



# Master Thesis

im Rahmen des  
Universitätslehrganges „Geographical Information Science & Systems“  
(UNIGIS MSc) am Interfakultären Fachbereich für GeoInformatik (Z\_GIS)  
der Paris Lodron-Universität Salzburg

zum Thema

## „Prediction Accuracy of Hot Spot Methods “

vorgelegt von

**Dipl.-Ing. (FH) Achim Hettel**

U102610, UNIGIS MSc Jahrgang 2012

Zur Erlangung des Grades

„Master of Science (Geographical Information Science & Systems) – MSc (GIS) “

Gutachter:

Ao. Univ. Prof. Dr. Josef Strobl

Anchorage, 29.12.2014

"Everything is related to everything else, but near things are more related than distant things." (Tobler, 1970)

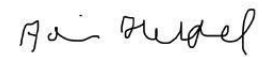
## ACKNOWLEDGEMENTS

I would like to thank the Police Department of Anchorage for providing me with the necessary crime data. I highly appreciated the supervision and advice by Prof. Michael Leitner with his knowledge covering this thesis topic.

Furthermore, I would like to thank the University of Salzburg UNIGIS team for providing an outstanding support throughout this master's program. Special thanks go to my wife for her patience and tolerance. Additionally, I would like to thank Ron Brown for his occasional distraction with a cup of coffee.

## DECLARATION OF ORIGINALITY

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.



Anchorage, 12/30/2014

Achim Hettel

## ZUSAMMENFASSUNG

### Vorhersagegenauigkeit von Hot-Spot Methoden

Hot-Spot Analysen von verschiedenen Kriminalitätsdelikten sind in den letzten Jahren bei Kriminalanalysten und Wissenschaftlern mehr in den Fokus gerückt. Besonders die Thematik, an welchen Standorten Kriminalität auftreten könnte, hat großes Interesse geweckt. Viele verschiedene Methoden sind entwickelt worden, um Kriminalitätsmuster zu identifizieren und um Wissenschaftler und Analysten bei der Vorhersage von Kriminalitätsereignissen, unter Einbeziehung von retrospektiven und zukünftigen Kriminalitätsdaten, zu unterstützen. Im Unterschied zu bisherigen Forschungen untersuchte diese Thesis unter anderem auch die dazugehörigen Täterwohnsitze.

Diese Thesis behandelte die Vorhersagegenauigkeit von verschiedenen Hot-Spot Methoden basierend auf einem Kriminalitätstyp und den dazugehörigen Täterwohnsitzen in Anchorage, USA. Das Hauptziel war, eine Hot-Spot Methode zu finden, die alle anderen Methoden unter Einbeziehung von drei verschiedenen Zeitreihen in ihrer Voraussagegenauigkeit übertrifft. Unter anderem wurden auch der Kriminalitätstyp und die Täterwohnsitze, bezüglich ihrer Vorhersagegenauigkeit miteinander verglichen und der Einfluss einer geringen Anzahl von Zukunftsdaten untersucht. Der erste Schritt beinhaltete die Untersuchung der globalen räumlichen Verbreitung für beide Datenmuster und für alle drei Zeitreihen. Um die Hot-Spots zu visualisieren, wurden neun verschiedene Methoden angewendet, räumlich analysiert und die Ergebnisse in die Vorhersagegenauigkeitsberechnung einbezogen. Die meisten vergangenen Untersuchungen mit ähnlicher Thematik kamen zum Ergebnis, dass die Kernel Density Estimation (KDE) Methode alle anderen Methoden übertraf. Die Ergebnisse dieser Thesis bestätigten, dass die KDE Methode die beste Vorhersagegenauigkeit bezüglich des Kriminalitätstyps besitzt. Die auf Zellenmatrix basierende  $G_i^*$  Methode zeigte jedoch bei den Täterwohnsitzen die beste Vorhersagegenauigkeit. Beim Vergleich des Kriminalitätstyps mit den Täterwohnsitzen erzielte der Kriminalitätstyp im Durchschnitt die besten Ergebnisse.

### *Zusammenfassung*

Die geringe Anzahl von zukünftigen Punktdaten bewirkte, dass einige Hot-Spot Methoden zur Berechnung der Vorhersagegenauigkeit nicht hinzugezogen werden konnten. Beim Vergleich der besten Vorhersagegenauigkeit aller Methoden, Zeitreihen und Datensätze, erzielte die KDE Methode zusammengefasst die besten Ergebnisse.

## ABSTRACT

### Prediction Accuracy of Hot Spot Methods

Crime hot spot mapping has gained increased interest during past several years among researchers and crime analysts. Especially the subject of forecasting where crime tends to occur has been focused on. Many different techniques have been developed to identify patterns of crime to support researchers and analysts to examine where crime may occur in the future by the use of retrospective crime patterns. So far, no research included offender residence patterns in its calculation.

This thesis focused on the prediction accuracy of different hot spot methods based on just one crime type and the related offender residences in the city of Anchorage, USA. The main goal was to find a cluster method which outperforms all other methods based on its prediction accuracy related to three different retrospective time periods. Additionally, the crime type and offender residence data were compared based on the outcome of the applied calculation and the influences of limited prospective point numbers were examined.

The first step in this research included the examination of the global distribution of both data patterns from all time periods. Nine different cluster techniques were applied to visually detect hot spots of crime events and offender residences. The related hot spots were then spatially analyzed and the results included into three prediction accuracy calculations.

Most previously conducted research on a similar topic concluded that the kernel density estimation method (KDE) has the best overall prediction accuracy performance. The results of this thesis confirmed that the KDE method outperforms all other methods in most of the different time periods containing the crime type. However, the grid based  $G_i^*$  method showed the best prediction accuracy performances related to the offender residence data pattern.

Comparing the two data sets, the crime event data set showed the highest prediction accuracy indexes on average. Another finding was that the influence of limited numbers of prospective data was causing several hot spot methods to fail calculating the prediction indexes. By comparing all best indexes throughout the time periods and data sets, the KDE method exceeded all other methods.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	II
DECLARATION OF ORIGINALITY .....	III
ZUSAMMENFASSUNG .....	IV
ABSTRACT.....	VI
TABLE OF CONTENTS.....	VII
LIST OF FIGURES .....	IX
LIST OF TABLES.....	XII
LIST OF ACRONYMS & ABBREVIATIONS.....	XIV
CHAPTER 1 INTRODUCTION .....	1
1.1 Motivation and Background.....	1
1.2 Research Statement & Objectives .....	2
1.3 Expected Results .....	3
1.4 Topics Not Covered .....	3
1.5 Target Groups.....	3
1.6 Thesis Structure.....	3
CHAPTER 2 LITERATURE REVIEW .....	5
CHAPTER 3 STUDY AREA AND DATA.....	7
3.1 Study Area.....	7
3.2 Data .....	8
CHAPTER 4 METHODOLOGY .....	12
4.1 Descriptive Spatial Statistics.....	12
4.2 Hot Spot Methods and Analysis.....	14
4.2.1 Parameter Settings and Classification Methods.....	14
4.2.2 Point Pattern Hot Spot Methods .....	17



4.2.3	Aggregated Hot Spot Methods.....	23
4.3	Prediction Accuracy Analysis .....	31
CHAPTER 5	RESULTS.....	33
5.1	Descriptive Spatial Statistics.....	33
5.2	Visual Results of Hot Spot Methods .....	35
5.2.1	Point pattern methods .....	35
5.2.2	Aggregated hot spot methods.....	52
5.3	Comparison of Prediction Accuracy .....	72
5.4	Summary of Prediction Accuracy .....	79
CHAPTER 6	DISCUSSION .....	83
6.1	Hot Spot Mapping .....	83
6.2	Prediction Accuracy .....	85
6.3	Recommendation.....	87
CHAPTER 7	CONCLUSION .....	88
LIST OF REFERENCES	.....	90

## LIST OF FIGURES

Figure 3.1: Study area city of Anchorage by census block groups.....	7
Figure 3.2: Spatial Join process, Source: ESRI online .....	9
Figure 3.3: Property theft from cars 2008 - 2013 .....	10
Figure 3.4: Offender residences 2008 - 2013.....	11
Figure 4.1: Nearest Neighbor Index; Source: C. Brown, online.....	13
Figure 4.2: Calculating the KDE of a point pattern, Source: (Gatrell et al., 1996) .....	22
Figure 4.3a: Thematic mapping census block groups.....	24
Figure 4.3b: Thematic Mapping census tracts .....	24
Figure 4.4: GTM table using 1,000 feet grid cell size .....	26
Figure 4.5: GTM with a 1,000 feet grid overlay .....	26
Figure 4.6: Deriving statistically significant hot-, cold-spots and spatial outliers using the Local Moran’s I statistic, Source: (ESRI, 2014).....	28
Figure 4.7: GI* matrix showing cells with values, Source: (Chainey and Ratcliffe, 2005) .....	30
Figure 5.1: Visual output of the Standard Deviational Ellipse of property theft from cars from 2008 - 2010 .....	34
Figure 5.2: Visual output of the Standard Deviational Ellipse of offender residences from 2008 - 2010 .....	34
Figure 5.3: Resulting table of Spatial Fuzzy Mode .....	35
Figure 5.4: Spatial Fuzzy Mode with multi rank output.....	35
Figure 5.5: Hot spot of property theft from cars from 2008 – 2010 using the Spatial Fuzzy Mode .....	36
Figure 5.6: Hot spot of property theft from cars from 2008 – 2012 using the Spatial Fuzzy Mode .....	36
Figure 5.7: Hot spot of property theft from cars from 2008 – June 2013 using the Spatial Fuzzy Mode .....	37
Figure 5.8: Hot spot of offender residences from 2008 – 2010 using the Spatial Fuzzy Mode .....	38
Figure 5.9: Hot spot of offender residences from 2008 – 2012 using the Spatial Fuzzy Mode .....	38

Figure 5.10: Hot spot of offender residences from 2008 – June 2013 using the Spatial Fuzzy Mode .....	39
Figure 5.11: STAC of property theft from cars from 2008 – 2010.....	40
Figure 5.12: STAC of property theft from cars from 2008 – 2012.....	40
Figure 5.13: STAC of property theft from cars from 2008 – June 2013 .....	41
Figure 5.14: STAC of offender residences from 2008 – 2010 .....	42
Figure 5.15: STAC of offender residences from 2008 – 2012 .....	42
Figure 5.16: STAC of offender residences from 2008 – June 2013 .....	43
Figure 5.17: NNH of property theft from cars from 2008 – 2010 .....	44
Figure 5.18: NNH of property theft from cars from 2008 – 2012 .....	44
Figure 5.19: NNH of property theft from cars from 2008 – June 2013.....	45
Figure 5.20: NNH of offender residences from 2008 – 2010 .....	46
Figure 5.21: NNH of offender residences from 2008 – 2012 .....	46
Figure 5.22: NNH of offender residences from 2008 – June 2013.....	47
Figure 5.23: KDE of property theft from cars from 2008 – 2010.....	48
Figure 5.24: KDE of property theft from cars from 2008 – 2012.....	48
Figure 5.25: KDE of property theft from cars from 2008 – June 2013 .....	49
Figure 5.26: KDE of offender residences from 2008 – 2010 .....	50
Figure 5.27: KDE of offender residences from 2008 – 2012 .....	50
Figure 5.28: KDE of offender residences from 2008 – June 2013 .....	51
Figure 5.29: GBTM of property theft from cars from 2008 – 2010 .....	52
Figure 5.30: GBTM of property theft from cars from 2008 – 2012 .....	52
Figure 5.31: GBTM of property theft from cars from 2008 – June 2013.....	53
Figure 5.32: GBTM of offender residences from 2008 – 2010 .....	54
Figure 5.33: GBTM of offender residences from 2008 – 2012 .....	54
Figure 5.34: GBTM of offender residences from 2008 – June 2013.....	55
Figure 5.35: GTM of property theft from cars from 2008 – 2010.....	56
Figure 5.36: GTM of property theft from cars from 2008 – 2012.....	56
Figure 5.37: GTM of property theft from cars from 2008 – June 2013.....	57
Figure 5.38: GTM of offender residences from 2008 – 2010.....	58
Figure 5.39: GTM of offender residences from 2008 – 2012.....	58
Figure 5.40: GTM of offender residences from 2008 – June 2013 .....	59
Figure 5.41: Local Moran’s I of property theft from cars from 2008 – 2010.....	60

Figure 5.42: Local Moran's I of property theft from cars from 2008 – 2012.....	60
Figure 5.43: Local Moran's I of property theft from cars from 2008 – June 2013 .....	61
Figure 5.44: Local Moran's I of offender residences from 2008 – 2010.....	62
Figure 5.45: Local Moran's I of offender residences from 2008 – 2012.....	62
Figure 5.46: Local Moran's I of offender residences from 2008 – June 2013 .....	63
Figure 5.47: Gi* census block groups of property theft from cars from 2008 – 2010 .....	64
Figure 5.48: Gi* census block groups of property theft from cars from 2008 – 2012 .....	64
Figure 5.49: Gi* census block groups of property theft from cars from 2008 – June 2013 .....	65
Figure 5.50: Gi* census block groups of offender residences from 2008 – 2010 .....	66
Figure 5.51: Gi* census block groups of offender residences from 2008 – 2012 .....	66
Figure 5.52: Gi* census block groups of offender residences from 2008 – June 2013 .....	67
Figure 5.53: Gi* grid of property theft from cars from 2008 – 2010 .....	68
Figure 5.54: Gi* grid of property theft from cars from 2008 – 2012 .....	68
Figure 5.55: Gi* grid of property theft from cars from 2008 – June 2013 .....	69
Figure 5.56: Gi* grid of offender residences from 2008 – 2010 .....	70
Figure 5.57: Gi* grid of offender residences from 2008 – 2012 .....	70
Figure 5.58: Gi* grid of offender residences from 2008 – June 2013 .....	71

## LIST OF TABLES

Table 3.1: Retrospective point count by time periods .....	8
Table 3.2: Prospective point count by time periods.....	8
Table 4.1: Input data type of point pattern methods and their parameter .....	15
Table 4.2: Parameter settings for the Spatial Fuzzy Mode .....	18
Table 4.3: Parameter settings for the STAC method .....	19
Table 4.4: Parameter settings for the NNH method.....	21
Table 4.5: Parameter settings for the KDE method .....	23
Table 4.6: Parameter settings for the GBTM method.....	25
Table 4.7: Parameter settings for the GTM method .....	27
Table 4.8: Parameter settings for the Local Moran’s I method .....	29
Table 4.9: Parameter settings for the Gi* census block group / Gi* grid methods .....	31
Table 5.1: Results of the NNI and the SDD of property thefts from cars and offender ... residences for different time periods.....	33
Table 5.2: Evaluation of prediction accuracy for property thefts from cars from 2008 - 2010.....	72
Table 5.3: Evaluation of prediction accuracy for property thefts from cars from 2008 – 2012.....	73
Table 5.4: Evaluation of prediction accuracy for property thefts from cars from 2008 – June 2013 .....	74
Table 5.5: Evaluation of prediction accuracy for offender residences from 2008 – 2010	76
Table 5.6: Evaluation of prediction accuracy for offender residences from 2008 - 2012	77
Table 5.7: Evaluation of prediction accuracy for offender residences from 2008 – June 2013.....	78
Table 5.8: Highest Hit Rate for property thefts from cars and their offender residence locations .....	79
Table 5.9: Highest PAI for property thefts from cars and their offender residence locations .....	80
Table 5.10: Highest RRI for property thefts from cars and their offender residence locations .....	81
Table 5.11: Highest prediction accuracies on average for property theft from cars and .. offender residence locations .....	81

Table 5.12: Summary of failing methods ..... 82

## LIST OF ACRONYMS & ABBREVIATIONS

Dec.	December
GBTM	Geographic Boundary Thematic Mapping
GIS	Geographic Information System
GTM	Grid Thematic Mapping
HR	Hit Rate
Jan.	January
KDE	Kernel Density Estimation
NNH	Nearest Neighbor Hierarchical
PAI	Prediction Accuracy Index
RRI	Recapture Rate Index
SD	Standard Deviation
SDD	Standard Distance Deviation
STAC	Spatial and Temporal Analysis of Crime

## CHAPTER 1      INRODUCTION

### 1.1    Motivation and Background

It can be agreed that most people were the victim of a crime like property theft at some point in their life. When crime happens, it occurs in places with a geographical connection. This is an important part in analyzing crime problems because crime has an inherent geographical quality (Chainey and Ratcliffe, 2005). Until the late 1970s, the scientific research of crime was mainly in the field of sociology and psychology (Georges, 1978). Even in the early days, police recognized the importance of the geographical aspect by sticking pins into maps to show crime locations.

With technology advancing, the understanding and the rising opportunities opened the door for new techniques like crime pattern identification, examination of relationships between crime and socio-economic and environmental aspects etc.

As Eck et al. (2005) stated, crime is not evenly distributed. There are areas with a higher concentration of crimes and others with no or little crime. These concentrations are caused by offender opportunities and the interaction of offender and victims (Cohen and Felson 1979; Cornish and Clarke 1986). Even when looking at daily life, people tend to avoid places where crime might interfere with their activities. Some people are for example driving another route to their destination in order to avoid high crime areas. Others choose their community, schools, recreation areas and stores based on their experience of high crime areas or neighborhoods with less socio-economical values.

Police and law enforcement are using this knowledge in their daily activities as well. They organize their routines like patrolling a specific area based on high or low crime occurrences. For problem-oriented policing it is also important to understand what causes high crime areas and to respond with crime reduction measures. Boba (2005) categorizes several types of crime analysis in conjunction with crime mapping which are administrative (ACA), tactical (TCA) and strategic (SCA) crime analysis. TCA involves the study of short-term (less than 6 months) criminal activities and potential events to identify trends and patterns. Long-term (more than 6 months) analysis of crime clusters and crime trend forecasting are part of the SCA. SCA is defined as the study of crime including the analysis of long-term patterns of specific crime activities (Boba, 2005).



Concentrations or clusters of high crime events are mainly referred to as hot spots.

Many different techniques were developed to identify patterns of crime such as spatial ellipse, thematic mapping of geographic boundary, grid thematic mapping, hierarchical clustering, continuous surface smoothing and local indicators of spatial association.

These methods help researchers and analysts to examine where crime may occur in the future by the use of retrospective crime patterns to support decision makers to organize and execute proactive approaches of crime prevention.

To calculate the prediction accuracy of different hot spot techniques, three measures were developed by different researchers. The Hit Rate, Prediction Accuracy Index (PAI) (Chainey et al., 2008) and the Recapture Rate Index (RRI) (Levin, 2008) were introduced. This can be accomplished by splitting the data into different time periods. The first part of the time span is used as retrospective data and the second part is used as prospective or “future” data for comparison. The different prediction accuracy calculations are distinct from each other by including different variables such as the area size or the ratio of crime events within hot spots.

The results are used for the comparison of different hot spot techniques based on their prediction accuracy capability. The majority of prediction accuracy research and practices are focused on different crime types as data sources. To my best knowledge, no research to date has focused on analyzing spatial patterns of offender residences based on different hot spot methods and their capability of prediction accuracy. Most of the research regarding prediction accuracy of hot spot methods includes counts of at least hundreds of events from the prospective data set. This thesis experimented with very small numbers of prospective crime events and offender locations by splitting the data into three different time periods for the retrospective and prospective part of the prediction accuracy calculation for comparison. By increasing the time span of the retrospective point data, the point counts of the “future” data will be automatically less. This can also affect the prediction accuracy and comparison of different hot spot methods.

## 1.2 Research Statement & Objectives

The main objective of this thesis is to analyze the prediction accuracy of several common hot spot methods based on different time periods of point data related to

property theft from cars and offender residences. Therefore this research includes two questions:

- What is the best method to predict where future crimes and offender residences may occur?
- What point data set has the highest prediction accuracy on average?

### 1.3 Expected Results

Research already exists that has examined the prediction accuracy of different cluster analysis methods. It is expected that at least one hot spot method will outperform all other methods based on the overall prediction accuracy index results. The quality of prediction accuracy will be different between the point data containing property theft from cars and offender residence. Additionally, some methods and prediction calculations will be affected by the limited point counts of the prospective data.

### 1.4 Topics Not Covered

Most research that has covered similar topics in the past includes multiple crime types and temporal aspects such as date and time. This thesis used only one crime type and one location type and did not analyze temporal patterns. Furthermore, not all cluster methods were covered, only the most common ones that have been used previously by other researchers.

### 1.5 Target Groups

The findings of this thesis could be useful to crime analysts and decision makers in police and law enforcement agencies as well as researchers in the academic field.

### 1.6 Thesis Structure

This research is structured into seven chapters. Chapter 1 includes the introduction. Chapter 2 covers the literature review and theories behind different hot spot techniques applied by researchers and analysts. The study area and the data used for this research are outlined in Chapter 3. Chapter 4 introduces the examination of the global distribution of

the data set. Following sections contain the analysis of different hot spot methods, their parameter settings and the introduction of three different prediction accuracy calculations. Chapter 5 presents the visual and prediction accuracy results of the analysis. Limitations of the applied methods, the interpretation of the results, and further recommendations are discussed in Chapter 6. At the end, a conclusion of the research findings is presented in Chapter 7.

## CHAPTER 2 LITERATURE REVIEW

The relationship between crime and geography was first researched and took notable attention in the 19<sup>th</sup> century by two French researchers who identified the relation between property crimes and their locations as well as the temporal connection to the crime (Ratcliffe, 2010). Since this early research was only based on large scale areas, Shaw and McKay (1942) examined the rates of juvenile delinquents in Chicago which were manually aggregated into smaller areas specified by community aspects.

Thanks to the advanced development of computer technology beginning with the last quarter of the 20<sup>th</sup> century, the majority of manual processing of crime data was replaced by complex computer programs to open the door for the development of new techniques to analyze the distribution of crimes.

Most law enforcement agencies are using computer programs like Geographic Information Systems to examine crime patterns and to develop preventive measures. Since the distribution of crimes is not random, many techniques were developed to analyze clusters of crime events, so called hot spots. Eck et al. (2005) categorizes these techniques in global statistical tests, hot spot mapping and Local Indicators of Spatial Association (LISA) (Anselin, 1995).

In 1996, the Illinois Criminal Justice Information Authority developed and published one of the first software suite called Spatial and Temporal Analysis of Crimes (STAC) to analyze patterns of crime. One of the simplest hot spot mapping techniques is thematic mapping of a geographic boundary (Chainey and Ratcliffe, 2005) where crime events are aggregated to administrative or statistical boundaries like census tracts to represent the thematic range of crime counts. The resulting map type is called “choropleth map” (Imhof, 1972) . Using this map type, only crime rates should be mapped, but not absolute numbers of crime (i.e., crime counts), unless the statistical boundaries consist of cells of a regular grid (cells are the same size). The disadvantage of this method is that administrative areas have different shapes and sizes and this leads to different results due to the spatial distribution of the underlying crime data. This problem is known as the Modifiable Area Unit Problem (MAUP) described by Openshaw (1984).

To overcome this problem is to generate uniform grids over the study area. Each grid cell can have crime counts or rates aggregated to each cell and can be thematically represented which is known as grid thematic mapping. Another preferred grid based

method is Kernel Density Estimation (KDE). This method creates a smooth continuous surface over the study area to represent crime density (Eck et al., 2005; Chainey and Ratcliffe, 2005; Chainey et al., 2008). These two methods can be used as hot spot methods, since the cells with the highest crime counts or rates (grid thematic mapping) and crime densities (KDE) can be defined as hot spots.

One of the oldest clustering methods is the Nearest Neighbor Hierarchical Clustering technique (King, 1967; Johnson, 1967). It identifies only points that are closer to each other than expected under spatial randomness (Eck et al., 2005) and groups these points based on their minimum number within a cluster which has to be defined by a user.

Most hot spot methods are focusing on high concentration of crimes relative to the study area but it cannot be ignored that the underlying population is influencing the distribution of most crimes too (Chainey and Ratcliffe, 2005) and should be included into common hot spot methods as crime rates instead of crime counts.

Since high crime areas can be small compared to the whole study area but large compared to its neighbors, Local Indicators of Spatial Association (LISA) were developed to examine the local associations between crime events (Anselin, 1995).

All reviewed hot spot methods have more or less the capability to predict where crime may occur in the future. To identify which method has the highest prediction accuracy, Chainey et al. (2008) introduced the Prediction Accuracy Index (PAI) to compare different hot spot techniques based on their prediction capability including the size of the hot spots. As a complement to the PAI, Levin (2008) developed the Recapture Rate Index (RRI) by calculating the ratio of total crime counts and crime counts within hot spots based on future and past events without taking the size of hot spots into consideration.

## CHAPTER 3 STUDY AREA AND DATA

### 3.1 Study Area

The area under study covers the city of Anchorage as a part of the Municipality of Anchorage. It is located in south-central Alaska at 61° north and 149° west. The total population of the municipality is based on the census data of the US Census Bureau (<http://quickfacts.census.gov/qfd/states/02/02020.html>) and it includes approximately 300,950 residents which represent around 41% of the states total population. Furthermore, the US Census Bureau listed the land area with 4,415 square kilometers and a population density of around 176 persons per square mile. Since the municipality has such a low population density, the study area in this research comprises only the city of Anchorage (see Figure 3.1) without the neighboring Towns of Eagle River and Girdwood

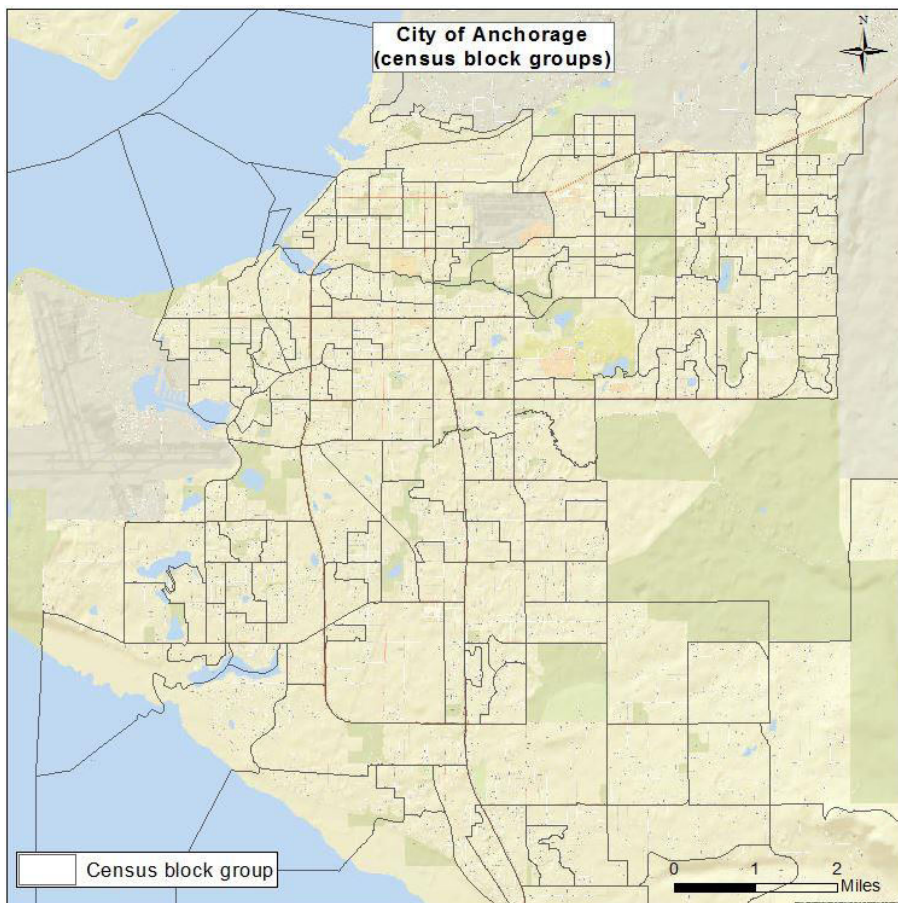


Figure 3.1: Study area city of Anchorage by census block groups

### 3.2 Data

The data were provided in an Excel format from the Anchorage Police Department (APD) based on an open record request. The file included property theft from cars and the residences of the arrested persons related to these crimes from January 2008 until December 2013.

Both sets of data were partially cleaned and organized and included the report number, report date, incident address, geographic location of the crime event, crime type, description of the stolen property, offender address and the coordinates of the address.

After cleaning the data, the date and the geographic coordinates from the crime events and offender residences were separated in three different retrospective (past) and prospective (future) time periods (Tables 3.1 and 3.2). The data were then imported into an ArcGIS software project and converted into shape files.

Event Type	Retrospective Point Count by Time Period		
	Jan. 2008–Dec. 2010	Jan. 2008–Dec. 2012	Jan. 2008–June 2013
Property Theft from Cars	154	253	282
Offender Residences	150	250	272

Table 3.1: Retrospective point count by time periods

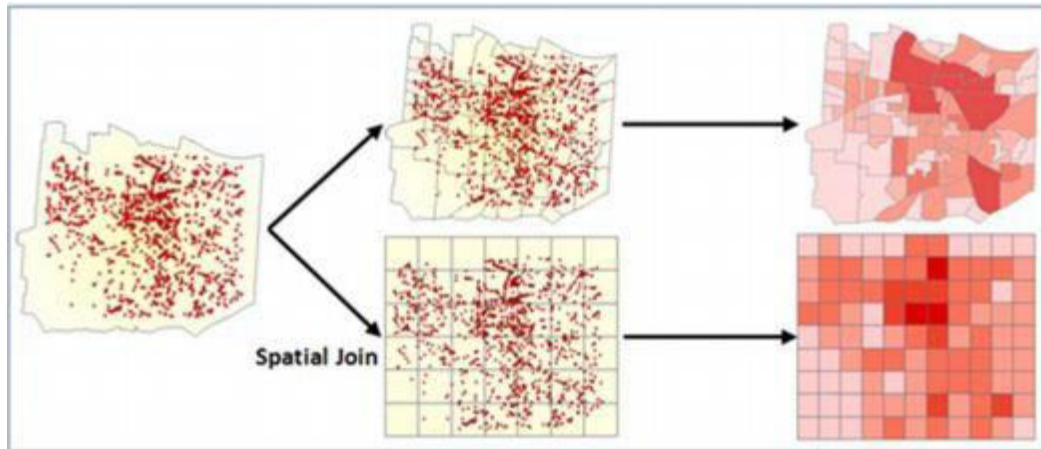
Event Type	Prospective Point Count by Time Period		
	Jan. 2011–Dec. 2013	Jan. 2013–Dec. 2013	July 2013–Dec. 2013
Property Theft from Cars	150	51	22
Offender Residences	151	50	23

Table 3.2: Prospective point count by time periods

This research also analyzed data aggregated to administrative boundaries. Thus, census block groups were downloaded from the Anchorage Information Technology Department web site (<http://munimaps.muni.org/moagis/download.htm>) and converted in a GIS shape

file format. The data contained demographic statistics for each census block group such as total population, population by different races, and the size of each census block group.

The crime and offender residence data were then aggregated to the census data as count values (Figure 3.2).



*Figure 3.2: Spatial Join process, Source: ESRI online*

The aggregated point count of both data has been calculated as crime rates based on the total population of each block group. Furthermore, additional spatial analysis processes like spatial selection and geo-processing, such as data projecting, were executed.

The following maps (Figures 3.3 and 3.4) show the point distribution of property theft from cars and offender residences from January 2008 until December 2013. These representations contain a Google Maps base layer and census block groups as a reference layer. Point maps are the simplest but also the most unreliable presentation of point distributions. Depending on the number of points, the spatial details of high concentration of point events are not clearly obvious.



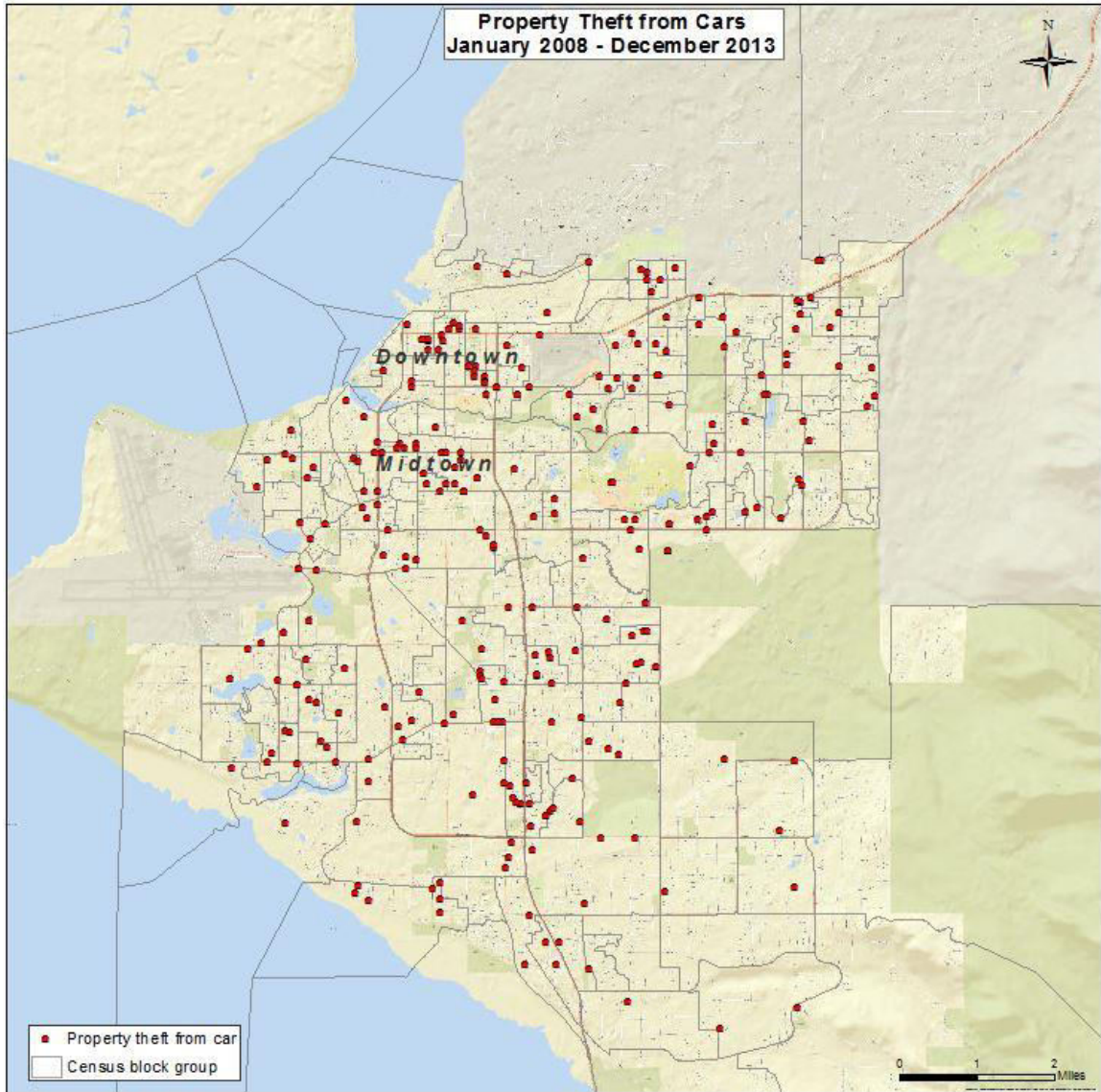


Figure 3.3: Property theft from cars 2008 - 2013

It can be seen that in Figure 3.3 there are some spatial concentrations of crime events. The down-town area from Anchorage shows strong evidences of clusters. The mid-town, northern and southern parts contain to some extent higher concentration of crimes.

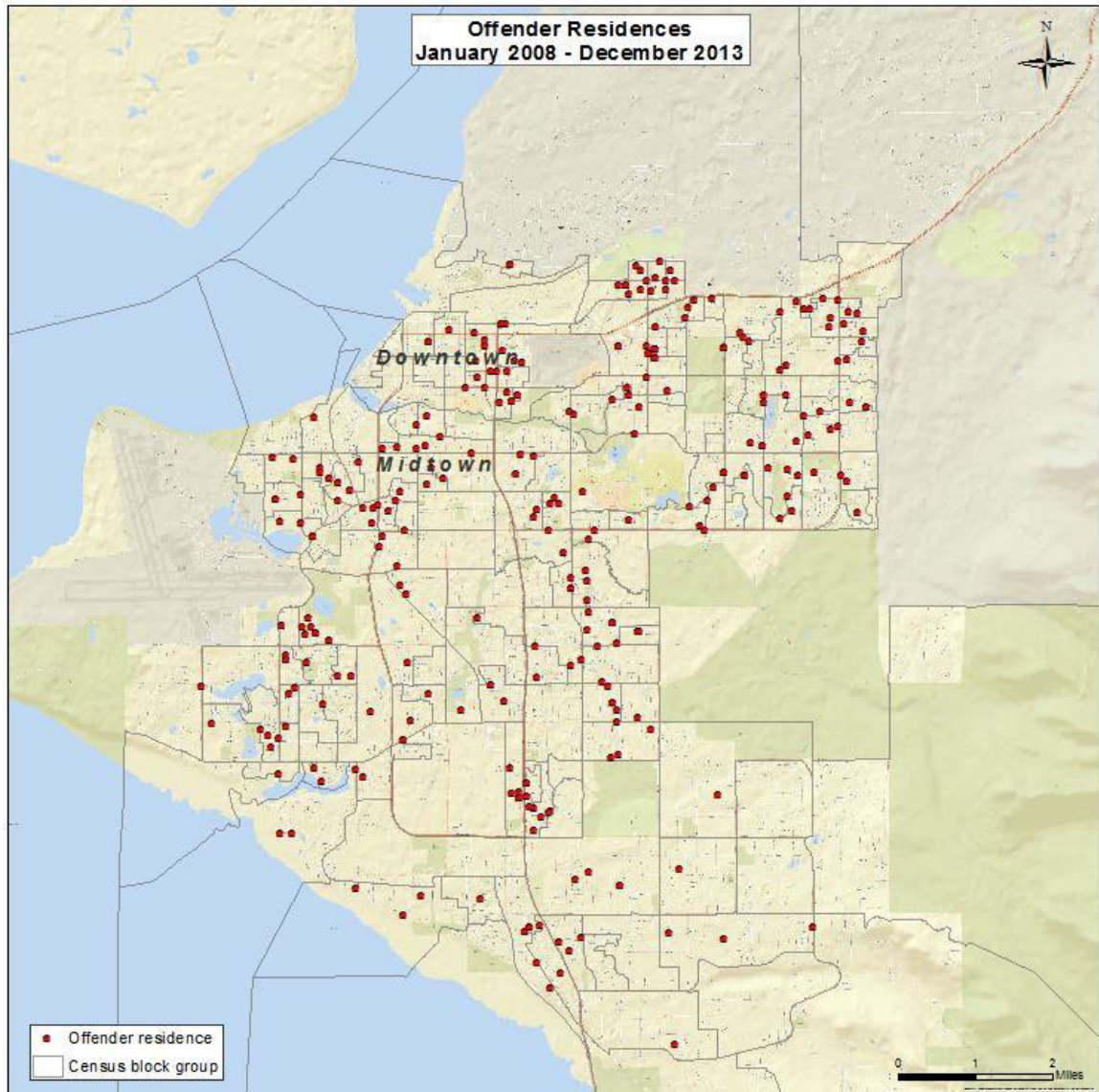


Figure 3.4: Offender residences 2008 - 2013

Comparing Figure 3.4 with Figure 3.3, the point distribution of offender residence locations shows a more clustered pattern. Smaller clusters are visible in the northern, eastern, western, and southern parts of Anchorage.

## CHAPTER 4      METHODOLOGY

### 4.1    Descriptive Spatial Statistics

Before the proposed cluster techniques were executed and analyzed it is recommended to determine the general pattern of the data based on global statistical tests. The Nearest Neighbor Index (Han and Gorman, 2013) is one of these methods to analyze the evidence of general clustering in the data (Eck et al., 2005) by comparing the distribution of the crime and the offender residence data against randomly distributed data with the same amount of points and located in the same study area. First, the NNI calculates the distance from each point to its nearest neighbor and summarizes these distances for all points and dividing them by the total point counts of the data. The same step as above is then executed with randomly distributed points. The final result is the ratio of the average nearest neighbor distance against the average random nearest neighbor distance. If a NNI result is less than 1 the data are showing a tendency towards clustered patterns. With a NNI of about 1, there is an evidence of randomly distributed data and a NNI of greater 1 is moving towards a uniform pattern in the point data (Figure 4.1).

The NNI is defined as:

$$NNI = \frac{d_{NN}}{d_{NN(ran)}} \quad (4-1)$$

where  $d_{NN}$  is the average distance of each point to its closest neighboring point and  $d_{NN(ran)}$  is the average distance of each point to its closest neighboring point in a complete spatial random point pattern.

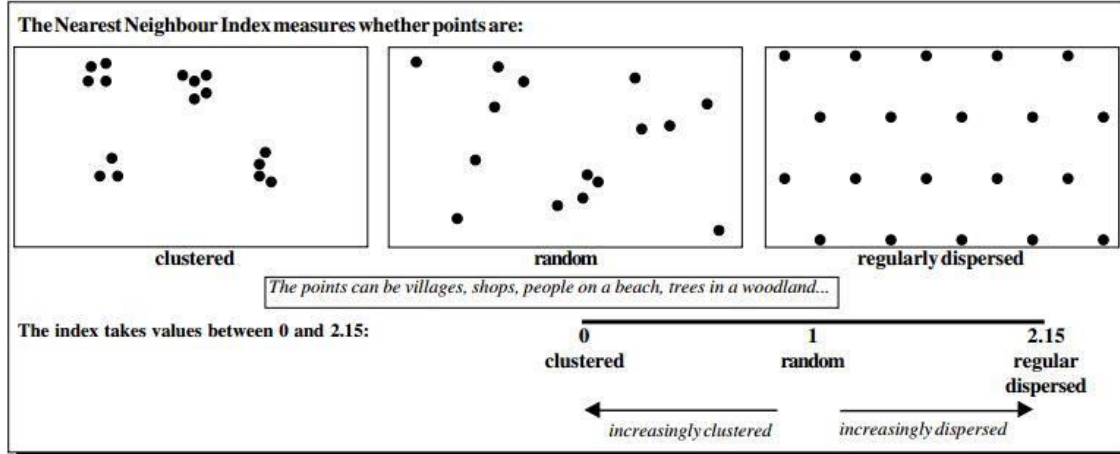


Figure 4.1: Nearest Neighbor Index; Source: C. Brown, online

The next global test included the calculation of the Standard Distance Deviation (SDD) that describes the dispersion of the data compared to different time lines and data sets.

The standard distance deviation (SDD) is defined as:

$$SDD = \sqrt{\sum_{i=1}^N \frac{(d_{iMC})^2}{N-2}} \quad (4-2)$$

where  $N$  is the total point counts and  $d_{iMC}$  defines the distance between each point ( $i$ ) and the mean center(MC).To create an unbiased estimate, 2 is subtracted from the total point counts.

The final global test included a visual output based on a standard deviational ellipse (Ebdon, 1988; Cromley, 1992) method and describes the level and the orientation of the data dispersion including the overall size.

The standard deviations for the x-axis and y-axis are defined as:

$$S_x = \sqrt{\frac{\sum[(x_i - \bar{X}) \cos \theta - (y_i - \bar{Y}) \sin \theta]^2}{N-2}} \quad (4-3)$$

$$S_y = \sqrt{\frac{\sum[(x_i - \bar{X}) \sin \theta + (y_i - \bar{Y}) \cos \theta]^2}{N-2}} \quad (4-4)$$

where  $\bar{X}$  and  $\bar{Y}$  are the means of X and Y respectively,  $\theta$  is the angle (in radians), and  $N$  is the number of points. To create an unbiased estimate, 2 is subtracted from the total point counts (Levin, 2013).

The results of the global spatial statistical tests can be found in Chapter 5.

## 4.2 Hot Spot Methods and Analysis

Since crime events are not evenly distributed (Eck et al., 2005), researchers and crime analysts are using different methods to detect and examine high concentrations of crime that are mainly referred to as hot spots.

This section is organized in four parts. The first part analyzes different parameter settings and the global distribution of the data sets. The second part examines four point pattern hot spot methods. The third part introduces five aggregated hot spot methods including their limitations, visual mapping results and the specific parameter settings for each hot spot method. Finally, the Hit Rate, the Prediction Accuracy Index (Chainey et al., 2008) and the Recapture Rate Index (Levin, 2008) are examined in the last part of this section.

All hot spot analysis methods were executed with the CrimeStat 4.0 software and ESRI's ArcGIS 10.2 software package.

### 4.2.1 Parameter Settings and Classification Methods

The hot spot methods included in this thesis differ also from each other based on their type of input data. An aggregated data type contains areal units with spatially neighboring point counts as attributes. Whereas a point type includes the point data set.

Table 4.1 represents different input data types of cluster methods and their related parameter settings.

Data Type	Method	Parameter
Point	NNH	Search radius, minimum points per cluster, convex hull
Point	Fuzzy Mode	Distance, threshold
Point	KDE	Kernel type, bandwidth, threshold
Point	STAC	Search radius, minimum points per cluster, threshold
Aggregated	Geographic Boundary Thematic Mapping	Areal units, threshold
Aggregated	Grid Thematic Mapping	Grid cells, threshold
Aggregated	Local Moran's I	Areal units, spatial weights
Aggregated	GI*	Areal units, spatial weights
Aggregated	Grid Gi*	Grid cells, threshold

Table 4.1: Input data type of point pattern methods and their parameter

Parameter Settings:

Parameter settings, including the search radius, also referred to as bandwidth, are a critical part in analyzing different cluster techniques described in this thesis. There are several suggestions from different researchers depending on what types of data are available. The resulting value of the median nearest neighbor distance multiplied by 6, 9, or 12 was mentioned by Brimicombe (2004) for point data in terms of bandwidth selection. For point data aggregated into geographic units, Cliff and Haggett (1988) suggested to use a global Moran's I correlogram to analyze how spatial autocorrelation changes with distance (Chainey and Ratcliffe, 2005). This thesis used K-order values of the mean nearest neighbor distance suggested by Williamson et al. (1999) to define the appropriate bandwidths. The bandwidth selection depends on the order values of the mean nearest neighbor distance between each point and its *n*th closest nearest neighbor (Chainey and Ratcliffe, 2005). It can be seen as a preferable method since it is related to the spatial distribution of the respective data.

The K-order mean nearest neighbor distance is defined as:

$$d_{k(ran)} = \frac{k(2k)!}{(2^k k!)^2 \sqrt{\frac{N}{A}}} \tag{4-5}$$

where  $k$  is the order and “!” is the factorial operation. The ratio of the observed  $k^{th}$  nearest neighbor distance to the  $k^{th}$  mean random distance defines the  $k^{th}$  nearest neighbor index (Levin, 2013).

These calculations are related to the Kernel Density Estimation (KDE), but the bandwidth selection for all different time periods and types of point events were calculated based on this algorithm as a starting point to experiment with different bandwidth values.

Some hot spot methods, examined in this thesis, are including the value of the cell size in their calculation depending on their different application results. For example, KDE uses the cell size to generate a grid across the point distribution to represent the results with more or less coarse or detailed (low or high resolution) looking maps. A good starting point for defining an appropriate cell size for this method was suggested by Ratcliffe (1999b) to divide the shorter side of the study area extend by 150. In terms of the Grid Thematic Mapping method, Chainey and Ratcliffe (2005) suggested to divide the shortest extent of the study area by 50. Since the limited number of point events affects the cell size determination, additional cell size calculations had to be executed to meet the requirements for the prediction calculation for specific methods.

#### Classification Methods:

Another crucial part to define hot spots is the determination of a thematic threshold which categorizes the resulting values of the hot spot methods into different classes. Most GIS software packages allow a user to choose between different classification schemes. In terms of crime mapping, there are several suggestions by researchers to define hot spots based on different classification schemes, such as the incremental multiples of the grid cell's mean (Chainey et al., 2002), the quantile classification (Chainey et al., 2008), and the standard deviation classification (Boba, 2005). It is also commonly accepted that the highest class defines the areas, where hot spots are located.

This research used mainly two classification schemes which are dependent on the numerical results of the different hot spot methods. For hierarchical clustering techniques, the highest hierarchy or rank was chosen as hot spots. For all the other techniques, researched in this thesis, the highest class of the standard deviation classification scheme was chosen to represent hot spots. Since this research calculates

the prediction accuracy for each hot spot method, the standard deviation classification was more useful in creating hot spots based on their count and area size.

The standard deviation describes the variation of the point events around the mean and is defined as follows:

$$SD = \sqrt{\frac{\sum(x-\bar{x})^2}{(n-1)}} \quad (4-6)$$

where  $x$  is each point,  $\bar{x}$  is the mean or average and  $n$  the total number of values.

#### 4.2.2 Point Pattern Hot Spot Methods

##### 4.2.2.1 The Spatial Fuzzy Mode

One of the simplest hot spot analysis method is the “Spatial Fuzzy Mode” and can be calculated with CrimeStat IV (Levin, 2013).

The algorithm is based on a user defined search radius around each visited point location and includes all crime events that fall within each circle. The size of the circle is a critical variable due to the amount of included points to avoid too few or too many clusters. The Fuzzy Mode calculates a circle with the most numbers of points inside the circle and represents the resulting circles in a table with different rank order and following variables:

1. A ranking value where the first rank defines the circle that has the highest number of points included inside its area. The second rank defines the circles with the second most points falling inside their area. The last rank constitutes of circles that include only one point inside their areas
2. The total number of points inside the respective circle areas
3. The X coordinate of each circle midpoint
4. The Y coordinate of each circle midpoint

Table 4.2 represents different input data types of cluster methods and their related parameter settings.



<b>Event Type / Time Line</b>	<b>Search Radius (in feet)</b>	<b>Thematic Threshold</b>
Property Theft from Cars 2008-2010	5,200	First rank
Property Theft from Cars 2008-2012	3,700	First rank
Property Theft from Cars 2008-June 2013	3,400	First rank
Offender Residence 2008-2010	4,000	First rank
Offender Residence 2008-2012	3,200	First rank
Offender Residence 2008-June 2013	2,900	First rank

Table 4.2: Parameter settings for the Spatial Fuzzy Mode

The bandwidth (search radius) selection (Table 4.2) for the different time periods was calculated based on the mean nearest neighbor distance algorithm to provide consistent parameter settings throughout the execution of different hot spot techniques analyzed in this thesis.

#### 4.2.2.2 Spatial and Temporal Analysis of Crime (STAC)

STAC is one of the first crime mapping application which was developed in 1989 by the Illinois Criminal Justice Information Authority (Levin, 2013). The spatial part of this method identifies clusters of points based of a user defined search radius, a minimum cluster, and cell size. The first step contains the creation of a user defined triangular or rectangular 20 x 20 grid that is laid over the study area. Then this method places a circle at every grid cell intersection based on the size of the defined search radius. STAC summarizes the points that fall within each circle area and specifies different ranks in descending order of the number of points falling inside each circle area. To avoid that the same points belong to multiple clusters, STAC repeatedly combines these points within the circles until no overlapping circles exist.

This routine is implemented in CrimeStat 4.0 and calculates as a result standard deviational ellipses or convex hulls. Furthermore, only the space algorithm is provided in CrimeStat 4.0, since the temporal part of STAC was not implemented.

STAC is not restricted to artificial or administrative boundaries, such as census tracts or police beats. It supports analysts and decision makers to focus on smaller areas within

different boundaries. It is worth to mention that there are also limitations due to the application guidelines of the parameter settings. It can be difficult for novices to calculate useful results (Eck et al., 2005). Furthermore, the STAC result based on standard deviational ellipses does not show the detailed spatial distribution of crime events (Ratcliffe and McCullagh 2001) and does not recognize events which are not within an ellipse for further comparison analysis (Eck et al., 2005).

Table 4.3 shows the STAC parameter settings of property theft from cars and the related offender residences.

STAC			
Event Type / Time Line	Cell Size (in feet)	Search Radius (in feet)	Thematic Threshold
Property Theft from Cars 2008-2010	N/A	2,900	First order clusters (min. 10 cluster)
Property Theft from Cars 2008-2012	N/A	2,400	First order clusters (min. 15 cluster)
Property Theft from Cars 2008-June 2013	N/A	2,400	First order clusters (min. 15 cluster)
Offender Residence 2008-2010	N/A	2,900	First order clusters (min. 10 cluster)
Offender Residence 2008-2012	N/A	2,700	First order clusters (min. 15 cluster)
Offender Residence 2008-June 2013	N/A	2,700	First order clusters (min. 15 cluster)

Table 4.3: Parameter settings for the STAC method

Defining a cell size was not necessary since non-aggregated point data are included in the calculation. Defining a suitable bandwidth was a time consuming experiment since the calculation of the mean nearest neighbor distance values did not deliver appropriate results for subsequent prediction index measurements. Therefore, the search radius values are ranging from 2,700 to 2,900 feet for each time period and crime event. The thematic threshold was set to first order clusters and a minimum number of 10 to 15 points per cluster.

#### *4.2.2.3 Nearest Neighbor Hierarchical Clustering (NNH)*

NNH is one of the oldest clustering method (King, 1967; Johnson, 1967).

It identifies only points that are closer than expected under spatial randomness (Eck et al., 2005) and groups these points based on their minimum number within a cluster that has to be defined by a user. This method defines different orders of clusters. The smallest are the so-called first-order clusters which will be grouped in the second step into second order clusters, etc. until one large single cluster is left which contains all points and all previous clusters or the clustering criteria fail (Levin, 2013).

The NNH method is included in the CrimeStat v4.0 software and offers next to default settings, two additional user-defined parameters. A user can set a threshold distance to recognize points that are within the selected distance to each other. The second parameter gives a user the opportunity to define a minimum number of points per cluster. Only if both criteria apply, selected points are clustered to first order clusters. Only Clusters, those are spatially closer than the threshold distance will be selected for higher level clustering (second order and higher level clusters).

Another option is to let the NNH method calculate a random nearest neighbor distance. This is based on the size of the study area, the number of points in the study area and a user-defined probability (p) value. For example, if  $p \leq 0.05$ , then there is a chance that only 5% of the point pairs would be within the random threshold distance. If a cluster contains more than two points, then the chance of selecting this cluster will be smaller than a cluster with two points (Levin, 2013).

Since Ratcliffe and McCullagh (1999b) pointed out the problem of representing hot spots in the form of standard deviational ellipse, convex hull polygons were chosen as an output format of the NNH calculation. Convex hulls have the advantages to reduce the area of clustered point groups, it forms smaller polygons to describe point clusters and shows more precise areas of cluster groups (Grubestic, 2006) compared to, e.g., administrative boundaries or ellipses.

It should be kept in mind that user-defined parameter settings are mostly dependent on the user's experience. For example, defining a minimum number of points per cluster can be rather a subjective than an objective decision.

Table 4.4 represents the NNH parameter of property theft from cars and the related offender residences.

NNH			
Event Type / Time Line	Cell Size (in feet)	Search Radius (in feet)	Thematic Threshold
Property Theft from Cars 2008-2010	N/A	5,200	First order clusters (min. 10 cluster)
Property Theft from Cars 2008-2012	N/A	3,700	First order clusters (min. 15 cluster)
Property Theft from Cars 2008-June 2013	N/A	3,400	First order clusters (min. 15 cluster)
Offender Residence 2008-2010	N/A	4,000	First order clusters (min. 15 cluster)
Offender Residence 2008-2012	N/A	3,200	First order clusters (min. 15 cluster)
Offender Residence 2008-June 2013	N/A	2,900	First order clusters (min. 15 cluster)

Table 4.4: Parameter settings for the NNH method

Defining a cell size was not necessary since the NNH method was using point data as an input. The search radius (or bandwidth) was defined by the first k-order mean nearest neighbor distance for each point type and time period separately and is consistent with other methods researched in this thesis. After experimenting with different numbers the thematic threshold was set between a minimum number of 10 and 15 points per cluster for first order clusters.

#### 4.2.2.4 Kernel Density Estimation (KDE)

KDE is a very popular hot spot technique which creates a smooth surface based on the density of point distributions across the study area (Chainey and Ratcliffe, 2005; Eck et al., 2005). It first generates a user defined grid over all point events. Starting from each grid cell, it calculates based on a moving three-dimensional kernel function the distance to each point within a user defined search radius (Figure 4.2). The different distances of points from a cell represent the weight values. As a result, the final cell value will be the sum of the different weights.

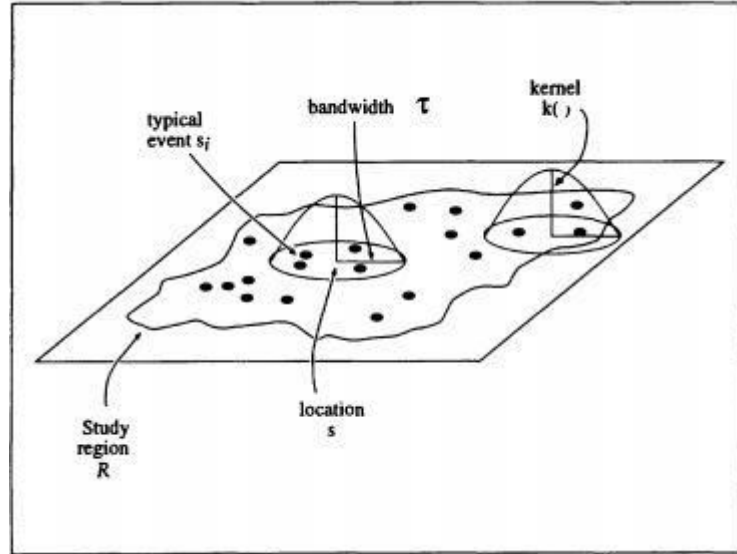


Figure 4.2: Calculating the KDE of a point pattern, Source: (Gatrell et al., 1996)

This research used a quartic kernel interpolation method which calculates a higher weight in the center of each point. The weight decreases with growing distances from each point. The quartic Kernel Density Estimation method is defined as follows:

$$\hat{\lambda}_{\tau}(s) = \sum d_i \frac{3}{\pi \tau^2} \left(1 - \frac{d_i^2}{\tau^2}\right)^2 \quad (4-7)$$

where  $d_i$  is the distance between point  $s$  and the observed location point event  $s_i$ . The summation is only over values of  $d_i$  which do not exceed  $\tau$ . The region of influence within which observed events contribute to  $\hat{\lambda}_{\tau}(s)$  is therefore a circle of radius  $\tau$  centered on  $s$ . At the site  $s$  (a distance of zero), the weight is simply  $3/\pi \tau^2$  and drops smoothly to a value of zero at distance  $\tau$  (Gatrell et al., 1996).

The grid cell size, the type of the kernel function and the search radius (bandwidth) are three critical parameters which are depended from the scale of the study area and the distribution of point events. Large cell sizes tend to be coarser looking and are more appropriate for large scale study areas. Fine cell sizes are more suitable for small scale areas where spatial details are more important for analysts. As a starting point, (Ratcliffe 1999b) suggested to divide the shortest extent of the study area based on a bounding rectangle by 150.

Since KDE is a smoothing method defined by different bandwidth values, calculated hot spot areas can still overlap with their smoothed edges into areas with no crime. Defining

an appropriate thematic threshold can be a challenge for some users since the influence of the visual attraction of the KDE results can be misleading.

Table 4.5 shows the KDE parameter settings of property theft from cars and the related offender residences

<b>KDE</b>			
<b>Event Type / Time Line</b>	<b>Cell Size (in feet)</b>	<b>Search Radius (in feet)</b>	<b>Thematic Threshold</b>
Property Theft from Cars 2008-2010	330	5,200	>2.6 std. dev.
Property Theft from Cars 2008-2012	330	3,700	>2.9 std. dev.
Property Theft from Cars 2008-June 2013	330	3,400	>2.9 std. dev.
Offender Residence 2008-2010	330	4,000	>2.6 std. dev.
Offender Residence 2008-2012	330	3,200	>2.6 std. dev.
Offender Residence 2008-June 2013	330	2,900	>2.6 std. dev.

Table 4.5: Parameter settings for the KDE method

The cell size was set to 330 feet for each time period and point event to keep a consistent resolution of the visualization (Table 4.5). The search radius (bandwidth) was defined based on the k-order mean nearest neighbor distance. This is the same bandwidth as was used for the other hot spot techniques researched in this thesis. The thematic threshold was set to the highest class of the standard deviation classification and ranges from >2.6 to >2.9 above the mean.

#### 4.2.3 Aggregated Hot Spot Methods

##### 4.2.3.1 Geographic Boundary Thematic Mapping (GBTM)

A well-known technique to visualize the distribution of point data is the geographic boundary thematic mapping method. Point events are aggregated into administrative or political boundaries such as census tracts, census block groups or borough boundaries.

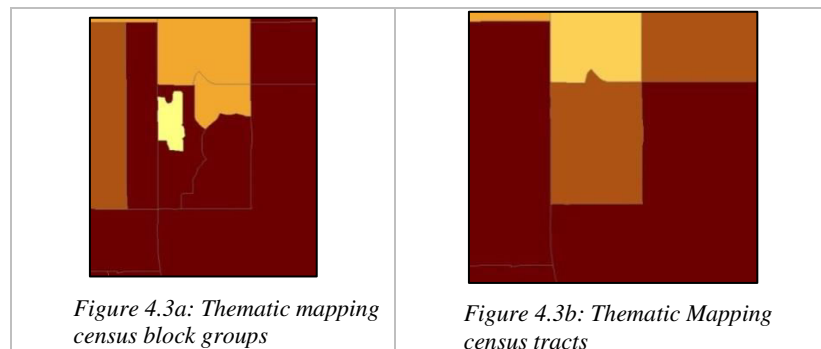
The sum of the aggregated points of each enumeration unit can be thematically displayed to describe the spatial pattern of point events (Eck et al., 2005).

A more common approach which considers the underlying population, in addition to the crime count or to the offender residence count calculation, results in crime or offender residence rates instead of point counts for each geographic unit. A higher population attracts more crime due to increasing opportunities and motivation for offenders. Business areas, including shopping locations have a higher fluctuation of people, including their property like cars or other personal belongings. The opportunities for theft, robbery, or other crime etc. are much higher in densely populated areas than in less populated areas.

This research used census block groups as administrative boundaries and crime rates as aggregated data, based on crime counts per 1000 people for each census block group.

Another reason to choose census block groups over census tracts or other enumeration units is related to the predictive accuracy calculations and the detection of smaller hot spot areas. The size of hot spot areas and the amount of predicted points in these areas are important parameters to calculate useful accuracy indexes and the visualization of a more detailed spatial distribution of point events.

The disadvantage of this method is that administrative areas can have different sizes, shapes. This can lead to different results due to the underlying spatial distribution of crime. For example, smaller geographic units like census block groups (Figure 4.3a) can be aggregated into larger units like census tracts (Figure 4.3b). This changes the shape and scale of these administrative units. In terms of thematic crime mapping, this can lead to different interpretations, since the underlying detailed crime pattern cannot be identified. This issue is known as a part of the Modifiable Area Unit Problem (MAUP) described by Openshaw (1984).



For example, in Figure 4.3b the total population is aggregated to different census geographies. Census tracts are larger and possess a lower spatial resolution than census block groups. Thus, a more detailed spatial distribution of the total population count is not possible.

Table 4.6 represents the GBTM parameter of property theft from cars and the related offender residences.

<b>Geographic Boundary Thematic Mapping</b>			
<b>Event Type / Time Line</b>	<b>Cell Size (in feet)</b>	<b>Search Radius (in feet)</b>	<b>Thematic Threshold</b>
Property Theft from Cars 2008-2010	N/A	N/A	> 2.9 SD
Property Theft from Cars 2008-2012	N/A	N/A	> 2.9 SD
Property Theft from Cars 2008-June 2013	N/A	N/A	> 2.9 SD
Offender Residence 2008-2010	N/A	N/A	>2.6 SD
Offender Residence 2008-2012	N/A	N/A	>2.6 SD
Offender Residence 2008-June 2013	N/A	N/A	>2.6 SD

Table 4.6: Parameter settings for the GBTM method

Table 4.6 shows thematic threshold values for two different crime types and three different time-frames that are applied to identify hot spots in geographic boundary thematic mapping. No definitions of cell size and search radius values (Table 4.3) were necessary, since the crime events were aggregated into administrative units. The highest class of the standard deviation (SD) classification was chosen to define hot spots. Thus, the thematic threshold ranges from >2.6 SD to >2.9 SD above the mean for the different crime types.

#### 4.2.3.2 Grid Thematic Mapping (GTM)

The problems regarding different sizes and shapes of administrative units (compare Section 4.2.3.1 GBTM)) are minimized by applying a grid based thematic mapping



method that requires a user to define a quadratic grid across the study area. The crime counts can be aggregated to each grid cell and thematically classified (Eck et al., 2005). Many crime analysts prefer crime hot spot indicators which are calculated by crime count per cell area. Based on the grid cell size the user can show more detailed spatial patterns within smaller or larger administrative boundaries.

The grid cell size is an important factor to detect hot spots based on this method because grid cells that are too coarse can miss details of the spatial distribution of crime events within a cell. A small cell size can represent too much spatial information in a specific area which may not be useful for a certain analysis goal. Additionally, a small cell size increases processing time and file size.

Compared to administrative boundary mapping, this method represents a more detailed spatial distribution of point events since the user can choose a suitable cell size to detect local hot spots. To provide a starting point, Chainey and Ratcliffe (2005) suggested to divide the longest extent of the study area by 50 to define an initial cell size. Following these suggestions, figures 4.4 and 4.5 represent a table with the aggregated count of crime events and the related reference grid.

FID	Shape	Join_Count	TARGET_FID	Id	REPOR
1879	Polygon	4	1879	0	905041
1681	Polygon	3	1681	0	800422
2406	Polygon	3	2406	0	800430
719	Polygon	2	719	0	900603
1602	Polygon	2	1602	0	800397
1610	Polygon	2	1610	0	800252
1787	Polygon	2	1787	0	1000101
1877	Polygon	2	1877	0	800520
2251	Polygon	2	2251	0	900404
2378	Polygon	2	2378	0	800023
2461	Polygon	2	2461	0	800297
2633	Polygon	2	2633	0	800314
2751	Polygon	2	2751	0	900009

Figure 4.4: GTM table using 1,000 feet grid cell size

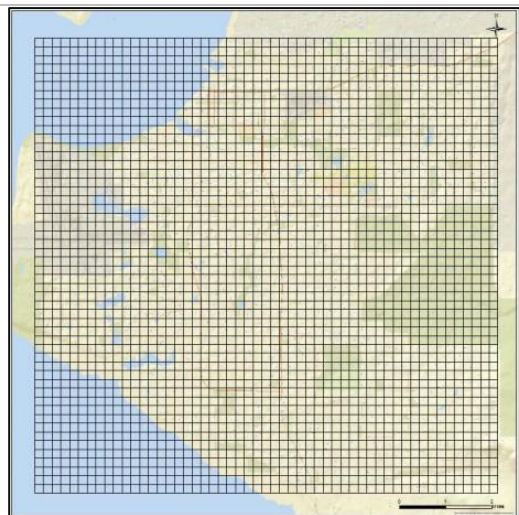


Figure 4.5: GTM with a 1,000 feet grid overlay

The “Join\_Count” field in the table (Figure 4.4) contains the aggregated crime counts. The suggested calculation by Chainey and Ratcliffe (2005) for an initial grid size resulted in a 1,000 feet cell size (Figure 4.5). It is evident that there are not high numbers of crime events falling within each cell. This cell size might be suitable to visually locate a

couple of small hot spots but these are too small to perform prediction calculations for the different time periods. The goal was to define a big enough cell size to provide enough data for the prediction calculations for the longest retrospective time periods. Therefore, after experimenting with different values, a cell size of 4,000 feet was chosen.

Table 4.7 shows the GTM parameter settings of property theft from cars and the related offender residences

<b>Grid Thematic Mapping</b>			
<b>Event Type / Time Line</b>	<b>Cell Size (in feet)</b>	<b>Search Radius (in feet)</b>	<b>Thematic Threshold</b>
Property Theft from Cars 2008-2010	4,000	N/A	> 2.9 std. dev.
Property Theft from Cars 2008-2012	4,000	N/A	> 2.9 std. dev.
Property Theft from Cars 2008-June 2013	4,000	N/A	> 2.9 std. dev.
Offender Residence 2008-2010	4,000	N/A	>2.9 std. dev.
Offender Residence 2008-2012	4,000	N/A	>2.9 std. dev.
Offender Residence 2008-June 2013	4,000	N/A	>2.9 std. dev.

Table 4.7: Parameter settings for the GTM method

The same cell size number and thematic threshold was defined for all the time periods and point events (Table 4.7). As mentioned in Chapter 4.2.1, the highest class of the standard deviation classification was chosen to define hot spots.

#### 4.2.3.3 Local Moran's I

Anselin's local Moran's I is the basis for the "Local Indicator of Spatial Association" (LISA) group, that contains different individual methods such as the Local Geary's C and the Getis and the Ord  $G_i^*$  statistics. These statistics apply the concept of spatial association. In short, this concept can be described as a test comparing the number crime events in an area similar with the count of events in neighboring areas. As an example, it is possible that high drug selling events in one neighborhood can influence crime rates in a neighboring area.

Global statistical tests like the global Moran's I can show limitations in analyzing the location, relative scale, shape, size, and extent of hot spots since data from the whole study area are included into the calculation without the local aspect (Chainey and Ratcliffe, 2005). To overcome these limitations, different extensions to LISA were developed to examine the relationship between a single point and its neighbors based on a specified distance.

As one of the oldest LISA statistic (Anselin, 1995), the local Moran's I method assigns a global Moran's I algorithm to each zone (polygon) and identifies similarities and dissimilarities based on covariance between neighboring zones (polygons) (Levine, 2013). The Local Moran's I value can be either positive or negative. If an enumeration unit such as a census block group has neighboring features with similarly high or low values in it, compared to the enumeration unit in the center, then Moran's I index will be positive and is considered as a part of a cluster. A negative local Moran's I index indicates that the feature in an enumeration unit has neighboring features with dissimilar values, then this is referred to as an spatial outlier.



Figure 4.6: Deriving statistically significant hot-, cold-spots and spatial outliers using the Local Moran's I statistic, Source: (ESRI, 2014)

Figure 4.6 shows the calculated output feature class including four attributes for each feature. Features surrounded by features with similar values have a high positive z-score. A low negative z-score defines a spatial significant outlier. The p-value is a probability that the observed spatial pattern was created by a random process. The p-value is the result of a standard deviation (ESRI, 2013) and must be small enough, in the context to the local Moran's I, to be determined as statistically significant. Summarized, a positive spatial autocorrelation defines spatial dependencies between locations whereas spatial independent locations are defined as a negative spatial autocorrelation.

The Local Moran's I calculation requires an intensity/weight value which is either the count of crimes or the crime rate of each enumeration unit (e.g. census block group). This

research used a crime rate as an intensity/weight value, which includes the total population into the calculation. No cell size and bandwidth values needed to be defined and the thematic threshold was set to >99% confidence (Table 4.8).

Table 4.8 represents the local Moran’s I parameter of property theft from cars and the related offender residences.

Local Moran’s I			
Event Type / Time Line	Cell Size (in feet)	Search Radius (in feet)	Thematic Threshold
Property Theft from Cars 2008-2010	N/A	N/A	>99% confidence
Property Theft from Cars 2008-2012	N/A	N/A	>99% confidence
Property Theft from Cars 2008-June 2013	N/A	N/A	>99% confidence
Offender Residence 2008-2010	N/A	N/A	>99% confidence
Offender Residence 2008-2012	N/A	N/A	>99% confidence
Offender Residence 2008-June 2013	N/A	N/A	>99% confidence

Table 4.8: Parameter settings for the Local Moran’s I method

#### 4.2.3.4 Getis and Ord $G_i^*$ Statistics

Another set of the LISA statistics which analyzes the local spatial autocorrelation of zones (polygons) are the  $G_i^*$  and  $G_i$  statistics. The difference between these two methods and the local Moran is that both  $G_i$  algorithms compare the local averages against the global averages in their calculation. The  $G_i^*$  statistic includes the attribute value of the point in its calculation, whereas  $G_i$  does not include this value and considers only the value of its nearest neighbors that lie inside the search distance. In other words, the  $G_i^*$  calculates if all values within a specified search distance are statistically different to the values of the whole study area (Chainey, 2008). When local spatial association exists, the calculated  $G_i^*$  value will be positive because a high number of crime events are closer together and the null hypothesis of complete spatial randomness can be

rejected. A negative  $G_i^*$  value indicates a low number of crimes are closer together. This thesis included the  $G_i^*$  method as one of the preferred hot spot technique.

1	1	1	5	0	0	0	1	0	0	0	0	0	0	3	2	
0	3	0	0	6	1	0	1	1	0	0	0	0	0	1	3	
5	0	0	0	1	9	5	0	0	3	0	0	1	0	1		
1	4	:												0	0	2
1	0	:	0	0	2	1	2							1	5	0
3	5	:	1	6	6	2	2							0	1	0
0	0	:	1	6	6	2	2							1	2	0
0	2	:	6	12	9	2	2							0	0	2
0	0	:	6	12	9	2	2							0	2	2
1	2	:	1	2	1	1	1							3	0	2
4	4	:	1	2	1	1	1							1	6	4
1	1	:	0	1	0	1	3							6	1	0
0	0	:	0	1	0	1	3							4	0	0
1	1	:	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	:	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	8	:	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4.7:  $G_i^*$  matrix showing cells with values, Source: (Chainey and Ratcliffe, 2005)

For example, Figure 4.7 is representing a grid cell matrix containing the crime counts within each cell which is identified by its centroid point. The centroid distance is defined by the distance of the centroid point (located in the center of the cell) (Chainey and Ratcliffe, 2005) of each cell to the centroid point of the immediate neighbor cell. The null hypothesis states that a certain point (cell) is not the center of a group of unusually high values centered on itself and its surrounding neighbors. In other words, the total crime count in a specific cell and in neighboring cells up to a certain distance is not significantly higher than anywhere else in the matrix (Chainey and Ratcliffe, 2005) and therefore randomly distributed. The above figure (Figure 4.7) shows unusually high values of 12 and the neighboring crime counts, which are also rather high. Therefore, the null hypothesis can be rejected.

The  $G_i^*$  method was executed with ESRI’s ArcGIS software. For further comparison, census block groups and grid cells were chosen as reference data including the total population in order to calculate the crime rate.

Table 4.9 shows the parameter settings of the  $G_i^*$  census block group and the  $G_i^*$  grid methods.

<b>Gi* census block group / Gi* grid</b>			
<b>Event Type / Time Line</b>	<b>Cell Size (in feet)</b>	<b>Search Radius (in feet)</b>	<b>Thematic Threshold</b>
Property Theft from Cars 2008-2010	-- / 4000	-- / --	>99% confidence
Property Theft from Cars 2008-2012	-- / 4000	-- / --	>99% confidence
Property Theft from Cars 2008-June 2013	-- / 4000	-- / --	>99% confidence
Offender Residence 2008-2010	-- / 4000	-- / --	>99% confidence
Offender Residence 2008-2012	-- / 4000	-- / --	>99% confidence
Offender Residence 2008-June 2013	-- / 4000	-- / --	>99% confidence

Table 4.9: Parameter settings for the Gi\* census block group / Gi\* grid methods

No cell size has been defined since the data were aggregated into census block groups, whereas the cell size of the grid based Gi\* method was set to 4,000 feet (Table 4.9).

After experimenting with different search radius values, the resulting maps were not suitable for further calculation of different prediction indexes. Thus, a search radius value was not defined. ArcGIS interprets an undefined search radius value as a default distance to ensure that each feature has at least one neighboring feature. The thematic threshold was set to >99%.

### 4.3 Prediction Accuracy Analysis

Since hot spot detection methods have become more sophisticated, researchers have started to develop different calculations how these techniques differ based on their capability to predict where crime may occur. One of the first calculations is the hit rate. This simple measure calculates the percentage of new point events within areas where crimes are predicted to occur (Chainey et al., 2008). The Hit Rate is defined as follows:

$$HR = \frac{n_2}{N_2} * 100 \tag{4-8}$$

where n<sub>2</sub> is the count of new crimes which falls within the calculated hotspots multiplied by 100 and then divided by the count of new crimes in the whole study area.

But this calculation does not consider the size of hot spot areas. For example, the hit rate could be 80%, with the hot spots could covering 80% of the whole study area. This would not be a helpful result for law enforcement agencies to plan tactical and strategical policing.

To include the size of the hot spots into the calculation, Chainey et al. (2008) introduced the Prediction Accuracy Index (PAI). This measure divides the hit rate by the size of the hot spot areas in respect to the size of the whole study area. The PAI is defined as follows:

$$PAI = \frac{\left(\frac{n}{N}\right)*100}{\left(\frac{a}{A}\right)*100} = \frac{Hit\ Rate}{Area\ Percentage} \quad (4-9)$$

where  $n$  is the number of future crimes within the hot spots from the retrospective data,  $N$  is the number of new crimes in the whole study area,  $a$  is the sum of the size of the generated hot spot areas and  $A$  is the size of the whole study area.

For example, when 25% of future crimes would fall into retrospective hot spots that make up 50% of the entire study area, the PAI result would be a value of 0.5. A higher PAI value indicates higher prediction accuracy of the hot spot method. The smaller the size of the hot spot areas compared to the whole study area a higher index can be expected.

As an complement to the PAI measure, (Levin (2008)) proposed the Recapture Rate Index (RRI). This calculation does not take any area size into account. Instead, it compares the count of future hot spot crime events to the past hot spot crime events divided by the ratio of all future crime events and past crimes.

The RRI is defined as follows:

$$RRI = \frac{n2/n1}{N2/N1} = \frac{Ratio\ Hot\ Spot\ Points}{Ratio\ Total\ Points} \quad (4-10)$$

where  $n1$  is the count of past hot spot crimes,  $n2$  is the count of future hot spot crime events,  $N1$  is the sum of all crime events from the past and  $N2$  is the total number of events from the future. A higher index indicates a more accurate hot spot technique.

## CHAPTER 5 RESULTS

This chapter contains the results of the global spatial statistical tests, the visual output of each hot spot technique, and the prediction accuracy results from all hot spot methods analyzed in this thesis. Chapter 5 is organized in three sections. Section 5.1 provides the results of the descriptive spatial statistical tests. Section 5.2 contains the visual result of all hot spot methods and Section 5.3 provides a comparison between all hot spot methods, and their prediction accuracy results including, the variables needed for the calculation. Section 5.4 presents a short summary of the findings.

### 5.1 Descriptive Spatial Statistics

The following table (Table 5.1) shows the results of the NNI and SDD calculations.

<b>Event Type / Time Line</b>	<b>NNI</b>	<b>Standard Distance Deviation (in feet)</b>
Property Theft from Cars 2008 - 2010	0.7657	17236.75
Property Theft from Cars 2008 - 2012	0.6988	16599.67
Property Theft from Cars 2008 - June 2013	0.6664	16607.80
Offender Residences 2008 - 2010	0.6358	17202.45
Offender Residences 2008 - 2012	0.5911	17594.83
Offender Residences 2008 – June 2013	0.5916	17470.41

*Table 5.1: Results of the NNI and the SDD of property thefts from cars and offender residences for different time periods*

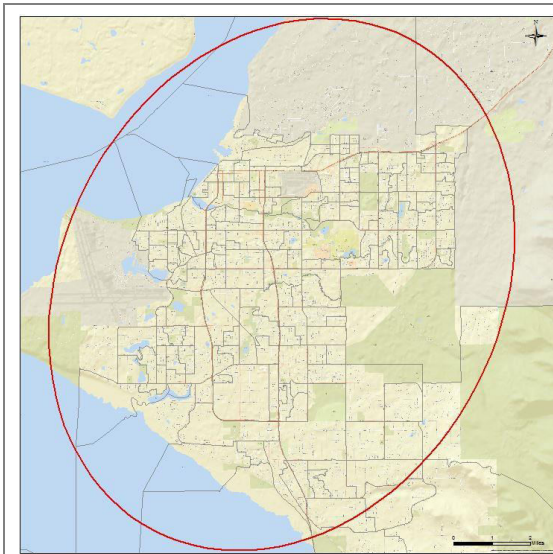
For every time period (Table 5.1) the NNI is less than 1, which is an evidence for clustering in the data. The NNI for all point events indicates that all time periods of



## Results

offender residence locations have a tendency of higher clustering compared to property theft from cars. The standard distance deviation show (Table 5.1) that the time periods of the crime type from 2008 – 2012 and the offender residences from 2008 – 2010 have the lowest dispersion around the mean center.

Following maps (Figures 5.1 and 5.2) show the results of the standard deviational ellipse calculation.



*Figure 5.1: Visual output of the Standard Deviational Ellipse of property theft from cars from 2008 - 2010*



*Figure 5.2: Visual output of the Standard Deviational Ellipse of offender residences from 2008 - 2010*

The visual output (Figures 5.1 and 5.2) was chosen for only one time period of both property theft from cars and the offender residence data. The standard deviational ellipses show nearly the same size and shape and cover the same area of Anchorage with a south-west and north-east direction of the data dispersion.

## 5.2 Visual Results of Hot Spot Methods

The following section contains the visualization of all hot spot methods based on two different crime data sets and three different time periods, including a brief interpretation.

### 5.2.1 Point pattern methods

#### 5.2.1.1 Fuzzy Mode

Figures 5.3 and 5.4 represent a part of the resulting table and created hot spot circles.

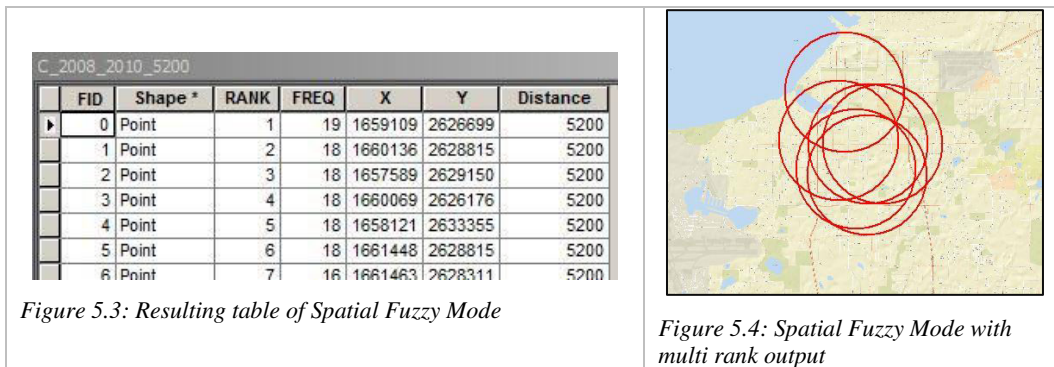


Figure 5.3: Resulting table of Spatial Fuzzy Mode

Figure 5.4: Spatial Fuzzy Mode with multi rank output

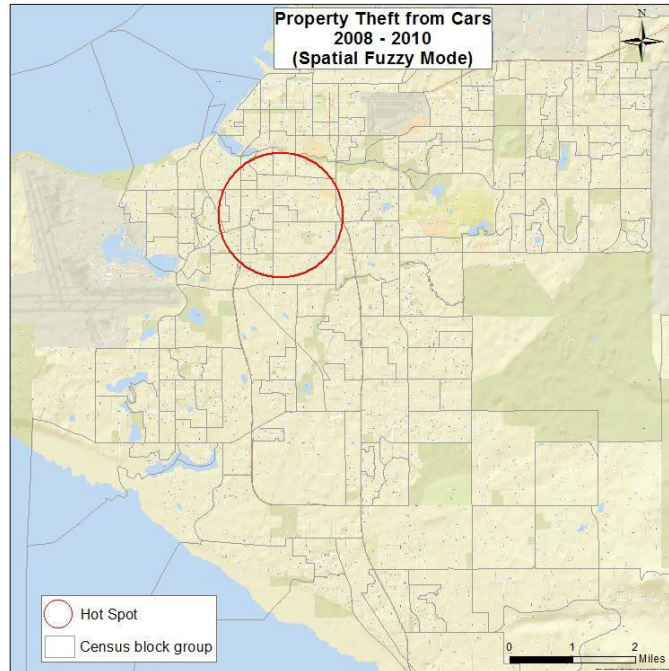
Figure 5.3 shows the resulting table containing rank 1 with the highest amount of points, namely 19. The second until the sixth rank contains 18 points each and so on.

Each search circle which overlaps with another circle (Figure 5.4) includes the amount of points within the overlapping area into the calculation. This is a disadvantage because it changes the frequency of the points at the locations and possibly the hierarchy as well (Levin, 2013) and can lead to misinterpretations based on the calculated ranks.

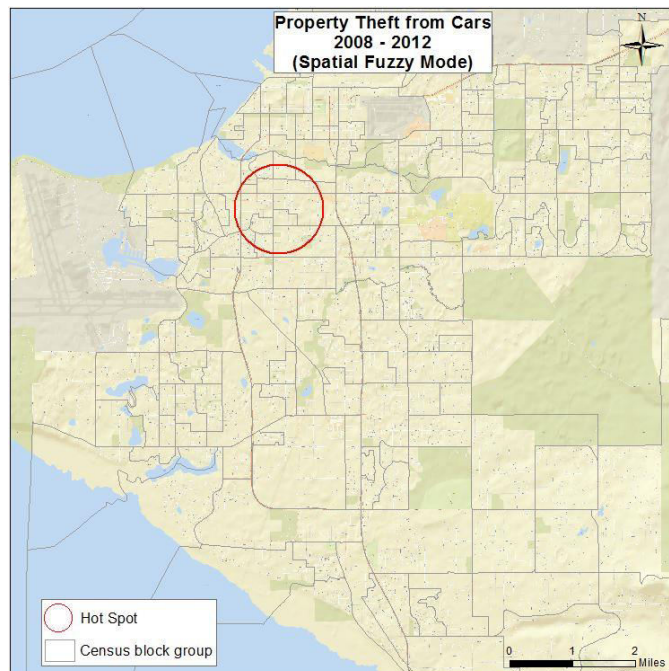
The results of the Spatial Fuzzy Mode for the different time periods show the highest rank with a unique value. Unfortunately, the following ranks are containing mostly the same values which made it difficult to define a thematic threshold including multiple ranks. The overlapping circles (Figure 5.4) defined by rank 1 to 6 (Figure 5.3) would create very big hot spot areas relative to the size of the study area and would not be suitable for the predictive accuracy of the different hot spot methods. Therefore, rank 1 was chosen for every time period as a thematic threshold to keep a consistent definition for all time periods and crime events. The radii of the hot spot circles were defined by the calculated bandwidth values.

*Results*

The following figures (Figures 5.5 - 5.7) show hot spots for all time periods of property theft from cars using the Spatial Fuzzy Mode method.

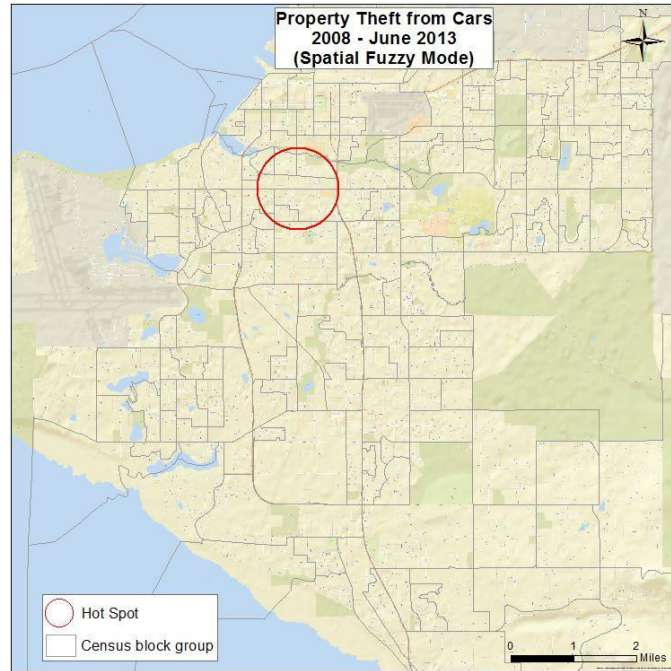


*Figure 5.5: Hot spot of property theft from cars from 2008 – 2010 using the Spatial Fuzzy Mode*



*Figure 5.6: Hot spot of property theft from cars from 2008 – 2012 using the Spatial Fuzzy Mode*

*Results*

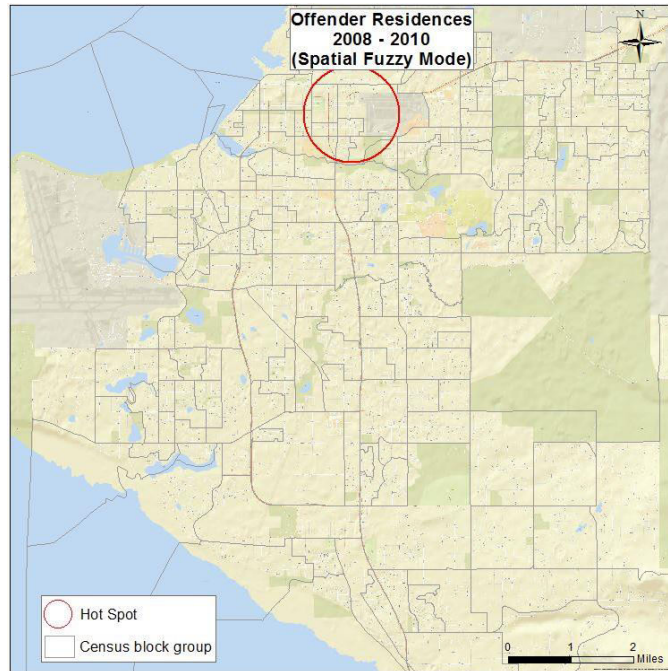


*Figure 5.7: Hot spot of property theft from cars from 2008 – June 2013 using the Spatial Fuzzy Mode*

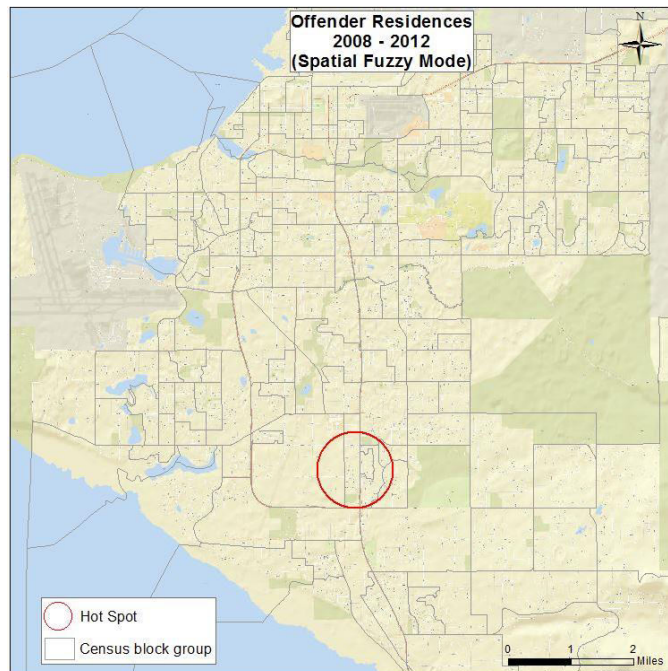
The hot spots are basically located in the mid-town area of Anchorage. The only differences are the sizes of the hot spots which are due to the radius of the bandwidth selection for each time period.

*Results*

The following figures (Figures 5.8 - 5.10) show hot spots for all time periods of offender residence locations using the Spatial Fuzzy Mode method.



*Figure 5.8: Hot spot of offender residences from 2008 – 2010 using the Spatial Fuzzy Mode*



*Figure 5.9: Hot spot of offender residences from 2008 – 2012 using the Spatial Fuzzy Mode*

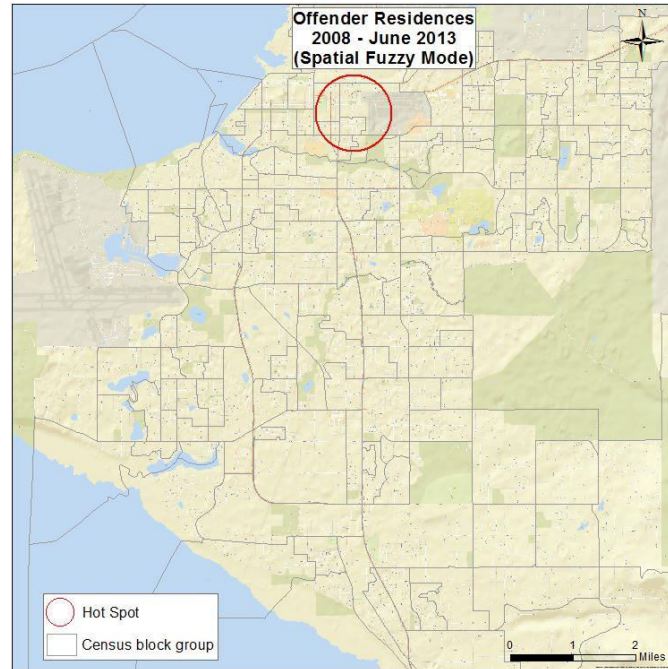


Figure 5.10: Hot spot of offender residences from 2008 – June 2013 using the Spatial Fuzzy Mode

The hot spot maps above (Figures 5.8 – 5.10) regarding offender residences are mostly representing hot spots in down-town Anchorage with one exception. Figure 5.9 shows the hot spot in the southern part of Anchorage. A logical explanation would be that the change from the three year (Figure 5.8) to the five year time period (Figure 5.9) increased the number of points in that particular area just enough to create a first rank hot spot. Interestingly, the five and ½ year time period (Figure 5.10) shows the same hot spot location as the three year time period which is an indication that more points were added at this location.

5.2.1.2 Spatial and Temporal Analysis of Crime (STAC)

Figures 5.11 – 5.13 visualize the spatial distribution of property thefts from cars using the STAC approach.

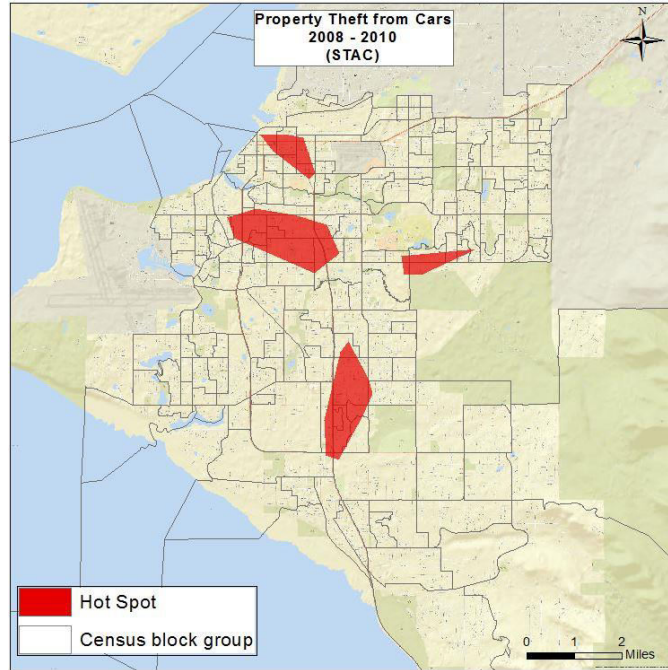


Figure 5.11: STAC of property theft from cars from 2008 – 2010

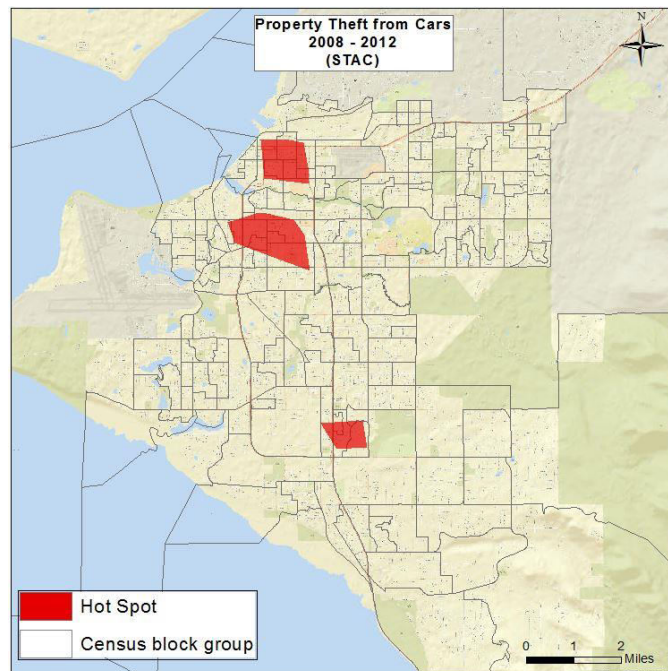


Figure 5.12: STAC of property theft from cars from 2008 – 2012

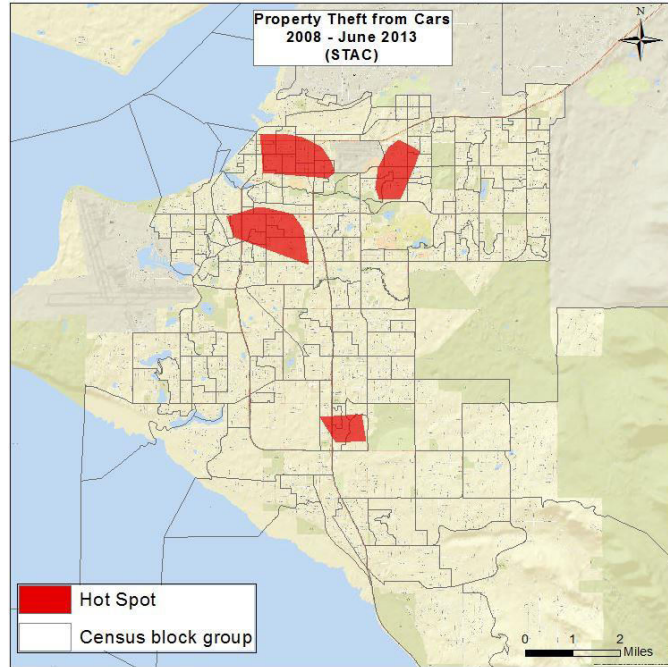


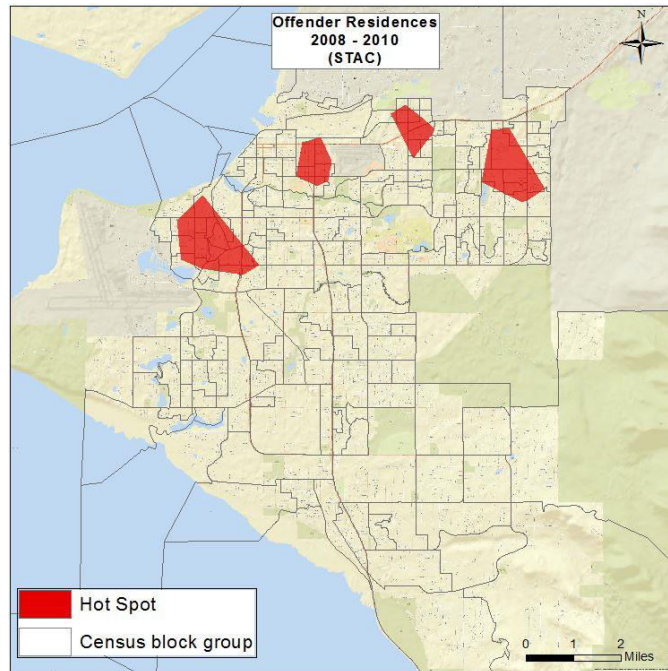
Figure 5.13: STAC of property theft from cars from 2008 – June 2013

The STAC routine calculated several big hot spot areas in down-town, mid-town, and in the southern part of Anchorage (Figures 5.11 – 5.13). The lowest number of hot spots is shown at the five year time period (Figure 5.12).

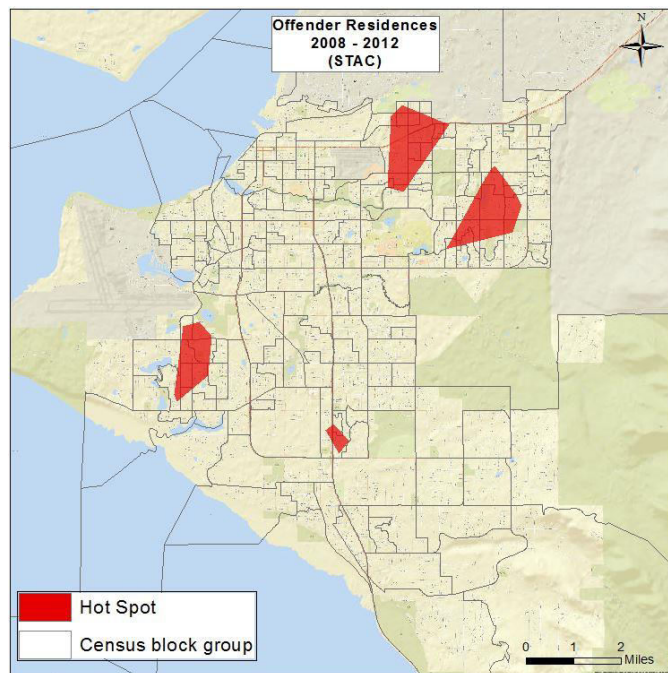


*Results*

Figures 5.14 – 5.16 visualize the spatial distribution of offender residences using the STAC approach.



*Figure 5.14: STAC of offender residences from 2008 – 2010*



*Figure 5.15: STAC of offender residences from 2008 – 2012*

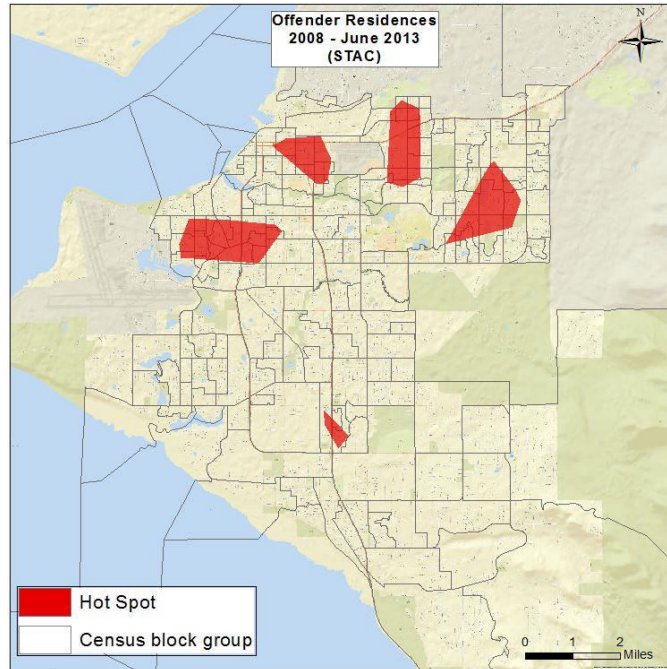


Figure 5.16: STAC of offender residences from 2008 – June 2013

The STAC results of offender residence locations are shown in figures 5.14 and 5.16, with hot spot locations in the down-town, western, northern, and eastern parts of Anchorage. The five year time period (Figure 5.15) shows hot spot location similar to above but without a hot spot in the down-town area.

5.2.1.3 *Nearest Neighbor Hierarchical Analysis (NNH)*

Figures 5.17 – 5.19 visualize the spatial distribution of property thefts from cars using the NNH approach.

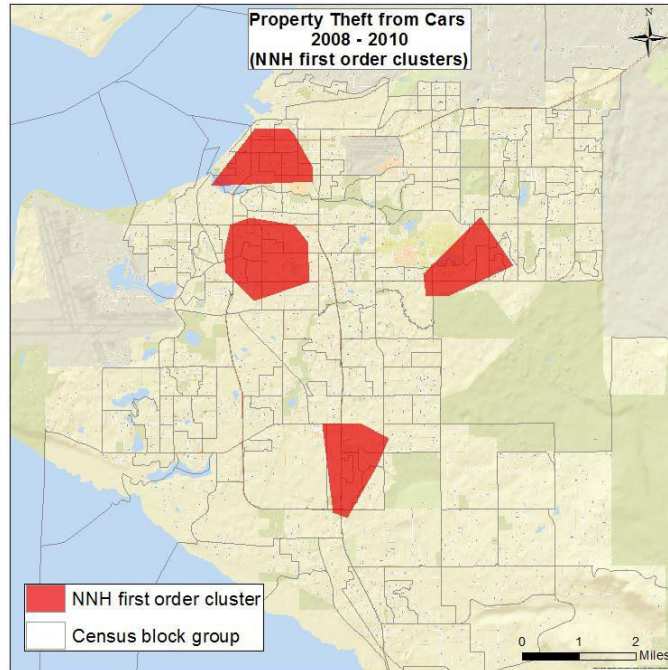


Figure 5.17: NNH of property theft from cars from 2008 – 2010

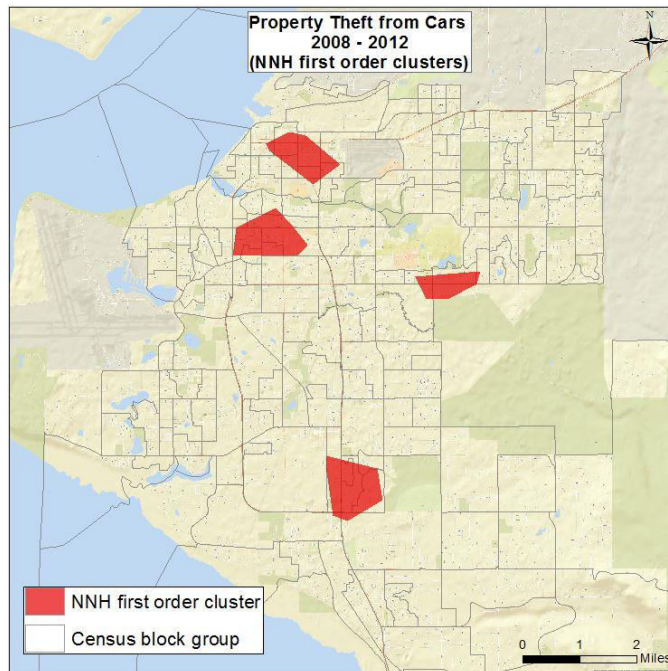


Figure 5.18: NNH of property theft from cars from 2008 – 2012

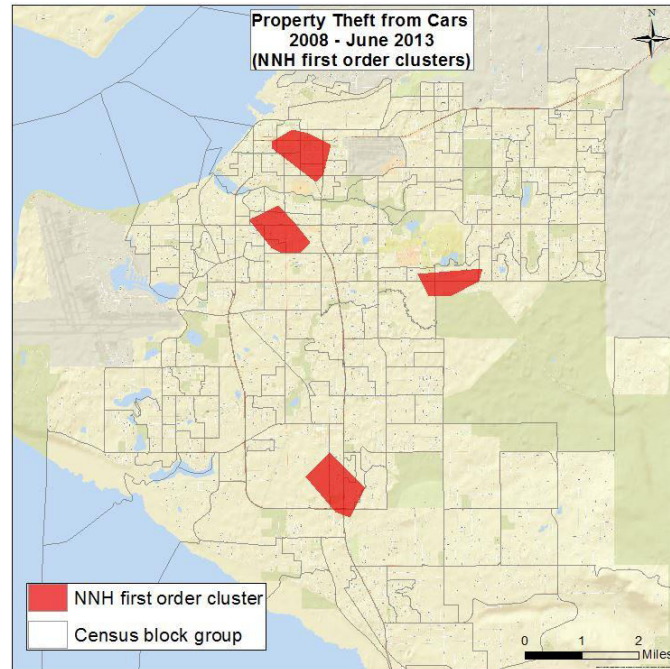
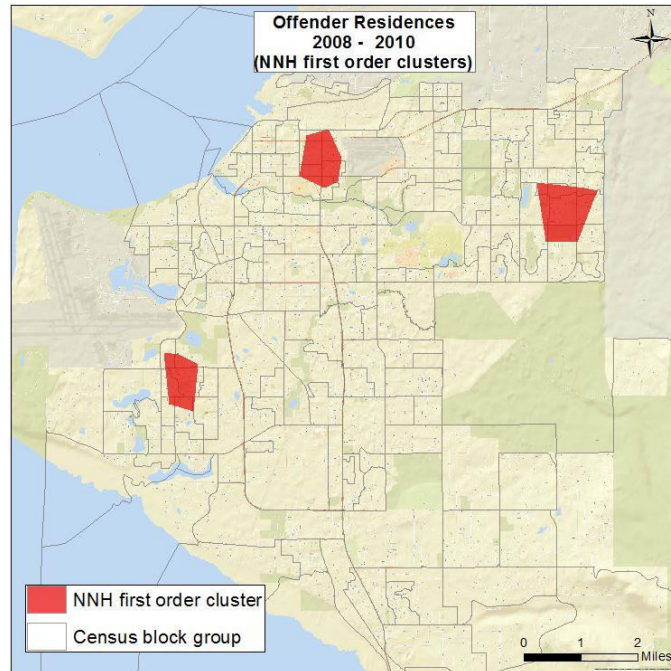


Figure 5.19: NNH of property theft from cars from 2008 – June 2013

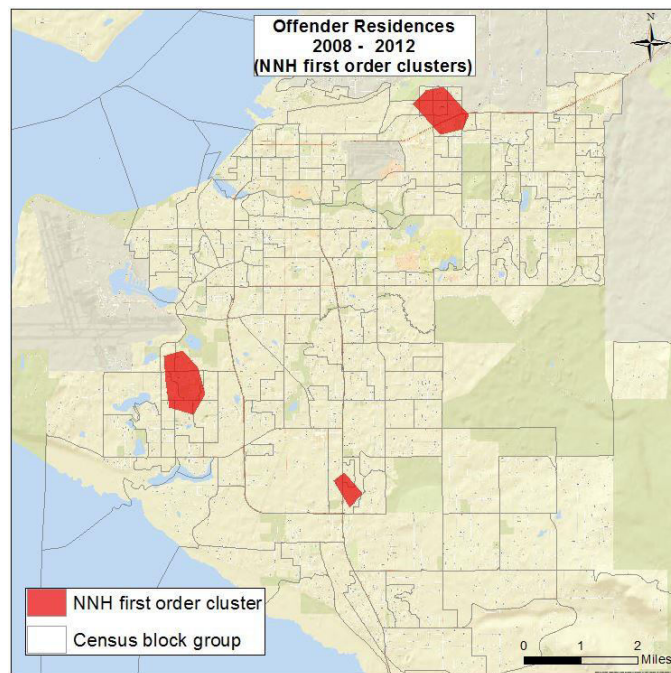
All maps contain the same amount of hot spots for each time period (Figures 5.17 – 5.19). The size of the hot spot areas are bigger in the three year time period (Figure 5.17) compared to the others. This can be explained by the low amount of points and the related bigger bandwidth value. The most hot spots for all maps are concentrated in the down-town and mid-town areas of Anchorage.

*Results*

Figures 5.20 – 5.22 visualize the spatial distribution of offender residences using the NNH approach.



*Figure 5.20: NNH of offender residences from 2008 – 2010*



*Figure 5.21: NNH of offender residences from 2008 – 2012*

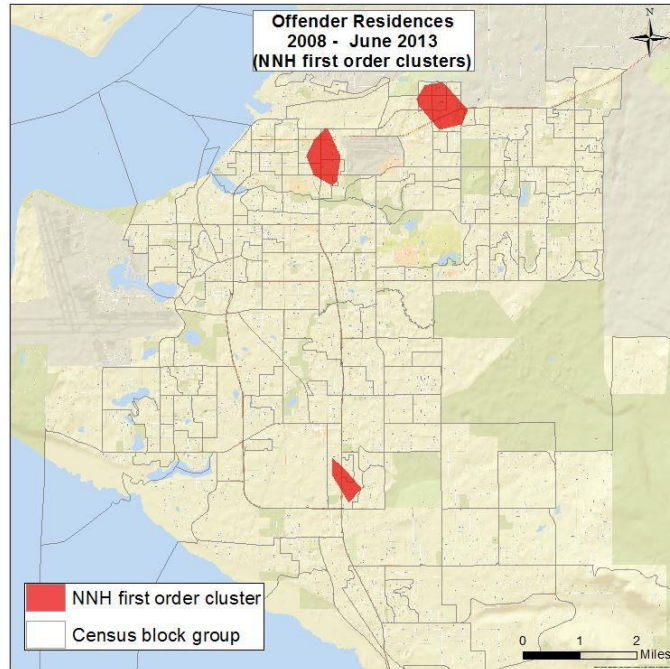


Figure 5.22: NNH of offender residences from 2008 – June 2013

Each of above map (Figures 5.20 – 5.22) shows just three hot spots. The clusters are smaller than the crime event clusters which can be an indication of a higher concentration of offender residence locations.

5.2.1.4 Kernel Density Estimation (KDE)

Figures 5.23 – 5.25 visualize the spatial distribution of property thefts from cars using the KDE approach.

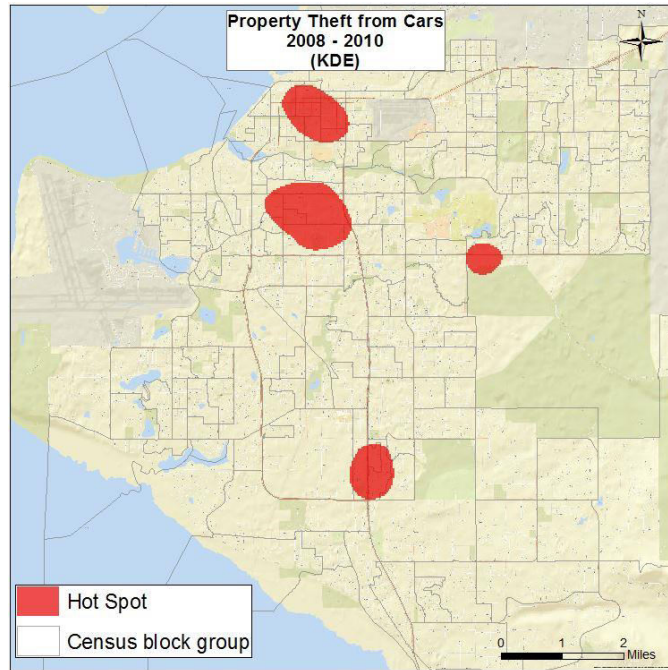


Figure 5.23: KDE of property theft from cars from 2008 – 2010

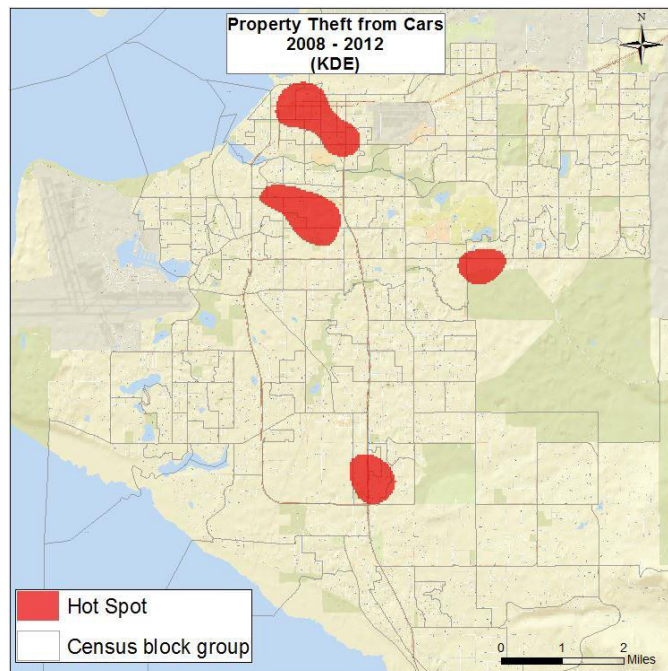


Figure 5.24: KDE of property theft from cars from 2008 – 2012

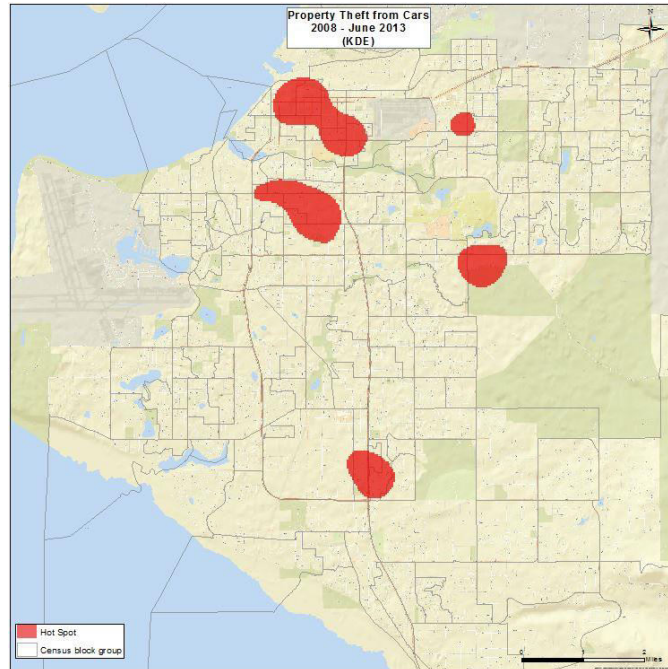


Figure 5.25: KDE of property theft from cars from 2008 – June 2013

One of the advantages of this hot spot technique is represented in the detailed visual output of hot spot areas. The KDE visualizes the spatial patterns of point events with round and smooth edges. Since only a low amount of point events are included, a higher bandwidth value was selected which resulted in bigger hot spot areas (Figures 5.23 – 5.25). The biggest hot spot areas are in down-town and mid-town and smaller areas in the southern and eastern parts of Anchorage for each time period. The five year time period (Figure 5.24) created a additional small hot spot in the north-eastern part of Anchorage.



Results

Figures 5.26 – 5.28 visualize the spatial distribution of offender residences using the KDE approach.

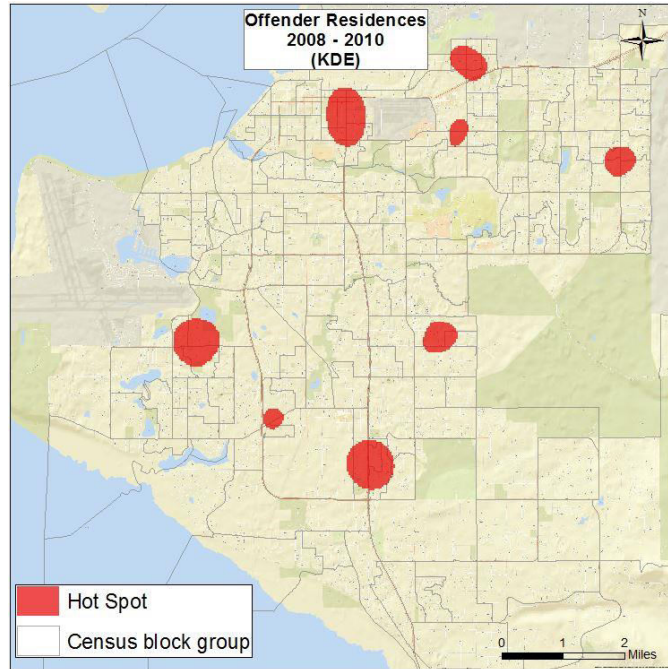


Figure 5.26: KDE of offender residences from 2008 – 2010

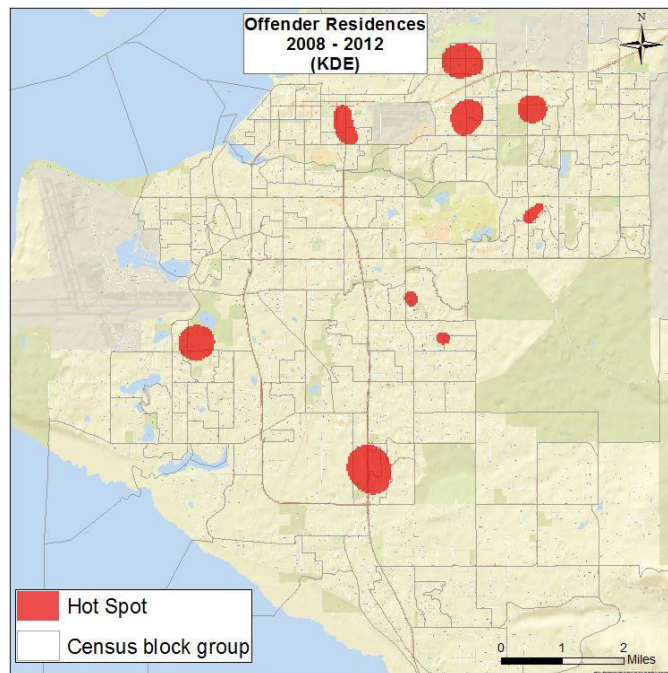


Figure 5.27: KDE of offender residences from 2008 – 2012

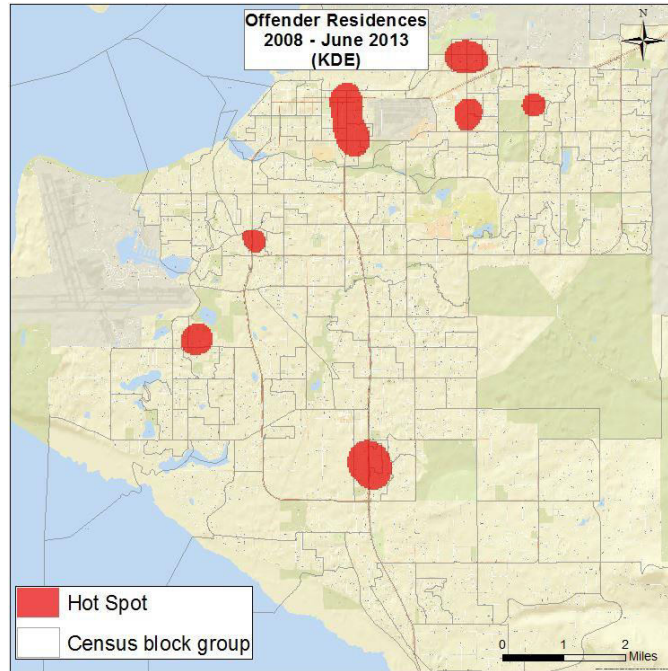


Figure 5.28: KDE of offender residences from 2008 – June 2013

All time periods for the offender residence hot spots (Figure 5.26 – 5.28) have noticeably more and smaller hot spot areas compared to the crime data. The bandwidth selection was smaller and the data seem to have a higher concentration of points. Bigger cluster sizes can be seen in down-town, in the western and southern parts of Anchorage. The five and 1/1 year time period (Figure 5.28) created a couple of small hot spots in the eastern area of Anchorage.

## 5.2.2 Aggregated hot spot methods

### 5.2.2.1 Geographic Boundary Thematic Mapping

Figures 5.29 – 5.31 visualize the spatial distribution of property thefts from cars using the Geographic Boundary Thematic Mapping approach.

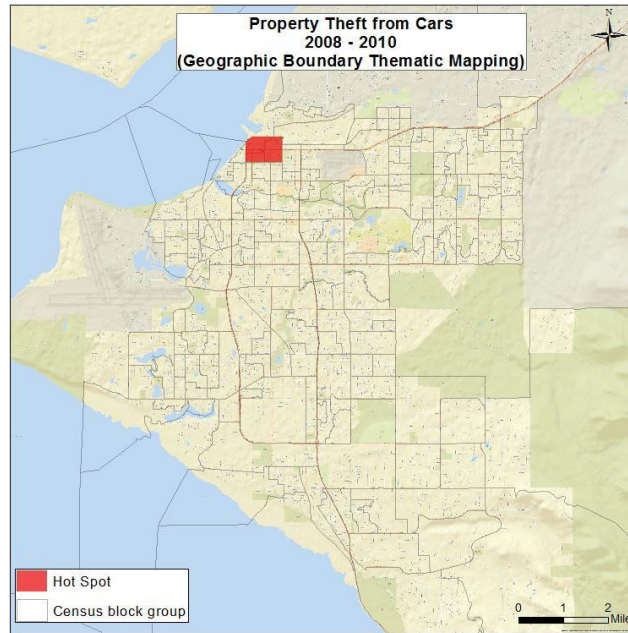


Figure 5.29: GBTM of property theft from cars from 2008 – 2010

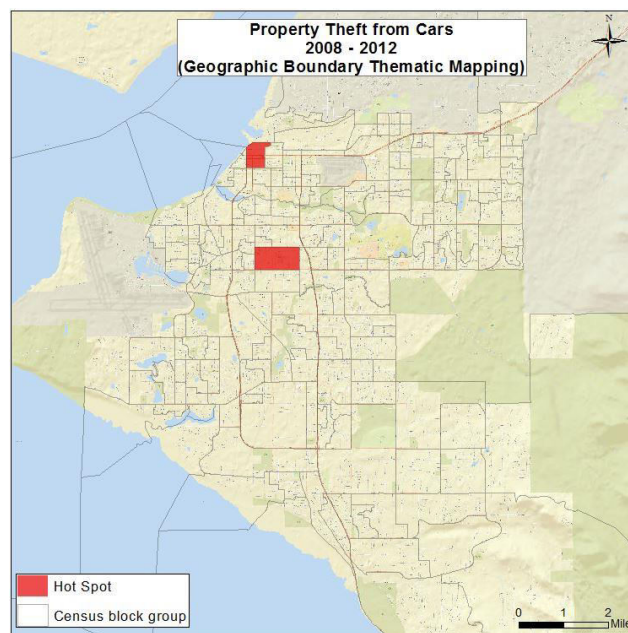


Figure 5.30: GBTM of property theft from cars from 2008 – 2012

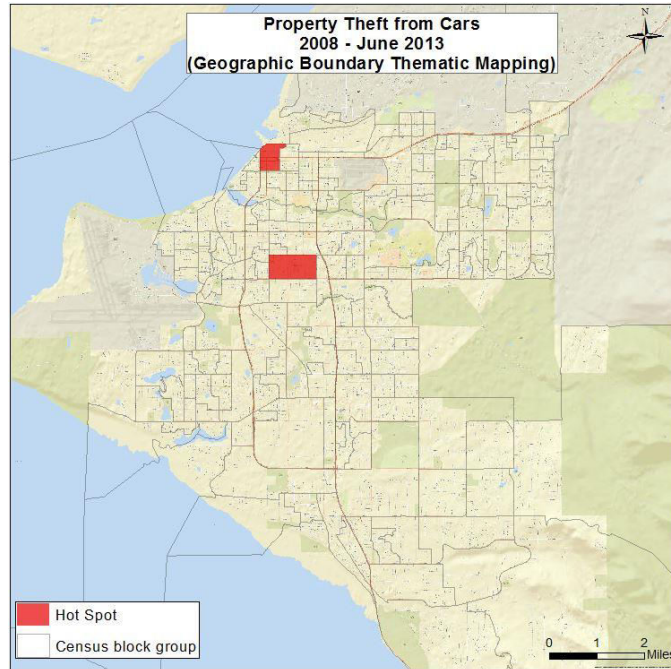
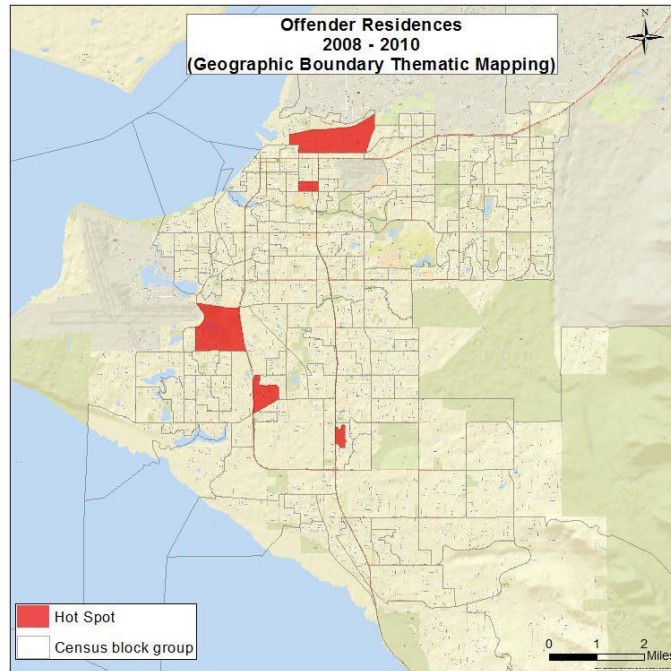


Figure 5.31: GBTM of property theft from cars from 2008 – June 2013

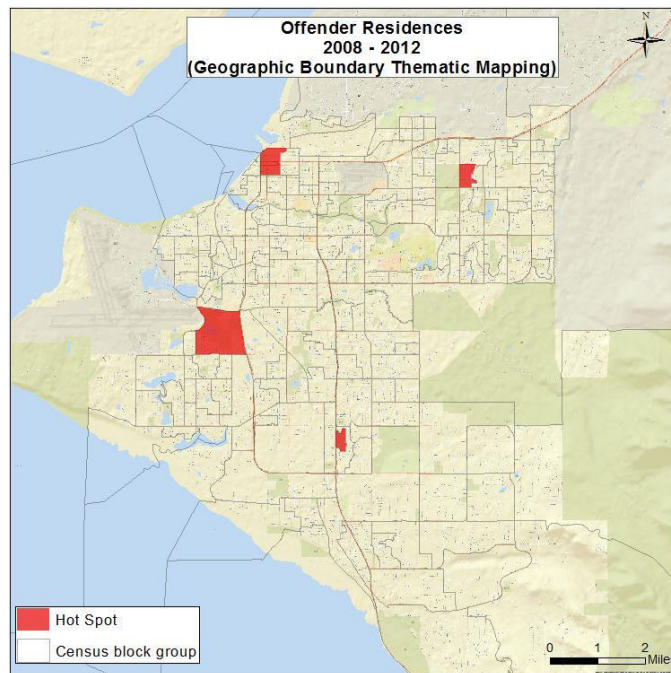
Figure 5.29 represents two census block groups located in down-town Anchorage as hot spots for the time period from 2008 – 2010. Figures 5.30 and 5.31 (2008 – 2012 and 2008 – June 2013) show the same results of crime hot spots. Also two census block groups are representing hot spot areas that are located in the down-town and mid-town areas. It can be agreed, that the last two time periods gained more crime events, in relation to the population, in the mid-town area of Anchorage.

*Results*

Figures 5.32 – 5.34 visualize the spatial distribution of offender residences from cars using the Geographic Boundary Thematic Mapping approach.



*Figure 5.32: GBTM of offender residences from 2008 – 2010*



*Figure 5.33: GBTM of offender residences from 2008 – 2012*

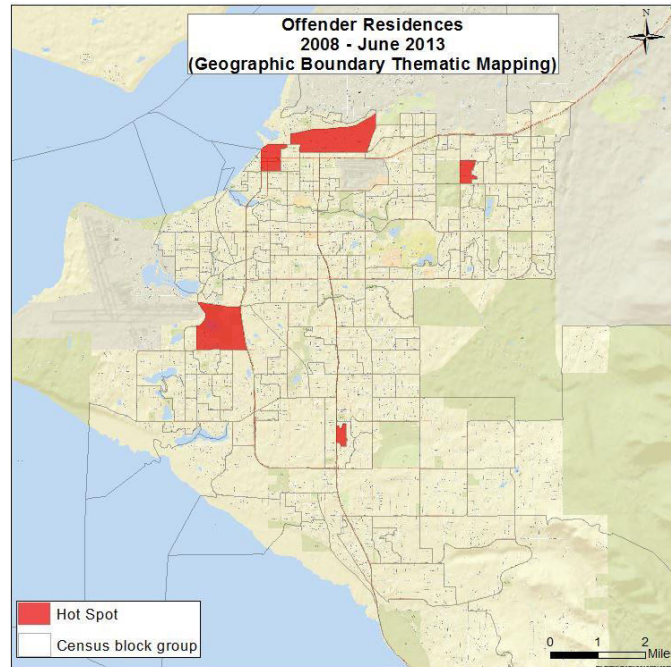


Figure 5.34: GBTM of offender residences from 2008 – June 2013

Figures 5.32 - 5.34 show more hotspots than the corresponding geographic boundary thematic maps of property theft from cars. Figure 5.32 represents five hot spots in the down-town, western part and northern part of Anchorage. When the time period is five years (Figure 5.33), a new hot spot was calculated in the eastern part of Anchorage. For the five and ½ years (Figure 5.34) time period there are changes in hot spots in down-town Anchorage.

### 5.2.2.2 Grid Thematic Mapping (GTM)

Figures 5.35 – 5.37 visualize the spatial distribution of property thefts from cars using the Grid Thematic Mapping approach.

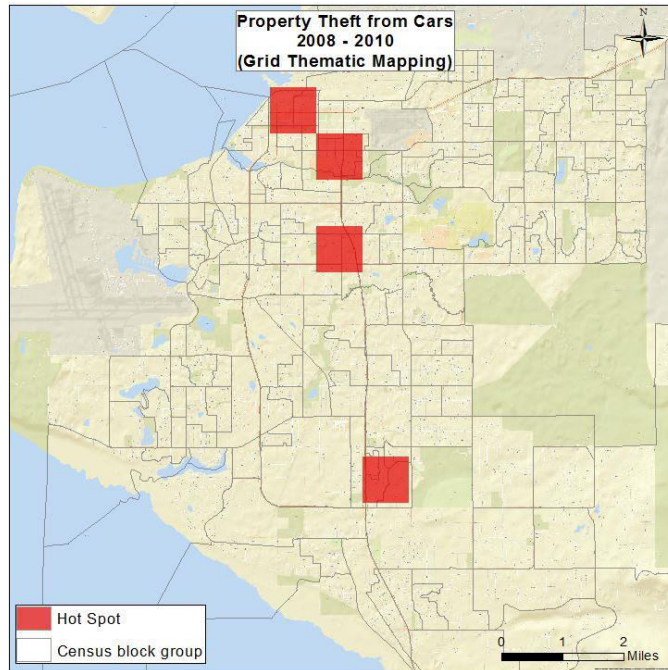


Figure 5.35: GTM of property theft from cars from 2008 – 2010

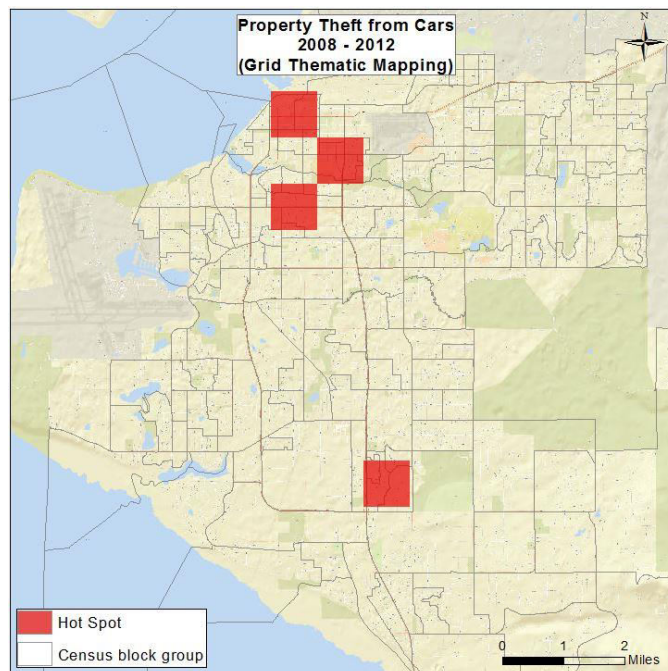


Figure 5.36: GTM of property theft from cars from 2008 – 2012

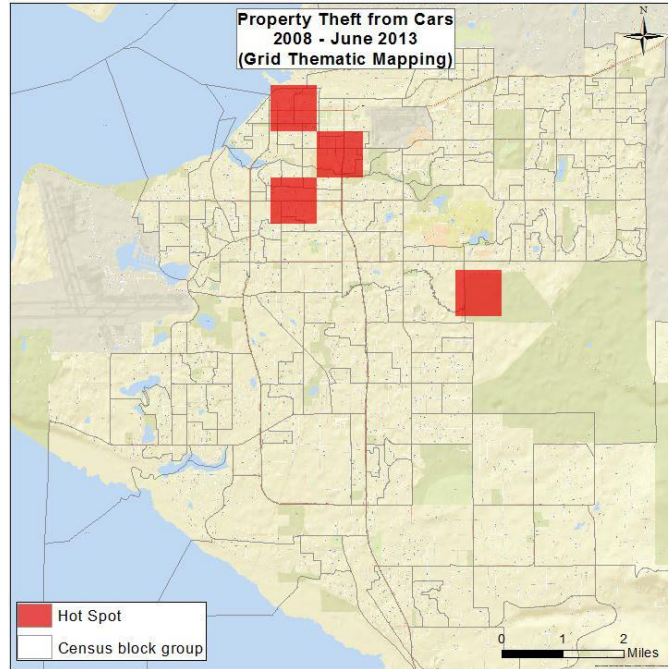


Figure 5.37: GTM of property theft from cars from 2008 – June 2013

All three time periods (Figures 5.35 – 5.37) are showing multiple hot spot grids in downtown and mid-town of Anchorage. A single hot spot grid is visible in the southern or eastern part of Anchorage for each time period.



Results

Figures 5.38 – 5.40 visualize the spatial distribution of offender residences using the Grid Thematic Mapping approach.

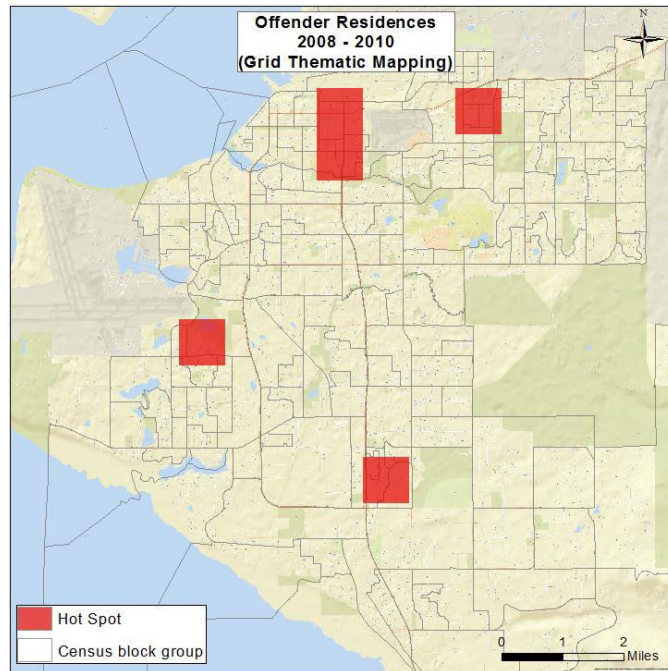


Figure 5.38: GTM of offender residences from 2008 – 2010

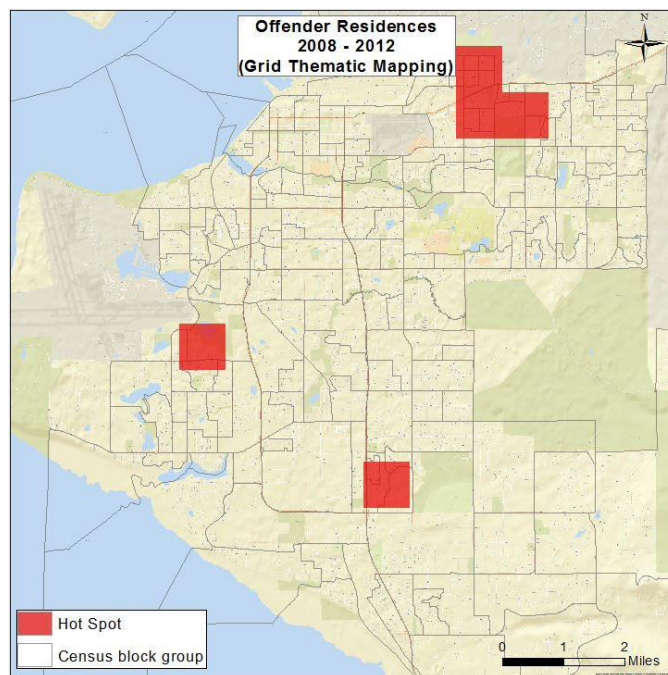


Figure 5.39: GTM of offender residences from 2008 – 2012

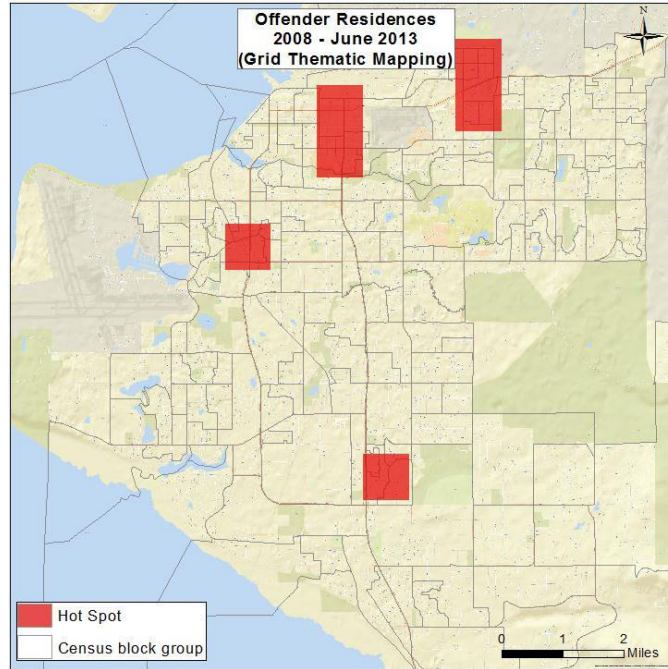


Figure 5.40: GTM of offender residences from 2008 – June 2013

Each of the above maps (Figures 5.38 – 5.40) is representing more hot spots as opposed to the hot spot maps of property theft from cars. Obviously, there are higher numbers offender residence locations within these hot spot areas.

5.2.2.3 Local Moran's I

Figures 5.41 – 5.43 visualize the spatial distribution of property thefts from cars using the Local Moran's I approach.

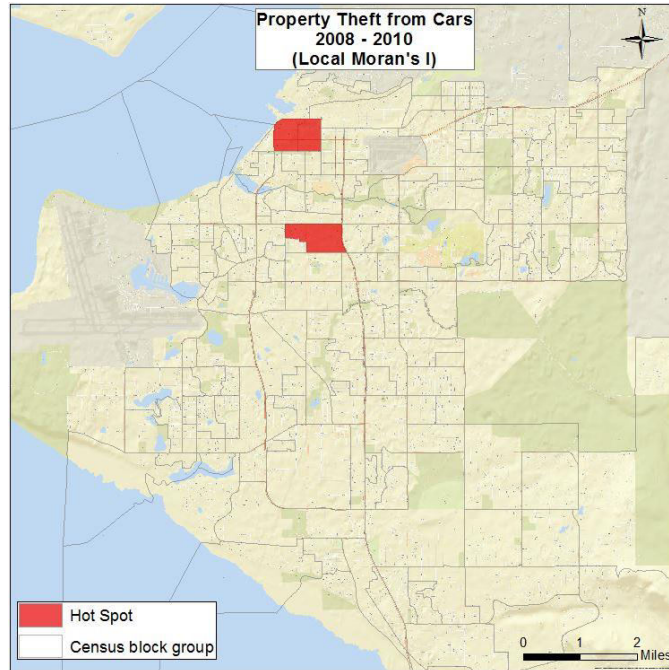


Figure 5.41: Local Moran's I of property theft from cars from 2008 – 2010

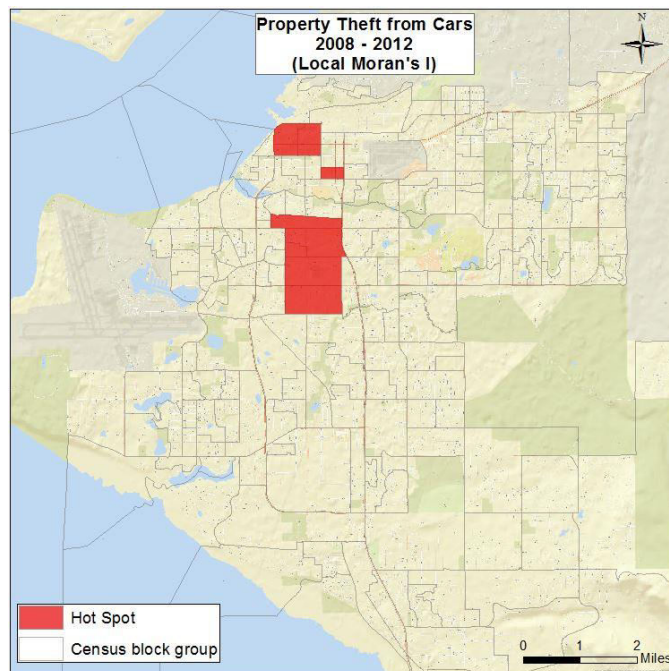


Figure 5.42: Local Moran's I of property theft from cars from 2008 – 2012

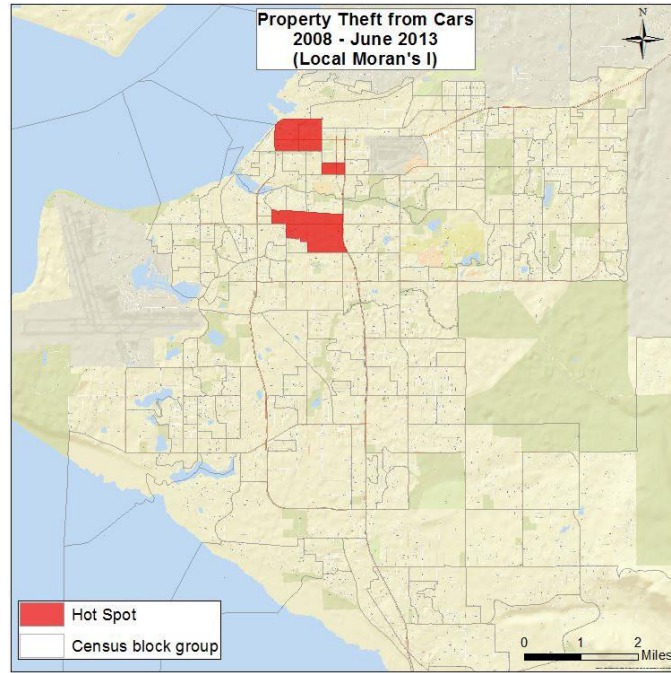
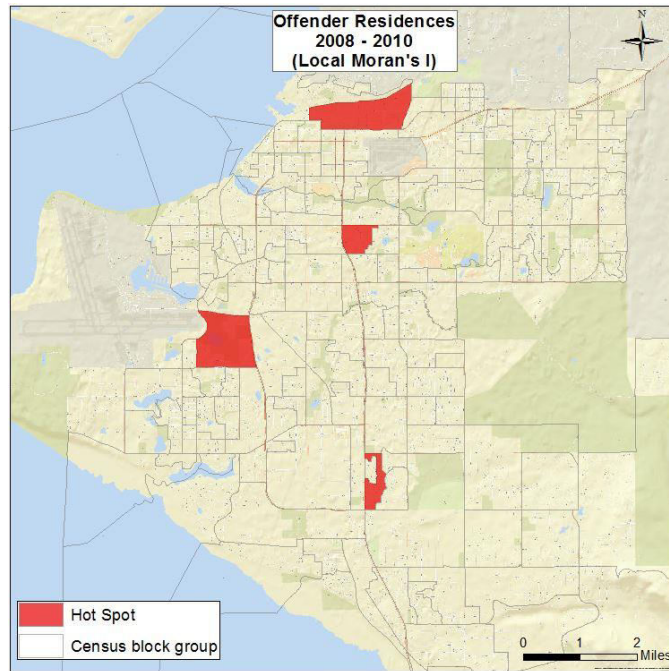


Figure 5.43: Local Moran's I of property theft from cars from 2008 – June 2013

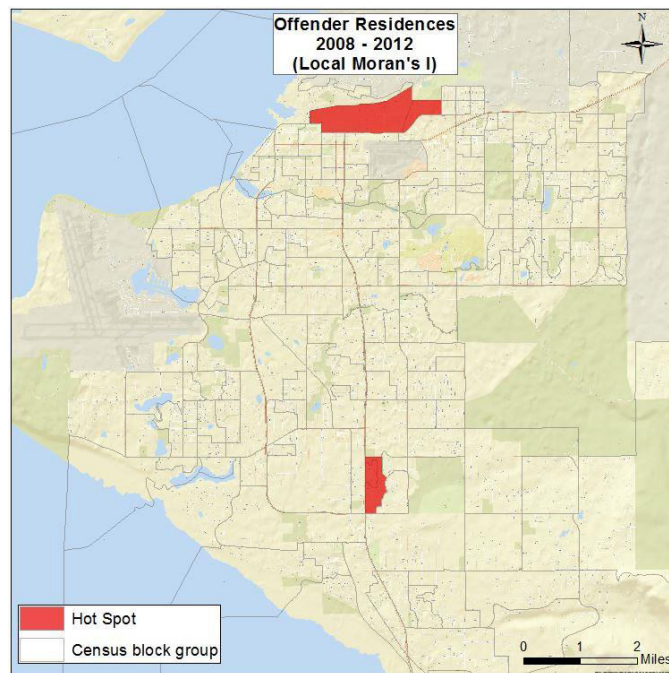
A limited number of hot spots are located in the census block group of down-town and mid-town Anchorage for the three and five year time periods (Figures 5.41 and 5.42). The five and ½ years' time period showed several hot spots in the mid-town area of Anchorage.

*Results*

Figures 5.44 – 5.46 visualize the spatial distribution of offender residences using the Local Moran's I approach.



*Figure 5.44: Local Moran's I of offender residences from 2008 – 2010*



*Figure 5.45: Local Moran's I of offender residences from 2008 – 2012*

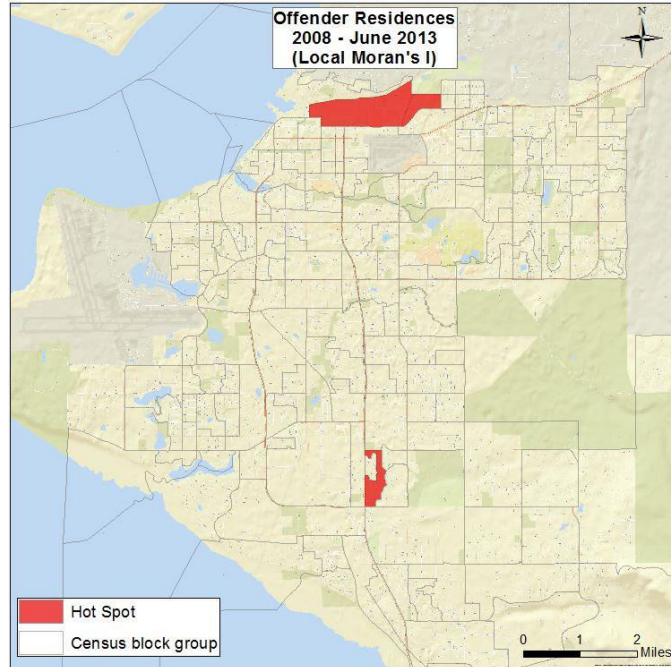


Figure 5.46: Local Moran's I of offender residences from 2008 – June 2013

Figure 5.44 represents hot spots in the northern, southern, and western areas of Anchorage. The five year and five and ½ year time periods (Figures 5.45 and 5.46) show two hot spots in the same location, which are in the northern and southern parts of Anchorage.

5.2.2.4 *Getis and Ord Gi\* Statistics (enumeration units)*

Figures 5.47 – 5.49 visualize the spatial distribution of property thefts from cars using the Gi\* census block group approach.

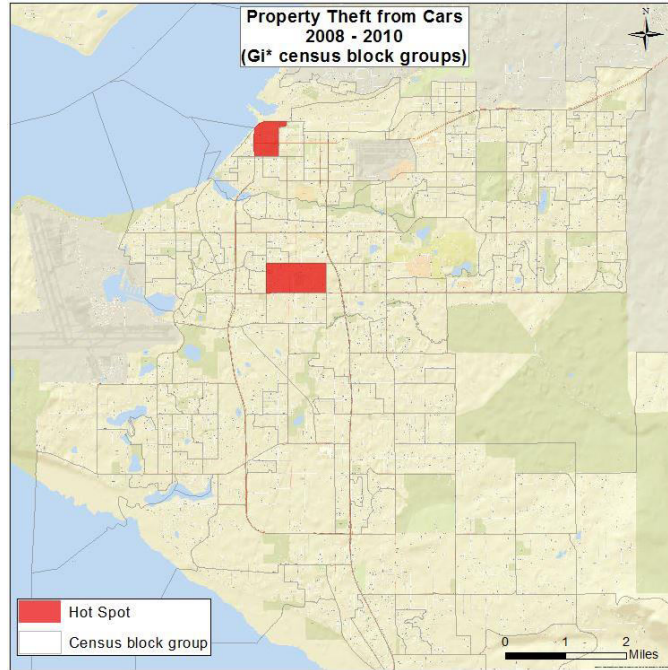


Figure 5.47: *Gi\** census block groups of property theft from cars from 2008 – 2010

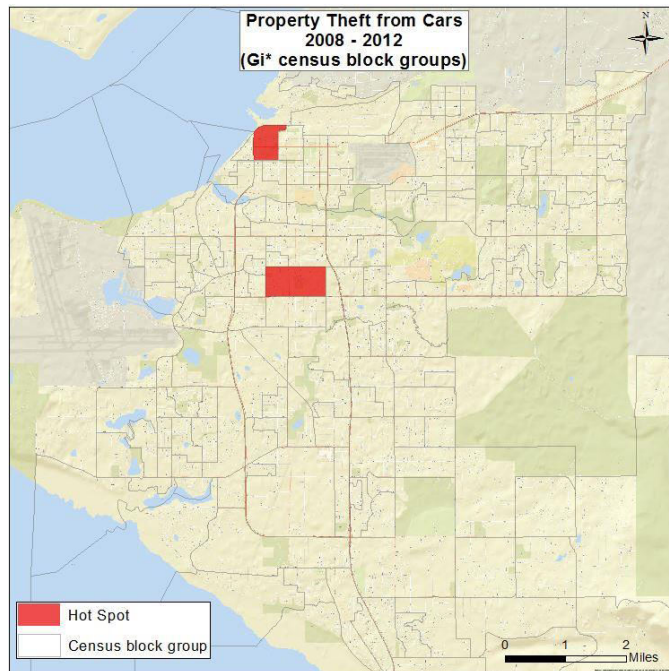


Figure 5.48: *Gi\** census block groups of property theft from cars from 2008 – 2012

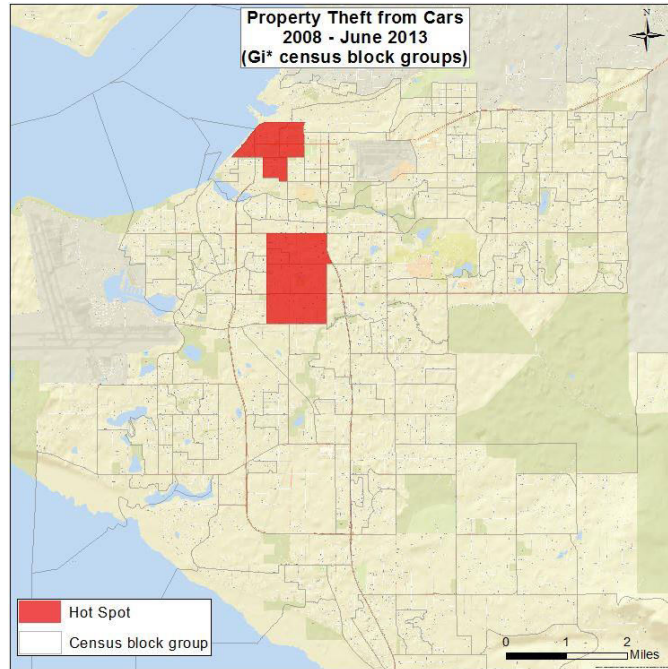


Figure 5.49: *Gi\** census block groups of property theft from cars from 2008 – June 2013

Figures 5.47 and 5.48 are representing two hot spots in the same locations in the downtown and mid-town area of Anchorage. Additional hot spots can be seen in the downtown and mid-town area based on the five and ½ years' time period (Figure 5.49).



Results

Figures 5.50 – 5.52 visualize the spatial distribution of offender residences using the Gi\* census block group approach.

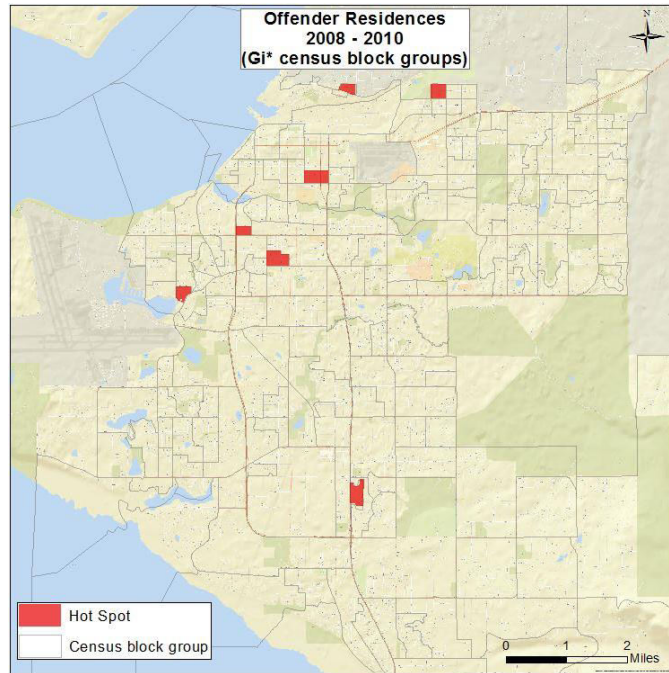


Figure 5.50: Gi\* census block groups of offender residences from 2008 – 2010

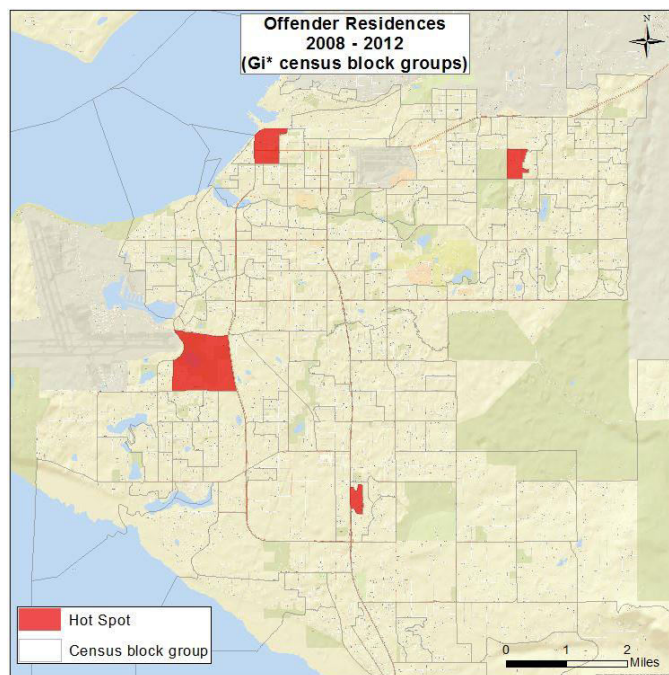


Figure 5.51: Gi\* census block groups of offender residences from 2008 – 2012

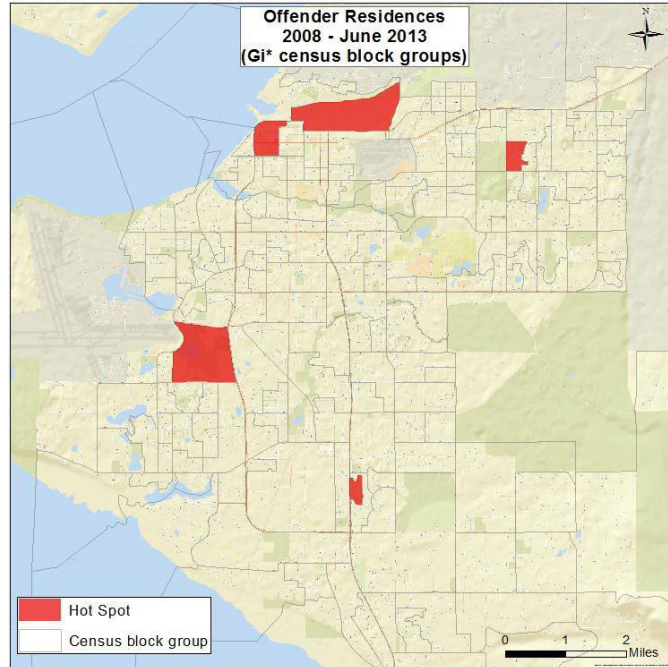


Figure 5.52: *Gi\** census block groups of offender residences from 2008 – June 2013

Several small hot spots are represented in Figure 5.50 located mainly in the northern and north-western parts of Anchorage. Figures 5.51 and 5.51 show a smaller number but bigger hot spots near the same areas as in Figure 5.50.

5.2.2.5 Getis and Ord  $G_i^*$  Statistics (grid)

Figures 5.53 – 5.55 visualize the spatial distribution of property thefts from cars using the  $G_i^*$  grid approach.

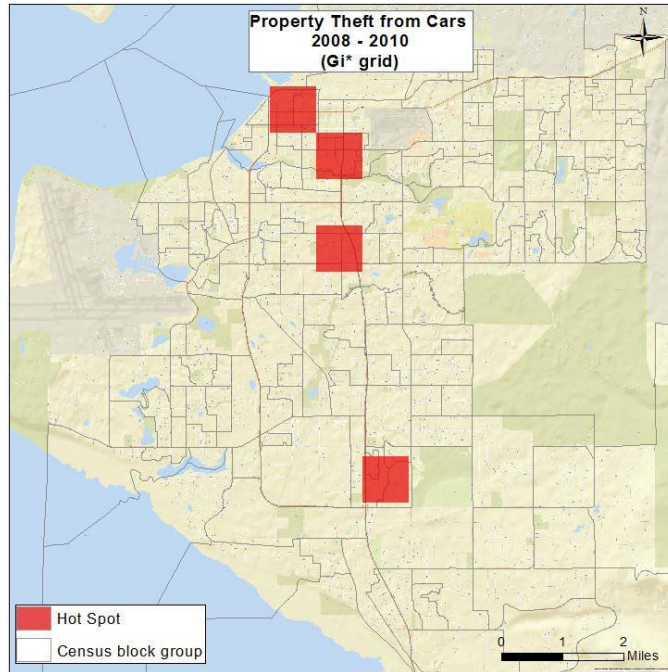


Figure 5.53:  $G_i^*$  grid of property theft from cars from 2008 – 2010

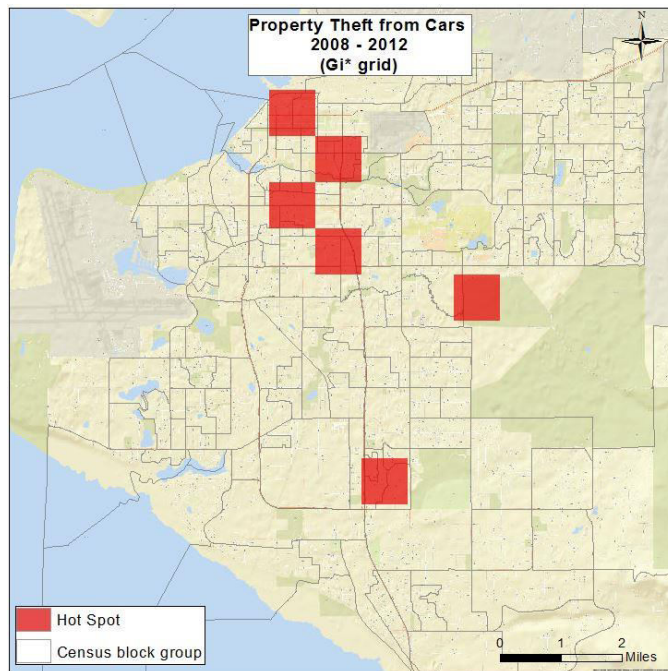


Figure 5.54:  $G_i^*$  grid of property theft from cars from 2008 – 2012

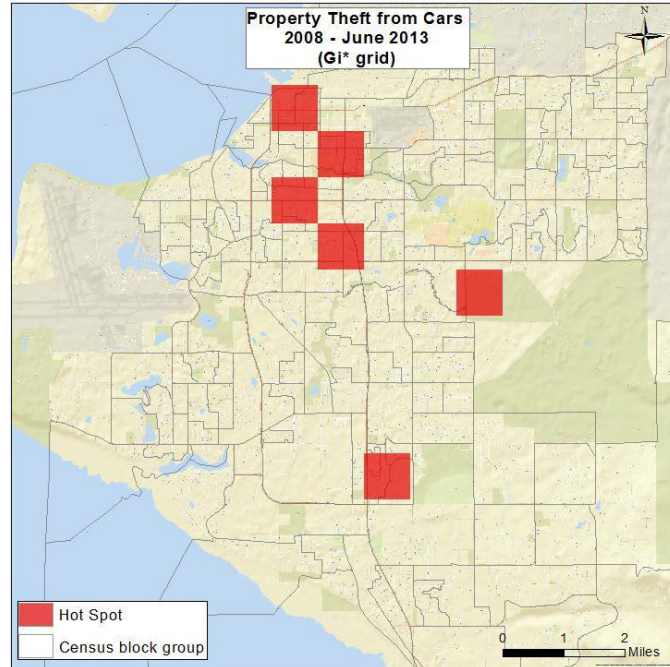


Figure 5.55: Gi\* grid of property theft from cars from 2008 – June 2013

Hot spots exist in the downtown, midtown and northern parts of Anchorage (Figure 5.53). Figures 5.54 and 5.55 show the same hot spot results. When making the time period longer, additional hot spots can be found in the eastern and mid-town areas (Figures 5.54 and 5.55).

Results

Figures 5.56 – 5.58 visualize the spatial distribution of offender residences using the Gi\* grid approach.

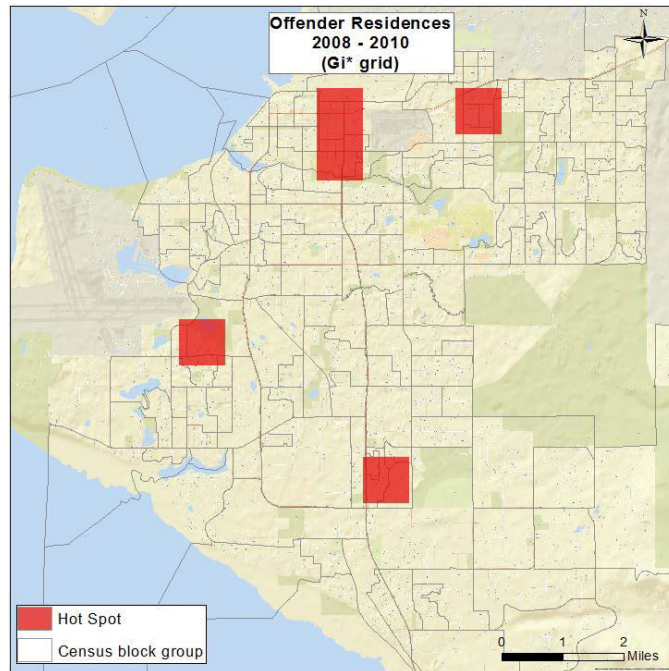


Figure 5.56: Gi\* grid of offender residences from 2008 – 2010

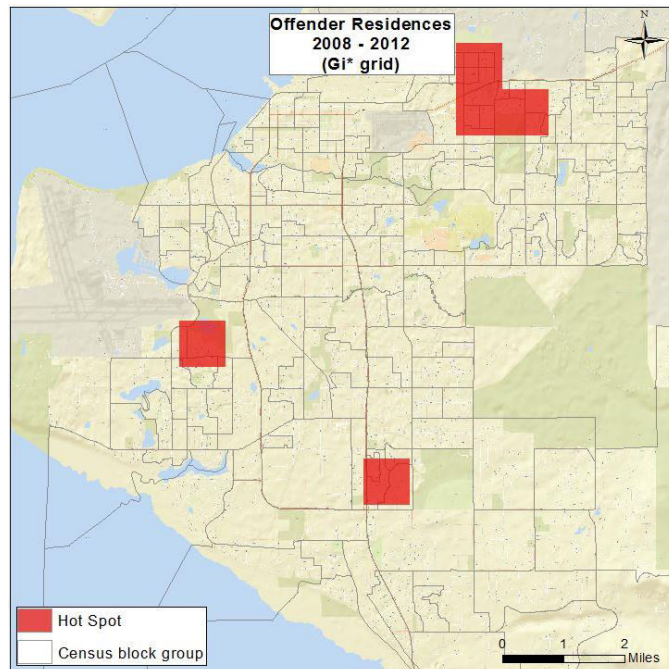


Figure 5.57: Gi\* grid of offender residences from 2008 – 2012

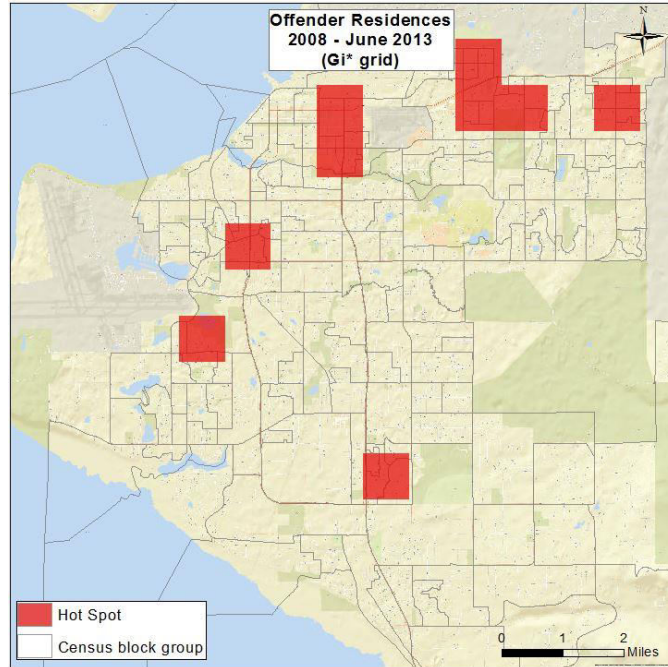


Figure 5.58: *Gi\** grid of offender residences from 2008 – June 2013

Figures 5.56 and 5.57 show the majority of hot spot locations in the western, southern, and north-eastern parts of Anchorage.

The most hot spots can be found for the longest time period which are visualized in Figure 5.58.

### 5.3 Comparison of Prediction Accuracy

The following tables (Tables 5.2 – 5.7) present all hot spot methods, the area size of the created hotspots, the retrospective and prospective point counts and the prediction accuracy results. Furthermore, the best prediction accuracy indexes are represented in bold numbers.

Table 5.2 shows the results of the prediction accuracy calculations of property theft from cars from 2008 – 2010.

Hotspot Method	Property theft from cars 2008 - 2010		Property theft from cars 2011 - 2013		Area (in square kilometer)		Prediction Accuracy		
	In 2008 – 2010 hotspots	In study area	In 2008 – 2010 hotspots	In study area	Area of 2008 – 2010 hotspots	Study area	Hit Rate	PAI	RRI
GBTM	9	154	5	150	1.92	271	3.33	4.70	0.57
GTM	24	154	22	150	5.95	251	14.67	6.19	0.94
Moran's I	19	154	11	150	2.29	271	7.33	<b>8.67</b>	0.59
GI* census block group	9	154	5	150	1.92	271	3.33	4.70	0.57
GI* grid	24	154	22	154	5.96	251	14.66	6.17	0.94
Fuzzy Mode	19	154	17	150	7.89	251	11.33	3.60	0.92
STAC	61	154	40	150	11.02	251	26.66	6.07	0.68
NNH	59	154	41	150	13.65	251	<b>27.33</b>	5.06	0.72
KDE	33	154	31	150	6.84	251	20.67	7.59	<b>0.97</b>

Table 5.2: Evaluation of prediction accuracy for property thefts from cars from 2008 - 2010

The first time period from Jan. 2008 – Dec. 2010 (Table 5.2) contains a total number of 154 thefts of property from cars. The future time period from 2011 – 2013 had a point count of 150 crime events. The total study area size with all aggregated cluster methods was 271 km<sup>2</sup> and 251 km<sup>2</sup> based on point pattern hot spot methods.

The number of retrospective crime events that are within the created hot spot areas of each cluster technique were ranging from 9 to 61 events. The STAC method had the highest count of 61 retrospective crime events falling inside its hot spot areas. The first

important variable to calculate the prediction accuracies are the number of future crime events covered by the retrospective hot spots. These numbers are ranging from 5 to 41 crimes. The NNH cluster technique covered the highest number of future crime events with 41, which resulted in the highest Hit Rate with a value of 27.33. Incorporating the hot spot and study area sizes into the calculation, the local Moran's I performed the best, with a PAI index of 8.67. The values of the RRI results were ranging between 0.57 and 0.97. The KDE had the highest RRI value of 0.97.

Table 5.3 presents the results of the prediction accuracy calculations of property theft from cars from 2008 – 2012.

Hotspot Method	Property theft from cars 2008 – 2012		Property theft from cars 2013		Area (in square kilometer)		Predictive accuracy		
	In 2008 – 2012 hotspots	In study area	In 2008 – 2012 hotspots	In study area	Area of 2008 – 2012 hotspots	Study area	Hit rate (in %)	PAI	RRI
GBTM	12	253	1	51	1.92	271	1.96	2.77	0.42
GTM	40	253	7	51	5.95	251	13.73	5.79	<b>0.88</b>
Moran's I	44	253	6	51	5.99	271	11.76	5.32	0.68
Gi* census block group	12	253	2	51	1.92	271	3.92	5.53	0.83
Gi* grid	54	253	9	51	8.92	251	<b>17.65</b>	4.96	0.83
Fuzzy Mode	24	253	2	51	3.99	251	3.92	2.46	0.42
STAC	67	253	7	51	6.44	251	13.73	5.35	0.52
NNH	81	253	9	51	6.58	251	<b>17.65</b>	6.73	0.56
KDE	63	253	9	51	6.50	251	<b>17.65</b>	<b>6.81</b>	0.71

Table 5.3: Evaluation of prediction accuracy for property thefts from cars from 2008 – 2012

Table 5.3 shows the second time period from Jan. 2008 – Dec. 2012 with a total number of 253 retrospective theft events and a total number of 51 future events. The total study area sizes of both categories of cluster methods was the same as in the first time span (Table 5.2). By increasing the length of the time period more points were added to the retrospective crime/offender residence counts. Therefore, the prospective crime counts included fewer numbers as opposed to the future point counts of the first time period from 2008 – 2010. The number of retrospective crime events that are fall inside the



Results

created hot spot areas were ranging from 12 to 81 events. The hot spot related point counts of future crime events varied between 1 and 9. The NNH cluster technique contained the most number of retrospective points (81 events) inside its retrospective hot spots as well as the highest point count (9 events) of future crimes together with the KDE and the Gi\* grid based method. Thus, NNH cluster, the Gi\* grid, and the KDE had the highest Hit Rate with a value 17.65. Including the hot spot and study area sizes into the calculation, the KDE method showed the highest PAI value of 6.81. The results of the RRI were ranging from 0.42 to 0.88. The GTM technique had the highest RRI index of 0.88. The Fuzzy Mode, GBTM, and GI\* areal methods show very low numbers of prospective crime events within their created retrospective hot spot areas. This could be an indicator that these methods will fail to calculate the prediction accuracy indexes in the following third time period.

Table 5.4 shows the results of the prediction accuracy calculations of property theft from cars from 2008 – June 2013.

Hotspot Method	Property theft from cars 2008 – June 2013		Property theft from cars July 2013 – Dec. 2013		Area (in square kilometer)		Predictive accuracy		
	In 2008 – June 2013 hotspots	In study area	In 2008 – June 2013 hotspots	In study area	Area of 2008 – June 2013 hotspots	Study area	Hit Rate (in %)	PAI	RRI
GBTM	14	282	0	22	1.92	271	0	0	0
GTM	46	282	2	22	5.95	251	<b>9.09</b>	3.83	0.54
Moran's I	34	282	2	22	3.19	271	<b>9.09</b>	<b>7.72</b>	<b>0.74</b>
Gi* census block group	14	282	0	22	1.92	271	0	0	0
Gi* grid	62	282	2	22	8.92	251	<b>9.09</b>	2.56	0.4
Fuzzy Mode	24	282	0	22	3.37	251	0	0	0
STAC	94	282	2	22	9.29	251	<b>9.09</b>	2.46	0.27
NNH	82	282	2	22	5.31	251	<b>9.09</b>	4.29	0.30
KDE	75	282	2	22	7.51	251	<b>9.09</b>	3.03	0.33

Table 5.4: Evaluation of prediction accuracy for property thefts from cars from 2008 – June 2013

## *Results*

The longest time period from Jan. 2008 - June 2013 is shown in Table 5.4 and represents a total number of 282 crime events from the past and only a point count of 22 events from the future. The total study area size was the same as in the other time periods. The retrospective point counts within hot spots were ranging from 14 to 94 crime events. The hot spots created from the STAC method contained the most points of past crime data with a total number of 94 events. The number of future crime events covered by the created hot spots ranged from 0 to 2 crimes. The NNH, KDE, STAC, Gi\* grid, local Moran's I and GTM technique covered the highest number of crime events, namely 2, from the "future" and thereby had the best Hit Rate result with values of 9.09. Including the hot spot and study area sizes into the calculation, the local Moran's I had the best PAI result with an index of 7.72. The values of the RRI results varied between 0.00 and 0.74. Local Moran's I had the highest RRI value of 0.74.

Since there were only a low number of points left for the future data, the GBTM, GI\*areal and Fuzzy Mode methods did not have any points within their retrospectively created hot spots. These methods had also a small size of hot spot areas, resulting in a low number of future crime events falling inside those small hot spot areas. Therefore, these methods failed to calculate different prediction accuracy indexes based on the limited number of points from the prospective data.

Table 5.5 presents the results of the prediction accuracy calculations of offender residences from 2008 – 2010.

Hotspot Method	Offender residence 2008 - 2010		Offender residence 2011 - 2013		Area (in square kilometer)		Predictive accuracy		
	In 2008 – 2010 hotspots	In study area	In 2008 – 2010 hotspots	In study area	Area of 2008 – 2010 hotspots	Study area	Hit rate (in %)	PAI	RRI
GBTM	19	150	9	151	5.60	242	5.96	2.57	0.49
GTM	29	150	26	151	7.43	203	<b>17.22</b>	4.70	0.89
Moran's I	14	150	9	151	5.69	242	5.96	2.53	0.64
Gi* census block group	12	150	8	151	1.13	242	5.30	<b>11.35</b>	0.66
Gi* grid	29	150	26	151	7.43	203	<b>17.22</b>	4.70	0.89
Fuzzy Mode	12	150	12	151	4.67	203	7.95	3.46	<b>0.99</b>
STAC	51	150	25	151	10.17	203	16.56	3.31	0.49
NNH	34	150	12	151	4.72	203	7.95	3.42	0.35
KDE	40	150	26	151	6.03	203	<b>17.22</b>	5.80	0.64

Table 5.5: Evaluation of prediction accuracy for offender residences from 2008 – 2010

Table 5.5 shows the first time period from Jan. 2008 – Dec. 2010 of offender residence locations with a point count of 150 residences as retrospective data and 151 residence locations as future data. The total study area size was 242 km<sup>2</sup> and 203 km<sup>2</sup> for the point data. The number of retrospective residence locations that are within the created hot spot areas of each cluster technique were ranging from 12 to 40 locations. The STAC technique had the highest retrospective residence counts of 51 locations within its created hot spots. The hot spot related point count of future offender residence locations varied between 8 and 26. The most number of points were covered from clusters that were created from the KDE, Gi\* grid, and GTM methods with a value of 26 points. Thus, the GTM and KDE had the highest Hit Rate with a value 17.22. Comparing the hot spot area sizes of the different cluster techniques, the GI\* had the lowest area size of 1.13 km<sup>2</sup> and eight residence locations falling into the retrospective hot spots, in order to obtain the highest PAI with a value of 11.35. The results of the RRI had a value range from 0.35 to 0.99. The Spatial Fuzzy Mode covered the same amount of “past” and “future” residence locations within its created hot spots and therefore gained the highest RRI value of 0.99. Table 5.6 presents the results of the prediction accuracy calculations of offender residences from 2008 – 2012.

Hotspot Method	Offender residence 2008 - 2012		Offender residence 2013		Area (in square kilometer)		Predictive accuracy		
	In 2008 – 2012 hotspots	In study area	In 2008 – 2012 hotspots	In study area	Area of 2008 – 2012 hotspots	Study area	Hit rate (in %)	PAI	RRI
GBTM	22	250	1	51	3.43	242	1.96	1.38	0.23
GTM	46	250	5	51	7.43	228	9.80	3.01	0.64
Moran's I	18	250	4	51	3.52	242	7.84	5.39	<b>1.09</b>
Gi* census block group	22	250	2	51	3.43	242	3.92	2.76	0.45
Gi* grid	46	250	5	51	7.43	228	9.80	3.01	0.64
Fuzzy Mode	17	250	2	51	2.99	228	3.92	2.99	0.59
STAC	81	250	5	51	9.69	228	9.8	4.04	0.31
NNH	49	250	3	51	3.16	228	5.88	4.24	0.31
KDE	54	250	7	51	4.61	228	<b>13.72</b>	<b>6.79</b>	0.65

Table 5.6: Evaluation of prediction accuracy for offender residences from 2008 - 2012

Table 5.6 shows the second time period from Jan. 2008 – Dec. 2012 with a total number of 250 retrospective offender residences and a total number of 51 future residence locations. The total study area sizes for point methods increased to 228 km<sup>2</sup> but stayed the same with aggregated methods, similar to the first time span (Table 5.5). By increasing the time period more points were added to the total number of retrospective locations. The STAC technique contained the most number of points, namely 81, in the retrospective hot spots. The point counts within these hot spots were ranging from 17 to 81 offender residences. The prospective crime counts showed fewer numbers as opposed to the prospective point counts of the first time period from 2008 – 2010 (Table 5.5) and varied between 1 and 7 residences. The KDE covered the most locations of prospective residence locations and yielded the highest RRI value of 13.72. Including the hot spot and study area sizes into the calculation, the PAI results were ranging between 1.38 and 6.79. The KDE method had the best PAI result with an index of 6.79. The results of the RRI had a value range from 0.23 to 1.09 led by local Moran's I with the highest score of 1.09. Similar to Table 5.5, it can be seen that the GBTM method covered the lowest number of prospective residence locations within its retrospective hot spot areas. This could be also

an indicator that this method would fail to calculate the prediction accuracy indexes in the following third time period.

Table 5.7 shows the results of the prediction accuracy calculations of offender residences from 2008 – June 2013.

Hotspot Method	Offender residence 2008 – June 2013		Offender residence July 2013 – Dec. 2013		Area (in square kilometer)		Predictive accuracy		
	In 2008 – June 2013 hotspots	In study area	In 2008 – June 2013 hotspots	In study area	Area of 2008 – June 2013 hotspots	Study area	Hit rate (in %)	PAI	RRI
GBTM	27	278	0	23	5.74	242	0	0	0
GTM	62	278	3	23	8.91	228	13.04	3.34	0.60
Moran’s I	15	278	1	23	3.31	242	4.35	3.18	0.83
Gi* census block group	27	278	1	23	5.69	242	4.35	1.85	0.46
Gi* grid	86	278	6	23	13.38	228	<b>26.09</b>	<b>4.44</b>	<b>0.87</b>
Fuzzy Mode	20	278	0	23	2.45	228	0	0	0
STAC	116	278	4	23	12.77	228	17.39	3.1	0.43
NNH	54	278	1	23	2.69	228	4.35	3.69	0.23
KDE	68	278	2	23	5.07	228	8.70	3.91	0.37

Table 5.7: Evaluation of prediction accuracy for offender residences from 2008 – June 2013

This is the third and longest time span of offender residence locations from Jan. 2008 – June 2013 (Table 5.7) represents a total number of 278 retrospective locations and a small point count of 23 prospective offender residence locations. The study area sizes remained the same as in the second time period (Table 5.6). The number of retrospective residence locations that were within the retrospective hot spot areas of each cluster technique were ranging from 15 to 116 points. The most point counts of 116 residences from the retrospective residence locations were covered by the respective hot spots created from the STAC algorithm. The Gi\* grid based method showed the highest Hit Rate, PAI, and RRI values in this particular time period. It can be also seen that the GBTM and Spatial Fuzzy Mode are two methods that could not cover future point data within their hot spots. Thus, these methods failed to calculate prediction accuracy indexes due to the limited point counts of their prospective data.

### 5.4 Summary of Prediction Accuracy

The following tables (Tables 5.8 – 5.12) summarize the best hot spot methods and crime data sets based on their prediction accuracy.

<b>Hit Rate</b>		
Time period / crime type	Highest Index (by time period)	Highest Index (by crime and offender residence data set)
Property theft from cars 2008 - 2010	NNH	NNH
Property theft from cars 2008 –2012	KDE Gi* grid NNH	
Property theft from cars 2008 – June 2013	KDE Gi* grid GTM Local Moran’s I STAC NNH	
Offender residence 2008 - 2010	KDE Gi* grid GTM	Gi* grid
Offender residence 2008 - 2012	KDE	
Offender residence 2008 – June 2013	Gi* grid	

Table 5.8: Highest Hit Rate for property thefts from cars and their offender residence locations

The best performing hot spot techniques based on their Hit Rate results are shown in Table 5.8 separated into the different time periods. Overall, the KDE and the Gi\* grid based method had the most numbers of highest hit rate followed by the NNH method. Due to the low number of points in the prospective data set, the time period from 2008 – June 2013 for property theft from cars included five different hot spots methods with the best Hit Rate results.

After totaling each data set, the NNH technique gained the highest Hit Rate index related to the crime type and the Gi\* grid obtained the highest index related to offender residence locations.

PAI		
Time period / point event	Highest Index (by time period)	Highest Index (by crime and offender residence data set)
Property theft from cars 2008 - 2010	Local Moran's	Local Moran's I
Property theft from cars 2008 - 2012	KDE	
Property theft from cars 2008 - June 2013	Local Moran's I	
Offender residence 2008 - 2010	Gi* census block group	Gi* census block group
Offender residence 2008 - 2012	KDE	
Offender residence 2008 - June 2013	Gi* grid	

Table 5.9: Highest PAI for property thefts from cars and their offender residence locations

Table 5.9 shows hot spot methods with the highest PAI related to different time periods and the two crime data sets. The Local Moran's I had the best results for two time periods for the property theft from cars, followed by the KDE. The Gi\* census block group, the KDE, and the Gi\* grid resulted in the highest PAI indexes in each time period for the offender residence locations.

Combining all time periods for each data set, the Local Moran's I had the highest PAI related to property theft from cars and the Gi\* census block group technique had the highest PAI related to offender residence locations.

RRI		
Time period / point event	Highest Index (by time period)	Highest Index (by crime and offender residence data set)
Property theft from cars 2008 - 2010	KDE	KDE
Property theft from cars 2008 –2012	GTM	
Property theft from cars 2008 – June 2013	Local Moran’s I	
Offender residence 2008 - 2010	Spatial Fuzzy Mode	Local Moran’s I
Offender residence 2008 - 2012	Local Moran’s I	
Offender residence 2008 – June 2013	Gi* grid	

Table 5.10: Highest RRI for property thefts from cars and their offender residence locations

In Table 5.10, the most number of highest RRI indexes related to the different time periods for property theft from cars were shared by the KDE, the GTM, and the Local Moran’s I methods. The spatial Fuzzy Mode, local Moran’s I and Gi\* grid techniques had the most numbers across the time periods related to the offender residence locations data set. Combining the time periods for each crime data set, the KDE had the highest RRI result related to property theft from cars and the local Moran’s method had the highest RRI related to offender residence locations.

Table 5.11 presents the average of all prediction index results related to each of the two crime data sets.

Prediction Accuracy Indexes	Property Theft from Cars	Offender Residence Locations
Hit Rate on average	10.59	9.12
PAI on average	4.53	3.67
RRI on average	0.57	0.55

Table 5.11: Highest prediction accuracies on average for property theft from cars and offender residence locations



*Results*

The property theft from cars data set achieved better results compared to the offender residence locations data set.

Table 5.12 shows those hot spot methods that failed to calculate the different prediction accuracy indexes.

<b>Failing hot spot methods</b>	
Point event	Hotspot Method
Property theft from cars	GBTM Gi* census block group Spatial Fuzzy Mode
Offender residence	GBTM Spatial Fuzzy Mode

*Table 5.12: Summary of failing methods*

It can be seen that these methods failed due to the low number of crime events in the shortest prospective time period for property theft from cars and offender residence locations. The GBTM and the spatial Fuzzy Mode techniques failed most often, followed by the Gi\* census block group method.

Summarized, the KDE method had the most count of highest prediction accuracy indexes than all other methods and the Gi\* grid had the highest averages of all prediction accuracy results throughout all time periods and across both crime data sets.

## CHAPTER 6      DISCUSSION

This chapter contains a discussion of the applied hot spot methods and addresses limitations based on the included data and executed methodologies. Furthermore, the results of the prediction accuracy indexes are discussed. This chapter closes with recommendations for future research topics.

### 6.1    Hot Spot Mapping

This research analyzed nine different cluster methods to detect hot spots of crime events and their related offender residences for making a prediction accuracy comparison between the hot spot methods in the city of Anchorage, USA. Before analyzing the hot spot and prediction accuracy methods, several pre-processing steps were executed to prepare the base data for the implementation of these algorithms. The first step was to split each of the data sets into three different time periods to determine the retrospective and prospective time span. Interestingly, since these data only included a limited number of points, a new research sub-topic occurred, namely how do the examined cluster methods and prediction accuracy indexes perform with very small numbers of prospective points. The next step was to aggregate the point data into areal units such as census block groups and user defined grid cells since the examined hot spot techniques included different types of base data. Furthermore, the aggregated crime events were separated into crime and offender residence rates.

The first statistical process included the determination if and how the different point data sets are distributed across the study area. In other words, it was found out whether the data show cluster characteristics and where these clusters are concentrated. It is agreed among researchers that crime data tend to be higher concentrated in different parts of the study area. To the best of my knowledge, no research described so far investigated if offender residence data share similar clustering behavior. After applying a NNI calculation for both crime data sets (crime and offender residence locations) and different time periods, the results showed that the offender residence data have a tendency of a higher concentration of point distributions throughout the different time periods as opposed to the crime location events. This can be evidence for strong underlying causes

to justify these higher concentrations of residence locations in areas with high poverty or unemployment rates.

After examining the descriptive statistic results, different hot spot methods were reviewed based on their parameter settings. One of the most crucial hot spot parameters is the bandwidth value, also referred to as the search radius. This research used the k-order mean nearest neighbor distance result for each time period and data set to calculate the bandwidth values. It was to be expected that the bandwidth values would be very high, since each data set contained only a low number of crime events compared to the study area size. Indeed, the results showed very high means of nearest neighbor distance values. Another critical parameter, especially for hierarchical methods, such as the NNH clustering method, is the definition of the minimum number of points per cluster. Its determination can be quite a challenge for an unexperienced user, since there are no published guidelines to determine an appropriate number of points to define a cluster.

Other methods, such as point pattern techniques require the user to define hot spots based on a thematic threshold. Several suggestions were made by different researchers and after experimenting with different classification schemes, the standard deviation classification was chosen for all cluster methods except for the hierarchical techniques, which were defined by their resulting first order clusters. Furthermore, the definition of the resulting output format was only necessary to be determined for hierarchical techniques. The choice was between standard deviational ellipses and convex hulls as output formats. Convex hulls were chosen because of their more detailed representation of the shape and size of hot spot areas. Basically, all parameter settings of the applied hot spot methods are more or less dependent on the level of the user's experience. This research also tried to apply the same parameter settings for the bandwidth selection and thematic thresholds whenever applicable throughout the different time periods and data sets.

The relatively low number and distribution of crime events made it difficult to create useful hot spots without losing the requirements for prediction accuracy calculations. In other words, the size and number of hot spot clusters are important variables for these calculations. This was done by a lot of experimenting and repetitive processes to keep a balance between the visual output and the resulting variables, which are being included into further calculations. A typical example was the GTM technique and the  $G_i^*$  grid based method, which were defined by the size of the grid cells and their point count distributions across the study area.

Suggestions were already made by different researchers to define the appropriate size of each grid cell. The size of the cells had to be big enough to include enough points that are within their boundaries. Unfortunately, the suggested methods did not aggregate enough points to identify useful hot spots based on thematic thresholds throughout the different time periods. This was somehow predictable, since the data did not have many crime events compared to most other research published with on a similar topic. As a result, the cell size of the GTM and the  $G_i^*$  grid method was bigger but had enough points within their boundaries to define useful thematic thresholds. Other methods, like the GBTM and the Local Moran's I, included point counts aggregated into areal units, such as census block groups. Since these areas are modifiable, the examination of the underlying data can be misleading when areal units of different scale and shape are being analyzed. This issue has been widely discussed and is referred to as the MAU problem (Openshaw, 1984).

At least one method showed limitations based on its algorithm. The Fuzzy Mode method included also points from overlapping neighboring circles into its algorithm. This had an influence on the resulting frequencies, ranking, and the choice of a useful thematic threshold. Therefore, only the highest rank, which included only one hot spot area, was chosen for each time span and data set to proceed with further prediction calculations.

After executing several hot spot methods and reviewing the resulting hot spot maps, several spatial analysis techniques were applied to obtain the required variables for the three different prediction accuracy index calculations. These spatial analysis results such as the point counts of retrospective and prospective data, the size of the hot spot areas, and the whole study area were presented in form of a table.

## 6.2 Prediction Accuracy

After reviewing all tables only listing the spatial analysis results, it was quite obvious that several hot spot methods would not calculate a prediction accuracy index. The total count of prospective points for the shortest time period for each of the two crime data sets included only 22 or 23 points. It was clear that at least several methods would not cover any points inside their created hot spots.

All three prediction accuracy calculations had the same number of prospective points within the created hot spot areas and the total count of prospective points as their common variables. For example, the Hit Rate includes the above mentioned variables

and the PAI includes the Hit Rate result and the size of the hot spot areas in relation to the whole study area size. The third prediction accuracy index (RRI) uses all point counts of the prospective and retrospective time periods in its calculation. The results of all indexes are also added to the table of the spatial analysis results for further analysis.

Before describing the hot spot method with the best prediction accuracy indexes, the examination of the tables provided some other interesting findings. Overall, hot spots for property theft from cars covered more prospective points throughout all time periods, whereas offender residence hot spots covered more prospective locations, especially in the last and longest retrospective time period.

Comparing the categories of different hot spot methods, the point pattern methods, including the hierarchical methods, showed overall the best performance, ahead of the grid-based and the areal unit based methods. Even the grid based  $G_i^*$  technique had in most cases higher prediction accuracy indexes than the areal unit  $G_i^*$  method. It was also interesting to find out which data set would gain better results related to the average of all prediction accuracy indexes throughout all time periods. Findings showed that property theft from cars achieved a better overall prediction accuracy compared to offender residence locations. This research also concluded that the NNH cluster method had the highest Hit Rate related to the property theft from cars and the  $G_i^*$  grid based technique had the best performance related to offender residence locations. The Local Moran's I achieved the best PAI results for property theft from cars, whereas the areal unit based  $G_i^*$  method outperformed all other cluster technique related to offender residence data. The last prediction accuracy index calculated in this research was the Recapture Rate Index (RRI). The KDE method showed the best results for the crime type and the Local Moran's I technique had the best RRI for the offender residence data set.

Another finding showed that small numbers of prospective data had the disadvantage that multiple hot spot methods were tied for the highest indexes, which is not useful in terms of comparing different hot spot methods with each other. Additionally, three cluster methods, such as the spatial Fuzzy Mode and the GBTM failed to calculate accuracy indexes for the third and longest retrospective time periods caused by low events of missing prospective data.

Overall, the KDE method achieved the most numbers of highest indexes.

On the other hand, the  $G_i^*$  grid based method achieved the best prediction accuracy indexes on average, throughout all time periods and data sets.

### 6.3 Recommendation

One recommendation would be to analyze how offender residence locations perform with additional different crime types included, since this research included only one type of crime. Another research study could include socio-economic correlations of offender residence locations using different regression models. The main purpose would be to identify possible causes for the spatial distribution of offender residence locations.

## CHAPTER 7 CONCLUSION

The main objective of this research was to compare the prediction accuracy indexes of several hot spot methods related to one crime type and offender residence locations of that crime type using different time periods. Finding the best hot spot method with the highest prediction accuracy indexes was not a new research topic, since most previous similar research studies were based on several different crime types. Instead, this study included solved property theft from cars and their related offender residence locations into the cluster and prediction accuracy calculations. Thus, it was expected that one data set will have higher overall prediction accuracy indexes compared to the other data set.

First, several preliminary global statistical techniques were applied to observe the general distribution of the point datasets that were divided into three different time periods. These periods included retrospective and the related prospective data for the calculation of the hot spot areas and the prediction indexes. Different parameter settings were examined and applied based on the complexity of different cluster techniques and their point distributions. The results of several spatial analysis techniques were then included into the prediction accuracy calculations. Besides addressing the main research objective, the results uncovered also some other interesting findings. Point pattern hot spot methods, including hierarchical techniques, showed a better performance compared to the aggregated grid cell and cluster methods based on enumeration units (e.g. census block groups). Since the number of crime events decreased for the prospective data set by extending the retrospective time periods, several hot spot methods failed to calculate the prediction accuracy indexes or had the same prediction accuracy indexes. A meaningful comparison of different hot spots across the different time periods was thus not possible.

Addressing the research objectives, most past similar research concluded that the KDE method has the best overall prediction accuracy performance. The results of this thesis confirmed that the KDE method outperforms all other methods for most of the different time periods for the specific crime type that was applied. However, the grid based Gi\* method showed the best prediction accuracy performance related to the offender residence data pattern.

### *Conclusion*

Comparing the two data sets, the crime event data set showed the highest prediction accuracy indexes, on average. By summarizing the best indexes throughout all time periods and data sets, the KDE method had a larger proportion of highest indexes than all other methods.





## LIST OF REFERENCES

- Anselin, L. (1995) 'Local Indicators of Spatial Association—LISA', *Geographical Analysis*, 27(2), 93-115.
- Boba, R. (2005) *Crime Analysis and Crime Mapping*, SAGE.
- Brimicombe, A. J. (2004) *On Being More Robust About 'Hot Spots'*, translated by Boston, Massachusetts.
- Chainey, S. (2008) 'Spatial significance hotspot mapping using the Gi\* statistic', [online], available: <http://www.popcenter.org/conference/conferencepapers/2010/Chainey-Gi-hotSpots.pdf> [accessed
- Chainey, S. and Ratcliffe, J. (2005) *Gis and Crime Mapping*, John Wiley & Sons Ltd.
- Chainey, S., Reid, S. and Stuart, N. (2002) *When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime*, London: Taylor & Francis.
- Chainey, S., Tompson, L. and Uhlig, S. (2008) 'The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime', *Security Journal*, 21, 4-28.
- Cliff, A. D. and Haggett, P. (1988) *Atlas of disease distributions : analytic approaches to epidemiological data*, Oxford : Basil Blackwell.
- Cohen , L. E. and Felson , M. (1979) 'Social Change and Crime Rate Trends: A Routine Activity Approach', *American Sociological Review*, 44, 588-605.
- Cornish , D. and Clarke , R. (1986) 'The Reasoning Criminal: Rational Choice Perspectives on Offending',
- Cromley, R. G. (1992) *Digital Cartography*, Englewood Cliffs, NJ: Prentice Hall.
- Ebdon, D. (1988) *Statistics in Geography*, Oxford: Blackwell.
- Eck, J., Chainey, S., Leitner, M., Cameron, J. and Wilson, R. (2005) *Mapping crime: Understanding Hotspots*, National Institute of Justice: Washington DC.
- ESRI (2013) 'ArcGIS Resources', *What is a z-score? What is a p-value?* [online], available: [www.http://resources.arcgis.com/en/help/main/10.1/index.html#/What\\_is\\_a\\_z\\_score\\_What\\_is\\_a\\_p\\_value/005p00000006000000/](http://resources.arcgis.com/en/help/main/10.1/index.html#/What_is_a_z_score_What_is_a_p_value/005p00000006000000/) [accessed
- ESRI (2014) 'ArcGIS Resources', *Cluster and Outlier Analysis (Anselin Local Moran's I)* [online], available:

[www.http://resources.arcgis.com/en/help/main/10.2/index.html#//005p0000000z000000](http://resources.arcgis.com/en/help/main/10.2/index.html#//005p0000000z000000) [accessed

Gatrell, A. C., Bailey, T. C., Diggle, P. J. and Rowlingson, B. S. (1996) 'Spatial point pattern analysis and its application in geographical epidemiology', *Institute of British Geographers*, 21(1), 256-274.

Georges, D. E. (1978) 'The Geography of Crime and Violence: A Spatial and Ecological Perspective', *Association of American Geographers*, (1; 78), 28.

GISC (2014) 'Anselin's Local Moran's I', [online], available: [www.http://giscollective.org/tutorials/gis-techniques/spatial-statistics/anselin-local-morans-i/](http://giscollective.org/tutorials/gis-techniques/spatial-statistics/anselin-local-morans-i/) [accessed

Grubestic, T. H. (2006) 'On The Application of Fuzzy Clustering for Crime Hot Spot Detection', *Journal of Quantitative Criminology*, 22(1).

Han, D. and Gorman, D. M. (2013) 'Exploring Spatial Associations between On-Sale Alcohol Availability, Neighborhood Population Characteristics, and Violent Crime in a Geographically Isolated City', *Journal of Addiction*, 2013, 6.

Imhof, E. (1972) *Thematische Kartographie*, Berlin - New York: Walter de Gruyter.

Johnson, S. C. (1967) 'Hierarchical Clustering Schemes', *Psychometrika*, 2, 241-254.

King, B. F. (1967) 'Step-Wise Clustering Procedures', *Journal of the American Statistical Association*, 62(317), 86-101.

Levin, N. (2008) 'The “Hottest” Part of a Hotspot: Comments on “The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime”', *Security Journal*, 21, 295 - 302.

Levin, N. (2013) CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v. 4.0), email to [accessed

Openshaw, S. (1984) *The Modifiable Areal Unit Problem*, Norwich, England: Geobooks.

Ratcliffe, J. and McCullagh, M. (2001) *Crime, Repeat Victimisation and GIS*. In Bowers, K. and Hirschfield, A (eds) *Mapping and Analysing Crime Data – Lessons from Research and Practice*, London: Taylor & Francis

Ratcliffe, J. H. (1999b) Hotspot Detective for MapInfo Helpfile version 1.0, email to [accessed

Ratcliffe, J. H. (2010) *Crime Mapping: Spatial and Temporal Challenges* Springer Science + Business Media, LLC.

Shaw, C. and McKay, H. (1942) 'Juvenile delinquency and urban areas',

Tobler, W. (1970) 'A computer movie simulating urban growth in the Detroit region', *Economic Geography*, 46(2), 234-240.

Williamson, D., McLafferty, S., Goldsmith, V., Mollenkopf, J. and McGuire, P. (1999) 'A Better Method to Smooth Crime Incident Data', *ESRI ArcUser Magazine*, 1-5.