A SURVEILLANCE MODELING AND ECOLOGICAL ANALYSIS OF URBAN RESIDENTIAL CRIMES IN COLUMBUS, OHIO, USING BAYESIAN HIERARCHICAL DATA ANALYSIS AND NEW SPACE-TIME SURVEILLANCE METHODOLOGY

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Youngho Kim, M.A.

*****

The Ohio State University
2007

Dissertation Committee

Professor Morton O’Kelly, Adviser
Professor Mei-Po Kwan
Professor Darla Munroe

Approved by

Adviser
Geography Graduate Program
ABSTRACT

This dissertation analyzes crime in both ecological and surveillance perspectives. In ecological perspective, many studies ignore spatial effects in the models, leading to inefficient and biased results. This dissertation, by applying Bayesian hierarchical analysis, accounts for spatial effect in the model and presents correct socio-demographic factors related to residential crime occurrences. In surveillance perspective, literature to date has limitations in presenting exact locations of crime hotspots and implementing continuous analysis over time. Use of population information is the main reason of the limitations in the literature. Because the population information is based on census administrative area unit and is only updated in decennial bases, corresponding hotspots involve approximations in both population size and their locations. However, this study handles the problem by applying a newly devised surveillance method, which uses only crime accounts over time without the use of population information. The goal of this dissertation is providing significant demographic factors of crime and crime hotspots in near real time base, which will contribute to crime control. This goal is achieved by 1) handling spatial autocorrelation and heterogeneity in the analysis, 2) visualizing spatial effects on a map, 3) enabling continuous surveillance over time, 4) providing precise crime hotspot locations, and 5) presenting local changes in clusters over time. The models presented in this dissertation is applied to residential crimes occurred in
Columbus, Ohio for the year 2000. Empirical results present significant demographic factors of residential crimes and locations of crime hotspots over time in near real-time framework.

Keyword: hierarchical Bayesian data analysis, surveillance, space-time surveillance, crime.
DEDICATION

To my grandmother in heaven and mother in Korea
ACKNOWLEDGMENTS

Over simplified, there are two groups of graduate students in every graduate program around the world: one is a brilliant group and the other is not necessarily so. The brilliant group has no problem at all regarding their eternal academic goals: Ph.D. and graduation. However, the mediocre group has a different situation. They can achieve the goals only if they meet a brilliant adviser.

Unfortunately, I did not belong to brilliant group. Seeing the past, I know for sure that I could not finish and graduate unless I met Dr. Morton O’Kelly. He made my graduate life from unfortunate to fortunate one. I really appreciate for his advice, endless efforts, encouragement, and sharp comments not only on my academic works, but also on my academic life.

I thank to Dr. Mei-po Kwan and Dr. Darla Munroe for their valuable comments and advise on my works. They also showed me how successful scholars behave including teaching, research, and interacting with students.

My special thanks go to my family: father, mother, sisters, aunt, and uncle. They have always been on my side and supported me.
VITA

January 29, 1975 ............................................. Born – Seoul, Republic of Korea

2000 ................................................................. B.A., Geography Education
                    Korea University

2003 ................................................................. M.A., Geography
                    The Ohio State University

2000 – present .................................................. Graduate Teaching and Research
                    Associate, The Ohio State University

PUBLICATIONS

Research Publication

1. Y. Kim, and M. O’Kelly, “A New Space-time surveillance of crime occurrence”
   Journal of Geographical Systems, in Review

2. Kim, Y. “An Analysis of Urban Economic and Social Status Effects to
   Residential Crimes using Bayesian Hierarchical Modeling”

3. Kim, Y. “Space-time measures of crime diffusion” In Artificial crime
   analysis systems: using computer simulations and geographic information
   systems, Edited by Lin Liu and John Eck, in press (2007)
FIELDS OF STUDY

Major Field: Geography
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>VITA</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Ecological analysis of crime using Bayesian hierarchical data analysis</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Prospective space-time surveillance of crime occurrences</td>
<td>5</td>
</tr>
<tr>
<td>2  LITERATURE REVIEW</td>
<td>11</td>
</tr>
<tr>
<td>2.1 Characteristics of urban crime</td>
<td>11</td>
</tr>
<tr>
<td>2.1.1 Urban communities in the aspects of crime occurrences</td>
<td>13</td>
</tr>
<tr>
<td>2.1.1.1 Social disorganization theory</td>
<td>15</td>
</tr>
<tr>
<td>2.1.1.2 Social control theory</td>
<td>16</td>
</tr>
<tr>
<td>2.1.1.3 Routine Activity Theory</td>
<td>18</td>
</tr>
<tr>
<td>2.1.2 Issues related to spatial crime data and their analytical perspectives in urban ecological analyses</td>
<td>21</td>
</tr>
<tr>
<td>2.1.2.1 Effects of spatial pattern in a spatial crime statistical analysis</td>
<td>22</td>
</tr>
<tr>
<td>2.1.2.2 Modifiable areal unit problem (Issues related using fixed boundary problems)</td>
<td>24</td>
</tr>
<tr>
<td>2.2 Surveillance Analysis</td>
<td>27</td>
</tr>
<tr>
<td>2.2.1 Clustering/cluster modeling: From space to space-time surveillance</td>
<td>29</td>
</tr>
<tr>
<td>2.2.1.1 Spatial Modeling</td>
<td>30</td>
</tr>
<tr>
<td>2.2.1.2 Space-time modeling</td>
<td>32</td>
</tr>
<tr>
<td>2.2.1.2.1 Retrospective Analyses</td>
<td>33</td>
</tr>
</tbody>
</table>
2.2.1.2.2  Prospective analyses ................................................................. 38

3  STUDY AREA AND DATA DESCRIPTION ................................................. 44
3.1  Columbus and State of Ohio .............................................................. 44
3.2  Crime Data ......................................................................................... 47

4  BAYESIAN HIERARCHICAL DATA ANALYSIS OF RESIDENTIAL CRIME
    OCCURRENCES .................................................................................. 51
4.1  Socio-demographic factors relevant to residential crimes ................. 52
4.2  Bayesian hierarchical modeling .......................................................... 54
4.2.1  General Framework ........................................................................ 54
4.2.2  Model specifications with random effect components .................. 55
4.3  Results ................................................................................................. 59
4.3.1  Model diagnostic and convergence .............................................. 59
4.3.2  Regression results ......................................................................... 61
4.3.3  Map decomposition ....................................................................... 65
4.4  Summary ............................................................................................. 70

5  PROSPECTIVE SPACE-TIME SURVEILLANCE ANALYSES OF
    RESIDENTIAL CRIMES ........................................................................ 72
5.1  Space-time Scan Statistic ................................................................. 72
5.1.1  Model description ........................................................................ 72
5.1.2  Numerical Results ................................................................. 79
5.2  An Application of Cumulative Sum version of Tango's Statistic .......... 86
5.2.1  Model description ........................................................................ 86
5.2.2  Numerical results ......................................................................... 89
5.3  An Application of a New Space-time Surveillance method .............. 92
5.3.1  Model description ........................................................................ 92
5.3.1.1  Conceptual description .......................................................... 92
5.3.1.2  Likelihood ratio tests ......................................................... 96
5.3.2  Numerical results ................................................................. 102
5.4  Model comparisons and Summary .................................................. 109

6  CONCLUSION ......................................................................................... 114
6.1  Results of the research...................................................................... 114
6.2  Future Research Direction .............................................................. 117

BIBLIOGRAPHY ....................................................................................... 123
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Effect of spatial autocorrelation on OLS estimates</td>
<td>24</td>
</tr>
<tr>
<td>3.1 Residential crimes and their UCR codes</td>
<td>47</td>
</tr>
<tr>
<td>4.1 Variable lists</td>
<td>52</td>
</tr>
<tr>
<td>4.2 Deviance summaries for the four hierarchical models</td>
<td>60</td>
</tr>
<tr>
<td>4.3 Summaries of Bayesian hierarchical regression models with posterior parameters and precision coefficients</td>
<td>62</td>
</tr>
<tr>
<td>4.4 Bayesian hierarchical model using the number of crime occurrences as a dependent variable</td>
<td>65</td>
</tr>
<tr>
<td>5.1 Results of space-time scan statistic ordered by cluster rank</td>
<td>83</td>
</tr>
<tr>
<td>5.2 List of low risk block groups included in the hotspots</td>
<td>85</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Wikstrom's diagram of key reasons for high crime rates in urban areas</td>
<td>14</td>
</tr>
<tr>
<td>2.2 A hypothetical model about criminal opportunities in offender's routine space by Brantingham and Brantingham</td>
<td>19</td>
</tr>
<tr>
<td>2.3 An Example of Knox test results</td>
<td>34</td>
</tr>
<tr>
<td>3.1 Columbus, Franklin County, and State of Ohio</td>
<td>46</td>
</tr>
<tr>
<td>3.2 Study area for Bayesian hierarchical analyses</td>
<td>50</td>
</tr>
<tr>
<td>3.3 Study area for prospective surveillance</td>
<td>50</td>
</tr>
<tr>
<td>4.1 Map decomposition of the Odds into the Covariates, Spatially structured, and Unstructured random effects</td>
<td>68</td>
</tr>
<tr>
<td>4.2 Map decompositions for the comparison between Model 1 and Model 4</td>
<td>69</td>
</tr>
<tr>
<td>5.1 Structure of spatial scan statistic</td>
<td>77</td>
</tr>
<tr>
<td>5.2 Structure of space-time scan statistic</td>
<td>78</td>
</tr>
<tr>
<td>5.3 Crime cases and Centroids organized by Census block group</td>
<td>80</td>
</tr>
<tr>
<td>5.4 Spatial pattern of crime hotspot distribution by space-time scan statistic</td>
<td>84</td>
</tr>
<tr>
<td>5.5 Patterns of crime hotspot distribution by CUSUM version of Tango's statistic</td>
<td>91</td>
</tr>
<tr>
<td>5.6 Generation of observed count for Date 4</td>
<td>94</td>
</tr>
<tr>
<td>5.7 Generation of expected count for Date 4 using average of previous counts</td>
<td>94</td>
</tr>
<tr>
<td>5.8 Area proportion of significant crime hotspots; $r = 0.5$, and $q = 3$ and 7</td>
<td>104</td>
</tr>
<tr>
<td>5.9 Area proportion of significant crime hotspots; $r = 1.0$, and $q = 3$ and 7</td>
<td>104</td>
</tr>
</tbody>
</table>
5.10 Locations and spatial patterns of crime hotspots for temporal unit 58 (June 17th to 19th ) ........................................................................................................................................................................ 107

5.11 Locations and spatial patterns of crime hotspots for temporal unit 78 (Aug. 19th to 21st ) ........................................................................................................................................................................ 108

6.1 A satellite image of crime hotspots (small scale) ........................................................................................................................................................................ 121

6.2 A satellite image of crime hotspots (large scale) ........................................................................................................................................................................ 122
CHAPTER 1

INTRODUCTION

1.1 Background

Map presentations and corresponding spatial analyses of crime incidents have a long history. Initiated from a key study by Shaw and McKay (1942) about locations of high crime areas in Chicago and corresponding neighborhood characteristics, spatial analyses of crime incidents have been the subject of many studies across several disciplines. Although there are numerous approaches to crime analyses, the common goal is to reduce and control crime incidents. For the goal, this study uses two approaches to crime analysis. One is an ecological analysis, which analyzes the relationship between crime incidents and socio-demographic characteristics of neighborhoods. By finding significant factors of crime occurrences, the analysis enables the reduction of crimes through generating preventive policies (Brantingham and Brantingham, 1991) and long-term urban planning associated with socio-demographic factors (Hancock, 2001). The second approach is a crime cluster analysis, which finds areas of higher crime incidents. The cluster analysis aims at providing information by which local councilors determine the areas of greatest need and correspondingly, police resources are more efficiently allocated.
These two approaches may seem disconnected to each other in analyzing crime occurrences. However this study asserts that the two approaches are tightly interconnected because crime occurrences are the results of human behavior which reflects both demographic features of a place (Bursik, 1988; Shaw and McKay, 1969) and crime opportunities a place provides at a given time (Cohen and Felson, 1979 p. 590). Consequently, a salient crime analysis should consider the two aspects of crime for the goal of crime analyses, which provides information to reduce crime. The two approaches in this study, therefore, aim to provide more reliable and empirical information than either one approach can. In particular, the ecological study would help improve relevant socio-demographic factors in the long term, and the clustering analysis enables rapid counter responses to sudden crime increases.

1.2 Ecological analysis of crime using Bayesian hierarchical data analysis

Ecological analyses of crime incidents have tried to explain why some neighborhoods have higher crime rates compared to others (Bursik and Grasmick, 1993; Sampson and Groves, 1989) and how neighborhood socio-demographic factors, such as relative deprivation, low socio-demographic status, and low economic opportunity, affect crimes (Block, 1979; Shaw and McKay, 1969). Recently, place-based theories, such as routine activity theory (Cohen and Felson, 1979), have drawn attention to relationship between space and crime. In particular, with the development of spatial modeling and statistical analyses, special features of spatial data have been considered (see, Anselin, 1988; Griffith and Anselin, 1988; Haining, 1990; Haining, 2003). Consisting of spatial
dependence and heterogeneity, spatial effects commonly cause inefficient or biased results in classical regression analyses. There have been many attempts to handle spatial effects, which stem from intrinsic structures of spatial data.

Spatial dependence, caused by loss of information from redundant cases among geographically nearby observations (Anselin, 1990; Haining, 1990 p. 40-41), invalidates statistical tests (Goodchild, 1996). In particular, spatial dependence causes significant tests misleading in regression models (Cliff and Ord, 1981; Kramer and Donninger, 1987). To handle spatial dependence, many models capture spatial autocorrelation in errors within a regression framework. For example, spatial regression models apply a spatial component, which reflects the spatial process of a dependent variable. Spatial filtering approaches by Getis (1990; 1995) spatially transform a dependent variable based on local $G_i$ statistic values. Another spatial filtering approach by Griffith (1996; 2000) applies eigenvectors generated from a spatial link matrix to capture spatial dependence.

Compared to spatial dependence, spatial heterogeneity indicates geographical variations of statistical properties in data (Anselin, 1990). Under spatial heterogeneity, most spatial statistical models generating global statistic values cannot account for regional variations of the values, leading to biased results (Anselin and Getis, 1992). Spatial heterogeneity is usually handled by specifying spatial or regional differences and providing spatially varying coefficient values. Some examples of techniques addressing spatial heterogeneity are the expansion models (Casetti, 1997; Jones and Casetti, 1992), locally different coefficient in spatial adaptive filtering (Foster and Gorr, 1986; Gorr and Olligschlaeger,
1994), geographically weighted regression (Brunsdon et al., 1996; Brunsdon et al., 1998),
and Bayesian estimation (Besag et al., 1991; Law and Haining, 2004).

Although there are many methods to handle such effects in spatial data, most methods are
incapable of considering both spatial dependence and heterogeneity. Particularly in
ecological crime analyses, the handling of spatial effects has only recently been
considered (Anselin et al., 2000; Messner and Anselin, 2004) in the process of explaining
regional differences of crime occurrences in the spatial context. Therefore, ecological
crime analyses require more specialized applications of spatial statistical tools.

Considering current limitations in literature about ecological crime analyses and
corresponding spatial effects, Bayesian hierarchical analysis handles both spatial
dependence and heterogeneity at the same time. In particular, random effect modeling in
Bayesian analysis has an advantage in that it can address not only spatial heterogeneity
(by applying unstructured random effects) but also spatial dependence (by applying
spatially structured random effects). In addition, Bayesian analysis can visualize spatial
effects extracted from crime data using map patterns of spatial effects. In spite of
substantial benefits of Bayesian analysis, there are very few applications to crime
analysis literature. Therefore, this study, which applies Bayesian hierarchical data
analyses to residential crime occurrences in Columbus, contributes to crime literature as
follows.
First, this study aims to provide a reliable analysis of residential crimes and corresponding significant socio-demographic covariates by accounting for spatial effects. In the process, this study compares several models, which are either with or without spatial components, so that the comparison among the models presents spatial effects on the analyses. The significant socio-demographic factors of residential crime expect to help city councilors and planners in the process of urban planning and management to reduce crime.

Second, a map decomposition technique is applied. The technique provides visualizations of spatial crime patterns and their spatial effects on the analysis. These visualizations help identify spatial crime patterns, such as high and low risk areas and show how the spatial crime patterns are formed in relationship with socio-demographic factors. In addition, the visualizations present how spatial effects influence crime rates in addition to relevant socio-demographic factors.

1.3 Prospective space-time surveillance of crime occurrences

The subject area of spatial clustering analysis has developed considerably in recent years (Lawson, 2005) and has led to greater use of geographical and statistical tools. Recently, combined with wide availability of regularly updated geo-referenced data, interest in spatial clustering analyses has been extended from spatial dimension (finding clusters in one time period) to space-time dimensions (analyzing multiple time periods in space). Considering that spatial patterns of phenomena are results of continuous occurrences over
time on space, studies focusing on only the spatial dimension ignore dynamically changing temporal patterns. Consequently, corresponding results only provide fixed information on hotspots despite their temporal variations. Considering these limitations, cluster analyses on space-time dimensions have many advantages because they can provide more realistic and adaptable information on hotspots, which changes over time. In particular, space-time cluster analyses enable more specific and detailed interpretations such as times, locations, and constantly changing sizes of hotspots.

Space-time cluster analyses are divided into retrospective and prospective analyses depending on their use of data and corresponding continuity of analyses (Sonesson and Bock, 2003). The two types of analyses are used for different purposes and answer different questions. Retrospective analyses are characterized by a single attempt to find clusters based on past and fixed information. Therefore, the resulting clusters indicate hotspots in a given study area in the past. Retrospective analyses can be applied to estimate prevalent spatial phenomena and their comparison in different regions.

Prospective analyses contrast to retrospective analyses by testing newly emerging clusters and their locations within near-real time bases under the availability of continuously updated data information (Rogerson, 1997). The emergence of prospective analyses reflects needs for repeated analyses designed for regularly updated data over time. Particularly, many cities and municipalities are maintaining crime or disease registries, and many organizations continuously collect information of various occurrences. In particular, incidents are added to the data registry on daily, weekly, or monthly bases,
depending on the type of occurrences and capacities of institutions. In spite of the large availability of the dataset, many existing space-time clustering analyses have been developed for retrospective analyses for temporally fixed data (Farrington and Beale, 1998 p.98; Lawson, 2005; Rogerson, 1997). However, the retrospective analyses have limitations for surveillance purposes because the retrospective analyses neither provide exact hotspot locations in a timely manner nor help take instant preventive actions. However, prospective analyses handle the problems by detecting newly emerging clusters, so that the analyses “offers the opportunity to provide proactive, early detection of raised incidence” (Elliott and Wartenberg, 2004) and enable ongoing collection, analysis, and interpretation of outcome-specific data for prevention and control (Thacker and Berkelman, 1988).

Given the aims and importance of surveillance, a proper geographical surveillance method needs to provide exact locations of clusters and their accurate spatial ranges. In addition, the surveillance method should be applicable in near real-time as on-going new data are registered. However, many currently existing surveillance methods do not necessarily fulfill these requirements because of several restrictions.

First, the use of population information in prospective surveillance commonly results in problems in generating expected counts. In many surveillance models (Kulldorff, 1997; Kulldorff, 2001; Kulldorff and Nagarwalla, 1995; Rogerson, 1997), the size of population at risk is commonly used to calculate the number of expected cases and, further, applied to likelihood ratio calculation and significance tests. Population information is generally
updated decennially. Consequently, the number of expected cases of an area is constant for these 10 years. If a prospective surveillance tests a study area in daily, weekly, monthly, or yearly bases, the analysis is based on an unrealistic assumption that the population is fixed for 10 years, completely disregarding changes in the population. The problem is directly related to the sensitivity of the analysis and large changes in underlying population may generate biased results.

Second, the use of area data involves approximations in the locations and the ranges of hotspots. Most currently available population data are based on Census administrative units. Since expected values are generated by population at risk information, area data are applied in many surveillance analysis. Even when point based individual case data are available, the point data must be aggregated based on the area units for the use of expected values. In the process, precision of hotspot locations and ranges are lost because Census administrative units impose a limit in spatial resolution. As a result, hotspots based on this administrative area unit do not necessarily present true spatial hotspot ranges. A prospective surveillance model may not require population information (e.g. Kulldorff et al., 2005). However, results of the analysis are still area-based with spatially approximated hotspot locations and their spatial ranges because the model uses administrative area units to aggregate cases and generate expected values.

Third, models (Rogerson and Sun, 2001; Rogerson, 1997; Rogerson and Yamada, 2004) testing for global changes in spatial patterns cannot provide locations and ranges of hotspots in the surveillance. The models measure significant deviations from prior spatial
patterns of occurrences, signaling alarms when spatially clustered and excessive occurrences are observed. However, the use of global spatial pattern measures, such as Tango’s statistic (Tango, 1995) and Nearest Neighbor statistic (Clark and Evans, 1954), cannot provide locations and ranges of hotspots. The problems restrict the applications of surveillance analyses, where the goals are to contribute to rapid and efficient responses for the emerged hotspots.

This study provides a new surveillance method for finding space-time clusters with an application to residential crime occurrences in Columbus, Ohio. The new method aims to handle the limitations of existing surveillance methods with several features. First, the number of previous crime occurrences generates the expected values rather than the underlying population at risk. For the purpose of prospective surveillance analyses, which detect sudden increase of a phenomenon, direct comparisons among occurrences over time appears to be more reasonable. In addition, not being dependent on population data enables continuous and more frequent surveillance analyses, given that the availability of population data encumbers prospective surveillance with its slow updates.

Second, an application of local windows for hotspot identification overcomes the problems of using area data, such as Census data structured by administrative area units. Particularly, user defined circular local windows are applied in the analysis. The use of the local windows enables to provide more precise hotspot locations and ranges by presenting hotspot areas with overlaid local window. Consequently, the shape of hotspots is neither necessarily circular nor bounded by administrative area units.
Third, the significance of the hotspots is tested by previous occurrences of the same area using bootstrap permutation. Doing repeated analyses in surveillance commonly causes multiple testing problems, which lead to type I error (indicating false hotspots). Bootstrap permutations makes it possible to reduce the problems.

Overall, the new surveillance measure provides insights into local changes over time in its application to residential crime occurrences. Consequently, the analysis contributes to more efficient police resource allocations by providing near-real time surveillance depending on data availability. In addition, this study provides valuable information for urban planning to reduce crime incidents. This research only presents an application to crime occurrences. However, the same measure can be applied to other fields such as epidemiology.
CHAPTER 2

LITERATURE REVIEW

2.1 Characteristics of urban crime

Urban areas are defined by their large population sizes and corresponding high population densities (US Census Bureau, 2007). These features of urban areas are reflected in their landuse patterns and demographic characteristics, such as diversified and compact landuse, population heterogeneity, and high levels of anonymity among residents.

In a crime perspective, the urban characteristics draw opportunities for crime occurrences, resulting in higher crime rates than rural areas (Herbert, 1982; Krivo and Peterson, 1996; Wikstrom, 1991). High crime rates in urban areas have been indicated by many studies. Balwin and Bottoms (1976) provide examples of a case study in Sheffield. Brantingham and Brantingham (1984 p. 152-155) present a positive relationship between crime rates and urbanization levels: “while generally speaking larger cities have higher crime rates than smaller towns, this relationship is, in most countries, not without its exceptions.”

Regarding demographic features of urban areas, population size and urban heterogeneity are the major factors in explaining urban crimes. Large population size, as a characteristic
feature of an urban area (Wirth, 1938), results in frequent stranger-to-stranger contacts among urban residents (Bottoms and Wiles, 1995; Wikstrom, 1995). Combined with various types of human activities in diverse space, contacts among strangers in different parts of urban space generally increase crime incident rates. However, it is not correct to generalize a simple positive relationship between urban population density and crime rates. This is because various urban landuse, such as residential, business, and entertainment districts, and their corresponding social circumstances generate different types of encounters, and the various encounters cannot necessarily lead crime occurrences. In this aspect, Wikstrom’s (1991) statement is valid that criminal opportunities and conditions are likely to vary between areas of a city, depending on its land use and its social composition.

Urban heterogeneity indicates various mixtures of people and corresponding various activities a city provides, such as industry, commerce, leisure, entertainment, and other social services (Bottoms and Wiles, 1995). Even though the term urban heterogeneity has a broad range of meanings as indicated above, it commonly refers to ethnic heterogeneity and mostly to population composition in criminology literature. Therefore, regarding urban heterogeneity, many crime studies provide the effect of population compositions to crime occurrences. In an early study, Shaw and McKay (1969 p. 153) referred to population composition as an important factor in crime occurrences because their analysis showed that delinquency rates were higher in areas of ethnic minorities. In Shaw and Mckay’s data (1969 p. 155), the delinquency rate in areas with over 70 percent ethnic
minorities was more than double the rate in areas of maximum heterogeneity (for comparison, 50-59 ethnic minorities).

2.1.1 Urban communities in the aspects of crime occurrences

Many studies claim that higher urban crime rate is not only the result of increased crime opportunities a city provides, but also the result of weaker informal social control\(^1\) caused by superficial human relationships. Figure 2.1 by Wikstrom (1991) shows key factors of high urban crime rates. The diagram explains the increased crime occurrences through urbanization process from the perspectives of greater crime opportunities and weaker social control.

\(^1\) The term, social control, generally defined as the capacity of a group to regulate its members according to desired principles – to realize collective, as opposed to forced, goals (Janowitz 1975).
Many theories in criminology provide greater crime opportunities and weaker social control of urban areas as key factors regarding high urban crime rates. In understanding urbanization and corresponding crime occurrences, the theories provide theoretical backgrounds on how crimes occur in urban areas. The following sub-chapters provide brief summaries of three major theories about urban crimes: social disorganization theory, social control theory, and routine activity theory.
2.1.1.1 Social disorganization theory

Social disorganization theory, originating from Shaw and McKay (1942), became a central component of American criminology for many years. This theory generated a number of related studies in the tradition and expanded the focus from delinquency to crime in many studies. Social disorganization theory explains the disruption of community’s social organization using three structural factors; low economic status, population heterogeneity, and residential mobility (Bursik, 1988). In particular, impoverished neighborhoods tend to experience high rates of population turnover; as economic status of residents improves, the neighborhoods are abandoned. The neighborhoods also tend to have high rates of population heterogeneity from rapid changes in population composition because many incoming populations are characterized as ethnic minorities and even illegal immigrants from foreign countries. Furthermore, relatively cheap housing price attracts these new incoming populations. With these two processes combined with population turnover, informal social structure, which maintains social order in a community, often fails to develop.

With regard to social disorganization theory, Bursik (1986) suggests that cultural fragmentation from population heterogeneity and instability (presented by population composition and residential mobility) makes teenagers incapable of keeping their traditionally descended moral values, which eventually lead to delinquent behaviors. In
addition, the disruption of social order disables communities to provide common values
to their residents and maintain effective social controls\(^2\) (Sampson and Groves, 1989).

2.1.1.2 Social control theory

In spite of the significance of social disorganization theory, later works indicated several
shortcomings of the theory (Bursik and Grasmick, 1993; Kornhauser, 1978). Particularly,
it was noted that social disorganization theory ignored inter-linked community structures
based on social network\(^3\). Given that the regulatory capacity of a community is formed
and enforced through interactions and communication among members (Sampson, 1986),
sufficient level of social networks is critical to form and keep social control in the
community. Consequently, social control becomes more important in the crime control of
a neighborhood through positive networks among residents. For instance, when residents
of a neighborhood form local social ties based on their social network, the increased
social control makes the residents better able to recognize strangers and are more apt to
engage in guardianship behavior against victimization (Skogan, 1986; Taylor et al., 1984).

---

\(^2\) In the context, social control does not imply formal regulation or forced conformity by institutions such as the police, but it indicates “the capacity of a group to regulate its members according to desired principles to realize collective, as opposed to forced, goals.” (Janowitz, 1975)

\(^3\) Many types of social networks are considered such as networks within a neighborhood, networks among residents and local institutions, and networks among local representatives of the neighborhood and external actors, institutions, and agencies.
Literature about social control provided its empirical effects on crime in various aspects. For the most part, the studies found the effects of social networks within/inter
neighborhoods to social controls and its corresponding influence to crime rates in the neighborhoods. Bursik (1988) and Bellair (1997) found that strong interactions between residents and local institutions enforce social network in a neighborhood. In particular, Bellair (1997) showed that the strong social networks formed strong social bonding among residents. In his later study, Bursik (1999) noted the influence of social control on individual levels that neighborhoods with high social control had strong disrespect for people with criminal records. Peterson et al. (2000) measured the effect of local institutions in crime restrictions. Peterson assumed that local institutions, as the centers of local social networks, would contribute forming social control in a neighborhood. In her study, the existence of local institutions did not have a significant effect on the levels of criminal violence. However, these institutions reduced violent crime in extremely deprived neighborhoods.

Social control demonstrates that local social ties often influence crime and violence. However, social control does not necessarily explain the mechanism by which social structure actually facilitates social control of neighborhood residents (Coleman, 1988; Coleman, 1990). Based on this limitation, Coleman provided the concept of social capital. In his explanation, social capital is a structural relationship which facilitates public goods by linking residents into a network. On a neighborhood scale, healthy acquaintance among members would result in social capital in the form of guardianship and
supervision for teenager groups in the neighborhood. In this aspect, social capital can be linked with social disorganization theory in a straightforward manner; lack of social capital is one of the primary features of socially disorganized communities (see Putnam, 1993)

### 2.1.1.3 Routine Activity Theory

Routine activity theory, suggested by Cohen and Felson (1979 p. 590), explains crime occurrences by the routine of people in any space and time. Routine activity theory is based on a crime opportunity\(^4\) that is mostly affected by the routine activities of people including offenders, victims, and surveilling guardians. In the theory, crimes occur in the place where the offender happens to be, as a result of his/her daily life-choices. Thus, the offender’s daily life patterns would influence the location of offending behavior even when he/she has already decided to commit an offence (Brantingham and Brantingham, 1981). In this context, Brantingham and Brantingham postulate that most offenders would not commit offences in unfamiliar areas. Instead, they argue that offences are most likely to occur where criminal opportunities intersect with familiar areas as shown in Figure 2.2.

---

\(^4\) The crime opportunity is decided by attractiveness of target and security to initiate criminal action (Bottom, 1994)
Several studies provide evidence of the Brantingham and Brantingham’s model (Brantingham and Brantingham, 1991 p. 239-251; Figlio et al., 1986 Part 2; Rhodes and Conly, 1981). Particularly, Rengert and Wasilchick’s (1985) interview study of imprisoned adult burglars in Delaware County, Philadelphia suggests how crime offenders’ acquaintance of space affects their preference for an offence location. When asked for an evaluation of areas as a potential burglary site, the responses for prepared areas were generally nearby areas from burglars’ residences even when these areas were remotely located from affluent neighborhoods. In addition, results from other questions confirm that the preferred crime targets are usually located within more familiar areas to burglars. Consequently, Rengert and Wasilchick (1985) confirmed that burglary sites were clustered in areas closest to the offenders’ normal routes to work and recreation. In
addition, Wilkstrom (1991 p. 203-206) study about offender’s rational of target attractiveness presented similar results, indicating nearby targets from residential areas becomes far more attractive than remote (even more profitable) targets. Studies about homicide show similar results that offenders are disproportionately involved in acts of violence near their home (Block, 1977; Curtis, 1974; Reiss and Roth, 1993).

Routine activity theory provides a theoretical background to study spatial patterns of crime, which have clusters. In other words, the theory brought huge interest in crime places rather than individual or community characteristics in crime analyses (Weisburd, 2005). Thus, in routine activity theory, criminal events require not only motivated offenders, but also suitable targets and the absence of a capable guardian such as a police officer. Spatial patterns of crimes, including crime hotspots, do not necessarily follow traditional ideas about crime and communities, which only focuses on offenders’ socio-demographic status. Instead, spatial crimes require more sophisticated and appropriate analysis methods, providing explanations why and how certain places have more crime occurrences in the context of crime opportunities.
2.1.2 Issues related to spatial crime data and their analytical perspectives in urban ecological analyses.

In quantitative spatial data analyses, intrinsic features of spatial data commonly become an obstacle in analyses. In particular, the presence of spatial pattern in the data are the major concerns. The presence or absence of spatial pattern is commonly indicated by the concept of spatial autocorrelation. Spatial autocorrelation presents the level of spatial association in the pattern (Anselin, 1988; Griffith, 1987). In particular, positive spatial autocorrelation exists when high values in a place tend to be associated with high values at nearby locations, or low values with low values for the neighbors. In contrast, when high values at a location are surrounded by nearby low values, or vice versa, negative spatial autocorrelation is present.

Haining (2003 p.21-22) provides four different spatial processes as factors generating spatial autocorrelation: diffusion process, exchange and transfer, interaction, and dispersal process. Though spatial distribution and dynamics of urban crimes are generally referred as main factors forming spatial autocorrelation in the patterns (Brantingham and Brantingham, 1981; Brown and Oldakowski, 1986; Bursik and Grasmick, 1993; Harries, 1980; Rengert, 1980), explicit causes of spatial autocorrelation are different depending on the location and the crime types. Messner et al. (1999) suggested that the diffusion process of crime was a major factor in spatial autocorrelation in their study of homicides in St. Louis. Brown (1982) explains spatial autocorrelation with spatial distributions of
offenders and corresponding socio-demographic factors in Chicago. Combined with routine activity theory, the daily routine of offenders is noted as another source of spatial autocorrelation (Felson, 1994). Sampson and Morenoff (2004) propose that improperly processed area boundaries for neighborhoods is a source of spatial autocorrelation. Given that neighborhoods are interdependent units characterized by a function relationship, boundaries set by census administrative units cannot properly reflect ecological features among different neighborhoods.

2.1.2.1 Effects of spatial pattern in a spatial crime statistical analysis

Combined with the identification of spatial autocorrelation, the effects of spatial autocorrelation have recently drawn interest in crime analyses. Given that spatially aggregated datasets, such as census area units, are commonly applied to ecological analyses of crime, it is not uncommon to observe spatial autocorrelation. In particular, when the aggregated area data separates a continuous (similar) crime variable, the level of spatial variation within the variable becomes dependent upon proximity among area units. Consequently, improperly processed area aggregation lead to spatial autocorrelation and cannot sustain internal uniformity of each area.

The main issues and problems related to spatial autocorrelation come from the violation of the assumption of independent observations in general statistical analyses. The
violation leads to the loss of information of a variable. When spatial autocorrelation is dominant, the information of the variable is less than what is expected from independent observations. It is because the information is redundant among the existing observations (Haining, 1990 p. 40-41). This loss of information may invalidate some statistical tests because there are fewer effective degrees of freedom in a test (Goodchild, 1996).

In the context of Ordinary Least Square (OLS) model, the presence of spatial autocorrelation in disturbances may be results of two different forms, one from model mis-specification and the other from existence of spatial process in a correctly specified model (Anselin, 1988; Anselin and Griffith, 1988; Griffith and Layne, 1999, p. 9-10; Haining, 1990, p. 21-26; McMillen, 2003; Tiefelsdorf, 2000). The results from a different source of spatial autocorrelation in a standard OLS model, $Y = X\beta + u$, are introduced in Table 2.1. Spatial autocorrelation in the disturbance can be caused from (A) underlying spatial processes in the disturbance under a correctly specified model or (B) model mis-specification formed by omitting important variables containing spatial pattern. In the spatial process perspective (A), certain spatial processes operate and assimilate disturbances by distance leading to over / under estimation of standard error. In the mis-specification perspective (2) and (4), the unestimated components in exogenous variables are represented by the error term. As a result, the correlation between disturbance and exogenous variables in the model causes bias in the parameter coefficient together with the biased standard error (Fomby et al., 1984, p. 204; Fox, 1997, p. 126-129). The biased standard error by spatial autocorrelation comes from violation in OLS assumption of
stochastic independence in the disturbance. Consequently, the biased standard error makes the inferential test invalid. At the same time, under the mis-specification perspective, the biased estimates of parameter coefficients make the model interpretation misleading.

<table>
<thead>
<tr>
<th>Mis-specification (B)</th>
<th>Spatial Process (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>β: unbiased</td>
</tr>
<tr>
<td></td>
<td>$SE_β$: correct</td>
</tr>
<tr>
<td>Yes</td>
<td>β: biased</td>
</tr>
<tr>
<td></td>
<td>$SE_β$: incorrect</td>
</tr>
</tbody>
</table>

($SE_β$: standard error of coefficient $β$)

Table 2.1 Effect of spatial autocorrelation on OLS estimates

2.1.2.2 Modifiable areal unit problem (Issues related using fixed boundary problems)

In using spatially aggregated area data, there is a general problem that many statistical properties of variables defined over a set of areal units vary according to how the spatial units of a study are constituted. Since spatial aggregation is a necessary step in generating
areal data for application of area based statistical analysis, the problem of dealing with areal data has been a crucial factor in spatial analysis. There have been many studies addressing this issue since Gehlke and Biehl (1934), who showed that differences could be observed between results obtained using data for areal units where different sets of zones were used.

The problem coming from aggregated spatial data involving various spatial scales and different surface partitioning schemes is called the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984; Openshaw and Taylor, 1979; Wong, 1996). The MAUP has two aspects: the scale effect and the aggregation/zoning effect. First, the scale effect implies different statistical results from the same set of data when the information is grouped at different levels of spatial resolution. Second, the zoning effect indicates that data on same scale can give various statistical results depending on how the areal unit’s boundaries are drawn. Since the result can be different depending on scale and zonation, the magnitude of the MAUP effect can be large. For example, a significant spatial pattern on one geographical scale and zonation may disappear in another. Furthermore, statistical relationships among different spatial units might increase or decrease without any tendency. More importantly, this indicates that research results at a given spatial configuration are not decisive but uncertain because the results could vary on levels of spatial resolution (scale effect) and configuration of zonation (aggregation/zoning effect) (Fotheringham and Wong, 1991; Openshaw, 1984).
In intuitive explanation without spatial statistical knowledge, the MAUP would be explained as a result of a mismatching and corresponding relationship between physical spatial unit\(^5\) and operational spatial units\(^6\) in the study area (Anselin and Getis, 1992). Once physical spatial units are different from operational spatial units, different spatial aggregation and zonation would always generate different results.

To understand the cause of MAUP from a spatial statistical aspect, it is necessary to understand the characteristics of spatial data. Related to this, Tobler’s first law of geography (1969) is widely known that spatial data are usually characterized by spatial dependence which can be defined as similarity in an attribute of nearby spatial units on one another. In addition, the similarity of an attribute is not homogeneous across geographic space, and may even vary by direction. Given the features of spatial data, as spatial aggregation proceeds to form larger units, the attribute values of all units will be represented by summary measures such as a mean or a weighted mean (Arbia, 1989; Wong, 1996). Consequently, this smoothing process causes uniqueness of each area unit, reduction of dissimilarity among units, and decreased variance for the whole area (Fotheringham and Wong, 1991). This is similar to the partitioning/zonation process. Utilizing different combinations of spatial neighbors results in different degrees of spatial dependence, rendering different analytical results. In this context, Arabia’s (1989)

\(^{5}\) Physical spatial unit is a physically decided areal unit by a researcher through aggregation and zonation in study area. Census areal units such as block, tract, and counties are examples.

\(^{6}\) Operational spatial unit is a spatial extent of a phenomenon’s spatial process such as spatial interaction, spatial externalities, and neighborhood effects.
comment that variance will never significantly change with any level or way of spatial aggregation when values are randomly distributed across space is particularly noteworthy.

Generally known effects of MAUP is summarized as follow under spatial aggregation process: (i) The variance of a variable decreases (Fotheringham and Wong, 1991; Wong, 1996); (ii) The correlation between two variables increases (Fotheringham and Wong, 1991; Gehlke and Biehl, 1934; Openshaw, 1984; Wong, 1996). Even though a higher correlation under a larger scale is generally accepted as a effect of MAUP, Flowerdew et al. (2001) pointed out that this is not universally true. Instead, they provided two circumstances in which higher correlations are observed as a result of aggregation: when grouping according to values of an attribute and when grouping by spatial proximity in which the units have positive spatial autocorrelation. However, aggregation of data which have negative spatial autocorrelation or totally random grouping (regardless of the values of an attribute or spatial location) will not change the correlation. (iii) R-square value of the linear regression model increases (Amerhein, 1995; Wong, 1996); (iv) Spatial autocorrelation decreases (Anselin and Getis, 1992).

2.2 Surveillance Analysis

The traditional focus of research and theory in criminology has been on individuals and communities (Sherman, 1995a). Criminologists have investigated why certain individuals
become criminals (e.g., see Gottfredson and Hirschi, 1990; Raine, 1993) and why some communities have unusually high crime rates than others (e.g., see Bursik and Grasmick, 1993; Sampson and Groves, 1989; Shaw and McKay, 1969). Usually, research on individuals and communities provides several significant factors relevant to crime, such as relative deprivation, low socio-economic status, and lack of economic opportunity. Traditional approach to crime emphasizes the role of social institutions such as family and community. Braga (2001) notes that in the traditional approach the ability of police to prevent crime had been considered not as important as the social institutions. For example, several studies (Greenwood et al., 1977; Spelman and Brown, 1984) supported the view of limited police role in crime control.

Recently, place-oriented crime prevention strategies began to play central roles in crime prevention research and policy (Eck and Weisburd, 1995). These place-oriented strategies are based on the findings that crimes do not occur evenly across urban landscapes. Instead, they are concentrated in relatively small places that generate large proportions of crime occurrences (Sherman et al., 1989; Weisburd et al., 1992). For instance, Sherman et al. (1989) found that only 3 percent of the addresses of the crime calls to the police in Minneapolis accounted for 50 percent of all the crime calls. In addition, similar results were shown in different locations, suggesting a very high concentration of crime in specific places (e.g., see Pierce et al., 1988; Weisburd and Green, 1994; Weisburd et al., 1992). These high concentrations of crime represent that many crime problems could be reduced more efficiently if police officers focused their attention on these deviant places.
Sherman, 1995b; Weisburd, 1997). Also, increasing evidence suggests that focused police interventions, such as directed patrol, proactive arrests, and problem-solving can produce significant crime prevention gains at crime hotspots (Sherman, 1997).

The use of hotspots to determine policing and crime prevention strategies has grown over recent years. Hotspots allow local governors to determine the areas of greatest need (Ratcliffe, 2004). In policing, there has been a large shift to the use of hotspots as the foundation for problem-oriented policing (Goldstein, 1990). More recently, the UK moved toward intelligence-led policing that utilizes hotspot information (e.g., see Maguire, 2000; Ratcliffe, 2002), and the US has been following this trend of growth in both intelligence and crime analysis practice (e.g., see Andrews and Peterson, 1990; Carter, 1990). For instance, the city of Columbus is practicing problem-oriented policing by focusing on hotspots of crime. Combined with the use of mapping visualization, crime hotspots enable both crime prevention practitioners and police to focus resources on the areas of most need and have a process for explaining their objective decision-making to others (Ratcliffe, 2004).

2.2.1 Clustering/cluster modeling: From space to space-time surveillance

Given the significance of crime hotspot analyses for crime control, this sub-chapter reviews clustering/cluster measures for crime hotspot analyses. Clustering/cluster models
are broadly divided into focused and general modeling by the use of pre-specified assumption of patterns. In addition, depending on their incorporation of either space only or space-time dimensions, analyses can be divided into space and space-time modeling. This sub-chapter revisits widely applied clustering/cluster measures by their features.

2.2.1.1 Spatial Modeling

Studies of the clustering of spatial phenomena have their focus on finding significantly high (or low) density of a phenomenon in a given time or space. Given that many human and physical processes on space are not homogenous, finding clusters has naturally become an interest of many studies, and thus have contributed spatial analysis in many disciplines. A common question in spatial clustering/cluster analyses is whether the observed patterns in a study area has risen arisen by chance or not. If clustering is observed, then another question can be derived: where is the cluster? Regarding these questions, the terms *cluster* and *clustering* should be defined and distinguished. *Cluster* indicates a collection of cases inconsistent with our null hypothesis, and *clustering* implies the overall propensity of cases to cluster together (Besag and Newell, 1991; Waller and Gotway, 2004). Therefore, Waller and Gotway (2004) explains that tests to detect *clusters* determines which collection of cases represents the most significant cluster, and a test of *clustering* provides a single assessment of the statistical significance of the pattern for the entire area.
Clustering/cluster tests can be further distinguished between *general* and *focused* tests (Besag and Newell, 1991). Where are *general tests* test for cluster/clustering anywhere in the study area, *focused tests* test for cluster/clustering around pre-specified factors that have been previously hypothesized to be associated with the phenomenon of interest.

Based on above distinctions, *general clustering tests* include Moran’s I (1948) and its variants, which provide a single value of the overall spatial patterns of a study area. *General cluster tests* include Geographical Analysis Machine (Openshaw and Charlton, 1987), the cluster evaluation permutation statistic (Turnbull et al., 1990), and the spatial scan statistic (Kulldorff and Nagarwalla, 1995), which provide specific size and location of clusters. These statistics are characterized by drawing a number of circular windows in a study area. Local spatial statistics, such as Local Indicators of Spatial Autocorrelation (LISA) (Anselin, 1995) and Local Statistics Model (LSM) (Getis and Aldstadt, 2004), would also belong to *general cluster test* category. The local spatial statistics quantify spatial autocorrelation or clustering by providing similarity in a given locality (e.g. local Moran’s I (Anselin, 1995)) or clusters of high or low values (e.g. local G statistics (Getis and Ord, 1992; Ord and Getis, 1995)). For the purpose of detecting spatial clusters, local G statistics would be more appropriate compared to local Moran’s I because, as pointed out by Zhang and Lin (2006), local Moran’s I distinguish between low value clusters and high values clusters only through Moran scatter plots. It should be noted that the
capability of local spatial statistic measures can be extended by applying appropriate spatial weight matrix (Aldstadt and Getis, 2006; Getis and Aldstadt, 2004).  

*Focused clustering tests* measure the effects of pre-specified factors on clustering. As an example of *focused clustering tests*, score test (Lawson, 1993; Waller et al., 1992) uses levels of exposure. The levels of exposures are approximated by inverse distances (Waller et al., 1992) or other variables (Lawson, 1993). Stone test (Stone, 1988), as another example, is characterized by not requiring specification of parametric exposure-occurrence relationship (Waller and Gotway, 2004 p. 265). Instead, it test if the risk stays same or decreases.  

**2.2.1.2 Space-time modeling**

This sub-chapter reviews the several existing space-time clustering measures in both retrospective and prospective analyses. As retrospective measures, Knox test, Mantel’s test, and Jacquez’s k-Nearest Neighbor (k-NN) test are presented. For prospective measures, a classical (temporal) prospective measure such as Shewhart charts and selected space-time prospective measures are reviewed, including Cumulative Sum (CUSUM) version of nearest neighbor tests and local Knox tests. It should be noted that other important space-time prospective measures, such as space-time scan statistics and CUSUM version of Tango’s statistics, are presented in Chapter 5 with applications.
2.2.1.2.1 Retrospective Analyses

Knox test (Knox, 1964), is the most widely used technique for testing space-time interaction and quantifies space-time interaction for individual-level data using pair of cases that are separated by less than the critical space and time distances. This implies that Knox test requires pre-specified space and time distances identifying critical distances. Thus, pairs of cases within critical distances in space and time are considered as space-time clusters. Figure 2.3 shows a scatterplot of pairs of cases specified by their distances in space and time. The vertical and horizontal lines in the plot show the critical space and time distances respectively, deciding clusters. Thus, the bottom left part of the plot shows the clustered pairs of cases in space-time.
The Knox test is defined as follow:

\[ X = \sum_{i=1}^{N} \sum_{j=1}^{i-1} s_{ij} t_{ij} \]

where \( s_{ij} \) is an index for spatial adjacency, 1 if distance between events \( i \) and \( j \) is less than critical space distance, and 0 otherwise. \( t_{ij} \) is index for temporal adjacency, 1 if distance between events \( i \) and \( j \) is less than critical time distance, and 0 otherwise. In the null hypothesis, the \( X \) is assumed to be independent and Poisson distributed with a
mean of $ST/n(n-1)$; $ST$ are the cases close in space and time. Thus, if $X$ is unusually large in comparison with theoretical distribution then independence is rejected in favor of attraction.

With its wide applications, the Knox test has many attractive features. Kulldorff and Hjalmars (1999) finds its advantages from its simplicity and sufficient information from cases with no need for controls. Also, many extensions have been proposed. Diggle et al. (1995) suggested how K-functions may be used as a diagnostic for possible space-time dependence. Baker (1996) introduced a way to choose critical values in space and time. Noted by Mantel (1967) that Knox test becomes biased under population shifts over time, Kulldorff and Hjalmars proposed unbiased version of the Knox tests to adjust this. Rogerson (2001) provided a local version of the Knox test in the application to surveillance methods. O’Kelly (1994), from the idea of Knox test, introduced optimal partitioning methods in generating coordinates of cluster centroids which optimize partitions based upon least space-time square distances.

Mantel’s test (Mantel, 1967) is mainly characterized with no requirement of specifying critical values for space and time. The test finds correlations among space and time distances by calculating the sum across all pairs of the time distance multiplied by the spatial distance.
\[ Z= \sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij} t_{ij} \]

Also Mantel’s test is standardized,

\[ r = \frac{1}{N^2 - N - 1} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{(s_{ij} - \bar{s})(t_{ij} - \bar{t})}{s_i s_j} \]

Where \( \bar{s} \) and \( \bar{t} \) are means of space and time distances, and \( s_i \) and \( s_j \) are corresponding standard deviations. \( r \) is a measure of correlation between space and time distances raged from –1 to 1. The significance of Mantel’s test under the null hypothesis, independence between space and time distance, is calculated by using Monte Carlo simulation method.

Though Mantel’s test does not require pre-specified critical space and time distance, Mantel’s test shares many similarities with the Knox test. Mantel’s test is a distance-based measure and does not require controls. Also, Mantel’s test, like original Knox test, can be biased by shifts in population size over time.

k-Nearest Neighbor (k-NN) test (Jacquez, 1996) determines whether pairs of cases that are the nearest neighbors in space tend also to be nearest neighbors in time. The method can adjust space-time scales at which interaction (clustering) occurs. In using individual data and space-time distances, the k-NN is similar to Knox test and Mantel’s test.

However, specification of the scale differentiates the k-NN test from the others, enabling
a focused test. The test statistic, $J_k$, is the number of case pairs that are $k$ nearest neighbors in both space and time. When space-time clustering exists, $J_k$ will be large because the nearest neighbors in space will also tend to be the nearest neighbors in time.

$$J_k = \sum_{i=1}^{N} \sum_{j=1}^{N} s_{ijk} t_{ijk}$$

Where $k$ is the pre-specified number of nearest neighbors, $s_{ijk}$ is the spatial nearest neighbor index ($s_{ijk} = 1$ for within nearest neighbor and $s_{ijk} = 0$ otherwise), and $t_{ijk}$ is the time nearest neighbor index ($t_{ijk} = 1$ for within nearest neighbor and $t_{ijk} = 0$ otherwise). The significant test is done by Monte Carlo simulation as applied in Mantel’s test. Given that k-NN test focuses on clustering in specified range of space-time distance ($k$ nearest neighbor), k-NN test can be considered as a local (proportional) version of Mantel’s test.

In spite of different statistical structures among Knox test, Mantel test, and k-NN test, they have similarities in that they measure and compare space-time distances among cases based on a temporally fixed whole dataset. As a result, while they can detect space-time clustering in past occurrences, they lack a capability to find the clusters in regularly updated datasets. In particular, Knox test finds space-time clusters based on pre-specified critical space and time distances. The critical distances are not adjustable to test the significance of newly added occurrences. In case of Mantel’s test, the correlation is decided by multiplied sum of all space and time distances. While, k-NN test adjust space-
time scales for clustering, k-NN test is similar to Mantel’s test in that it uses space-time covariance of temporally fixed dataset.

2.2.1.2.2 Prospective analyses

Prospective surveillance detects clusters by combining prior data with continuously updated information. Until recently, prospective analyses have mainly focused on a temporal dimension in surveillance (Strat, 2005). Within an area of homogeneous spatial patterns, temporal prospective tests find excessive cases in each period from the assumption of independent and normal distribution. However, because of their neglect of the spatial dimension, prospective surveillance has been given less attention, compared to space-time retrospective analysis (Farrington and Beale, 1998). In this part, the classical (temporal) prospective method such as Shewhart charts (Shewhart, 1931) and selected space-time prospective surveillance measures are reviewed, including CUSUM version of nearest neighbor tests (Rogerson and Sun, 2001) and local Knox tests (Rogerson, 2001). It should be noted that other space-time prospective surveillance measures are compared in Chapter 5 with applications, including Kulldorff’s scan statistics, CUSUM version of Tango’s statistics.
Shewhart charts (Shewhart, 1931), as a simplest form of control chart, triggers alarm when the value exceeds predetermined control limits. This indicates that the process level has shifted from its previous level and out of control. The control limits are usually decided by multiples of the standard deviation of the study variable.

Moving average charts (Stern and Lightfoot, 1999) extends Shewhart chart by taking group of observations into account rather than the last observation.

\[ y_t = \frac{1}{m} \sum_{k=0}^{m-1} x_{t-k} \]

where \( m \) is the number of past observations in account used in the moving average. Since \( m \) decides the effect and frequency of alarms, proper choice of \( m \) determines the suitability of the chart.

The exponentially weighted moving average (EWMA) chart (Hunter, 1986) assigns less weight to data as they become older observations and assigns more weight to recent observations:

\[ y_t = \lambda x_t + (1-\lambda)y_{t-1} \]

\[ = \lambda \sum_{j=1}^{t} x_j (1-\lambda)^{t-j} \]

where \( 0 < \lambda \leq 1 \) is the EWMA weighting parameter and \( y_0 = 0 \). Since each observation is cumulatively weighted by \( (1-\lambda) \), the equation is decomposed into
\[ y_t = \lambda x_t + (1-\lambda)[\lambda x_{t+1} + (1-\lambda)[\lambda x_{t+2} + (1-\lambda)(\ldots)]] \]

forming exponentially weighted moving average. As \( \lambda \) increases, the previous data have less influence, and recent data have more influence. A value for \( \lambda \) is usually chosen between 0.1 and 0.5.

The nearest neighbor statistic (Clark and Evans, 1954) finds clusters among points by comparing the observed mean of the distances between neighboring points and the expected distances between nearest neighbors in a random pattern. The expected distance is given by

\[ r_e = \frac{1}{2\sqrt{\rho}} \]

where \( \rho = N/A \) is the density of points with the number of points (N) and the size of area (A). The nearest neighbor statistic, \( R \), is the ratio between the observed and the expected distance.

\[ R = \frac{r_{\text{obs}}}{r_e} \]

When clustering in the point pattern exists, the observed mean distance becomes less than the expected distance making \( R \) value less than 1. The minimum value of \( R \) is zero, which occurs when all points are at a single location. Given that the standard deviation of the mean distance between nearest neighbors in a random pattern is

\[ \sigma = \frac{0.26}{\sqrt{N\rho}}, \text{ normal} \]
approximation of the nearest neighbor statistic, $Z_i$ can be generated, and it is ready to
apply a cumulative sum statistic.

The local Knox test (Rogerson, 2001) identifies specified pairs of observations when the
original Knox test (Knox, 1964) proves significant clustering in space and time. While
Knox tests are categorized as a retrospective analysis, the local Knox test can be
considered as a prospective analysis with its application to a cumulative sum method.
Consequently, local Knox test, under cumulative sum framework, enables a prospective
surveillance to compare any increase in the Knox statistic to what would be expected. A
significant local Knox statistic for an observation implies that the observation has more
links that are both close in space and time than expected for the observation. Thus, the
local Knox test provides specific spatial and temporal information of the influential
observation which contributes global clustering in space and time.

Let $n_s(i)$ be the number of observations that are close to observation $i$ in space, $n_t(i)$ the
number that are close in time, and $n_{st}(i)$ the number that are close together in both space
and time. When the distribution of $N_{st}(i)$ under the null hypothesis that space and time
distance between observations are equally likely, the null distribution of $N_{st}(i)$ is given
as a weighted sum of hypergeometric distribution with parameters $n-1$, $n_s(i)$, $n_t(i)$:
\[ p\{N_{st}(i) = n_{st}(i)\} = \frac{1}{n} \sum_{j=1}^{n} \binom{n_{j}(i)}{n_{st}} \binom{n-1-n_{j}(i)}{n_{j}(i) - n_{st}} \]

where \( n_{j}(i) \) is the number of points that are close in time to point \( i \) when point \( i \) is assigned the \( j \)th value of time. And the expectation and variance are

\[
E[N_{st}(i)] = \frac{2n_{st}n_{j}(i)}{n(n-1)},
\]

\[
V[N_{st}(i)] = \frac{2(n-1)n_{st} \sum_{j=1}^{n} n_{j}(i)^{2}n_{j}(i)\{n-1-n_{j}(i)\}}{n(n-1)^{2}(n-2)}
\]

With expectation and variance, the number of observations that is close both in space and time, \( n_{st}(i) \), can be approximately normalized. However, since \( N_{st}(i) \) is a discrete variable, a correction factor is applied to make the quantities on the left-hand side and right-hand side approximately equal.

\[
z_{i} = \frac{n_{st}(i) - E\{N_{st}(i)\} - 0.5}{\sqrt{V\{N_{st}(i)\}}}
\]

For the purpose of testing changes in continuous patterns by combining prior data, the listed prospective tests are differentiated from retrospective analyses in generating expected values from prior observations, which also change over time. In particular, Shewart charts, moving average charts, and exponentially weighted moving average
charts focus on changes in observations over time. Though these tests are similar in their
main structure which compares previous values to continuously updated observations,
they have minor differences in generating expected values such as average and
exponentially weighted average. The nearest neighbor statistic and the local Knox test in
Cumulative sum frame works, as spatial prospective tests, are distinguished by measuring
spatial patterns and their changes over time. While basic frame works, which comparing
prior values to the expected, are identical to other prospective tests, the spatial
prospective tests do not measures the intensity of observation, but measure the spatial
patterns of observations.
CHAPTER 3

STUDY AREA AND DATA DESCRIPTION

3.1 Columbus and State of Ohio

The city of Columbus, as the third largest and state capital of Ohio, has 711,470 populations for the year 2000. Columbus is characterized with its rapid population growth. From 1980, Columbus has grown about 10% in population size, which is more than twice to the state average in the same period. Currently, Columbus is one of the fastest growing large cities in the nation in terms of population. The economic index shows both diversity and stability. The economy of Columbus has relatively higher rates in tertiary and quaternary economic sectors than that of the state. 2000 Census information shows that about 30% of total 692,339 employees are hired in service industry, which is about 3% higher than that of the state average. At the same time, the city’s 2.4% unemployment rate is lower than the state (4.0%) and the national rates (4.1%).

Other demographic index of Columbus shows minor differences from the state and the nation averages. In ethnic composition, Columbus has higher racial diversity compared
with the State and the nation. Columbus has 75.5% white, 17.9% black, and the others composed of Asian, Pacific Islander, and others. Compared with the State of 85% (white) and 11.5% (black) or the nation of 75% (white) and 12% (black), Columbus has more racial minorities indicating higher ethnic heterogeneity as a middle sized urban areas in the U.S. In case of sex, sex is similar to that of national average with 49% male and 51 female. Age and marital status show a relatively larger size of single population between 25 to 35 age group than both the state and the nation average have. Combined with racial composition, age, sex and marital status of Columbus present the characteristics of a middle size city in the US. These urban features are also represented in the housing types with relatively higher rate of renter occupied housing units (40.2%) ; the state (30%) and the nation (34%).

One of geographical features of Columbus is that several incorporated cities in Franklin County are located inside the city boundary as shown in Figure 3.1. These incorporated cities have their own municipalities and handle local businesses. However, it should be noted that for many incorporated cities some administrative issues and public services are shared with Columbus such as Police service.
Figure 3.1 Columbus, Franklin County, and State of Ohio
3.2 Crime Data

The original crime data were collected and reported by the police department. Organized by Uniform Crime Reporting (UCR) codes of 119 different types of crime, the data contain specific address information, census tract numbers (based upon 1990 census tract configuration), and exact dates of occurrences. According to the crime classifications of UCR program, the residential crimes are composed of robberies and burglaries targeted at residential units. Under UCR definitions, robbery is defined as “The taking or attempting to take anything of value from the care, custody or control of a person or persons by force or threat of force or violence and/or putting the victim in fear”, and burglary is defined as “The unlawful entry (forceful or not) of a structure to commit a felony or theft” (Federal Bureau of Investigation, 2000). The corresponding UCR codes of residential crimes are in Table 3.1.

<table>
<thead>
<tr>
<th>Crime Types</th>
<th>UCR Codes</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robbery</td>
<td>3150</td>
<td>Robbery using Gun Targeted at Residence</td>
</tr>
<tr>
<td></td>
<td>3250</td>
<td>Robbery using Knife Targeted at Residence</td>
</tr>
<tr>
<td></td>
<td>3350</td>
<td>Robbery using Others Targeted at Residence</td>
</tr>
<tr>
<td></td>
<td>3450</td>
<td>Robbery using Hands Targeted at Residence</td>
</tr>
<tr>
<td>Burglary</td>
<td>5111</td>
<td>Forcible Entry to Residence at Day time</td>
</tr>
<tr>
<td></td>
<td>5112</td>
<td>Forcible Entry to Residence at Night</td>
</tr>
</tbody>
</table>

Table 3.1 Residential crimes and their UCR codes
Residential crimes have a merit for the spatial crime data analyses because of its reliable crime location information. Compared with other crime types such street robbery or highway theft, residential crimes provide more accurate address information of crime location from the victims’ addresses. For the spatial referencing of the addresses, each crime cases were geocoded based on street network map of the Franklin County. About 83% of the data were successfully geocoded.

Figure 3.2 and Figure 3.3 show study areas applied for both hierarchical Bayesian analysis and prospective surveillance analyses. For the two analyses, two different subsets of the residential crime data are applied. The data applied to hierarchical Bayesian analyses (Figure 3.1.) were organized by the coverage area of 19 Columbus police precincts. Since the data is created and managed by Columbus police department, the use of the police precincts is justified. As a result, the study area is not completely overlaid by the city boundary missing several incorporated areas such as Upper Arlington, Bexley, Whitehall, Worthington, and Upper Arlington. These incorporated areas maintain their own police departments, and crime cases occurred in the incorporated areas are managed by their police departments.

In addition to the police precincts, Census tract areas were considered, defining study area. Since Census tract boundaries do not match the police precincts, Census tract areas completely included in the precincts are selected for the study area. Figure 3.2 present corresponding study areas for hierarchical Bayesian analysis.
For prospective surveillance analyses (Figure 3.3), central city areas are selected to minimize edge effects in the analyses. In the surveillance analyses, circular windows are used to identify crime hotspots. Given that the city of Columbus and police precinct boundary is irregular, crime hotspots in nearby the city edge areas become vulnerable to edge effects. In addition, other municipalities intensify the edge effects further because they are presented as empty holes in the study area. From this reason, the selected central city area in Figure 3.3 presents contiguous areas with relatively regular boundary to reduce potential edge effects.
Figure 3.2 Study area for Bayesian hierarchical analyses

Figure 3.3 Study area for prospective surveillance
CHAPTER 4

BAYESIAN HIERARCHICAL DATA ANALYSIS OF RESIDENTIAL CRIME OCCURRENCES

This chapter analyzes urban residential crimes using a Bayesian hierarchical data analysis model to handle the spatial effects of crime data in the analyses. In particular, this chapter accounts for spatial autocorrelation and spatial heterogeneity in the analyses in explaining neighborhood socio-demographic effects on residential crimes. Regarding Bayesian hierarchical model, this chapter applies random effect components to account for spatial effects in the model (Besag et al., 1991; Law and Haining, 2004). The random effects capture the spatial effects in residuals. In particular, spatially structured random effects capture spatial autocorrelation, and unstructured random effects capture spatial heterogeneity. Among many neighborhood socio-demographic factors, this study focuses on the three characteristic features of urban areas: economic deprivation, residential stability, and the number of available targets. The use of the three covariates is justified by crime theories such as Social disorganization theory (Bursik, 1988) and Opportunity theory (Felson and Clarke, 1998).
4.1 Socio-demographic factors relevant to residential crimes

In analyzing residential crimes, 3 principal components were generated from 10 neighborhood socio-demographic variables. The 10 initial variables were selected as they have been used in many crime studies (see, Land et al., 1990). The data were extracted from Summary File 1 and Summary File 3 of the 2000 US Census. The 10 variables have strong collinearity in spite of their unique effects in estimating residential crimes.

Principal component analysis handles the collinearity problem by generating independent and coherent subsets of variables. The resulting principal components reflect underlying processes and account for the correlations among the variables (Tabachnick and Fidell, 2006). Table 4.1 shows a list of the components applied in this study and the corresponding relevant variables.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
<th>Corresponding Variable</th>
<th>Relevant Sociological Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rcrime_R (Dep. var)</td>
<td>Residential crime Rate</td>
<td>(No. of residential crimes / No. of households)</td>
<td></td>
</tr>
<tr>
<td>Comp. 1</td>
<td>Poverty</td>
<td>Income, Families in poverty, and unemployment</td>
<td>Social disorganization</td>
</tr>
<tr>
<td>Comp. 2</td>
<td>Residential Stability</td>
<td>Rate of household living in the same house more than 5 years</td>
<td>Social disorganization</td>
</tr>
<tr>
<td>Comp. 3</td>
<td>No. of Targets</td>
<td>No. of households</td>
<td>Opportunity theory</td>
</tr>
</tbody>
</table>

Table 4.1 Variable lists
The independent variables, represented by each principal component, reflect social theories that explain the relationship between crimes and urban neighborhoods. The use of poverty and population instability variables is justified by social disorganization theory (Bursik, 1988; Shaw and McKay, 1942). This theory proposes that poverty (Component 1) is a factor that attracts and motivates people to engage in criminal behaviors. Additionally, this theory indicates that population instability (Component 2) increases with population heterogeneity, weakening social networks in a neighborhood. Considering that strong social networks function as a social guardian in preventing delinquent behaviors and controlling crime occurrences (Coleman, 1988; Coleman, 1990), weak social networks caused by population instability would increase crime occurrences in a neighborhood.

Research indicates a positive relationship between crime rates and criminal opportunities (Herbert, 1982; Krivo and Peterson, 1996; Wikstrom, 1991). In particular, an urban environment characterized by a high level of anonymity among residents, high population density, and large population size explains higher crime rates in urban areas than rural areas. Felson and Clarke (1998) note that crime opportunity is a major factor in crime occurrences, regardless of crime types. Although crime opportunities vary depending on target qualities, space, time, and the offenders’ preferences, the number of available targets is one of the most important factors in crime occurrence. Therefore, this study applies the number of household (Component 3) as a proxy for residential crime opportunities.
4.2 Bayesian hierarchical modeling

4.2.1 General Framework

Unlike the conventional statistical inference which derives the average estimates of parameters, Bayesian hierarchical modeling produces parameter estimates for each individual analysis unit by borrowing information from all of the analysis units (Carlin and Louis, 2000; Spiegelhalter et al., 2002). In the analysis process, the model is called hierarchical because it uses multiple levels of analyses in an iterative way. Whereas a simple Bayesian model is defined using a prior probability $p(\theta)$ and likelihood $p(y|\theta)$ to compute a posterior distribution $p(\theta|y)$,

$$p(\theta|y) = \frac{p(\theta,y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta) \quad (4.1)$$

Bayesian hierarchical model requires another prior probability $p(\phi)$ to specify the other prior probability $p(\theta)$. Consequently, the prior distribution in the simple Bayesian model $p(\theta)$ is replaced by joint prior distribution $p(\theta|\phi)p(\phi) = p(\theta,\phi)$, generating the joint posterior distribution $p(\phi,\theta|y)$ as follows:
\[ p(\phi, \theta \mid y) \propto p(y \mid \theta, \phi) p(\theta \mid \phi) p(\phi) \\
= p(y \mid \theta) p(\theta \mid \phi) p(\phi) \\
= p(y \mid \theta) p(\theta, \phi) \] (4.2)

Regarding the equation (4.2), the key hierarchical part of the model is that \( \phi \) is not known thus has its own prior distribution \( p(\phi) \), which is usually specified as a vague prior with large variance (Gelman et al., 2003 p. 124). In addition, the condition
\[ p(y \mid \theta, \phi) = p(y \mid \theta) \]
is holding because the data distribution \( p(y \mid \theta, \phi) \) depends only on \( \theta \).

The application of Bayesian hierarchical modeling is largely attributable to the development of Markov Chain Monte Carlo (MCMC) methods and software such as WinBUGS (Carlin and Louis, 2000). MCMC is a general method drawing parameter values \( \theta \) from prior distribution \( p(\theta) \) and correcting these draws to approximate the target posterior distribution \( p(\theta \mid y) \) (Gelman et al., 2003). Thus, MCMC makes the posterior distribution convergent, fitting Bayesian models that would otherwise be computationally intractable.

### 4.2.2 Model specifications with random effect components

Given the dependent variable (residential crime rates) are proportional, the number of crime occurrences can be modeled as a conditional \( \text{Bionomial}(n, p) \) distribution, using the
number of households $n$ and an estimated crime rate $p$. In addition, a logit transformation enables the estimation of rate values to approximate a normal distribution. In the simplest Bayesian logistic regression form, the model is defined by Model 1:

$$
y_i \sim \text{Binomial}(n_i, p_i)$$

$$\logit(p_i) = \ln(p_i/(1-p_i))$$ (4.3)

Model 1:

$$\logit(p_i) = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i}, \quad i = 1,\ldots,n$$ (4.4)

where $p_i$ is the crime rate for census tract areas $i$. $X_1 \text{ to } X_3$ are the independent variables, generated by the principal component analysis from 10 socio-economic variables. Model 1 would have similar results in a standard logit regression model because Model 1 does not account for spatial dependence and heterogeneity in the dependent variable. Consequently, Model 1 may have inefficient standard errors of coefficients under standard logit regression framework, and the resulting inference tests can be misleading. In Bayesian modeling, the logit regression (Model 1) can be extended by incorporating random effects. Model 2, 3, and 4 present extended Bayesian models with random effects accounting for heterogeneity and spatial dependence as follows:

Model 2:

$$\logit(p_i) = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + U_i, \quad i = 1,\ldots,n$$ (4.5)

Model 3:

$$\logit(p_i) = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + S_i, \quad i = 1,\ldots,n$$ (4.6)
Model 4:

\[
\text{logit}(p_i) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + U_i + S_i, \quad i = 1, \ldots, n
\] (4.7)

where \(U_i\) is a random effect component to capture heterogeneity, and \(S_i\) is a spatially structured random effect component to capture spatial dependence. In the equations, Model 2 only accounts for heterogeneity in residential crime rates, Model 3 for spatial dependence, and Model 4 for both heterogeneity and spatial dependence.

Random effects \(U_i\) is structured by an independent normal random variable with mean zero and variance \(\sigma_u^2\). The prior distribution for \(\sigma_u^2\) is set by gamma \(\sigma_u^2 \sim \text{Gamma}(\alpha, \beta)\).

Because of no prior information for the precision of the random effects, a vague prior using Gamma(0.001, 0.001) was applied. The use of the vague prior with large variance are commonly applied in many hierarchical Bayesian analyses (e.g. Law and Haining, 2004; Zhu et al., 2006). For sensitivity tests, other vague priors were applied such as Gamma(0.0001, 0.0001) and Gamma(0.01, 0.01), but no significant differences were found in results.

For the spatial component \(S_i\), intrinsic conditional autoregressive (ICAR) model (Besag et al., 1991) was applied. Many spatial models can be applied such as simultaneous autoregressive (SAR) model (LeSage, 1997; LeSage, 2000), spatial lag (LAG) model (Cliff and Ord, 1981), or conditional autoregressive (CAR) model (Besag and Newell,
1991; Haining, 1990). However, ICAR model has relative advantages compared to other models because ICAR model can be structured by normal random variables and is supported by WinBUGS/GeoBUGS for an implementation (Spiegelhalter et al., 2003; Thomas et al., 2004). The distribution of ICAR model for $S_i$ is specified as follow for the vector of a normal random variable $S = (S_1, ..., S_n)$ when $S_i$ and $S_j$ s are neighboring ($S_i|S_j$):

$$S_i|S_j = \text{Normal}(\bar{S}_i, \frac{\sigma^2_i}{m_i})$$

$$\bar{S}_i = \frac{\sum_{j=1}^{N_i} W_{ij} S_j}{\sum_{j=1}^{N_i} W_{ij}}, \quad m_i = \sum_{j=1}^{N_i} W_{ij}$$

(4.8)

where $W_{ij}$ is a contiguity matrix with $W_{ij} = 1$ when $i$ and $j$ are neighbors and $W_{ij} = 0$ otherwise. Also, diagonal values, $W_{ii}$, are set to 0. $\sigma^2_i$ is a variance parameter that controls the variance of $S_i$. Given $\bar{S}_i$, $S_i$ has a normal distribution with a conditional mean formed by the average of neighboring $S_j$ s. Also $m_i$, the number of neighbors of $S_i$, indicates that the conditional variance of $S_i$ is inversely proportional to the number of neighbors. A vague prior is applied to the prior distribution for $\sigma^2_i$ using Gamma(0.001, 0.001).
4.3 Results

4.3.1 Model diagnostic and convergence

Deviance Information Criterion (DIC), devised by Spiegelhalter et al. (2002), was used to compare the four models. DIC is a generalization of the Akaike Information Criterion (Akaike, 1973) based on the posterior distribution of the deviance statistic (Zhu et al., 2006). The deviance, defined as \(-2\log(\text{likelihood})\), is given for Binomial likelihood by:

\[
D(\theta) = -2 \sum_i \{y_i \log(p_i) + (n_i - y_i) \log(1 - p_i)\} \tag{4.9}
\]

The DIC, which accommodates both a model fit indicator ($\bar{D}$) and the number of effective parameters ($p_D$), is defined as follow:

\[
\text{DIC} = \bar{D} + p_D = 2\bar{D} - D(\bar{\theta}) \tag{4.10}
\]

where $\bar{D}$ ($=E(D(\theta))$) is the posterior mean of the deviance $D(\theta)$, and $D(\bar{\theta})$ ($=D(E(\theta))$) is a deviance evaluated by posterior means of relevant parameter values $\theta$. $p_D$ is evaluated as:

\[
p_D = \bar{D} - D(\bar{\theta}) \tag{4.11}
\]
Table 2 shows the deviance summaries for the four hierarchical models applied in the study. Given that a smaller DIC value indicates better fit, the models with random effects (Model 2, 3, and 4 with unstructured $U$, spatially structured $S$, or both) perform far better than the fixed effect model (Model 1). In particular, Model 2 shows a better fit with a smaller DIC compared to Model 3. It indicates that heterogeneity is a more dominant process than spatial dependence in crime patterns in the study area. Overall, given the smallest DIC value, Model 4 can be regarded as the best model in estimating the dependent variable. However, the number of effective parameters $p_D$ in Model 4 shows that Model 4 is not as parsimonious as Model 2 with a bigger number of effective parameters.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\bar{D}$</th>
<th>$D(\bar{\theta})$</th>
<th>$P_D$</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) $Xb$ (Fixed effects only)</td>
<td>4033.590</td>
<td>4029.620</td>
<td>3.972</td>
<td>4037.560</td>
</tr>
<tr>
<td>(2) $Xb + U$ (Fixed + heterogeneity)</td>
<td>1211.190</td>
<td>1039.010</td>
<td>172.188</td>
<td>1383.380</td>
</tr>
<tr>
<td>(3) $Xb + S$ (Fixed + dependence)</td>
<td>1232.580</td>
<td>1051.150</td>
<td>181.433</td>
<td>1414.010</td>
</tr>
<tr>
<td>(4) $Xb + U + S$ (Full model)</td>
<td>1182.720</td>
<td>986.658</td>
<td>196.057</td>
<td>1378.770</td>
</tr>
</tbody>
</table>

Table 4.2 Deviance summaries for the four hierarchical models

Parameter convergence was mainly assessed by a visual inspection of the trace plots for each parameter. Trace plots confirmed convergence of the parameters in the models by providing posterior distribution of parameters with a constant variance and a stable mean. Complex models with random effects, such as Model 2, 3, and 4, required more iterations
for convergence. The most complex model (Model 4) was converged by 20,000 iterations. Therefore, the sample for the first 20,000 iterations was discarded as a burn-in period, and each chain was executed for 20,000 more iterations with acceptable Monte Carlo errors, which is less than about 5% of the sample posterior standard deviation.

4.3.2 Regression results

Table 3 reports coefficient values for the four Bayesian hierarchical regression models from the parameters after the burn-in period. The table presents posterior means, standard deviations, Monte Carlo errors, 2.5% percentiles, medians, and 97.5% percentiles (those two percentile values show 95% posterior credible interval).

The four models in this study provide notable consistent coefficient values. In spite of minor deviations in coefficient values among models, signs of the mean coefficients are constant. However, minor deviations in coefficients are found in the residential stability variable (Component 2) and the crime opportunity variable (Component 3). Compared to model 1 and 4, Model 2 has a slightly increased coefficient value for the residential stability variable (component 2), along with Model 3 for the crime opportunity variable (component 3). Given that Model 4 provides the best model fit, our interpretation of regression results are based on Model 4.
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>S.D.</th>
<th>MC error</th>
<th>2.5%</th>
<th>Median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>intercept</td>
<td>-3.395</td>
<td>0.01208</td>
<td>0.00011</td>
<td>-3.418</td>
<td>-3.395</td>
<td>-3.371</td>
</tr>
<tr>
<td></td>
<td>Comp. 1</td>
<td>0.516</td>
<td>0.00962</td>
<td>0.00008</td>
<td>0.497</td>
<td>0.516</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>Comp. 2</td>
<td>-0.062</td>
<td>0.00944</td>
<td>0.00006</td>
<td>-0.080</td>
<td>-0.062</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>Comp. 3</td>
<td>-0.119</td>
<td>0.01293</td>
<td>0.00009</td>
<td>-0.144</td>
<td>-0.119</td>
<td>-0.093</td>
</tr>
<tr>
<td>Model 2</td>
<td>intercept</td>
<td>-3.635</td>
<td>0.05736</td>
<td>0.00263</td>
<td>-3.748</td>
<td>-3.635</td>
<td>-3.526</td>
</tr>
<tr>
<td></td>
<td>Comp. 1</td>
<td>0.626</td>
<td>0.05752</td>
<td>0.00301</td>
<td>0.514</td>
<td>0.627</td>
<td>0.742</td>
</tr>
<tr>
<td></td>
<td>Comp. 2</td>
<td>-0.006</td>
<td>0.05581</td>
<td>0.00278</td>
<td>-0.117</td>
<td>-0.006</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>Comp. 3</td>
<td>-0.156</td>
<td>0.07333</td>
<td>0.00373</td>
<td>-0.301</td>
<td>-0.156</td>
<td>-0.014</td>
</tr>
<tr>
<td>Model 3</td>
<td>intercept</td>
<td>-3.598</td>
<td>0.06678</td>
<td>0.00349</td>
<td>-3.732</td>
<td>-3.598</td>
<td>-3.469</td>
</tr>
<tr>
<td></td>
<td>Comp. 1</td>
<td>0.348</td>
<td>0.06671</td>
<td>0.00395</td>
<td>0.215</td>
<td>0.349</td>
<td>0.473</td>
</tr>
<tr>
<td></td>
<td>Comp. 2</td>
<td>-0.123</td>
<td>0.0731</td>
<td>0.00466</td>
<td>-0.269</td>
<td>-0.121</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>Comp. 3</td>
<td>-0.027</td>
<td>0.06882</td>
<td>0.00341</td>
<td>-0.162</td>
<td>-0.028</td>
<td>0.106</td>
</tr>
<tr>
<td>Model 4</td>
<td>intercept</td>
<td>-3.612</td>
<td>0.145</td>
<td>0.01097</td>
<td>-3.858</td>
<td>-3.632</td>
<td>-3.266</td>
</tr>
<tr>
<td></td>
<td>Comp. 1</td>
<td>0.506</td>
<td>0.08223</td>
<td>0.00528</td>
<td>0.353</td>
<td>0.503</td>
<td>0.665</td>
</tr>
<tr>
<td></td>
<td>Comp. 2</td>
<td>-0.048</td>
<td>0.06947</td>
<td>0.00413</td>
<td>-0.184</td>
<td>-0.047</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>Comp. 3</td>
<td>-0.101</td>
<td>0.09211</td>
<td>0.00578</td>
<td>-0.273</td>
<td>-0.104</td>
<td>0.086</td>
</tr>
</tbody>
</table>

Table 4.3 Summaries of Bayesian hierarchical regression models with posterior parameters and precision coefficients

The economic deprivation variable (component 1) demonstrates a positive impact (0.506) on the residential crimes rates. As shown in the literature about other urban crime studies (e.g., Land et al., 1990), economic deprivation functions as a strong motivator for residential crime occurrences in Columbus. Considering the spatial patterns of crime rates and economic status, wealthy areas have lower crime rates despite their attractiveness as targets. Given that impoverished individuals become potential crime offenders (Bursik, 1988), this result confirms the routine activity theory (Brantingham and Brantingham, 1984 p. 362) that offenders travel short distances to their crime targets located nearby.
offenders’ neighborhoods. The short travel distances of the potential offenders result in positive spatial dependence in crime occurrences because adjacent areas of the offender’s residence become crime targets, leading to clusters of high crime rates.

The residential stability variable (component 2) shows negative influence (-0.048) on the residential crime rates. Residential stability indicates a level of social networks and a cultural fragmentation in a neighborhood (Bursik, 1988; Coleman, 1988; Coleman, 1990). In spite of its negative coefficient sign, the influence to the residential crime is not as strong as other variables, with a value very close to zero. Results imply that residential stability does not have strong influence on crime occurrences.

The number of households (component 3) displays a negative relationship (-0.101) with the residential crime rates. This result is quite opposite to other literature because the results indicate that neighborhoods with a large number of households have less crime rates. In urban crime analyses, it is a common understanding that large population sizes increase crime rates by providing more opportunities such as more targets, high population density, and high levels of anonymity among the population (Herbert, 1982; Krivo and Peterson, 1996; Wikstrom, 1991).

Regarding this opposite result to literature, two perspectives should be considered. First, a scale of the analysis matters. Describing urban crimes, literature commonly indicates a
positive relationship between crime rates and population sizes. Compared with rural areas, urban areas, which characterized by large population size, tend to have higher crime rates. However, this assumption does not necessarily fit within city areas. Different neighborhoods within a city can have various crime rates regardless of their population size. This would be related to MAUP issues, seeing inconstant relationship among variables in different scales. For instance, a study on large areas with several urban and rural areas, the analysis results would present a relationship (positive in literature) between population size and crime rate. However, in a study of small areas, such as one small sized city or one rural community does not necessarily present the same relationship. In addition, several literature present inverse or nonsignificant effects of population size on crime rates in city level analyses (Bailey, 1984; Blau and Blau, 1982; Blau and Golden, 1986; Jackson, 1984).

Second, crime rate is occasionally confused with the number of crime counts. Crime rates are calculated by dividing crime counts with underlying population size. Consequently, when a neighborhood with large population size has a lower crime rate than a neighborhood with small population size, it does not necessarily mean the neighborhood with large population size has less crime occurrences than the neighborhood with small population. To account for this issue, a Bayesian hierarchical model with the number of crime occurrences as a dependent variable is applied using the same set of independent variables to Model 4. The Table 4.4 presents the result.
<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean</th>
<th>S.D.</th>
<th>MC error</th>
<th>2.5%</th>
<th>Median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>intercept</td>
<td>50.34</td>
<td>3.912</td>
<td>0.2963</td>
<td>42.27</td>
<td>51.01</td>
<td>55.86</td>
</tr>
<tr>
<td>(Crime</td>
<td>Comp. 1</td>
<td>17.57</td>
<td>3.168</td>
<td>0.2189</td>
<td>12.77</td>
<td>16.56</td>
<td>24.7</td>
</tr>
<tr>
<td>counts)</td>
<td>Comp. 2</td>
<td>-8.146</td>
<td>1.928</td>
<td>0.08515</td>
<td>-12.2</td>
<td>-8.439</td>
<td>-4.152</td>
</tr>
<tr>
<td></td>
<td><strong>Comp. 3</strong></td>
<td><strong>17.87</strong></td>
<td>2.706</td>
<td>0.1592</td>
<td>11.75</td>
<td>17.83</td>
<td>22.56</td>
</tr>
</tbody>
</table>

Table 4.4 Bayesian hierarchical model using the number of crime occurrences as a dependent variable

The result shows that the number of crime occurrences is positively related to the number of household with a coefficient value 17.87. Also, the signs of the other variables are same to the crime rate model. Therefore, this result indicates that residential crime increases as the number of household increases. However, the rate of crime increase (as indicated in Model 4 with crime rate) is smaller than the rate of population increase in the study area.

4.3.3 Map decomposition

One of the main advantages of Bayesian hierarchical analyses is that a dependent variable can be decomposed into covariates \( X_i \beta s \), unstructured effects \( U_i s \), and spatially structured effects \( S_i s \). This decomposition enables us to visualize identifiable high or
low risk areas and the extends of risk areas, driven by covariates, unstructured, or spatially structured random effects (Haining, 1990; Law and Haining, 2004). Figure 4.1 shows the result of map decomposition for Model 4 using posterior means of $X_i\beta$ s for covariates, $U_i$ s for unstructured effect, and $S_i$ s for spatially structured effects. The odds of residential crime rates for Model 4 are decomposed as follows:

$$\frac{p_i}{1-p_i} = \exp(-3.612 + 0.506X_{1,i} - 0.048X_{2,i} - 0.101X_{3,i} + U_i + S_i)$$

$$= \exp(-3.612 + 0.506X_{1,i} - 0.048X_{2,i} - 0.101X_{3,i}) \exp(U_i) \exp(S_i)$$

where $\exp(X)$ denotes exponential function of $X$. Figure 4.1 clearly reveals that the covariates (fixed effects) does not fully account for the residential crime rates. Instead, the unstructured effects (heterogeneity) and the spatially structured effects (spatial dependence) effectively fill in unaccounted parts of the dependent variable in the estimation. Neither the spatial dependence nor the heterogeneity is dominant over the other in the map pattern. However, the map pattern of spatially structured effects shows local clusters of similar values, and the map pattern of unstructured effects shows heterogeneity. Considering that the areas of high values in both the spatial and the heterogeneity effects imply larger unaccounted proportions in crime rates by covariates, the areas may require further investigation for a better analysis of the residential crime rates.

Finally, Figure 4.2 presents the maps of estimated crime counts and posterior means of residuals for Model 1 and Model 4. The residuals for both the models enable us to assess
the overall goodness-of-fit of the models and compare the model fits. The residuals are generated as follow:

\[
\hat{Y}_i = \hat{p}_i h_i
\]

\[
e_i = Y_i - \hat{Y}_i
\]

\[
\hat{p}_i = \frac{\exp(-3.395 + 0.516X_{1,i} - 0.062X_{2,i} - 0.119X_{3,i})}{1 + \exp(-3.395 + 0.516X_{1,i} - 0.062X_{2,i} - 0.119X_{3,i})}, \quad \text{(for Model 1)}
\]

\[
(4.13)
\]

\[
\hat{p}_i = \frac{\exp(-3.612 + 0.506X_{1,i} - 0.048X_{2,i} - 0.101X_{3,i} + U_i + S_i)}{1 + \exp(-3.612 + 0.506X_{1,i} - 0.048X_{2,i} - 0.101X_{3,i} + U_i + S_i)}, \quad \text{(for Model 4)}
\]

\[
(4.14)
\]

where \( h_i \) is the number of household, \( \hat{p}_i \) is the estimated crime rates, and \( \hat{Y}_i \) is the estimated crime occurrences for the tract \( i \).

The map patterns in Figure 4.2 confirm that Model 4 makes far better estimations with smaller residual values than Model 1. The map patterns of estimated crime counts are almost same to the map patterns of observed crime counts in Model 4. Also, it should be noted that the residual patterns in Model 4 show no similarities between nearby local values, implying no spatial dependence in the residuals. However, Model 1 indicates positive spatial dependence with similarities in local clusters.
Figure 4.1 Map decomposition of the Odds into the Covariates, Spatially structured, and Unstructured random effects
Figure 4.2 Map decompositions for the comparison between Model 1 and Model 4
4.4 Summary

In this study, a Bayesian hierarchical model was applied to analyze residential crime rate in Columbus, Ohio for the year 2000. Results indicate that three covariates (poverty, residential stability, and the number of households) were important factors accounting for residential crime rates. Among the covariates, poverty was the most significant factor accounting for residential crimes. In addition, the number of households showed a negative effect to crime rates, implying that crime opportunities were not necessarily related to residential crimes in Columbus. Based on the results, it was interpreted that the negative effects of the household sizes was attributable to intensified police surveillance in recent years and corresponding decrease in violent crimes in U.S. However, a further investigation is necessary in providing a more complete explanation.

The four models applied in this study demonstrated that unstructured random effects and spatially structured random effects adjusted the coefficient values of covariates by capturing unaccounted components in the dependent variable. Significance of the random effects (both unstructured and spatially structured) was confirmed by Deviance Information Criterion (DIC). However, the DIC results indicated that heterogeneity was more dominant than spatial dependence in the analysis.
The map decomposition in this study enabled us to visualize contribution of regression components in estimating residential crime rates. Areas with high values in both the unstructured component and the spatially structured component indicated that the crime rates are mainly derived by spatial and heterogeneity effects rather than covariates.
CHAPTER 5

PROSPECTIVE SPACE-TIME SURVEILLANCE ANALYSES OF RESIDENTIAL CRIMES

This chapter presents empirical results and comparisons of several prospective surveillance analyses methods. With minor methodological differences, the applied surveillance methods have the same goal to provide exact spatial and temporal information of hotspots in a near-real time base. Regarding the purpose of prospective surveillance methods, this chapter focuses on the limitations of existing measures, and presents a new surveillance method, which handles the limitations. In the process, main algorithms of the surveillance methods are presented. Their capabilities and inabilities are critically evaluated together with their empirical results.

5.1 Space-time Scan Statistic

5.1.1 Model description

The space-time scan statistic (Kulldorff, 2001) applied in this chapter is an extension from a spatial scan statistic (Kulldorff, 1997; Kulldorff and Nagarwalla, 1995), originally
designed for hotspot identification in temporally fixed spatial patterns. The space-time scan statistic enables prospective surveillance analyses by extending its scan windows from a circular shape (in spatial scan, Kulldorff, 1997; Kulldorff and Nagarwalla, 1995) to a cylindrical shape (in space-time scan, Kulldorff, 2001). In the process, the original assumption about the spatial pattern of occurrences were extended from Complete Spatial Randomness (CSR, in spatial scan) to Complete Spatio-temporal Randomness (CSTR, in space-time scan). Even though a family of analysis methods has been developed for different models of underlying clusters, such as Bernoulli model and Poisson model, Poisson model are used for this analysis. Because crime events are assumed to occur by an inhomogeneous Poisson process (the number of occurrences in a region over a time interval is expected to be proportional to the corresponding population), the use of Poisson model is reasonable.

The basic model is drawing gradually expanding circular/cylindrical windows on every case location over the study area. The cylindrical windows in space-time scan statistics present space with their bases and time with their heights. Figure 5.1 and Figure 5.2 present conceptual structures of the both scan statistics. Under the Poisson assumption, the null hypothesis assumes that the number of occurrences inside circular/cylindrical windows is same to outside areas. The observed number of counts is compared to the expected number of occurrences, which is generated by the population size at risk. Consequently, if the observed number of counts is significantly larger than the expected, the corresponding circular/cylindrical window areas become hotspots. For this model, the
likelihood ratio statistic is as follows for both spatial and space-time scan statistics (Kulldorff, 1997; Kulldorff, 2001),

\[
\sup_{z \in \mathcal{Z}} \left( \frac{n(Z)}{\mu(Z)} \right)^{n_G} \left( \frac{n(G) - n(Z)}{\mu(G) - \mu(Z)} \right)^{n_G - n_z} I\left( \frac{n(Z)}{\mu(Z)} \geq \frac{n(G) - n(Z)}{\mu(G) - \mu(Z)} \right)
\]

(5.1)

where \( n(G) \) is the total number of observed cases, and \( \mu(G) \) is the total number of expected cases within a study area \( G \). Since \( \mu(G) \) is the expected number under the null hypothesis, \( \mu(G) \) is equal to \( n(G) \). \( I(G) \) is an indicator function. The likelihood ratio value is calculated for every possible window, and the window associated with the significantly higher than the expected value becomes a cluster.

For a significance test, Monte Carlo simulation is applied, where cases are randomly simulated by population size at risk. For example, the number of cases in the study area is selected from the population, and labeled as cases. For each replicate, the number of counts in the window is counted, and p-value is obtained by the rank of the observed likelihood ratio value compared the likelihood ratio values of the replicas. When a Monte Carlo study consisted of 999 replicas, total 1000 likelihood ratio values exist, combining 999 replicas and the original data itself. If the observed likelihood ratio value for the data is among the 50 highest of the 1000 values, then this means the window area is significant at the 5 percent level.
As a prospective surveillance method, the space-time scan statistic overcomes the limitations of spatial scan statistics (see, Kulldorff, 2001). In temporal perspectives, the space-time scan statistic can detect emerging clusters by accounting for temporal dimensions of hotspots. In addition, the space-time scan statistic adjusts multiple testing problems, which are caused by repeated analyses, by finding one maximized likelihood among cylinders.

In spite of the advantages as a prospective surveillance measure, the space-time scan statistic has problems at the same time. First, as shown in Figure 5.1 and Figure 5.2, a point represents a population size and the number of occurrences in an administrative Census unit. Naturally, significant cluster areas, which are generated by the points, contain spatial approximation. This spatial approximation makes the analyses results less precise to provide ranges of hotspots. Regarding the issue, Kulldorff and Nagawalla (1995) noted, “Individuals actually outside the defining circle (cylinder), but lying within cells whose centroids lie inside the circle (cylinder), are included in the zone. Similarly, individuals actually inside the circle (cylinder), but lying within cells whose centroids are outside the circle (cylinder), are excluded.” Second, in generating a maximum possible cluster size, using a gradually expanding window may detect a larger cluster than the true cluster by absorbing surrounding regions where there is no elevated risk (Tango and Takahashi, 2005). Particularly, given that not all clusters are necessarily circular shaped, gradually expanding circular/cylindrical windows may not correctly detect noncircular
clusters. Third, dependence on population size for the generation of expected cases is unrealistic for frequent and continuous prospective surveillance applications. Because population information is only updated in decennial bases, weekly, monthly, or yearly based surveillance should depend on a fixed population assumption, which is far from reality.
Figure 5.1 Structure of spatial scan statistic
Figure 5.2 Structure of space-time scan statistic
5.1.2 Numerical Results

Combined with crime data, population data from Census block group for year 2000 are applied. Because space-time scan statistic generates expected cases from population size at risk, the use of population data is crucial in the analysis. However, it should be noted that population information applied in the study are fixed throughout the year. As a result, the expected cases are constant in each block group for the entire year. Census block group is the smallest area unit available with population information. As a result, the use of the block group units minimizes area approximation in the resulting crime hotspots. In the application, centroids are generated for each block group. Each centroid presents its population size in the corresponding block group and crime cases. It should be noted that block groups are not perfectly matched to original study area. Therefore, block groups completely within the study area are selected and applied. Figure 5.3 represents crime cases, Census block group, and corresponding centroids in the study area.
Figure 5.3 Crime cases and Centroids organized by Census block group
Analyzing the space-time crime patterns in the city, nine significant hotspot areas were identified. Hotspots were decided by 5% significance level, selecting p-values less than 0.05. Table 5.1, Table 5.2, and Figure 5.4 shows the detailed information and location of the hotspots. Figure 5.4 shows that large groups of hotspot areas observed in central city (cluster 1) and central north area (cluster 2), each containing 86 and 41 block groups, and other small size hotspots. The rank of hotspots in Table 5.1 are ordered by their relative sizes because their p-values are all identical after Monte Carlo simulation. In the largest hotspots (Cluster 1 and 2), the problem of detecting a larger cluster than the true cluster was observed, which was noted in the previous section as a potential problem of applying extending windows in scan statistics. In particular, some block group areas in Cluster 1 and Cluster 2 do not have elevated risks. Instead, surrounding high-risk areas absorbed the low risk areas, generating large hotspots. Table 5.2 shows the lists of low risk block groups in Cluster 1 and Cluster 2. In spite of high relative risk values in Table 5.1, the block groups in Table 5.2 have less than expected crime occurrences as shown in the relative risk values. Overall, 9 out of 86 block groups (10.5%) in Cluster 1 and 11 out of 41 block groups (26.8%) in Cluster 2 are mis-represented as crime hotspots.

In temporal aspects, the End-Date column in Table 5.1 should be noted. The column has all the same values as Dec. 31st. It means that the once formed space-time hotspots last until the end of the study period. This result is caused by the unrealistic assumption that the number of crime cases decided by the number of the population sizes at risk and the corresponding generation of expected values. Since the number of population size is
fixed over time, the expected number of crimes is constant correspondingly. In addition, the CSTR assumption of the space-time scan measures requires temporally independence among crime occurrences. These assumptions do not fit into the observed crime patterns in this study. Observed temporal crime patterns in the study shows intensive occurrences in the summer period and gradual decline afterward. As a results, space-time hotspots in the Table 5.1 show that space-time clusters are mostly formed in summer periods (particularly in July) and last until December. As a result, winter periods, such as November and December, which have even less cases than the average are included in the space-time hotspots. The result implies the problems of absorbing low-risk periods after high-risk periods by applying temporally expanding cylinders. in temporal dimension. This is the same problems observed in spatial dimensions caused by applying spatially expanding windows.
<table>
<thead>
<tr>
<th>Cluster Rank</th>
<th>Start_Date</th>
<th>End_Date</th>
<th>p-value</th>
<th>No. of Area Units</th>
<th>Radius</th>
<th>Observed</th>
<th>Expected</th>
<th>Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2000/7/4</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>86</td>
<td>3.77</td>
<td>4778</td>
<td>2606.16</td>
<td>1.83</td>
</tr>
<tr>
<td>2</td>
<td>2000/7/2</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>41</td>
<td>1.82</td>
<td>1761</td>
<td>1286.72</td>
<td>1.37</td>
</tr>
<tr>
<td>3</td>
<td>2000/7/5</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>1</td>
<td>0.00</td>
<td>148</td>
<td>47.22</td>
<td>3.13</td>
</tr>
<tr>
<td>4</td>
<td>2000/7/4</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>1</td>
<td>0.00</td>
<td>91</td>
<td>22.35</td>
<td>4.07</td>
</tr>
<tr>
<td>5</td>
<td>2000/7/7</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>1</td>
<td>0.00</td>
<td>138</td>
<td>48.01</td>
<td>2.87</td>
</tr>
<tr>
<td>6</td>
<td>2000/7/4</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>1</td>
<td>0.00</td>
<td>109</td>
<td>39.00</td>
<td>2.79</td>
</tr>
<tr>
<td>7</td>
<td>2000/7/3</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>1</td>
<td>0.00</td>
<td>65</td>
<td>18.44</td>
<td>3.52</td>
</tr>
<tr>
<td>8</td>
<td>2000/7/4</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>1</td>
<td>0.00</td>
<td>54</td>
<td>14.44</td>
<td>3.74</td>
</tr>
<tr>
<td>9</td>
<td>2000/7/2</td>
<td>2000/12/31</td>
<td>0.00100</td>
<td>1</td>
<td>0.00</td>
<td>38</td>
<td>8.38</td>
<td>4.53</td>
</tr>
</tbody>
</table>

Table 5.1 Results of space-time scan statistic ordered by cluster rank
Figure 5.4 Spatial pattern of crime hotspot distribution by space-time scan statistic
<table>
<thead>
<tr>
<th>Block Group Codes</th>
<th>Observed</th>
<th>Expected</th>
<th>Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>390490032001</td>
<td>57</td>
<td>69.67</td>
<td>0.818</td>
</tr>
<tr>
<td>390490020002</td>
<td>27</td>
<td>36.56</td>
<td>0.739</td>
</tr>
<tr>
<td>390490020004</td>
<td>24</td>
<td>29.68</td>
<td>0.809</td>
</tr>
<tr>
<td>390490020003</td>
<td>12</td>
<td>18.87</td>
<td>0.636</td>
</tr>
<tr>
<td>390490056101</td>
<td>14</td>
<td>19.98</td>
<td>0.701</td>
</tr>
<tr>
<td>390490018103</td>
<td>9</td>
<td>24.33</td>
<td>0.370</td>
</tr>
<tr>
<td>390490018105</td>
<td>17</td>
<td>28.11</td>
<td>0.605</td>
</tr>
<tr>
<td>390490083301</td>
<td>35</td>
<td>46.41</td>
<td>0.754</td>
</tr>
<tr>
<td>390490029001</td>
<td>5</td>
<td>26.93</td>
<td>0.186</td>
</tr>
<tr>
<td>Cluster 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>390490010001</td>
<td>16</td>
<td>20.82</td>
<td>0.769</td>
</tr>
<tr>
<td>390490007202</td>
<td>31</td>
<td>32.98</td>
<td>0.940</td>
</tr>
<tr>
<td>390490007102</td>
<td>25</td>
<td>28.27</td>
<td>0.884</td>
</tr>
<tr>
<td>390490006002</td>
<td>19</td>
<td>21.63</td>
<td>0.879</td>
</tr>
<tr>
<td>390490006001</td>
<td>9</td>
<td>23.83</td>
<td>0.378</td>
</tr>
<tr>
<td>390490007103</td>
<td>20</td>
<td>21.94</td>
<td>0.912</td>
</tr>
<tr>
<td>390490010007</td>
<td>25</td>
<td>37.54</td>
<td>0.666</td>
</tr>
<tr>
<td>390490010005</td>
<td>23</td>
<td>26.30</td>
<td>0.875</td>
</tr>
<tr>
<td>390490006004</td>
<td>10</td>
<td>23.17</td>
<td>0.432</td>
</tr>
<tr>
<td>390490007301</td>
<td>36</td>
<td>47.85</td>
<td>0.752</td>
</tr>
<tr>
<td>390490007302</td>
<td>19</td>
<td>19.66</td>
<td>0.967</td>
</tr>
</tbody>
</table>

Table 5.2 List of low risk block groups included in the hotspots
5.2  An Application of Cumulative Sum version of Tango's Statistic

5.2.1  Model description

Rogerson (1997) devised a prospective surveillance analysis measure by adapting Tango’s statistic (Tango, 1995) to Cumulative Sum (CUSUM) statistics. Using Tango’s statistic as a measure of spatial pattern, Rogerson’s test applies CUSUM statistic with expected values conditional upon Tango’s statistic from the previous observations. Thus, the test assumes that a current spatial pattern of occurrences is not different from previous spatial patterns, meaning fixed spatial patterns of a phenomenon over time. CUSUM statistics originally devised for quality control in manufacturing processes can be extended to prospective surveillance tests by combining spatial pattern measures (Rogerson, 1997). Although this chapter applies only CUSUM version of Tango’s statistics, other spatial pattern measures can be applied including nearest neighbor statistic (Rogerson and Sun, 2001) and Knox tests (Rogerson, 2001).

Introduced by Page (1954), CUSUM statistic detects sudden changes in the mean value of a quantity of interest. CUSUM statistic assumes a independent normal distribution as a null hypothesis, following $x_t \sim N(\mu_t, \delta_t^2)$. The statistic is defined iteratively as:
\[ y_t = \max \left( 0, y_{t-1} + \left( \frac{x_t - \mu_t}{\sigma_t} - k \right) \right) \]

\[ y_0 = 0 \]

(5.2)

where \( k \) is a parameter, usually set equal to 0.5. The CUSUM statistic cumulates deviations from the mean that exceed \( k \) standard deviations. When \( y_t \) exceeds a critical value \( h \), it means there’s a significant change in a process. The choice of \( h \) determines the Average Run Length (ARL), which indicates average number of observations until a change is occurred under the null hypothesis. Thus, high \( h \) values result in long ARL resulting in the stochastically less number of false alarms. By adjustments, CUSUM statistic can be used for data following Poison or Exponent distributions. Mostly, the adjustments transform original data distribution into normal.

Tango statistics applied to Rogerson’s prospective surveillance tests assumes Poisson random process of no clustering. Therefore, the expected number of cases is generated proportionally to population size, \( \xi_i \)

\[ H_0: E[N_i] = \lambda \xi_i, \quad i = 1, 2, \ldots, m \]

(5.3)

where \( \lambda \) is a disease rate across entire populations. Based on the expected cases, Tango’s statistic, \( C_G \), for a general test of clustering is:

\[ C_G = (r - p)^T A (r - p) \]

(5.4)
where \( \mathbf{r} \) and \( \mathbf{p} \) are \( n \) by 1 vectors containing the observed and expected proportions of cases in \( n \) regions respectively. \( \mathbf{A} \) is an \( n \) by \( n \) matrix containing elements \( a_{ij} \) that measures a closeness from region \( i \) to region \( j \). \( a_{ij} \) is defined as a function of the distance \( d_{ij} \) between two regions:

\[
a_{ij} = \exp(-d_{ij}/\tau), \quad i \neq j
\]

For the diagonal values of \( \mathbf{A} \), unit value is applied as \( a_{ii} = 1 \). A scale parameter \( \tau \) decides the size of cluster. Regarding the scale parameter \( \tau \), Tango argues that results are not change drastically by the choice of \( \tau \) (1995).

Rogerson’s surveillance test generates \( Z_i \), which is deviations of Tango’s statistic from its conditional expectation, and applies to CUSUM statistic. Theoretically, \( Z_i \) value implies changes in spatial pattern under approximated normal distribution. Therefore, \( Z_i \) is a normal-transformed Tango’s statistic value. However, Rogerson (1997) indicated that \( Z_i \) may not follow normal distribution in the application of small sample sizes for surveillance purpose. The equation for \( Z_i \) generation is as follows:

\[
Z_i = \frac{C_{Gi} - E\left[ C_{Gi}|C_{Gi-i} \right]}{\sigma_{C_{Gi}|C_{Gi-i}}} \quad (5.6)
\]
where $E[C_{G,i}|C_{G,i-1}] = \mathbf{p}^T \mathbf{u}$ and $\sigma^2_{C_{G,i}|C_{G,i}} = \mathbf{p}^T \left( \text{diag} \mathbf{u} \mathbf{u}^T \right) - \left( \mathbf{p}^T \mathbf{u} \right)^2$. The $\mathbf{u}$ is an $m \times 1$ vector containing $k$ as an element $u_k = \left( \mathbf{r}_{i-1}(k) - \mathbf{p} \right)^T \mathbf{A} \left( \mathbf{r}_{i-1}(k) - \mathbf{p} \right)$. The $\mathbf{r}_{i-1}(k)$ denotes the proportion of cases in each regions, given $\mathbf{r}_{i-1}$ and given that case $i$ is located in region $k$.

The application of $Z_i$ to CUSUM statistic is straightforward. Instead of $\frac{X_i - H_i}{\delta_i}$ component in CUSUM equation, $Z_i$ value is plugged in. Since large positive $Z_i$ value implies substantial deviation in Tango’s statistic from its conditional expectation, CUSUM statistic signals alarms whenever $Z_i$ exceeds a critical limit.

5.2.2 Numerical results

Figure 5.5 shows the result of Rogerson’s CUSUM version of Tango’s statistic (Rogerson, 1997). The result is generated using critical value ($h = 4.3$), threshold ($k = 0.5$), risk weight ($\tau = 20$), and batch size ($n = 1$). The parameters applied use the same values in Rogerson’s study (1997) except the batch size. Unit batch size is applied because each temporal unit has plenty of cases. For the generation of expected cases, fixed population sizes are applied, assuming no population changes over time. As widely acknowledged, the spatial CUSUM statistic is a general clustering test, where the statistic

89
is based on an assessment of changes over the entire area (Rogerson, 1997). Therefore, even though Tango’s statistic accounts for spatial patterns of occurrences in CUSUM process, the result only provides single values indicating significant period of time in surveillance without location information of the hotspots. Even though several other CUSUM based surveillance statistics were developed (Rogerson, 2005; Rogerson and Sun, 2001; Rogerson and Yamada, 2004), most of them are general clustering tests, which lack location information of clusters. In addition, it should be noted that expected cases are generated by underlying population, causing a problem for frequent and continuous analyses because of population data availability.

In Figure 5.5, an emergence of significant crime cluster is observed for the period from September 21\textsuperscript{st} to 24\textsuperscript{th}. The significant increase of crime occurrences is interpreted by a large number of incoming migrations of the OSU students to Columbus at the start of autumn quarter. In addition, given that the test assumes a fixed population size over time, the increased student population should make considerable effects on the test results. In particular, significant crime clusters are supposed to be caused by the difference between assumed population size and observed population size in September. While CUSUM version of Tango’s statistic accounts for spatial patterns and temporal trends in crime occurrences, the results cannot provide the exact interpretation because of the lack of location information. Therefore, more reliable interpretation would be significantly increase spatial autocorrelation patterns compared to previous periods occurred in the study area from September 21\textsuperscript{st} to 24\textsuperscript{th}. 

90
Figure 5.5 Patterns of crime hotspot distribution by CUSUM version of Tango's statistic
5.3 An Application of a New Space-time Surveillance method

5.3.1 Model description

5.3.1.1 Conceptual description

Many cluster surveillance analysis methods (Kulldorff, 2001; Kulldorff and Nagarwalla, 1995; Openshaw, 1994; Openshaw and Charlton, 1987; Turnbull et al., 1990), use numerous overlapping circular windows to find the extent of high intensity areas. A window with a significantly larger number of counts than expected becomes a cluster. Testing significance, many spatial cluster analysis methods use expected counts, which are generated from the underlying population at risk. In other words, the methods assume that the population size decides the number of occurrences under a null assumption of constant risk for all individuals. Therefore, an area with an unusually large number of occurrences compared to at risk population becomes a cluster.

The new space-time surveillance methods devised in this study, however, hypothesizes that areas keep the same occurrence level from a previous period of time. Accordingly, this method detects areas of significantly increased occurrences, compared to previous occurrences in the same area. This means that this new space-time surveillance method
assumes temporal changes in the occurrences rather than fixed. In application perspective, 
the method requires neither population at risk information nor corresponding expected 
values. Instead, the method derives expected values from previous observations. Figure 
5.6 and 5.7 provide conceptual demonstration about generating expected values, using 
previous occurrences.
Figure 5.6 Generation of observed count for Date 4

Figure 5.7 Generation of expected count for Date 4 using average of previous counts
Both Figure 5.6 and 5.7 are composed of point locations of occurrences over a four-day period. Applying the surveillance analysis on Date 4, I draw circular windows of a given radius centered on all the Date 4 occurrence locations. For each window, the expected values are the average counts of previous dates in the same circular window. For instance in the Figure 5.6, a window is presented in Date 4, which contains four occurrences. The expected value is generated as shown in Figure 5.7 by drawing the window at the same location on previous times and averaging the number of counts:

\[
\text{Expected value for the window in Date 4} = \frac{1 (\text{Date 1}) + 3 (\text{Date 2}) + 5 (\text{Date 3})}{3} = 3
\]

Observed counts for the window in Date 4 = 4

The surveillance model requires a spatial parameter, \( r \), and a temporal parameter, \( q \), for implementation. The spatial parameter, \( r \), decides the radius of circular windows, and the temporal parameter, \( q \), specifies previous time periods in generating expected values. In Figure 5.7, \( q \) is equal to 3 because 3 previous days’ observations are applied to generate expected counts. In deciding reasonable spatial \( r \) and temporal \( q \) parameters, users should have a level of understanding about the applied dataset. Applying several parameter combinations are recommended for an exploratory spatial data analysis.
Many cluster analysis methods find maximum spatial ranges of clusters, using sequentially increased spatial parameters for circular/cylindrical windows (e.g. Kulldorff, 1997; Openshaw and Charlton, 1987; Turnbull et al., 1990). Even though the self-increasing spatial parameters might be more convenient in analyses, there are many problems in finding maximum spatial ranges of clusters as discussed in chapter 5.1. As an alternative, a square pyramid shaped window (Iyengar, 2005) can be considered. The square pyramid shape would have an advantage for capturing diffusion process over time particularly in lattice data. However, many geographical clusters are neither circular nor square. Therefore, when identifying spatial ranges of clusters in a surveillance analysis, I recommend to use overlaid multiple windows of an appropriate size rather than one or two largest clusters. Regarding the temporal parameter, clusters generated by very large temporal parameters may be easily affected by seasonal trends when data have significant seasonal patterns. In reverse, applying very small parameter values cannot differentiate random fluctuations from the true clusters.

5.3.1.2 Likelihood ratio tests

The likelihood ratio test in this study assumes the Poisson distribution. The Poisson assumption has been applied to many point pattern analyses (Griffith and Haining, 2006). In this study, since the number of occurrences in a circular window is very small compared to the number of cases outside the window, the use of Poisson distribution assumption is reliable. Under the null hypothesis, we assume that the number of observed
cases within a circular window is same to the expected number of cases. The alternative hypothesis assumes the higher number of observed cases within a window than the expected:

\[ H_0: E(N(Z)) = \mu(Z), \text{ for all } Zs \]  
\[ H_1: E(N(Z)) > \mu(Z), \text{ for some } Zs \]

where \( N(Z) \) is the random number of cases within a circular window \( Z \), and \( \mu(Z) \) is the null expected number of cases for the window generated from previous observations.

Let \( Z_{\alpha, t} \) be a window centered on location of a case \( \alpha \) at a time \( t \). Then, \( n(Z_{\alpha, t}) \) is the number of cases within the window \( Z_{\alpha, t} \), including the case \( \alpha \) itself. Since \( \alpha \) is a location of observed case at time \( t \), we might not have a case at the same location \( \alpha \) in other times such as \( t-1 \) and \( t-2 \). However, by locating a circular window centered on location \( \alpha \) in other times, we can estimate the expected number of cases. The expected value of a window \( Z_{\alpha, t} \) is calculated as:

\[ \mu(Z_{\alpha, t}) = \frac{1}{q} \left( n(Z_{\alpha, t-1}) + \ldots + n(Z_{\alpha, t-q}) \right) \]  

(5.9)

The expected values are decided by temporally proximate values. High values in previous times result in high expected values, and low values in previous times generate low
expected values. Since the expected values are flexibly decided by cases in previous times, the likelihood ratio test does not require spatially and/or temporally independent distribution assumption, which is, in many times, violated by temporal and spatial trends in empirical datasets. For example, Kulldorff (2005) generates expected values form both spatial and temporal random distribution assumption. In addition, the sequentially updated expected values, as time progresses, enable one to detect clusters in sequential and ongoing processes by combining prior data with information as it is received. However, It should be noted that the analysis could miss slowly evolving clusters when the parameter $q$ is not large enough to clearly show the changing trend

Even though the equation (5.9) presents a simple average value for the observations of past $q$ periods, it can be further extended by allowing more weights on recent past values and less weights on further past values. For example, exponentially weighted moving average methods (Montgomery, 1996) would be applicable because the current value of a process is being compared with weighted average of past values.

Under the Poisson assumption, the likelihood ratio test, originally derived by Kulldorff and Nagarwalla (1995), is applied as a test statistic and given by:
where \( n(G) \) is the total number of observed cases, and \( \mu(G) \) is the total number of expected cases within a study area for past \( q \) periods. \( I(\cdot) \) is an indicator function. The likelihood ratio value is calculated for every possible window, and the window associated with the significantly higher than the expected value becomes a cluster. Even though the likelihood ratio equation (5.10) is same to that of Kulldorff and Nagarwalla (1995), its application is slightly modified. Rather than detecting most likely clusters by finding maximum likelihoods (as in Kulldorff and Nagarwalla (1995)), this study calculates the local likelihood ratio for each potential cluster. In addition, rather than considering all the past observations (as in Kulldorff and Nagarwalla (1995)), this study utilizes observations within past \( q \) periods. Since it is clear that the denominator of the likelihood ratio test statistic does not depend on the case configurations, the test statistic can be equivalently defined only by its numerator (for detail, see Kulldorff, 1997). The test statistic can be simplified to:

\[
\sup_{z \in Z} \left( \frac{n(Z)}{\mu(Z)} \right)^{a_2} \left( \frac{n(G)-n(Z)}{\mu(G)-\mu(Z)} \right)^{a_2} I \left( \frac{n(Z)}{\mu(Z)} > \frac{n(G)-n(Z)}{\mu(G)-\mu(Z)} \right) \]

(5.11)
The distribution of the test statistic for individual windows does not have a simple analytical form for significance tests; because the test statistic is not based on CSR process, and population at risk information is not applied in the analysis.

To overcome these problems, this study adopts the bootstrap permutation to generate an approximate distribution of the test statistic for each individual window. The bootstrap permutation is a widely applied random sampling method from an unknown probability distribution when estimating a parameter of interest on the basis of given observation (Davison and Hinkley, 1997; Efron and Tibshirani, 1993). The bootstrap permutation can simply be explained as resampling from original data to create replica of datasets with replacement (Westfall and Young, 1989).

Since this study uses local likelihood ratios for each individual observation, each likelihood value requires permutations for significance test. For each permutation, a new sample of \( m(G) \) observations is drawn with replacement from the original sample locations \( n(G) \) for past \( q \) time periods. In the permutation, each observation has same probability of being drawn. For the new sample, the expected value and corresponding likelihood is calculated. For example, testing significance for a window \( Z_{a,t} \), the window \( Z_{a,t} \) is located on the permutated new sample of \( m(G) \), an expected value is calculated as done in equation (5.9), and a likelihood is calculated as shown in equation (5.11). The
sample size of $m(G)$, is decided for each time period because the number of occurrences
is different for a different time. $m(G)$ for a window $Z_{a,t}$ is calculated as:

\[
m(G) = n(G_{\alpha,t}) + \ldots + n(G_{\alpha,q})
\]

(5.12)

where $n(G_{\alpha,t})$ is the total number of cases at time $t-1$. As a result, when a bootstrap
consists of 999 permutations, 1000 samples (including the actual data) are available for a
significance test. For each sample, the likelihood ratio (equation 5.10) is calculated for
the window $Z_{a,t}$. If the likelihood ratio for the actual data is higher than 50th highest
among the permutated likelihood ratio values for the window $Z_{a,t}$, then the window $Z_{a,t}$
becomes a statistically significant cluster at the 0.05 level. More generally, $p$-values for
hypothesis testing are defined as $p = r / (1+sim)$ where $r$ is the rank and $sim$ is the
number of bootstrap permutations. This process adjusts the $p$-values and reduces the false
significance rates caused by the multiplicity of testing. This method is used in scan
statistics (Kulldorff, 2001). However, it should be noted that the original sample for the
bootstrap permutations is continuously updated over time, making the original sample
less redundant. The bootstrap permutation in this study is computationally intensive but
appealing in that, unlike the other methods such as Bonferroni and Sidak adjustments,
correlations and distributional characteristics are incorporated into the adjustments
(Westfall and Young, 1989; Westfall and Young, 1993).
5.3.2 Numerical results

Depending on the data features and application purposes, the spatial surveillance methods in this study may be applied with various parameter settings. For the application, we set the spatial parameter $r$, which decides the radius of circular window, to 0.5 and 1.0 kilometers. The temporal parameter $q$, which specifies previous observations used for expected values, are set to 3 and 7. It should be noted that the original dataset is aggregated by 3-day intervals to reduce computational load and prepare enough occurrence cases. As a result, 366 days for the year 2000 are aggregated to 122 temporal units, and the temporal parameter value 3 ($q = 3$) presents 9 days, and 7 ($q = 7$) presents 21 days. For instance, when we have $r = 0.5$ and $q = 3$, the past 3 temporal units’ (9 days) observations within a circular window of 0.5km bandwidth are applied to generate expected values.

Bootstrap application consists of 999 replicates, each of which involves null hypothesis with no space-time cluster. We order the collection of 1000 values coming from the 999 replicates and the data itself with the highest value assigned 1. This means that the smallest $p$-value we could get is 0.001. In the result, circular windows with $p$-value 0.001 are considered as significant crime hotspots.
Since the circular windows are generated based on crime locations, some circular windows may cross the study area boundary causing edge effects. The edge effects are handled by excluding circular windows whose extent is beyond the study area boundary.

Figure 5.8 and Figure 5.9 show the surveillance results for the residential crime occurrences using different sets of spatial and temporal parameters. For two temporal parameters, $q = 3$ and $q = 7$, Figure 5.8 is generated using a spatial parameter $r = 0.5$, and Figure 5.9 is generated using $r = 1.0$. In the both figures, vertical axis indicates the proportion of hotspot area size out of total study area size, and horizontal axis provides corresponding temporal unit. As a result, higher Y values indicate larger hotspot areas in a given time.
Figure 5.8 Area proportion of significant crime hotspots; $r = 0.5$, and $q = 3$ and $7$

Figure 5.9 Area proportion of significant crime hotspots; $r = 1.0$, and $q = 3$ and $7$
Comparing the two figures, the high spatial parameter \((r = 1.0)\) in Figure 5.9 shows larger hotspot area size than the low spatial parameter \((r = 0.5)\) in Figure 5.8. The larger area size in Figure 5.9 is explained by larger circular windows from the high spatial parameter \((r = 1.0)\) because large circular windows tend to absorb surrounding regions with less crime occurrences.

Regarding temporal parameter \(q\), the high temporal parameter \((q = 7)\) results in larger hotspot areas in both Figure 5.8 and 5.9. Intrinsic features of crime data explain the relationship between the hotspot area sizes and temporal parameters. Usually, areas with a high crime risk tend to have high crime occurrences in near past time periods (Sherman, 1995a). For example, in analyzing the number of crime occurrences in a given area in a day, \(n\left(Z_{a,t}\right)\), the expected number of crime occurrences generated by the previous 2 days, 

\[
\frac{1}{2}\left(n\left(Z_{a,t-1}\right)+n\left(Z_{a,t-2}\right)\right),
\]

is usually more similar to \(n\left(Z_{a,t}\right)\) than the expected number generated by 5 days, \(\frac{1}{5}\left(n\left(Z_{a,t-1}\right)+...+n\left(Z_{a,t-5}\right)\right)\), or more. Consequently, the expected values generated by the low temporal parameters \((q =3)\), becomes similar (sometimes higher) to observed crime occurrences in this study, making the likelihood ratio less significant. However, the expected values from the high temporal parameters \((q =7)\) tend to generate likelihood ratio more significant because the longer interval from the high temporal parameter levels off the similar (or higher) number of crime occurrences in near
past crime periods. Overall, the high temporal parameters generate larger hotspot areas with significantly different expected values compared to the low temporal parameters.

Map visualization of the space-time surveillance results is presented in Figure 5.10 and 5.11. Map display of the surveillance result is crucial as it indicates the locations and ranges of hotspots. This study composed of 122 temporal units provides 122 spatial hotspot patterns for each combination of spatial and temporal parameters. However, it is not feasible to present all the map patterns in this paper. Therefore, Figure 5.10 and Figure 5.11 provide samples of the resulting map patterns for temporal unit 58 (June 17th to 19th) and 78 (Aug. 19th to 21st). Temporal unit 58 and 78 are contrasted by relatively large and small cluster area sizes respectively.
Figure 5.10 Locations and spatial patterns of crime hotspots for temporal unit 58 (June 17th to 19th)
Figure 5.11 Locations and spatial patterns of crime hotspots for temporal unit 78 (Aug. 19th to 21st)
Areas of significant circular windows indicate crime hotspots in both Figure 5.10 and Figure 5.11. With minor differences in their area sizes, the hotspot patterns for each temporal unit are similar in spite of different spatial and temporal parameter settings. Figure 5.10 and 5.11 show that hotspots are located in several different places, and their shapes are not circular in many places. In the perspective, finding maximum spatial ranges of a circular window using sequentially increasing spatial parameters (Kulldorff, 1997; Kulldorff et al., 2005; Turnbull et al., 1990) becomes questionable for both spatial hotspot analyses and space-time surveillance analyses. The largest hotspot does not necessarily provide reliable locations and ranges of hotspots for the purpose of surveillance, which requires instant and pinpointing responses such as medical treatments for infectious diseases. In case of crime surveillance, crime hotspot requires more frequent and intensive police patrols in the areas. Therefore, it is strongly recommended to select spatial parameter properly to match the hypothesized cluster size together with the purpose of study. There is no doubt that the power of finding a cluster is maximized when the window, which is circular in this study, is chosen to match the size and shape of the actual cluster (Rosenfeld and Kak, 1982).

5.4 Model comparisons and Summary

This chapter presented three prospective space-time surveillance tests and their results. Though they have the same goals for surveillance, which provide precise spatial and
temporal information of hotspots in near-real time base, the results have minor
differences among the tests because of their intrinsic features. Particularly, the results can
be compared in three issues as follows:

• Capability of continuous surveillance

• Provision of hotspots’ location information

• Precision of identified hotspots

Regarding the capability of continuous surveillance, while all the tests were applicable
for continuous surveillance with short time intervals, both space-time scan statistics and
CUSUM version of Tango’s statistics showed limitations from the unrealistic assumption
of Completer Spatio-temporal Randomness (for space-time scan statistics) and
independent normal distribution (for CUSUM version of Tango’s statistics). The
assumptions imply that crime occurrences in the study areas have minor deviations from
the expected numbers of cases, which are generated by population size. However,
because of the unavailability of frequently updates in population information, the results
cannot account for changes of underlying populations, which might have impact on crime
occurrences. For example, massive migrations into some city areas or huge residential
resettlements caused by suburban developments and urban gentrification cannot be
accounted for. Figure 5.5 reflects the limitation in CUSUM version of Tango’s statistics,
where significant crime clusters are accounted for by migrations of the OSU students at
the beginning of autumn quarter. Further, constant expected values for crime cases over
time make the test results insensitive to subtle changes in urban crime patterns. Crime patterns commonly have both seasonal and spatial trends. Therefore, certain periods in a year and certain areas in the city may have predominantly more crime occurrence than others. These spatio-temporal crime patterns cannot be considered in the analyses because of the fixed and constant expected values, commonly leading to larger hotspot areas over longer periods than true crime hotspots. Figure 5.4, Table 5.1, and Table 5.2 reflect the limitations with unusually long hotspot periods and large size. The large cluster, presented as a primary cluster in Figure 5.4, is not quite helpful to allow local councilors to determine the areas of greatest needs for crime control because the large area size over a long period obscures the exact ranges of true clusters.

The new surveillance method introduced in chapter 5.3. handled the continuous surveillance issue by generating expected values from previous occurrences. The use of expected values derived from previous dates’ occurrences naturally adjusts population changes in the study areas and does not require fixed expected assumptions over time, which are commonly unrealistic. Consequently, for the capability of the continuous surveillance issue, the new surveillance method is found to be the most reliable for prospective surveillance analyses.

Regarding the availability of hotspot location information, both the space-time scan statistic and the new surveillance measure provides location information with spatial
ranges of hotspots. Figure 5.4, Figure 5.10, and Figure 5.11 present visualization of
spatial patterns of the hotspots. Because the two tests analyze local patterns of crime
occurrence values, they can present individual significant hotspots and their locations.
However, CUSUM version of Tango’s statistics could not provide location information
because of its structural dependence on global measure of spatial patterns. Even though
several other CUSUM based surveillance statistics were developed (Rogerson, 2005;
Rogerson and Sun, 2001; Rogerson and Yamada, 2004), those are all general clustering
tests, which lacks location information of the clusters. Therefore, the model can only
present more significant periods in the analysis and set off alarms in spite of its
assessment of spatial patterns in the prospective surveillance process. By applying
individual tests for several independent areas, Rogerson (1997) presented significant
hotspot areas and temporal extents. However, the results cannot be compared between the
areas about information on which areas are more vulnerable because the areas were
calculated independently. In addition, applying the same measures commensurate with
the number of area units is inefficient, considering the purpose of surveillance.

Since the space-time scan statistic and the new surveillance measure present location
information, only these two models are comparable about the precision of identified
hotspots. As indicated in Chapter 5.1, the space-time scan statistic uses centroids to
present area units. It implies that a centroid included in a hotspot window is considered a
hotspot depending on whether its corresponding area unit is fully covered by the window
or not. Consequently, at the edge of hotspot areas, the centroid based hotspot detection
mechanism has spatial approximation. Some areas might be evaluated as hotspots even when proportional areas are covered by hotspot windows, and some others might be considered as hotspots even when the centroids are located outside the areas. These hotspot area approximations might be trivial when the area units in the study is small (i.e., Census block groups). With large area units such as Census tracts or counties, the result, however, may be critically imprecise. The new surveillance measure does not use centroids. Consequently, the measure does not approximate spatial ranges of hotspot areas, providing more precise analysis results.
CHAPTER 6

CONCLUSION

6.1 Results of the research

In spite of the various approaches applied, many crime studies have the same goal: to reduce and control crime occurrences in our living space. Some studies focus on ecological analyses, answering questions such as which human socio-demographic factors cause crime occurrences, while others investigate crime hotspots to enable efficient relocation of police resources. In this context, this study has two goals. First, by properly handling spatial effects, such as spatial autocorrelation and heterogeneity, this research finds significant socio-demographic factors of residential crimes in Columbus. Second, this research presents a new surveillance method, which enables continuous and precise surveillance analysis with the application to residential crimes, along with the comparison to the results of other existing measures.

Chapter 2 proposed theoretical backgrounds for urban ecological crimes analyses and space-time surveillance analyses. First, several significant urban sociological theories and
special features of spatial data were presented. The urban sociological theories explain how and why urban areas have higher crime rates in the context of socio-demographic features, social ties, and crime opportunities urban residents faced. Second, several surveillance models were summarized by their statistical features and assumptions. The surveillance models were categorized by their capacity to handle space or space-time dimensions of data. In the spatial model, test to detect cluster determines which collection of cases represents the most significant cluster, and a test of clustering provides a single assessment of the statistical significance of the pattern for the entire area. In both test of cluster and clustering, general and focused tests are distinguished by the use pre-specified factors (focused) or not (general).

Chapter 4 analyzed residential crimes in Columbus using Bayesian hierarchical data analysis. The applied Bayesian hierarchical modeling handled spatial heterogeneity and dependence. The results showed that the residential crime data have significant spatial dependence patterns and heterogeneity patterns, and the map decomposition technique visualized these patterns. Among the applied covariates, poverty was the most significant factor that contributed to economically deprived areas having higher residential crime rates. Contrasting to common expectations, the number of households showed a negative effect to crime rates. In addition, the map decomposition technique showed that some areas have high crime rates purely from spatial effects of crime dates rather than socio-demographic covariates.
Chapter 5 analyzed and compared the results of three prospective space-time surveillance of the residential crimes in Columbus: space-time scan statistic, CUSUM version of Tango’s statistic, and a new space-time surveillance method. Space-time scan statistic captures emerging clusters and adjusts multiple testing problems by applying cylindrical windows for testing prospective space-time crime hotspots. However, its dependence on population information disabled continuous surveillance analyses over time. Because of area based analytical structure, the space-time scan statistic had problems to provide precise spatial and temporal ranges in crime hotspots. In addition, the use of gradually expanding windows contained spatial approximation in resulting hotspots. However, since the CUSUM statistic is a general clustering test which assesses the changes of spatial pattern over the entire area, the test results cannot provide crime hotspot locations. In addition, the assumption of fixed population over time did not provide robust results when the study area had incoming populations.

The new surveillance method, which is newly devised and applied to the same dataset, was inspired by temporal extension of circular windows to cylinders by Kulldorff (2001) and temporal updates of spatial analyses in spatial CUSUM by Rogerson (1997). The applied model showed that it overcame the limitations of both space-time scan statistic and CUSUM version of Tango’s statistic. The model can be applied continuously over time without the limitation from population data availability and does not have hotspot area approximations leading to precise spatial hotspot ranges. In addition, the model is relatively more robust to the effects of underlying population changes because it is based
on previous crime occurrences rather than a fixed population size. Because the new surveillance method does not account for population size, the method may have a critique that it may find false clusters, which are caused by sudden large population increase. In the situation, other surveillance methods will continuously find false clusters until population information is updated; it can be even several years considering slow (decennial) population updated in Census data. In comparison, the new surveillance methods may find false clusters once. However, the clusters would be adjusted afterward and find true clusters because expected values are updated based on previous observations.

The new surveillance method has drawbacks at the same time. Selecting proper values for a spatial and a temporal parameter, $r$ and $q$, is no explicit criterion has been developed yet. An appropriate spatial parameter should reflect the size of the study area and the intensity of occurrences. A dynamically adjustable temporal parameter which accounts for temporal dependence would be desirable for further studies.

6.2 Future Research Direction

In spite of its contributions to current literature about prospective space-time surveillance modeling, It can be concluded that the model can still be further developed and extended based on the empirical needs in the field as follow.
First, automatic calculations of optimal spatial and temporal parameters, \( r \) and \( q \), are desired. The parameter calculation should consider the temporal frequency and the spatial intensity of the cases. With properly devised parameters, the surveillance analyses can generate robust results in finding space-time clusters.

Second, given that location information of individual crime cases is registered with street addresses, street network based crime hotspot analyses would generate superior results rather than using circular hotspot windows. Already, several studies present the application of network based Voronoi diagram (Okabe et al., 2006; Okabe and Satoh, 2006). The application of a network based model would contribute to generating more flexible hotspot shapes rather than circles and produce precise results by reducing the spatial approximation in the significant crime hotspots.

Third, the future development of surveillance software is expected. Availability of easily applicable software is important because general surveillance users may not have proficiency in statistical modeling and computer programming to implement the model. There are several software available for the application of hotspot analyses (Levine, 2006) and prospective surveillance (Kulldorff and Information Management Services Inc, 2006; Lee et al., 2006). However, considering the limitations of current surveillance measures, the software has the same limitations. When prospective hotspot results is
combined with corresponding optimal police routing information, the newly developed software may provide a valuable tool for police, which are “being asked to do the same, or even more, in times of fiscal constraint” (Ratcliffe, 2001). I believe that it is a very attractive contribution geographers can make.

Finally, enhanced visualization with a real-world image can be considered. Combined to a rapid development of technologies in 3 dimensional mapping and satellite images, presentation of crime hotspots can be further extended. Figure 6.1 and Figure 6.2 presents visualizations of crime hotspots on real-world images as preliminary examples. These visualizations enable straightforward recognition of the locations and the ranges of the hotspots particularly for the residents of the areas. In addition for the urban counselors and police officer, the visualization help consider different land use types so that they can focus on residential areas in locating their crime control effort, excluding irrelevant areas such as rivers and vacant forests.

The application of the Bayesian hierarchical data analysis and the new surveillance method presented in this study are not limited to crime analyses. Instead, it is expected to contribute to many other fields dealing with both spatially and temporally referenced point data, such as epidemiology, by finding space-time clusters of a disease. In addition, it should be acknowledged that this research is just an initial step to understand crime as
an example of space-time phenomena. I hope that this study may lead to further interests
and developments in the field.
Figure 6.1 A satellite image of crime hotspots (small scale)
Figure 6.2 A satellite image of crime hotspots (large scale)
BIBLIOGRAPHY


Block, 1979, Community, environment, and violent crime, *Criminology*, 17, 46-57.


Federal Bureau of Investigation, 2000, Crime in the United States, FBI, Washington, DC.


128


Kulldorff, M. and Information Management Services Inc. 2006. *SaTScan v7.0: Software for the spatial, temporal, and space-time scan statistics*.


Page, E. S., 1954, Continuous inspection schemes, Biometrika, 41, 100-115.


Zhang, T. and Lin, G., 2006, A supplemental indicator of high-value or low-value spatial clustering, Geographical Analysis, 38, 209-225.