2014

The spatial-temporal prediction of various crime types in Houston, TX based on hot-spot techniques

Shuzhan Fan
Louisiana State University and Agricultural and Mechanical College

THE SPATIAL-TEMPORAL PREDICTION OF VARIOUS CRIME TYPES IN HOUSTON, TX BASED ON HOT-SPOT TECHNIQUES

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirement for the degree of
Master of Science

in

The Department of Geography and Anthropology

by
Shuzhan Fan
B.E., Central South University, 2012
August 2014

ACKNOWLEDGMENTS

First and foremost, I would like to present my appreciation to my major advisor, Dr. Michael Leitner, who enlightened and guided me during my master's study and this thesis research. My thesis would not have been achieved and completed without his aid and guidance. Also, I would like to thank Dr. Fahui Wang, Dr. Xuelian Meng and Dr. Bin Li for serving as my thesis committee members and providing me with valuable advice.

I would also like to extend my gratitude to the Department of Geography and Anthropology for funding me as a graduate assistant during my master's study, without which I would not have the opportunity to come to LSU and pursue my master's degree. In addition, my great appreciation is extended to Luke Driskell, the computer management person in the department, who gave me valuable suggestions while I was working in the CADGIS lab. I also want to thank Weijie Wang, with whom I had many enlightening conversions related to the topic of my thesis research and from whom I learned a lot. Finally, many thanks go to Aimee Moles, a PhD candidate in the department, who I had many conversions with about my thesis research and who provided me with some valuable information.

Last but not least, I am most grateful to my families and my girlfriend. My families are always behind me and being my strong backing. My girlfriend's supporting words helped me get through some tough times. This thesis could be dedicated to them as a gift for their unconditional and everlasting love to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A series of hotspot mapping theories and methods have been proposed to predict where and when a crime will happen. Each method has its strengths and weaknesses. In addition, the predictive accuracy of each hotspot method varies depending on the study area, crime type, parameter settings of each method, etc. The predictive accuracy of hotspot methods can be quantified by three measures, which include the hit rate, the predictive accuracy index (PAI), and the recapture rate index (RRI). This thesis research applied eight hotspot mapping techniques from the crime analysis field to predict crime hotspot patterns. In addition, these hotspot methods were compared and evaluated in order to possibly find a single best method that outperforms all other methods based on the three predictive accuracy measures. Identifying the single best method is carried out for all Part1 Crimes combined and individually, for five of the nine Part 1 Crime. In addition to the spatial analysis, a spatial–temporal analysis of the same crime dataset was conducted to investigate the distribution of crime clusters from both the space and time dimensions. The reported crime data analyzed in this study are from the city of Houston, TX, from January 2011 to December 2012. The results show that the predictive accuracy is affected by both the hotspot mapping method and the crime type, although the crime type has a more moderate effect. Considering the use of the three predictive accuracy measures, the kernel density estimation could be identified as the method which could most accurately predict the overall Part1 Crimes for the city of Houston. The nearest neighbor hierarchical clustering and kernel density estimation could be identified as the methods which are best at predicting each of the five crime types examined based on PAI and RRI, respectively. Also, spatial-temporal analysis indicates that more crimes occurred during September to December, 2011 around the center and in the southwestern part of the city of Houston, TX.

CHAPTER 1 INTRODUCTION

When crime analysts in law enforcement agencies conduct crime analysis, including crime prediction, a key element centers on where crimes tend to occur. Like some other human involved activities (traffic accidents, disease outbreaks, gentrification, etc.), crime incidents are not distributed randomly throughout space. Their distribution is dense at some locations while sparse at others. This feature of crime events distribution was described as an ‘inherent geographical quality’ by Chainey and Ratcliffe (2005) and was explained by theories such as the ecology of crime (Brantingham and Brantingham, 1984) or routine activities (Cohen and Felson, 1979), and others. The places where crime events are relatively densely distributed are called hotspots. Crime hotspots are referred to as areas where crimes concentrate spatially (McLafferty et al., 2000; Eck et al., 2005).

The concept of a hotspot is widely used in our daily life. Being aware of which places are safer and which places are with a higher risk of being a victim of crime, people visit or live in some locations while they avoid others. Based on the knowledge of risks of victimization, people make choices of the communities they live in, the schools they send their children to, or the recreation area they spend their weekend in, etc. In some western countries, people living in some neighborhoods need to install a closed-circuit television (CCTV) to secure their house and deter potential offenders. In other neighborhoods they do not have to worry about their properties even if they forgot to lock their door during the day. The hotspot concept is also of critical importance to policing and patrolling actions. Provided with information about the specific spread of hotspots, police commanders could then make more appropriate decisions about where and when to allocate limited manpower resources to the places where patrolling demands are at the highest.

Hotspot analysis is at the center of the analysis of crime, and hotspot mapping is paid most attention among crime mapping.

Hotspot mapping is an effective and widely used analytical technique which uses retrospective crime data to identify crime hotspots. Hitherto a number of hotspot mapping techniques have been proposed and applied to identify crime clusters. These include spatial ellipse, thematic mapping of geographic boundaries, quadrat thematic mapping, interpolation and continuous surface smoothing methods, and local indicators of spatial association (LISA) statistics mapping, among others. These visualization techniques possess both strengths and weaknesses. To better assess the quality of these techniques to forecast the occurrence of future crime events, three different standard measures which are commonly referred to as predictive accuracy measures have been proposed. The hit rate is one of the earliest and most used measures. It is calculated as the percentage of crime events that falls within hotspot areas produced from retrospective crime data. Another measure is the Predictive Accuracy Index (PAI) which takes both the effect of the hit rate and the size of the study area and the crime hotspots into consideration. In addition, Levine (2008) provided the Recapture Rate Index (RRI) as an adjustment to the PAI. To compare how accurately these techniques work to predict where and when crimes may occur in the future, each predictive accuracy measure (hit rate, PAI, RRI) is calculated in this thesis research to represent the relative accuracy level of each technique. Also, the literature indicates that crime types have an effect on the predictive accuracy (Chainey et al, 2008; Hart and Zandbergen, 2012). For this reason, the three predictive accuracy measures (hit rate, PAI, RRI) will be computed and examined for five different crime types, including aggravated assault, auto theft, burglary, larceny-theft, and robbery.

While hotspot mapping reveals the inherent spatial characteristics of crime events, it fails to reveal their temporal features. For example, based on the routine activity theory, which put an emphasis on the place or environment where offenders commit crimes instead of on the characteristics themselves, the occurrence of a criminal event requires ‘the convergence in space and time of likely offenders, suitable targets and the absence of capable guardians against crime’ (Cohen and Felson, 1979). However, due to the scarce availability of GIS functionalities and corresponding theories and applications, the integration of spatial and temporal analysis of crime have been traditionally neglected or little researched by both academics and professional practitioners (Ratcliffe, 2002a). McCullagh (2006) states that ‘emphasis is usually placed on the spatial hotspot with only simplistic attempts to tie in temporal changes because of the complexities involved’. To include time into the analysis, the Kulldorff’s scan statistics analysis (Kulldorff et al., 1998) will be used to investigate the space-time patterns of crime incidents. Also, a hotspot plot which was first devised by Townsley (2008) will be created so that the reader could ‘assess temporal profiles of individual hotspots at the micro and macro level; compare the importance and temporal signature of different hotspots; and relate the results of the temporal analysis at both macro and micro levels to baseline measure’ (Townsley, 2008).

The remainder of this thesis is organized into six chapters:

Chapter 2 includes a literature review of the theories, methods, techniques, and applications of the relevant practices done by crime analysts or academic researchers. This review includes discussion about spatial hotspot mapping methods and spatial-temporal hotspot analysis and mapping methods.

Chapter 3 outlines the study area and datasets used in this thesis research. Also, the preprocessing of the data, especially geocoding, will be discussed.

Spatial hotspot mapping techniques will be introduced in Chapter 4. This chapter includes the following three sections. Three measures of predictive accuracy are introduced and discussed in the first section. Next, eight hotspot mapping methods and their parameter settings will be discussed in the second section. The third section will talk about the effect that crime types have on hotspot techniques' predictive accuracy.

Chapter 5 will discuss spatial-temporal analysis of crime data. It contains the following two sections, namely the hotspot plot and the spatial-temporal scan statistic.

Results are shown in Chapter 6. Implications of the results will also be discussed in this chapter.

In the final Chapter 7 the results from Chapters 4 and 5 will be summarized. Limitations of the research and future research directions will be discussed. The possible implications of the results from this research for the Houston Police Department will be highlighted.

CHAPTER 2 LITERATURE REVIEW

For the police and governmental administrations, crime analysis which is based at the sub-jurisdiction level is paid particular attention to. This is referred to as the Strategic Crime Analysis (SCA). SCA focuses on cluster analysis in order to produce information that can be used for resource allocation, beat configuration, the identification of non-random patterns in criminal activity, and unusual community conditions (Hart & Zandbergen, 2012). Hotspot analysis is one of the most popular techniques used in SCA. Crime hotspots are areas where crimes tend to concentrate in space and/or time. The common understanding is that a hotspot is an area that has a greater than average number of criminal events, or an area where people have a higher than average risk of victimization (Chainey & Ratcliffe, 2005). Hotspot techniques have the unique characteristics that they can identify spatial and/or temporal clusters and their ad hoc boundaries as well as predict future events. Such clusters vary depending on the geographic scales (jurisdictions, blocks, streets, specific addresses, etc.) as well as temporal scales (years, seasons, months, days, hours, etc.).

The use of hotspot mapping has gained its popularity both from crime prevention practitioners and academics. In some western countries such as England, the U.S., and Australia, hotspot mapping techniques have been increasingly adopted by law enforcement agencies and police officers (Gottlieb et al., 1994; Maguire, 2000; Ratcliffe, 2002c; Seddon and Napper, 1999). The reason for the increasing trend to apply hotspot mapping can be partly explained to the limited fiscal budget provided to law enforcement agencies. This method offers the agencies a way to assist with allocating their limited resources or manpower to the areas where a crime is more likely to happen.

In the academic area, hotspot mapping has been increasingly drawn attention by the advance of both hotspot mapping theories and techniques. Different theories have been developed by a variety of researchers to help find theoretical explanations for the definition and cause of hotspots. These theories range from the social ecology of crime to theories on routine activities and repeat victimization (Anselin et al., 2000). In addition, the advance of Geographic Information Systems (GIS) has prompted the further development of hotspot mapping techniques. A variety of crime analysis tools available in GIS make it easier and attract more researchers both in more practical and theoretical academic fields to focus on the research of hotspot mapping. A detailed literature review of these hotspot mapping techniques, including spatial and spatial-temporal hotspot mapping, will be discussed next.

Spatial crime hotspot mapping techniques have witnessed their development alongside huge innovations in information technology (IT). Some of these spatial techniques are associated with the spatial arrangement and the size of the subdivisions inside the study area (e.g. districts, blocks, census tracts, etc.). Thematic mapping is the simplest method regardless of what spatial arrangement and size of subdivision is. One problem occurs when this method is applied to statistical or administrative areas such as census blocks. The individual units of these different spatial subdivisions (census blocks versus census tracts) have different shapes and boundaries, i.e. a different spatial arrangement. The main problem is that different spatial arrangements of such statistical / administrative areas result in hotspot maps that differ from each other. This problem is referred to as the Modifiable Areal Unit Problem (MAUP). The effect of MAUP cannot be neglected when methods associated with administrative / statistical areas are applied (Chainey et al., 2008; Openshaw, 1983; Ratcliffe, 2004).

A simple solution to the MAUP would be the use of a regular grid imposed onto the study area. Grid thematic mapping is among one of the commonly used methods that produce grid maps. Each grid cell has a uniform size and shape. In addition, each grid cell has a value, usually crime counts, assigned to it. The value could also be a density value such as crime rates (Eck et al., 2005). Kernel density estimation (KDE) also imposes a regular grid onto the study area and uses a three-dimensional kernel function to visit each grid cell and to calculate a density value assigned to each grid cell (Eck et al, 2005). This method has been viewed by several researchers as the most suitable method for the purpose of visualization (Chainey and Ratcliffe, 2005) and as the most accurate method for predicting future crime incidents (Chainey et al., 2008).

The improvement of computing power has also spurred the development of some computer programs in crime analysis. One of earliest software packages used was the Spatial and Temporal Analysis of Crimes (STAC) to identify crime hotspots (Illinois Criminal Justice Information Authority, 1996). The output of a crime hotspot is displayed as ellipses. Though STAC has been used by many crime prevention practitioners and crime analysts, weaknesses exist in this method. One such weakness is that the distribution of crime clusters does not necessarily form an ellipse. This may create misleading results to the police decision makers who may use these results to allocate limited patrol manpower (Bowers and Hirschfield, 1999; Chainey et al., 2008; Ratcliffe, 2002b).

As for now, STAC has been integrated to the widely used crime analysis program CrimeStat 4.0 (Levine, 2013). CrimeStat 4.0 is usually used by crime analysts and practitioners to investigate the distribution of point patterns data (crime event locations), which means, the input data should be point data, or centroids when polygon data were used (where a centroid represents the geometric center of the corresponding area). This program contains a series of functionalities to examine crime point patterns data, including hotspot mapping techniques. Nine hotspot mapping techniques are provided by the program. These are mode, fuzzy mode, nearest neighbor hierarchical clustering, risk-adjusted nearest neighbor hierarchical clustering, STAC, K-means clustering, local Moran's I, Getis Ord local "G", and kernel density estimation. Each technique requires the user to enter suitable parameters.

Another problem in crime mapping is related to the heterogeneity of the study area. In some urban geographic spaces (e.g. the city of Houston as explored in this thesis research), some areas may have a number of crimes which is small compared to the entire study area, but relatively large compared to its local neighbors. This area which has a local cluster pattern is referred to as a local hotspot. Measures designed to detect these local hotspots are called Local Indicator of Spatial Association (LISA) statistics (Anselin, 1995; Ord and Getis, 1995; Getis and Ord, 1996; Ratcliffe and McCullagh, 1999). They include the local Moran's I, the Local Geary's C, Gi and the Gi* statistics. Among these LISA statistics, the local Moran's I and the Gi* received the most attention (Chainey and Ratcliffe, 2005). The difference between these two statistics is that the local Moran's I is based on covariance and identifies Moran's I value for each zonal area so that the area can be examined as being different or similar to its neighborhoods. The Gi* compares local averages to global averages. Some other techniques are also available to produce spatial crime hotspots. These include, but are not limited to the Nearest Neighbor Hierarchical Clustering (Levine, 2004), K-Means clustering, spatial scan statistic, etc.

There is at least one more thing in the discussion of spatial hotspot mapping techniques that needs to be paid particular attentions to. This is related to some other factors which may affect the spatial distribution of crimes (e.g. population density, income, density of housing, etc.). For example, in an area which has a spatially concentrated numbers of larceny-theft crimes and which has a large volume of population residing, working, or visiting (e.g. the downtown area in a city, or a recreation center), motivated crime offenders are more likely to find potential targets to commit crimes (Chainey and Ratcliffe, 2005). An area with more people in it tends to attract more criminals, thus more crimes occur. Hence, the population distribution has to be considered in research related to the spatial and/or temporal distribution of crime. One solution is to use crime rates rather than crime counts as the value used to create hotspot maps. Examples include risk-based thematic mapping, risk-adjusted nearest neighbor hierarchical clustering, etc.

Much effort has been devoted to studying the relevance of space in identifying patterns of crime or crime clusters. The eight hotspot mapping techniques discussed in this thesis research may just represent the "tip of the iceberg" of the large volume of work that has been contributed to this topic. By contrast, temporal analysis has received much less attention. In fact, if crime analysts or crime prevention practitioners do not consider the temporal factor of crime analysis, at all, they may provide incomplete, biased, or even misleading results to police officers or law enforcement agencies. According to the routine activity theory (Cohen and Felson, 1979), a motivated crime offender is more likely to commit a crime when he/she encounters a suitable target (or victim) under the circumstance of the absence a guardian. The factors which result in the occurrence of crime have to meet both in the dimension of space and time. Many activities like traffic rush hours or the difference between workload during weekdays and weekends present changes in the temporal pattern. Felson and Paulson (1979) thus reasoned that certain types of crime tend to concentrate at certain times of day/weak/year. Several but not too many studies have been carried out to address differences in crime concentrations across different temporal scales (Johnson et al., 2008; Felson and Paulson, 2002; Paulson and Robinson, 2004).

Some work has examined crime changes over periods of time, either to look at long-trend changes such as years or seasons (Block, 1984; Lebeau, 1992) or to look at short-trend changes such as weeks, days or intra-days (Bowers et al., 1998; Johnson et al., 1997; Ratcliffe and McCullagh, 1998). There exist a series of techniques to detect spatial-temporal patterns of crime clusters. According to a comparative study of spatial-temporal hotspot analysis techniques used in the area of security informatics conducted by Zeng et al. (2004), two types of spatial-temporal hotspot analysis and mapping techniques have gained more popularity among researchers and practitioners. One was developed by the advance of different scan statistics which are primarily applied to the realms of public health and epidemic prevention (Kulldorff, 2001). The other one was built upon the growing of data clustering analysis and its variations. Among these two types of spatial-temporal hotspot techniques, scan statistics and nearest neighbor hierarchical clustering received most attention (Leitner and Helbich, 2011).

CHAPTER 3 DATA AND GEOCODING

3. 1 The Study Area and the Spatial Data

The study area of this research consists of the jurisdiction of the Houston Police Department (HPD), which is the primary law enforcement agency serving the City of Houston and which overlaps with several other law enforcement agencies such as the Harris County Sheriff's Office and the Harris County Constable Precincts. On a geographic scale, the boundary of the HPD districts extends from -95.784602°W to -95.000783°E and from 30.126094°N to 29.519338°S (see Figure 3.1 below).

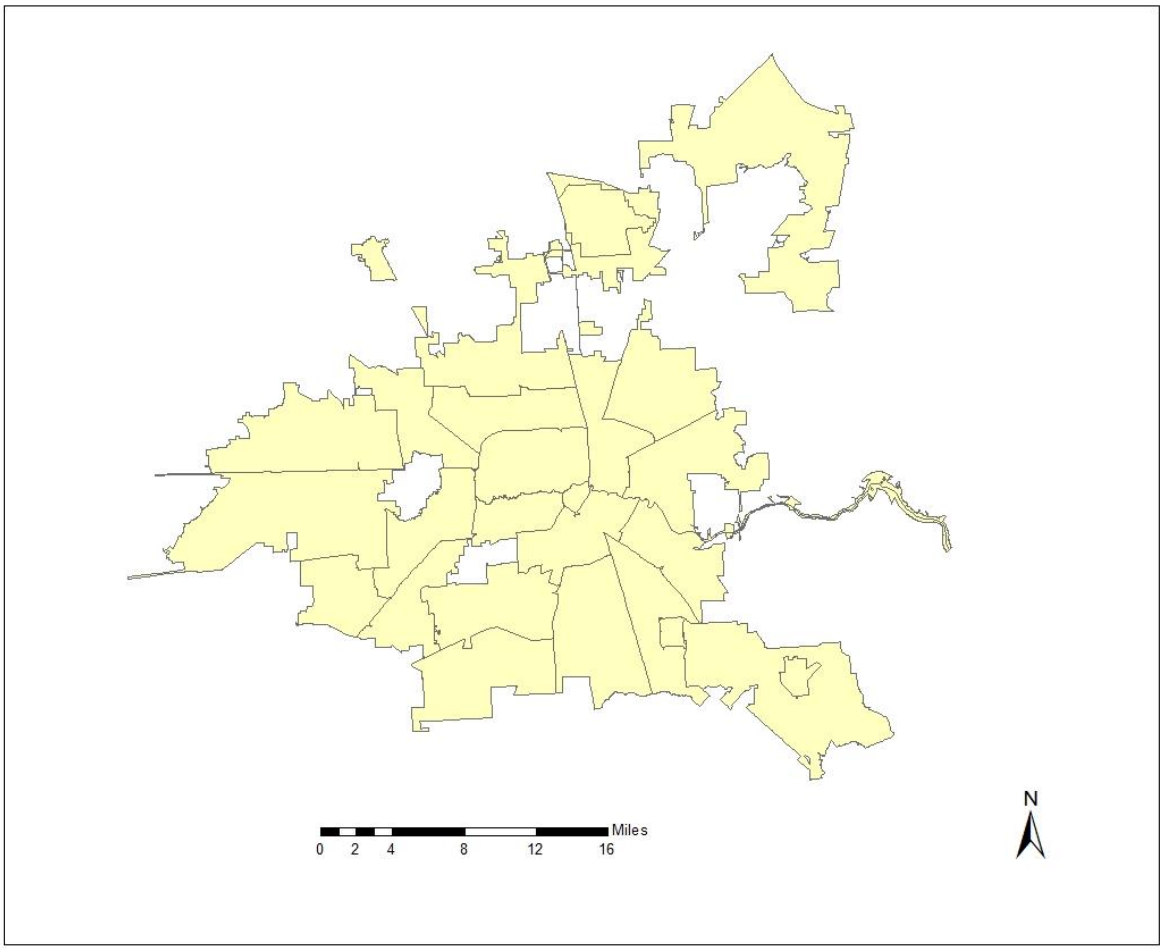


Figure 3. 1 Geographic boundary of the study area, the City of Houston

In order to geocode crime incidents onto a street network map, the census tract shapefiles for 2010 were downloaded for free online as part of the products of the City of Houston GIS Release, which is also known as COHGIS (http://gisdata.houstontx.gov/cohgis). The COHGIS data release contains administrative places, roads, boundaries, blocks, and census tracts datasets, etc. Compared to the commonly used TIGER/Line shapefiles, which can be downloaded through the U.S. Census Bureau website (http://www.census.gov/geo/maps-data/data/tiger-line.html), the COHGIS not only includes geographic data, but also include some demographic data such as population, race, house unit, etc. For the purpose of this research, the population information of 2010 is required to conduct risk-based hotspot methods that include the risk-based thematic mapping and risk-adjusted nearest neighbor hierarchical clustering method. Also, the boundaries of the COHGIS data correspond to the spatial extend of the crime data which is to be discussed in the next section. The boundary of the TIGER/Line shapefile includes the entire Harries County, where the city of Houston is located. There would have been a need to do “clip” to narrow the study area down to the city extent when using the TIGER shapefile.

3. 2 The Crime Data

The crime data used in this research could have been obtained from the Houston Police Department (HPD) website (http://www.houstontx.gov/police/cs/stats2.htm). However, crime data were collected free of charge from the HPD through the Texas Public Information Act by submitting an open record request. Acquiring the crime data through an open record request results in a more complete and accurate dataset, than the one available at the HPD website. The crime data set includes all reported crimes classified according to the Federal Bureau of Investigation’s (FBI) Uniform Crime Reporting (UCR) program. This research will investigate nine Part 1 Crimes which include murder and non-negligent manslaughter, manslaughter by negligence, forcible rape, robbery, aggravated assault, burglary, larceny-theft, auto theft, and arson. Only Part 1 Crimes are included in this thesis research because these crimes are taken as more serious than others in crime analysis and the data sources are more reliable. The police are usually on the scene to record these types of crimes. Table 3.1 shows the UCR codes for the nine Part 1 Crimes.

Table 3. 1 UCR classification offenses codes for Part 1 Crimes

UCR Classification Offenses for Houston Police Department

Part 1 Crimes (Part 1 crimes, except for 01 & 09, are included in the Crime Index.)

Violent Crimes

00	Murder And Non-negligent Manslaughter
01	Manslaughter By Negligence (Usually not included with other Part 1 Crimes)
02	Forcible Rape
03	Robbery
04	Aggravated Assault (Class I)

Non-Violent Crimes

05	Burglary
06	Larceny – Theft (Includes Burglary of Motor Vehicles)
07	Auto Theft
09	Arson (This includes only those Arsons which also have other offenses. The Houston Fire Department Arson. Arson is included with Crime-Index Crimes in the Modified Crime Index)

In addition to the almost 50 offense types (Part 1 and Part 2 Crimes, and Other Offenses), the data set includes the offense date and time, police beat, and the actual street address, where the offense took place. A complete set of crime data for the selected nine Part 1 Crimes from January 2011 to December 2012 will be used in this research.

The original crime data are provided in either a Microsoft Office Access Database format or a Microsoft Excel format and are limited to those crime events which are known to the police. The 2011 crime dataset includes a total of 131,707 recorded crime incidents and the 2012 dataset 130,218 recorded incidents. Table 3.2 lists the number of crime incidents by crime type and by year.

Table 3. 2 Number and percentage of crimes for nine Part 1 Crime types for the year 2011 and 2012

UCR Code	Type of Crime	Number of Crimes and Percentage 2011		2012	
00	Murder and Non-negligent Manslaughter	226	(0.17%)	245	(0.19%)
01	Manslaughter By Negligence	17	(0.01%)	44	(0.03%)
02	Forcible Rape	820	(0.62%)	640	(0.49%)
03	Robbery	8435	(6.4%)	9394	(7.21%)
04	Aggravated Assault	12484	(9.48%)	11310	(8.69%)
05	Burglary	27783	(21.09%)	26579	(20.41%)
06	Larceny-Theft	68978	(52.37%)	67893	(52.14%)
07	Auto Theft	12826	(9.74%)	13948	(10.71%)
09	Arson	138	(0.1%)	165	(0.13%)
	All Part I Crimes	131707		130218	

Table 3.2 shows that larceny-theft takes up more than 50% of all Part 1 Crimes. Robbery, aggravated assault, burglary and auto theft make up almost 50% of all Part 1 Crimes, while the proportion of murder and non-negligent manslaughter, manslaughter by negligence, forcible rape and arson total less than 1%. This may be explained by the fact that the four crime types whose proportion of crimes of all Part 1 Crimes is less than 1% are all violent crimes. The occurrence of a violent crime is less likely to take place than a non-violent crime. A law enforcement agency branch may receive a couple of burglary reports during a single day, but may receive only one murder report every other day or days.

3. 3 Geocoding

Geocoding is a process to transfer indirect geocodes (e.g. place names, zip codes, census tracts, etc.) to direct geocodes (e.g. x and y coordinates, latitude and longitude). In my thesis research, the indirect geocodes are the names of addresses where crime incidents occurred. The direct geocodes are the X and Y coordinates of the crime locations. The crime incidents must be geocoded onto the street map for the purpose of hotspot mapping.

After the acquisition of the crime data set and the street network data (the TIGER/Line shapefile), geocoding can then be accomplished using ArcGIS 10.2. The street network data contain all roads information (e.g. names, addresses, ranges, city, etc.) for a county. They are part of the product of TIGER/Line shapefiles and can be downloaded from U.S. Census Bureau website (http://www.census.gov/geo/maps-data/data/tiger-line.html).

Several parameters require to be specified in order to perform geocoding correctly and appropriately. According to Leitner and Helbich (2011), who did a spatial-temporal analysis in the City of Houston to study the impact of hurricanes on crime, the spelling sensitivity was set to 80, the minimum candidate score and the minimum match score were set to 75 and 60, respectively. These three parameters are utilized jointly in ArcGIS for geocoding to help find an appropriate and accurate match address for each crime incident location. The matched or tied point will be assigned an address which has the highest match score from the candidate addresses and the unmatched point will not be assigned an address. The same user-defined geocoding parameter settings as in Leitner and Helbich (2011) are applied in this research and are shown in Figure 3.2.

Using this set of parameters, the match rates for the nine crime types and the total of all Part 1 Crimes are all close to or above 95%. According to Ratcliffe (2004), this is a sufficiently high match rate. In comparison, an increase of the minimum match score to 80 and keeping the other parameters unchanged would have resulted in match scores of less than 90%. Table 3.3 presents the match rates after geocoding.

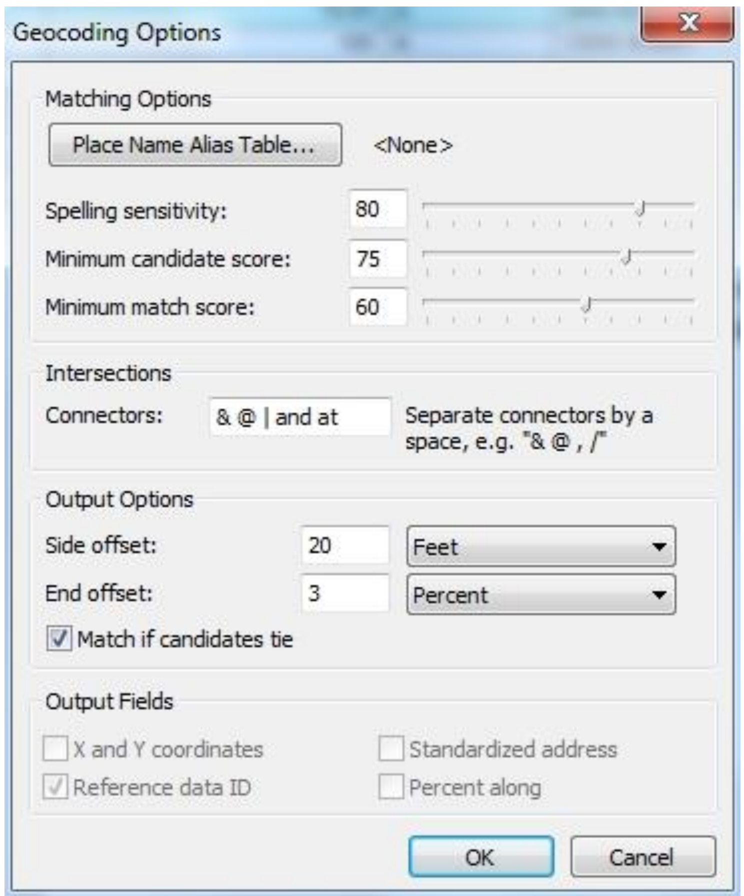


Figure 3. 2 Geocoding parameters setting window in ArcGIS 10.2

After geocoding, all crime locations with unmatched addresses were removed and not included in the subsequent analysis of this research. Table 3.4 shows the number of crime incidents and the corresponding percentages for nine crime types and the overall Part 1 Crimes after completion of the geocoding process.

After geocoding, robbery, aggravated assault, burglary, larceny-theft, and auto theft again total to close to 99% of all Part 1 Crimes. Since after geocoding all crime incident locations are assigned X and Y coordinates, crime locations can now be used to conduct spatial and temporal hotspot analysis.

Table 3. 3 Match rates for nine Part 1 Crime types for the year 2011 and 2012 after geocoding

UCR Code	Type of Crime	Match Rate 2011	Match Rate 2012
00	Murder and Non-negligent Manslaughter	99%	96%
01	Manslaughter By Negligence	94%	89%
02	Forcible Rape	95%	96%
03	Robbery	96%	96%
04	Aggravated Assault	97%	97%
05	Burglary	96%	96%
06	Larceny-Theft	94%	94%
07	Auto Theft	96%	95%
09	Arson	95%	95%
	All Part I Crimes	95%	95%

Table 3. 4 Number and percentage of crimes for nine Part 1 Crime types for the year 2011 and 2012 after geocoding

UCR Code	Type of Crime	Number of Crimes and Percentage 2011	Number of Crimes and Percentage 2012
00	Murder and Non-negligent Manslaughter	223 (0.18%)	235 (0.19%)
01	Manslaughter By Negligence	16 (0.01%)	39 (0.03%)
02	Forcible Rape	786 (0.63%)	616 (0.50%)
03	Robbery	8128 (6.51%)	9043 (7.31%)
04	Aggravated Assault	12024 (9.63%)	10892 (8.81%)
05	Burglary	26732 (21.41%)	25593 (20.70%)
06	Larceny-Theft	64580 (51.71%)	63782 (51.59%)
07	Auto Theft	12258 (9.82%)	13269 (10.73%)
09	Arson	131 (0.10%)	157 (0.13%)
	All Part I Crimes	124878	123626

CHAPTER 4 SPATIAL PREDICTIVE HOTSPOT MAPPING METHODS

This research will use eight hotspot mapping techniques to create hotspot maps based on 2011 Part 1 crimes data and then predict crime incidents for 2012. The hotspot crime maps for 2011 and the predicted crime maps for 2012 will be utilized to compare and evaluate the eight techniques so that it may be possible to find a single best method that outperforms all others.

Identifying the single best hotspot method is accomplished for two violent crime types (robbery and aggravated assault) and for three non-violent crime types (burglary, larceny-theft, and auto theft). According to Table 3.4, each of the other four crime types (murder and non-negligent manslaughter, manslaughter by negligence, forcible rape, and arson) possesses very low counts and makes up less than 1% of the total of all nine Part 1 Crimes. The individual sample sizes for these four crime types are too small to reasonably conduct some of the hotspot techniques (i.e. STAC or nearest neighbor hierarchical clustering method).

Previous research has revealed that the accuracy of predictive crime hotspot mapping depends in part on the predictive crime mapping techniques as well as the types of crime. The measures of predictive accuracy include the hit rate, the Predictive Accuracy Index (Chainey et al, 2008), and the Recapture Rate Index (Levine, 2008).

This chapter consists of the following three parts. First, the three predictive accuracy measures will be discussed in detail. Second, the impact of hotspot crime mapping methods on the predictive accuracy using all the Part 1 Crimes data for the year 2011 and 2012 will be analyzed. Finally, the predictive accuracy will be assessed for five different crime types.

4. 1 Measures of Predictive Accuracy

The first measure of predictive accuracy is the hit rate. This measure is calculated as the percentage of new crimes that occur within the areas where crimes are predicted to occur (Chainey et al, 2008). The higher the hit rate, the more accurate the hotspot technique is. This measure is easy to calculate and to understand. However, the larger the hotspot area, the higher the likelihood is that a higher number of future crimes would fall into it. The hit rate does not thus take the area of the hotspot into consideration. This could make the results less meaningful to law enforcement agencies. For instance, a hit rate can be calculated that exceeds 90%, but the hotspot areas also make up more than 90% of the study area. It is unlikely for the police to patrol such a large area because of limited resources and manpower. Thus, a measure which considers the size of hotspots vis-à-vis the size of the study area is needed to better evaluate the predictive accuracy. This is accomplished with the next measure, which is the Predictive Accuracy Index.

Predictive Accuracy Index (PAI) was first introduced by Chainey et al (2008). It was created to address the problem the hit rate may produce. In other words, the PAI takes the sizes of hotspots and the study area into consideration. It is defined as the ratio of the hit rate to the proportion of the study area that consists of hotspots in the retrospective year (Hart and Paul, 2012). The formula (4-1) is as follows:

$$PAI = \frac{hit\ rate}{proportion\ of\ hot\ spot\ area} = \frac{n/N}{a/A} \qquad (4-1)$$

where *n* is the number of new crime incidents which fall into predicted hotspot areas from the retrospective year, *N* is the number of new crimes in the whole study area, *a* is the total area occupied by hotspots, and *A* is the size of entire study area. Compared to the hit rate, the PAI could weaken the effect of study area on producing meaningless information to police's tactical determination. Again, a larger PAI value means a hotspot mapping method that is more accurate for predicting crime.

The third predictive accuracy measure is the Recapture Rate Index (RRI). It was proposed by Levine (2008) in a response to Chainey et al.'s newly invented PAI. The RRI does not take the sizes of hotspots or the study area into consideration. The index is calculated by dividing the ratio of hotspot crime counts for 2011 and 2012 by the ratio of the total number of crimes for each year (see formula 4-2 below):

$$RRI = \frac{hotspot\ crime\ ratio}{total\ crime\ ratio} = \frac{n1/n2}{N1/N2} \qquad (4-2)$$

where $n1$ is the number of crimes in hotspot areas for year 2011, $n2$ is the number of new crime incidents for year 2012 which took place in predicted hotspot areas, $N1$ is the total number of crimes for year 2011, and $N2$ the total number of crimes for year 2012. Similar to the hit rate and the PAI, a larger RRI corresponds to a more accurate hotspot mapping method for crime prediction.

After introducing the three measures of predictive accuracy, the eight hotspot methods will be discussed one by one in much detail.

4. 2 Hotspot Methods and Parameters

Eight hotspot mapping methods were selected in this research to create hotspot maps. These eight methods were chosen because of their availability (for example in ArcGIS or in other programs that are easily accessible), popularity (whether they have been commonly applied by other crime analysis researchers or practitioners), and their comprehensiveness (this set of eight methods includes two risk-based hotspot mapping methods in order to consider the effect of population density on crime prediction). The selected eight methods include risk-based thematic mapping, grid thematic mapping, spatial and temporal analysis of crime (STAC), nearest neighbor hierarchical clustering (NNHC), risk-adjusted nearest neighbor hierarchical clustering, kernel density estimation, local Moran's I statistic, and Gi* statistic (Table 4.1). The type of data and the mapping result vary for different methods. Points and administrative polygons are two types of data used and census tracts, grids and grids are three forms of mapping results.

The data used in this section are the reported crime events for 2011 and 2012 in Houston, TX. Since the effect that crime types have on hotspot technique's predictive accuracy will be studied in the next section (Section 4.2), the total number of Part 1 Crimes data was analyzed in this section.

Table 4. 1 Polygon and point pattern analysis methods and their corresponding outputs

Methods	Data Type	Hotspot Mapping Results
Thematic Mapping	Polygon	Census Tracts
Risk-Based Thematic Mapping	Polygon	Census Tracts
Grid Thematic Mapping	Point	Grids
STAC	Point	Ellipse
NNH	Point	Ellipse
Risk-Based NNH	Point	Ellipse
Kernel Density Estimation	Point	Grids
Local Moran's I	Polygon	Census Tracts
Gi*	Point	Grids

4. 2. 1 Non-Risk-Based Methods

1. Grid Thematic Mapping

The grid thematic mapping technique is put forward to deal with the problem of the effect of different sizes and shapes of enumeration areas on crime counts or crime rates. This is accomplished by placing a uniform grid over the study area with each grid cell having the same size and shape (usually a square). Different to risk-based thematic mapping, where each area has a crime rate associated with it, in grid thematic mapping each cell can display a value that is either a crime count or a crime rate. It is possible to display crime counts with this mapping approach, since all cells of the regular grid have the same size and shape.

One critical part in successfully performing grid thematic mapping is to choose an appropriate cell size. Coarse cell sizes may fail to display the detailed spatial information within each cell and thus the resulting map may become less useful to law enforcement agencies (Chainey and Ratcliffe, 2005). Too fine cell sizes may create a larger volume of data and may present too much information to police decision makers that they can hardly rely on to make appropriate tactical decisions. Researchers have provided guidelines on how to select a possible grid cell size. One guideline is to divide the distance in the longest extent of the study area by 50, and use the resulting value as a starting point in choosing the right cell size. This guideline was suggested by Chainey and Ratcliffe (2005). After some experimenting, 200 meters was finally selected as the grid cell size for the grid-based thematic mapping method. In addition, the threshold was set to the 90% percentile, which separates the 10% highest from the 90% lowest crime score or crime rates. Cells with the 10% highest crime counts or crime rates are defined as hotspots.

2. Spatial and Temporal Analysis of Crime (STAC)

The Spatial and Temporal Analysis of Crime (STAC) method is one of the earliest tools available for crime analysis (Illinois Criminal Justice Information Authority, 1996). It was initially developed as two computer programs, which include the Time Analyzer and the Space Analyzer. The Space Analyzer is aimed to help crime analysts find and locate the hotspot areas by creating ellipses placed over the study area. Now this function was integrated into CrimeStat 4.0, which is a software specifically developed to perform spatial and temporal crime incidents analysis (Levine, 2004). The Time Analyzer helps police identify when the particular type of crime is most likely to occur. The time analysis function was not provided in CrimeStat 4.0. As Eck et al. stated (2005), this method has several drawbacks. One major drawback is that the spatial distribution of crime hotspots does not naturally form an ellipse, which is the output created by STAC. A second drawback is that STAC is a technique more suitable to a crime analyst, who has a good knowledge of the technique as well as of the data. It is somewhat difficult for a novice to correctly specify the parameters used in STAC.

To perform STAC in CrimeStat 4.0, several parameters are required to be entered. Among those is the cell size which was set to 200 meters. This is consistent with other methods performed in this thesis, such as grid thematic mapping or kernel density estimation. The hotspot threshold in STAC is the number of points which could form a cluster, which was set to 15 points.

3. Nearest Neighbor Hierarchical Clustering

The nearest neighbor hierarchical clustering (NNHC) method uses a hierarchical clustering routine to create a hierarchy of hotspots based on several user-defined criteria, including the minimum number of points that a cluster should consist of. NNHC is based on the nearest neighbor analysis technique and hotspots consist of groups of points that are closer than expected under spatial randomness (Eck et al., 2005). The clustering routine will repeat until all points are grouped into a single cluster or the clustering criteria fail (Levine, 2004). The clustering criteria are based on two parameters, which need to be selected by the user.

The first parameter is the minimum number of points, which requires that a hotspot should at least contain this number of points to be considered a hotspot. The other parameter is the threshold distance. In CrimeStat 4.0, there are two choices available for setting the threshold distance. They are the fixed threshold distance and the random threshold distance. For the random threshold distance, which is the default one, the user has to specify a significance level. For example, if a 'p less than 0.05' significance level is selected, then only 5% of all pairs of points (two points consist of one pair) will have a distance which is smaller than the threshold distance. For the fixed distance, a specific distance value, e.g. 100 meters, has to be entered. A point will only be considered to be included into a hotspot if the distances between this point and other point or points are all smaller than the specified threshold distance.

Only if the both criteria are met will a point be grouped into a first-order cluster/hotspot. Then, the process continues with first-order clusters to be clustered to the second-order, third order, etc. clusters, until one of the criteria fails.

The search radius for this method was set to 250 meters. In addition, 15 points were chosen as the minimum number of points to form a first order nearest neighbor cluster.

4. Kernel Density Estimation

The kernel density estimation has been agreed by several researchers as being the most suitable hotspot mapping technique (Chainey and Ratcliffe, 2005; Chainey et al., 2008). It is also a very popular method among crime analysis practitioners. It is one of the continuous surface smoothing methods which interpolate values based on intensity values of known points. It works by first imposing a regular grid with a specified cell size over points across the study area. Then, a user-defined three-dimensional kernel function of a user-defined search radius will visit each point and calculate densities for all the cells within the search radius. The final kernel density estimate for one cell is then calculated by summing up all values obtained from all kernel density functions for that particular cell. This method is preferred by many practitioners in part due to its nicely visualized mapping results and its availability in most spatial analysis and GIS software packages.

CrimeStat 4.0 provides several kernel functions to be used. Different kernel function will yield different density values. The quartic kernel function was selected for this thesis research, since it is a rather popular selection (Chainey et al., 2002; Chainey and Ratcliffe, 2005; Eck et al., 2005). Also, the cell size and the bandwidth (search radius) are required to be entered to successfully perform this hotspot mapping method. The appropriate selection of these parameters is of vital significance for the results of this method. Researchers have proposed a series of guidelines on how to determine these parameters (Ratcliffe, 2004; Chainey and Ratcliffe, 2005). To be consistent with the parameter settings from the other hotspot methods, the cell size and the search radius were set to 200m and 250m, respectively. The thematic threshold was greater than three standard deviations.

5. Local Moran's I

Local indicators of spatial association (LISA) are a set of statistics, which are widely employed by crime analysts. These statistics are proposed because traditional global statistics which explores the spatial association across the whole study area offer little insight into the location, relative scale, size, shape and extent of hotspots (Chainey and Ratcliffe, 2005). Instead global statistics just provide a general examination of the spatial relationships of crime events in the study area. LISA statistics, however, were developed to study the spatial association between one point and its neighbors within a user-defined distance. The local Moran's I and the Gi* statistics are two of the most commonly used LISA statistics by researchers and practitioners.

The local Moran's I is based on covariance and identifies a Moran's I value for each zonal area so that the area can be examined as being different or similar to its neighbors. The definition of "neighbors" has to be specified by users. It can be either adjacent areas or areas negatively weighted based on the distance from the observation area (Anselin, 1995).

In terms of parameter settings, the local Moran's I requires a Z value (e.g. intensity or weight) to be specified. This intensity value, was set as the number of crime counts. The cell size was set to 200m. The thematic threshold was set to larger than 99.9% significance, which means that there is a 1 in 1000 chance to commit a type I error that is the null hypothesis will be rejected, even though it is true.

6. Gi*

The Gi and the Gi* statistics are another set of LISA statistics. The difference between these two statistics is that the Gi* statistic considers the effect of the value of the point itself in the calculation of the Gi* values, while Gi does not. Gi* is more popular to be utilized by crime researchers and analysts. It was thus selected instead of Gi as one of two hotspot mapping methods in this thesis research as one hotspot mapping method.

Different from the local Moran's I, the Gi* statistic compares local averages to global averages. This statistic can be calculated and displayed in ArcGIS 10.2 using the "Hotspot Analysis – Getis and Ord Gi*" tool. This requires the user to enter a threshold distance. According to the instructions provided by Chainey (2008), the lag distance or threshold distance can be calculated as the distance of the diagonal of one cell. The cell size was determined to be 200m, resulting in a threshold distance of 283m. The resulting Gi* values are actually Z scores, which is calculated as the distance of the observation from the mean, standardized by the standard deviation. Z scores can be further used to evaluate the statistical significance. The same as the local Moran's I, the thematic threshold for the Gi* statistic was aslo set to larger than 99.9% significant.

4. 2. 2 Risk-Based Methods

1. Risk-Based Thematic Mapping

Thematic mapping is also called graduated color or choropleth mapping. It is widely used for showing administrative or enumeration areas by cartographers and crime analysts in order to obtain an overview of the spatial distribution of crime incidents. It works by assigning graduated colors to different statistical areas. In crime analysis, these areas are usually associated with attributes such as crime rates.

Thematic mapping method requires users to specify a classification scheme whereby areas with similar values are grouped together. In ArcGIS, several classification methods are provided. They include natural breaks, equal interval, quantile, standard deviation, manual classification, etc. Choosing an appropriate classification method and the corresponding class boundaries is important in crime analysis research. Different classification schemes will place crime events into different categories, and will change classification boundaries.

After the classification scheme is specified, the risk-based thematic map can then be produced based on the crime rates associated with each statistical area. Crime rate, rather than crime count, is used as the value based on which a thematic map is created because it is more appropriate for the purpose of crime analysis.

It is common-sense knowledge that a densely populated area tends to have a larger number of people living and working in, which are potential victims to criminals. The larger volume of victims may attract additional crime offenders. Hence, a higher amount of crimes may be committed within this area. For example, a downtown area usually witnesses a higher number of crimes (both violent and nonviolent) compared to a suburb due to its large amount of people visiting, working or living in it. In addition, shopping districts are more likely to be attractive places for crime offenders to commit crimes like larceny-theft and auto theft. This is because a big flow of people together with a large parking lot become possible targets for offenders. This reveals a fact that population density may be somewhat related to crimes that occur within the statistical or administrative area. To assess the effect of this factor on the predictive accuracy, a new field in the GIS attribute table called crime rate was added, defined as the counts of crime per 100,000 people.

To decide on which points or areas can be regarded as hotspots, a thematic threshold value needs to be specified. A thematic threshold is a value which crime analysts use to separate hotspot crime areas from other areas. For hotspot mapping techniques (e.g. risk-based thematic mapping, grid thematic mapping, kernel density estimation, local Moran's I, and Gi*), which produce a hotspot map with several categories, from lowest to highest, usually the highest class will be regarded as the hotspot class. Enumeration areas or grids falling within this class are hotspots.

By examining the statistical distribution of the crime data and through a trial and error process, the threshold was set at greater than one standard deviation for the risk-based thematic mapping method. All census tracts with a crime rate of greater than three standard deviations are classified as hotspots. All crime rates falling into the hotspot class can be utilized for the calculation of the three predictive accuracy measures.

2. Risk-Adjusted Nearest Neighbor Hierarchical Clustering

The risk-adjusted nearest neighbor hierarchical clustering (risk-adjusted NNHC) method is developed on the basis of the nearest neighbor clustering (NNHC) routine, which is discussed above and the kernel density estimation, which is discussed below. The risk-adjusted NNH clustering method introduces an intensity or weight field. For many police purposes, for example, as discussed in risk-based thematic mapping, the population distribution plays an important role in where crime hotspots occur. In this research, the intensity field is the population of each census tract. The risk-adjusted NNH clustering routine will dynamically adjust the threshold distance based on the distribution of the population rather than relying on the user-defined threshold distance. The clusters of points which are closer than what would be expected according to a baseline population will then be identified by the routine as risk-based hotspots (Levine, 2004).

The risk-adjusted NNH clustering routine utilizes the kernel density estimation to implement the dynamic adjustment of the threshold distance. This requires the user to specify several parameters for the kernel density routine. The parameters are set to the same values as the kernel density method discussed next. Also, to be consistent with the NNH clustering technique, the threshold of the minimum number of points was set to 15 and first order clustering was used.

4. 3 Comparison of Predictive Accuracy Measures Across Crime Types

The dataset in this research contains nine Part 1 Crime types. However, as shown in Table 3.4, after geocoding, only five of the nine crime types possess more than 5% of the total number of crimes each. These are robbery, aggravated assault, burglary, larceny-theft and auto theft. This section will study how crime types affect predictive accuracy testing the same eight hotspot mapping techniques as applied in Section 4.2. To be consistent across each crime type, the parameters selected for each hotspot mapping technique remain the same. The three measures of predictive accuracy were calculated for each combination of any one of the five crime types and eight hotspot mapping techniques. Table 4.2 shows the results of the three predictive accuracies for each of the 40 combinations (5 crime types x 8 mapping techniques).

The results clearly show that different crime types have an effect on the predictive accuracy. For example, hit rates for larceny-theft are higher than for any of other four crime types. This may be because larceny-theft has by far the highest percentage (52%) among all five crime types. However, when using the PAI, robbery tends to be as accurate or more accurate than any of the other four crime types. Finally, the RRI is again highest for larceny-theft.

It is also interesting to answer the questions which crime type has a higher predictive accuracy for one particular hotspot mapping technique, or which hotspot mapping technique is more accurate at predicting future crimes for any or most of the crime types. To answer the first question, the STAC method can be taken as an example. When using STAC as the hotspot mapping technique for all five crime types studied, the predictive accuracy is higher for larceny-theft than for any of the other four crime types. In order to answer the second question, it can be shown that the NNH clustering and the kernel density estimation outperform all other mapping techniques at predicting future crime events across all five crime types.

Table 4. 2 Results of three measures of predictive accuracy for any combination of five crime types and eight hotspot mapping techniques

	Robbery	Aggravated Assault	Burglary	Larceny-Theft	Auto Theft
	Hit Rate (%)				
Risk-Based Thematic Mapping	0.26	0.16	0.15	2.69	0.95
Grid Thematic Mapping	26.16	24.08	33.42	49.58	30.04
STAC	9.90	7.17	6.89	9.84	7.03
NNHC	10.87	14.21	22.72	47.02	13.92
Risk-Adjusted NNHC	1.17	2.93	10.73	26.62	4.25
KDE	18.18	19.15	19.29	23.24	20.53
Local Moran's I	48.84	51.31	32.34	24.70	30.22
Gi*	9.63	7.90	14.31	15.52	9.93
	PAI				
Risk-Based Thematic Mapping	0.09	0.06	0.06	0.46	0.29
Grid Thematic Mapping	27.20	23.61	14.33	15.12	21.98
STAC	12.78	8.26	8.67	16.08	9.82

(Table 4.2 continued)

	Robbery	Aggravated Assault	Burglary	Larceny-Theft	Auto Theft
	PAI				
NNHC	54.39	36.78	19.96	15.78	49.93
Risk-Adjusted NNHC	34.10	34.33	17.47	13.93	25.39
KDE	29.26	23.67	19.04	28.80	24.07
Local Moran's I	2.87	1.92	1.01	1.73	1.26
Gi*	32.06	26.85	19.19	23.34	27.94
	RRI				
Risk-Based Thematic Mapping	0.47	0.69	0.95	1.09	0.96
Grid Thematic Mapping	0.72	0.76	0.81	0.93	0.81
STAC	0.57	0.53	0.56	0.63	0.54
NNHC	0.52	0.60	0.59	0.68	0.60
Risk-Adjusted NNHC	0.58	0.88	1.01	1.23	0.94
KDE	1.01	1.10	1.11	1.15	1.12
Local Moran's I	0.92	0.97	0.98	1.01	0.97
Gi*	0.81	0.82	0.89	0.95	0.87

CHAPTER 5 SPATIAL-TEMPORAL HOTSPOT MAPPING METHODS

In this chapter, the discussion of the spatial analysis of crime will be extended to spatial-temporal analysis. Similar to the previous chapter on spatial analysis, which compared eight crime hotspot mapping techniques to explore the spatial distribution of five crime types, in temporal analysis mapping techniques have been widely adopted to identify temporal patterns of crime. One simple idea is to use to compare a pair of timestamps to detect changes of crime clusters in the temporal dimension. For example, in a research conducted by Leitner and Helbich (2011) to investigate the impact of Hurricane Rita and Hurricane Katrina on crime, the Kulldorff's scan statistics was used to detect spatio-temporal crime clusters over two periods, namely before and after the landfall of both hurricanes. Another example is given by Bowers and Johnson (2003), who developed statistical testing structures to access crime prevention before and after some specific measures have been implemented.

Choosing a pair of timestamps could produce problems of underestimating the importance of time in the distribution of crime clusters, particularly for distinguishing stable and fluid clusters (Nakaya and Yano, 2010). Consequently, larger time periods have been chosen by some researchers. A time interval of an hour, day, week, month, season, or year are most commonly used by researchers. For example, Rengert's study (1997) concluded that crime cluster patterns varied based on different periods of time within one day. Nakaya and Yano (2010) chose one month as the time interval in their study to explore a 3-D hotspot mapping method for visualizing crime clusters.

In this research, the data were provided by the Houston Police Department on a monthly basis. The dataset ranges from Jan. 2011 to Dec. 2011 (12 months). Thus it was decided to use one month as the time interval. Figure 5.1 shows the reported monthly numbers of crimes (all Part 1 Crimes) in Houston, TX in 2011.

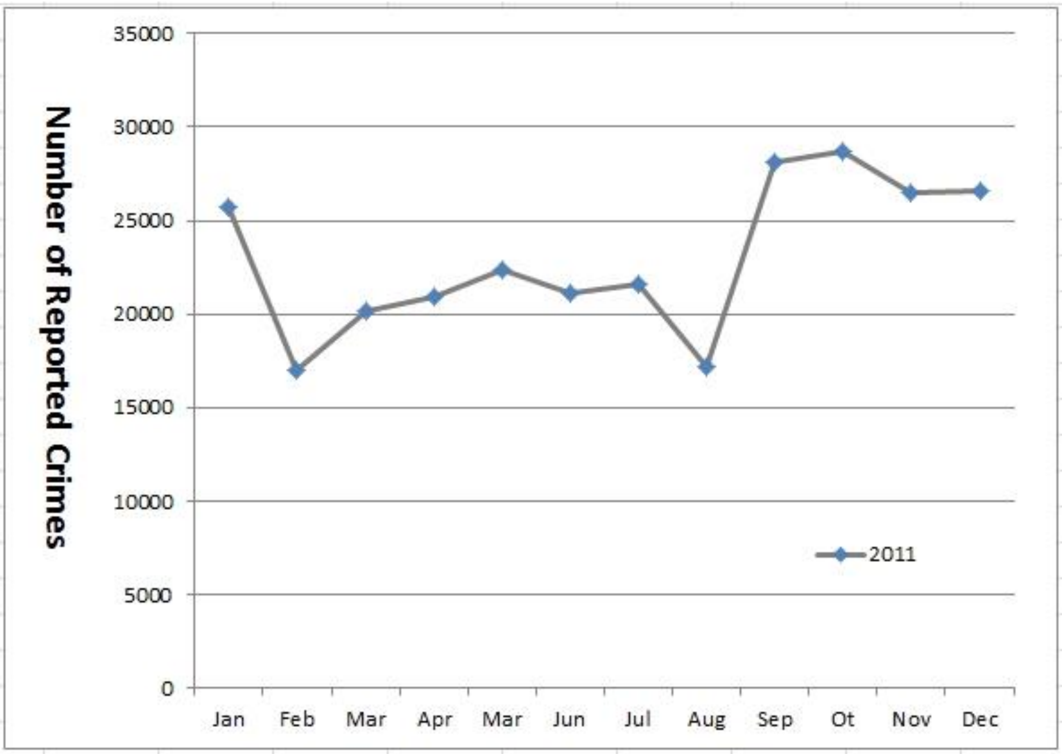


Figure 5. 1 Monthly trends of Part 1 Crimes in Houston, TX in 2011

5. 1 Hotspot Plot

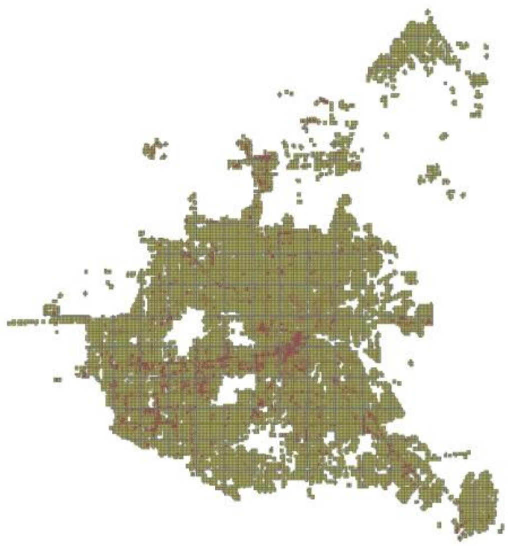
The hotspot plot is a visualization method which aims to present spatial analysis with consideration of the distribution of events in time within hotspots (Townsley, 2008). Different from other spatial-temporal hotpot analysis and mapping methods such as Kulldorff's scan statistic, hotspot plots focus more on visualizing data and communicating information to users efficiently. As stated by Townsley (2008), several criteria need to be met in order to assure this method is useful. First, it should not be complicated to be implemented. Second, it should allow time patterns to be presented at various hotspot levels. And, third, it should be able to be compared with other hotspot maps. Intuitively, the hotspot plot comprises three parts that include the long term trend in crime, the intra-day trend in crime, and the spatial crime clusters map (Townsley, 2008).

Based on the availability of the dataset, one month as the long term trend (over 12 months) and one hour as the short term trend (over 24 hours) were used. The kernel density is chosen to be used as the technique to produce spatial hotspot maps. The datasets used in this section are the all Part 1 Crimes. The results are shown in Figure 5.2.

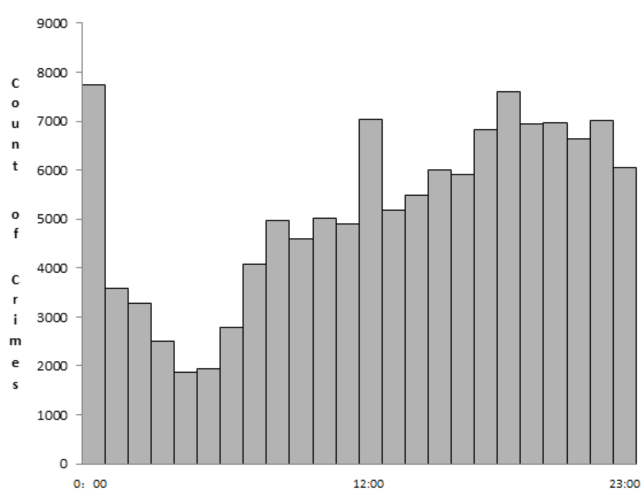
Looking at the long term trend plot on the top, there is a clearly increasing trend staring from September, to January, 2011. The kernel density estimation map shows that more crimes are clustered in the center, west, and southwest of the city. The short term trend plot at the bottom indicates that crime trend in one day can be split into three sections. Crimes decrease after 12 o'clock in the midnight and start to bounce back in the morning and reach a spike at noon. Then the high crime counts continue till the midnight.

The results from the hotspot plot (Figure 5.2) can be used by the police decision makers to determine when and where a police patrol allocation is most needed. For example, the hotspot in Figure 5.2 shows that crimes were more likely to take place in the western and central part of the city of Houston, TX after 12 pm to the midnight, from October to January. The police commanders and law enforcement agency officers could rely on these results to allocate their manpower and schedule the patrol shifts which could be most likely to prevent crimes.

a) Long term trend (month) hotspot plot for 2011



b) Kernel density estimation surface of crime for 2011



c) Short term trend (hour) hotspot plot for 2011

Figure 5. 2 Hotspot plot for all Part 1 Crimes in 2011 using the entire study area

5. 2 Space-Time Scan Statistic

The space-time scan statistic has been one of the most widely used methods in the analysis of spatial-temporal data. It is derived from the space scan statistic which is aimed to identify spatial clusters by imposing circular windows with various radii to scan across the study area (Kulldorff, 1997). Each circular window with a particular radius assigned to it will cover sets of neighboring areas and a likely candidate of including a hotspot or cluster. In accordance with Kulldorff (1997), the formula to calculate the spatial scan statistic is as follows (formula 5-1):

$$S = \frac{max_z L(Z)}{L0} \qquad (5-1)$$

where S is the spatial scan statistic. Z is the set of circles of the scanning windows. L(Z) is the likelihood ratio for circle Z. $L0$ is the likelihood ratio under the null hypothesis. S is essentially the maximum likelihood ratio of all circles divided by the likelihood ratio computed from the null hypothesis. Thus, the cluster contained in the circle with the maximum likelihood scan statistic is also the most likely cluster. Furthermore, in order to test the distribution of the test statistic, whose actual distribution remains unknown, Monte Carlo simulations are utilized. Under the null hypothesis that cases within the study area taking place at random following a user-defined model, the program then calculates values of the scan statistic for both the real dataset and the simulated datasets (Zeng et al., 2004). If the calculated value of the scan statistic of the real dataset is more than 95% of all the values, then the identified cluster or hotspot is significant at 95% level.

The spatial-temporal scan statistic is based on the spatial scan statistic. The spatial scan statistic is viewed as a 2D crime map, which uses a circular window scanning the study area. While after adding a time factor the spatial-temporal scan statistic employs a 3-D cylinder to scan the area both horizontally and vertically. The circular window now serves as the base of the cylinder and time is measured by the height.

In this research the Kulldorff's spatial-temporal scan statistic is used to detect crime clusters in space and time. The software used to apply Kulldorff's scan statistic is SaTScan which was developed by Kulldorff (Kulldorff, 2001, 2005). The input data are X, Y coordinates (the spatial component) and the date (day) when the crime happened (the temporal component).The space-time permutation model was chosen in the analysis. Other settings were not changed from the defaults provided in SaTScan. Figure 5.3 shows the selected settings in SaTScan. The dataset used here is all Part1 Crimes from January 2011 to December 2011. One month was selected as the temporal unit. The calculation in SatScan is very time-consuming. It took more than 62 hours on a computer (i5-2400QM CPU, 3.10 GHz, 8 GB RAM) to perform the Kulldorff's spatial-temporal scan statistic. To visualize the data in ArcGIS, the results were joined with point data. The datasets for the five crime types were also analyzed in SatScan. The results of the analysis will be discussed in the following chapter.

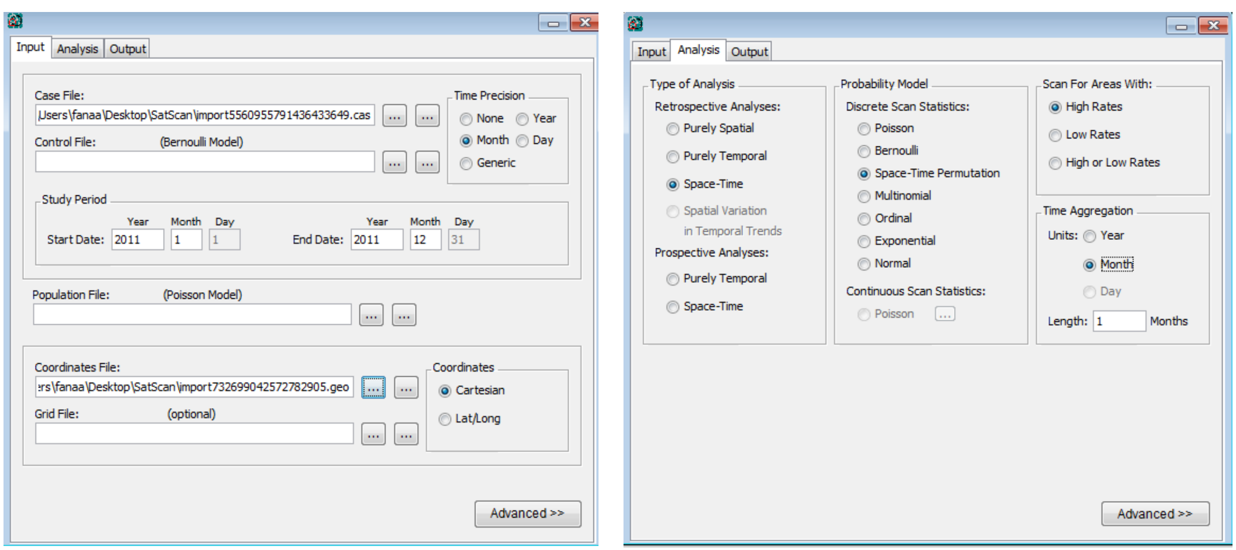


Figure 5. 3 Parameter settings windows in SaTScan

CHAPTER 6 RESULTS

6. 1 Hotspot Techniques

Since different techniques are based on different theories, concepts, and set of parameters, their resulting outputs, namely, the hotspot maps, are thus somewhat different from each other. The statistically significant hotspot area produced using one technique maybe lacking statistical significance using another method or even turning into a coldspot, when considering the population at risk. The study area for all methods is the same, namely, the City of Houston, TX. The hotspots produced by risk-based thematic mapping and local Moran's I use census tracts as their unit of observation. Grid thematic mapping, Gi* and KDE show their results in the form of a regular grid. Finally, STAC, NNH clustering and risk-adjusted NNH clustering methods exhibit their results in the form of ellipses.

After having compiled all hotspot maps, the three measures of predictive accuracy (hit rate, PAI, and RRI) can be computed. The formulas for all three measures were given in Section 4.1. Table 6.1 lists the parameter settings for each cluster method. Table 6.2 presents the results of the three predictive accuracy measures across the eight hotspot crime mapping techniques.

When interpreting the hit rate as one measure to assess the predictive accuracy for various hotspot methods, it obviously needs to be kept in mind that the four methods, which produced the highest number of hotspots and largest hotspot sizes, are better at predicting future crime events, since a higher number of new crime events would be located inside these retrospective hotspots. In contrast, the PAI, which takes the study area and the hot spot sizes into consideration, yields the best results with the kernel density estimation and the Gi* statistic. Finally, the RRI predicts future crimes the best with the risk-adjusted nearest neighbor hierarchical clustering method and the kernel density estimation.

Table 6. 1 Hotspot mapping methods parameters

Methods	Parameters		
	Cell Size	Search Radius	Threshold
Risk-Based Thematic Mapping	N/A	N/A	Greater than 1 standard deviation
Grid Thematic Mapping	200m	N/A	10%
STAC	200m	750m	15 points, first order
NNHC	N/A	250m	15 points, first order
Risk-Adjusted NNHC	200m	250m	15 points, first order
KDE	200m	250m	Greater than 3 standard deviation
Local Moran's I	N/A	N/A	Greater than 99.9%
Gi*	200m	283m	Greater than 99.9%

It should be kept in mind that these results are based upon a large dataset consisting of nine different crime types. These results may be applied by the police for tactical decision making. For example, if the results are presented to the general police officer in the city of Houston and the main purpose is to reduce overall crime for the entire city, the results shown in this section might be potentially suitable. However, if the purpose is to effectively allocate resources by a police decision maker in order to control the number of one particular crime or crimes, then additional studies about the effect of individual crime type on the predictive accuracy needs to be studied.

Table 6. 2 Measures of predictive accuracy for eight hotspot mapping methods

	Crimes in 2011		Crimes in 2012		Total (km^2)		Predictive Accuracy		
Hotspot Mapping Techniques	In 2011 Hotspot	In Study Area	In 2011 Hotspot	In Study Area	Area of 2011 Hotspot	In Study Area	Hit Rate (%)	PAI	RRI
Risk-Based Thematic Mapping	1890	124022	1979	122785	94.26	1625	1.61	0.28	1.06
Grid Thematic Mapping	64889	124251	61550	123028	65.63	1571	50.03	11.98	0.96
STAC	15389	124878	9323	123626	11.25	1571	7.54	10.58	0.61
NNHC	100398	124878	65879	123626	105.34	1571	53.29	7.95	0.66
Risk-Adjusted NNHC	38058	124878	48558	123626	85.31	1571	39.28	7.23	1.29
KDE	23565	124878	28070	123626	18.44	1571	22.71	19.34	1.20
Local Moran's I	57488	124022	55836	122785	401	1625	45.47	1.84	0.98
Gi*	21628	124251	20580	123626	18.86	1571	16.65	13.87	0.96

It can be seen from the results in Figures 6. 1 – 6. 8 that the local Moran's I, grid thematic mapping, NNH clustering and risk-adjusted NNH clustering yield more hotspots than the other methods.

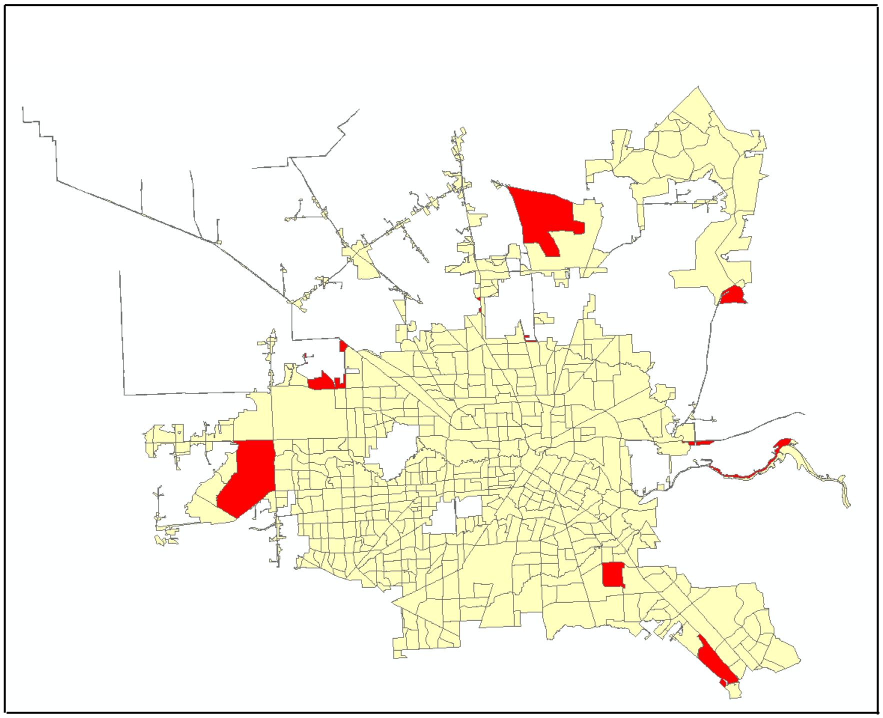


Figure 6. 1 Hotspot mapping results for risk-based thematic mapping technique for 2011

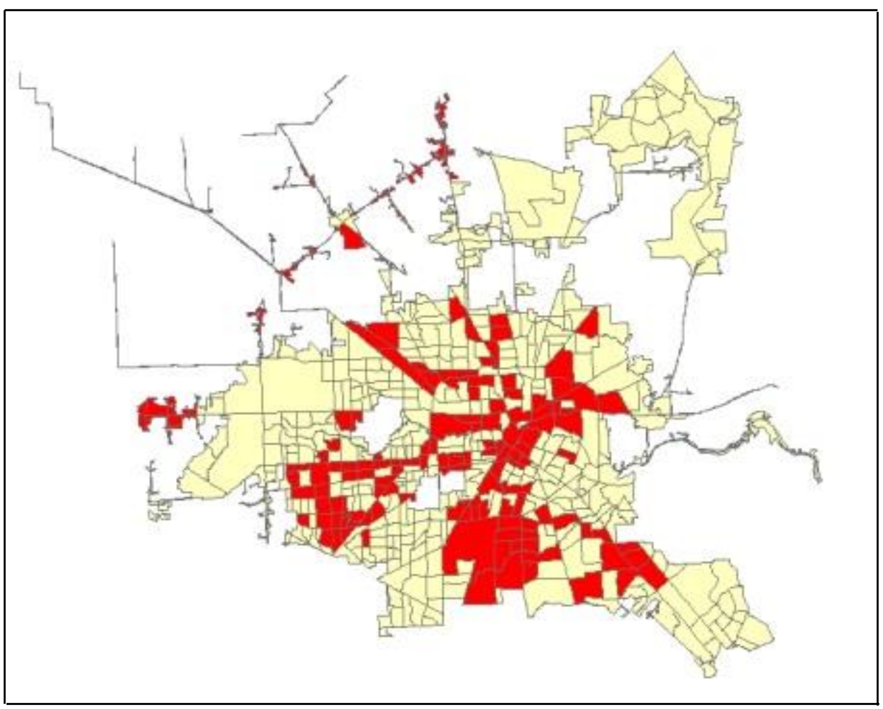


Figure 6. 2 Hotspot mapping results for local Moran's I mapping technique for 2011

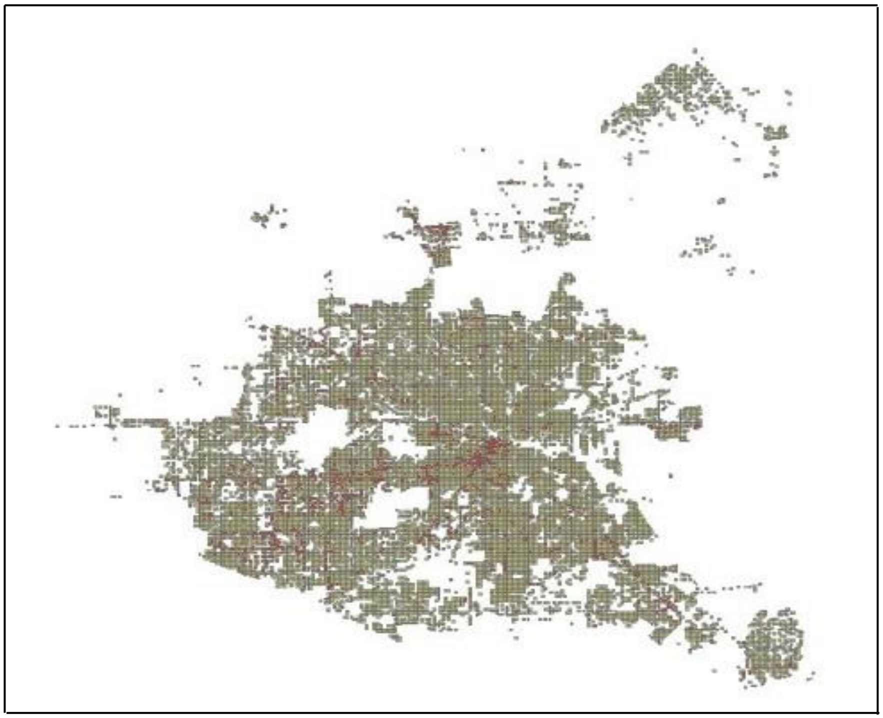


Figure 6. 3 Hotspot mapping results for grid thematic mapping technique for 2011

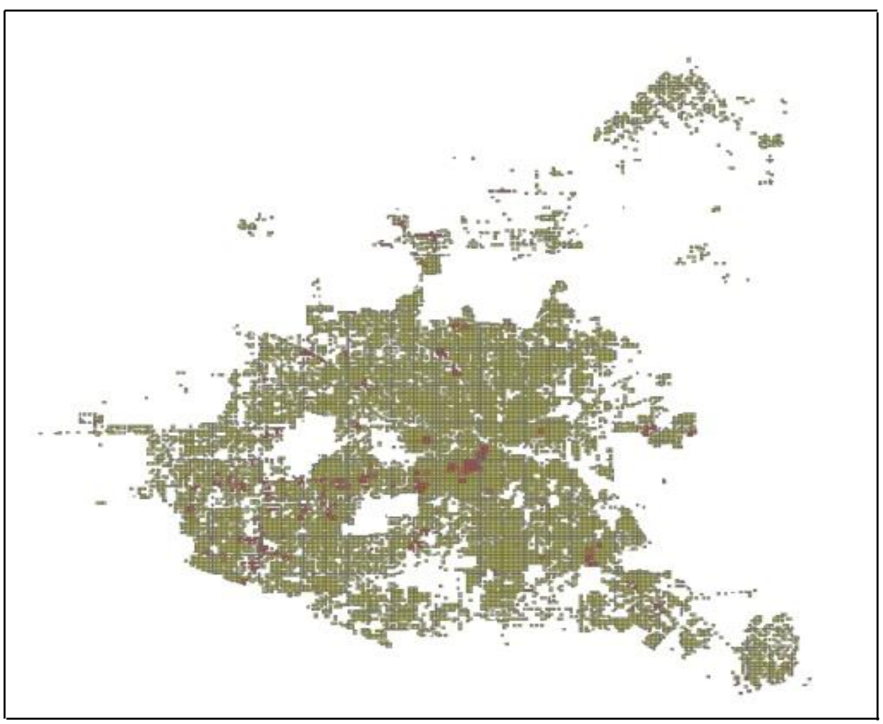


Figure 6. 4 Hotspot mapping results for Gi* mapping technique for 2011

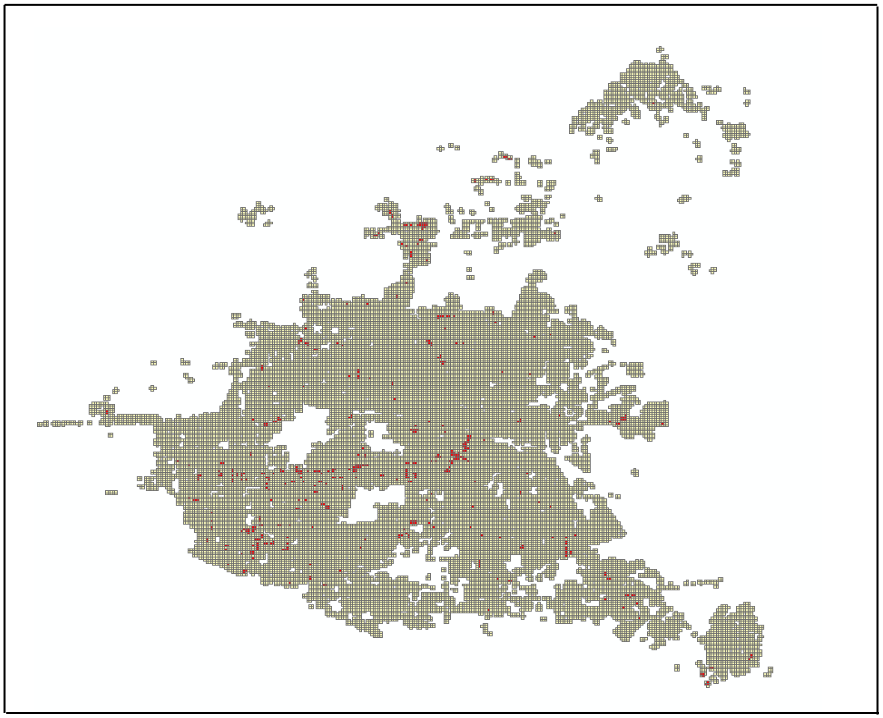


Figure 6. 5 Hotspot mapping results for kernel density estimation mapping technique for 2011

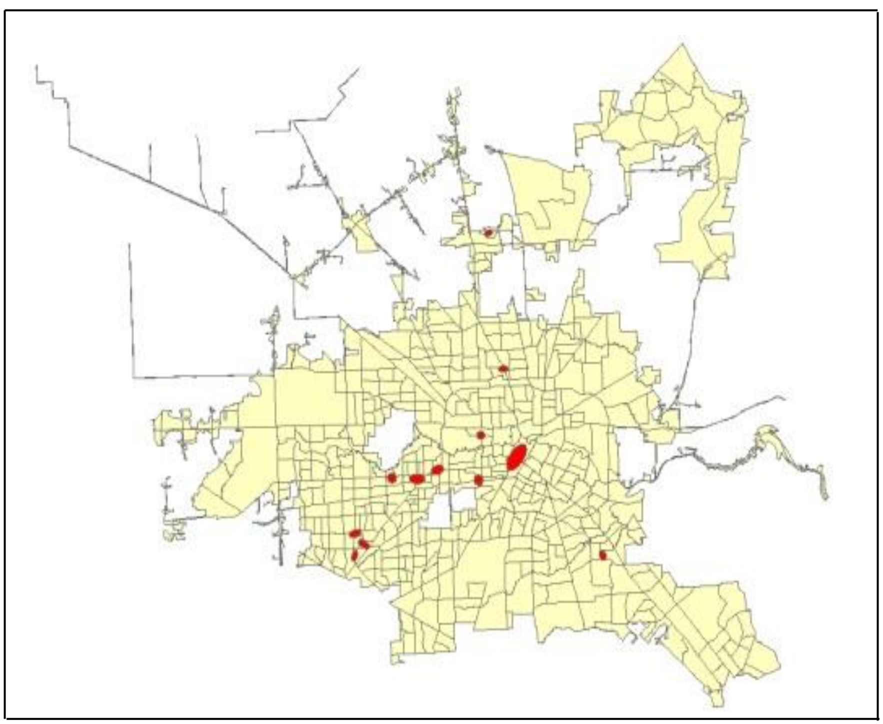


Figure 6. 6 Hotspot mapping results for spatial and temporal analysis of crime mapping technique for 2011

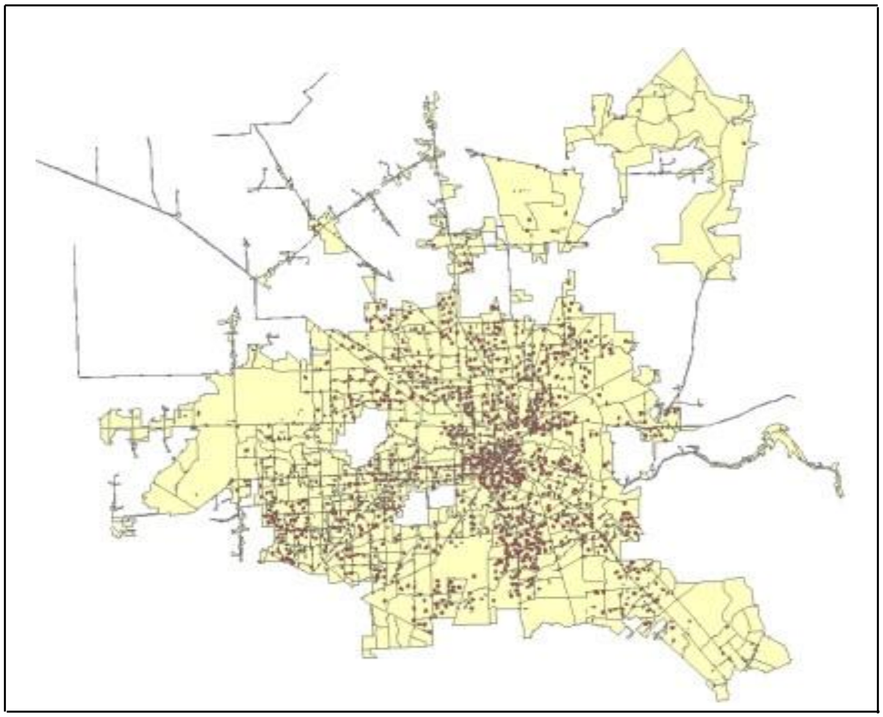


Figure 6. 7 Hotspot mapping results for nearest neighbor hierarchical clustering mapping technique for 2011

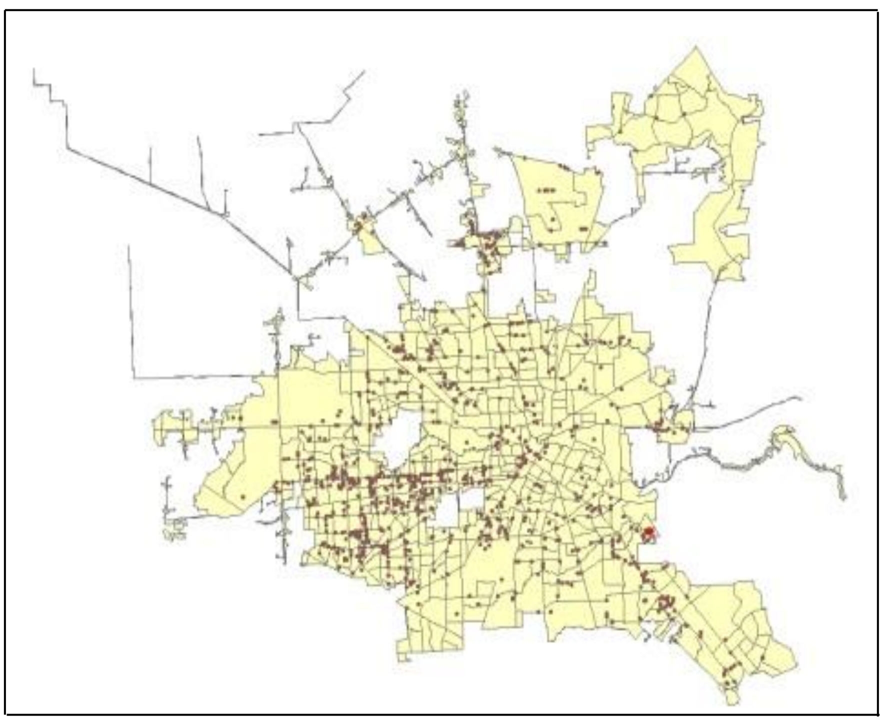


Figure 6. 8 Hotspot mapping results for risk-adjusted neighbor hierarchical clustering mapping technique for 2011

6. 2 Crime Type

When taking crime type into consideration, the three predictive accuracy measures change substantially across eight hotspot methods. But from the perspective of hotspot methods, the three measures vary moderately across five crime types. The results of the three predictive accuracy measures by nine crime types and eight hotspot methods are presented in Table 4.2. In general, hit rate and PAI for robbery appear to be higher among five crime types. When using RRI as the predictive accuracy measure, however, larceny-theft is the crime type which can be predicted more accurately.

One objective in this thesis research is to find a single best hotspot method which is better at predicting future crime events. A modified version of Table 4.2 is shown in Table 6.3, Table 6.4 and Table 6.5 in order to locate the best method for each individual crime type based on three predictive accuracy measures.

By examining the hotspot methods' ability to predict future crime events across five crime types, findings are different from the above and may provide valuable advice to police decision makers. Kernel density estimation method is consistently the best method at predicting future crime events for all five crime types when RRI is used as the predictive accuracy measure. Nearest neighbor hierarchical clustering method could be generally regarded as the most accurate hotspot method for crime prediction when PAI is the measure. When hit rate serves as the predictive accuracy measure, the best hotspot method varies for different crime types at predicting crime incidents in the future.

Table 6. 3 The hit rate for the combination of five crime types and eight hotspot mapping techniques. The value in bold represents the highest value among the eight hotspot methods for each crime type

	Robbery	Aggravated Assault	Burglary	Larceny-Theft	Auto Theft
Risk-Based Thematic Mapping	0.26	0.16	0.15	2.69	0.95
Grid Thematic Mapping	26.16	24.08	33.42	49.58	30.04
STAC	9.9	7.17	6.89	9.84	7.03
NNHC	10.87	14.21	22.72	47.02	13.92
Risk-Adjusted NNHC	1.17	2.93	10.73	26.62	4.25
KDE	18.18	19.15	19.29	23.24	20.53
Local Moran's I	48.84	51.31	32.34	24.70	30.22
Gi*	9.63	7.90	14.31	15.52	9.93

Table 6. 4 The PAI for the combination of five crime types and eight hotspot mapping techniques. The value in bold represents the highest value among the eight hotspot methods for each crime type

	Robbery	Aggravated Assault	Burglary	Larceny-Theft	Auto Theft
Risk-Based Thematic Mapping	0.09	0.06	0.06	0.46	0.29
Grid Thematic Mapping	27.27	23.61	14.33	15.12	21.98
STAC	12.78	8.26	8.67	16.08	9.82
NNHC	54.39	36.78	19.96	15.78	49.93
Risk-Adjusted NNHC	34.1	34.33	17.47	13.93	25.39
KDE	29.26	23.67	19.04	28.80	24.07
Local Moran's I	2.87	1.92	1.01	1.73	1.26
Gi*	32.06	26.85	19.19	23.34	27.94

Table 6. 5 The RRI for the combination of five crime types and eight hotspot mapping techniques. The value in bold represents the highest value among the eight hotspot methods for each crime type

	Robbery	Aggravated Assault	Burglary	Larceny-Theft	Auto Theft
Risk-Based Thematic Mapping	0.47	0.69	0.95	1.09	0.96
Grid Thematic Mapping	0.72	0.76	0.81	0.93	0.81
STAC	0.57	0.53	0.56	0.63	0.54
NNHC	0.52	0.6	0.59	0.68	0.60
Risk-Adjusted NNHC	0.58	0.88	1.01	1.23	0.94
KDE	1.01	1.10	1.11	1.15	1.12
Local Moran's I	0.92	0.97	0.98	1.01	0.97
Gi*	0.81	0.82	0.89	0.95	0.87

6. 3 Spatial-Temporal Analysis of Crime Data

This thesis research utilized the hotspot plot and Kulldorff's space-time scan statistic to analyze the distribution of crime hotspots in space and time. The dataset are all reported Part 1 Crimes and five individual crime types from the city of Houston, TX in 2011. Both methods used one month as the temporal unit. Results from the two spatial-temporal analysis and mapping methods indicate that all Part 1 Crimes in Houston are most likely to occur around the center and southwest of Houston from September, 2011 to December, 2011.

Figures 6. 9 – 6.14 visualize the statistically significant spatial-temporal clusters detected by SatScan.

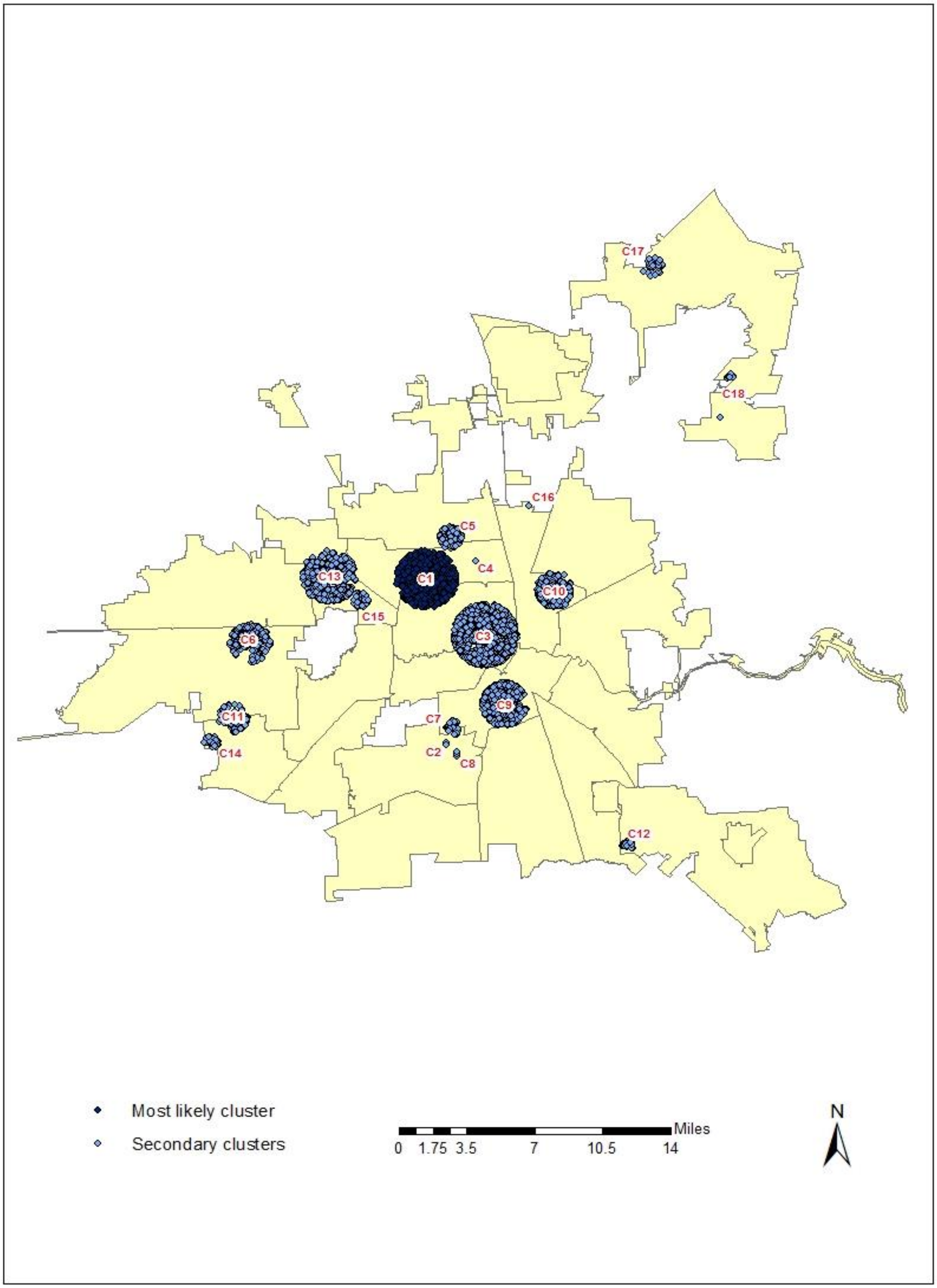


Figure 6. 9 Spatial-temporal clusters of all Part 1 Crimes for the city of Houston from Jan. 2011 to Dec. 2011

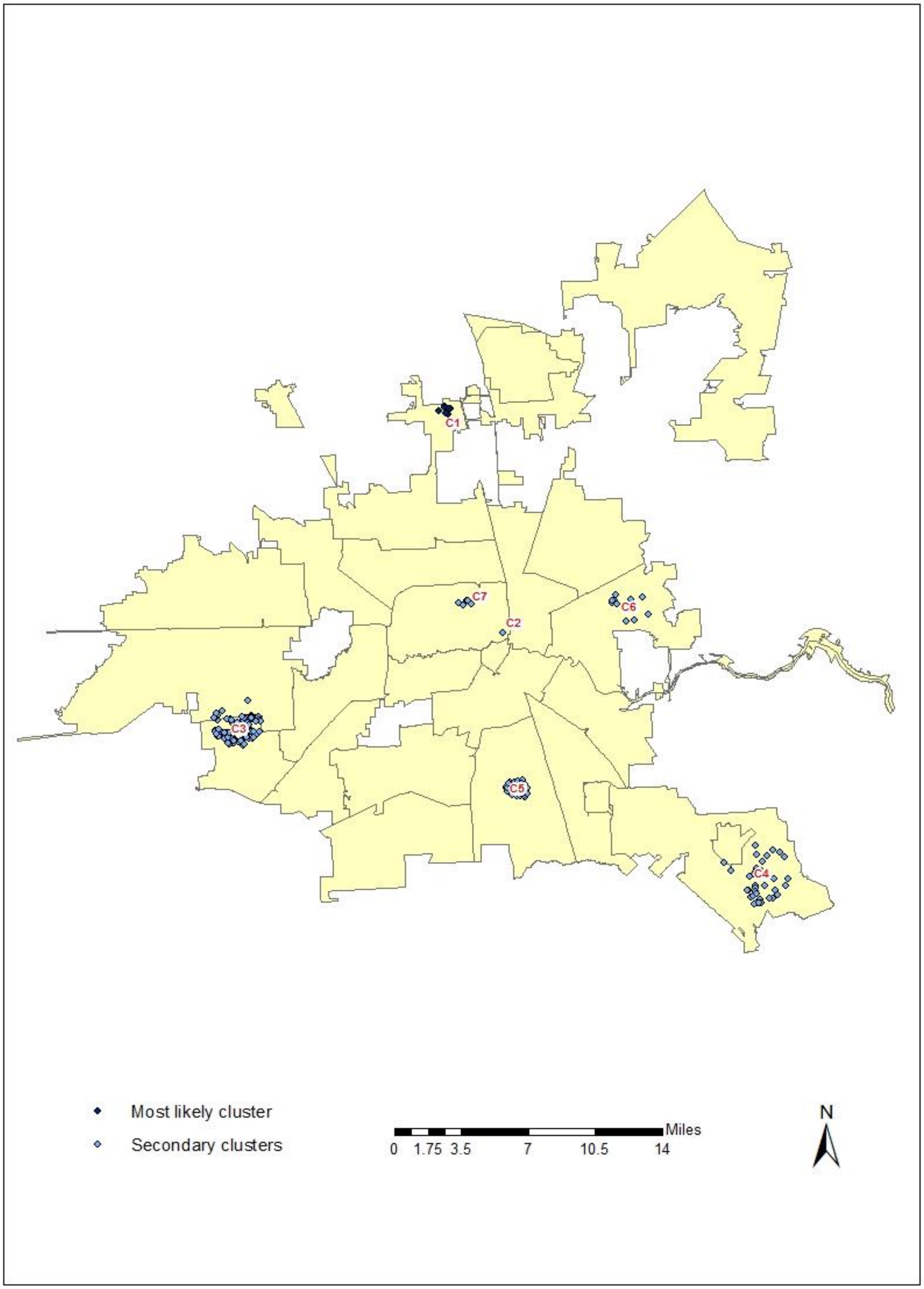


Figure 6. 10 Spatial-temporal clusters of aggravated assault for the city of Houston from Jan. 2011 to Dec. 2011

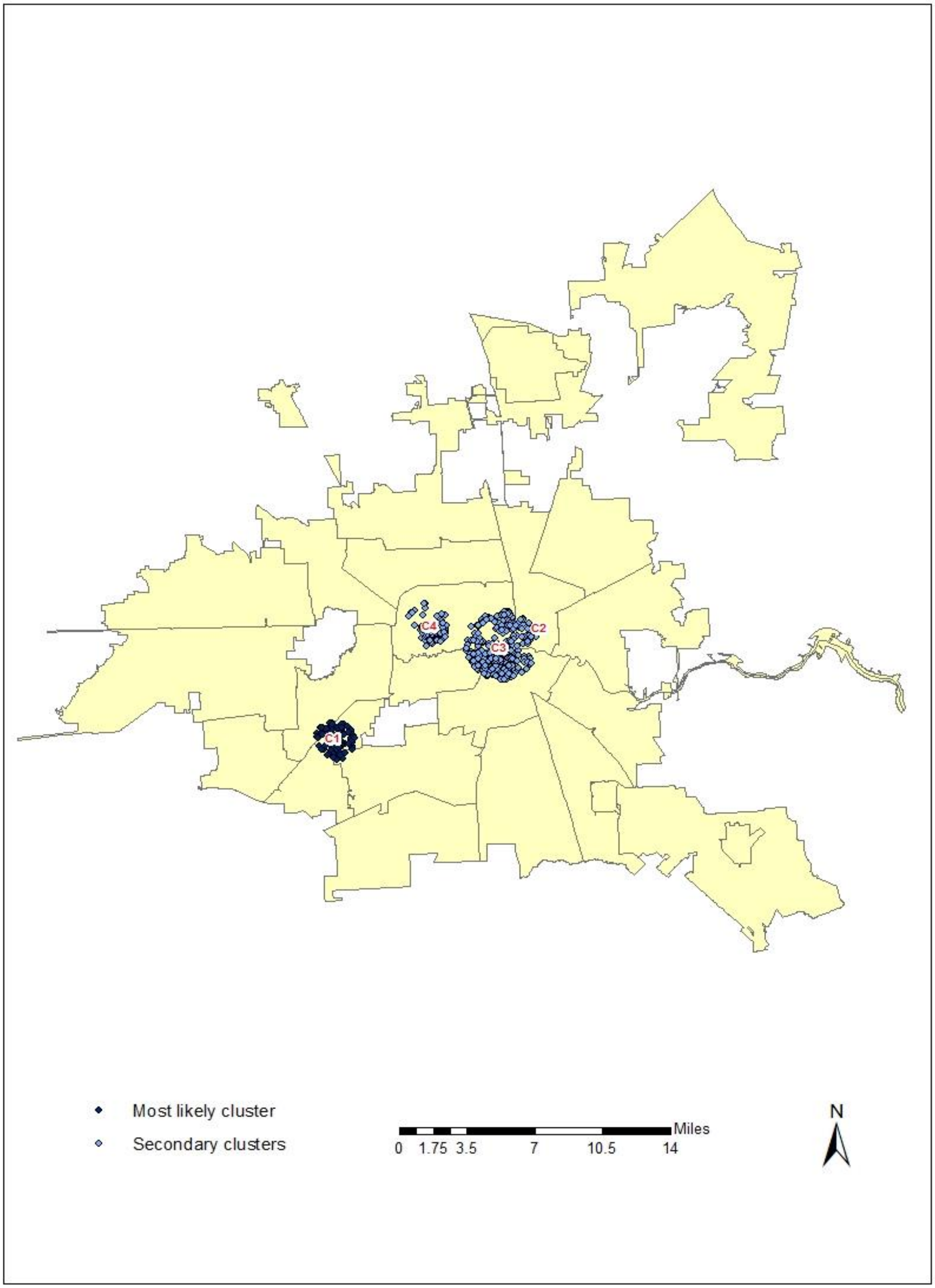


Figure 6. 11 Spatial-temporal clusters of auto theft for the city of Houston from Jan. 2011 to Dec. 2011

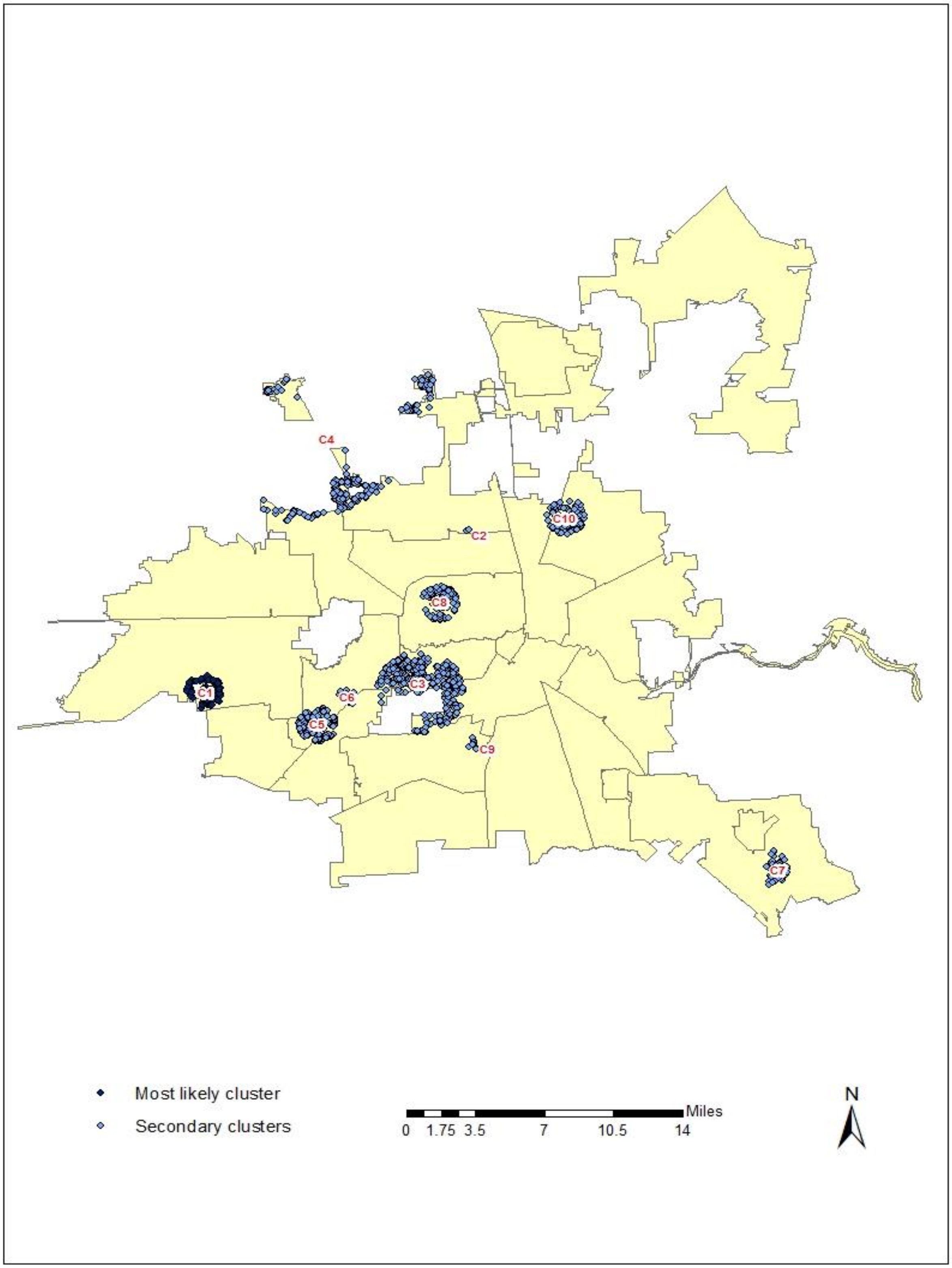


Figure 6. 12 Spatial-temporal clusters of burglary for the city of Houston from Jan. 2011 to Dec. 2011

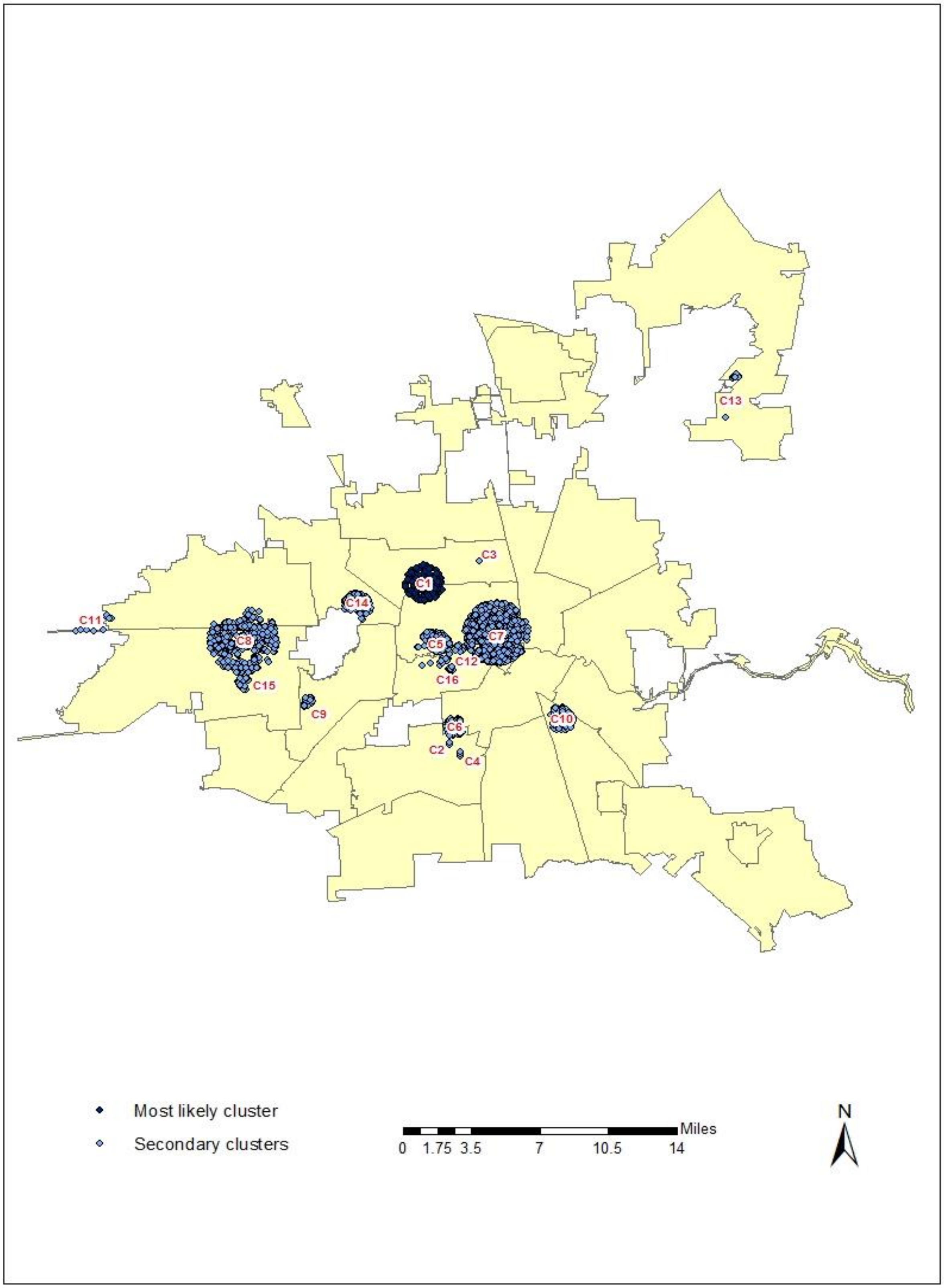


Figure 6. 13 Spatial-temporal clusters of larceny-theft for the city of Houston from Jan. 2011 to Dec. 2011

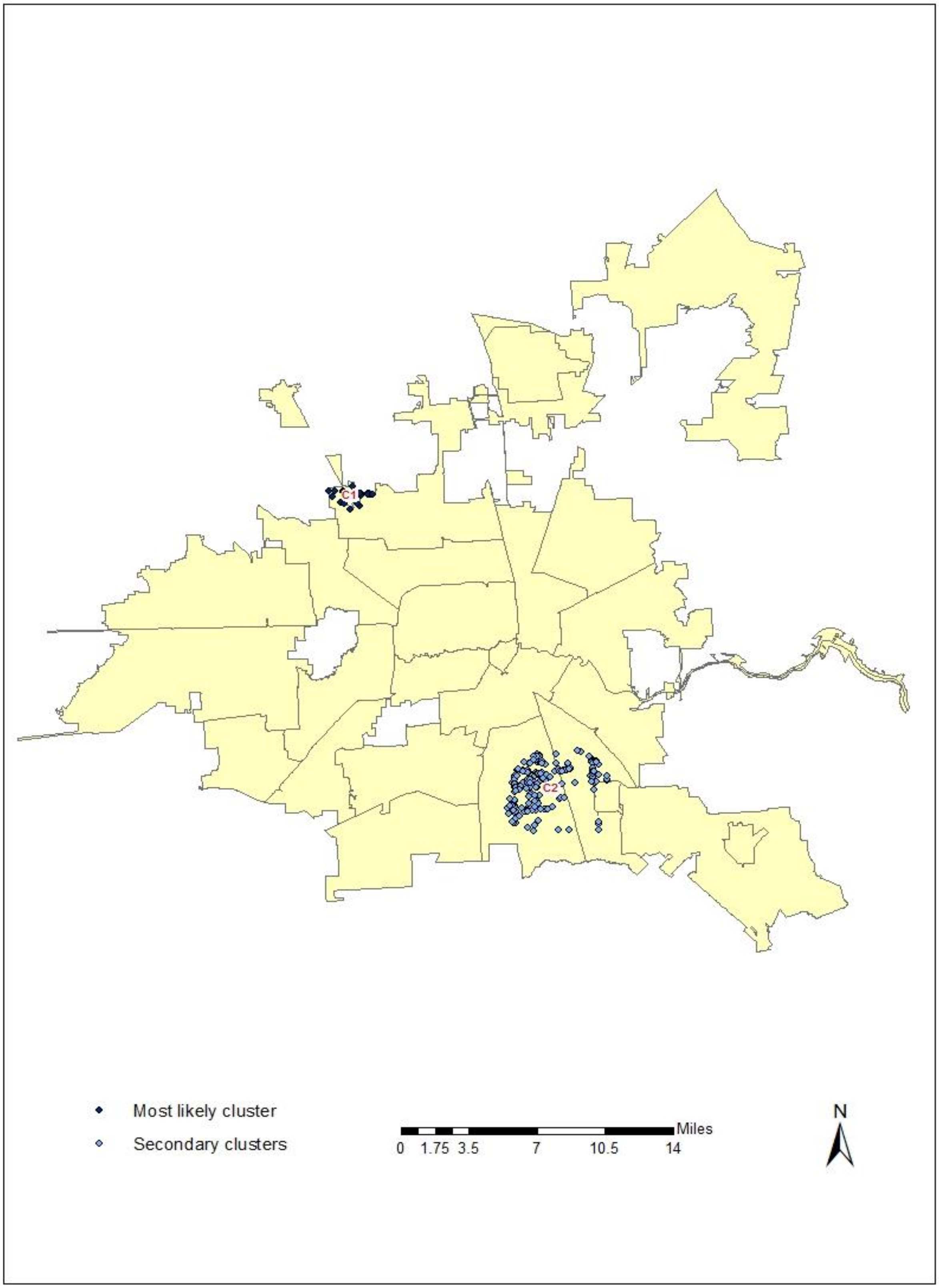


Figure 6. 14 Spatial-temporal clusters of robbery for the city of Houston from Jan. 2011 to Dec. 2011

The cluster with the smallest p value is the most likely cluster, which means this cluster is least likely to be due to chance. The secondary clusters are other detected clusters whose p values are also less than the user-defined significance level (here, 0.05). For the spatial-temporal analysis of all Part 1 Crimes, the most likely cluster is located in central north Houston. The other 17 secondary clusters are mostly located in the central, western and southern parts of the city. The most likely cluster lasts through October. The time periods for all secondary clusters are listed in Table 6. 6.

Table 6. 6 Spatial-temporal scan statistic results using SatScan

Crime Type	Cluster ID	Time Period	P value
All Part1 Crimes	C1	10/1 to 10/31	0.0000000020
	C2	3/1 to 4/30	0.0000000025
	C3	4/1 to 5/31	0.00000051
	C4	10/1 to 12/31	0.00000099
	C5	2/1 to 2/28	0.0000036
	C6	11/1 to 12/31	0.000062
	C7	10/1 to 10/31	0.000093
	C8	3/1 to 3/31	0.000097
	C9	4/1 to 4/30	0.0013
	C10	4/1 to 4/30	0.0016
	C11	9/1 to 9/30	0.0048
	C12	9/1 to 9/30	0.0058
	C13	1/1 to 2/28	0.011
	C14	12/1 to 12/31	0.012
	C15	7/1 to 8/31	0.013
	C16	2/1 to 5/31	0.018
	C17	3/1 to 3/31	0.029
	C18	9/1 to 9/30	0.046
Aggravated Assault	C1	5/1 to 5/31	0.0055
	C2	8/1 to 8/31	0.0086
	C3	8/1 to 8/31	0.012
	C4	9/1 to 11/30	0.020
	C5	6/1 to 6/30	0.024
	C6	8/1 to 8/31	0.039
	C7	5/1 to 5/31	0.050

(Table 6. 6 continued)

Crime Type	Cluster ID	Time Period	P value
Auto Theft	C1	11/1 to 11/30	0.0000033
	C2	11/1 to 11/30	0.00036
	C3	4/1 to 7/31	0.011
	C4	3/1 to 3/31	0.031
Burglary	C1	9/1 to 11/30	0.00067
	C2	2/1 to 2/28	0.00083
	C3	1/1 to 2/28	0.0050
	C4	11/1 to 12/31	0.0072
	C5	5/1 to 7/31	0.010
	C6	7/1 to 7/31	0.011
	C7	10/1 to 11/30	0.017
	C8	10/1 to 10/31	0.022
	C9	9/1 to 10/31	0.026
	C10	6/1 to 6/30	0.030
Larceny Theft	C1	10/1 to 10/31	0.000010
	C2	3/1 to 4/30	0.00001065
	C3	10/1 to 12/31	0.00001027
	C4	3/1 to 3/31	0.000011
	C5	11/1 to 12/31	0.000017
	C6	10/1 to 10/31	0.000047
	C7	4/1 to 5/31	0.000065
	C8	10/1 to 12/31	0.00022
	C9	3/1 to 4/30	0.0020
	C10	6/1 to 7/31	0.0037
	C11	5/1 to 5/31	0.0047
	C12	1/1 to 1/31	0.0051
	C13	9/1 to 9/30	0.012
	C14	7/1 to 7/31	0.024
	C15	7/1 to 8/31	0.035
	C16	5/1 to 5/31	0.040
Robbery	C1	6/1 to 6/30	0.0022
	C2	10/1 to 10/31	0.013

For the spatial-temporal analysis using SatScan of five individual crime types, the results vary from crime type to crime type. For auto theft, burglary, and larceny-theft, more clusters were detected. They are primarily located in the center and southwest of the city. Several small clusters detected by SatScan spread across the entire study area for aggravated assault crimes.

Only two clusters were identified for robbery. They are located in northwest and southeast of the city. The time periods for the five crime types could vary. They are also presented in Table 6. 6.

When examining the results in more detail, the time period for the most likely cluster (C1) for the crime types examined are ranging from September to December, 2011, except for aggravated assault and robbery. And the clusters are mainly distributed in the central and southwestern part of the city of Houston. Also, aggravated assault and robbery are two exceptions (the reason may be due to less amount of records of crimes for these two crime types). These results using SatScan correspond to the results using hotspot plot.

CHAPTER 7 CONCLUSION

With the advance of Geographic Information Systems (GIS) and crime theories, crime hotspot mapping and analysis have been drawn increasing attention. Crime researchers and practitioners have put a lot of effort into studying how crime hotspot mapping can be used to assist police decision makers with allocating their limited resources and manpower to areas where crime events are most likely to occur. This thesis research used all 2011 and 2012 reported Part1 Crimes data from the city of Houston, TX. Eight hotspot mapping methods were employed to produce hotspot maps and their corresponding predictive accuracies for all Part 1 Crimes combined. In addition, nine individual crime-type hotspot maps were created and the predictive accuracies were calculated. For each crime type, the "best" method among the eight hotspot mapping techniques was identified, after comparing the predictive accuracy results across the eight mapping techniques with each other. In addition, spatial-temporal analysis using hotspot plots and Kulldorff's space-time scan statistic were performed for the same crime dataset, and study area. Maps showing crime clusters which were statistically significant, both spatially and temporally, were created.

The results from this research could provide valuable suggestions for law enforcement agencies in Houston to adapt their decision-making strategy based on the type of crime involved. For example, if an area is predicted to have a high rate of robbery, then a deterrent force, such as the armed police patrol, should be used to control this area. Also, the hotspot map and its predictive accuracy for all crime types combined will help the police allocate their limited resources more effectively and efficiently. For instance, if an area is predicted to have a high rate of multiple crime types, then this place should be paid most attention to by the police. If one area is predicted to have a high rate of burglary, but another is predicted to have a similarly high rate of assault, then the area with the predicted high rate of assault should receive more patrols in the future.

The results in this thesis research indicate that the type of hotspot mapping method chosen markedly affects the predictive accuracy. Moreover, by using different measures of predictive accuracy, the extent to which hotspot methods affect predictive accuracy results varies, as well. For example, the hit rate yields the best predictions with the grid thematic mapping method. However, the kernel density estimation (KDE) method predicts future crime incidents the best if the PAI and the RRI are applied. Since the KDE method also yields a hit rate, this method could thus be identified as the most accurate method at predicting all Part 1 Crimes combined.

The kernel density estimation and the nearest neighbor hierarchical clustering are the two methods which result in the highest RRI and PAI across the five crime types selected. In contrast, for the hit rate, no single hotspot method consistently possesses the highest prediction across the five crime types.

In terms of the temporal factor, the spatial-temporal analysis shows that the spatial-temporal clusters vary for different crimes. Crimes were more likely to concentrate in the central and southwestern part of the city of Houston, TX.

One issue which has to be drawn particular attention to is related to the sampling method. In most of the social work study, the dataset used in the analysis consist of all observational records, which is to say, no sampling process was conducted to select the dataset to be analyzed. The approach used in this thesis research could be considered to be a social work approach. While in the field of engineering, a random design study is usually conducted to randomly select the records to be included in the analysis. The experimental design requires the knowledge of statistics. Further research could be focusing on this engineering approach.

Of course, this thesis research has some limitations. First, the crime data analyzed are limited to the nine Part 1 Crime types, which may not provide useful information for the analysis of other crime types. Second, the study area of this research is limited to the city of Houston, TX. The implications from the results of this research may thus not be applicable to other urban study areas. Third, although in this research the effect of hotspot methods and crime types on predictive accuracy has been investigated, other issues (e.g. study area, parameter settings, threshold selection, geocoding quality, etc.) may also contribute to the resulting predictive accuracy. Finally, the time span of the spatial-temporal analysis is two years, which may not be sufficient for preforming a credible and accurate spatial-temporal hotspot map for predicting future crimes.

Accordingly, future research could emphasize the following aspects. First, variations of other factors, such as the study area, parameter settings, and the threshold selection could be examined to investigate the effect that these factors have on the ability to predict future crimes. To implement this, crime data from alternative urban study areas should be evaluated, and a series of different sets of parameter settings and threshold selections should be investigated and their predictive accuracy results compared with each other. Second, Part 1 Crimes can also be categorized as violent or non-violent crimes. Redoing the analysis from this research with these two crime categories could also be carried out. Hotspot methods not selected for this thesis research could also be applied. Finally, for spatial-temporal analysis, cluster maps for each of the five of the nine individual crime types could be produced rather than just for the overall Part 1 Crimes combined.

REFERENCES

Anselin, Luc. "Local indicators of spatial association—LISA." *Geographical Analysis* 27.2 (1995): 93-115.

Anselin, Luc, et al. "Spatial analyses of crime." *Criminal Justice* 4 (2000): 213-262.

Bowers, Kate. "Exploring links between crime and disadvantage in north-west England: an analysis using geographical information systems." *International Journal of Geographical Information Science* 13.2 (1999): 159-184.

Bowers, Kate J., and Shane D. Johnson. "Measuring the geographical displacement and diffusion of benefit effects of crime prevention activity." *Journal of Quantitative Criminology* 19.3 (2003): 275-301.

Bowers, Kate J., Shane D. Johnson, and Ken Pease. "Prospective hot-spotting: the future of crime mapping?" *British Journal of Criminology* 44.5 (2004): 641-658.

Brantingham, Paul J., and Patricia L. Brantingham. *Patterns In Crime*. New York: Macmillan, 1984.

Chainey, Spencer. Webpage: http://www.ucl.ac.uk/scs/people/academic-research-staff/spencer-chainey/Slides/Spatial_significance_mapping (2008).

Chainey, Spencer, and Jerry Ratcliffe. GIS and Crime Mapping. John Wiley & Sons, 2005.

Chainey, Spencer, Svein Reid, and Neil Stuart. *When Is A Hotspot A Hotspot? A Procedure For Creating Statistically Robust Hotspot Maps of Crime*. Taylor & Francis, London, England, 2002.

Chainey, Spencer, Lisa Tompson, and Sebastian Uhlig. "The utility of hotspot mapping for predicting spatial patterns of crime." *Security Journal* 21.1 (2008): 4-28.

Cohen, Lawrence E., and Marcus Felson. "Social change and crime rate trends: A routine activity approach." *American Sociological Review* (1979): 588-608.

Eck, John, et al. "Mapping crime: Understanding hotspots." (2005): 1-71.

Gottlieb, Steven, Sheldon I. Arenberg, and Raj Singh. *Crime Analysis: From First Report to Final Arrest*. Montclair, CA: Alpha Publishing, 1994.

Hart, Timothy C., and Paul A. Zandbergen. "Effects of data quality on predictive hotspot mapping." (2012).

Illinois Criminal Justice Information Authority. "Spatial and temporal analysis of crime." State of Illinois (1996).

Johnson, Shane D., and Kate J. Bowers. "The burglary as clue to the future the beginnings of prospective hot-spotting." *European Journal of Criminology* 1.2 (2004): 237-255.

Kulldorff, Martin. "A spatial scan statistic." *Communications in Statistics-Theory and Methods* 26.6 (1997): 1481-1496.

Kulldorff, Martin. "Prospective time periodic geographical disease surveillance using a scan statistic." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 164.1 (2001): 61-72.

Kulldorff, Martin, et al. "Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico." *American Journal of Public Health* 88.9 (1998): 1377-1380.

Leitner, Michael, and Marco Helbich. "The impact of hurricanes on crime: a spatio-temporal analysis in the city of Houston, Texas." *Cartography and Geographic Information Science* 38.2 (2011): 213-221.

Levine, Ned. "The "hottest" part of a hotspot: comments on" the utility of hotspot mapping for predicting spatial patterns of crime"." *Security Journal* 21.4 (2008): 295-302.

Levine, Ned. "CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 4.0)." Ned Levine & Associates, Houston, TX and the National Institute of Justice, Washington, DC (2013).

Maguire, Mike. "Policing by risks and targets: Some dimensions and implications of intelligence-led crime control." *Policing and Society: An International Journal* 9.4 (2000): 315-336.

McCullagh, Michael J. "Detecting hotspots in time and space." *ISG06* (2006).

McLafferty, Sara, Doug Williamson, and Philip G. McGuire. "Identifying crime hotspots using kernel smoothing." *Analyzing crime patterns: Frontiers of Practice* (2000): 77-85.

Nakaya, Tomoki, and Keiji Yano. "Visualising Crime Clusters in a Space Time Cube: An Exploratory Data Analysis Approach Using Space Time Kernel Density Estimation and Scan Statistics." *Transactions in GIS* 14.3 (2010): 223-239.

Openshaw, Stan. *The modifiable areal unit problem*. Vol. 38. Norwich: Geo Books, 1983.

Ord, J. Keith, and Arthur Getis. "Local spatial autocorrelation statistics: distributional issues and an application." *Geographical Analysis* 27.4 (1995): 286-306.

Ratcliffe, Jerry H. "Aoristic signatures and the spatio-temporal analysis of high volume crime patterns." *Journal of Quantitative Criminology* 18.1 (2002a): 23-43.

Ratcliffe, Jerry H. "Damned if you don't, damned if you do: crime mapping and its implications in the real world." *Policing and Society* 12.3 (2002b): 211-225.

Ratcliffe, Jerry H. "Intelligence-led policing and the problems of turning rhetoric into practice." Policing & Society 12.1 (2002c): 53-66.

Ratcliffe, Jerry H. "Geocoding crime and a first estimate of a minimum acceptable hit rate." *International Journal of Geographical Information Science*18.1 (2004): 61-72.

Ratcliffe, Jerry H. "HotSpot Detective for MapInfo Help file Version 2.0." (2004).

Rengert, George F. "Auto theft in central Philadelphia." *Policing for prevention: Reducing crime, public Intoxication and injury, Crime Prevention Studies 7* (1997).

Seddon, K. and Napper, R. "Integrated crime management model discussion paper." Unpublished, New South Wales Police Service, Sydney(1999).

Tiefelsdorf, Michael, and Barry Boots. "The exact distribution of Moran's I." *Environment and Planning A* 27 (1995): 985-985.

Townsley, Michael. "Visualising space time patterns in crime: the hotspot plot." *Crime Patterns and Analysis* 1.1 (2008): 61-74.

Zeng, Daniel, Wei Chang, and Hsinchun Chen. "A comparative study of spatio-temporal hotspot analysis techniques in security informatics." *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference*. IEEE, 2004.

VITA

Shuzhan Fan received his bachelor's degree of geomatics engineering from Central South University in China in 2012. He served as a graduate assistant during his master's study in the Department of Geography and Anthropology at Louisiana State University. His main research area is crime analysis, spatial statistics, GIS and remote sensing. He also earned a minor in applied statistics from the Department of Experimental Statistics. He received a scholarship from Austrian Marshall Plan Foundation and will conduct his research in Austria in the summer, 2014. He will continue his PhD study starting from August, 2014 under the guidance of Dr. Michael Leitner in the Department of Geography and Anthropology at LSU.