction methods on a crime data set from an U.S. city (Houston) and an Austrian city (Vienna) with the best suited statistical techniques being applied and compared. Chosen spatial cluster methods are kernel density estimation, nearest neighbor hierarchical clustering, local Moran's I, and Getis-Ord Gi*. For the second approach for crime prediction the near-repeat analysis was carried out with the software program near repeat calculator. Results from the spatial cluster methods are evaluated and interpreted with the Prediction Accuracy Index and the Hit rate. Also, a visual interpretation of the results was done.

Table of Contents

List of Abbreviations

BKA	*Bundeskriminalamt*
GIS	*Geografisches Informationssystem*
KDE	*kernel density estimation*
LMI	*local Moran's I*
MAUP	*Modifiable Area Unit Problem*
NNHC	*nearest neighbor hierarchical clustering*
PAI	*Prediction Accuracy Index*
UCR	*Uniform Crime Reporting*
U.S.	*United States of America*

1. Introduction

1.1 Motivation

Crime prediction is an important topic for every country, because it can help to lower the crime rate, which leads to a more secure, peaceful, and stable society. These are fundamental rights of every person and should be guaranteed. The primary goal of this bachelor thesis is to find and test methods to predict where and when crime events happen and to possibly prevent them, before they can happen. It is hoped that this research can make a contribution to crime prevention efforts.

The main reason why this research project is done in the U.S. at the Louisiana State University in Baton Rouge is, because the external supervisor, Professor Michael Leitner is an expert in crime modelling and mapping in general, and in the prediction of crime, in particular. In addition, criminal predictive analytics is still in its infancy in Austria. However, the Austrian police is very keen on predictive crime modelling approaches and would like to implement those approaches into their proactive decision-making processes.

1.2 Problem definition

In 2012 there were 548,027 reported cases of criminal offences in Austria (Statistik Austria 2013). Looking at this number of crimes, there is definitely a need to predict when and where crime happens and to possibly prevent it, before it occurs.

Exemplary crime prediction approaches already exist in the U.S., especially for cities, where they can be used to optimize the prediction of crime by using statistical analysis integrated in software programs. These programs are often based on Geographic Information System (GIS) technology.

The Geographic Information System (GIS) has been used in Austria since 2004 at the Federal Criminal Police Office to visualize crimes. Criminal analysts can then find hot spots and recognize connections and reasons for crimes being committed. It is also used for measures of prevention. Spatial analyses are used for the visualization of crime situations and crime centers, evaluation of offender groups, data, etc. (Kampitsch et al. 2008, p. 1). However, approaches for crime prediction, like in the U.S., have not yet been applied in Austria. So it would be important to propose, adapt, and integrate a suite of criminal predictive policing approaches into Austrian law enforcement agencies.

The analysis of crime is important to find out about the reasons why crime happens and to reduce crime on the basis of crime prediction and then crime prevention. When it is known beforehand, where and when crime will happen, then it can possibly be prevented, for example, by increasing the number of police officers, by installing surveillance equipment, or even by targeted urban planning, including more light in dark places, etc.

1.3 Method of solution

Crime has a geographical quality (Chainey and Ratcliffe 2005 cited Chainey et al. 2008) and does not happen randomly. In fact there are certain places of high crime concentration and these are called hot spots (Chainey et al. 2008, pp. 1).

In this research different methods for finding hot spots for predicting crime were described and individual methods were selected and used to find the best suitable technique for the different datasets and the two selected cities. Evaluation tools like the hit rate and the prediction accuracy index (PAI) are used to compare the individual methods and to find out which is the best method for crime prediction. This research will analyse differences and similarities between Houston and Vienna and recommendations will be made if methods can be applied successfully for the Austrian capital.

1.4 Expected results

The expected results of this research are:

- Summary of existing methods for spatial crime prediction
- Comparison of methods and evaluation of their advantages and disadvantages
- Practical use of the best methods on data sets from one large U.S. and one large Austrian urban area
- Comparison of the results from crime prediction methods between the U.S. and Austria
- Identifying useful methods for crime prediction in Austria
- Visualization of prediction results of the different methods used

1.5 Structure of the thesis

The second chapter of this thesis deals with the theoretical basis for hot spot detection and analysis and describes the theory behind crimes and cluster analysis methods. It ends with the description of mapping types and other crime modelling techniques. Chapter 3 describes the methodology according to the stated problem definition and the method of solution in detail. It also covers the implementation of the methods. In chapter 4 the results are shown as well as their evaluation and interpretation. Chapter 5 includes the discussion with a critical reflection about the methods and the entire work, and if the goals have been reached. The summary and conclusions, as well as future research are discussed in the last chapter.

2. Theoretical background

This chapter describes the relevant theories behind hot spot analysis and deals especially with cluster analysis methods. It then describes best practice projects.

2.1 Methods for crime prediction and theories

There are different crime **hot spot theories**, which form the theoretical basis for explaining the occurrence of crime in the form of spatial and spatiotemporal clusters. They include place theories, street theories, neighborhood theories, and others (FBI 2012). Another set of theories for explaining why crime happens at particular places include the routine activity theory, landmark discoveries, social disorganization theory, social efficacy, broken windows theory, and crime opportunity theories (Filbert 2008, p. 5). Relevant theories will be discussed below and their usefulness for crime prediction, as a theoretical basis, will be assessed.

Today there exist **three important sets of methods for crime prediction**: First, crime density mapping, crime cluster, or hot spot mapping; second, near-repeat analysis, which is based on the fact that there is an increased likelihood of crime to happen in areas, where crime has happened before; and third, Risk Terrain Modelling (RTM), which applies so-called criminogenic factors to predict where and when future criminal events are most likely to happen. The focus in this bachelor thesis will be on the first two methods, which are also referred to as retrospective prediction methods. This means that where and when crime happens is "predicted" for the immediate past and then the predicted (modeled) crime events are compared with the actually happened crime events. If both the predicted and the actual crimes happen in similar places and at similar times, then it can be assumed that those methods can also be used to predict crimes for the near (actual) future. The RTM approach, which is a prospective crime prediction method will not be discussed, since this approach will be researched in the bachelor thesis by Milena Kocher.

There are several spatial statistical tools for the spatial and spatiotemporal prediction of crime events in a retrospective manner. Such tools will be discussed next in more detail.

As a preliminary analysis, a standard method to determine the type of the crime pattern (random, clustered, or regular) can be used. An example for this method is the **nearest neighbor analysis statistic** (Chainey and Ratcliffe 2005, p. 126). It is a global statistic that tests for whether **spatial clustering** exists in the study area (Eck et al. 2005, p. 17).

The **nearest neighbor statistic** is a distance statistic for point pattern data sets and determines the degree of clustering of the points. It compares characteristics between pairs of closest points with distances between points that are randomly placed (Chainey and Ratcliffe 2005, p. 126).

The nearest neighbor analysis measures the area per point by the spacing between points (Lee and Wong 2001, p. 72).
To get the nearest neighbor index you have to calculate the nearest neighbor distance. That is the distance from one point to the closest other point. The distances between all pairs of points are calculated and the average of all distances is taken as average nearest neighbor distance. The nearest neighbor index tells you about the difference of the distances found between the nearest points of the chosen data set and the distances between random points (Levine 2013a, p. 6).
The scale for the nearest neighbor index ranges from 0 (completely clustered) over 1 (random) to 2.149 (completely dispersed). So 0 would mean that all points are at the exact same location. A result of 2 or more indicates dispersion between points (Lee and Wong 2001, p. 75).
The next step is to **determine and analyze crime hot spots**. The most important methods include grid cell analysis, such as the kernel density estimation, and several methods that can be summarized under the name cluster analysis (Grubesic and Murray 2001, p. 3).
To determine crime hotspots Jefferis and Mamalian (1998) and Craglia et al. (2000, both cited Grubesic 2006) recommend different techniques like visual interpretation of maps, choropleth mapping, spatial autocorrelation, kernel density estimation, and cluster analysis (Grubesic 2006, p. 3).
A **hot spot** is defined as an area with a higher concentration of criminal events than can be expected by chance (Filbert 2008, p. 4).
Another definition comes from Braga & Weisburd (2010 cited Levine 2013b, p. 6): "Typically called *hot spots* or *hot spot areas*, these are concentrations of incidents within a limited geographical area that appear over time".
Crime tends to accumulate in specific areas, which are called hot spots. Police efforts should be concentrated at these hot spots to reduce crime. This has been shown in practical studies. Police efforts could be patrols at hot spots or solving the underlying problem, for example bad lighting at a place. Crime incident data will be analyzed to find hot spots. Crime maps depict data and show patterns and hot spots (Filbert 2008, p. 4).
There are many techniques for hot spot analysis, with the more advanced techniques being calculated with GIS software extensions (Chainey and Ratcliffe 2005, pp. 145).
Hot spot analysis forms the foundation why there are crimes happening at special places. Environmental factors like the shape of an area or proximity to certain services have an influence on the likelihood of crime (Greenburg and Rohe, 1984 cited Grubesic 2006, p. 2).
Hot spots can be classified into different types like repeat places hot spots, repeat victimization hot spots, repeat street hot spots, and neighborhoods and other area hot spots (Eck et al. 2005, pp. 10).

Crime hot spot theories

Hot spots can cover places, streets, or whole areas. The selection which types to choose depends on the research question someone wants to answer (Eck et al. 2005, p. 9).

Hot spots can vary in size and this influences the mapping and analysis techniques that can be used (Filbert 2008, p. 5).

Place theories explain why crime happens at certain places. Crimes are mapped as points and the units of analysis are addresses, street corners, etc. (Eck et al. 2005, p. 3).

Street theories discuss crime incidences related to specific streets or blocks. Analysis can be done for street segments or highway sections, which are best displayed as lines on maps. **Neighborhood theories** want to explain differences among neighborhoods and cover large areas. So the hot spots are best displayed as polygons. Other **large area theories** contain a few cities or regions. **Repeat victimization theories** want to explain why someone is often a victim (Eck et al. 2005, pp. 3).

Place theories are important for the analysis of this author. Address data are available and the crimes can be mapped as points. So the theories for repeat places hot spots will be described closer now.

A place is a small location, like an address or a house. Places have a single owner and a function, for example as a residence or school (Eck et al. 2005, p. 4).

When you look at a high-crime area it can be seen that crime is concentrated at a few places and not everywhere in the area. So crime incidents are best displayed by points (Eck et al. 2005, p. 4).

Causes for that phenomenon are tried to be discovered with different theories. To explain the causes for crime concentration at special places the routine activity theory can be used. This theory identifies crime opportunities as a reason for crime (Filbert 2008, p. 5).

The Routine activity theory focuses on the behavior of the place manager (that are persons who own the place or are their representatives). This theory looks at how behavior is being regulated at specific places. If behavior is regulated, crime incidents will be minimized. If behavior is not regulated, crime incidents are more likely to happen at this place. A regulation has three effects: 1. It prevents crime because of early intervention (for example the alcohol provided to people is limited to a certain amount); 2. More people who search for a well-regulated location come here (and these people tend to make no problems); and 3. Less people who look for a weakly regulated location go there (these people will more likely produce trouble). A non-functioning place management is a core problem and repeat places (that are places where crimes happen often) are in many cases staying the same over time (Eck et al. 2005, pp. 4).

There are different ways of police solutions for the existing problems. For example, situational crime prevention can be used (e.g., increased lighting) to help reduce crime at hot spots. Also more police patrols that stay for 10 to 15 minutes can help reduce crime (Filbert 2008, p. 5).

2.2 Cluster analysis methods

Cluster analysis is a popular approach for detecting hot spots (Grubesic and Murray 2001, p. 2).

Cluster analysis techniques are grouping nearby incidents together into clusters (Levine 2013b, p. 7). A cluster has the characteristics of homogeneity and separation (Cormack 1971 and Gordon 1999 cited Everitt 2011, p. 7).

Homogeneity refers to the similarity inside of a cluster (the closeness of points) and separation applies to the outside of the cluster (other points are farther away) (Everitt 2011, p. 7).

2.2.1 Types of Cluster Analysis Methods

Levine (2013b, pp. 8) describes the following classification of cluster analysis methods in the CrimeStat User Handbook:

1. Point locations

Spatial Mode and Spatial Fuzzy Mode can be calculated in CrimeStat. Places with the highest number of incidents are counted and declared as hot spots. For more detail refer to Levine 2013b, p. 7.

2. Hierarchical techniques

The hierarchical clustering methods are explained in more detail in subchapter 2.2.3 of this thesis.

3. Partitioning techniques or non-hierarchical techniques are also explained in more detail in subchapter 2.2.3 in this thesis.

4. Scan statistics use a search circle for each incident point or node of a grid. For more detail refer to Levine 2013b, p. 9.

5. Density techniques search for dense concentrations. One example of this technique is the kernel density estimation. For more detail refer to Levine 2013b, p. 9.

This technique is also described in more detail in subchapter 2.3.1 kernel density estimation.

6. Clumping techniques rely on the partitioning method, but it is possible that a point can belong to two or more clusters. This is not possible for a partitioning technique. For more detail consult Levine 2013b, p. 9.

7. Risk-based techniques use an “at risk” variable like for example population.

8. Zonal clustering methods are finding zones with high or similar degree of an attribute. See Levine 2013b, p. 11 for more detail.

9. Miscellaneous techniques exist, but they are not that often used. More detail about these techniques are provided in Levine 2013b, p. 11.

2.2.2 Optimization criteria

Optimization criteria should be decided upon and they inform about which method to use. Also is important:

- which type of cluster someone wants to have (for example, if individual points can belong to two different clusters, or not)
- if there are other variables except x and y, that should be considered (weighting or intensity variables)
- how the distance between points is measured
- the number of clusters (if there is a fixed number or an undetermined number)
- the geographical scale
- the starting points of cluster locations (defined from the user or mathematically)
- the different optimization routines (algorithms)
- the visual display (convex hull, ellipse, etc.)

(Levine 2013b, pp. 11).

2.2.3 Divisions of cluster analysis methods

Cluster methods can also be divided into global and local methods. Global methods deal with the measurement of the average tendency while local methods identify individual clusters. Local methods have more advantages because they can find the location of clusters and they are measuring the significance for all found clusters (Quick and Law 2013, p. 4).

There exist hierarchical and non-hierarchical (or partitioning) clustering methods. In any case a (n x n) matrix is first calculated with the degree of similarities between every data pair. Similar pairs are grouped together in clusters. The nearest neighbor distance is often used as a similarity measure in hierarchical clustering methods (Grubesic and Murray 2001, pp. 5).

The nearest neighbor measure takes the distances between two points and the average distance that is found between all points and compares them. A new cluster is built if the distance between two points fulfils a defined criterion (Grubesic 2006, p. 4).

First-order and second-order clustering is done until all sub-clusters are included in a single cluster. The hierarchical clustering analysis shows crime concentrations in small geographical areas and connections between the first- and second-order hierarchies (Grubesic and Murray 2001, pp. 5).

The hierarchical clustering methods can be divided into agglomerative and divisive techniques. CrimeStat uses an agglomerative hierarchical method with nearest neighbor clustering (Grubesic 2006, p. 4).

While the agglomerative method starts with every single incident and groups them together to cluster until only one cluster is left or the criterion failed, the divisive method does the opposite. It starts with one group with all incidents in it and split them into clusters (Grubesic 2006, pp. 4).

The problem with hierarchical clustering techniques is that the previous step (joining of incidents or splitting) cannot be undone, even when it is not correct (Grubesic 2006, p. 5).
Non-hierarchical or partitioning cluster methods split incidents into a predefined number of groups (k) (Grubesic 2006, p. 5). One example for such a method is the k-means cluster technique (Grubesic 2006, p. 6).
A disadvantage for the non-hierarchical techniques is to find the best number of clusters because the examiner has to define it first (Grubesic and Murray 2001, pp. 5).
Also there is the fundamental problem that incidents are assigned to clusters on an all-or-nothing-basis. That involves problems with outliers or intermediate incident locations. Outliers influence the building of a cluster (Grubesic 2006, pp. 9).

2.3 Detailed discussion of cluster analysis methods used in this thesis

In this chapter the chosen methods are explained in more detail. It covers four spatial cluster analysis methods and one spatio-temporal cluster analysis method. From all spatial cluster methods, the kernel density estimation, nearest neighbor hierarchical clustering, local Moran´s I, and Gi* are picked. The near repeat calculator which tells about the near repeat phenomenon was chosen for spatio-temporal analysis.

2.3.1 Kernel density estimation

The kernel density estimation is a popular method to identify crime hot spots. A grid is placed over a point pattern map of crime incidents. Every grid is assigned a density value, which is calculated based on the number of crime incidents located inside and in the immediate vicinity of that cell. A high density value represents a high number of crime incidences, whereas a low value represents a low number of crime incidences. Density values can be subsequently classified and visualized with different colors in the final grid map (Filbert 2008, p. 6).
Interpolation allows generalizing specific locations of crime to an area with the help of density estimates. This is a Z-value and shows the intensity. Kernel density estimation is used for discrete point data (Levine 2013b, p. 2).
The kernel density estimation is also an interpolation technique for visualizing the crime distribution and identifying hot spots. A continuous surface represents the density of crimes (Eck et al. 2005, p. 32).

Kernel density estimation is “the most suitable spatial analysis technique for visualizing crime data” (McGuire and Williamson 1999; Williamson et. al. 1999, 2001; Chainey et al. 2002; Chainey and Ratcliffe 2005; Eck et al. 2005 cited Quick and Law 2013, p.5).

2.3.2 Nearest neighbor hierarchical clustering

This method uses the nearest neighbor analysis as a basis and its goal is to identify clusters of incidents (CrimeStat3_Workbook.pdf, p. 52).
The nearest neighbor hierarchical clustering method groups incidents together that are close to each other. All points will be grouped until only one cluster is left or the criterion for the clustering was not fulfilled (Levine 2013b, p. 21).

2.3.3 Local Moran´s I

Local Moran´s I (LMI) has often been used in crime research in different fields. LMI uses a contiguity matrix for detecting spatial relationships (Quick and Law 2013, p. 6).
It compares nearby values with the overall mean (Quick and Law 2013, p. 4).
The output of LMI consists of five classes (insignificant result of clustering, High-High, Low-Low, High-Low and Low-High clusters). The High-High category for example means that there are high values which neighbors have also high values – this stands for a hot spots. In comparison to that a High-Low cluster would be a spatial outlier, where a high value is surrounded by lower values (Quick and Law 2013, p. 6).

For the calculation of the local Moran´s I the data points that represent individual incidents have to be aggregated to spatial areas (zones or polygons) (Levine 2013c, p. 4).
The distance between the centroids (the centers of the zones, which are depicted as single points) is observed. Also an attribute variable is needed (for example the number of assaults within a zone) (Levine 2013c, p. 6).
The local Moran´s I tells if zones (or polygons) are similar or different to their neighbors (Levine 2013c, p. 7).
A high attribute value zone next to other high attribute value zones result in a positive local Moran´s I value and shows similarity of the zone to the other zones.
A high attribute value in a zone next to low values will lead to a negative local Moran´s I value. This indicates dissimilarity between the zones. In this case the zone would be a spatial outlier (Levine 2013c, p. 8).
The local Moran´s I indicates the similarity or dissimilarity of a zone (or polygon or raster cell) to its neighbors (Levine 2013c, p. 8).
Local Moran´s I compares zones to their neighboring zones and tells which are unusually low or high (Bruce and Smith 2011, p. 63).

2.3.4 Getis-Ord Statistic (Gi*)

The Gi* statistic is analyzing the sum of nearby values. It is similar to the local Moran´s I (Quick and Law 2013, p. 4).

It also deals with zones (or polygons) and looks at polygons within a certain distance.

Features are observed within the context to nearby features. A high value-feature surrounded by other high value-features can be a hot spot. Their local sum will be compared to the sum of all features in the data set (ArcGIS 9.3 Tool Help 2009b).

2.4 Other crime modelling techniques: the near repeat analysis and the Risk Terrain Modeling method

Other crime modeling techniques, besides the already discussed hot spot mapping methods, are the near repeat analysis and the Risk Terrain Modeling approach. The Risk Terrain Modeling approach shows regions with their risk values for a particular crime to happen. For example, possible risk factors in the study from Irvington, NJ included gang members, bus stops, schools, bars, etc. (Kennedy et al. 2011, pp. 13). The near repeat analysis is also used in this research, so it will be explained in more detail next.

The **near repeat analysis** considers the circumstance that in the immediate surroundings of a place where a crime has happened, the chance of a future crime to happen is increased. This is referred to as the near repeat phenomenon. In other words, this approach also takes the temporal component between crimes into consideration. An analysis of the New Jersey Police showed a 500% higher chance of a near repeat crime to happen for the next seven days at the place where a crime has happened before and a 153% higher chance for the surrounding area for the next 14 days (Kennedy et al. 2011, pp. 13).

The near repeat calculator is a software program to look for near repeat phenomena. A near-repeat phenomenon was identified in earlier research for burglary that means that there is a risk for repeated burglaries in the surroundings of the first burglary for the near future. Also near-repeat patterns for gun shootings have been detected (Ratcliffe and Rengert 2008, p. 1).

Youstin et al. (2011) described in the article "Assessing the generalizability of the near repeat phenomenon" the findings of near repeat patterns for the crime types shootings, robbery, and auto theft. The authors found near repeat patterns for all crime types, but every crime type has a special spatiotemporal pattern (Youstin et al. 2011, p. 1).
Near repeats for robbery lie within one day and up to a distance of 500 meters (Grubesic and Mack 2008 cited Youstin et al. 2011, p. 3).

The near repeat phenomenon is a special kind of repeat victimization (Youstin et al 2011, p. 2).
The most important types of repeat victimization are true repeat victimization and near repeat victimization. True repeat victimization would mean that the same place (for example a house) or person is again the

target of a crime. Near repeat victimization means that targets are not exactly the same as before but they are located in the immediate proximity (in time and space). Research results proved that the risk of being again a victim is significantly higher than for other persons or places (Youstin et al. 2011, p. 4).

The difference between repeat victimization and hot spots (which are specific areas with a higher amount of crime) lies therein that hot spots are including multiple targets of crime and various crime types. Crimes in hot spots are also only near in space but may not be near in time (Youstin et al. 2011, p. 4).
Much research has been done about the repeat victimization for the crime type burglary, but nearly all crime types (only murder or manslaughter is an obvious exception) include repeat victimization (Townsley, Homel and Chaseling 2003 cited Youstin et al. 2011, p. 4).

Townsley et al. found out in their study that no diversity of the houses (appearance and construction) leads to fewer repeat victimization. Diversity offers a choice to offenders and the same attractive targets can be again the victim of a burglary. There is no reason why a house that shows no real differences to the surrounding houses should be a preferable target (Townsley et al. 2003, p. 1).

There are two explanations for repeat victimization: First, risk heterogeneity, which means that persons or targets have special characteristics that increase the risk of becoming a victim again. Second, state dependence, that includes the context where the victimization takes place. Locations have certain characteristics that support crime while there are other locations that are not so suitable as crime areas (Caplan et al. 2013, p. 6).

From a happened crime in the past nearby incidents in the future can be predicted because of the existence of the near repeat phenomenon (Bowers & Johnson 2005 cited Caplan et al. 2013, p. 6).

Near-repeat analysis includes the time and can tell with a distinct statistical confidence level about new crimes happening in the near future in the neighborhood of past crimes (Short et. al 2009 cited Caplan et al. 2013, p. 10).

2.5 Mapping types

Chainey et al. (2008) lists point mapping, thematic and grid thematic mapping, spatial ellipses, and kernel density estimation as mapping types for visualizing hot spots (Chainey et al. 2008, p.1).

Hot spot mapping uses data from past incidents to possibly tell where crime will happen next. When looking at retrospective patterns of past crimes it should help to predict crime places for the future (Chainey et al. 2008, p. 2).

There are different mapping types to visualize a concentration of crimes. A simple method is **point distribution mapping**, where each point represents a crime event on a map. But there are disadvantages, since it can be difficult to identify the actual hot spot, e.g., the shape or the size or the location of the hot spot (Chainey and Ratcliffe 2005, pp. 148).
Every incident is in the same color as its associated cluster (Grubesic 2006, p. 10).
In addition, there is also a visualization problem when many crimes occur at the exact same location (Chainey and Ratcliffe 2005, p. 149).

An alternative to point distribution mapping is **choropleth mapping (or geographic boundary thematic mapping)** which displays the counts of crimes or crime rates inside their geographic region. Problems of this method are that it can be difficult to determine the exact hotspot area within a particular geographic region. For example in the case that the entire region has a high crime rate, there may be also smaller sub-regions where little or no crime has happened. In addition, differently sized and shaped boundaries for a chosen area may lead to different results. This is referred to as the well-known modifiable areal unit problem (MAUP).
Finally, class boundaries should be determined in a way that they are easy to understand from the map viewer (Chainey and Ratcliffe 2005, pp. 150).
MAUP also occurs when for example detected hot spots on a census block level are not even found when someone looks at the regional level of crime (Harries 1999 cited Grubesic 2006, p. 2).
Different cluster patterns are found on different scale levels (Quick and Law 2013, p. 18).

Another approach is **grid thematic mapping**, which uses a grid with regularly sized cells (mostly rectangles or quadrats), for which the number of crimes falling into each cell are aggregated. A point-in-polygon operation is used to calculate the number of crime events in each cell. This method tends to show a crime pattern better. However, it is difficult to choose the best grid size and the MAUP can also be an issue (Chainey and Ratcliffe 2005, pp. 153).

The grid also influences how hot spots look like (because of the cell boundaries) (Chainey et al. 2008, p.5). The author of this thesis used grids for local Moran´s I and Getis-Ord Gi* methods.

Maps for repeat places

Dot maps are the best choice for mapping repeat places crimes. Graduated symbols are one option, but the different size of the dots could lead to overlapping of other points. A better solution would be color gradient dots (for example from yellow through red), showing the extent of crimes happened at a place. Also repeat address mapping (RAM) would be possible, where only a few percent of the hot spot addresses are mapped (for example the 10 percent with the most crimes). The disadvantage of this displaying method is that not all hot spots are shown on the map (Eck et al. 2005, pp. 11).

Density maps and spatial statistics can be used to find hot spots. One such example is the **kernel density map**. This map can be done with the software CrimeStat, for example. Density maps are made out of a point map of crime incidents. An arbitrary grid is put on the map, where each grid cell has a score that represents the number of incidents (or points) (Filbert 2008, p. 6).

2.6 Best-practice projects

Exemplary crime prediction approaches already exist in the U.S., especially for cities, which can be used to optimize the prediction of crime by using statistical analysis integrated in software programs. These programs are often based on Geographic Information System (GIS) technology.

2.6.1 PREDPOL

One such example is "PREDPOL", a technology used at the Los Angeles Police Department to predict crimes using so-called hot spots. These are places, where crime has a high chance to occur. PREDPOL marks hot spots with boxes on a map and the police officers patrol more often in this area (PREDPOL 2013a). The program works with a model that was used to predict aftershocks after an earthquake. It is a model based on a computer algorithm that uses data of location, time, and type of crime (Risling 2012). The results in the Los Angeles Division where it was first tested show a reduction of the crime rate by 13% in the four months after first using the PREDPOL software (PREDPOL 2013b).

2.6.2 Computer Statistics Program (COMPSTAT) in New Jersey

The New Jersey report highlights city's crime hot spots. Mapping of hot spots helped to detect crime in Jersey City, NJ. The crime rate is very high in Jersey City and every three hours a crime incident is reported. Most of them are theft or robbery. The incidents were mapped and four areas with a high concentration of crime were found. These four crime hot spots have their own underlying crime problem. When a hot spot is detected, a closer look may identify the underlying problem. One of the crime hot spots is for example a shopping mall, where most crimes are shoplifting and theft. Another is a housing area where car break-ins are often reported. Measures to reduce crime are housing demolition or renovation in affected areas and increased police officers patrol in these areas. Also buildings that are not so high should be built in downtown in the future and more money should be invested in better street lighting, and a greater presence of police officers. COMSTAT is used at the Jersey City police since 2006. It originally came from COMPSTAT from the New York City Police Department (Geography & Public Safety 2008, pp. 12).

COMPSTAT can track crime incidents and generate pin maps. The police department can then analyze the provided information and react appropriately to the results. (Jersey City Police Department 2012, 2013).

3. Methodology

This chapter comprises the problem definition in more detail as well as the method of solution. The project area and which geodata have been used for the analyses, are described.

3.1 Problem definition

Crime could possibly be prevented by using crime prediction methods. To do this, an overview of existing statistical methods to predict crime is discussed and differences between these methods are explored by applying them to two different urban areas, including one Austrian (Vienna) and one U.S. city (Houston, TX). By comparing the methods between these two cities, differences between the U.S. and the Austrian city will be apparent. Existing crime prediction methods, which have already been successfully applied in the U. S., could be integrated into the Austrian police agencies and could possibly lead to crime prevention. Maybe some methods need to be adapted for their use in Austria or new crime prediction methods have to be developed.

3.2 Method of solution

How the research questions will be answered is shown in figure 1 in a conceptual model. First an overview over existing methods (cluster methods and near-repeat analysis) will be gained. A few cluster methods are selected and then implemented. The most appropriate parameters for the methods must be found. Then, the results are displayed and edited in ArcGIS and evaluated with the prediction accuracy index (PAI) and the hit rate. In a final step the results will be interpreted to find differences and similarities between the U.S. and the Austrian city.

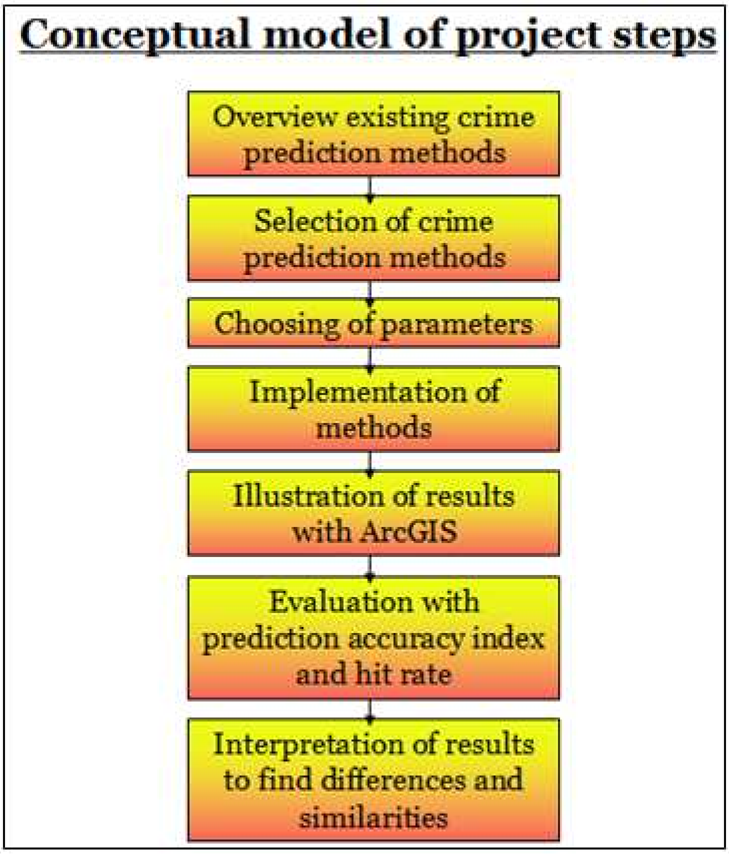


Figure 1: Conceptual model for this research project

3.3 Project area

The project area covers two cities, namely Houston, a U.S. city located in the state of Texas and Vienna, which is the biggest city in Austria, and also the capital city of Austria. These two cities were selected because they have a similar population total. In general, it is difficult to compare cities of two countries because of the differences, especially what is called a crime and how it is defined, which will be described closer in the subchapter Geodata.

Houston has about 2.1 million inhabitants (status 2013) and an area of about 1,500 km² (World Population Statistics 2013).

Vienna has about 1.7 million people (status 2013) and covers an area of about 415 km². The population density in Houston is 1,400/ km² and 4,200/ km² for Vienna (Lukacsy and Fendt 2013).

3.4 Geodata

The used geodata will be explained in detail in this chapter.

One of the most important steps in statistical analysis is the selection of data on which the analysis is based upon. That depends also on the data which is available. For a comparison between two countries an important first step is to find out how the laws about crime look like and how they differ and how the individual crime types are exactly defined. Therefore, the author of this thesis looked at the law texts of the United States, especially at the laws of Texas, where Houston lies. In comparison to that the law texts of Austria were studied and the classification of crimes as well as their definition, for example robbery, was examined.

3.4.1 Data for Houston, Texas, U.S.

In a first step, a selected number of crime types were chosen and, in a second step, it was checked if the selected crime data could be acquired. Data for Houston come from the Houston Police Department and they were provided by a LSU student in geography, Shuzhan Fan, who also needed similar crime data for Houston for his master thesis. Crime data sets from the city of Houston have been collected from the Houston Police Department through the Texas Public Information Act.

The crime data in the United States are collected as part of the uniform crime reporting program (short UCR). This is an effort to gain crime data for the entire United States, so crimes are assigned to eight part 1 crimes and 18 part 2 crimes, as well as 24 other offenses categories and other data. The data are collected from the state and then forwarded to the FBI´s national UCR Program or the data are directly sent from agencies to the national Program. The purpose is to gain statistics for crime for the whole United States, which are published every year. These crime data are free and open-access that means that anybody can get these crime data for research or other purposes (FBI 2004, p. 8).

The UCR include aggregate numbers of selected crime types. The participation of police departments is voluntary. However, the UCR statistics cover over 95 percent of the whole population of the United States (Boba 2008, p. 4).

Of interest for the author of this thesis are the data for Houston for the years 2011 and 2012. The data for the year 2013 are not completely available at the time of the writing of this thesis. The data were collected in excel-files and access-files and are sorted by months. They were all transferred to excel-files and made consistent with each other.

3.4.2 Data for Vienna, Austria

Crime data for the city of Vienna are provided by the Austrian Federal Criminal Police Office (Bundeskriminalamt (BKA)) and is in the form of address-level crime locations. The data are from the Security Monitor (SIMO) which includes all crimes reported in Austria. The author looked at the crime data of Houston to find the best fitting crime types of Austria. Then the chosen crime types were selected from SIMO for this study.

Selected Crime types

For Houston the selected crime types are robbery, aggravated assault, other assaults and burglary. The exact UCR codes are:

For violent crimes:

03 robbery
04 aggravated assault
08 other assaults
For non-violent crimes:
05 burglary

For Austria the selected crime types are:
For comparison with burglary from the UCR:
§ 129 "Diebstahl durch Einbruch oder mit Waffen" (burglary), including the selected subcategories:

- Firmen- und Geschäfts-Einbruchdiebstahl (burglary of companies and stores),
- Wohnhaus-Einbruchdiebstahl (burglary into houses),
- Keller-Einbruchdiebstahl (burglary into cellars)
- Wohnungs-Einbruchdiebstahl (burglary into flats)

§ 109 "Hausfriedensbruch" (domestic disturbance)

For comparison with assault from the UCR:
§ 83 "Körperverletzung" (assault)
§ 84 "Schwere Körperverletzung" (aggravated assault)
§§ 85-89 „KV mit schweren Dauerfolgen", „KV mit tödlichem Ausgang", „Absichtliche schwere KV", „Fahrlässige KV", „Gefährdung der körperlichen Sicherheit" (other forms of assault)

For comparison with robbery from the UCR:
§ 142 "Raub" (robbery)
§ 143 "Schwerer Raub" (aggravated robbery)

3.4.3 Burglary

Burglary,UCR 05, is defined as follows:
"The unlawful entry of a structure to commit a felony or a theft" (FBI 2004, p. 28).
A structure is defined as a building, also including for example a garage or a cabin.
Burglary is divided into burglary with forcible entry, with unlawful entry, and no force and with attempted forcible entry (FBI 2004, p. 28).
The penal code or criminal code (Strafgesetzbuch) in Austria contains the following categories of relevant crimes for burglary:
Burglary is not a crime on its own in Austria, but is combined into burglarytheft (§ 129) in one category. Theft is a crime treated in Austria in §§ 127-131. The best solution was to take burglary as crime type and to leave out theft because that is a crime type on its own. So in the data set of Vienna the subcategories "Fahrraddiebstahl" (bicycle theft), "Handy-Diebstahl" (mobile phone theft), "KFZ-Einbruch" (car burglary), "KFZ-Entfremdung" (car theft), "Schi-Diebstahl" (ski theft), "Taschendiebstahl" (pickpocketing) will be deleted from the data set for a correct and

comparable analysis with the UCR burglary. Burglary in the US does not include cars, so therefore that subset was deleted. The subcategory "Rest" (rest) will also be deleted because the data should only contain categories which are definitely correct and fit to the comparative data.
The remaining subcategories of § 129 "Einbruchdiebstahl" (burglarytheft) are "Firmen- und Geschäfts-Einbruchdiebstahl" (burglary of companies or stores), "Wohnhaus-Einbruchdiebstahl" (burglary into houses), "Keller-Einbruchdiebstahl" (burglary into cellars), and "Wohnungs-Einbruchdiebstahl" (burglary into flats).
Burglary into cellars could be put on a level with storage facility, stable, etc., meaning a place where things are kept, which is included in the U.S. burglary data. Additionally, breaking into a cellar is also breaking into the house because it is a part of the house. That is why this subcategory was included.

Definition of burglarytheft, § 129 "Diebstahl durch Einbruch oder mit Waffen":
A theft where someone is breaking into a building, a vehicle, a residence, a locked room, etc., including breaking open a box or a lock, or carrying a weapon or another mean to overcome or prohibit the resistance of a person (Jusline 2014i).

In Austria exists the § 109 "Hausfriedensbruch" (domestic disturbance) in the criminal code, which is similar to the definition of burglary (UCR 05) of the US law. Therefore, these data were also included to increase the comparability of data of different countries.
The **definition for § 109 Hausfriedensbruch** (domestic disturbance) is the following:
"Wer den Eintritt in die Wohnstätte eines anderen mit Gewalt oder durch Drohung mit Gewalt erzwingt, ist mit Freiheitsstrafe bis zu einem Jahr zu bestrafen." (The entry into the home of another with force or with threat of force) (Jusline 2014h).

3.4.4 Assault

Assault is covered in the U.S. as part 1 crime aggravated assault (UCR 04) and also as part 2 crime other assaults (UCR 08).
Definition for aggravated assault:
"An unlawful attack by one person upon another for the purpose of inflicting severe or aggravated bodily injury." (FBI 2004, p. 23)
Aggravated assaults can include weapons like firearms, knifes, and others or attacks with hands, fists, feet, etc. (Uniform Crime Reporting Handbook, p. 23).
Aggravated would include assaults or attempts to kill or murder somebody (FBI 2004, p. 24).
Definition for other assaults:

Simple, not aggravated assaults are summed up under the title "other assaults". An assault is "an unlawful attack by one person upon another" (FBI 2004, p. 23).

Opposite to this, assault in Austria is comprised in § 83 "Körperverletzung" (assault) and § 84 "Schwere Körperverletzung" (aggravated assault) as well as in the sub-categories in §§ 85-87. In the data sets for Vienna, assault is summed up and cannot be split, therefore both aggravated assault and other assaults, is chosen and combined from the Houston data set.

Definitions: § 83 Körperverletzung (assault):
Assault: When somebody hurts a person at the body or damages his/her health (Jusline 2014a).

§84 Schwere Körperverletzung (aggravated assault):
Aggravated assault: If the attack leads to a more than 24 days lasting health injury or the injury or health damage is heavy (Jusline 2014b).

§ 85 Körperverletzung mit schweren Dauerfolgen:
An assault that leads to a damage or injury for a long time or forever for example the loss of the faculty of speech, loss of visual faculty, etc. (Jusline 2014c).

§ 86 Körperverletzung mit tödlichem Ausgang:
An assault that leads to death (Jusline 2014d).

§ 87 Absichtliche schwere Körperverletzung:
Aggravated assault with intention (Jusline 2014e).

§ 88 Fahrlässige Körperverletzung:
Bodily injury caused by negligence (Jusline 2014f).

§ 89 Gefährdung der körperlichen Sicherheit:
Provoke danger for the life, the health, or the physical safety of another person (Jusline 2014g).

3.4.5 Robbery

Robbery is a part 1 crime in the U.S. (UCR 03) and is defined as follows:
"A person commits an offense if, in the course of committing theft and with the intent to obtain or maintain control of the property, he:

1. intentionally, knowingly, or recklessly causes bodily injury to another;
2. or knowingly threatens or places another in fear of imminent bodily injury or death" (Houston Police Department 2014).

Aggravated robbery would mean, that

"A person commits an offense if he/she commits robbery and he/she:

1. causes serious bodily injury to another;
2. uses or exhibits a deadly weapon; or
3. causes bodily injury to another person or threatens or places another person in fear of imminent bodily injury or death, if the other person is:
    a. 65 years of age or older; or
    b. a disabled person"

(Houston Police Department 2014).

Robbery includes a theft of something but also implies force or threat of force (FBI 2004, p. 21).
In Austria, robbery is mentioned under § 142 "Raub" (robbery) and § 143 "Schwerer Raub" (aggravated robbery) in the criminal code and the **definitions** are:

§ 142 Raub (robbery):
Who takes away with force or threat of force against body or life a foreign belonging of somebody for enriching himself or a third person (Jusline 2014j).

§ 143 Schwerer Raub (aggravated robbery):
Who commits robbery as part of a criminal coalition or with usage of a weapon attempts an aggravated robbery (Jusline 2014k).
In the data sets for Vienna robbery and aggravated robbery is combined. In the United States everything is included in the part 1 category robbery (UCR 03).

3.5 Implementation

This chapter deals with the realisation of the first problem definition. The software used is described, as well as the data preparation which was necessary for the execution of the analyses and also for visualizing the data in ArcGIS. An overview of the data sets is provided and the implementation of the selected cluster methods (spatial and spatio-temporal) is described. Finally, it is also stated why these particular methods were chosen.

3.5.1 Software used

Software tools for identifying hot spots are CrimeStat, GeoDa, the near repeat calculator, ArcGIS, and the Spatial Analyst module of ArcGIS (Eck et al. 2005). For this project, the following **software** packages will be used. First, ArcGIS and its Spatial Analyst module. In addition, CrimeStat and the near repeat calculator will be used. CrimeStat and near repeat calculator are free of charge and open-source software. Near repeat calculator will be used for the near repeat analysis. Prediction results will be shown in tables and graphics, but also will be visualized in the form of maps using ArcGIS.

CrimeStat IV

CrimeStat is an open source software program for spatial statistics and was developed by Dr. Ned Levine et al. It allows analyzing and visualizing crime places and is used by police departments and criminal justice and researchers. The latest version is CrimeStat 4.0.

Input files can be dbf, point shape, or Ascii files. They represent the incident locations. Spherical or projected coordinates are allowed. Statistical methods that the program can calculate are spatial distribution statistics (such as mean center, center of minimum distance, etc.), spatial autocorrelation statistics (Moran´s I, Geary´s C etc.), and distance analysis statistics (nearest neighbor analysis and Ripley´s K statistic).

The hot spot analysis section covers mode, fuzzy mode, hierarchical nearest neighbor clustering, hot spot analysis of zones, etc.

Also spatial modeling methods can be used with CrimeStat, like interpolation techniques (kernel density estimation), space-time analysis (Knox and Mantel indices, etc.), journey to crime analysis, etc. There are also methods for regression modeling, discrete choice modeling, time series forecasting, crime travel demand modeling included. Results of the analysis are shape-files, which can be visualized in GIS programs like ArcMap (Levine 2013e).

ArcGIS 10.2

ArcGIS is a commercial geographic information system software program from ESRI and the latest version which was also used for this research is 10.2. Different analyses as well as the maps were made with the spatial analyst extension in ArcGIS.

ArcGIS has hundreds of different analytical tools for solving a wide variety of problems. One example is the use of spatial statistics for analyzing the distribution of points (ArcGIS Resources).

Near-Repeat Calculator

The Near-Repeat Calculator software was developed by Jerry H. Ratcliffe of the Department of Criminal Justice at Temple University, Philadelphia, PA (Ratcliffe 2008, p. 4).

The software searches for unusual patterns in the connection of space and time between all data points. A spatial bandwidth and number of spatial bands have to be chosen (Ratcliffe 2008, pp. 6).

Temporal patterns are investigated with a temporal bandwidth and temporal bands. Temporal patterns are the number of days between crimes, for example, when a burglary happened at a neighboring house only a few days later than the burglary at the first house close to it (Ratcliffe 2008, pp. 7).

Monte Carlo iterations are used to produce a random pattern of spatial points. The date values are randomly placed in this case. This pattern shows no near repeat process and it is compared to the data set which someone wants to observe. When the difference between the patterns is large and the statistical significance of 0.05 or lower is reached, a near repeat phenomenon was detected. P= 0.01 would be an acceptable result (with a chance of an error of one in a hundred) and P = 0.001 would be the best result (with a chance of an error of one in a thousand) but it would take very long to process 1,000 random iterations for the random pattern (Ratcliffe 2008, p. 8).

Distance settings can be Manhattan or Euclidean. Manhattan would be the going from one point to the other point first in the horizontal direction and then in the vertical direction. Euclidean would be the direct distance or as the crow flies (most direct way from one point to the other). The Manhattan distance is more correct for (U.S.) street networks, because that represents more the reality than the Euclidean distance (Ratcliffe 2008, pp. 8).

The output consists of two files, a summary htm file and a comma-separated-values output file (Ratcliffe 2008, p. 9).

Results are significant when events are close to each other in both time and space. That is the case when there is a 95 percent or higher chance that the pattern does not occur randomly (Ratcliffe 2008, pp. 9).

The near repeat calculator offers also the function of determining the first, original event and the near repeat event (the second event that follows after an original event). The user can choose the spatial and temporal frame. The output is a comma-separated values file (csv). You can see how many times an event was the originator and how many times it was the following event of a near repeat pair. It is possible to identify clusters or hotspots and also high-count originator events (Ratcliffe 2008, pp. 12).

3.5.2 Data preparation

Data were provided in different formats and had to be prepared for the usage with a geographic information system like ArcGIS and for the use with software programs for calculating the cluster analyses, which are CrimeStat, the near repeat calculator and SaTScan.

3.5.2.1 Geocoding of Houston data

From the crime types the author of this thesis chose 03- robbery, 04-aggravated assault, 05-burglary and 08-other assaults. Aggravated assault and other assaults were grouped together as “assault” to make it comparable to the crime type “Körperverletzung” in Austria. For each crime type a separate file for geocoding was made with the aid of the Census 2010 TIGER/Line data.

Data for Harris County were downloaded for the year 2010 from the website http://www.census.gov/cgi-bin/geo/shapefiles2010/main. The edges-shapefile (lines of street net in Houston) was clipped to the boundary of the city of Houston (2010_Tracts-shapefile with polygons). The data are from the U.S. Bureau of the Census and the coordinate system is NAD 1983. The data include for example the State Fips Code, the County Fips Code and the Census Tract Code.

The crime data for Houston had to be edited so they could be used for statistical analysis and with GIS. Originally, the crime data had no coordinates, so they had to be generated by geocoding. The exact address existed in the collected data for Houston. For the geocoding process the address had to be put into a single column instead of several columns (for street name, house number, etc.). The data had also to be associated with the correct month, because a few data sets were recorded one or several months later than they actually happened. The data were sorted by the different crime types for 2011 and 2012.

The offense code column shows the type of crime with the UCR code. The author received a list with the UCR codes, but this list only included the first two numbers. This allows you to tell which crime type it is but not any further classification. For the analysis it will be enough to know the crime type.

Data were combined per year and separated by crime type. Additionally to the offense code there were two columns added (see table 1) which included the abbreviated offense code (i.e., the first two numbers of the UCR-Code) and a column “crime type”, which shows the crime type as a text (for example, robbery).

Offense Code	Offense Code Short	Crime Type
04014	04	aggravated assault

Table 1: Added columns for geocoding of Houston data

For geocoding, a column named “CountyFP” was added. This stands for County Fips Code (see table 2). For the city of Houston, which belongs to Harris County, the Fips Code is 201.

Address	CountyFP
01725 DES JARDINES ST	201
01445 LAKESIDE ESTATES DR	201
05000 PERSHING	201
08112 HOFFMAN	201
12401 PALMFREE	201
02000 STUDEWOOD	201

Table 2: The county Fips Code for geocoding

FIPS is short for Federal Information Processing Standard and these are codes for definitely identifying geographic entities, for example states, counties, etc. (U. S. Census Bureau, Geography Division 2014)
The author of this thesis tried to improve the result of the geocoding by adding different columns like state zip, state name, etc. but it had no effect on the final result.
The final excel-files were imported into ArcMap and with the tool "Table to Table" the excel-file was transformed into a dbf-file. The tool „Create Address Locator" was used and the option "US Address – Dual Ranges" for adding the data was selected. Dual Ranges means that also the side of the street where a house stands is important and this will be considered. For the reference data the clipped edges-shapefile from the TIGER/line Census data 2010 was chosen. For "Left City or Place" and "Right City or Place" Countyfp was chosen as the referring column.
By right-clicking on the table the option "Geocode Addresses" can be selected. A USLocator for Houston is now created with the chosen crime type data as the address table. For Street or Intersections the field "Address" has to be chosen and for City or Placename "CountyFP".
By clicking on Geocoding Options additional options can be chosen. The minimum match score was set to 60, the remaining options were left as defaults. Different option settings influence how many addresses can be matched.

As a result the percentage of matched data sets can be seen at the end of the geocoding process. This is the percentage of the original data that can be used and will be applied for in the subsequent analyses. In the article "The Impact of Hurricanes on Crime: A Spatio-Temporal Analysis in the City of Houston, Texas" a match score of 60 was also applied for a crime data set for Houston (Helbich and Leitner, 2011). Therefore, a match score of 60 was also chosen in this research. Overall, a match rate of about 90% could be reached for all different crime types.
It was also searched for a shapefile with the city boundary of Houston for the analyses (local Moran´s I and Gi*). The shapefile for Houston is included in the Census 2010 Data Sets. It contains a shapefile for the census tracts for 2010, where only the census tracts with the right county number (201 for Harris County) were selected. The data are from the website http://mycity.houstontx.gov/home/cohgis.html.

The final geocoding results for Houston are the following:
For robbery for the year 2011 a match rate of 93% could be reached. For burglary in 2011 94% matched addresses were reached and the geocoding of assaults (aggravated assault and other assaults combined) for 2011 of Houston also yielded a result of 94% of matched addresses. No data correction was necessary because no points were lying outside of Houston. After the geocoding the data sets with a score under 60 were removed. All of the removed crime data had a score of 0 that means there was no matching found at all for these addresses.
The geocoding was also done for Houston for 2012. For robbery and burglary 2012 a match rate of 94 % could be reached. For assault 90 % of matched addresses were reached.

3.5.2.2 Vienna data preparation

For Vienna the following crime type data sets were collected and prepared: "Körperverletzung" (assault), "Raub" (robbery), "Einbruchsdiebstahl" (burglary-theft), and "Hausfriedensbruch" (domestic disturbance). Of the § 129 "Einbruchsdiebstahl" (domestic disturbance) only four categories were chosen, in order to fit to the data of Houston. Those four categories include "Wohnungs- Einbruchsdiebstahl" (burglary into flats), "Keller-Einbruchsdiebstahl" (burglary into cellars), "Firmen-und Geschäfts-Einbruchdiebstahl" (burglary of companies or stores) and "Wohnhaus-Einbruchsdiebstahl" (burglary into dwelling houses). Then "Einbruchsdiebstahl" (burglary-theft) and "Hausfriedensbruch" (domestic disturbance) were grouped together. Also a column named „crimetype" was included with the name of the crime type for easier identification to the data sets. Data for Vienna were separated for the year 2011 and 2012.

For the integration into ArcGIS a new File Geodatabase was created and the Excel-tables included with "Import Table (Single)". The coordinates are in MGI Austria Lambert and the workspace was changed to that coordinate system.
Also data correction was conducted. For example, points that were lying outside of Vienna were deleted to improve the data quality. For some reason, a few points were found that were located in Graz and other cities of Austria. With the "Display XY Data" Tool a new layer was added to ArcMap where the data sets of the table are displayed as points on the map. With the Editing-Tool the incorrect data sets were deleted and only data sets which are located in Vienna remained.
To use the data with the CrimeStat software they needed to be in dbf-format. The shapefiles include dbf files too, so a shapefile can be used for data input. In CrimeStat only x and y are the variables that are being used.
It was also searched for a shapefile with the boundary of Vienna for the analyses (used for local Moran´s I and Gi*). The shapefile for Vienna was

downloaded from the website http://publicdata.eu/dataset/land-wien-landesgrenzen/resource/04c37591-bbc5-46e8-9994-9d0b07920976.

3.5.2.3 Data Overview

The crime type files of the two cities include the following amounts of crime records (table 3 and 4):

Vienna 2011:	
Burglary:	14,986 records
Assault:	15,334 records
Robbery:	2,736 records
Houston 2011:	
Burglary:	26,732 records
Assault:	48,539 records
Robbery:	8,128 records

Table 3: Total incidents for 2011 (data used)

Vienna 2012:	
Burglary:	19,726 records
Assault:	15,787 records
Robbery:	2,614 records
Houston 2012:	
Burglary:	25,593 records
Assault:	44,320 records
Robbery:	9,043 records

Table 4: Total incidents for 2012 (data used)

3.5.3 Spatial cluster methods

Four spatial cluster methods were selected and analysed for 2011 and applied to every crime type file of Houston and Vienna. The four chosen spatial cluster methods are described in detail in the following sub-chapters.

3.5.3.1 Why were these methods chosen?

There exist many different techniques for cluster analysis as described in the theoretical background chapter. Some of the cluster analysis methods were chosen for this project.

When using different cluster analysis methods they reveal varying groups and patterns (Levine 2013b, p. 12).

The results can then be compared and evaluated to find out which method provides the best result and where methods agree or disagree in terms of the location of hot spots.

Quick and Law (2013) mention that up to now there is no single local cluster analysis method that is proven better than other methods. It is

important to use a method that describes the studied phenomenon best and helps find the best practical applications (Quick and Law 2013, p. 3).
The kernel density estimation interpolates crime resulting in a density map, with a risk surface being developed. It is among the most used type of map done by crime analysts (Smith and Bruce 2008, p. 17).
The risk surface means that there does not need to be a crime on every point of the study area, but the risk surface defines the chance that a crime could happen at a particular point (Smith and Bruce 2008, p. 67).
The kernel density estimation does not use a sample of the data, but the entire data set (Smith and Bruce 2008, p. 67).

As described in the theoretical background chapter non-hierarchical clustering methods would need to define the quantity of clusters in advance. Hierarchical clustering methods do not need to define the number of clusters a-priory and this is the main reason that the nearest neighbor hierarchical clustering is chosen.
The nearest neighbor hierarchical clustering technique uses Euclidean distances to measure distances between neighboring points (Grubesic 2006, p. 4). In general, the Euclidean distance is more often used than any other distance measuring method.
The local Moran´s I was often used in crime research concerning violence and property crime, similar to this project (Quick and Law 2013, p. 6). So this method was also chosen. The Gi* statistic is another local spatial autocorrelation method (Hinman et al. 2006, p. 2). It was used with other spatial relationship methods to compare the results with and to find differences and find out which method is better suited.

3.5.3.2 Kernel density estimation

The kernel density estimation is calculated in CrimeStat. The following parameters were chosen. For the method of interpolation normal was chosen; the choice of bandwidth is fixed with 250 meters; the area units are points per square kilometres and the output units are absolute densities. Several different settings were tried (for example adaptive bandwidth) the chosen parameters delivered the best results.
Absolute densities are chosen because it is good to use for comparisons between different crime types (Levine 2013d, p. 22).
A bandwidth of 250 meters shows enough detailed results.
For the calculation of a kernel density estimation a reference file has to be created first.
With the reference file a grid file is placed over the study area for analysis with grid-based systems (Heraux 2007, p. 3).
A rectangle was drawn in ArcMap over the study area that includes the whole area and the coordinates of the corner points were used for the reference file in CrimeStat.
A reference file with 250 columns represents a good value for the cell size (Bruce and Smith 2011, p. 27).

The output file of the calculation is a shapefile, which can be shown in ArcMap. The z-scores, which represent the density values, with a value of 0 were deleted and the result then fit to the boundary of Vienna. All results are shown in chapter 4.

3.5.3.3 Nearest neighbor hierarchical clustering

In CrimeStat a special algorithm is used for the NNHC method. A threshold distance is one cluster criterion and points that are closer than this distance are clustered together. It is possible to define a minimum number of points for a cluster. This is the second cluster criterion. These two cluster criteria influence how many clusters will be found for the first-and higher order clusters. (Levine 2013b, p. 21)

One option is to select a random distance or a fixed distance as criterion. Random distances are based on a random distribution, while a fixed distance is more subjective, because it is chosen by the analyst (Smith and Bruce 2008, p. 58).

Random distances were tried for every data set, but the Houston data sets assault and burglary were too large for the calculation in CrimeStat (an error message "out of memory" occurred in the middle of the calculation). This is an error on the software side, which is unable to handle such large data sets. Alternatively, various fixed distances were tried and 0.5 kilometres was chosen as the distance for all Houston data sets. According to Levine (2013b, p. 27) the fixed threshold distance should be under 0.5 miles that are 0.8 km.

The threshold distance was set to 0.05 (or 5%) of the distances.

For the minimum number of points 1 to 2 percent of the total number of data sets would be a good value. This was tried but good results with a few clusters were found by reducing the minimum number of points to the percentages that can be seen in table 5.

Ellipses or convex hulls can be chosen as output units. Convex hulls are more accurate that is why the convex hull was selected (Smith and Bruce 2008, p. 58).

The finally chosen parameters can be seen in table 5.

Nearest Neighbor Hierarchical Clustering:					
	total data	NNHC-minimum	percentage	distance:	output unit:
Houston 2011					
Assault:	48,539	243	0,50%	fixed: 0.5 Kilo	km
Robbery:	8,128	61	0,75%	fixed: 0.5 Kilo	km
Burglary:	26,732	134	0,50%	fixed: 0.5 Kilo	km
Wien 2011					
Körperverletzung:	15,334	100	0,65%	random	km
Raub:	2,736	27	1,00%	random	km
Einbruchsdiebstahl:	14,986	25	0,17%	random	km

Table 5: Parameters selected for NNHC method

3.5.3.4 Local Moran´s I

The local Moran´s I is calculated in ArcGIS. It identifies clusters and also spatial outliers. The output is the local Moran´s I value as well as a Z score and a p-value, which both can inform about the statistical significance (ArcGIS 9.3 Tool Help 2009a).

At first a raster for the project area has to be compiled. This was done with the "Create Fishnet" tool. For both cities a raster with a cell width and a cell height of 500 meters was calculated. This seemed to deliver the best result.

Then the point data layer was joined to the created fishnet with the "Spatial Join" tool. Parameters were Join_One_To_One and the Match option was "Contains". The last parameter means that all features in the join feature class (the point data layer) are selected, if a target feature includes them (the raster cell). If a point or more points lie in a raster cell, they will be joined to this cell (polygon) (ArcGIS 9.3 Tool Help 2009b). Keep all target features was selected, so that no points get lost and all entry features are kept. Polygon was chosen as the geometry type. A new "Join_Count" field was created which informs how many join features have been joined with the target features.

Under "Statistics" the "Join_Count" field was checked if all points have been correctly aggregated to the raster cells. Then the rectangle raster grid was fit to the boundary shapefile of the particular city with the "Clip" tool. Again the number of points was checked under "Statistics".

The creation of the raster was necessary for calculating the local Moran´s I. The tool in ArcGIS is called "Cluster and Outlier Analysis (Anselin local Morans I). The raster layer is the input feature class and the input field parameter is the "Join_Count" field that includes the points in every raster cell. For Conceptualization of Spatial Relationships the default setting is chosen, which is "inverse distance". This means that closer features have a larger influence than features further away (ArcGIS 9.3 Tool Help 2009a). Distance is again calculated with euclidean distance.

With no stated "distance band or threshold distance" value the default neighborhood search threshold was automatically chosen. That is, the minimum distance that every feature has is one neighbor (ArcGIS 9.3 Tool Help 2009c). For Vienna this default threshold distance was 500 meters and for Houston it was 592 meters.

There is no value recommended but one should set a threshold distance and should use the default value as a starting point for analyses (ArcGIS 9.3 Tool Help 2009c).

Two other distances were tried, 1000 m and 800 m. For NNHC a distance of 500 meters was used and for kernel density estimation a distance of 250 meters. For this reason, the author decided to take a distance that is

closer to the distances selected for the other two cluster methods. The results with an 800 m distance can be seen in chapter 4 (Results).
In CrimeStat IV a distance value of up to a mile (1.6 km) is chosen by default from the software for calculating the local Moran´s I and this is a typical distance for crime analysis (Levine 2013c, p.9).
As explained in the next subchapter Gi*, the best result was achieved with a distance of 800 meters.

3.5.3.5 Getis-Ord statistic Gi*

Gi* is calculated in ArcGIS. The output is a z score, which specifies the statistical significance of clustering for each feature (ArcGIS Desktop Help 2009).

For the Gi* calculation the same raster as was already used for the local Moran´s I calculation is used as input feature class and input field. The default settings were kept, which is for the conceptualization of spatial relationships "fixed distance band".
This method is most suitable for point data. Choosing a value for the fixed distance is recommended, especially for large data sets. The distance should include at least one neighbor for each feature, with an average of about eight neighbors for each feature and no feature should contain all other features (ArcGIS Desktop Help 2009).
The default distance ensures that every feature has at least one neighbor. So the distance to select should be higher than the default distance. When there are spatial outliers in the data set this could lead to features with many thousand neighbors. This should be avoided. A Spatial Weights Matrix can be used to define the number of neighbors and also the distance to which points are searched for. Spatial outliers would have a wider distance in that case in order to have the defined number of neighbors. (ArcGIS Tool Help 2009c).

A spatial weights matrix was calculated with a distance of 800 (and also 1,000) meters and a minimum number of eight neighbors were chosen. The result was compared with the result for the fixed distance band with 800 (and also 1,000) meters. No differences were found, so there do not seem to be any spatial outliers that influence the results. Therefore, the fixed distance band method without a spatial weight matrix was used.

The default distance for Houston is 593 meters and for Vienna 500 meters. For kernel density estimation a bandwidth of 250 meters was chosen and for the nearest neighbor hierarchical clustering a distance of 500 meters was selected for Houston and a random distance for Vienna. Again, the author chose to take a similar distance for the Gi* statistics. The results for an 800 meters distance showed the same clusters as the results for 1,000 or 1,300 meters distances. With higher distances, more raster cells are included and the clusters get larger. So the author selected 800 meters distance for this study.

3.5.4 Spatio-Temporal cluster method – near repeat calculator

The main settings in the software are the spatial bandwidth and the number of bands, the temporal bandwidth and the number of bands, and the significance level. Distance settings can also be chosen, which includes the Manhattan or the Euclidean distance (Ratcliffe 2008, p.5).

Only csv (comma separated value) – files as data input can be used by the near-repeat calculator. Data should be organized into rows with 3 columns that includes the x-coordinate, the y-coordinate, and a date. The date should stand for the date when the crime happened. The coordinates can be in any projection and in meters or feet. Latitude and longitude geographic coordinates cannot be processed by the near repeat calculator (Ratcliffe 2008, pp. 5).

The spatial bandwidth and the number of spatial bands have to be chosen. Different settings should be tested to find the right parameter setting. Near repeat patterns occur mostly only for a few blocks or a few hundred meters. So the effect is limited to a small area and the bandwidth has to be chosen accordingly. The manual recommends ten spatial bands as a starting point (Ratcliffe 2008, pp. 6).

A temporal bandwidth of a week (7 days) or a month (28 or 30 days) is recommended. Common settings for temporal bands would be 12 temporal bands for a 30 day temporal bandwidth to cover a year, or 13 temporal bands for 14 day temporal bandwidth to cover half of a year (Ratcliffe 2008, pp. 7).

The near repeat calculator software only needs three types of information in the following form, including the "x-coordinate, the y-coordinate, and he date in the form TT/MM/JJ". So it was necessary for the author to adapt the data to this particular supported format to use the near repeat calculator. Coordinates can be in any projection but they have to be without decimal places. The data format is also limited to csv-files. Getting the right date format for the data of Vienna was no problem but the data for Houston was in the format "YYMMDD" and had to be cut into pieces and brought into the right form. This was done in Excel with different formulas (Chain, Left, Right). In the end the data were opened with the program Notepad++ to check the output. In addition, the separator had to be changed from ";" to ",".

The near repeat phenomenon for burglary was calculated for the different seasons for 2011. So the data had to be split up for this analysis and was divided into the standard dates of the seasons. An overview can be seen in table 6:

winter	**21.12.2010 – 20.3.2011**
spring	**21.3.2011 – 20.6.2011**
summer	**21.6.2011 – 22.9.2011**
autumn	**23.9.2011 – 20.12.2011**

Table 6: Selected dates for the near repeat analysis of the four seasons for 2011

At the beginning, various settings were used, for example, a 30 day- time period for different seasons combined with 1 block length (125m) or 2 block lengths (250m), but no appropriate results were returned. In addition, many settings found nearly the same results (1 repeat victimization pattern which means that there were crimes happening at the same location within a time slot of 14 days after the original crime incident), no matter the spatial parameters (125m or 250m) and the temporal parameters (7 or 14 days). The block length is the length from one street intersection to the next. An average block length for Houston and Vienna was measured in ArcGIS and determined to be 400 feet. The same block length is also used in the manual of the near repeat calculator. In terms of the temporal bandwidth, the percentage value for the 14 days result – time period was not as high as for the 7 day- time period. For this reason, it was decided to use a 400 feet spatial bandwidth and a 7 days temporal bandwidth as the input parameters in the near repeat calculator. This parameter selection is also similar to the study by Caplan et al. (2013, p. 10), which results in one repeat victimization pattern and for some seasons also in one near repeat victimization pattern.

The finally chosen parameters were $p<0.05$ (as recommended in the manual), one block length (400 feet – in this case 125 meters, as recommended in the manual), a time period of 7 days bandwidth with 24 bands (based on the article by Caplan et al. (2013)), resulting in a total time period of 168 days or about half of a year), and Euclidean distance (more appropriate for Vienna).

For the calculation of the near repeat originator and the repeat counter the same settings as for the near repeat calculations were chosen, namely one block length (125m) and 1 to 7 days for temporal comparison.

Because the near repeat calculator does not visualize the results in form of maps, this step was done by the author with the results being visualized in ArcGIS. The results from the near repeat calculator software are htm-files with a description and tables of the findings in it. The results for the originator and the repeat counter are csv-files, which were adapted in Microsoft Excel to have the data in the right format for importing them into ArcGIS. The data were split up into a dbf-file for all results, one with the originators, and one with the near repeats. Then, the dbf-tables were imported into the existing geodatabases with the function "Import/ Table (single)". With "Display xy data" the data can be displayed as a shapefile with points. That was done for all tables and for the originators a "buffer" was calculated with 125 meters, which was the chosen search radius (spatial bandwidth) in the near repeat calculator program for near repeats.

The originator points, the near repeat points, and the buffers are visualized in ArcGIS as the final result.

3.6 Summary

The chapter methodology described the problem definition and the method of solution in more detail. In addition, the project areas (Houston and Vienna) are explained together with the used geodata that included crime types from both cities, which matched according to their definitions in the law texts. Burglary, assault, and robbery are the three crime types being selected and analyzed. In the implementation subchapter the used software is described and how the data had to be prepared for the analyses. For example, the Houston data had to be geocoded. It is then explained why the specific methods were chosen and which parameters were chosen for each method. The selected spatial cluster methods are the kernel density estimation, nearest neighbor hierarchical clustering, local Moran's I, and Gi*. For the near repeat analysis the near repeat calculator was used and the chosen settings are also described.

4. Results and Interpretation

4.1 Results of spatial cluster methods

In this chapter the results of the four spatial cluster analysis methods are shown.

4.1.1 Kernel density estimation result

The classes of the results for the kernel density estimation can be seen in figure 2.

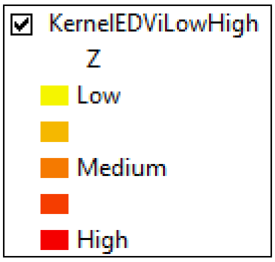


Figure 2: Selected classes for the kernel density estimation

Usually, the results of the kernel density estimation are labelled from "low" to "high", so the results can be better understood (Bruce and Smith 2011, p. 98).
This was also done in this research. It is easier to describe the result with a scale with low and high because it is hard to explain the meaning of the Z-score (Bruce and Smith 2011, p. 98).

The results of the KDE method for all three crime types can be seen in figures 3 – 5.

Burglary:

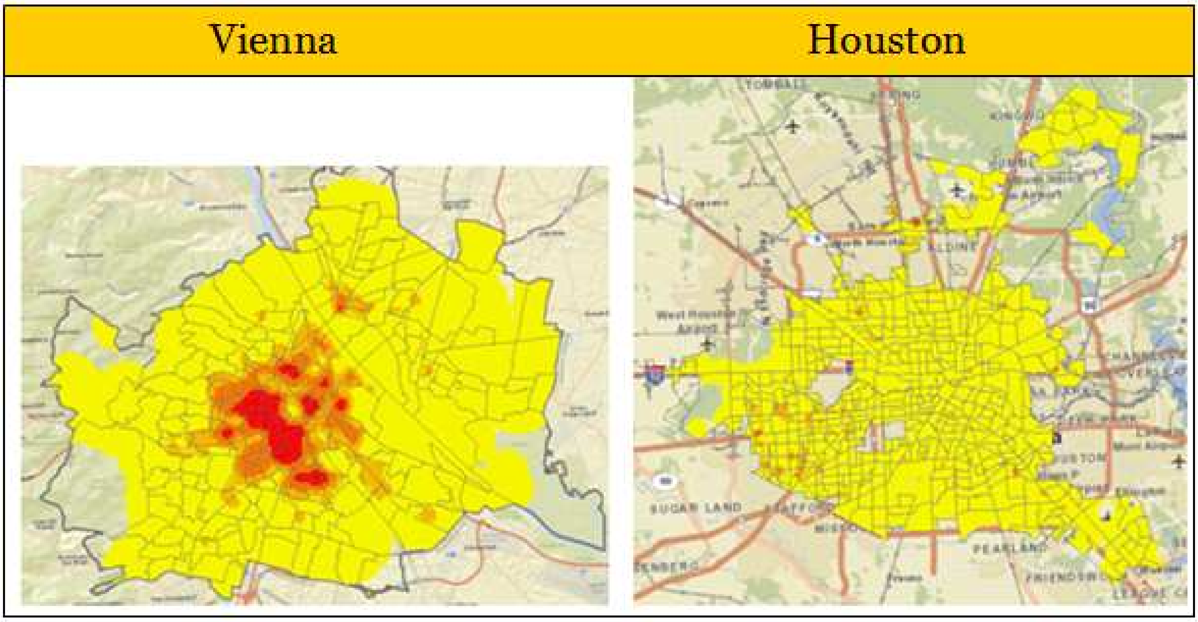


Figure 3: The results of the kernel density estimation for burglaries for Vienna and Houston

Assault:

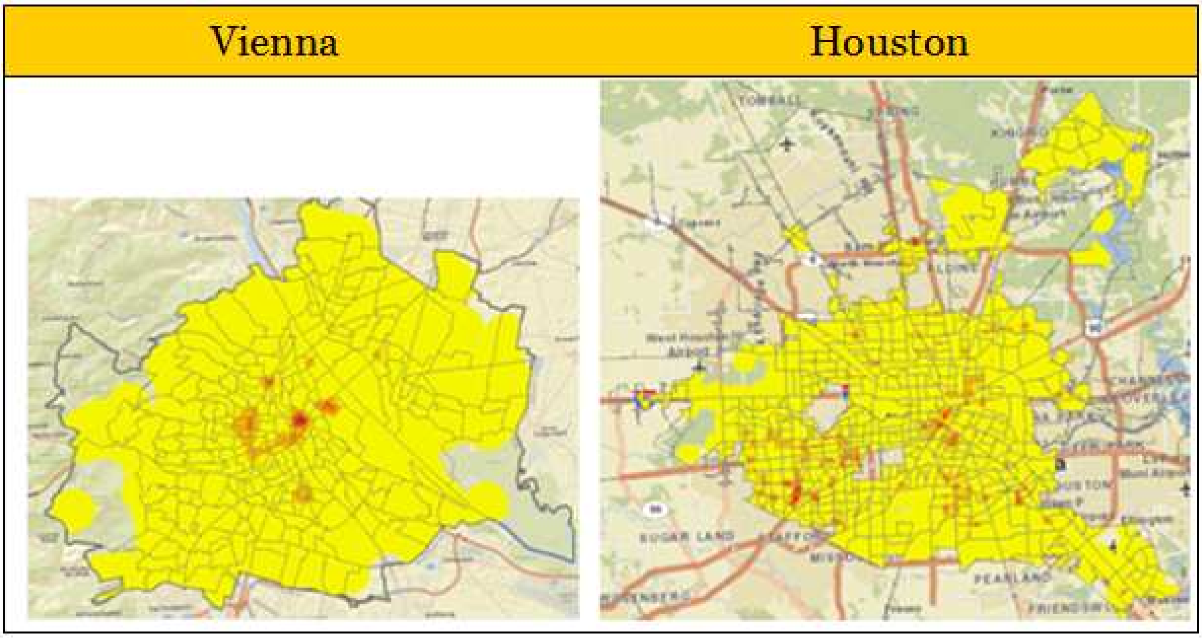


Figure 4: The results of the kernel density estimation for assaults for Vienna and Houston

Robbery:

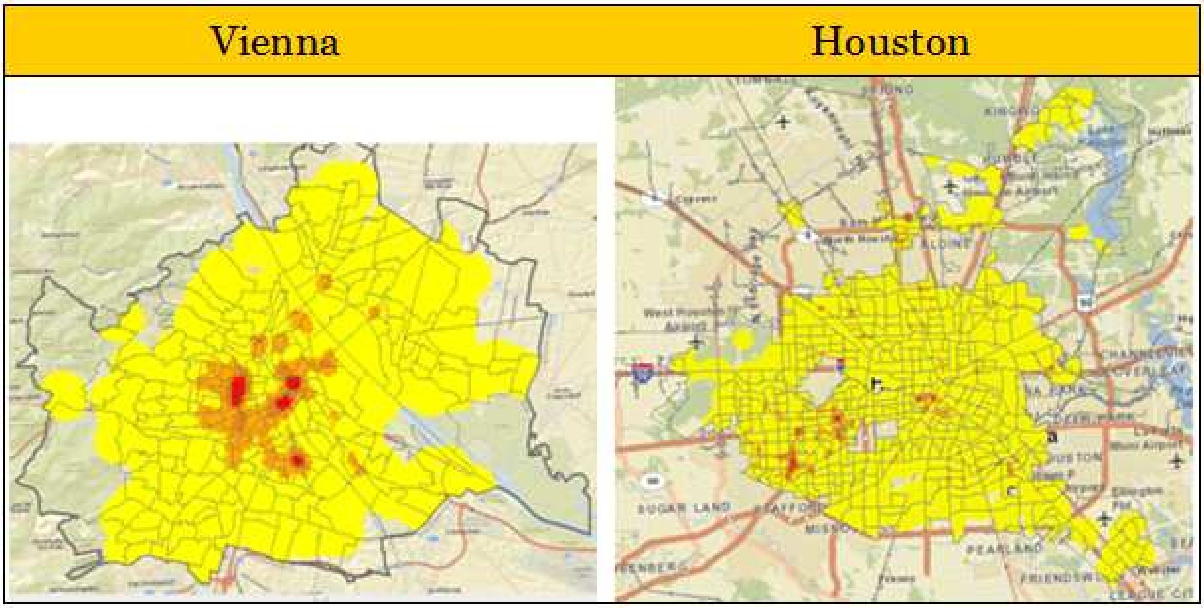


Figure 5: The results for the kernel density estimation for robbery for Vienna and Houston

In figures 3 – 5 the kernel density estimation results can be seen. The red areas in the map define hot spots, i.e., the places with the highest concentration of incidents. The Z-scores were divided by an equal interval classification with 5 categories, so the highest 20 % of the data are shown as hot spots.

4.1.2 Nearest neighbor hierarchical clustering result

The results of the NNHC method for all three crime types can be seen in figure 6 – 8.

Burglary:

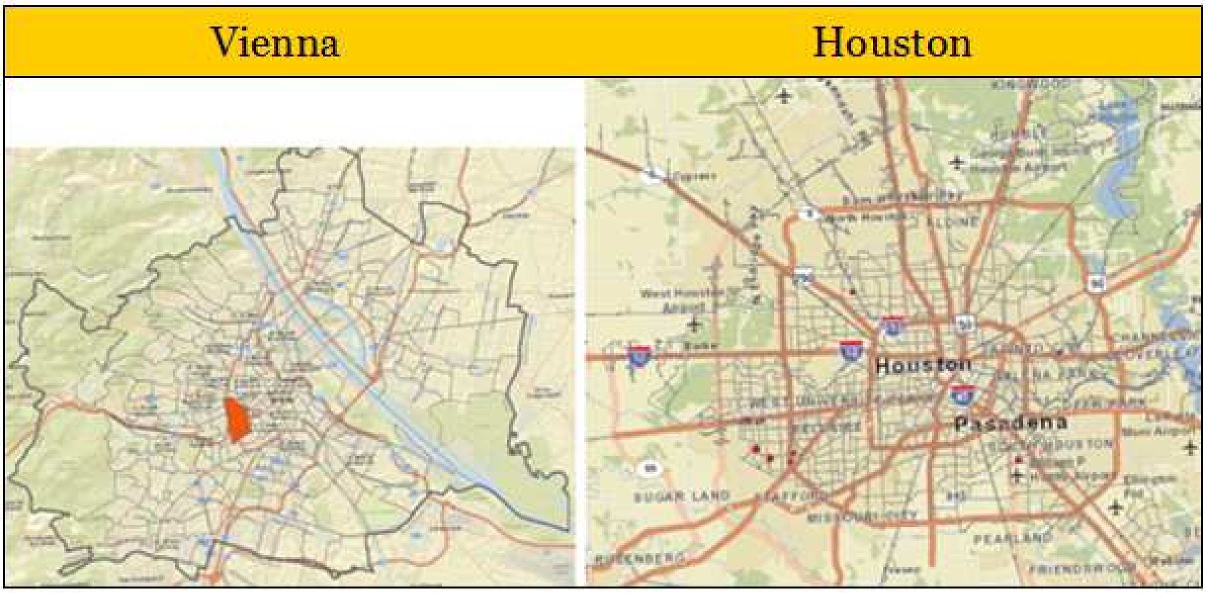


Figure 6: The results of the NNHC for burglaries for Vienna and Houston

Assault:

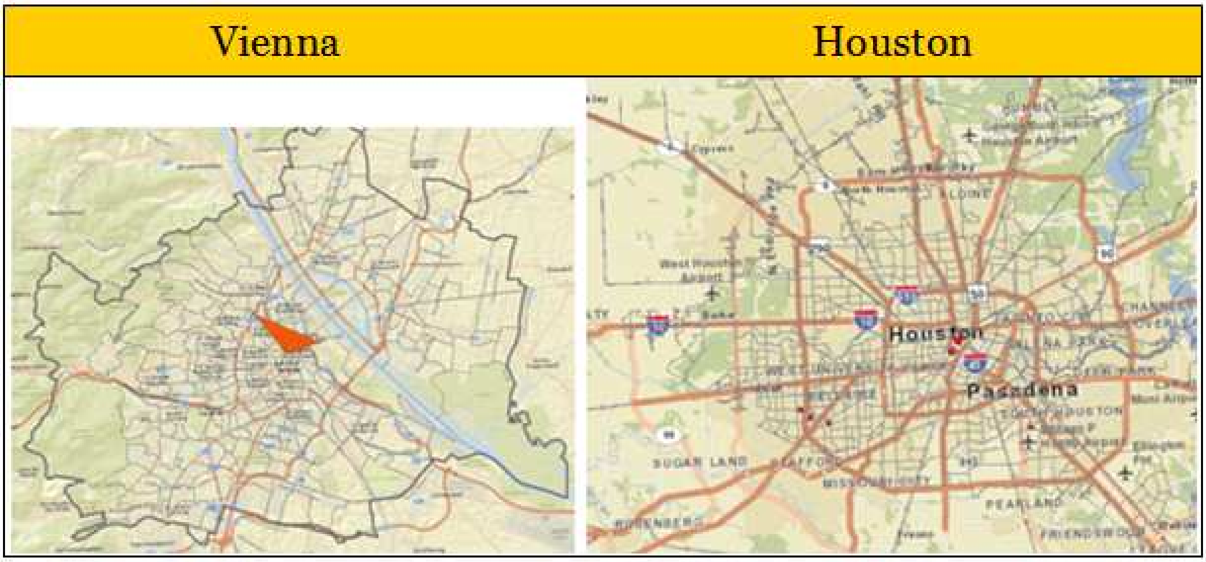


Figure 7: The results of the NNHC for assaults for Vienna and Houston

Robbery:

Figure 8: The results for NNHC for robberies for Vienna and Houston

The results for the nearest neighbor hierarchical clustering show the following number of clusters (see table 7).

	Cluster 1. order	Cluster 2. order
Houston 2011		
Assault:	7	-
Robbery:	8	-
Burglary:	9	-
Vienna 2011		
Assault:	5	1
Robbery:	6	1
Burglary:	15	1

Table 7: Number of clusters of the NNHC method found for Vienna and Houston

4.1.3 Local Moran’s I result

The categories described in the results of the local Moran's I can be seen in figure 9.

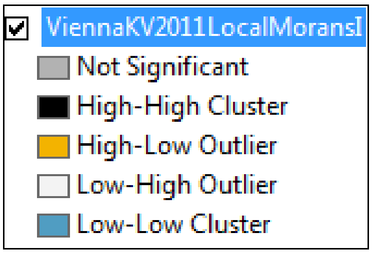


Figure 9: Selected classes for the local Moran’s I

The results of the LMI method for all three crime types can be seen in figures 10 – 12.

Burglary:

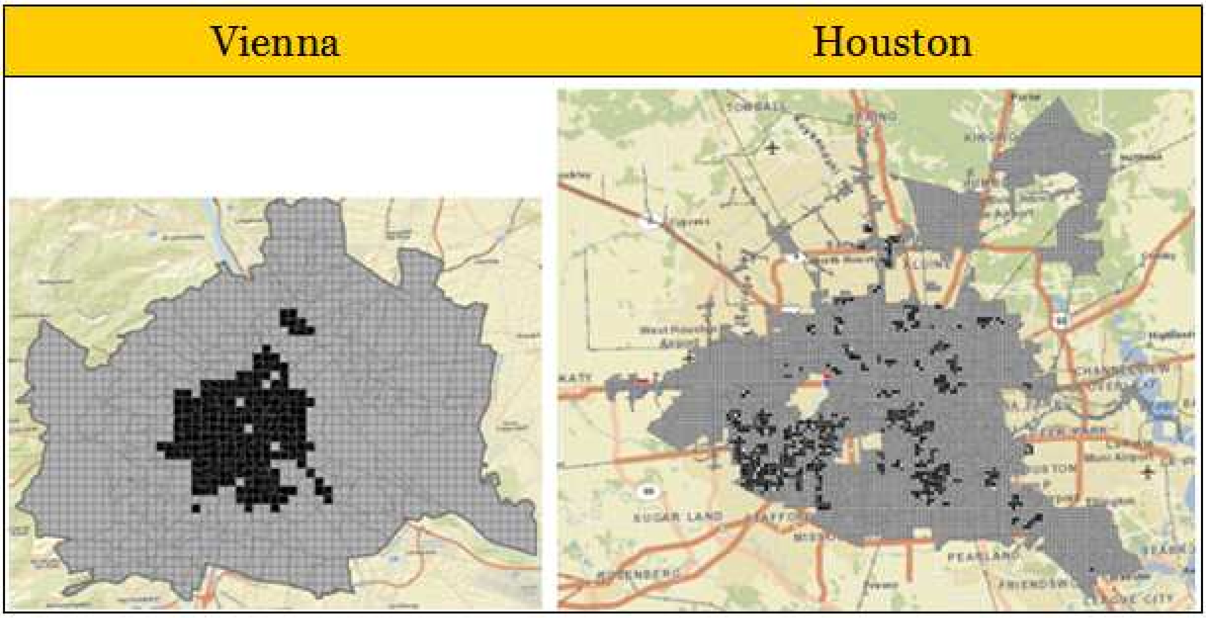


Figure 10: The results for LMI for burglaries for Vienna and Houston

Assault:

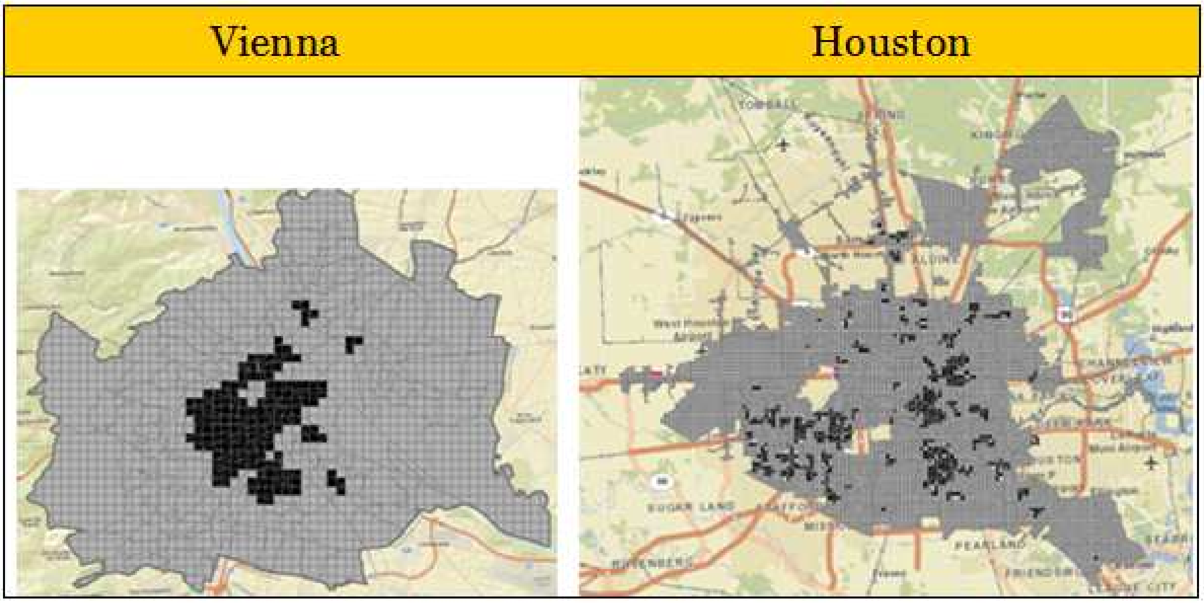


Figure 11: The results for LMI for assaults for Vienna and Houston

Robbery:

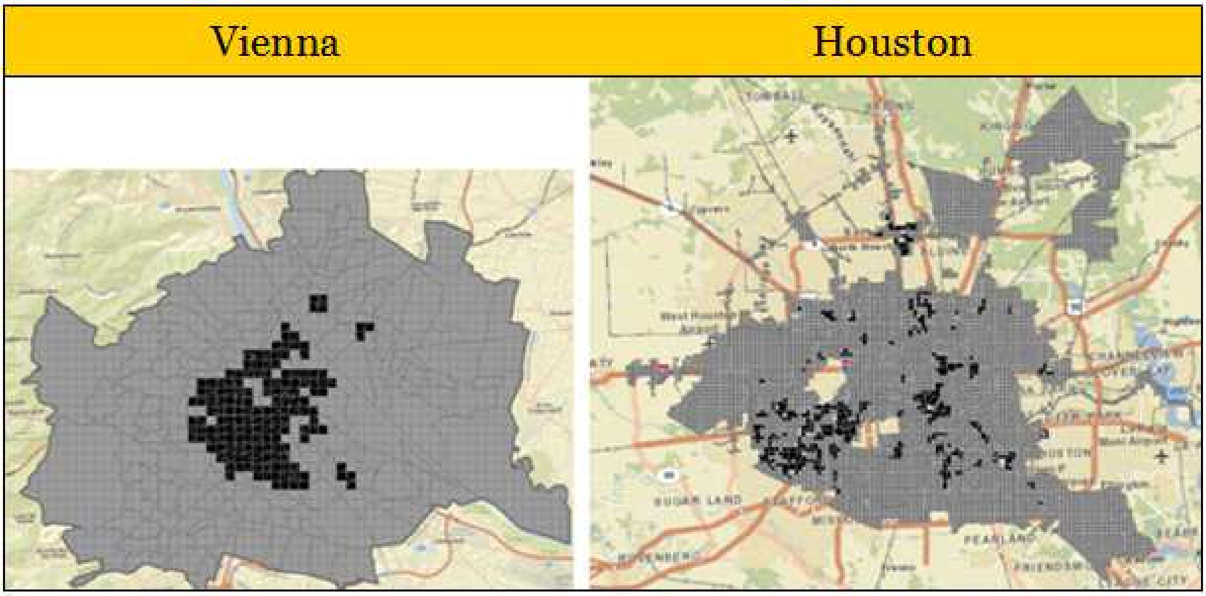


Figure 12: The results for LMI for robberies for Vienna and Houston

Only high-high (hot spot) clusters were found and values in the "not significant" category for Vienna. For Houston also a few high-low outliers (spatial outliers) were found. This means that a raster cell with a high incident value is surrounded by cells with low values.

4.1.4 Getis-Ord Gi* result

The categories described in the results for the Gi* can be seen in figure 13.

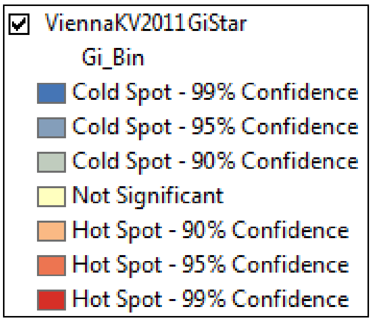


Figure 13: Selected classes for the Gi*

The results of the Gi* method for all three crime types can be seen in figures 14 – 16.

Burglary:

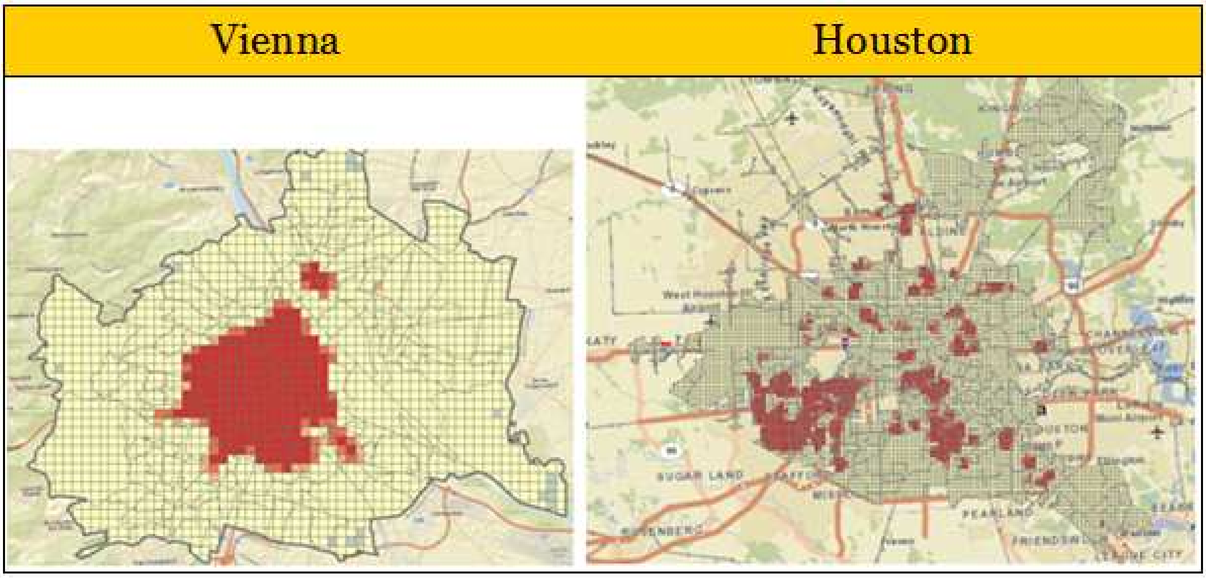


Figure 14: The results for Gi* for burglaries for Vienna and Houston

Assault:

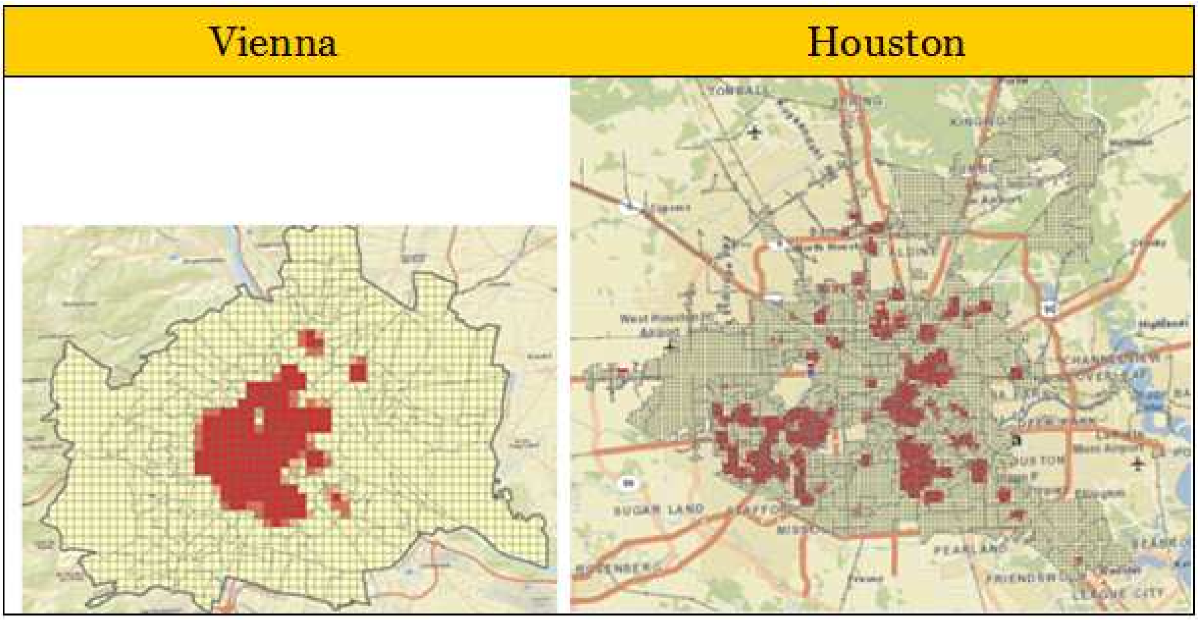


Figure 15: The results for Gi* for assaults for Vienna and Houston

Robbery:

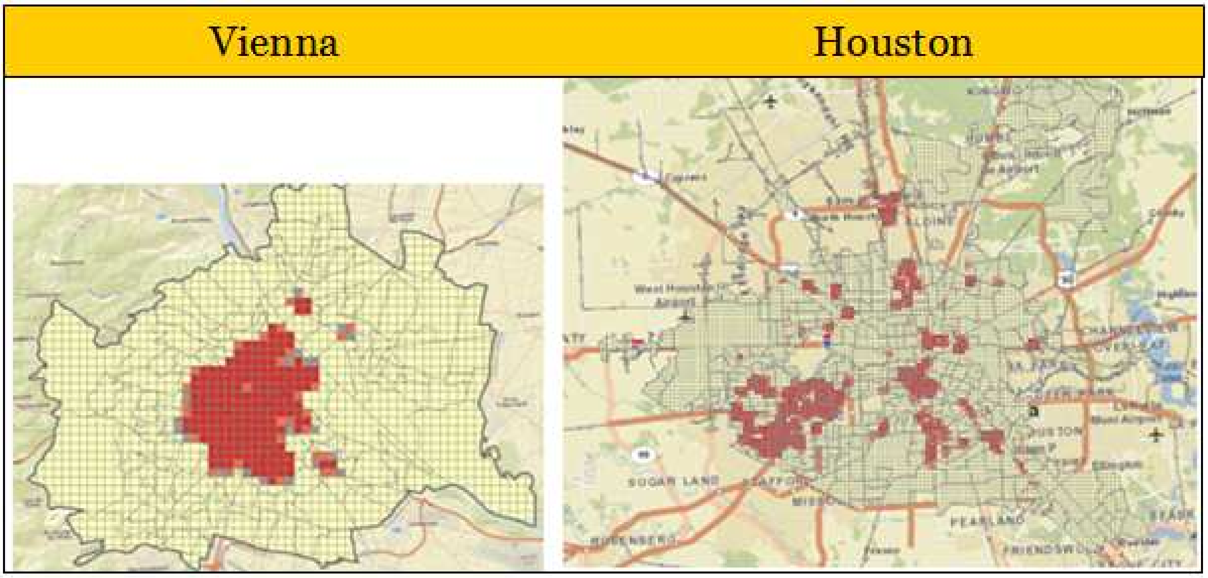
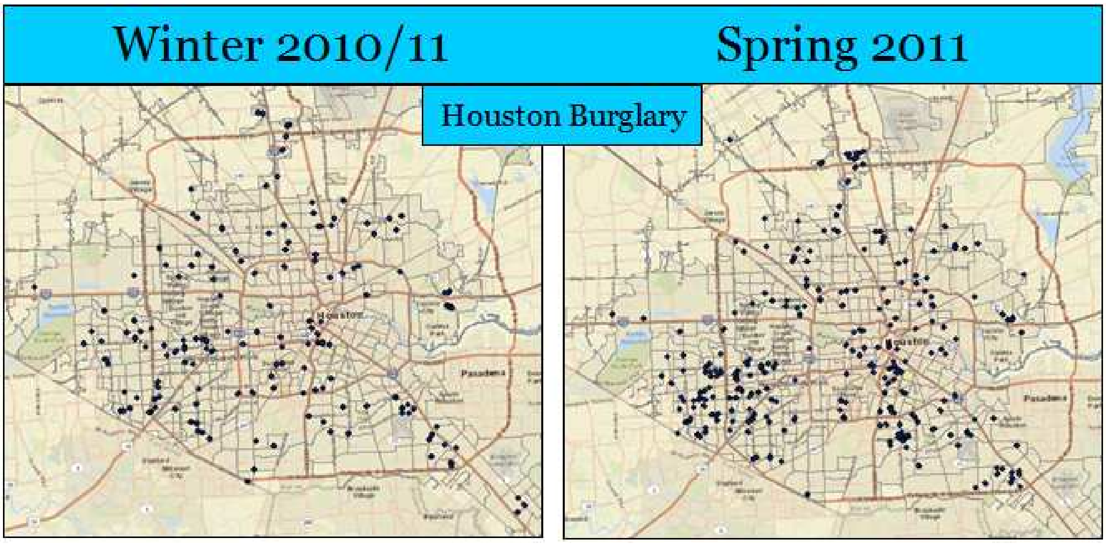


Figure 16: The results for Gi* for robberies for Vienna and Houston

The results for Vienna show many hot spots in the centre of the town, which are surrounded by a few cold spots for robberies. There are also a few cold spots at the edges of the city for burglaries. Results for Houston only show hot spots for all crime types, which seem to be spatially distributed across the entire city and located in similar areas.

4.2 Results of spatio-temporal cluster method – near repeat calculator

The final results displayed in ArcGIS can be seen in the following figures (17, 18).



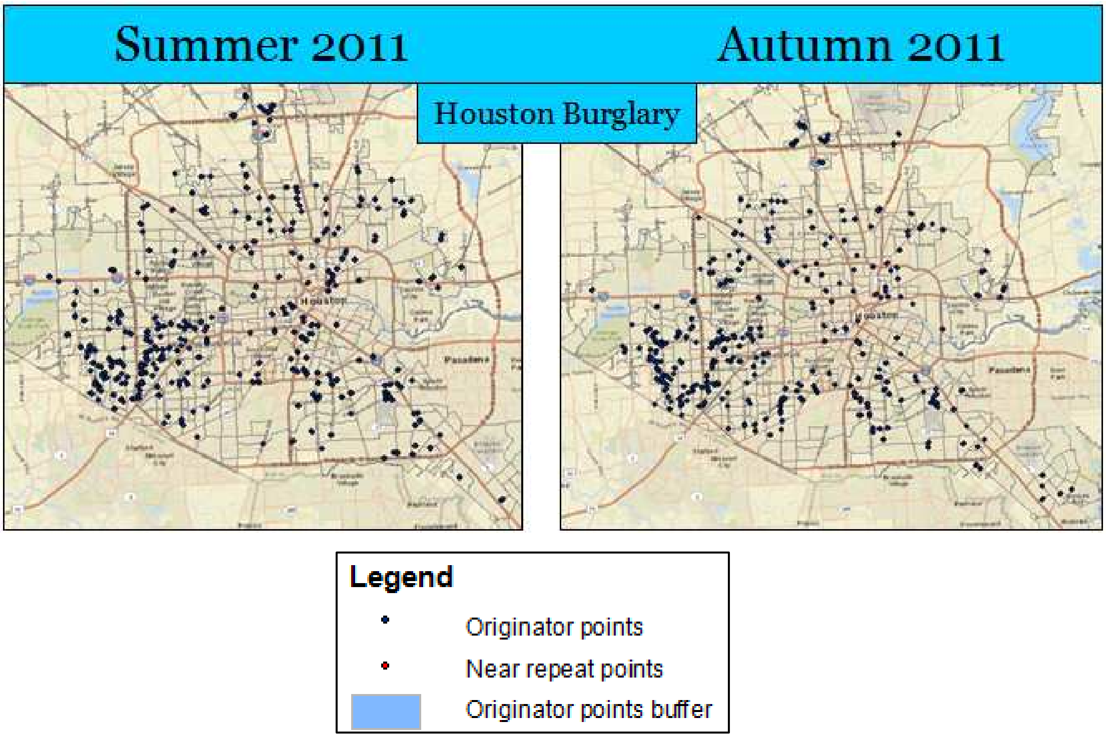
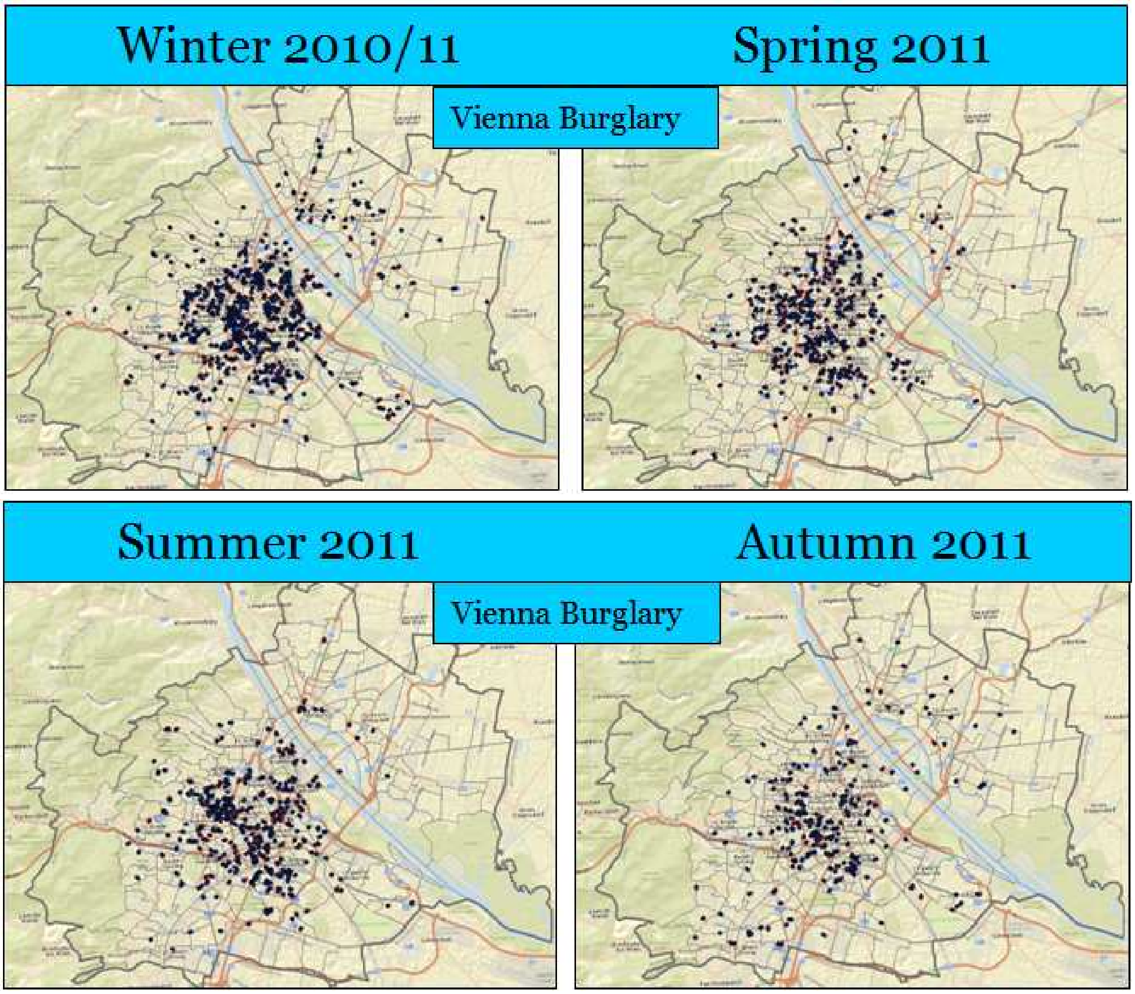


Figure 17: The results of the near – repeat analysis for all four seasons in 2011 for Houston for burglaries



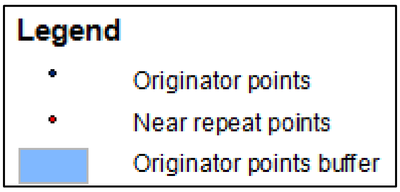


Figure 18: The results of the near – repeat analysis for all four seasons in 2011 for Vienna for burglaries

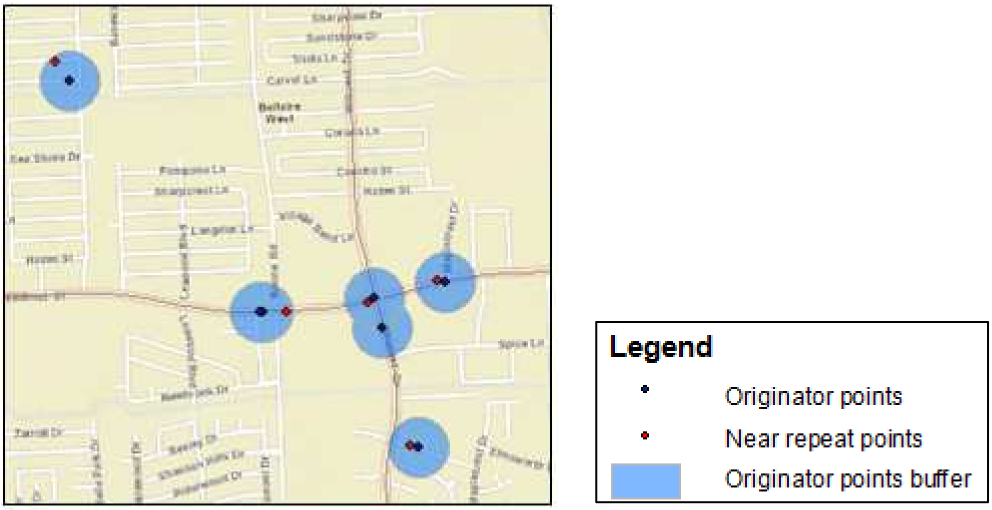


Figure 19: Detail results for burglaries for Houston for spring 2011

The originator points indicate the initial burglary event and are displayed in dark blue. Additional crimes that happen after that initial event in the spatial neighborhood (within a 125 m search radius shown with the light blue buffer) of the originating point and in the temporal proximity (7 days) after the originator incident are near repeat points. They are shown in red in figure 19.

So the results in this figure show, where a near repeat burglary has taken place within seven days and in proximity of 125 meters.

4.3 Evaluation of the four spatial cluster methods results with the PAI and the Hit Rate

Hot spot mapping is used to predict crime by using retrospective crime data to show areas of high crime concentration.

The article "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime" covers the prediction accuracy of hot spot mapping methods (Chainey et al. 2008). Results of methods used with already occured crimes shall show where crimes will happen next. The predicted crime data (for 2012) already exists for this thesis, so it is possible to say if the prediction was correct or incorrect. Results of the study by Chainey et al. (2008) show that there are differences between the results of the different methods and that there are also differences between crime types. Kernel density estimation had the best results from the different mapping techniques tested (spatial ellipses, choropleth mapping, grid thematic mapping and KDE were compared) (Chainey et al. 2008, p.1, 3).

The study by Chainey et al. (2008) and other previous studies that used mainly visual comparisons found out that different mapping techniques lead to different results concerning the found hot spots (place, size and shape) (Chainey et al. 2008, p.2). The study by Chainey et al. (2008) tests if the prediction ability for different crime types is the same (Chainey et al. 2008, p.2).

Measures are needed that inform how good a method is for the prediction of crime incidents. One measure is the hit rate that is the percentage of currently happened crime within previously predicted hot spot areas. One disadvantage is that the hit rate does not include the size of these areas and is therefore not as accurate as other prediction methods (Chainey et al. 2008, p. 10).

The Prediction Accuracy Index (PAI) considers the hit rate but takes the study area size also into account. The PAI is calculated as the hit rate divided by the area (Chainey et al. 2008, pp. 10). The higher the prediction accuracy index the better the hot spot prediction of the method is.

As can be seen in the formula, the PAI considers the number of incidents of the current time period that fall into the predicted hot spots of the previous time period as a proportion of the total number of incidents in the current time period and divides this by the proportion of the entire study area that are hotspots (Chainey et al. 2008, pp. 11).

$$\boxed{\text{Hit Rate} = \frac{n}{N} * 100}$$

Formula 1: Calculation of the Hit Rate

The Hit Rate is calculated as the percentage of the current crime incidents that fall into the hot spots from the previous time period. The result of the Hit Rate is in percent.

$$\boxed{PAI = \frac{(n/N)}{(a/A)}}$$

Formula 2: Calculation of the Prediction Accuracy Index (PAI)

n = number of incidents in predicted hot spots (incidents of 2012 in hot spots of 2011)
N = number of incidents in study area (from 2012)
a = area of hot spots (in km² from 2011)
A = total study area (in km²)

4.3.1 Workflow PAI and Hit Rate

The individual parts of the formula had to be found out in ArcGIS. “n” is the number of crime incidents from 2012 that fall into the predicted hot

spots from 2011. First, a layer with the right hot spot category had to be built with "Selection/ Create Layer from selected features".
For kernel density estimation the Z-scores were divided into 5 classes with equal interval, so that every class contains 20% of the data. From the highest class the layer for the hot spots was built.
For the nearest neighbor hierarchical clustering method the convex hulls as polygons exist and they were used as the hot spot category. For Vienna the clusters of first and second order were merged to gain the total area of hot spots.
The local Moran's I has five categories (insignificant, High-High, Low-Low, High-Low and Low-High). The High-High category represents the hot spots, so that category was chosen.
The Gi* has several categories showing significant hot spots. Significant hot spots with 95 and 99% confidence level were chosen as hot spot category.
The layer with the selected hot spot categories from the different methods and the pointlayer from the 2012 crime incidents were overlapped and cut with the tool "Intersect". "N" is the total number of incidents for 2012 and can be found in the pointlayer for Houston and Vienna. For the areas of the hot spots "a" a new column was added in the shapefile and it was calculated with the tools "Calculate Field" (!shape.area!) and the sum was gained with the tool "Summary Statistics". "A" for the entire area of the city was taken from the boundary-shapefile (the sum at Statistics). Then the PAI was calculated with Microsoft Excel.

4.3.2 Results of the PAI and the Hit Rate for burglary

The results of the different hot spot methods are shown in the next four figures (20 to 23) for the crime type burglary. The figures include the crime incidents that happened in 2012 and the hot spots calculated with the crime data from 2011.

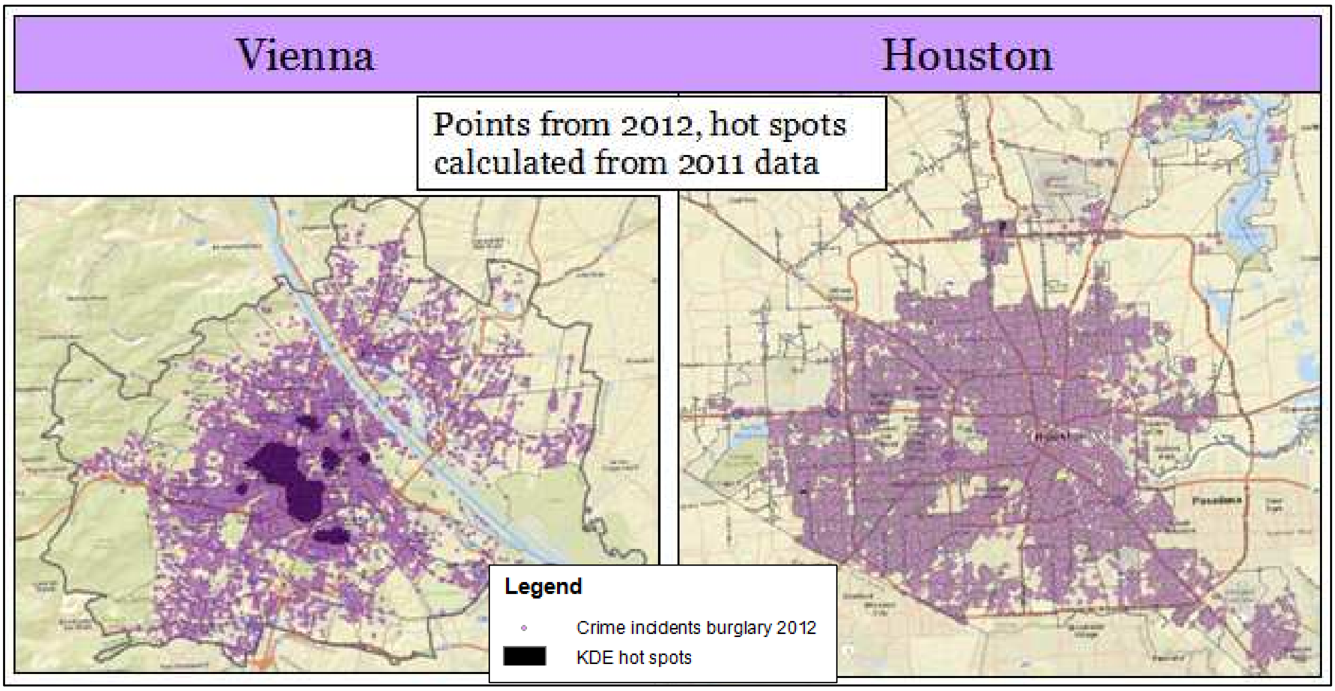


Figure 20: Kernel density estimation hot spots for burglary for 2011 and burglary incidents for 2012

The results of the **kernel density estimation** had the Z-scores divided into 5 classes with equal intervals and the class with the highest 20% of the density values was chosen as the hot spot areas. The hot spot areas are the black squares in figure 23. It can be seen that larger areas were found in Vienna, especially in the town centre with smaller clusters in the East, South, and West. In Houston only two very small clusters were found with one cluster located high up in the North and a second cluster located far in the West.

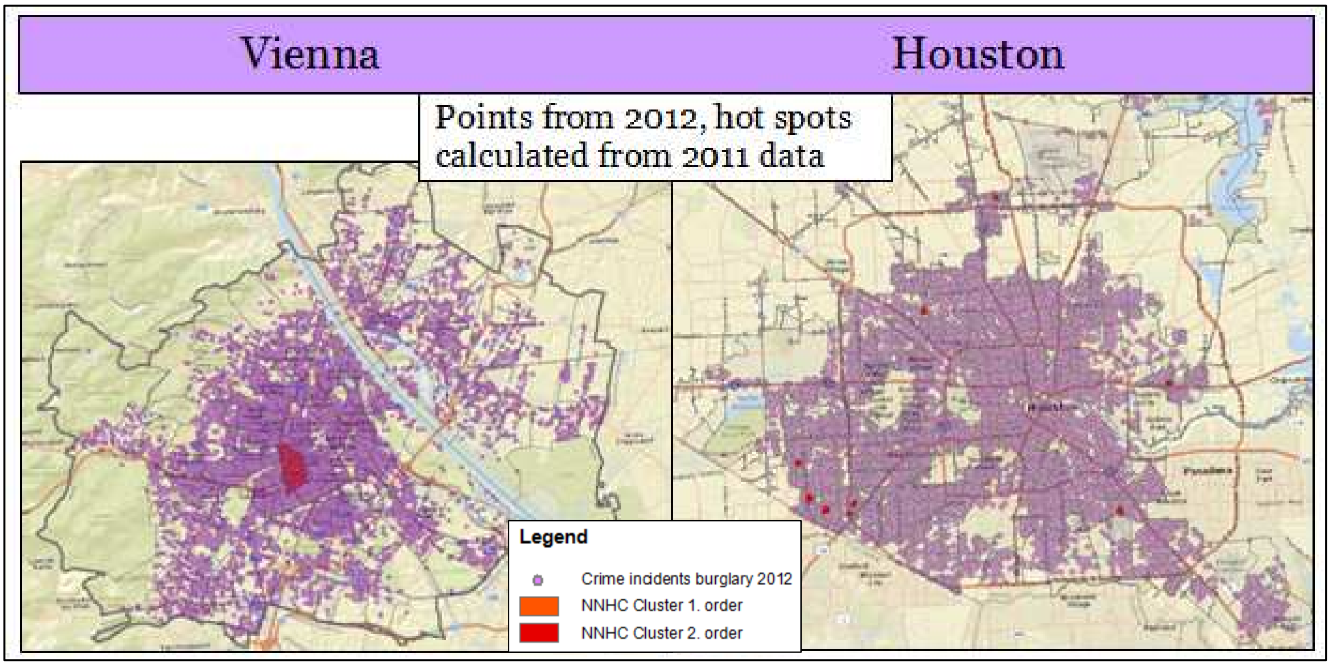


Figure 21: Nearest neighbor hierarchical clustering hot spots for burglary for 2011 and burglary incidents for 2012

The resulting hot spots of the **nearest neighbor hierarchical clustering** method for 2011 cover small areas. For Vienna first-order and second-order clusters were found. The first-order clusters (in red) are very small and are located at the edge of the second-order cluster (in orange). Again, all clusters lie in the town centre of Vienna. In Houston nine first-order clusters were found and most of them lie in the Southwest, two are located in the North, and two are located in the East.

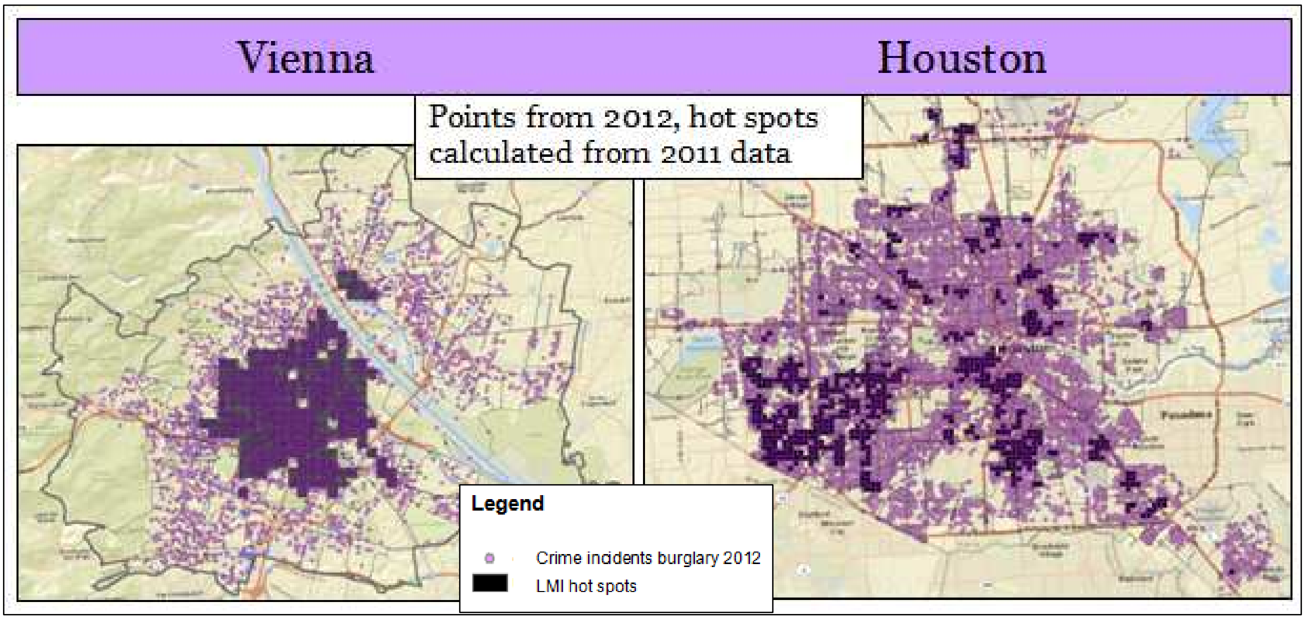


Figure 22: Local Moran's I hot spots for burglary for 2011 and burglary incidents for 2012

The results for the **local Moran's I** show bigger areas of hot spots. It can be seen that hot spots in Vienna are not only centred in the middle of the town but that there is also a hot spot on the other side of the Danube in the North. In Houston the hot spots are far more spread out over the entire city, many of them are located in the Southwest and in the South.

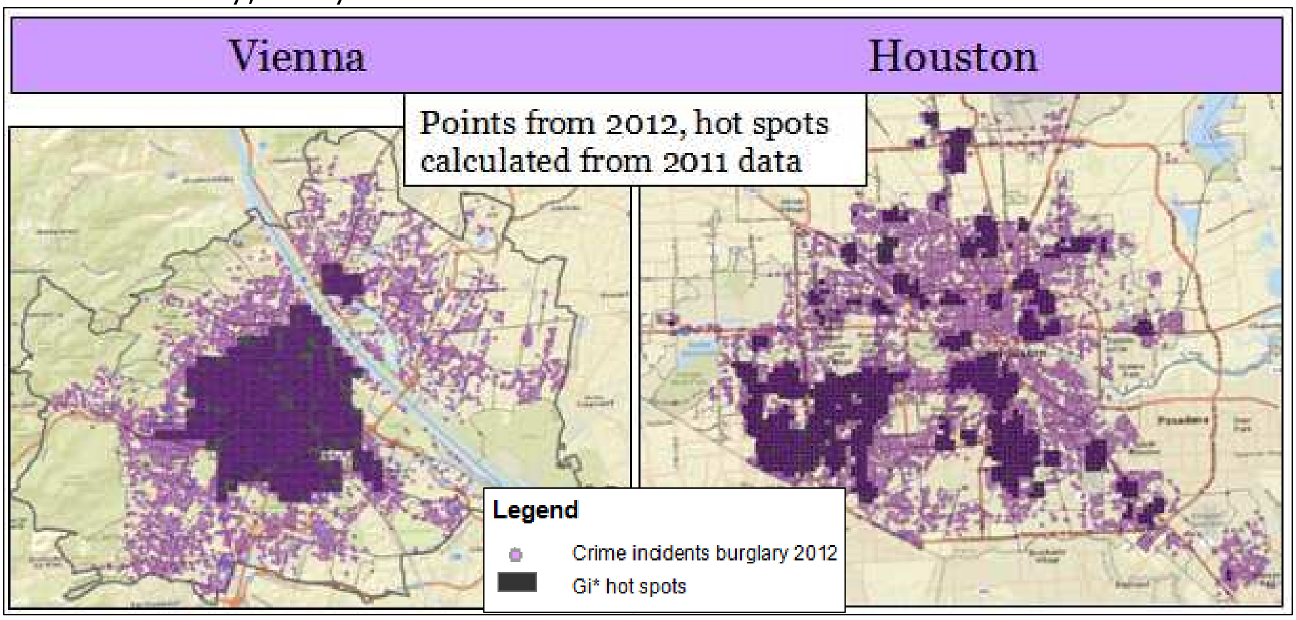


Figure 23: Gi* hot spots for burglary for 2011 and burglary incidents for 2012

The **Gi*** results also show a larger part of the study area as hot spot areas. In Vienna nearly every hot spot is again focused on the centre with a smaller hot spot located further away from the centre in the North, similar to the results found with the local Moran's I. In Houston, hot spot areas are again spread out over the entire city, but in contrast to the local Moran's I hot spots, the Gi* hot spot areas are more connected to each other.

The results of the PAI and the Hit Rate for the crime type burglary are pointed out in table 8.

Burglary

PAI	Vienna	Houston
KDE	**8,13**	**34,63**
NNHC	0,0082	21,58
LMI	4,83	4,32
Gi*	5,11	3,06
Hit Rate:		
KDE	4,79	1,43
NNHC	4,39	5,09
LMI	59,46	41,93
Gi*	**62,89**	**49,55**

Table 8: Results of the PAI and the Hit Rate for burglary

For the crime type burglary the kernel density estimation had the highest PAI with 8.13 for Vienna and 34.63 for Houston. The Gi* method had the best results for the Hit Rate.
The correctly predicted crime incidents per area for the two cities are shown in the following two figures (figures 24 and 25).

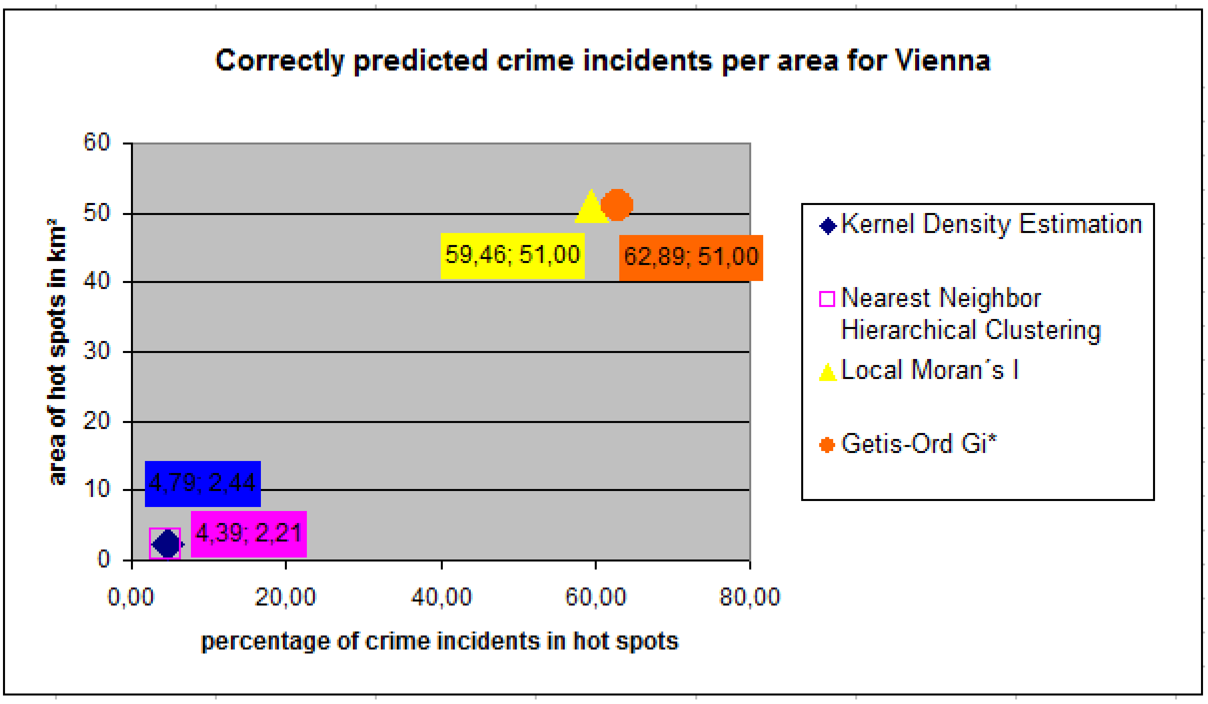


Figure 24: Correctly predicted crime incidents per area for Vienna for burglary

In the case of burglary, the local Moran's I and the Gi* correctly predicted a high percentage of crime incidents in 2012 (59.46 % for the LMI to 62.89 % for the Gi*), which cover a large proportion of the total study area (51 km² for both LMI and Gi*). The kernel density estimation and the nearest neighbor hierarchical clustering methods predicted a lower percentage of crime incidents for 2012 (4.79 % for the KDE and 4.39 % for the NNHC) with the area of the hot spots being a smaller proportion compared to the study area (2.44 km² for the KDE and 2.21 km² for the NNHC.

The results for Houston can be seen in the next figure 25:

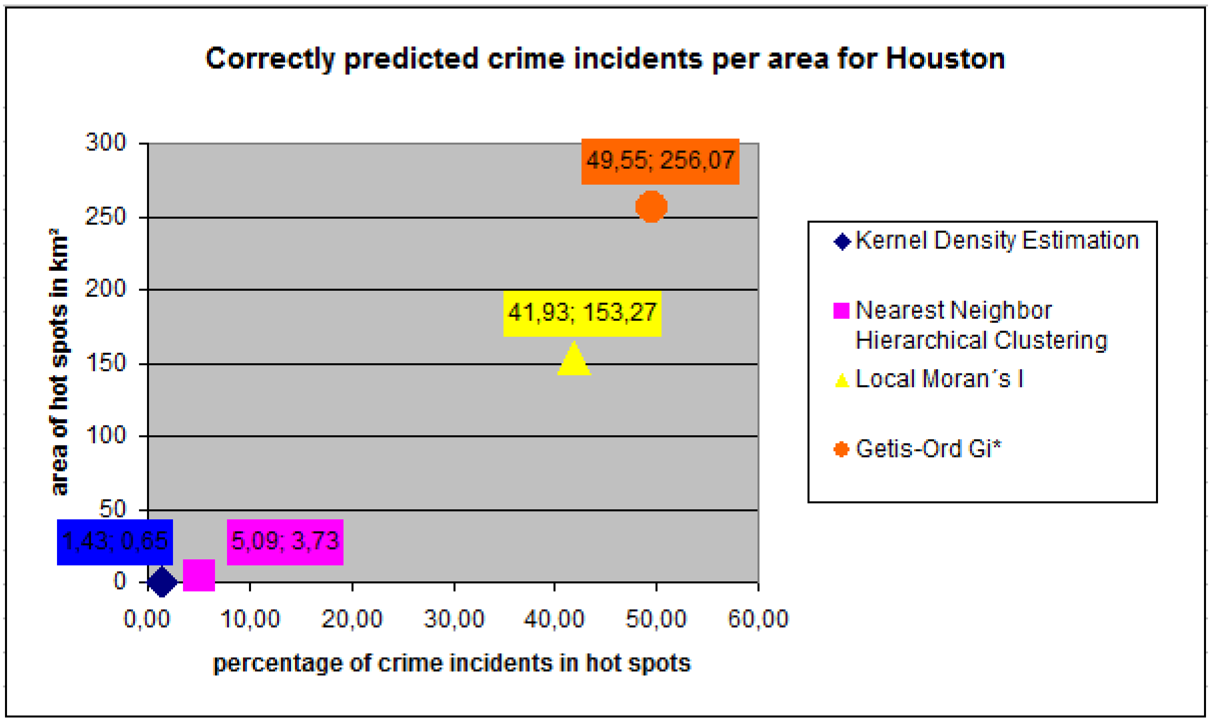


Figure 25: Correctly predicted crime incidents per area for Houston for burglary

The KDE and the NNHC hot spots cover again only a small proportion of the total study area (0.65 km² for KDE and 3.73 km² for NNHC) and found a rather low percentage of correctly predicted crime incidents for 2012 (1.43 % for the KDE and 5.09 % for the NNHC). The LMI and the Gi* hot spots cover a larger area of the study area (153.27 km² for the LMI and 256.07 km² for the Gi*) and predicted a larger percentage of crime incidents for 2012 (41.93 % for the LMI and 49.55 % for the Gi*). Although the LMI and the Gi* predicted a higher percentage of crime incidents, the hot spot areas that they cover of the total area is so big that the results are not so good as they seem. The KDE and the NNHC were more accurate in predicting crime incidents for 2012 because their hot spots covered only a very small amount of the total study area.

When the percentage of the study area covered by hot spots is calculated it is possible to compare the results of the two cities with each other. This percentage is calculated as follows:

Percentage of the study area covered by hot spots = area of hot spots/ study area size * 100

The resulting hot spot area percentages (see table 9) show that the local Moran's I and the Gi* cover between 10 % and 17 % of the total study area, however, the kernel density estimation and the nearest neighbor hierarchical clustering cover a very small proportion (less than 1 percent) of the total area for both cities. The best method for the prediction accuracy index was the kernel density estimation for both cities, because the predicted hot spot areas cover a very small proportion of the study areas.

- **Vienna, Austria:**
  - Total area 415 km²

covered percentage of study area:	
KDE	0.59%
NNHC	0.53%
LMI	12.29%
Gi*	12.29%

- **Houston, Texas:**
  - Total area 1,500 km²

covered percentage of study area:	
KDE	0.04%
NNHC	0.25%
LMI	10.22%
Gi*	17.07%

Table 9: The percentage of the study area covered by hot spots for burglary

4.3.3 Results of the PAI and the Hit Rate for assault

The results of the different hot spot methods are shown in the next four figures (26 to 29) for the crime type assault. The figures include the crime incidents that happened in 2012 and the hot spots calculated with the crime data from 2011.

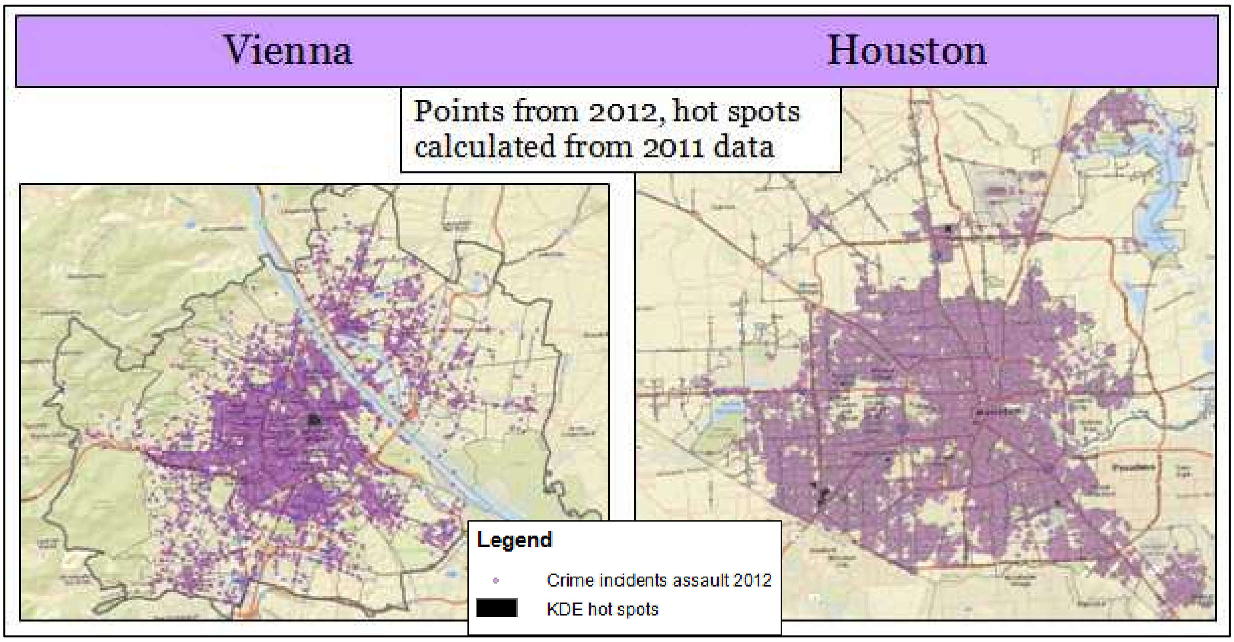


Figure 26: Kernel density estimation hot spots for assault for 2011 and assault incidents for 2012

The results of the **kernel density estimation** show only small areas and in Vienna all hot spots lie in the centre of the city. In Houston the hot spots are more divided and are located in the North, East, Southeast, Southwest and the centre of the town. The highest concentration of hot spots lies in the Southwest of the city.

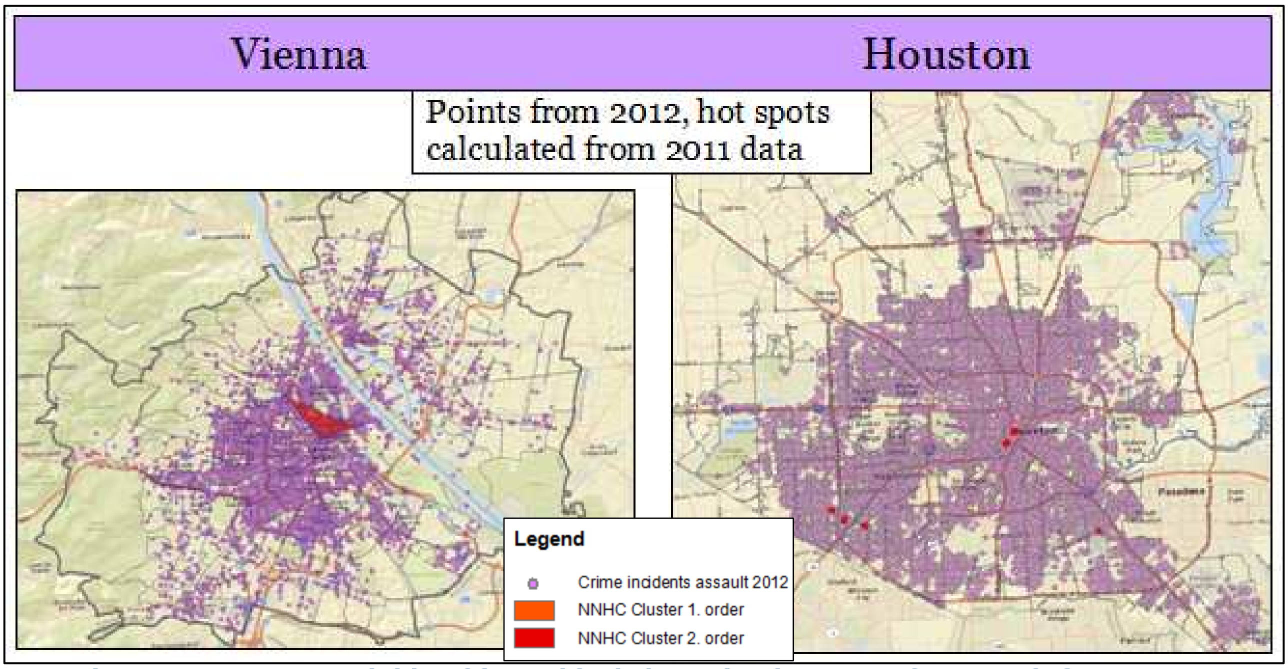


Figure 27: Nearest neighbor hierarchical clustering hot spots for assault for 2011 and assault incidents for 2012

The resulting hot spots of the **nearest neighbor hierarchical clustering** method for 2011 cover also very small areas. For Vienna first-order and second-order clusters were found. The first-order clusters (in red) are very small and are located at the edge of the second-order cluster (in orange). Again, all clusters lie central in the town centre but are shifted more to the Northeast of Vienna in comparison to the other crime types. In Houston seven first-order clusters were found especially in the centre and the Southwest of the town. Isolated clusters are found in the Southeast and in the North.

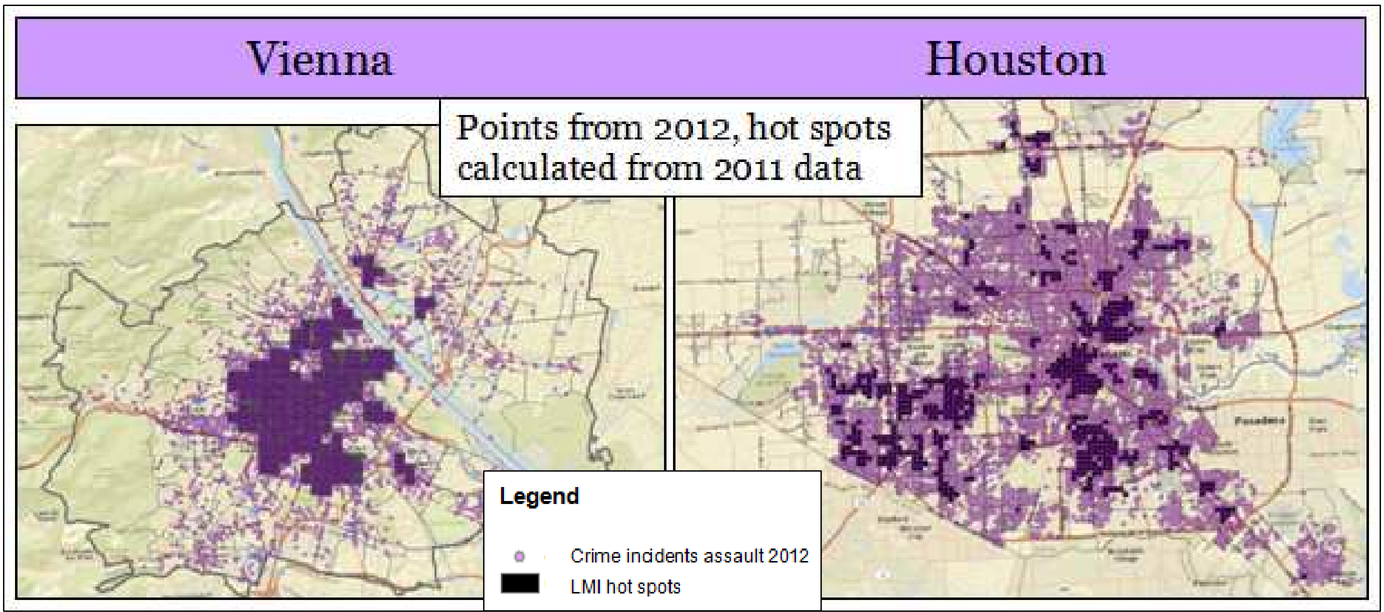


Figure 28: Local Moran’s I hot spots for assault for 2011 and assault incidents for 2012

The results for the **local Moran's I** show a bigger area of hot spots. Here it can be seen that hot spots in Vienna are not only centred in the middle of the town but that there are small hot spots on the outskirts on the other side of the Danube. In Houston the same area as before (but bigger) is the result for hot spots and there are also more hot spots spread out over the whole city in comparison to Vienna.

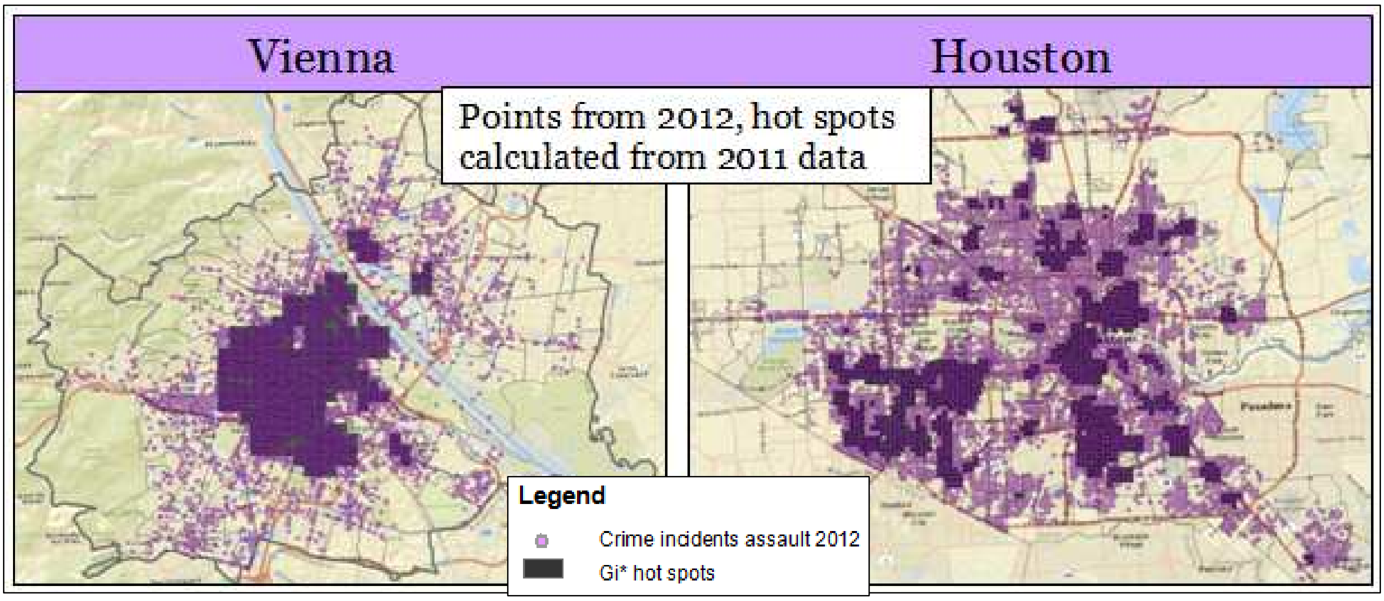


Figure 29: Gi* hot spots for assault for 2011 and assault incidents for 2012

The **Gi*** results also show a larger part of the study area as hot spot areas. In Vienna nearly every hot spot is again focused on the centre with two smaller hot spots located further away, similar to the results found with the local Moran's I. In Houston, the same hot spot areas were found but they are more connected to each other than in the local Moran's I method.

The results of the PAI and the Hit Rate for the crime type assault are included in table 10.

Assault

PAI	Vienna	Houston
KDE	**109,39**	17,17
NNHC	16,46	**20,00**
LMI	6,56	5,11
Gi*	5,68	3,44
Hit Rate:		
KDE	2,88	3,13
NNHC	10,19	4,34
LMI	64,48	48,50
Gi*	**68,52**	**55,49**

Table 10: Results of the PAI and the Hit Rate for assault

For the crime type assault the kernel density estimation had the highest PAI with 109.39 for Vienna and the nearest neighbor hierarchical clustering with 20.00 for Houston. The Gi* method had again the best results for the Hit Rate.
The correctly predicted crime incidents per area for the two cities are shown in the following two figures (figure 30 and 31).

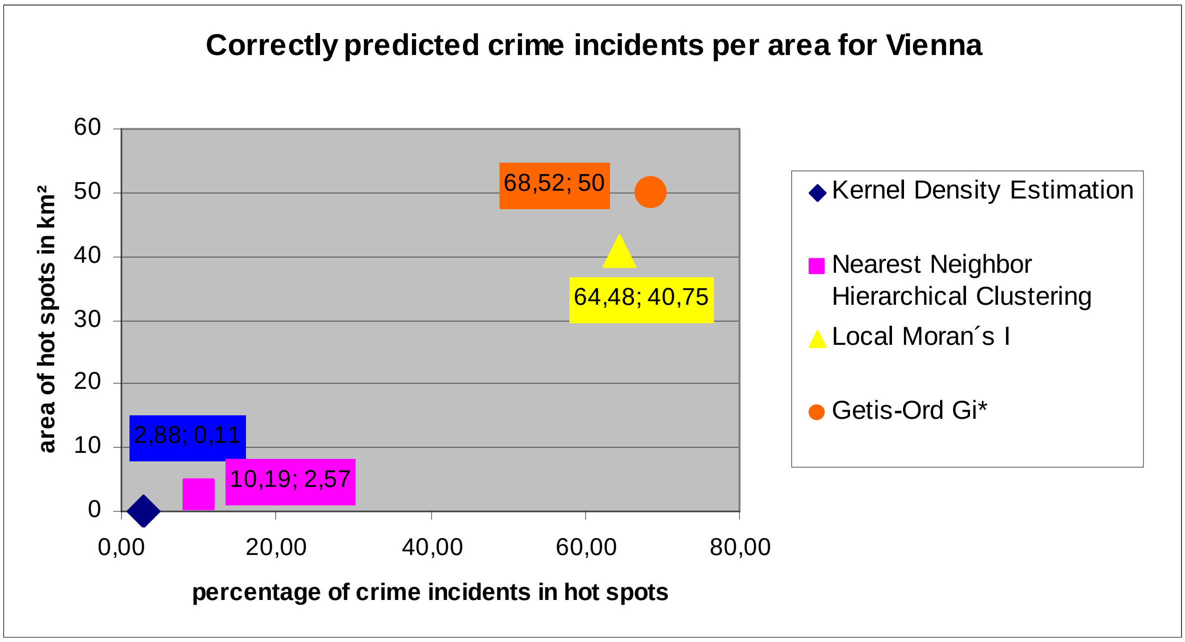


Figure 30: Correctly predicted crime incidents per area for Vienna for assault

In the case of assault, the local Moran's I and the Gi* correctly predicted a high percentage of crime incidents in 2012 (64.48 % for the LMI to 68.52 % for the Gi*), which cover a large proportion of the total study area (40.75 km² for the LMI and 50 km² for the Gi*). The kernel density estimation and the nearest neighbor hierarchical clustering methods predicted a lower percentage of crime incidents for 2012 (2.88 % for the KDE and 10.19 % for the NNHC) and found not so much incidents in this area but the area of the hot spots is again very small compared to the study area (0.11 km² for the KDE and 2.57 km² for the NNHC).

The results for Houston can be seen in the next figure 31:

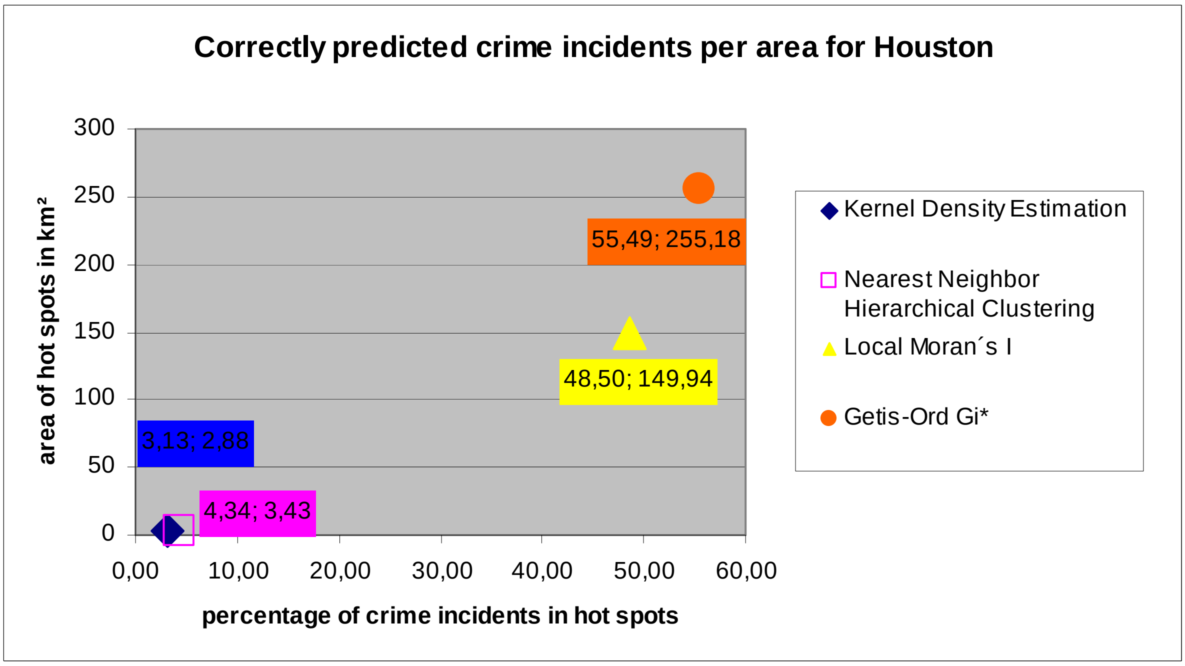


Figure 31: Correctly predicted crime incidents per area for Houston for assault

The KDE and the NNHC hot spots cover again only a small proportion of the total study area (2.88 km² for KDE) and 3.43 km² for NNHC) and found a rather low percentage of correctly predicted crime incidents for 2012 (3.13 % for the KDE and 4.34 % for the NNHC). The LMI and the Gi* hot spots cover a larger area of the study area (149.94 km² for the LMI and 255.18 km² for the Gi*) and predicted a larger percentage of crime incidents for 2012 (48.50 % for the LMI and 55.49 % for the Gi*).

The resulting hot spot area percentages (see table 11) show that the local Moran's I and the Gi* methods cover between 10 % and 17 % of the total study area, however, the kernel density estimation and the nearest neighbor hierarchical clustering cover a very small proportion (less than 1 percent) of the total area for both cities. The best method for the prediction accuracy index was the kernel density estimation for both cities, because the predicted hot spot areas cover a very small proportion of the study areas.

- **Vienna, Austria:**
  - Total area 415 km²

covered percentage of study area:	
KDE	0.03%
NNHC	0.62%
LMI	9.82%
Gi*	12.05%

- **Houston, Texas:**
  - Total area 1,500 km²

covered percentage of study area:	
KDE	0.19%
NNHC	0.23%
LMI	10.00%
Gi*	17.01%

Table 11: The percentage of the study area covered by hot spots for assault

4.3.4 Results of the PAI and the Hit rate for robbery

The results of the different hot spot methods are shown in the next four figures (32 to 35) for the crime type robbery. The figures include the crime incidents that happened in 2012 and the hot spots calculated with the crime data from 2011.

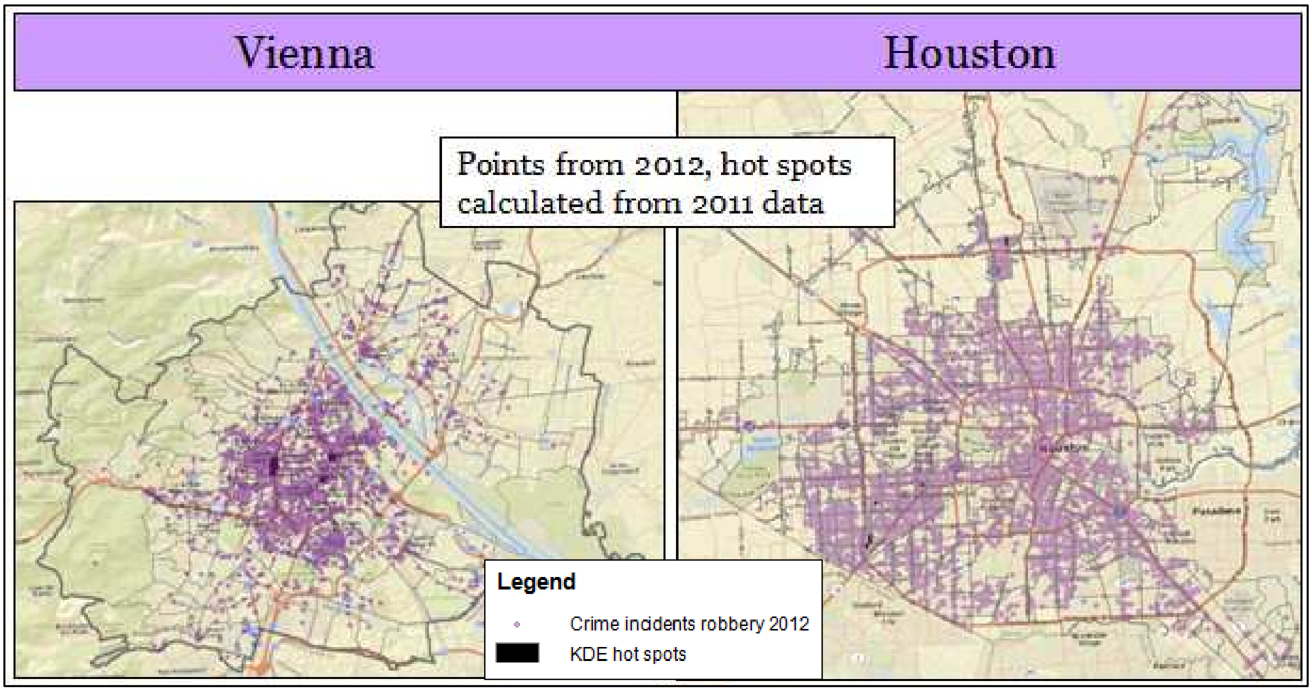


Figure 32: Kernel density estimation hot spots for robbery for 2011 and robbery incidents for 2012

The results of the **kernel density estimation** had the Z-scores divided into 5 classes with equal intervals and the class with the highest 20% of the density values was chosen as the hot spot areas. The hot spot areas are the black squares in figure 35. It can be seen that small areas were found and in Vienna they all lie in the centre of the city. In Houston the hot spots are more divided and are located in the Southwest and in the centre of the town.

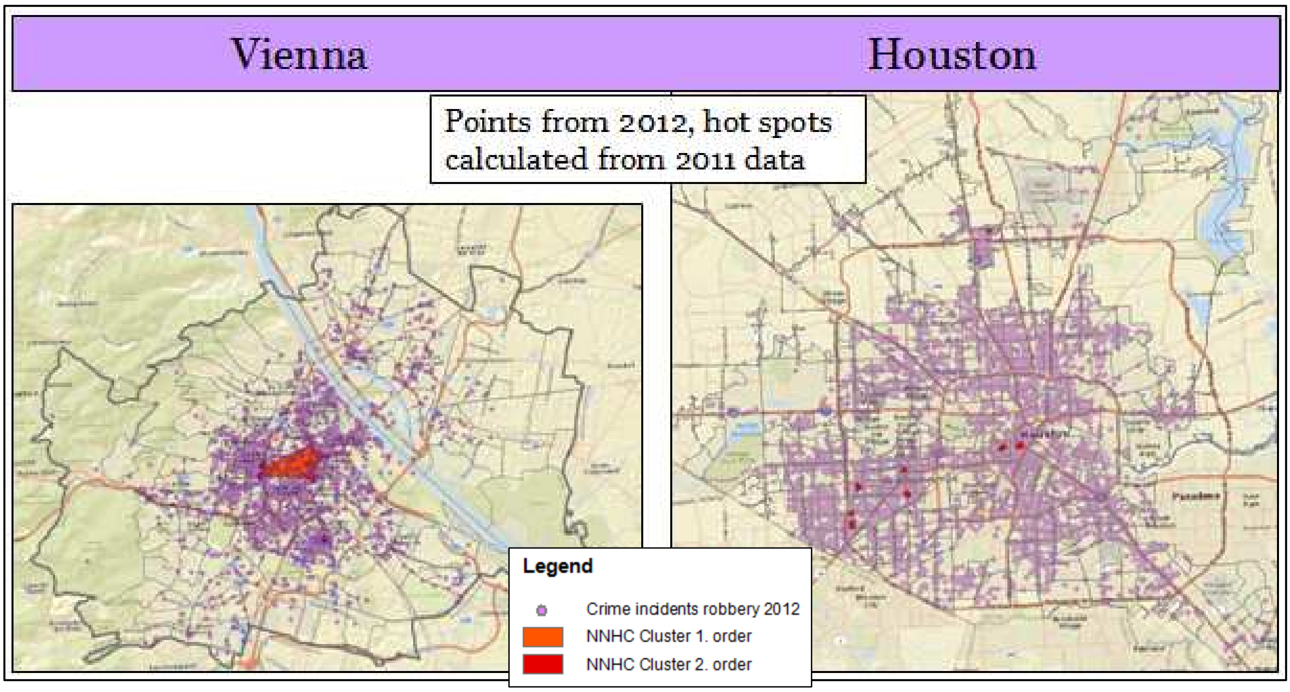


Figure 33: Nearest neighbor hierarchical clustering hot spots for robbery for 2011 and robbery incidents for 2012

The resulting hot spots of the **nearest neighbor hierarchical clustering** method for 2011 cover very small areas. For Vienna first-order and second-order clusters were found. The first-order clusters (in red) are very small and are located at the edge of the second-order cluster (in orange). Again, all clusters lie in the town centre of Vienna. In Houston eight first-order clusters were found but they are too far away from each other to build a second-order cluster. Most of them are located in the Southwest of the town.

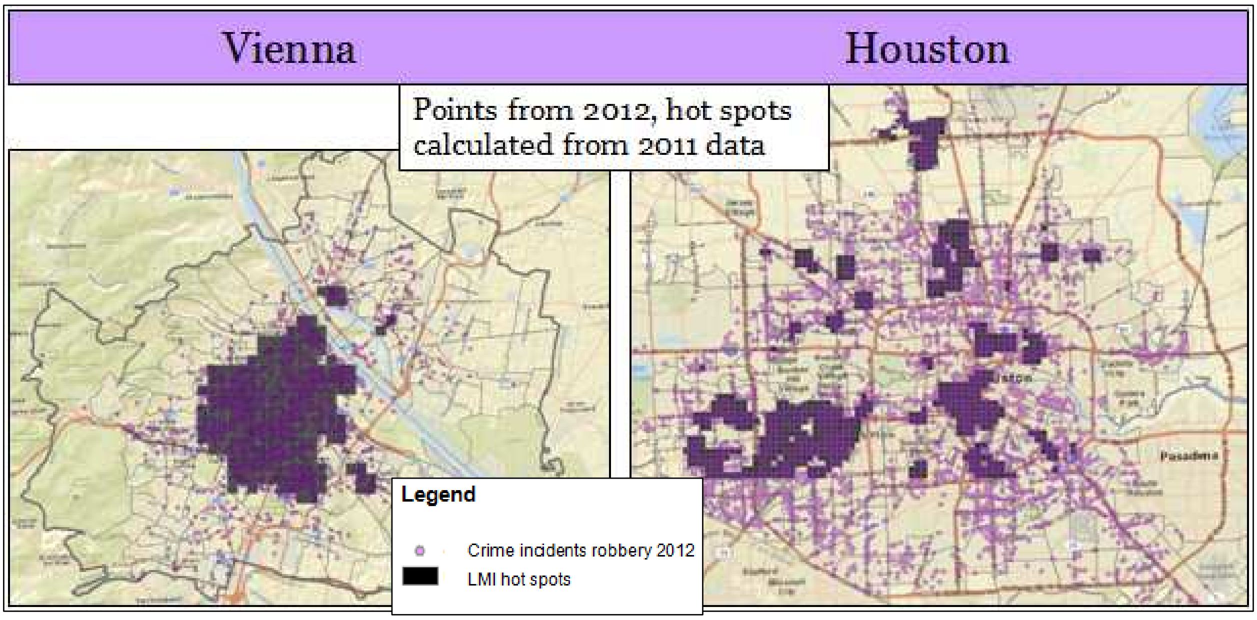


Figure 34: Local Moran's I hot spots for robbery for 2011 and robbery incidents for 2012

The results for the **local Moran's I** show a bigger area of hot spots. It can be seen that hot spots in Vienna are not only centred in the middle of the town but that there are also small hot spots on the outskirts on the other side of the Danube. In Houston the same area as before (but bigger) is the resulting hot spots and there are also more hot spots spread out over the whole city in comparison to Vienna.

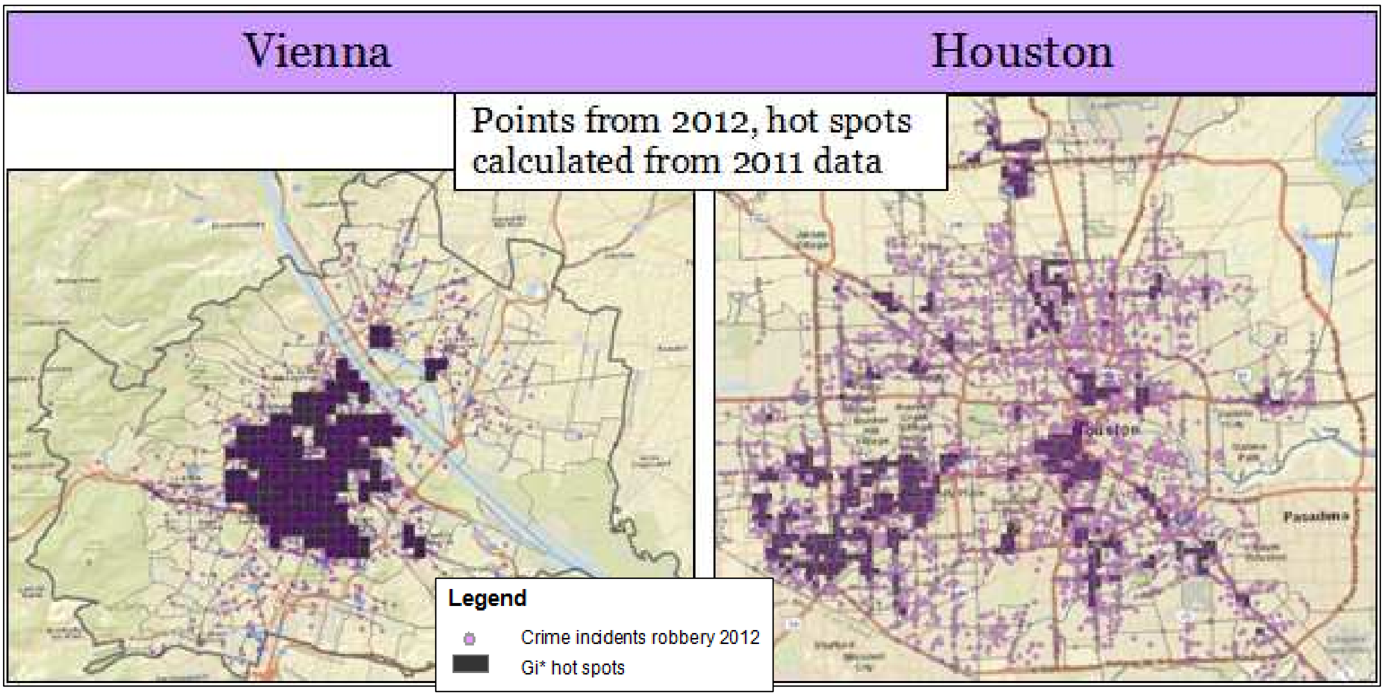


Figure 35: Gi* hot spots for robbery for 2011 and robbery incidents for 2012

The **Gi*** results also show a larger part of the study area as hot spot areas. In Vienna nearly every hot spot is again focused on the centre with smaller hot spots located further away from the centre, similar to the results found with the local Moran's I. In Houston, hot spot areas are again spread out over the entire city, but in contrast to the local Moran's I hot spots, the Gi* hot spot areas are more connected to each other.
The results for the crime type robbery are shown in table 12.

Robbery

PAI:	Vienna	Houston
KDE	**41,76**	14,1
NNHC	14,39	**27,21**
LMI	7,19	3,08
Gi*	5,99	1,98
Hit Rate:		
KDE	5,32%	0,83%
NNHC	11,97%	6,09%
LMI	68,52%	23,39%
Gi*	**74,83%**	**26,18%**

Table 12: Results of the PAI and the Hit Rate for robbery

For Vienna the kernel density estimation had the highest PAI with 41.76. For Houston the best method was the nearest neighbor hierarchical clustering with a PAI of 27.21.
The Gi* had the best Hit Rate for both cities. But the hit rate does not consider the area that is covered from the found hot spots and therefore it is not such a reliable tool for prediction measuring than the prediction accuracy index. But compared with the area of the found hot spots the correctly predicted crime incidents per area for the two cities can be illustrated (figure 36):

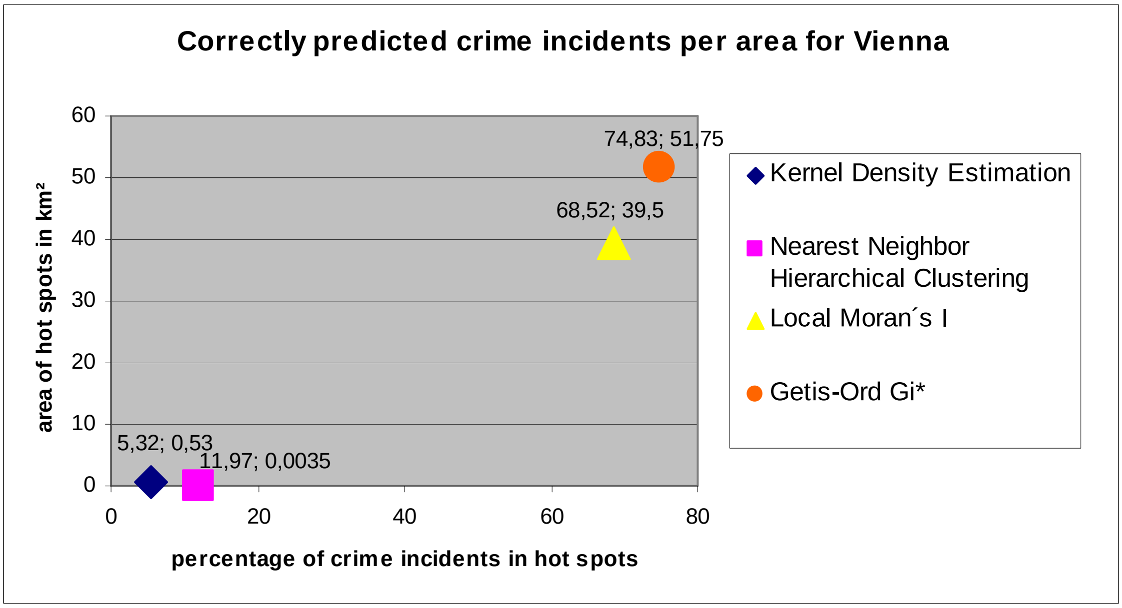


Figure 36: Correctly predicted crime incidents per area for Vienna for robbery

In the case of robbery, the local Moran's I and the Gi* correctly predicted a high percentage of crime incidents in 2012 (68.52% for the LMI to 74.83% for the Gi*), which cover a large proportion of the total study area (39.5 km² for the LMI and 51.75 km² for the Gi*). The kernel density estimation and the nearest neighbor hierarchical clustering methods predicted a lower percentage of crime incidents for 2012 (5.32 % for the KDE and 11.97 % for the NNHC) with the area of the hot spots being a smaller proportion compared to the study area (0.53 km² for the KDE and 0.0035 km² for the NNHC). So the last two methods are more accurate in predicting hot spots in total because the covered areas are very small.

For Houston nearly the same results were found when the area of the found hot spots is compared to the total study area in percent (see figure 37).

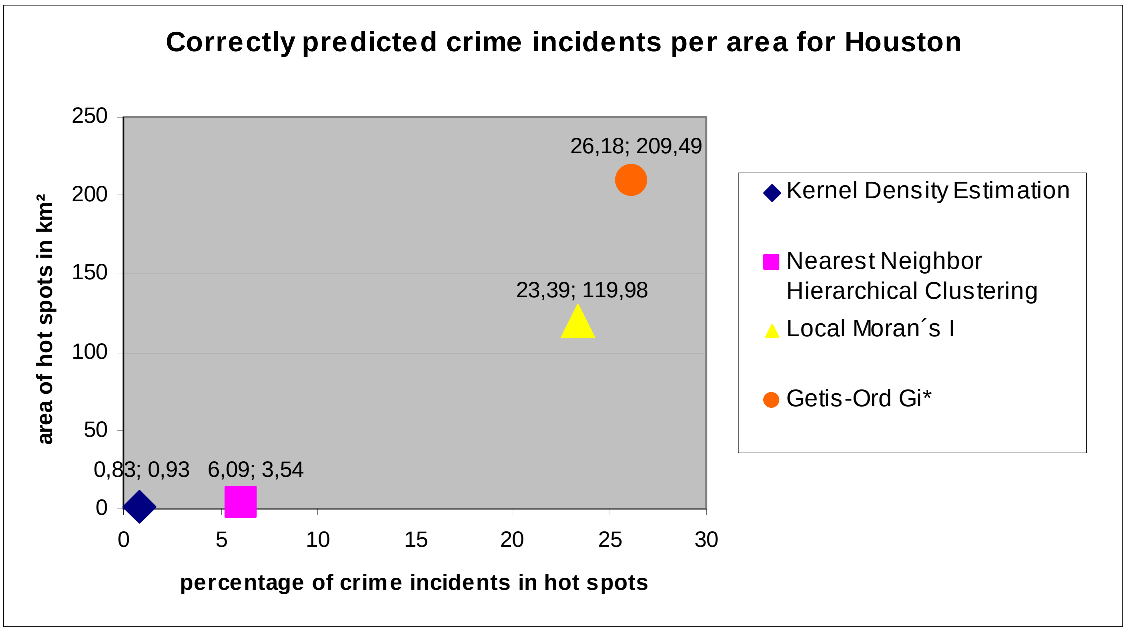


Figure 37: Correctly predicted crime incidents per area for Houston for robbery

The KDE and the NNHC hot spots cover again only a small proportion of the total study area (0.93 km² for KDE and 3.54 km² for NNHC) and found a rather low percentage of correctly predicted crime incidents for 2012 (0.83 % for the KDE and 6.09 % for the NNHC). The LMI and the Gi* hot spots cover a larger area of the study area (119.98 km² for the LMI and 209.49 km² for the Gi*) and predicted the following percentage of crime incidents in hot spots (23.39 % for the LMI and 26.18% for the Gi*).

The resulting hot spot area percentages (see table 13) show that the local Moran's I and the Gi* cover between 8 % and 14 % of the total study area, however, the kernel density estimation and the nearest neighbor hierarchical clustering cover a very small proportion (less than 1 percent) of the total area for both cities. The best methods for the prediction accuracy index were the kernel density estimation and the nearest neighbor hierarchical clustering, because the predicted hot spot areas cover both only a very small proportion of the study areas.

- **Vienna, Austria:**
  - Total area 415 km²

covered percentage of study area:	
KDE	0.13%
NNHC	0.0008%
LMI	9.52%
Gi*	12.47%

- **Houston, Texas:**
  - Total area 1,500 km²

covered percentage of study area:	
KDE	0.062%
NNHC	0.236%
LMI	8.00%
Gi*	13.97%

Table 13: The percentage of the study area covered by hot spots for robbery

4.4 Interpretation of results

In this subchapter the used crime data as well as the results of the four spatial cluster methods and the near-repeat analysis are interpreted.

4.4.1 Interpretation of Crime Data and Statistics

The author of this thesis analyzed data of reported crimes (these are crimes known to the police). In fact, there are more crimes happening and there are certain crime types that are not so often reported (for example domestic violence, rape, drug crimes, etc.). On the other hand, other crime types are not as affected from not reporting the crime to the police. Those crime types include motor vehicle theft, murder, etc. (Boba 2008, p. 2).

The analyzed crime types in this thesis were burglary, robbery, and assault. Robbery is happening in public and reports are often made because of insurance issues, which is also true for burglary. These two crime types must be well known and often reported to the police. Assaults can happen in public as well as in private. Private assaults may not be that often reported to the police than public assaults.

Important is also the exact date and time when a crime was permitted (Boba 2008, p. 3).

For robbery and assault the exact date and time are often known, for burglary there is often only an estimation given because no one has seen the crime (Boba 2008, p. 3).

4.4.2 Interpretation of the four used cluster methods

When comparing the results described in subchapter 4.1 for the two cities, differences and similarities become apparent. Differences between Houston and Vienna were found in the amount of crime incidents per 100,000 of the population (i.e., the crime rate):

$$\boxed{\text{Crime rate} = \text{number of crime incidents /population number} * 100{,}000}$$

Formula 3: Calculation of the crime rate

For 2011 there were:

- 1.4 times more burglaries
- 2.4 times more robberies
- 2.6 times more assaults in Houston than in Vienna

For 2012 there were:

- 1.1 times more burglaries
- 2.8 times more robberies
- 2.3 times more assaults in Houston than in Vienna

So the crime rate is somewhat higher in Houston than in Vienna. Robbery in Vienna is more concentrated in one area that is the centre of the city. In

Houston, robbery is more spread out over small areas in different parts of the city. In Vienna, a large hot spot for burglary is located in the town centre. But this hotspot is more orientated to the South than the robbery hot spot. Burglary in Houston also covers small area clusters that are mostly located in the Southwest/ West of the city. Assault hot spots in Vienna are located in the town centre, but more orientated to the North than the robbery hot spot. The resulting hot spots for Houston are again spread out over the city, with a concentration in the Southwest, the centre and also a small hot spot in the North.
The best method for predicting crime incidents measured with the PAI was the kernel density estimation for Vienna and the nearest neighbor hierarchical clustering method for Houston for robbery and assault. For the crime type burglary the best method for both cities was the kernel density estimation.

Similarities between Houston and Vienna can also be found. The used cluster methods showed similar results across different methods and crime types for both cities. The LMI and the Gi* resulted in many hot spot areas covering a large proportion of the study area. The KDE and the NNHC predict only a few hot spot areas, which are rather small.
The different sizes of the two study areas do not have such a big impact on the results. For example, when calculating the percentage of hot spot sizes, the percent values are similar for Houston and Vienna. For all crime types, the proportion of the study area covered by the resulting hot spots are between 8 % and 17 % for the LMI and the Gi* and less than 1 % for the KDE and the NNHC. From this point of view, it makes sense to compare these cities. Overall, the cluster methods applied in this thesis are applicable to crime prediction for Vienna and the methods that predict smaller areas with a higher prediction accuracy should be used. These methods include the KDE and the NNHC.

4.4.3 Interpretation of the Near-repeat analysis

In this subchapter the results of the near-repeat analysis for Houston and Vienna are interpreted for the four seasons (see table 14).

Houston:	winter 2010/11	spring 2011	summer 2011	autumn 2011
repeat victimization pattern	same location 0 to 7 days 72 %	same location 0 to 7 days 38 %	same location 0 to 7 days 51 %	same location 0 to 7 days 61%
near repeat victimization pattern	-	-	-	1 to 125 meters 0 to 7 days 26 %
Vienna:	**winter 2010/11**	**spring 2011**	**summer 2011**	**autumn 2011**
repeat victimization pattern	same location 0 to 7 days 317 %	same location 0 to 7 days 442 %	same location 0 to 7 days 363 %	same location 0 to 7 days 338 %
near repeat victimization pattern	-	-	1 to 125 meters 0 to 7 days 41 %	1 to 125 meters 0 to 7 days 38 %

Table 14: Overview of results of the near-repeat calculator for Vienna and Houston

Repeat victimization means that on the same location where one incident has already happened, the chance for another incident to happen in a short period of time is significantly greater. For example, for Houston for autumn 2011 there was a repeat victimization pattern found at the same place for up to 7 days after an initial incident with a 61 percent greater chance of another incident.

Near repeat victimization means that additional incidents happen at a specific distance away from the first incident but not exactly at the same location. For Houston, for autumn 2011 there was a near repeat victimization pattern found. Within 1 to 125 meters of an initial incident, near repeats are overrepresented for up to 7 days. The chance of another incident is about 26 percent greater than if there were a random pattern. Looking at table 8 the results are shown in an overview and there are obviously not so much near repeat patterns found. However, repeat victimization patterns where found everywhere. In Vienna the percentage of another incident to happen is extremely high (317 % to 442 %) than compared to Houston (38 % to 72 %). This is the case, because the incidents in Vienna are nearly all concentrated in the centre of the city, while in Houston the incidents are spread over the whole city.

The results below (figure 38) show where a near repeat burglary has taken place in Houston and in Vienna within seven days and in a surrounding of 125 meters for the four seasons in 2011.

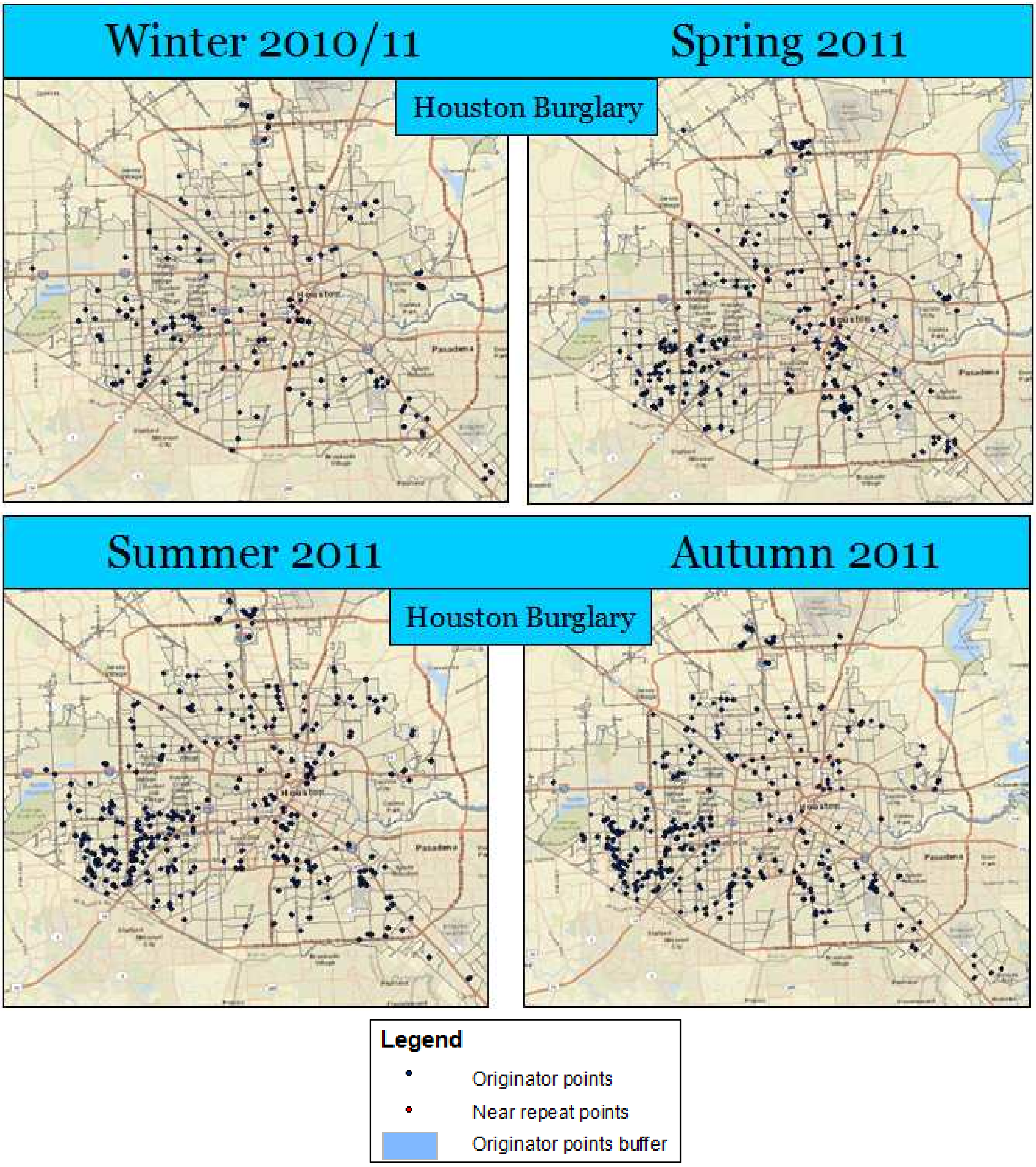


Figure 38: Results of the near – repeat analysis for the four seasons of 2011 for Houston for burglaries

In figures 38 and 40 the originator points can be seen as black dots. The near repeat points are so close to the originator points and the buffer is also so small, that they can't be seen in this figure, but an example map inset can be seen in figure 39. The near repeat points are scattered over the city showing concentrations in similar areas across all four seasons. Many near repeats are located in the Southwest of Houston, in the centre of the town to the south, farther away in the North, and the Southeast and to a lesser degree also in the East.

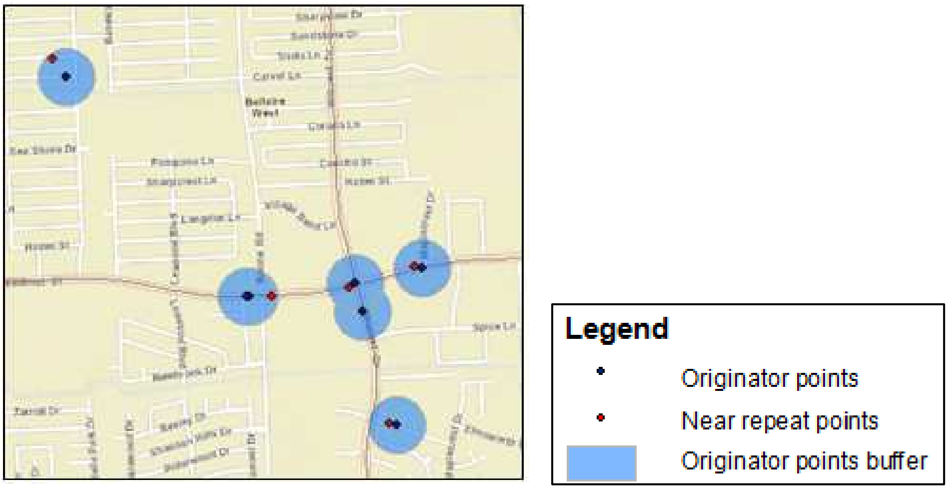


Figure 39: Detail results for burglaries for Houston for spring 2011

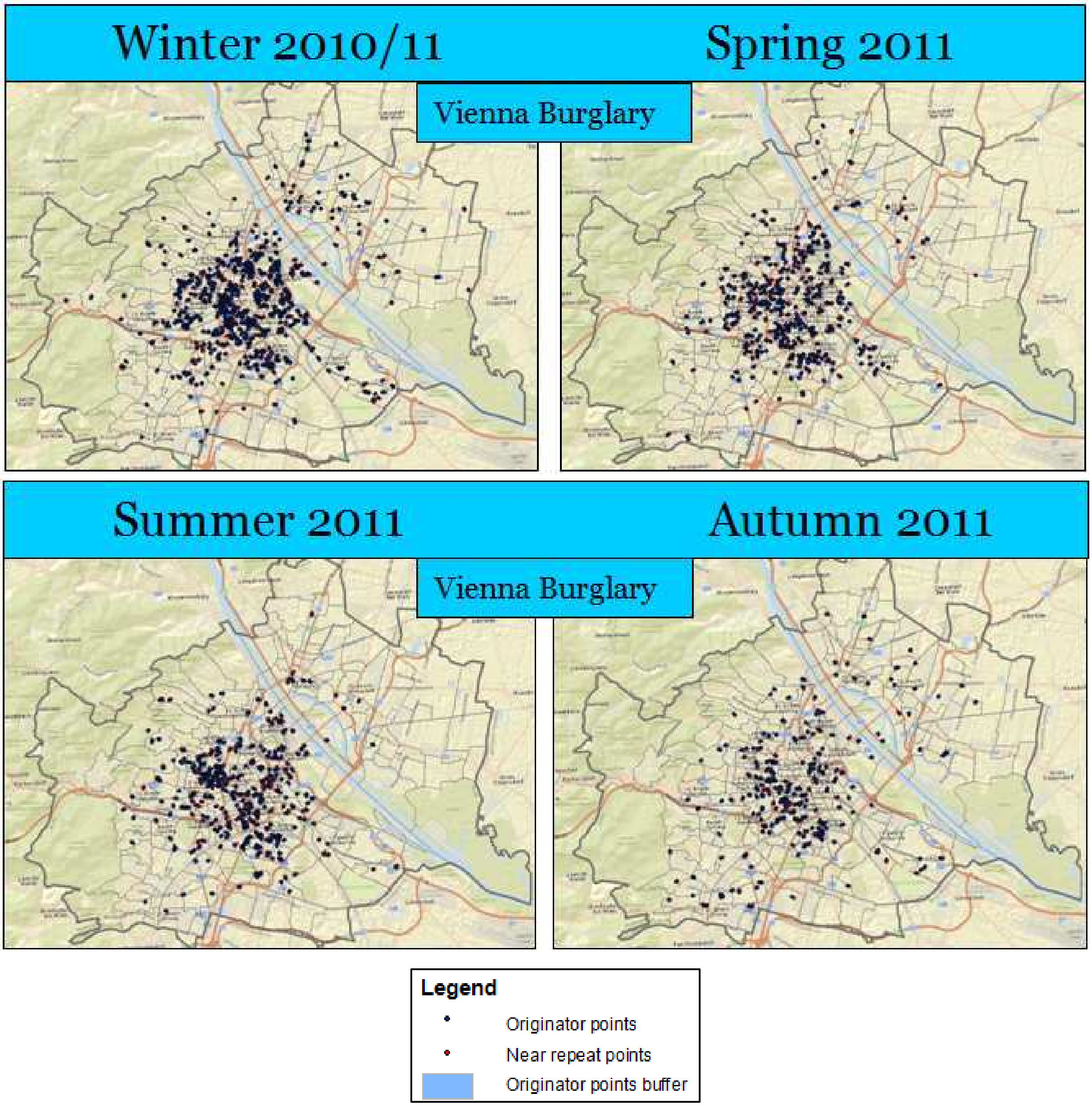


Figure 40: Results of the near – repeat analysis for the four seasons of 2011 for Vienna for burglaries

The results for the four seasons for Vienna show many cases of near repeats especially in the centre of the city. Also some points on the other side of the Danube in the district of Floridsdorf were found.

Quantity of originator points:			
Vienna:		Houston:	
Winter	1067	Winter	274
Spring	586	Spring	430
Summer	575	Summer	530
Autumn	387	Autumn	613

Table 15: Quantity of originator points for near repeats for burglaries for Vienna and Houston

Considering the quantity of the identified originator points (see table 15) it can be said, that there are more near repeats in Vienna than in Houston in proportion to the overall crime incidents, which are 1.4 times higher in Houston. Especially in winter there are nearly twice as many near repeat burglaries in Vienna than in the other seasons. Autumn shows the smallest amount of near repeat burglaries in Vienna. In comparison to that, Houston is completely different, with the winter season showing the smallest amount of originator points for burglaries.

5. Discussion

In this section a critical reflexion on the used methods of solution and results is done. The question is answered if the results and goals for this research have been accomplished.

5.1 Critical reflexion

In general it has to be said that it is difficult to compare two different cities with each other. Also it is possible that there are cities which can be compared better with each other than the two selected cities. But crime data of the two cities were made available, which allowed a comparison between the two cities. Both Vienna and Houston are big cities and they have a similar population size. Austrian and U.S. cities vary in many points (street network and structure of the city, hinterland of the city, etc.) which should be kept in mind and which can have a more or less big influence on the differences in the outcomes of this thesis. For this reason it was important that whenever possible, the same parameter settings for the same cluster methods were used for both cities. This allowed for a better comparison of the results of the cluster methods with each other. But due to the differences it was also hard to find parameter settings which fit both cities equally well.

5.2 Are the applied cluster and prediction evaluation methods appropriate?

Overall, the used methods delivered good and meaningful results. Nearly each result for different crime data on which the methods were performed showed the same tendency. The local Moran's I and the Gi* found many hot spot areas which cover a large proportion of the total study area. The kernel density estimation and the nearest neighbor hierarchical clustering predicted only a few hot spot areas and did not find a large number of incidents in this hot spot area. However, the hot spot areas are rather small compared to the study area (less than 1 %). So, overall, the kernel density estimation and the nearest neighbor hierarchical clustering are more accurate in predicting hot spots, because the covered hot spot areas are very small, which results in a rather low Prediction Accuracy Index. The results of the two methods (KDE and NNHC) are useful for the police to set further actions, because the predicted areas are very small and can be better controlled through patrols etc. It is thus easier to focus on them. The used methods appear useful for applying them to Austria too, not only in Houston. The results show that the KDE and the NNHC methods correctly predicted between 2.88 % and 11.97 % of all reported crimes for Vienna for 2012. In comparison to Vienna, the results for Houston showed that the same two methods correctly predicted between 0.83 % and 6.09 % for 2012. So, the results have a similar tendency for both cities. All results lead to the conclusion that an U.S. city and an Austrian city are comparable to a certain extent.

5.3 Have the expected results and goals of the thesis been reached?

Expectations and goals have been reached to a certain extent. It is difficult to compare two cities, and especially two cities of two different countries, because of so many influencing factors, which are also often unknown. Many factors have to be put into consideration. The different laws for crimes in the countries are just one important point, also the different behavior and attitude of people living there is an important part, as well. The different possibilities to commit crimes can also be different. For example, different security systems for shops or residential buildings, different means to attack somebody or defend yourself (for example firearms in U.S. are allowed, but not in Austria, some types of weapons are more common than others, etc.), different degree of police measures (type and extent of inspections of police, different prevention measures), different environmental structures of the cities, different infrastructure types, etc. When considering all of these factors that influence the results, then it must be assumed that the results show a tendency and should not be taken as a fixed and absolutely correct number. A tendency was found and this tendency shows that there exist similar results for both cities which lead to the conclusion that U.S. and Austria can be compared at a basic level. The near repeat analysis for burglary showed differences in the spatio-temporal results for Houston and Vienna. Different seasons are most critical for repeat burglary (winter for Houston and spring for Vienna). In general, the percentage that another incident happens at the same place in the near future is much higher in Vienna than it is in Houston. Near-repeat patterns were only found for autumn 2011 in

Houston and summer and autumn 2011 in Vienna. In Vienna twice as many crimes are detected as repeat or near repeat points than in Houston. This can be partially explained because of the smaller size of Vienna, where everything is more concentrated, including crime.

6. Conclusions and future work

This chapter summarizes the results and describes the conclusion of the results as well as the possibilities for future work.

6.1 Summary

In this Bachelor Thesis, two cities, namely Houston in Texas, U.S. and Vienna in Austria were compared by using four different crime prediction cluster methods and one spatio-temporal method to find out about crime hot spots and crime distribution. Goals were to find out which methods, including their implementation and evaluation, would be appropriate for applying to Austria. Results were shown and compared to find differences and similarities between the two cities and to find out if cities in the United States and in Austria can be compared, at all. This is important, because crime prediction methods are used in the U.S. already a lot, but in Austria their application is just at the beginning. Four cluster methods were chosen, including the kernel density estimation, the nearest neighbor hierarchical clustering, the local Moran’s I, and the Getis-Ord Gi*. The local Moran's I and the Gi* found many hot spot areas which cover a large proportion of the total study area. The kernel density estimation and the nearest neighbor hierarchical clustering predicted only a few hot spot areas and did not find many crime incidents in this area but the total area of all hot spots is very small compared to the entire study area. For this reason, the last two methods are more accurate in predicting hot spots in total because the covered areas are very small. The spatio-temporal method (near- repeat analysis) showed repeat patterns for all seasons in both cities. Near-repeat patterns were only found for autumn 2011 in Houston and in summer and in autumn 2011 in Vienna. In Vienna, twice as many incidents are detected as repeat or near repeat points than in Houston. This can be explained because of the smaller size of Vienna, where crime is more concentrated. Differences and similarities were found between the cities. Differences between Houston and Vienna can be found in the crime rates. In 2011, the burglary rate was 1.4 times higher in Houston than in Vienna, the robbery rate, 2.4 times higher, and the assault rate, 2.6 times higher. So, the crime rate is higher in Houston than in Vienna. All three selected crime types (burglary, assault, and robbery) are more concentrated in one area in Vienna, which corresponds to the centre of the city. In Houston, the selected crime types are more spread out over small areas in different parts of the city. The best methods for predicting crime incidents evaluated with the Prediction Accuracy Index were the kernel density estimation and the nearest neighbor hierarchical clustering. Similarities between Houston and Vienna are the similar results across the different cluster methods for both cities. Specifically, the LMI and the Gi* found many hot spot areas, which cover a large proportion of the entire study area. The KDE and the NNHC predict only a few hot spot areas which are very small.

6.2 Conclusions

The different sizes of the two study areas have not such a big impact on the results. When calculating the percentage of hot spot areas, the percentage values are similar for Houston and for Vienna. So, it makes sense to compare the two cities. The prediction methods are applicable to predict crime for Vienna. And the methods that predict smaller areas with a higher prediction accuracy should be used. These are the KDE and the NNHC.

In general, the results have a similar tendency for both cities. All results lead to the conclusion that an U.S. city and an Austrian city are comparable to a certain extent.

An adaption of the methods was necessary, since for the KDE meaningful parameter settings had to be found with which it was possible to visualize the results, which are not too general and not too detailed. In general, it was tried to use the same parameter settings of the used methods for both cities, so that the results of the cluster methods are comparable. Only for the NNHC, different parameter settings (minimum number of points and distance) had to be chosen because otherwise no clusters would have been found. The results were more detailed for Houston because the city size is larger. This resulted in hot spots for Houston, which are sometimes very small, but for comparison purposes, it was necessary to choose the same parameter setting. The size of Houston leads to these differences. The methods were chosen such that the results are more precise for Vienna than for Houston.

6.3 Future work

It was tried to apply many methods in this research. Other spatio-temporal cluster methods, like the Gi* or Kulldorff´s Scan statistics (SaTScan) could have also been tested and the results compared to the results of the near repeat calculator. Also risk-based spatial methods, like the risk-based nearest neighbor hierarchical clustering method, could have been applied and compared to the other spatial cluster methods.

The near repeat calculator could have also been used to find near repeats for 2012 for Houston and Vienna and then the results of 2011 and 2012 could have been compared. However, there was no time left to do this in this thesis. It would have also been interesting to search for near repeat patterns within the other used crime types robbery and assault. Unfortunately, there was also not enough time to do that in this thesis work.

References

ArcGIS 9.2 Desktop Help 2009. Hot Spot Analysis (Getis-Ord Gi*) (Spatial Statistics) [online]. Available from: http://webhelp.esri.com/ArcGISdesktop/9.2/index.cfm?TopicName=Hot_Spot_Analysis_(Getis-Ord_Gi*)_(Spatial_Statistics) [Accessed 12 May 2014].

ArcGIS 9.3 Tool Help 2009a. How Cluster and Outlier Analysis: Anselin Local Moran's I (Spatial Statistics) works [online]. Available from: http://resources.esri.com/help/9.3/ArcGISengine/java/gp_toolref/spatial_statistics_tools/how_cluster_and_outlier_analysis_colon_anselin_local_moran_s_i_spatial_statistics_works.htm [Accessed 12 May 2014].

ArcGIS 9.3 Tool Help 2009b. How Hot Spot Analysis: Getis-Ord Gi* (Spatial Statistics) works [online]. Available from: http://resources.esri.com/help/9.3/ArcGISdesktop/com/gp_toolref/spatial_statistics_tools/how_hot_spot_analysis_colon_getis_ord_gi_star_spatial_statistics_works.htm [Accessed 12 May 2014].

ArcGIS 9.3 Tool Help 2009c. Modeling spatial relationships [online]. Available from: http://resources.esri.com/help/9.3/ArcGISdesktop/com/gp_toolref/spatial_statistics_toolbox/modeling_spatial_relationships.htm [Accessed 12 May 2014].

ArcGIS Resources. What is ArcGIS? Available from: http://resources.ArcGIS.com/en/help/getting-started/articles/026n00000014000000.htm#ESRI_SECTION2_5421710764D94239AE362080003260AB [Accessed 12 May 2014].

Boba, R. 2008. Interpreting crime data and statistics, Chapter 8. In: Gwinn, S. L., Bruce, C., Cooper, J. P., and Hick S., second edition. Exploring Crime Analysis – Readings on essential skills. International Association of Crime Analysts. North Charleston, South Carolina: BookSurge, LLC. Available from: http://www.iaca.net/ExploringCA/2Ed/exploringca_chapter8.pdf [Accessed 12 May 2014].

Bruce, C. W., and Smith, S. C. 2011. Spatial statistics in crime analysis – using CrimeStat III. Edition 2.1, National Institute of Justice. Available from: http://www.icpsr.umich.edu/CrimeStat/workbook/CrimeStat_Workbook.pdf [Accessed 12 May 2014].

Caplan, J. M., Kennedy, L. W., and Piza, E. L. 2013. Joint utility of event-dependent and environmental crime analysis techniques for violent crime forecasting. Crime & Delinquency [online]. Available from: http://cad.sagepub.com.libezp.lib.lsu.edu/content/59/2/243.full.pdf+html

Via LSU account [Accessed 12 May 2014].

Chainey, S., and Ratcliffe, J., 2005. GIS and Crime Mapping. Chichester, England: John Wiley & Sons Ltd.

Chainey, S., Tompson, L., and Uhlig, S. 2008. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime [online]. Security Journal, 21. Available from: http://discovery.ucl.ac.uk/112873/1/PREPRINT_-Chainey%2C_Tompson_%26_Uhlig_2008.pdf [Accessed 12 May 2014].

City of Houston 2014. Download GIS datasets (COHGIS; February 2014Q1) [online]. Available from: http://mycity.houstontx.gov/home/cohgis.html [Accessed 12 May 2014].

Eck, J., Chainey, S., Cameron, J. G., Leitner, M., and Wilson, R. E. 2005. Mapping Crime: Understanding hot spots. U.S. Department of Justice, National Institute of Justice, Special Report, August 2005. Available from: https://www.ncjrs.gov/pdffiles1/nij/209393.pdf [Accessed 12 May 2014].

Everitt, B. 2011. Cluster Analysis 5th Edition. Available from: http://eds.a.ebscohost.com.libezp.lib.lsu.edu/eds/detail?vid=5&sid=5edd43d1-9a4b-4e81-8eaf-d656bc8a0fbe%40sessionmgr4001&hid=4203&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#db=nlebk&AN=354105 Via LSU account (Online Book) [Accessed 12 May 2014].

FBI 2004. UCR Uniform Crime Reporting Handbook. U.S. Department of Justice, Federal Bureau of Investigation. Available from: http://www.fbi.gov/about-us/cjis/ucr/additional-ucr-publications/ucr_handbook.pdf [Accessed 12 May 2014].

FBI 2012. Uniform Crime Report, FBI Releases 2012 Crime Statistics. Available from: http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/resource-pages/download-printable-files [Accessed 12 May 2014].

Filbert, K. 2008. Targeting Crime in Hot Spots and Hot Places, Geography & Public Safety – A quarterly bulletin of applied geography for the study of crime & public safety, Volume 1, Issue 1, February 2008. Available from: http://www.nij.gov/topics/technology/maps/gps-bulletin-v1i1.pdf [Accessed 12 May 2014].

Geography & Public Safety, 2008. New Jersey Report Highlights City's Crime Hot Spots. Geography & Public Safety Volume 1 Issue 1 [online]. Available from: http://www.nij.gov/topics/technology/maps/documents/gps-bulletin-v1i1.pdf [Accessed 12 May 2014].

Grubesic, T., and Murray, A. 2001. Detecting Hot Spots Using Cluster Analysis and GIS; Proceedings from the Fifth Annual International Crime Mapping Research Conference, Dallas, Texas. Available from: http://www.tonygrubesic.net/hot_spot.pdf [Accessed 12 May 2014].

Grubesic, T. H., 2006. On the application of fuzzy clustering for crime hot spot detection. Journal of Quantitative Criminology [online], Vol. 22, No. 1. Available from: http://eds.b.ebscohost.com.libezp.lib.lsu.edu/eds/pdfviewer/pdfviewer?sid=53daac07-911d-4a67-a15d-2a2f5ec3c761%40sessionmgr115&vid=7&hid=107 Via LSU Account [Accessed 12 May 2014]

Heraux, C. G. 2007. Software review: spatial data analysis of crime – a review of CrimeStat III. Social Science Computer Review [online]. Available from: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/60433/Heraux%20-%20CrimeStat%20III%20Review.pdf?sequence=1 [Accessed 12 May 2014].

Hinman, S. E., Blackburn, J. K., and Curtis, A. 2006. Spatial and temporal structure of typhoid outbreaks in Washington, D.C., 1906-1909: evaluating local clustering with the Gi* statistic. International Journal of Health Geographics [online]. Available from: http://www.ij-healthgeographics.com/content/5/1/13 [Accessed 12 May 2014].

Houston Police Department 2014. Robbery [online]. 2014. Available from: http://www.houstontx.gov/police/robbery/ [Accessed 12 May 2014].

Jersey City Police Department 2012, 2013. COMPSTAT [online]. Available from: http://www.njjcpd.org/node/5 [Accessed 12 May 2014].

Jusline 2014a. § 83 StGB Körperverletzung [online]. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=83&mvpa=92 [Accessed 12 May 2014].

Jusline 2014b. § 84 StGB Schwere Körperverletzung [online]. Verlag, Ort. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=84&mvpa=93 [Accessed 12 May 2014].

Jusline 2014c. § 85 StGB Körperverletzung mit schweren Dauerfolgen [online]. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=85&mvpa=94 [Accessed 12 May 2014].

Jusline 2014d. § 86 StGB Körperverletzung mit tödlichem Ausgang [online]. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=86&mvpa=95 [Accessed 12 May 2014].

Jusline 2014e. § 87 StGB Absichtliche schwere Körperverletzung [online]. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=87&mvpa=96 [Accessed 12 May 2014].

Jusline 2014f. § 88 StGB Fahrlässige Körperverletzung [online]. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=88&mvpa=97 [Accessed 12 May 2014].

Jusline 2014g. § 89 StGB Gefährdung der körperlichen Sicherheit [online]. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=89&mvpa=98 [Accessed 12 May 2014].

Jusline 2014h. § 109 StGB Hausfriedensbruch [online]. Verlag, Ort. Available from: http://www.jusline.at/109_Hausfriedensbruch_StGB.html [Accessed 12 May 2014].

Jusline 2014i. § 129 StGB Diebstahl durch Einbruch oder mit Waffen [online]. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=129 [Accessed 12 May 2014].

Jusline 2014j. § 142 StGB Raub [online]. Available from: http://www.jusline.at/142_Raub_StGB.html [Accessed 12 May 2014].

Jusline 2014k. § 143 StGB Schwerer Raub [online]. Available from: http://www.jusline.at/index.php?cpid=ba688068a8c8a95352ed951ddb88783e&lawid=11&paid=143&mvpa=160 [Accessed 12 May 2014].

Kampitsch, K., Leitner, M., and 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 12 May 2014].

Kennedy, L., Caplan, J., and Piza, E. 2011. Hotspot Mapping, Near Repeat Analysis, and Risk Terrain Modeling - Joint Operational Utility [online]. Available from:
http://www.ucl.ac.uk/secret/secret-student-seminars-11/hotspot-mapping
[Accessed 12 May 2014].

Land Wien 2012. publicdata.eu, ESRI Shapefile (WFS) – Land Wien - Landesgrenzen [online]. Available from: http://publicdata.eu/dataset/land-wien-landesgrenzen/resource/04c37591-bbc5-46e8-9994-9d0b07920976 [Accessed 12 May 2014].

Leitner, M., and Helbich, M. 2011. The impact of hurricanes on crime: a spatio-temporal analysis in the city of Houston, Texas. Cartography and Geographic Information Science [online], Vol. 38, No. 2, 2011, pp. 213-221. Available from:
http://www.geog.uni-heidelberg.de/md/chemgeo/geog/lehrstuehle/gis/leitner_helbich_2011.pdf
[Accessed 12 May 2014].

Levine, N. 2013a. CrimeStat IV, Part II: Spatial Distribution, Chapter 4. Available from:
https://www.ncjrs.gov/pdffiles1/nij/grants/242960-242995.pdf [Accessed 12 May 2014].

Levine, N. 2013b. CrimeStat IV, Part III: Hot Spot Analysis, Chapter 7. Available from:
https://www.ncjrs.gov/pdffiles1/nij/grants/242960-242995.pdf [Accessed 12 May 2014].

Levine, N. 2013c. CrimeStat IV, Part III: Hot Spot Analysis of Zones, Chapter 9. Available from:
https://www.ncjrs.gov/pdffiles1/nij/grants/242960-242995.pdf [Accessed 12 May 2014].

Levine, N. 2013d. CrimeStat IV, Part IV: Spatial Modeling I, Chapter 10. Available from:
https://www.ncjrs.gov/pdffiles1/nij/grants/242960-242995.pdf [Accessed 12 May 2014].

Levine, N. 2013e. National Institute of Justice - CrimeStat: Spatial Statistics Program for the Analysis of Crime Incident Locations. Available from: http://nij.gov/topics/technology/maps/Pages/crimestat.aspx [Accessed 12 May 2014].

Lukacsy, M., and Fendt, C. 2013. Wien in Zahlen 2013 [online]. Magistrat der Stadt Wien, Wograndl Druck, Mattersburg. Available from: http://www.wien.gv.at/statistik/pdf/wieninzahlen.pdf [Accessed 12 May 2014].

PREDPOL 2013a. Looking Ahead, Not in the Rear View Mirror [online]. Available from:
http://www.predpol.com/technology/ [Accessed 12 May 2014].

PREDPOL 2013b. Scientifically Proven Field Results [online]. Available from: http://www.predpol.com/results/ [Accessed 12 May 2014].

Quick, M., and Law, J. 2013. Exploring Hotspots of Drug Offences in Toronto: A Comparison of Four Local Spatial Cluster Detection Methods. Revue canadienne de criminology et de justice pénale [online]. Available from: http://eds.a.ebscohost.com.libezp.lib.lsu.edu/eds/pdfviewer/pdfviewer?sid=5edd43d1-9a4b-4e81-8eaf-d656bc8a0fbe%40sessionmgr4001&vid=8&hid=4203 Via LSU account [Accessed 12 May 2014].

Ratcliffe, J. H. 2008. Near Repeat Calculator – Program Manual for Version 1.3. Temple University, Philadelphia, PA and the National Institute of Justice, Washington, DC. Available from: http://www.temple.edu/cj/misc/nr/access.asp?ac=emsub [Accessed 12 May 2014].

Ratcliffe, J. H., and Rengert, G. F. 2008. Near-repeat patterns in Philadelphia shootings. Security Journal [online], 21, 58-76. Available from: http://www.palgrave-journals.com/sj/journal/v21/n1/abs/8350068a.html [Accessed 12 May 2014].

Risling, G. 2012. SFGate - Sci-fi policing: predicting crime before it occurs [online]. Available from: http://www.sfgate.com/default/article/Sci-fi-policing-predicting-crime-before-it-occurs-3725708.php [Accessed 12 May 2014].

Smith, S. C., and Bruce, C. W., 2008. CrimeStat III User Workbook. The National Institute of Justice. Washington, DC. Available from:
[Accessed 12 May 2014].

Statistik Austria 2013. Polizeiliche Kriminalstatistik von 1990 bis 2012 [online]. Available from: http://www.statistik.at/web_de/statistiken/soziales/kriminalitaet/anzeigen_polizeiliche_kriminalstatistik/index.html [Accessed 12 May 2014].

Townsley, M., Homel, R., and Chaseling, J. 2003. Infectious burglaries – A test of the near repeat hypothesis. British Journal of Criminology 43/3 [online]. Available from: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CCoQFjAA&url=http%3A%2F%2Fwww.researchgate.net%2Fpublication%2F29457289_Infectious_Burglaries_A_Test_of_the_Near_Repeat_Hypothesis%2Ffile%2F60b7d51eb901bb1af7.pdf&ei=RWdwU_2THMfioATbgoKICA&usg=AFQjCNGKhjveecvrFtjdwx2FjgbYAMkhfg&bvm=bv.66330100,d.cGU [Accessed 12 May 2014].

U. S. Census Bureau, Geography Division 2014. State & County QuickFacts - American National Standards Institute/Federal Information Processing Standards (ANSI/FIPS). Available from: http://quickfacts.census.gov/qfd/meta/long_fips.htm [Accessed 12 May 2014].

World Population Statistics 2013. Houston Population 2013 [online]. Available from: http://www.worldpopulationstatistics.com/houston-population-2013/ [Accessed 12 May 2014].

Youstin, T. J., Nobles, M. R., Ward, J. T., and Cook, C.L. 2011. Assessing the generalizability of the near repeat phenomenon. Criminal Justice and Behavior 2011 38 [online]. Available from: http://cjb.sagepub.com.libezp.lib.lsu.edu/content/38/10/1042.full.pdf+html [Accessed 12 May 2014].

List of Figures

List of Tables

List of Formulas