SPATIAL ANALYSIS OF CRIMINAL INCIDENTS: IMPLEMENTATION OF GIS AND SPATIAL STATISTICS

by

ABDULRAHMAN ABDULQADER, B. Planning & Statistics (UAE University)

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Science in Mathematics (Specialization: Statistics and Probability)

School of Mathematics and Statistics, Carleton University
Ottawa-Carleton Institute of Mathematics and Statistics
Ottawa, Ontario, Canada

© Copyright 2006 Abdulrahman Abdulqader
NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

AVIS:
L’auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l’Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L’auteur conserve la propriété du droit d’auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n’y aura aucun contenu manquant.
ABSTRACT

This thesis consolidates theory for several tools - from GIS, spatial statistics, and data mining - useful for analysing criminal incidents to discover crime patterns, recognize extraordinary concentrations of crimes or hot spots that show clusters of criminal events, or characterize criminal behaviour in the different neighbourhoods. Their use is illustrated with Ottawa Police Services data for 2001 and 2004.

Chapter 1 presents motivation, problem, and literature review. Chapter 2 presents various maps and introduces the GIS and some statistical manipulation of GIS output. Chapter 3 presents statistical spatial data analysis for irregular lattice type data subject to high variability; various data representations, comparison methods, and models assess large-scale variability (trend) and small-scale variability (spatial dependency). Chapter 4 uses spatial regression modeling to test for spatial dependence due to an omitted spatial lag or spatially autoregressive errors. Parameter estimation is addressed whether or not there is significant spatial dependence. Stochastic processes are used in Chapter 5 to check for spatial structure or complete randomness in the data; analysis and testing for randomness or clustering is done in terms of spatial point patterns. If there is structure, the pattern of location of events in a particular region, the intensity of the spatial point pattern, and ways to test whether or not there is clustering are given as well as ways to identify the clustering patterns. Chapter 6 discusses limitations of this research and gives future directions.
ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to HH Lt.Gen. Sheikh Saif Bin Zayed Al Nahyan, the Minister of Interior, United Arab Emirates for giving me the opportunity to continue my studies. This reflects his concerns for enhancing the scientific level of Abu Dhabi Police Headquarters’ officers and their ability and tactics in executing their security responsibilities in order to protect society from the danger of crimes that threaten it.

I would also to thank my supervisor, Prof. Shirley E. Mills for her excellence guiding me through all portions of my research. Her endurance, kindness, and follow-up of my improvement were extremely valued. I have comprehensively benefited from working under her supervision.

I would like also to acknowledge Ottawa Police Services for giving me data and access to the facilities required for my research.

I would like to express my appreciation to Prof. E. J. Norminton for his assistance in preparing this manuscript in $\TeX$ format.

Lastly I would like to give special thanks to my wife for all her encouragement, patience, and taking care of me during my Masters studies.
# Table of Contents

Acceptance

Abstract

Acknowledgments

Table of Contents iv

List of Tables ix

List of Figures x

Chapter 1: Introduction 1

1.1 Motivation 1

1.2 Geo-referenced data 1

1.2.1 Geostatistical data 2

1.2.2 Lattice data 2

1.2.3 Point patterns 3

1.3 Background 3

1.4 Purpose 4

1.5 Software 5

Chapter 2: Data Visualization: Crime Mapping 6

2.1 Historical Background 6

2.2 Types of Mapping 7

2.2.1 Manual Pin Mapping 7

2.2.2 Computer Mapping 7
2.3 Geographic Information System (GIS) 8
2.4 Crime Analysis Mapping 9
  2.4.1 The Map Purpose 10
  2.4.2 Audience 11
  2.4.3 Data Types 11
    2.4.3.1 Tabular data 11
    2.4.3.2 Geographic (spatial) data 12
  2.4.4 Data Management 13
    2.4.4.1 Metadata 13
    2.4.4.2 Geocoding 13
  2.4.5 General Data Integrity Issues 15
  2.4.6 GIS Components 17
    2.4.6.1 Data representation 18
    2.4.6.2 Data of geographic features 22
    2.4.6.3 GIS toolbox 24
    2.4.6.4 Visual display/mapping 25
    2.4.6.5 Querying GIS data 26
  2.5 Examples of Crime Analysis Map Production 29

Chapter 3: Statistical Analysis of Geostatistics and Lattice Data 40
  3.1 Non-residential Breaking and Entering Data 41
  3.2 Nonspatial Data Analysis 44
3.2.1 Stem and leaf plots
   3.2.1.1 Results for 2004
   3.2.1.2 Results for 2001
3.2.2 Choropleth maps
   3.2.2.1 Results for 2004
   3.2.2.2 Results for 2001
3.2.3 Relative Risk Map
   3.2.3.1 Results for 2004
   3.2.3.2 Results for 2001
3.2.4 Probability Map
   3.2.4.1 Results for 2004
   3.2.4.2 Results for 2001
3.3 Exploratory Spatial Data Analysis
   3.3.1 Spatial smoothing
   3.3.2 Median Polish
   3.3.3 Transformations
      3.3.3.1 Results for 2004
      3.3.3.2 Results for 2001
   3.3.4 Trend Removal
      3.3.4.1 Results for 2004
      3.3.4.2 Results for 2001
List of Tables

Table 3.1. Ward Information for Ottawa 43
Table 3.2. Results of Different Maps for Nonresidential B&E data - 2004 47
Table 3.3. Results of Different Maps for Nonresidential B&E data - 2001 49
Table 3.4. Estimates of overall effect, row effects and column effects - 2004 62
Table 3.5. Estimates of overall effect, row effects and column effects - 2001 63
Table 4.1. OLS model output for 2004 data 85
Table 4.2. Regression diagnostics – OLS model for 2004 data 86
Table 4.3. Spatial autocorrelation diagnostics – OLS model 88
Table 4.4. ML estimation results, spatial lag model, ZI04 88
Table 4.5. Diagnostics, spatial lag model, ZI04 89
Table 4.6. OLS model output for 2001 data 91
Table 4.7. Regression diagnostics – OLS model for 2001 data 92
Table 4.8. Spatial autocorrelation diagnostics – OLS model 92
Table 5.1. Counts by quadrat 98
Table 5.2. Frequency distribution of quadrat counts re Fig.5.2 99
Table 5.3. Frequency distribution of quadrat counts re Fig.5.3 100
Table 5.4. Quadrat frequency distribution and computations 114
List of Figures

Figure 2.1. A computer map showing the location of Carleton University 8
Figure 2.2. Single events locations of auto-theft as point features 18
Figure 2.3. Different geographic features represented by different lines 19
Figure 2.4. Polygon features 20
Figure 2.5. Aerial photograph 21
Figure 2.6. Digital orthophotograph of Ottawa 22
Figure 2.7. Point data from GIS 23
Figure 2.8. Detail of event at particular location 24
Figure 2.9. Layers of GIS 26
Figure 2.10. Query on B & E 27
Figure 2.11. Query on B & E in eastern Ottawa 27
Figure 2.12. Map with eastern area data in blue 28
Figure 2.13. Buffer around a school and crimes within 28
Figure 2.14. Highlighting where particular crimes occurred 29
Figure 2.15. B & E in Dalhousie St. area in second half of 2004 30
Figure 2.16. Graduated size point map of Figure 12.15 31
Figure 2.17. Shaded land use parcels with number of incidents re Fig. 2.15 32
Figure 2.18. Final detailed map re Figure 2.15 33
Figure 2.19. Thefts from vehicle Ottawa, 2001-2004 34
Figure 2.20. Graduated size point map -thefts from Ottawa, 2001-2004 35
Figure 2.21. Choropleth map for thefts from vehicle by census tract, Ottawa, 2001-2004

Figure 2.22. Graduated color map—census tracts shaded by the total number of thefts from vehicle per 1000 population

Figure 2.23. Modified Fig. 2.22 with census tracts, city boundaries and scale

Figure 3.1. A map showing the 21 wards of Ottawa

Figure 3.2. Stem-and-leaf plot of \( \{ W_i, i = 1, \ldots, 21 \} \), 2004

Figure 3.3. Stem-and-leaf plot of \( \{ W_i, i = 1, \ldots, 21 \} \), 2001

Figure 3.4. Map of Weighted Nonresidential B&E \( \{ W_i \} \), 2004

Figure 3.5. Map of Weighted Nonresidential B&E \( \{ W_i \} \), 2001

Figure 3.6. The Relative Risk of Ottawa wards in 2004

Figure 3.7. The Relative Risk of Ottawa wards in 2001

Figure 3.8. Poisson probabilities for Ottawa’s Nonresidential B&E in 2004

Figure 3.9. Poisson probabilities for Ottawa’s Nonresidential B&E in 2001

Figure 3.10. Stem-and-leaf plot of \( \{ Z_i, i = 1, \ldots, 21 \} \), 2004

Figure 3.11. Stem-and-leaf plot of \( \{ Z_i, i = 1, \ldots, 21 \} \), 2001

Figure 3.12. Contour map of \( Z_i \) for 2004

Figure 3.13. Perspective plot of \( Z_i \), for 2004

Figure 3.14. Low resolution map of the Freeman-Tukey transformed \( NR_i \)

Figure 3.15. A Choropleth map of Weighted Median-polish fit - 2004

Figure 3.16. A Choropleth map of Weighted Median-polish fit - 2001

Figure 3.17. Stem-and-leaf plot of weighted median polish residuals - 2004
Figure 3.18. Histogram of weighted median polish residuals -2004

Figure 3.19. Stem-and-leaf plot of weighted median polish residuals - 2001

Figure 3.20. Variogram estimator based on standardised residuals -2004

Figure 3.21. Variogram estimator based on standardised residuals -2001

Figure 4.1. Ward connectivities

Figure 4.2. The 4 wards with 0 connectivity

Figure 4.3. The 2 wards with 8 connectivities

Figure 4.4. Plot of residuals against predicted values - OLS 2004

Figure 4.5. Plot of residuals against predicted values - OLS 2001

Figure 5.1. B & E incidents in Central East Ottawa 2004

Figure 5.2. Partitioning into quadrats

Figure 5.3. Filner partitioning into quadrats

Figure 5.4. Kernel estimate map, $\tau = 60$

Figure 5.5. Kernel estimate map, $\tau = 120$

Figure 5.6. Kernel estimate map, $\tau = 180$

Figure 5.7. Plot of $\hat{G}(w)$ versus distance $w$

Figure 5.8. Plot of $\hat{F}(x)$ versus distance $x$

Figure 5.9. Plot of $\hat{G}(w)$ against $\hat{F}(x)$

Figure 5.10. $\hat{K}$ function for B & E

Figure 5.11. Plot $\hat{L}(h)$ against $h$ for B & E

xii
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.12</td>
<td>Test for CSR based on $\hat{G}(w)$</td>
<td>118</td>
</tr>
<tr>
<td>5.13</td>
<td>$\hat{L}(h)$ and its simulation envelope for B &amp; E data</td>
<td>120</td>
</tr>
<tr>
<td>5.14</td>
<td>B &amp; E incidents in Central East Area of Ottawa, 2004</td>
<td>121</td>
</tr>
<tr>
<td>5.15</td>
<td>B &amp; E incidents in Central East Area of Ottawa, 2004</td>
<td>122</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Motivation

The use of up-to-date statistical practice to analyze criminal data is crucial in all law enforcement activities. This thesis is motivated by the fact that Abu Dhabi Police General Headquarters is interested in developing its statistical system pursuant to the latest statistical techniques in order to address a variety of threats that are facing its community and to be able to develop strategies to fight and prevent crimes. Ottawa Police Services data has been used for illustrative purposes because currently data from Abu Dhabi Police General Headquarters is not geo-referenced and hence is inadequate for the analyses presented in this thesis.

Most spatial analyses of crime data do not attempt to model the individual criminal behaviour but rather look at the aggregated actions of several criminals and try to understand the relationship between the crime and the location. Widespread use of geographic information systems (GIS) increases the ability to examine the role of location in criminal activity.

The primary reason for studying spatial statistics in law enforcement is to answer the question “how many crimes happened where” instead of just answering the question of only “how many crimes happened”. This is now possible because of the availability of geo-referenced data, which allows us to answer the “where” part.

1.2 Geo-referenced data

Geo-referenced data refers to the fact that the data quantities \( \{s_1, ..., s_n\} \) are associated with specific geographic locations \( \{Z(s_1), ..., Z(s_n)\} \). This association usually comes in a \( d \)-dimensional positive volume Euclidian space \( \mathbb{R}^d \) and usually \( d = 2 \) or \( 3 \). The spatial data follows a general
spatial model, i.e., let \( s \in \mathbb{R}^d \) denote a generic data location, and \( Z(s) \) is a random potential datum at spatial location \( s \). Let \( D \subset \mathbb{R}^d \) and \( s \) varies over its index to generate the random process

\[ \{Z(s), \ s \in D\}. \]

There are three main components of general spatial processes. The data types are distinguished by the nature of the spatial domain \( D \).

### 1.2.1 Geostatistical data

Geostatistics applications began with ore-reserve estimation, and the prefix "geo" indicates that the statistical techniques were designated to deal initially with the earth. \( D \) is a fixed subset of \( \mathbb{R}^d \) and \( Z(s) \) is a random vector at location \( s \in D \). Geostatistical-type problems are distinguished from lattice and point pattern-type problems by the ability of the spatial index \( s \) to vary continuously throughout \( D \) which is a subset of \( \mathbb{R}^d \) in the random field process \( \{Z(s), s \in D\} \). The primary tool in geostatistical analysis is the variogram.

### 1.2.2 Lattice data

\( D \) is a fixed set of a countable number of points in \( \mathbb{R}^d \) and \( Z(s) \) is a random vector at location \( s \in D \). Lattice refers to a finite or countable collection of spatial sites (usually regions, areas, counties, wards, etc.) where observations are made, and the lattice could be regular or irregular. The spatial sites in a lattice are typically identified by specifying the centroid or the seat of each region or area by using its longitude \((x)\) and latitude \((y)\), and usually are represented by a map showing the index of each area. Another aspect of lattice data is that of neighborhoods, which are important for building a model of spatial dependence and spatial regression.
1.2.3 Point patterns

$D$ is a point process in $\mathbb{R}^d$ and $Z(s)$ is a random vector at location $s \in D$. When the important variable is the location of the event, the analysis of point patterns arise. Basic issues of interest in this situation are whether or not the points of interest exhibit complete randomness, do the points cluster together in some behaviour, or possibly do the points occur with a kind of regularity throughout the area (Cressie, 1993).

1.3 Background

Current efforts to discover spatial relationships and associations between crimes are limited because crime analysts are accustomed to manually acquire data, link records, and map the outcome. This limits awareness to major crimes only. Tools have been developed to help in crime analysis, particularly spatial analysis. But they are lacking for they do not help in acquiring data sets according to the users' needs, analyses are not automated, and GIS proficiency is required. ReCAP (Brown, 1998) is software that, in addition to solving the insufficiency of other tools, provides numerous extra utilities in order to analyze the data. The system is an interactive shared information and decision support system that uses databases, geographic information system (GIS), and standard statistical tools to analyze, predict and display crime patterns.

Spatial data mining can provide law enforcement agencies with the necessary tools to analyze crime data. Chawla et al. (2000) compared classical data mining and spatial data mining to develop a spatial regression model which uses geo-spatial data and incorporates spatial properties into both classical statistical analysis and data mining frameworks. The model includes the idea of spatial autocorrelation according to the first law of geography “Everything is related to everything
else, but nearby things are more related than the distant things." When both classical and spatial models are applied to real world data, the spatial regression model has been shown to be efficient and significantly better than the classical model.

Nowadays a huge amount of criminal data is collected without a prior hypothesis or experimental structure. Estivill-Castro and Lee (2001) incorporated clustering and association-rule mining techniques in algorithms to discover spatial patterns and autonomously detect cause-effect associations among layers of point and area records, combined groups of spatial concentrations, and generated hypotheses. In addition to complementing the usual statistical analysis on geo-referenced data, the algorithms raise the capacity of GIS to develop into intelligent pattern spotting rather than just to serve as a repository to store, manipulate and retrieve data.

Xue and Brown (2003) analyzed criminal incidents as spatial choice processes and used this to discover the distribution of peoples’ behaviours in space and time.

1.4 Purpose

The general purpose of this thesis is to present an illustrative analysis of criminal incidents that integrates a layer-based GIS environment and spatial statistical analysis. This work will bring together Geographic Information Systems (GIS) and spatial analysis techniques so that one can discover crime patterns, recognize extraordinary concentrations of crimes or hot spots that show clusters of criminal events, or characterize criminal behaviour in the different neighborhoods of concern. The main contribution of this thesis is that it gathers together the theory for several tools from GIS, spatial statistics, and data mining that are useful in analyzing criminal incidents and it uses Ottawa Police Services data for 2001 and 2004 to illustrate application of the theory.
This thesis will examine:

i) Data visualization using GIS software

ii) Analysis of geostatistics and lattice data

iii) Analysis of data using spatial regression methods

iv) Application of spatial point pattern analyses

1.5 Software

The software used in this thesis is the following:

• ArcGIS Ver: 9.0 which includes the ArcMap Ver: 9.0

• R Ver.2.0

• R Ver.2.1

• R Ver.2.2

• GeoDa™ Ver: 0.9.5-i5 (Aug 3, 2004)
Chapter 2: Data Visualization: Crime Mapping

2.1 Historical Background

Mapping by itself has a broad history; however, crime mapping in particular goes back in time to near the beginning of the 1800s, as social theorists started to produce maps that consisted of single symbol points and graduated area maps to demonstrate their research and theories regarding crimes. For the purpose of policing and crimes, maps originally were used for looking at important issues like crime, demographic characteristics, or poverty (Boba, 2001). Crime mapping started in France and spread to England. The New York City police department was one of the first departments that used crime mapping, beginning in the 1900s. The department used simple hardcopy pin maps to chart criminal activity and pins of different colours represented types of crimes (Harries, 1999).

Sociologists at the University of Chicago used graduated area maps to observe crimes and criminal behaviour during the 1920s and 1930s, particularly juvenile delinquency and associated social characteristics. The first computerized crime mapping was in the 1960s. In the 1980s, additional sophisticated technologies and desktop PCs became available and were more commonly used. However, slow processing speeds and poor printing quality limited the quality of computerized mapping at that time. Around the same time, intellectuals started to inspect the spatial characteristics of crime in addition to how location characteristics could be involved in criminal activity at particular places and environmental criminology theory came into being. In the 1990s, geographic information systems (GIS) became generally available for personal computers, and in the late 1990s, law enforcement agencies and criminologists generally used the GIS (Boba, 2001).
2.2 Types of Mapping


2.2.1 Manual Pin Mapping

In the last 100 years, police officers have fixed pins into paper maps displayed on walls, where each pin represented a crime location (Ratcliffe, 2004). Maps have been a practical instrument for crime analysts for a long time. These maps have in general served to illustrate at a glance which neighbourhoods are affected by severe crimes (Lodha and Verma, 1999). Even if they are helpful, manual wall maps offer inadequate efficacy since they can only show a small amount of data, they are hard to keep updated, they are not accurate, and they are not easy to read. For instance, these types of maps do not easily demonstrate date, time, character of incidents, etc. although dissimilar coloured pins can represent different types of crime. Each month pins must be removed in order to update a manual wall map. If there is no way to keep the record of the previous month’s map, the information illustrated on the previous map is lost. As a result, comparisons of one month to the next are not possible. Finally, when a manual pin map displays a large amount of data it becomes incomprehensible because of the many pins and/or holes (Boba, 2001).

2.2.2 Computer Mapping

The computer is used to put a point at a particular location just as if someone would put a pin on a wall map, so a computer map is similar to a wall map. Therefore, computer maps have limitations similar to wall maps (Boba, 2001). To illustrate, Figure 2.1 is an example of a computer mapping where a map has been taken from Microsoft® Streets & Trips (a computer-mapping...
program), which locates restaurants, ATMs, hotels and other points of interest, in addition to an address if entered. The surrounding area map will appear with a pin to locate the address. It is possible to zoom in and zoom out of this kind of map; the geographic features (e.g., roads, places, parks) are fixed and cosmetic only.

Figure 2.1. A computer map showing the location of Carleton University

For example, when using Microsoft® Streets & Trips to locate an address, clicking on a location will not give more information behind that point, such as a particular address or other demographic information. Thus computer mapping does not allow any deeper study than manual pin mapping, even if visually interesting and simple to apply.

2.3 Geographic Information System (GIS)

A Geographic Information System (GIS), according to (Mena, 2003), is “an interactive mapping system that permits information layering to produce detailed descriptions of conditions and analyses of relationships among variables. A GIS is based on drawing different spatial distributions of data and overlaying them on one another to find interrelated points. Conditions, or filters, can be
used in a GIS to refine searches at any level an analyst chooses”. It can be used for analyzing physical space, assigning perspective, and producing visual images of different styles of data in map layouts. Data displayed in the form of a map facilitates understanding of the significance of where, when, and by whom crimes are committed (Stoe et al, 2003). It also has the capability to change, visualize, query, and analyze tabular and geographic (spatial) data. The difference between manual pin maps and the GIS is that GIS allows the analyst to examine the data beyond the geographic features, to join various features, to control the data and maps, and to perform statistical calculations. It allows the user to create anything from a simple point map to a three-dimensional visualization of spatial or temporal data (Boba, 2001). There are many different brands names of GIS software packages, like ArcGIS®, MapInfo®, GeoMedia®, Atlas GIS®, and Maptitude®.

2.4 Crime Analysis Mapping

Crime Analysis Mapping (CAM) is the combination of crime analysis techniques with the geographic information system in order to focus on the spatial perspective of criminal activities and other law enforcement operations (Boba 2001). Crime Analysis Mapping can be a problem-solving tool and it can also identify new problems facing law enforcement. In addition to providing a visual perspective for a study, it can identify responses that will make the fight against criminal activities more effective, support the assessment and evaluation systems, and supply a general suggestion aiding the arrangement of partnerships. Three factors must taken into consideration in order to produce accurate and effective crime analysis maps: (i) the map purpose, (ii) the map audience, and (iii) the data types used in the map (Velasco and Boba, 2000). These factors usually determine the nature of the map that will be used and the presentation method.
2.4.1 The Map Purpose

Using Crime Analysis Mapping gives one the ability to help in all steps of the problem solving sequence. If the purpose of a crime analysis map is to assist in the recognition of a specific problem, crime analysts may map chosen data to discover patterns of activity that have, until that time, gone unnoticed. For instance, to decide if there is a pattern to commercial burglary incidents, analysts may map a past month's data for every commercial burglary that happened in the city. Analyzing for clusters of activities may identify a group of incidents occurring in a neighbouring area. Mapping for successive months may show pattern changes. In this case, the crime analysis mapping purpose is to discover a problem i.e., in a specific area, 'Is there a sudden increase in commercial burglary incidents?' Studying a particular problem in depth is another purpose. Crime Analysis Mapping can be used to carry out a complete spatial analysis of the issue to examine it more closely from different viewpoints by using a mixture of data. For example, daytime residential burglaries may have significantly increased. To determining where the activity has increased the most, crime analysis mapping has been used. In addition to burglary data, the locations of schools, school absentee information, and identified offenders plus demographic information may used to decide if neighbourhoods with the greatest increase are similar to one another. Finally, the purpose of the map might be a presentation, and the complexity or simplicity of the map would depend on the level of the data and the analyses that are necessary. Moreover, when a problem is recognized and analyzed, the presentation of a map becomes important, and it is important to set up maps that are suitable for the purpose (Velasco and Boba, 2000). For instance, a comprehensive map of every hotspot location may be required for patrol officers and investigators who need specific information that focus their
problem solving efforts. On the other hand, monthly maps of a city representing reported crimes are useful for presenting general information to the public regarding the variation in hotspots of reported crime.

2.4.2 Audience

It is necessary to think about the audience who will use the information when deciding the purpose of a crime analysis map. Maps can be developed in a way to be suitable for different audiences according to their requirements, prospects, and skill level. These groups might consist of law enforcement staff, community groups, city council members, the ordinary public, etc. Political, legal, and ethical concerns have to be balanced with the requirements of users of the information when creating a crime analysis map. For instance, due to individual privacy rights, law enforcement agencies cannot distribute to the public a detailed map of the houses of every identified gang member that also contains personal information. But such maps are practical to gang enforcement officers collecting intelligence information. Maps may include an annotation such as 'law enforcement use only' or 'external use' to indicate the anticipated audience (Velasco and Boba, 2000).

2.4.3 Data Types

Crime analysis maps usually involve two (2) types of data: tabular data and geographic (spatial) data.

2.4.3.1 Tabular data

Tabular data is a table that contains a list of records accompanied by geographic information or some type of address about the variables. The tabular data source holds information on events that
are not mainly geographic however it regularly has a geographic element (Boba, 2001). Crime data as well as calls-for-service data are the types of tabular data that are usually mapped in law enforcement. For instance, the data may have information on crime incidents and calls-for-service activity like the category of activity, time, date, priority, etc. Other law enforcement tabular data contains accidents, identified criminals, field information relative to criminal incidents, and citizen survey results representing victimization or fear of crime. Generally at least one geographic component is common for both types, such as address and district. According to the particular circumstances, further tabular data from outside of law enforcement might be contained in the map, like census data, school information, and land use information (Velasco and Boba, 2000).

2.4.3.2 Geographic (spatial) data

This data is primarily geographic; namely, it explains phenomena that are naturally geographic. Although these kinds of data may consist of non-geographic information, its main focal point is the geographic features (Velasco and Boba, 2000). Even though geographic data has related tabular data (attributes), the tabular data do not have a value exterior to a GIS (Boba, 2001). Law enforcement examples of geographic data include neighbourhood boundaries, gang territorial boundaries, zone boundaries, and district boundaries. Census tract data consist of information about population in addition to further demographic variables but this data is based on the geographic boundaries of the census tracts. Additional cases of geographic data consist of streets, land parcels, rivers, roads, block clusters, and metropolitan boundaries (Velasco and Boba, 2000).
2.4.4 Data Management

2.4.4.1 Metadata

Metadata is a term that refers to data about data (or to the data that illustrates the analysts work). It is important that each police department record the set of procedures that sketches out how crime analysis is conducted rather just trusting it to the analyst's memory. This reduces duplication of efforts and can provide useful training tools for new analysts. Moreover, the metadata contains an outline of data cleanup and handling procedures, rules of sharing work with others, and maintaining a record of products and files produced. Metadata provide information about the databases used (Boba, 2001). The principles of metadata have been developed under the patronage of the Federal Geographic Data Committee (FGDC), a division that organizes the improvement of the National Spatial Data Infrastructure (NSDI) (Harries, 1999).

2.4.4.2 Geocoding

Incorporating tabular data with geographic (spatial) data to establish the geographic locations of records in a table is a procedure called "geocoding" (Velasco and Boba, 2000). Alternatively, it refers to the procedure of turning an address into a point on a map (Ratcliffe, 2004). The job of geocoding is to associate tabular data with a location on the surface of the earth to visualize the spatial properties of the data. It corresponds to insertion of a pin into a map at the correct point. When geocoding in a GIS, data associated with the data points is also available. The address is the common geographic unit of analysis used to make the geocoding possible (Boba, 2001). Geocoding is necessary for crime mapping because it is the most frequently used technique of setting crime or crime-related data in the GIS. The information that allows linking the database and the map are the
street addresses or any locational attributes that are usually assigned with crime records. The address is represented by x-y coordinates after geocoding the crime locations, either by latitude and longitude decimal degrees, or by state-plane x-y coordinates which are identified from a specific origin by feet or meter measurements (Harries, 1999). To know where the data points should be located on the map, the computer translates the x-y coordinates and represents their locations. The process of geocoding is discussed in detail in a report produced by the Crime Mapping Laboratory Police Foundation as a part of work carried out for the Office of Community Oriented Policing Services (COPS) and can be found at (http://www.cops.usdoj.gov/). The report is entitled, “Geocoding in Law Enforcement: Final Report (2000)”. According to this report, the general outline of geocoding is as follows:

1: Prepare the geographic and tabular files for geocoding

2: Specify geocoding parameters

3: Geocode

4: Review results

5: If necessary, re-specify parameters and geocode again.

There can be a few problems with the geocoding system and some causes for errors. The following are examples of these (Ratcliffe, 2001).

- The geocoding source database file is out-of-date, where new addresses are not known or the source database is not updated with the new addresses.

- Street names are misspelled or abbreviated.

- The target address cannot be matched to the geocoded address, because the street address name
of a road in the source database may not reveal the local name of the road.

- Address duplication.
- Non-existent addresses.
- The subdivision of line section may not reveal precisely the geography of the road or housing layout.
- The unnecessary additional that data complicate the address field, and makes the address-file noisy.
- Geocoding the non-address locations.
- Ambiguity or vague addresses in the target address file can make geocoding impossible.

2.4.5. General Data Integrity Issues

Tabular and geographic data of police departments are related to different data integrity issues (Boba, 2001). Below is a general overview of the major data integrity issues.

a) Data Entry

The people who carry out data entry in a law enforcement location may not be aware that others are using the data that they create; this can lead to carelessness and result in producing unreliable data. Proper training, making individuals aware of the uses and importance of the data, and technology (address-cleaning software), can reduce data entry errors.

b) Timeliness

One of the most important concerns in crime analysis is to obtain timely data. However, due to individuals’ inaccuracy and technological complexities, it is not always possible to obtain such data.
To solve this problem it is necessary to work with departments in order to come up with solutions for improving data timeliness and to show the necessity and the value of up-to-date data for crime analysis.

c) Data Validity

Like any other data, law enforcement data tends to have validity problems. Validity refers to whether the data reflects the matter to be studied. For instance, many law enforcement units use calls-for-service data as a proxy for crime. This is tricky, since generally not every call that is dispatched as a call-for-service is actually associated with a crime. To illustrate, consider a “robbery” call-for-service. There is a general uncertainty regarding the difference between a “burglary” and a “robbery”. A person who has been *burglarized* may call the police to report that he has been *robbed* while actually he has not. If the dispatcher does not change the name of the call after the actual crime is found out, and he leaves the entry as the person has initially reported it, then the call-for-service would not accurately correspond to the crime that happened. Also using call-for-service as a proxy for crime is tricky since the date and time would show when the police department has received the call, not when the crime really took place. Therefore, analysis of these types of data would give invalid outcomes.

d) Reliability

Reliability refers to the degree to which repeated measures of the same data result in the same readings. A policy or law change is an example of a reliability issue that is related to law enforcement data, such as obligatory arrest for domestic violence crimes. Often, a sharp increase or
decrease is related to the fact that the data do not correspond to the same phenomenon anymore. As a result, the data are not reliable. Clearly, reliability problems can significantly influence the results of crime analysis.

e) Data Transfer Process

Unintentionally losing or reformatting data during the data transfer process can have an effect on data quality and integrity. Data compatibility is also an issue related to the data transfer process. Law enforcement agencies are notorious for having data in several different formats. Furthermore transferring and joining these data is a time-consuming and annoying process.

f) Data Confidentiality/Privacy

Since crime analysts work with police data, they are in charge of protecting the data and those who are represented by the data. Data created and used in crime analysis usually is protected under policies on privacy and confidentiality. Yet new conditions of sharing information (like the development of the Internet and the use of mapping) require further and more detailed policies. Consequently, law enforcement departments must address particular crime analysis concerns in their data-protection arrangement. Crime analysis units need an explicit policy that addresses general issues that are related to requests for crime analysis information.

2.4.6 GIS Components

The major components of a GIS consists of five (5) parts: data representation, data features, data visualization, scale, and data querying.
2.4.6.1 Data representation

Real world data have an endless variety of shapes. All of them, however, are represented in a GIS as one of four (4) features – point, line, polygon, and image.

a) Point feature

A point is a zero-dimensional shape that illustrates a geographic feature which is too tiny to be portrayed as line or area, and is stored with the attribute as an x-y coordinate (Zeiler, 1999). A point feature in the GIS corresponds to a pin located on a wall map. Different symbols are applied to describe different crimes places (Boba, 2001). Figure 2.2 shows single event locations of auto-theft as point features. Depending on the type of the crime, different symbol styles are used (AUTOMOBILE, FROM VEHICLE, TRUCK) to shows the points on the map.

![Figure 2.2. Single events locations of auto-theft as point features](image)

b) Line feature

A line feature is a one-dimensional shape that corresponds to a geographic feature that is too
slim to be portrayed as area, and is stored with the attribute as sequence of structured x-y coordinates. The line feature can be straight, curved, oval, or splined (Zeiler, 1999). Figure 2.3 shows different types of geographic features such as roads, railways, and power lines that are represented by lines in a GIS. Every line feature is indicated by a single symbol so that it can be distinguished from the other lines.

![Image of different geographic features represented by lines]

**Figure 2.3.** Different geographic features represented by different lines

c) **Polygon feature**

A polygon feature is a two-dimensional shape that corresponds to several geographic features such as a sequence of parts that surround a region and form a closed area (Zeiler, 1999). Polygon features can represent large areas like continents or small areas like buildings. The following map, Figure 2.4, is an example of polygon features used in law enforcement. The largest (red) polygon

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
represents the city boundary, the blue are districts, and the gray are neighbourhoods groups.

![Figure 2.4. Polygon features](image)

d) **Image feature**

An image feature is a digitized picture taken from an aircraft or a satellite and positioned inside the GIS coordinate system so that there are x-y coordinates linked with it. The following image, Figure 2.5, is an illustration of an aerial photograph. In this picture,
the details of the streets, buildings, and environmental features such as landscaping are clearly seen.

There is a difference between an aerial photography that is just an image and a digital orthophotography that is an image joined with geometric characters of the map. Orthophotographs, such as roads and other features, can be viewed with other layers when they are placed inside the GIS (Boba, 2001). The following image, Figure 2.6, represents a digital orthophotograph with the related street network, parklands, and rivers.
2.4.6.2 Data of geographic features

A wall map and a computer map give a picture of points, lines, and polygons however they do not include data linked with the features and are not simply manipulated. The real power of a GIS is that it has the capability to view, query, relate, and manipulate the data behind the geographic features, because every type of feature has a table of data that explains it. Points, lines, and polygons are stored in a GIS as a data table. However, a digital orthophotograph does not have a related data table worthy of analysis although it has x-y coordinates. In a GIS, simply clicking on a particular feature can produce the data table linked with it (Boba, 2001). The following, Figure 2.7, is an example of point data taken from a GIS. The blue point in the map is a robbery incident that took place on Tuesday May 25, 2004 at 5:22 pm. The highlighted row in the table is the data corresponding to that point. Hence, each row in the data table describes one point on the map and
each column describes one variable about the data.

Moreover, one can use the *identify* option for each individual point by clicking on that point to get the details about it, as the following Figure 2.8 shows.
2.4.6.3 GIS toolbox

The functionality of manipulating the geographic data is the most important benefit of a GIS.

Below are some components that facilitate data manipulation.

a) Connectivity

Connectivity is the capability of the GIS to know that when two line-segments come together they form an intersection. It mainly refers to streets or other linear features. Connectivity lets
intersections be placed on the map, which is most essential in police data, where incident information usually lists an intersection as the location of an incident (Boba, 2001).

b) Contiguity

Contiguity represents adjacent areas by adjacent polygons. The ability of the GIS to distinguish adjacent polygons enables querying, selecting polygons surrounded by others, and visualization of complex land units (Boba, 2001).

c) Geometry

Geometry refers to the fact that all features in the GIS are associated at least with one set of x-y coordinates. The x-y coordinates of map features, particularly points, are important in law enforcement data because the location of a crime is then more accurate and there is no need for geocoding (Boba, 2001).

2.4.6A Visual display/mapping

a) Thematic approach

In a GIS, data is separated into themes or layers. A thematic layer is a group of common geographic components, such as crime locations, road networks, a collection of parcel boundaries, or digital air photos or satellite images. The group of themes are like layers in a stack. Each theme can be handled as a data group independent of the other themes, and each has its own set of symbols such as points, lines, polygons, images and so on. Since the different independent themes are geo-referenced, they can be placed on top of one another and can be joined in a common map display (Arctur and Zeiler, 2004). Data coming from two different sources would not be included in the same layer or theme, such as point data of crime and calls-for-service. This is a thematic
approach to managing data. The thematic approach technique permits the data to be analyzed and visualized independently or together. In the following Figure 2.9, each layer represents a different kind of feature and a different kind of data.

![Figure 2.9. Layers of GIS](image)

Since some layers are not transparent (and if placed on top would hide the underneath layer) it is important to be aware of how the layers are stacked (Boba, 2001).

b) Scale

Scale is the ratio or relationship between the dimension of features on a map and the equivalent space on the ground and is generally expressed as a ratio or fraction. If the scale of the map is 1:10,000, it implies that features on the map are ten thousand times smaller than their actual size. The proper scale is related to both the purpose of the map and the type of displayed data (Ormsby et al, 2004).

2.4.6.5 Querying GIS data

The enormous power of a GIS is that it does not only show where objects are, but also tells much about them. However, it is not possible to present everything that is known about a feature on
a single map. That is why queries are helpful to retrieve the unseen information about the features.

a) Tabular queries

Query expressions are formulated in GIS software to choose features both in the data table and on the map. For example, one can create a query to choose only the break and entry (B&E) incidents that occurred in the eastern area of Ottawa region (see Figure 2.10).

![Figure 2.10. Query on B & E](image)

After running this query, the eastern area data have been selected and highlighted with blue in the table that appears as Figure 2.11 and appear as blue on the map in Figure 2.12.

![Figure 2.11. Query on B & E in eastern Ottawa](image)
b) Spatial queries

A spatial query allows querying map features according to their location on the map or their location relative to other features. For example, we can draw a 1000 meters buffer around a school and select the crimes within this area (see Figure 2.13).

In addition, a feature could be selected according to its spatial relationship to another feature. We can select neighbourhoods where particular crimes occurred (see Figure 2.14); the neighbourhoods are highlighted in yellow.
2.5 Examples of Crime Analysis Map Production

Crime analysis maps are presented in different ways for different situations, depending on choice of relevant data, type of map to use, and purpose and audience for the map presentation. The following examples illustrate the procedures for producing different kinds of crime analysis maps. The final map at the end of each example will contain all of the proper design elements that explain how each map would finally be presented, such as legends and scale bars, roads, etc.

Example 1: Break and Enter Incidents

A number of Break and Enter incidents have happened recently in the neighbourhood of Dalhousie Street in the Lowertown and Byward Market areas in Ottawa. This led to the Crime Analysis Section collecting information on all Break and Enter incidents that have occurred in the area during the second half of 2004. The purpose was to report to the Break & Enter Response Team
(BERT) to assist them in finding out if a crime pattern existed, and hence to possibly identify the suspects in the area. In addition, the Community Policing Centers and neighbourhood officers in the area are provided with this data in order to help them in crime prevention programs and Neighbourhood Watch programs. The audiences are (BERT) officers, Community Policing officers, neighbourhood officers, and other police personnel. The outcome of information collection in the neighbourhood of Dalhousie Street was thirty (30) incidents (see Appendix.A: Table A.1). A map representing the land-use parcels has been used because the data was collected from a relatively small geographic area. Points in red (Figure 2.15) symbolize the break and enter incidents.

Unlike centerline road maps, the land-use parcel maps are important tools that allow detailed

Figure 2.15. B & E in Dalhousie St. area in second half of 2004

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
spatial analyses of a small area where crimes locations are close. To see if more than one incident took place at a particular location, a graduated size point map (Figure 2.16) has been used.

![Graduated size point map of Figure 2.15](image)

**Figure 2.16.** Graduated size point map of Figure 2.15

We get one location with three incidents, three locations with two incidents and the rest are associated with one incident only. A further way of representing the data is to shade the whole land-use parcel to show the occurrence of a criminal incident. By shading the parcels according to type of each land-use parcel and inserting numbers onto each land-use parcel to specify the number of incidents (Figure 2.17), we get a visual perspective of the study area, helping the police force community policing office to direct patrols in that area and to strengthen crime prevention efforts.
As these maps show, point maps are practical for investigating crime patterns and hotspots and supporting short-range problem-solving efforts. In addition, maps representing land-use parcels allow for a comprehensive study of particular small areas. Maps showing the exact locations of buildings within every land-use parcel would be practical for analyzing patterns in the occurrence and techniques of break and enter. Figure 2.18 is the final map.
A comprehensive summary of the incidents comes with the final map; it might include information such as date, case number, time, address, and modus operandi (Velasco and Boba, 2000).

**Example 2: Theft from Vehicle**

Although they slightly decreased in 2004, theft from vehicle incidents showed high rates during the last four years, where they represented 25.36% of all reported theft incidents in Ottawa during this period. Supervisors in law enforcement asked for information to see which part of the city is mainly affected by thefts from vehicle incidents. Patrol officers can use the information to...
enhance surveillance in the region. Furthermore, community policing officers use it to teach people how to stop auto theft related crimes in their neighbourhoods. The purpose of the analyses is to identify the more problematic areas for theft from vehicle for the purpose of starting an education campaign. Census tracts, city boundary, and tabular crime data are used in the data analysis. Because the purpose of the map is to decide which part of the city has the most thefts from vehicle, a map of the whole city, portraying streets, has been used. The study region is large, so it is not required to use detailed maps, like land-use parcels or streets since point symbols will cover a lot of the land-use parcels and streets. The following Figure 2.19 is a single symbol map depicting the reported thefts from vehicle for the last four years in Ottawa region.

![Figure 2.19. Thefts from vehicle Ottawa, 2001-2004](image)

The map is messy and it is not easy to identify accurately the high frequency areas, which is the purpose for making the map. Moreover, point data, which are geocoded to the same location,
will be stacked on top of each other and there would be no way to decide if more than one theft from vehicle took place at a specific place. Thus this map does not precisely describe theft from vehicle incidents, and it does not meet the requirements of the audience because it is almost impossible to discriminate one incident from another. A graduated size point map depicting all thefts from vehicle for the last four years in Ottawa area is used in Figure 2.20 as follows:

![Graduated size point map - thefts from Ottawa, 2001-2004](image)

**Figure 2.20. Graduated size point map - thefts from Ottawa, 2001-2004**

The use of a graduated size point map has clarified the difficulty of deciding whether several incidents have occurred in a particular location; yet, the map is still jumbled. Also, it is practically impossible to make a distinction between crimes happening at nearby locations, which prevent
further analysis. At this time in the analysis, it is necessary to remember the objective of the map and its viewers. The goal of the analysis is to help determine which part of the city has been most exposed to the occurrence of thefts from vehicle incidents in order for law enforcement people to focus their training and enforcement efforts. Clearly, a point map does not appear to be useful representing data with such a high amount of incidents. A choropleth map that uses colours to represent the different values for sections within the study region has been used to represent thefts from vehicle by census tract (see Figure 2.21 below).

![Choropleth map for thefts from vehicle by census tract, Ottawa 2001-2004](image)

Since the objective of the map is just to emphasize problem areas, more than particular locations, a choropleth map is useful in this example. Managers in law enforcement and community
policing officers tend to direct effort at general regions instead of particular neighbourhoods. However, choropleth maps could wrongly suggest that crimes are occurring homogeneously within the census tract. Geographic layers, block groups (smaller census areas), precincts, or reporting districts can also be used depending on the purpose and availability of the data (Velasco and Boba, 2000). The above map is likely matching the target of the analysis, since census tracts shaded with dark colors represent the high frequency areas for thefts from vehicle. On the other hand, the above map does not consider other factors like size, population, or number of dwellings units in each tract. One might suppose larger census tracts may have more vehicles and consequently more vehicle crime; similarly census tracts with a higher population or more dwelling units may have more vehicle crimes as well. Therefore, a map based solely on frequencies may not precisely reflect theft from vehicle crime if a common attribute is not used, such as area size, population, or number of dwellings units in the area. The following Figure 2.22 is a graduated color map that depicts census tracts shaded by the total number of thefts from vehicle per 1000 population.
Figure 2.22. Graduated color map-census tracts shaded by the total number of thefts from vehicle per 1000 population.

The map shows that there are two census tracts (the darkest) with high numbers of thefts from vehicle per one thousand of population. This may suggest police administrators focus their efforts to fight these crimes in these two areas. The following Figure 2.23 is the final map.
Theft From Vehicle in Ottawa Per 1000 Population
From 2001 to 2004

Figure 2.23. Modified Fig.2.22 with census tracts, city boundaries and scale.

The final map could be attached to a report including the total number of thefts from vehicle that happened per census tract during the last four years and other information like the incidents by year and month in addition to weekdays. This could be used to establish an overall view of the theft from vehicle problem in the city and accordingly education and prevention efforts could be directed to the most challenging areas.
Chapter 3: Statistical Analysis of Geostatistics and Lattice Data

In this chapter we present a statistical analysis of geostatistics and lattice data using the number of nonresidential breaking and entering (B & E) relative to the base of number of households at risk in the same year, for the wards of Ottawa for the years 2004 and 2001. The risk to the population of households and the spatial properties of the nonresidential units are the targets of our study in this chapter. Particular issues of concern are irregularly spaced data and high variability of the underlying number of households in wards, $HH_i$. Since the attribute of interest may display geographical variability, we begin with exploratory maps. To facilitate comparisons between wards we introduce a probability map that standardizes rates onto a probability scale and an index of deviation that quantifies deviation from a common average $\{\lambda_i/HH_i\} = p$ where $\lambda_i$ is the average count in ward $i$. First we need to transform the data to stabilize variance. Then, with our data on a lattice, we introduce a model which expresses the transformed data as an additive function of spatial location and spatial dependence. We first examine large-scale variation which gives an indication of spatial trend; we use spatial smoothing to obtain robust estimates of row and column effects and a fixed overall mean effect. Mapping this large-scale variation shows overall spatial trend and ignores small-scale variation. When we remove the spatial trend from our observed data, we isolate residuals that estimate small-scale variation (which consists of spatially dependent noise). Spatial analysis is then performed on these in order to address the small-scale variation and we introduce the concept of the variogram, which is used to measure of spatial dependence.
3.1 Non-residential Breaking and Entering Data

There are three (3) types of B & E: those in households (residential), those in public establishments (nonresidential or commercial), and those which are called "unlawfully in house". In Ottawa, breaking and entering is one of the most common forms of crime. Police statistics for the year 2004 registered 4,883 incidents (with a rate of 571.8 per 100,000 residents) and it occupies second place after theft incidents in the category of crimes against property (Ottawa Police Service, Annual Statistics – 2004). About 61% of breaking and entering offences took place in residential properties, 38% in nonresidential or commercial places, and 1% elsewhere (1). Some nonresidential properties consist of Single Home/House or Dwelling Unit or Commercial Dwelling Unit which by their nature look residential but in fact are commercial. These units interfere with the target of burglars. The risk to the population of households and the spatial properties of the nonresidential units are the target of our study in this chapter. As is recognized in time series analysis, we believe that data which are close together usually exhibit higher dependence than that which are farther apart. Data here is of an irregular lattice type where lattice refers to a countable collection of spatial sites or regions; it can be formed by specifying a particular feature common to each region. For example, in the case of the Ottawa area, ward centroid might be the chosen feature where we record its longitude (x) and latitude (y). It usually is represented by a map which is a primary component of any spatial analysis.

(1) Data source: Ottawa Police. This data represents a 30% random sample of data from the Ottawa Police Service Records Management System for data collected between 2000-01-01 and 2004-12-31.
The following Figure 3.1 shows a map of Ottawa wards numbered according to the City of Ottawa categorization as given in Table 3.1.

![Map of Ottawa wards](image)

**Figure 3.1.** A map showing the 21 wards of Ottawa

The concern here is with data at a finite number of spatial locations (21 wards) that are irregularly spaced: \( \{Z(s_i), i = 1, \ldots, n\} \), where \( \{s_i, i = 1, \ldots, n\} \) are the data locations in \( d \)-dimensional Euclidean space \( \mathbb{R}^d \) (for our data \( d = 2 \)). Here we analyze for the years 2004 and 2001 the number of nonresidential B&E relative to the base of number of households at risk in the same year, for the wards of Ottawa. Let \( \{NR_i, i = 1, \ldots, 21\} \) and \( \{HH_i, i = 1, \ldots, 21\} \) denote, respectively, the number of nonresidential B&E and the number of households in 21 wards of Ottawa. The following Table 3.1 (next page) shows the detail of the 2004 data.
<table>
<thead>
<tr>
<th>WARD</th>
<th>i</th>
<th>x</th>
<th>y</th>
<th>Area</th>
<th>Perimeter</th>
<th>Population</th>
<th>2001</th>
<th>NRB&amp;E</th>
<th>RB&amp;E</th>
<th>Population</th>
<th>2004</th>
<th>NRB&amp;E</th>
<th>RB&amp;E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orleans</td>
<td>1</td>
<td>38000</td>
<td>5040000</td>
<td>25.5</td>
<td>29.2</td>
<td>47940</td>
<td>15694</td>
<td>2</td>
<td>52</td>
<td>48201</td>
<td>16294</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>Innes</td>
<td>2</td>
<td>38000</td>
<td>5035000</td>
<td>40.4</td>
<td>33.0</td>
<td>38763</td>
<td>12812</td>
<td>0</td>
<td>37</td>
<td>39094</td>
<td>13292</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Bell/Nepean</td>
<td>3</td>
<td>36000</td>
<td>5015000</td>
<td>171.4</td>
<td>67.6</td>
<td>57644</td>
<td>20249</td>
<td>4</td>
<td>56</td>
<td>68097</td>
<td>24207</td>
<td>5</td>
<td>44</td>
</tr>
<tr>
<td>Kanata</td>
<td>4</td>
<td>35000</td>
<td>5025000</td>
<td>148.9</td>
<td>66.4</td>
<td>63290</td>
<td>21742</td>
<td>7</td>
<td>55</td>
<td>69926</td>
<td>24584</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>WestCarleton</td>
<td>5</td>
<td>33500</td>
<td>5025000</td>
<td>667.4</td>
<td>117.7</td>
<td>18219</td>
<td>6439</td>
<td>2</td>
<td>25</td>
<td>19227</td>
<td>6855</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Goulbourn</td>
<td>6</td>
<td>35000</td>
<td>5005000</td>
<td>274.2</td>
<td>67.5</td>
<td>25155</td>
<td>8434</td>
<td>1</td>
<td>9</td>
<td>28900</td>
<td>9668</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Bay</td>
<td>7</td>
<td>35500</td>
<td>5025000</td>
<td>52.1</td>
<td>35.7</td>
<td>47011</td>
<td>20880</td>
<td>1</td>
<td>49</td>
<td>46598</td>
<td>20731</td>
<td>1</td>
<td>39</td>
</tr>
<tr>
<td>Baseline</td>
<td>8</td>
<td>36000</td>
<td>5025000</td>
<td>14.3</td>
<td>18.4</td>
<td>41775</td>
<td>16031</td>
<td>2</td>
<td>54</td>
<td>42063</td>
<td>16515</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>Knoxdale/Merivale</td>
<td>9</td>
<td>36500</td>
<td>5020000</td>
<td>14.4</td>
<td>25.8</td>
<td>39902</td>
<td>14779</td>
<td>4</td>
<td>46</td>
<td>39489</td>
<td>14749</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Gloucester/Southgate</td>
<td>10</td>
<td>37500</td>
<td>5020000</td>
<td>249.1</td>
<td>69.4</td>
<td>53648</td>
<td>19987</td>
<td>2</td>
<td>55</td>
<td>58261</td>
<td>21853</td>
<td>3</td>
<td>63</td>
</tr>
<tr>
<td>BeaconHill/Cyrville</td>
<td>11</td>
<td>37500</td>
<td>5035000</td>
<td>19.7</td>
<td>25.9</td>
<td>35137</td>
<td>13098</td>
<td>0</td>
<td>36</td>
<td>35831</td>
<td>13792</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>Rideau/Vanier</td>
<td>12</td>
<td>36500</td>
<td>5035000</td>
<td>8.0</td>
<td>15.1</td>
<td>44093</td>
<td>22890</td>
<td>6</td>
<td>105</td>
<td>45260</td>
<td>23869</td>
<td>4</td>
<td>91</td>
</tr>
<tr>
<td>Rideau/Rockcliffe</td>
<td>13</td>
<td>37000</td>
<td>5035000</td>
<td>19.8</td>
<td>25.9</td>
<td>39373</td>
<td>17742</td>
<td>4</td>
<td>82</td>
<td>40232</td>
<td>18347</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>Somerset</td>
<td>14</td>
<td>36500</td>
<td>5030000</td>
<td>6.5</td>
<td>11.4</td>
<td>37944</td>
<td>21460</td>
<td>3</td>
<td>62</td>
<td>38250</td>
<td>21916</td>
<td>3</td>
<td>83</td>
</tr>
<tr>
<td>Kitchissippi</td>
<td>15</td>
<td>36000</td>
<td>5030000</td>
<td>15.1</td>
<td>16.4</td>
<td>38359</td>
<td>17389</td>
<td>3</td>
<td>82</td>
<td>38985</td>
<td>17727</td>
<td>1</td>
<td>59</td>
</tr>
<tr>
<td>River</td>
<td>16</td>
<td>37000</td>
<td>5025000</td>
<td>27.1</td>
<td>32.8</td>
<td>47714</td>
<td>19356</td>
<td>3</td>
<td>47</td>
<td>49076</td>
<td>20162</td>
<td>2</td>
<td>62</td>
</tr>
<tr>
<td>Capital</td>
<td>17</td>
<td>37000</td>
<td>5030000</td>
<td>10.8</td>
<td>19.2</td>
<td>35219</td>
<td>14953</td>
<td>5</td>
<td>50</td>
<td>35632</td>
<td>15335</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>AltaVista</td>
<td>18</td>
<td>37500</td>
<td>5030000</td>
<td>20.3</td>
<td>19.4</td>
<td>46203</td>
<td>19162</td>
<td>7</td>
<td>56</td>
<td>46004</td>
<td>19240</td>
<td>5</td>
<td>44</td>
</tr>
<tr>
<td>Cumberland</td>
<td>19</td>
<td>39000</td>
<td>5030000</td>
<td>310.2</td>
<td>79.5</td>
<td>24752</td>
<td>8241</td>
<td>1</td>
<td>25</td>
<td>30605</td>
<td>10520</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Osgoode</td>
<td>20</td>
<td>38000</td>
<td>5010000</td>
<td>385.7</td>
<td>82.8</td>
<td>18746</td>
<td>6484</td>
<td>1</td>
<td>27</td>
<td>20852</td>
<td>7215</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Rideau</td>
<td>21</td>
<td>36000</td>
<td>4995000</td>
<td>418.8</td>
<td>94.9</td>
<td>13307</td>
<td>4644</td>
<td>1</td>
<td>20</td>
<td>13791</td>
<td>4891</td>
<td>0</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 3.1. For each ward entries show: ward name, number, coordinates, area, perimeter, population, number of households, number of nonresidential B&E, number of residential B&E, for the years 2001 and 2004.

Data source: City of Ottawa, and Ottawa Police, where B&E represents a 30% random sample of data from the Ottawa Police Service Records Management System for data collected between 2000-01-01 and 2004-12-31. Area, perimeter, and coordinates are calculated by ArcGIS.
3.2 Nonspatial Data Analysis

We define the following:

\( NR_i \), number of nonresidential B&E in ward \( i \),

\( HH_i \), number of households in ward \( i \),

We discriminate between the two (2) wards in 2001 with zero (0) incidents of nonresidential B&E but different number of households by using raw observed weights:

\[
W_i = \frac{10000(NR_i + 1)}{HH_i}, i = 1, ..., 21.
\]

When the attribute of interest may display geographical variability, the first step in the data analysis is an exploratory mapping.

3.2.1 Stem and leaf plots

We use stem and leaf plots to check for symmetry and peakedness of the distribution of \( W_i \).

3.2.1.1 Results for 2004

For 2004 data from Table 3.1 (previous page) there are nine (9) wards with zero (0) incidents of nonresidential B&E but with different numbers of households. The stem-and-leaf plot of \( \{W_i\} \), shown in Figure 3.2, indicates a slightly skewed distribution.

```
0  |  56778
1  |  0001223588
2  |  015
3  |  1
4  |  2
```

Figure 3.2. Stem-and-leaf plot of \( \{W_i, i = 1, ..., 21\} \). -2004 (The decimal point is at the |.)
3.2.1.2 Results for 2001

Figure 3.3 shows the stem-and-leaf plot of \( \{W_i\} \) for 2001 data. It appears to be moderately symmetric and somewhat flat in distribution.

<table>
<thead>
<tr>
<th>Digit</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>88</td>
</tr>
<tr>
<td>1</td>
<td>05999</td>
</tr>
<tr>
<td>2</td>
<td>134458</td>
</tr>
<tr>
<td>3</td>
<td>1147</td>
</tr>
<tr>
<td>4</td>
<td>0237</td>
</tr>
</tbody>
</table>

Figure 3.3. Stem-and-leaf plot of \( \{W_i, i = 1, \ldots, 21\} \) - 2001
(The decimal point is at the |.)

3.2.2 Choropleth maps

A choropleth map of \( \{W_i, i = 1, \ldots, 21\} \) could be misleading due to large variation between wards with a small number of households and wards with a large number of households (Cressie 1993).

3.2.2.1 Results for 2004

Figure 3.4 is a crude map of weighted nonresidential B&E \( \{W_i\} \) in Ottawa in the year 2004, and the results are shown in Table 3.2.
Figure 3.4. Map of Weighted Nonresidential B&E $\{W_i\}$ in 2004.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Ward Number</th>
<th>Nonresidential B&amp;E Rate per 10000 Household</th>
<th>Weight ((W_i))</th>
<th>Expected ((E_i))</th>
<th>Relative Risk ((RR_i))</th>
<th>Probability Map (\rho_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.613723</td>
<td>1.430294</td>
<td>0</td>
<td>0.239239</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.752332</td>
<td>1.166777</td>
<td>0</td>
<td>0.311369</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>2.065518</td>
<td>2.478622</td>
<td>2.1249</td>
<td>235.3052</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.813537</td>
<td>1.220306</td>
<td>2.157993</td>
<td>92.67871</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1.458789</td>
<td>0.601735</td>
<td>0</td>
<td>0.547861</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1.03434</td>
<td>0.848661</td>
<td>0</td>
<td>0.427988</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.482369</td>
<td>0.964739</td>
<td>1.819775</td>
<td>54.95184</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0.60551</td>
<td>1.21102</td>
<td>1.449693</td>
<td>68.98012</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0.678012</td>
<td>1.294673</td>
<td>0</td>
<td>0.273988</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>1.372809</td>
<td>1.830412</td>
<td>1.918265</td>
<td>156.3913</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0.725058</td>
<td>1.210667</td>
<td>0</td>
<td>0.297998</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>1.675814</td>
<td>2.094767</td>
<td>2.09523</td>
<td>190.9098</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0.545048</td>
<td>1.610507</td>
<td>0</td>
<td>0.199786</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>1.368863</td>
<td>1.825151</td>
<td>1.923795</td>
<td>155.9418</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0.564111</td>
<td>1.128223</td>
<td>1.556083</td>
<td>64.26393</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>0.991965</td>
<td>1.487948</td>
<td>1.769828</td>
<td>113.0053</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0.652103</td>
<td>1.304206</td>
<td>1.346112</td>
<td>74.28801</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>2.598753</td>
<td>3.118503</td>
<td>1.688895</td>
<td>296.0516</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0.95057</td>
<td>0.92345</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2.772003</td>
<td>4.158004</td>
<td>0.633336</td>
<td>315.7884</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0</td>
<td>2.044572</td>
<td>0.429334</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.2. Results of Different Maps for Nonresidential B&E data.-2004

Note that small probabilities relate to extreme areas that are either significantly high risk or significantly low risk.

### 3.2.2.2 Results for 2001

Figure 3.5 is a choropleth map of \(\{W_i, i = 1, \ldots, 21\}\), a crude map of weighted nonresidential B & E \((W_i)\) in Ottawa in the year 2001, and the results are shown in Table 3.3.
Figure 3.5. Map of Weighted Nonresidential B&E ($W$) in 2001.
<table>
<thead>
<tr>
<th>Ward Number</th>
<th>Nonresidential B&amp;E Rate per 10000 Household</th>
<th>Weight</th>
<th>Expected (Ei)</th>
<th>Relative Risk (RRi)</th>
<th>Probability (pi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1.2743724</td>
<td>1.9115586</td>
<td>2.8714531</td>
<td>69.65115</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.0000000</td>
<td>0.7805183</td>
<td>2.3441479</td>
<td>0.000000</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1.9754062</td>
<td>2.4692577</td>
<td>3.7048588</td>
<td>107.96633</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>3.2195750</td>
<td>3.6795143</td>
<td>3.9780256</td>
<td>175.96669</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3.1060724</td>
<td>4.6591086</td>
<td>1.1781118</td>
<td>169.76318</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.1856770</td>
<td>2.3713540</td>
<td>1.5431270</td>
<td>64.80348</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.4789272</td>
<td>0.9578544</td>
<td>3.8203097</td>
<td>26.17589</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1.2475828</td>
<td>1.8713742</td>
<td>2.9331123</td>
<td>68.18696</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>2.7065431</td>
<td>3.3831788</td>
<td>2.7040401</td>
<td>147.92680</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1.0006504</td>
<td>1.5009756</td>
<td>3.6569220</td>
<td>54.69080</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0.0000000</td>
<td>0.7634753</td>
<td>2.3964759</td>
<td>0.000000</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>2.6212320</td>
<td>3.0581040</td>
<td>4.1880694</td>
<td>143.2645</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>2.2545373</td>
<td>2.8181716</td>
<td>3.2461655</td>
<td>123.2223</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>1.3979497</td>
<td>1.8639329</td>
<td>3.9264295</td>
<td>76.40529</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>1.7252286</td>
<td>2.3003048</td>
<td>3.1815788</td>
<td>94.29281</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>1.5499070</td>
<td>2.0664527</td>
<td>3.5414710</td>
<td>84.71056</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>3.3438106</td>
<td>4.0125727</td>
<td>2.7358760</td>
<td>182.75682</td>
</tr>
<tr>
<td>18</td>
<td>7</td>
<td>3.6530634</td>
<td>4.1749295</td>
<td>3.5059758</td>
<td>199.65911</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>1.2134450</td>
<td>2.4268899</td>
<td>1.5078148</td>
<td>66.32114</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1.5422579</td>
<td>3.0845157</td>
<td>1.1863452</td>
<td>84.29250</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>2.1533161</td>
<td>4.3066322</td>
<td>0.8496896</td>
<td>117.69004</td>
</tr>
</tbody>
</table>

Table 3.3. Results of Different Maps for Nonresidential B&E data - 2001.

Note that small probabilities relate to areas that are either significantly high or significantly low risk.

### 3.2.3 Relative Risk Map

The raw observed rate of nonresidential B&E per 10000 households \(\{10000(NR_i)/HH_i\}\) (see Table 3.2 for the 2004 results and Table 3.3 for the 2001 results) could be misleading because such rates may differ widely from ward to ward since the variability of \(NR_i\) is a function of \(HH_i\). We
consider other approaches to try to better highlight the irregular areas at the same time as accounting
for the variability of the underlying number of households in each of the areas. A straightforward
approach is to map some measure of relative risk, which is common in epidemiology or medical
geography. If the observed count in each of the wards is \( \{NR_i, i = 1, \ldots, 21\} \) which take place in the
corresponding region with number of households \( \{HH_i, i = 1, \ldots, 21\} \), and if we suppose that the
overall rate of occurrence is constant for all wards and that the wards are independent, then the
realistic assumption is that the counts \( NR_i \) are observations on independent Poisson random
variables with the expected values \( E_i \) (Bailey and Gatrell 1995). Then \( E_i \) is given by:

\[
E_i = HH_i \times \hat{p},
\]

where \( \hat{p} = \sum_{i=1}^{21} NR_i / \sum_{i=1}^{21} HH_i \). For example, for 2004, \( \hat{p} = 8.778039e - 05 \).

Instead of mapping the raw rate for each ward, we map the relative risk \( RR_i \), where

\[
RR_i = NR_i / E_i \times 100.
\]

### 3.2.3.1 Results for 2004

The results for 2004 are listed in Table 3.2 and the resulting map is shown below in Figure
3.6. Since the observed count of each ward is divided by its expected value and multiplied by 100,
the value of 200 suggests that the observed count \( NR_i \) in ward \( i \) is double \( E_i \).
In our case, although the map clearly shows the high and low relative risk of the wards, ward number 20 has 2 recorded NRs with 0.63 expected, yielding a relative risk of 315.8, while ward number 7 has 1 recorded NRs with 0.48 expected, but yielding a relative risk of 54.95. Adding only a single observed case has increased the estimate significantly. This is because the number of households in wards is highly variable (4891 < HH_i < 24584). This means that E_i varies widely from ward to ward and this has not been taken into account. To solve this problem it is suggested that a probability map that standardizes the rates onto a probability scale be used to facilitate comparisons.

3.2.3.2 Results for 2001

The relative risk map of \( RR_i = \frac{NR_i}{E_i} \times 100 \) for the 2001 nonresidential B&E yields the map shown in Figure 3.7. and the results are listed in Table 3.3.
Examining the map in Figure 3.7 and the data in Table 3.3 we see a strange result. Ward number 5 (West Carleton area) was at high risk. While it had 1.178 as its expected rate (the second smallest expected rate), its relative risk is 169.76 for 2001, the fourth largest, This unusual result is due to the fact that $H_i$ and thus $E_i$ varies broadly from ward to ward and this has not been taken into consideration. To solve this problem it is suggested that a probability map that standardizes the rates onto a probability scale be used to facilitate comparisons.

### 3.2.4 Probability Map

Under the assumption that the count in each ward $i$ is Poisson with mean values

\[ \{ \lambda_i, i = 1, \ldots, 21 \}, \text{ and } \lambda_1/HH_1 = \cdots = \lambda_{21}/HH_{21} = p, \]

where $E_i$ was determined previously, an index of deviation from equal $\{ \lambda_i/HH_i \}$ can be defined as:
Small values of $\rho_i (< 0.2)$ indicate that the $NR_i$ rate for a ward is unusually high or low.

### 3.2.4.1 Results for 2004

Computing the Poisson probabilities for 2004 yields the probability map shown in Figure 3.8. We note that now the extreme values in the previous relative risk map are replaced with more reasonable evaluations of how extreme these areas are on a probability scale.

The map shows the clusters of wards.

### 3.2.4.2 Results for 2001

Using Poisson probabilities yields the probability map shown in Figure 3.9; again the extreme values in the relative risk map have been replaced now with more reasonable evaluations of
how extreme these areas are on the probability scale.

The map shows some clustering of wards. If we compare the 2004 and 2001 maps we note that the areas at high or low risk were further away from the downtown in 2001 while in 2004 these moved closer to the downtown (as we note from the map previously shown in Figure 3.8).

3.3 Exploratory Spatial Data Analysis

3.3.1 Spatial smoothing

When the data are on a regular grid and come from an underlying continuous population, (Cressie, 1993) gives a spatial data analysis that is resistant to outlying data values. For a two-dimensional \( r \times s \) grid with data \( \{Z_{kl}; k = 1, \ldots, r; l = 1, \ldots, s\} \) a first decomposition can be

\[
Z_{kl} = \mu + \alpha_k + \beta_l + \delta_{kl},
\]

where \( \{\alpha_k\} \) and \( \{\beta_l\} \) are fixed row and column effects respectively, \( \mu \) is the fixed overall mean effect, and \( \{\delta_{kl}\} \) represents the random errors. The resistant method proposed is a median polish. For
analysis of the nonresidential B & E incidents we should note that Ottawa wards are irregularly
spaced and the data represent regions; the location of the value $W_i$ is given by the location
coordinates $(x_i, y_i)$ with a fixed origin. We should attempt the decomposition $W_i = \mu_i + \delta_i$, where $\mu_i$
is a function of the location $(x_i, y_i)$. Since our data are counts and we transformed the data, we can
consider a more general decomposition

$$g(W_i) = \mu_i + v_i.$$ 

3.3.2 Median Polish

For spatial model

$$Z(s) = \mu(s) + \delta(s), s \in D$$

the mean $\mu(s)$ is now a function of the spatial location and spatial dependence is captured in $\delta(s)$. A
low resolution map is produced by overlaying a grid on the original data and shifting data points to
the nearest grid node. Let the rows of the grid be indexed as $k = 1, \ldots, p$ and the columns of the grid
be indexed as $l = 1, \ldots, q$. Then the gridded data is $\{Y_{kl}; k = 1, \ldots, p; l = 1, \ldots, q\}$. Note that the $Y$s at
the $(k, l)^{th}$ node are not always one (it could be more). Moreover, if a whole row (column) of the grid
has no observations, then that row column is ignored. Now that $\{Y_{kl}\}$ have the structure of a
two-way table, we can do the median polish. The algorithm for median polish (Cressie, 1993) is as
follows:

For $i = 1, 3, 5, \ldots$, we define

$$Y^{(i)}_{kl} = Y^{(i-1)}_{kl} - \text{med}\{Y^{(i-1)}_{kl}, l = 1, \ldots, q\}, \quad k = 1, \ldots, p + 1; \quad l = 1, \ldots, q,$$

$$Y^{(i)}_{k,q+1} = Y^{(i-1)}_{k,q+1} + \text{med}\{Y^{(i-1)}_{kl}, l = 1, \ldots, q\}, \quad k = 1, \ldots, p + 1$$
and for \( i = 2, 4, 6, \ldots \), we define

\[
Y_{ki}^{(l)} = Y_{ki}^{(l-1)} - \text{med}\{Y_{ki}^{(l-1)}, k = 1, \ldots, p\}, \quad k = 1, \ldots, p; \quad l = 1, \ldots, q + 1,
\]

\[
Y_{p+1,l}^{(l)} = Y_{p+1,l}^{(l-1)} + \text{med}\{Y_{ki}^{(l-1)}, k = 1, \ldots, p\}, \quad l = 1, \ldots, q + 1
\]

where \( \text{med}\{y_1, \ldots, y_n\} \) is the median of \( y_1, \ldots, y_n \). To start the algorithm, assume

\[
Y_{ki}^{(0)} = \begin{cases} 
Y_{ki}, & k = 1, \ldots, p; \quad l = 1, \ldots, q \\
0 & \text{elsewhere.}
\end{cases}
\]

The median polish proceeds by repeating extraction of row and column medians until convergence. Assuming convergence, the estimated effects are

\[
\tilde{a} = Y_{p+1,q+1}^{(\infty)},
\]

\[
\tilde{r}_k = Y_{k,q+1}^{(\infty)}, \quad k = 1, \ldots, p,
\]

\[
\tilde{c}_l = Y_{p+1,l}^{(\infty)}, \quad l = 1, \ldots, q,
\]

with the property that

\[
Y = \tilde{a} + \tilde{r}_k + \tilde{c}_l + Y_{ki}^{(\infty)}, \quad k = 1, \ldots, p; \quad l = 1, \ldots, q
\]

where \( \{Y_{ki}^{(\infty)}, k = 1, \ldots, p; l = 1, \ldots, q\} \) is \( p \times q \) table of residuals , and \( \{\tilde{r}_k\}, \{\tilde{c}_l\} \) and \( \{\tilde{a}\} \) are row effects, column effects and overall effect, respectively.

### 3.3.3 Transformations

Dealing with the number of nonresidential households that were involved in B & E, \( NR_i \) can be considered as the outcome of a binomial experiment. Then the number \( NR_i \) out of the total number of households \( HH_i \) in ward \( i \) is the sum of \( HH_i \) spatially dependent Bernoulli random
variables. Independence between the Bernoulli experiments means $NR_i \sim Bin(HH_i, p_i)$, $i = 1, \ldots, 21$, where it is of concern to determine whether the average, $E(NR_i/HH_i) = p_i$, is heterogenous or not. Note that the parameter of interest is also contained in $Var(NR_i/HH_i) = p_i(1 - p_i)/HH_i$. If there is local spatial dependence between the Bernoulli experiments, $E(NR_i/HH_i) = p_i$ is still the parameter of interest but now $Var(NR_i/HH_i) = \{(p_i(1 - p_i)/HH_i)\{1 + k_i\}$, where $k_i$ accounts for the local spatial covariance. It is reasonable then to try to find a transformation of the data which gets rid of the dependence of the variance on the mean. Following Cressie (1993), the following transformations are considered:

$$W_i = 10000(NR_i + 1)/HH_i$$

$$Z_i = (10000(NR_i)/HH_i)^{1/2} + (10000(NR_i + 1)/HH_i)^{1/2}$$

$$V_i = log(W_i).$$

To compare the stability of the variance associated with each of the above formulae, we partitioned the group of 21 wards into 4 similar subsets with roughly $(1/HH_i)$ for each subset in order to eliminate the effect of $HH_i$. We consider $Z_i$ the Freeman-Tukey square root transformation because it shows more stability than the usual square-root transformation $(W_i)^{1/2}$. The general purpose of using a transformation is to interpret symmetry, homoskedasticity, and lack of interaction in the two-way data decomposition as a property of additivity at all scales. With the Freeman-Tukey transformation

$$Z_i = (10000(NR_i)/HH_i)^{1/2} + (10000(NR_i + 1)/HH_i)^{1/2}, \quad i = 1, \ldots, 21,$$

the mean-variance relation is removed and $Var(Z_i) \approx \sigma^2/HH_i$. $Z_i$ is considered as the square-root transformation of the data. Hence, we could write
\[ Z_i = a + r_{k(i)} + c_{l(i)} + HH_i^{1/2} \delta_i \]

where \( r_{k(i)} \) is the \( k \)th row effect and \( c_{l(i)} \) is the \( l \)th column effect and \( a \) is the overall mean effect, and \( \{HH_i^{1/2} \delta_i\} \) represents the transformed random errors.

### 3.3.3.1 Results for 2004

Figure 3.10 shows the stem-and-leaf plot of \( \{Z_i, i = 1, \ldots, 21\} \), the Freeman-Tukey transformed NR\(_i\) rates, for 2004 data.

\[
\begin{array}{c|c}
0 & 78899 \\
1 & 00247899 \\
2 & 02557 \\
3 & 047 \\
\end{array}
\]

Figure 3.10. Stem-and-leaf plot of \( \{Z_i, i = 1, \ldots, 21\} \) for 2004 data. (The decimal point is at the |).

### 3.3.3.2 Results for 2001

Figure 3.11 shows the stem-and-leaf plot of \( \{Z_i, i = 1, \ldots, 21\} \), the Freeman-Tukey transformed NR\(_i\) rates, for 2001 data.

\[
\begin{array}{c|c}
0 & 99 \\
1 & 7 \\
2 & 25556778 \\
3 & 002455789 \\
4 & 0 \\
\end{array}
\]

Figure 3.11. Stem-and-leaf plot of \( \{Z_i, i = 1, \ldots, 21\} \) - 2001. (The decimal point is at the |).

We note from Figures 3.10 and 3.11 that this transformation symmetrizes the data and stabilizes variances, which is what we are seeking to do.
3.3.4 Trend Removal

If we look at the contour and perspective plots of the transformed data in Figures 3.12 and 3.13, we see that in 2004 there are higher values in the south area of Ottawa and lower values in the north area. Clearly there is a trend effect in the data, what is called large-scale variation.

Figure 3.12. Contour map of $Z_i$ for 2004
Now considering that the data is made up of 2 components, i) large-scale variation which consist of spatial trend and ii) small-scale variation which consists of spatially dependent noise, each part could then be isolated and interpreted. As mentioned before, the idea of transforming to achieve additivity over all scales and the two-dimensional (spatial) coordinate representation of the ward locations lead us obviously to a two-way additive decomposition for the spatial trend. Overlaying a rectangular grid with 60×50 kilometers grid spacing onto Ottawa yields 12 east-west transects and 10 north-south transects and 12×10 nodes of the grid. A ward is associated with the node that is closest to its centroid. Note that some nodes may have no wards due to the irregularity of the regions.

Figure 3.13. Perspective plot of $Z_i$ for 2004
3.3.4.1 Results for 2004

Figure 3.14 shows for 2004 the Freeman-Tukey transformed NR\textsubscript{i} rates (Z\textsubscript{i}) associated with each of the nodes. Rows or columns without observations are not listed.

\[\begin{array}{cccccccc}
1, & 0.7659070 \\
2, & 2.4097458 \\
3, & 0.7737537 \\
4, & 1.8680068 \\
5, & 0.8104409 & 3.588701 & 1.8680068 & 1.9076498 \\
6, & 0.7737537 & 2.9489711 & 3.4330827 \\
7, & 2.4860976 & 1.9076498 & 0.8104409 \\
8, & 3.7024079 & 0.8104409 \\
9, & 2.4860976 & 2.8876883 & 0.7659070 \\
10, & 0.7659070 \\
11, & 0.7901744 \\
\end{array}\]

Figure 3.14. Low resolution map of the Freeman-Tukey transformed NR\textsubscript{i} rates (Z\textsubscript{i}) associated with each of the 12 \times 10 nodes of grid overlaid on the Ottawa ward map, 2004

We can now carry out a median polish on the matrix of the nodes, weighting each ward rate proportional to its \((HH_{i})^{1/2}\), in order to eliminate the row, column and overall effects. We obtain a set of stationary residuals that capture the large-scale variation. Wards 9 and 11 (Knoxdale-Merivale and Beacon Hill-Cyrville) have unacceptably high residuals so they are removed from the calculation. The median-polish gives the decomposition \[Z_i = \tilde{\alpha} + \tilde{\gamma}_{(i)} + \tilde{\epsilon}_{(i)} + R(x_i, y_i),\] where \[R(x_i, y_i) = HH_{i}^{-1/2} \delta(x_i, y_i).\] Table 3.4 gives the estimates derived from the median-polish for 2004.
\[ \tilde{a} = 1.887828 \quad \tilde{r}_{10} = -1.12192133 \]
\[ \tilde{r}_1 = -1.12192133 \quad \tilde{r}_{11} = -1.09765386 \]
\[ \tilde{r}_2 = 0.52191751 \quad \tilde{c}_1 = -1.077387 \]
\[ \tilde{r}_3 = -1.11407455 \quad \tilde{c}_2 = 0.000000 \]
\[ \tilde{r}_4 = -0.01982153 \quad \tilde{c}_3 = 1.700873 \]
\[ \tilde{r}_5 = 0.0000000 \quad \tilde{c}_4 = 0.000000 \]
\[ \tilde{r}_6 = 1.85180309 \quad \tilde{c}_5 = 0.0000000 \]
\[ \tilde{r}_7 = 0.01982153 \quad \tilde{c}_6 = 0.0000000 \]
\[ \tilde{r}_8 = 1.81457959 \quad \tilde{c}_7 = -1.097209 \]
\[ \tilde{r}_9 = 0.59826935 \]

Table 3.4. Estimates of overall Effect (\( \tilde{a} \)), Row Effects (\( r_1, \ldots, r_{11} \)), and Column Effects (\( c_1, \ldots, c_8 \))

Using these estimates we produce a map of \( \{ \tilde{\mu}_i = \tilde{a} + \tilde{r}_{ki} + \tilde{c}_{li}, \quad i = 1, \ldots, 21, \} \) which appears below as Figure 3.15. This is a smoothed map of the region that has accounted for unequal variation.

Figure 3.15. A Choropleth map of Weighted Median-polish fit for Ottawa’s Nonresidential B&E in 2004

The clustering observable in Figure 3.15 is similar to that seen previously in the probability
map.

3.3.4.2 Results for 2001

In order to capture the large-scale variation and small-scale variation, we carry out a median polish on the matrix of the nodes. (Overlaying a rectangular grid with 60×50 kilometers grid spacing onto Ottawa yields 12 east-west transects and 10 north-south transects, and 12×10 nodes of the grid, and the ward is associated with the node that is closest to its centroid.) Weighting each ward rate proportional to its \((HH_i)^{1/2}\) in order to eliminate the row, column and overall effects, we get a set of stationary residuals. The median-polish gives the decomposition

\[
Z_i = \tilde{a} + \tilde{r}_k + \tilde{c}_i + R(x_i, y_i),
\]

where \(R(x_i, y_i) = HH_i^{-1/2} \delta(x_i, y_i)\). Table 3.5 gives the estimates derived from the median-polish of the transformed data for 2001.

<table>
<thead>
<tr>
<th>(\tilde{a})</th>
<th>(\tilde{r}_{10})</th>
<th>(\tilde{r}_{11})</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.536403</td>
<td>0.0000000</td>
<td>-0.8502870</td>
</tr>
<tr>
<td>(\tilde{r}_1)</td>
<td>0.0000000</td>
<td>-0.7995163</td>
</tr>
<tr>
<td>1.8766789</td>
<td>-0.6291103</td>
<td>0.6631598</td>
</tr>
<tr>
<td>(\tilde{r}_2)</td>
<td>-0.6291698</td>
<td>0.3</td>
</tr>
<tr>
<td>(\tilde{r}_3)</td>
<td>0.153999</td>
<td>0.2661706</td>
</tr>
<tr>
<td>(\tilde{r}_4)</td>
<td>0.8102029</td>
<td>-0.5811783</td>
</tr>
<tr>
<td>(\tilde{r}_5)</td>
<td>0.9864661</td>
<td>-0.5841164</td>
</tr>
<tr>
<td>(\tilde{r}_6)</td>
<td>0.2483874</td>
<td>0.0000000</td>
</tr>
<tr>
<td>(\tilde{r}_7)</td>
<td>-0.5841164</td>
<td></td>
</tr>
<tr>
<td>(\tilde{r}_8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\tilde{r}_9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5. Estimates of overall Effect (\(\tilde{a}\)), Row Effects (\(r_1, ..., r_{11}\)), and Column Effects (\(c_1, ..., c_8\)) -2001

Using the fitted values, we map \(\{\tilde{\mu}_i = \tilde{a} + \tilde{r}_k + \tilde{c}_i; \quad i = 1, ..., 21,\}\) which results in the smoothed map in Figure 3.16; this map accounts for unequal variation.
In Figure 3.16 we observe clustering, where Ottawa rural areas form one cluster and the urban area has two clusters.

### 3.3.5 Median-Polish Residuals

The weighted-median polish residuals are \( R(x_i, y_i) = Z_i - \bar{a} - \tilde{r}_{k(i)} - \tilde{c}_{k(i)} \) and estimate the small-scale variation. Spatial analysis is therefore performed on these in order to address the small-scale variation. Stationary error models could be fit to the detrended and standardized data \( \{ \sqrt{HH_i} R(x_i, y_i), i = 1, \ldots, 21 \} \).

#### 3.3.5.1 Results for 2004

By examining the stem-and-leaf plot of the standardized (weighted) residuals \( \sqrt{HH_i} R(x_i, y_i) \) we identify the two (2) wards that are outliers - Wards 9 and 11 (Knoxdale-Merivale and Beacon Hill-Cyrville) as mentioned previously. For 2004, Figure 3.17 shows the stem-and-leaf plot for the
weighted median-polish residuals \( \{ \sqrt{HH_i} R(x_i,y_i), \ i = 1,\ldots,21 \} \) and Figure 3.18 shows the histogram for the weighted median-polish residuals \( \{ \sqrt{HH_i} R(x_i,y_i), \ i = 1,\ldots,21 \} \).

-2 | 2
-1 |
0 | 6600
0 | 0000000000000046
1 |
2 | 2

Figure 3.17. Stem-and-leaf of weighted median-polish residuals \( \{ \sqrt{HH_i} R(x_i,y_i), \ i = 1,\ldots,21 \} \) - 2004
(The decimal point is at the |.)

![Histogram of weighted median-polish residuals](image)

Figure 3.18. Histogram of weighted median-polish residuals \( \{ \sqrt{HH_i} R(x_i,y_i), i = 1,\ldots,21 \} \) - 2004.

3.3.5.2 Results for 2001

To deal with small-scale variation spatial analysis is performed on the weighted-median polish
residuals. Figure 3.19 shows the plot of stem-and-leaf plot of these weighted median-polish residuals \( \{ \sqrt{HH_i} R(x_i, y_i), \ i = 1, \ldots, 21 \} \). By examining the stem-and-leaf plot we note that are no outliers in the 2001 data.

![Stem-and-leaf plot](image)

Figure 3.19. Stem-and-leaf of weighted median-polish residuals \( \{ \sqrt{HH_i} R(x_i, y_i), \ i = 1, \ldots, 21 \} \) - 2001.

(The decimal point is 2 digit(s) to the right of the |.)

### 3.3.6 The Variogram

The variogram serves as a measure of spatial dependence among the data \( \{ Z(s_i), i = 1, \ldots, n \} \). Let \( \{ Z(s), s \in D \subset \mathbb{R}^d \} \) be the spatial process defined on a domain \( D \) of the d-dimensional space \( \mathbb{R}^d \) and suppose that differences of variables lagged \( h \) apart vary in a way that depends only on \( h \).

That is,

\[
\text{var}(Z(s + h) - Z(s)) = 2\gamma(h) \quad \text{for all } s, s + h \in D. \tag{3.1}
\]

Usually the spatial index \( s \) is two- or three-dimensional (for our data, \( d = 2 \)). The quantity \( 2\gamma(\bullet) \), which is only a function of the difference between the spatial locations \( s \) and \( s + h \), is called the variogram. The variogram must satisfy the conditional negative semi-definiteness condition,

\[
\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \ 2\gamma(s_i - s_j) \leq 0, \quad \text{for any finite number of spatial locations } \{s_i, i = 1, \ldots, k\}, \text{and real numbers } \{a_i, i = l, \ldots, k\} \text{ satisfying } \sum_{i=1}^{k} a_i = 0. \text{ When } 2\gamma(h) = 2\gamma^*(||h||), \text{ the variogram is said to be isotropic. Replacing (3.1) with the stronger assumption} \]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[
\text{cov}(Z(s + h), Z(s)) = C(h) \quad \text{for all } s, s + h \in D, \tag{3.2}
\]

and taking the mean function to be constant, i.e.

\[
E(Z(s)) = \mu \quad \text{for all } s \in D, \tag{3.3}
\]

will define the class of second-order stationary processes in \( D \), with covariance function \( C(\bullet) \).

Conditions (3.1) and (3.3), which define the class of intrinsically stationary processes, contain the class of second-order stationary processes. Assuming only (3.2), \( \gamma(h) = C(0) - C(h) \); i.e., the semivariogram (half of the variogram) is related very simply to the covariance function.

Suppose observations are on an intrinsically stationary process, a process that satisfies (3.1) and (3.3), taken at \( n \) spatial locations \( \{s_i : i = 1, \ldots, n\} \). Because of (3.3),

\[
\text{var}(Z(s + h) - Z(s)) = E((Z(s + h) - Z(s))^2).
\]

The method-of-moments estimator of the variogram \( 2\gamma(h) \) is

\[
\hat{\gamma}(h) = \sum_{N(h)} (Z(s_i) - Z(s_j))^2 / |N(h)|, \quad h \in \mathbb{R}^d, \tag{3.4}
\]

where the average in (3.4) is taken over \( N(h) = \{(s_i, s_j), s_i - s_j = h\} \) and \(|N(h)|\) is the number of distinct elements in \( N(h) \). For irregularly spaced data, \( N(h) \) is usually modified to

\( \{(s_i, s_j), s_i - s_j \in T(h)\} \), where \( T(h) \) is a tolerance region of \( \mathbb{R}^d \) surrounding \( h \).

### 3.3.6.1 Results for 2004

For our 2004 data, an estimate of the variogram (which is a measure of spatial dependence) is computed based on the standardized residuals \( \{\sqrt{HH_i} R(x_i, y_i), i = 1, \ldots, 21, i \neq 9, 11\} \). An isotropic variogram estimator \( 2\gamma^*(h) \) is calculated and plotted as a function of distance \( h = ||h|| \) as shown in Figure 3.19. Joining these points and observing the slopes of these lines, we see positive spatial
dependence at distances up to 10 kilometers (corresponding mainly to the urban areas of Ottawa). Between 10 and 20 kilometers we see observe negative spatial dependence and between 20 and 40 kilometers the spatial dependence is essentially zero. At distances between 40 and 50 kilometers we observe negative spatial dependence and at a distance of approximately 40 kilometers we begin to see the separation of a cluster of high nonresidential B & E wards in the rural area from those wards in the urban area of Ottawa.

![Figure 3.20 Variogram estimator $2\gamma^*(h)$ based on standardised residuals $\sqrt{HH_i}R(x_i,y_i), \ i = 1, \ldots, 21, \ i \neq 9, 11$] for 2004 data

3.3.6.2 Results for 2001

Again, for 2001, an estimate of the variogram is computed based on the standardized residuals $\{\sqrt{HH_i}R(x_i,y_i), \ i = 1, \ldots, 21\}$. An isotropic variogram estimator $2\gamma^*(h)$ is calculated and plotted as a function of distance $h = ||h||$ as shown in Figure 3.20. Here we observe positive spatial dependence at distances up to 10 kilometers (corresponding to the urban areas of Ottawa). From 10 to 20 kilometers the spatial dependence is negative. These results are similar to 2004. However, in
2001, spatial dependence is close to zero for distances between 20 and 30 kilometers and at a distance of between 30 and 50 kilometers, we again see negative spatial dependence and the separation of a cluster of wards in the rural area from those wards in the urban area of Ottawa.

Figure 3.21. Variogram estimator $2\gamma^*(h)$ based on standardised residuals $\sqrt{HH_i}R(x_i,y_i), \ i = 1,...,21, \ i \neq 9,11$ - 2001.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Chapter 4: Spatial Regression Analysis

We now wish to test the hypothesis of no spatial dependence due to an omitted spatial lag or due to spatially autoregressive errors. Whether or not we find significant spatial dependence, we will also wish to estimate the parameters of our best-fitting model. We will begin by considering spatial autocorrelation.

4.1 Spatial autocorrelation

There are two forms of spatial autocorrelation: a) positive spatial autocorrelation where similar values for a random variable tend to cluster in space and b) negative spatial autocorrelation where locations tend to be surrounded by neighbours with very dissimilar values. Negative spatial autocorrelation implies a checkerboard pattern of values and does not always have a meaningful interpretation. Positive spatial autocorrelation implies that a sample contains less information than an uncorrelated counterpart. Correlation is expressed in terms of the following moment condition:

\[ \text{Cov}(y_i,y_j) = E(y_iy_j) - E(y_i).E(y_j) \neq 0 \quad \text{for} \quad i \neq j \]  

(4.1)

where \( y_i \) and \( y_j \) are observations on a random variable at locations \( i \) and \( j \) in the 2-dimensional space. It becomes spatial when the pairs of \( i,j \) locations for which the correlation is nonzero have a meaningful interpretation in terms of spatial structure. In order to estimate the potentially \( N \times N \) covariance terms or correlations directly from the data, we need to impose sufficient constraints on the \( N \times N \) covariance matrix such that a sufficient, finite number of parameters can be estimated. If the \( N \) observations are considered as fixed in space there is not enough information in the data to estimate the \( N \times N \) interactions. There are two main approaches to impose constraints on the interaction. The first one is to express the strength of correlation of all pairs of locations as a

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
continuous function of the distance between them in a so-called variogram as we have done in the previous chapter. The second one is to exploit the neighbourhood set for each point of data in order to specify a spatial stochastic process in an approach called the lattice perspective.

4.2 Spatial weights

A spatial weights matrix $W$ is an $N \times N$ positive symmetric matrix where the elements of $W$ are $w_{ij} = 1$ when $i$ and $j$ are neighbours, and $w_{ij} = 0$ otherwise, and its diagonal elements are set to zero. For simplicity of explanation, the weights matrix is standardized so that the elements of a row sum to one. The elements of a row-standardized weights matrix therefore equal $w_{st}^r = w_{ij} / \sum_j w_{ij}$. This guarantees that all weights are between 0 and 1 and helps in interpreting operations with the weights matrix as an averaging of neighbouring values. It also guarantees that the spatial parameters in the spatial stochastic process are similar between models. In order to impose valid constraints in a maximum likelihood estimation framework, the spatial autoregressive parameters must be constrained to lie in the interval $l/\omega_{\min}$ to $l/\omega_{\max}$, where $\omega_{\min}$ and $\omega_{\max}$ are respectively the smallest and largest eigenvalues of the matrix $W$. For a row-standardized weights matrix, the largest eigenvalue is always $+1$, which facilitates interpretation of the autoregressive coefficient as a correlation. Although, the original matrix is symmetric, there is a side effect to row standardization, i.e., the resulting matrix is likely to become asymmetric since $\sum_j w_{ij} \neq \sum_i w_{ji}$.

There are several ways to specify which elements are nonzero in the spatial weights matrix. The traditional approach relies on the spatial arrangement of the observations, designating areal units as neighbours when they have a border in common (first-order contiguity) or are within a given
distance of each other so that \( w_{ij} = 1 \) for \( d_{ij} \leq \delta \), where \( d_{ij} \) is the distance between units \( i \) and \( j \), and \( \delta \) is a distance cutoff value (called distance-based contiguity). In general, weights may be specified to express any measure of "potential interaction" between units \( i \) and \( j \). Typically, the parameters of the distance function are set \textit{a priori}, like \( \alpha = 2 \), to reflect a gravity function, and they are not estimated jointly with the other coefficients in the model. When they are estimated jointly, the resulting specification will be highly nonlinear. It is important to keep in mind that, irrespective of how the spatial weights are specified, the resulting spatial process must satisfy the necessary regularity conditions so that asymptotics may be invoked to obtain the properties of the estimators and test statistics.

### 4.3 Spatial Lag Operator

For an irregular spatial structure, the concept of a spatial lag operator is used, which consists of a weighted average of the values at neighbouring locations. The weights are fixed and exogenous, analogous to a distributed lag in time series. A spatial lag operator \( Wy \) is the product of a spatial weights matrix \( W \) with the vector of observations on a random variable \( y \). Each element of the resulting spatially lagged variable equals \( \sum_j w_{ij} y_j \), which is a weighted average of the \( y \) values in the neighbour set \( S_i \), since \( w_{ij} = 0 \) for \( j \not\in S_i \). Row standardization of the spatial weights matrix ensures that a spatial lag operation smooths neighbouring values, since the positive weights sum to one.

### 4.4 Spatial Lag Dependence

Spatial lag dependence in a regression model is analogous to the insertion of an autoregressive term \( (y_{t-1}) \) for the dependent variable in a time-series framework. This is referred to as a mixed regressive spatial autoregressive model.
\[ y = \rho Wy + \beta X + \epsilon \]  

(4.2)

where \( y \) is a \( N \times 1 \) vector of observations on the dependent variable, \( Wy \) is the corresponding spatially lagged dependent variable for weight matrix \( W \), \( X \) is a \( N \times K \) matrix of observations on the explanatory variables, \( \epsilon \) is a \( N \times 1 \) vector of error terms, \( \rho \) is the spatial autoregressive parameter, and \( \beta \) is a \( K \times 1 \) vector of regression coefficients. The presence of the spatial lag term \( Wy \) on the right side of the above model results in a nonzero correlation with the error term, similar to the existence of an endogenous variable, but different from a lagged dependent variable in the time-series case. (In the latter model, \( y_{t-1} \) is uncorrelated with \( \epsilon_t \) in the absence of correlation in the errors; in contrast, \( (Wy)_t \) is always correlated with \( \epsilon_t \), irrespective of the correlation structure of the errors.) Moreover, the spatial lag for a given observation \( i \) is not only correlated with the error term at \( i \), but also with the error terms at all other locations. Therefore, unlike what holds in the time-series case, an ordinary least-squares (OLS) estimator will not be consistent for this specification. This can be seen from a slight reformulation of the model:

\[ y = (I - \rho W)^{-1}X\beta + (I - \rho W)^{-1}\epsilon. \]

The matrix inverse \( (I - \rho W)^{-1} \) is a full matrix, not triangular as in the time-series case where dependence is only one-directional, yielding an infinite series that involves error terms at all locations, \( (I + \rho W + \rho^2 W^2 + \rho^3 W^3 + \ldots)\epsilon \). It follows that \( (Wy)_i \) contains the element \( \epsilon_i \) as well as other \( \epsilon_j, j \neq i \). Thus,

\[ E[(Wy)_i, \epsilon_i] = E[(I - \rho W)^{-1}\epsilon, \epsilon_i] \neq 0. \]

The spatial dynamics embedded in the structure of the spatial process model (4.2) determine
the form of the covariance between the observations at different locations (i.e., the spatial autocorrelation). For the mixed regressive spatial autoregressive model, this can be seen to equal

\[(I - \rho W)^{-1}\Omega(I - \rho W')^{-1},\]  

where \(\Omega\) is the variance matrix for the error term \(\varepsilon\) (note that for a row-standardized spatial weights matrix, \(W = W'\)). Without loss of generality, the latter can be assumed to be diagonal and homoskedastic (or, \(\Omega = \sigma^2 I\)) and hence

\[Var[y] = \sigma^2 (I - \rho W)^{-1}(I - \rho W')^{-1}\]

The resulting variance matrix is full, implying that each location is correlated with every other location, but in a fashion that decays with the order of contiguity (the powers of \(W\) in the series expansion of \((I - \rho W)^{-1}\).

The implication of this particular variance structure is that the simultaneity embedded in the \(Wy\) term must be explicitly accounted for, either in a maximum likelihood estimation framework, or by using a proper set of instrumental variables. When a spatially lagged dependent variable is ignored in a model specification, but present in the underlying data generating process, the resulting specification error is of the omitted variable type. This implies that OLS estimates in the nonspatial model (i.e., the "standard" approach) will be biased and inconsistent.

Two situations can be distinguished in order to interpret a significant spatial autoregressive coefficient \(\rho\). First, a significant spatial lag term indicates substantive spatial dependence. Specifically, it measures the extent of spatial overflow or diffusion. This interpretation is valid when the items under consideration match the spatial unit of observation and the spillover is the result of a theoretical model. Second, the spatial lag model may be used to deal with spatial autocorrelation that results from a mismatch between the spatial scale of the phenomenon under study and the spatial
scale at which it is measured, and the expected positive spatial autocorrelation will in fact result in the sample containing less information than a truly "independent" sample of observations. The inclusion of a spatially lagged dependent variable in the model specification is a way to correct for this loss of information. It permits the proper interpretation of the significance of exogenous variables in the model after the spatial effects have been filtered out.

More formally, the spatial lag model may be re-expressed as

\[(I - \rho W)y = \lambda \beta + \epsilon\]  \hspace{1cm} (4.3)

where \((I - \rho W)y\) is a spatially filtered dependent variable, i.e., one with the effect of spatial autocorrelation taken out. This is roughly similar to first-differencing of the dependent variable in time series, except that a value of \(\rho = 1\) is not in the allowable parameter space for (4.1) and thus \(\rho\) must be estimated explicitly.

### 4.5 Spatial Error Dependence

Another way to introduce spatial autocorrelation in a regression model is to specify a spatial process for the disturbance terms. The most common specification is a spatial autoregressive process in the error terms:

\[y = \lambda \beta + \epsilon\]  \hspace{1cm} (4.4)

i.e., a linear regression with error vector \(\epsilon\), and

\[\epsilon = \lambda W \epsilon + \xi\]  \hspace{1cm} (4.5)

where \(\lambda\) is the spatial autoregressive coefficient for the error lag \(W \epsilon\) (to distinguish the notation from the spatial autoregressive coefficient \(\rho\) in a spatial lag model), and \(\xi\) is an uncorrelated and (without

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
loss of generality) homoskedastic error term. Alternatively this may be expressed as

\[ y = \lambda \beta + (I - \lambda W)^{-1}\xi. \]  (4.6)

The resulting error covariance will be nonspherical and efficient estimators are obtained by taking advantage of the particular structure of the error covariance implied by the spatial process. Different spatial processes lead to different error covariances, with varying implications about the range and extent of spatial interaction in the model. With the model as in (4.6) the error covariance is

\[ E[\epsilon e'] = \sigma^2 (I - \lambda W)^{-1}(I - \lambda W)^{-1} = \sigma^2 [(I - \lambda W)^T(I - \lambda W)]^{-1}, \]  (4.7)

which is a structure identical to that for the dependent variable in the spatial lag model. Therefore, a spatial autoregressive error process leads to a nonzero error covariance between every pair of observations, but decreasing in magnitude with the order of contiguity. Furthermore, the complex structure in the inverse matrices in (4.7) yields nonconstant diagonal elements in the error covariance matrix, thus inducing heteroskedasticity in \( \epsilon \), irrespective of the heteroskedasticity of \( \xi \).

We have a much simpler situation for the case of autocorrelation in the time-series context, where the model is written as \( \epsilon_t = \lambda \epsilon_{t-1} + \xi_t \). Therefore, this is a special case of (4.5) with

\[ W = W^T = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \]

where each observation is connected to only its immediate past value. For this case,

\[ \text{Var}(\epsilon_t) = \sigma^2(I - \lambda^2) \] for all \( t \). so autocorrelation does not induce heteroskedasticity.

There is a great degree of similarity between a spatial lag dependence model and a spatial
error dependence model, as suggested by the error covariance structure. After premultiplying both sides of (4.6) by \((I - \lambda W)\) and moving the spatial lag term to the right side, a spatial Durbin model results in

\[
y = \lambda Wy + X\beta - \lambda WX\beta + \xi. \tag{4.8}
\]

This model has a spatial lag structure (but with spatial autoregressive parameter \(\lambda\) from (4.5)) with a well-behaved error term \(\xi\). However, the equivalence between (4.4), (4.5) and (4.8) imposes a set of nonlinear common factor constraints on the coefficients. For (4.8) to be a proper spatial error model, the coefficients of the lagged exogenous variables \(WX\) must equal the negative of the product of the spatial autoregressive coefficient \(\lambda\) and the coefficients of \(X\), for a total of \(K\) constraints.

Spatial error dependence may be interpreted as a nuisance (with parameter \(\lambda\) as the nuisance parameter) in the sense that it reflects spatial autocorrelation in measurement errors or in variables that are otherwise not crucial to the model (i.e., the spillover of the "ignored" variables across the spatial units of observation). It causes a problem of inefficiency in the regression estimates, which may be remedied by increasing the sample size or by exploiting consistent estimates of the nuisance parameter. The spatial autoregressive error model can also be expressed in terms of spatially filtered variables, but slightly different from (4.3). After moving the spatial lag variable in (4.8) to the left hand side, the following expression results:

\[
(I - \lambda W)y = (I - \lambda W)X\beta + \xi. \tag{4.9}
\]

This is a regression model with spatially filtered dependent and explanatory variables and with an uncorrelated error term \(\xi\) similar to first differencing of both \(y\) and \(X\) in time-series models. As in the spatial lag model, \(\lambda = 1\) is outside the parameter space and thus \(\lambda\) must be estimated jointly with
the other coefficients of the model.

4.6 Maximum Likelihood Estimation

Maximum likelihood estimation for regression models that incorporate spatial autocorrelation in the form of a spatial lag or a spatial error term assumes a joint normal density for the errors in the model, from which the likelihood function is derived. An important aspect of this likelihood function is the Jacobian of the transformation, which is $|I - \rho W|$ and $|I - \lambda W|$ in the spatial lag and spatial autoregressive error models, respectively, with $\rho$ and $\lambda$ as the autoregressive coefficient and $W$ as the spatial weights matrix. The spatial Jacobian is not the determinant of a triangular matrix but of a full matrix. This would normally complicate computational matters considerably, however it can be expressed as a function of the eigenvalues $\omega_i$ of the spatial weights matrix so

\begin{equation}
|I - \rho W| = \prod_{i=1}^{N} (1 - \rho \omega_i).
\end{equation}

Using this simplification, under the normality assumption, the log-likelihood function for the spatial lag model follows in a straightforward manner as

\begin{equation}
L = \sum \ln(1 - \rho \omega_i) - \frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{(y - \rho Wy + \lambda)'}{(y - \rho Wy + \lambda)} \frac{1}{2\sigma^2}
\end{equation}

Since it ignores the Jacobian term, the ordinary least squares estimate does not correspond to the maximum likelihood estimate. From the usual first-order conditions, the ML estimates for $\beta$ and $\sigma^2$ in a spatial lag model is
\[ \beta_{ML} = (X'X)^{-1}X'(I - \rho W)y \]  

(4.12)

and

\[ \sigma^2_{ML} = \frac{(y - \rho Wy)^t(X'y - \rho Wy + \beta_{ML})}{N}. \]  

(4.13)

Conditional upon \( \rho \), these estimates are simply OLS applied to the spatially filtered dependent variable and the explanatory variables in (4.3). Substitution of (4.12) and (4.13) in the log-likelihood function yields a concentrated log-likelihood as a nonlinear function of a single parameter \( \rho \):

\[ L_c = \frac{N}{2} \ln \left( \frac{(e_0 - \rho e_L)'(e_0 - \rho e_L)}{N} \right) + \sum_i \ln(1 - \rho \omega_i) \]  

(4.14)

where \( e_0 \) and \( e_L \) are residuals in a regression of \( y \) on \( X \) and \( Wy \) on \( X \), respectively. A maximum likelihood estimate for \( \rho \) is obtained from a numerical optimization of the concentrated log-likelihood function \( L_c \). The resulting estimates have the usual asymptotic properties, including consistency, normality, and asymptotic efficiency. The asymptotic variance matrix follows as the inverse of the information matrix

\[
\text{AsyVar}[\rho, \beta, \sigma^2] = \left[ \begin{array}{ccc}
\text{tr}[W_A]^2 + \text{tr}[W_A'W_A] & \frac{W_A'W_A}{\sigma^2} & \frac{X'W_A\beta}{\sigma^2} \\
\frac{X'W_A\beta}{\sigma^2} & \frac{X'X}{\sigma^2} & 0 \\
\frac{tr(W_A)}{\sigma^2} & 0 & \frac{N}{2\sigma^4}
\end{array} \right]^{-1}
\]  

(4.15)

where \( W_A = W(I - \rho W)^{-1} \).

Maximum likelihood estimation for models with spatial error autocorrelation can be approached by considering them as special cases of general parametrized nonspherical error terms, for which \( E[\varepsilon \varepsilon'] = \sigma^2 \Omega(\theta) \), with \( \theta \) as a vector of parameters. From (4.15) for a spatial autoregressive
error term, it follows that

\[ \Omega(\lambda) = [(I - \lambda W)'(I - \lambda W)]^{-1}. \] (4.16)

Under the assumption of normality, the log-likelihood function takes on the usual form:

\[ L = \frac{-1}{2} \ln|\Omega(\lambda)| - \frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{(y - \lambda \beta)'\Omega(\lambda)^{-1}(y - \lambda \beta)}{2\sigma^2} \] (4.17)

with \( \Omega(\lambda) \) as in (4.16). First-order conditions yield the familiar generalized least-squares estimates (GLS) for \( \beta \), conditional upon \( \lambda \):

\[ \beta_{ML} = [X'\Omega(\lambda)^{-1}X]^{-1}X'\Omega(\lambda)^{-1}y \] (4.18)

For a spatial autoregressive error process, \( \{\Omega(\lambda)^{-1}\} = (I - \lambda W)'(I - \lambda W) \), so for known \( \lambda \), the maximum likelihood estimates are equivalent to OLS applied to the spatially filtered variables in (4.9). For other forms of error dependence, the GLS expression (4.18) will involve the inverse of an \( N \times N \) error covariance matrix. To obtain a consistent estimate for \( \lambda \), an explicit optimization of the likelihood function must be carried out. One approach is to use the iterative solution of the first order

\[ \text{tr} \left[ \left( \frac{\partial \Omega^{-1}}{\partial \lambda} \right) \Omega \right] = e' \left( \frac{\partial \Omega^{-1}}{\partial \lambda} \right) e \] (4.19)

where \( e = y - \lambda \beta \) are GLS residuals. For a spatial autoregressive error process,

\[ \frac{\partial \Omega^{-1}}{\partial \lambda} = -W' - W + \lambda W'W. \] The solution of condition (4.19) can be obtained by numerical means. Alternatively, the GLS expression for \( \beta \) and a similar solution of the first-order conditions for \( \sigma^2 \) can be substituted into the log-likelihood function to yield a concentrated log-likelihood as a nonlinear function of the autoregressive parameter \( \lambda \), i.e.,

\[ L_C = \frac{N}{2} \ln \left( \frac{u'u}{N} \right) + \sum_i \ln(1 - \lambda \omega_i) \] (4.20)
with \( u'u = y_Ly_L - y_L'X_L[X_L'X_L]^{-1}X_L'y_L \) and \( y_L \) and \( X_L \) as the spatially filtered variables \( y - \lambda W y \) and \( X - \lambda W y \), respectively. The Jacobian term follows from \( \ln|\Omega(\lambda)| = 2 \ln|I - \lambda W| \). The asymptotic variance for the ML estimates in general form is block diagonal between the regression (\( \beta \)) and error variance parameters \( \sigma^2 \) and \( \theta \). The variance block for the error parameters is

\[
\text{AsyVar}[\sigma^2, \lambda] = \begin{bmatrix}
\frac{N}{2\sigma^4} & \frac{\text{tr}(W_b)}{\sigma^2} \\
\frac{\text{tr}(W_b)}{\sigma^2} & \text{tr}[W_B]^2 + \text{tr}[W'_bW_B]
\end{bmatrix}^{-1}
\] (4.21)

where \( W_b = W(I - \lambda W)^{-1} \). Due to the block-diagonal form of the asymptotic variance matrix, knowledge of the precision of \( \lambda \) does not affect the precision of the \( \beta \) estimates. Thus, if the latter is the primary interest, the complex inverse and trace expressions in (4.21) need not be computed. In a straightforward way, a significance test for the spatial error parameter can be based on a likelihood ratio test. The fit of spatial process models estimated by means of maximum likelihood procedures should not be assessed by the traditional \( R^2 \), which will be misleading in the presence of spatial autocorrelation. Instead, the fit of the model should be assessed by comparing the maximized log-likelihood or an adjusted form that takes into account the number of parameters in the model, such as the familiar AIC (Anselin and Bera, 1998).

Considering the models (4.2) for Spatial Lag Dependence and (4.6) for Spatial Error Dependence, where and \( \rho \) and \( \lambda \) are scalar autoregressive parameters, the diagnostic tools used to identify a spatially autoregressive error term or an erroneously omitted spatial lag are the Lagrange Multiplier (LM) tests:

\[
LM_\lambda = \frac{(\hat{e}'W\hat{e}/\hat{\sigma}^2)^2}{T},
\]
and

$$LM_\rho = \frac{(\hat{\epsilon}'W\hat{\epsilon}/\hat{\sigma}^2)^2}{NJ},$$

with

$$J = \frac{1}{N\hat{\sigma}^2}[((WX\hat{\beta})'M(WX\hat{\beta}) + T\hat{\sigma}^2)]$$

where $LM_\lambda$ has the spatial error model as the alternative hypothesis and $LM_\rho$ has a spatial lag model as the alternative hypothesis. $M = I - X(X'X)^{-1}X'$, $T$ is the trace of the matrix $((W' + W)W)$, $\hat{\epsilon} = My$ are the OLS residuals, $\hat{\sigma}^2 = \hat{\epsilon}'\hat{\epsilon}/N$, and all other symbols are as previously defined. It should be noted that three (3) different estimators are appropriate, depending on the model specification: OLS for the specification with i.i.d. error terms and without a spatially lagged dependent variable, and maximum likelihood estimators for the spatially autoregressive error model (MLERR) and for the model including a spatially lagged dependent variable (MLLAG). The test statistics are distributed as $\chi^2$ with one degree of freedom.

The classical approach towards specification can be summarized as follows:

1. Estimate the initial model $y = X\beta + \epsilon$ by means of OLS.
2. Test the hypothesis of no spatial dependence due to an omitted spatial lag or due to spatially autoregressive errors, using $LM_\rho$ and $LM_\lambda$ respectively.
3. If both tests are not significant, the initial estimates from step 1 are used as the final specification. Otherwise proceed to step 4.
4. If both tests are significant, estimate the specification pointed to by the more significant of the two tests. For example if $LM_\rho > LM_\lambda$ then estimate ((4.2) in Spatial Lag Dependence) using...
MLLAG. If \( LM_P < LM_x \) then estimate (4.6) in Spatial Error Dependence) using MLERROR.

Otherwise, proceed to step 5.

5. If \( LM_P \) is significant but \( LM_x \) is not, estimate (4.2) in Spatial Lag Dependence) using MLLAG.

Otherwise proceed to step 6.

6. Estimate (4.6) in Spatial Error Dependence) using MLERROR. (Florax et. al, 2003)

### 4.7 Model Fitting for Ottawa Data

We illustrate the model fitting above with reference to Ottawa data. Based on 10 kilometers we construct a weight matrix. The connectivities are as in the following Figure 4.1, which describes the distribution of locations (the number of observations (wards) in each category is shown at the top of the corresponding bar) by number of neighbours (shown in the legend).

![Figure 4.1. Wards connectivities](image)

For example, we can see that there are 4 wards with 0 neighbours in the bar at left, and they correspond to the wards 6, 19,20 and 21 as shown in Figure 4.2 below:
and 2 wards with 8 neighbours in the bar at the right side of the chart, and they correspond to the
ward 14 and 17 as shown in Figure 4.3 below. (See Figure 3.1 for the ward numbering and
mapping.)

We now apply the classical OLS approach for the model specification and we find that:

4.7.1 For 2004 data:

- $Z_{I04}$ is the Freeman-Tukey transformed data of nonresidential B&E
- as the dependent variable, and
- $R_{04TR}$ is the Freeman-Tukey transformed data of residential B&E,
- AREA is the area of each ward,
PERIMETER is the perimeter of each ward,

POP2004 is population of each ward, and

HH04SQ is the number of households per square kilometer in each

ward

are the independent variables. The OLS output appears as Table 4.1.

Table 4.1. OLS model output for 2004 data.

The regression output shows a respectable fit, with an adjusted $R^2$ of 0.51, and

Prob(F-statistic) = 0.005. The plot of the residuals against the predicted values appears as Figure 4.4.
Figure 4.4. Plot of the residuals against the predicted values-OLS

In this graph, we look for evidence of funnel-like patterns, suggesting a relation between the spread of the residuals and the predicted value, but there is no evidence of this in Figure 4.4. Formal testing for heteroskedasticity can be used to supplement this visual inspection. Table 4.2 below displays diagnostics provided in the regression output window and consists of a set of three traditional measures: the multicollinearity condition number, the Jarque-Bera test for non-normality, and diagnostics for heteroskedasticity.

<table>
<thead>
<tr>
<th>REgression Diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td>MULTICOLLINEARITY CONDITION NUMBER</td>
</tr>
<tr>
<td>TEST ON NORMALITY OF ERRORS</td>
</tr>
<tr>
<td>Jarque-Bera</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>DIAGNOSTICS FOR HETEROskedASTICITY</td>
</tr>
<tr>
<td>RANDOM COEFFICIENTS</td>
</tr>
<tr>
<td>Breusch-Pagan test</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>Koenker-Basset test</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>SPECIFICATION ROBUST TEST</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>20</td>
</tr>
</tbody>
</table>

Table 4.2. Regression diagnostics – OLS model for 2004 data

The multicollinearity condition number is not a test statistic by itself but is a diagnostic to
suggest problems with the stability of the regression results due to multicollinearity (i.e. where the explanatory variables are too correlated and provide insufficient separate information). Typically, an indicator over 30 is suggestive of some problems (Anselin, 2005); in our case it is 33.1 which is only slightly over 30 and thus not too problematic. The Jarque-Bera test for normality of the errors is distributed as a $\chi^2$ statistic with 2 degrees of freedom. In our case, there is suggestion of some non-normality of the errors but it is not severe (with $p > 0.05$).

The next three diagnostics are common statistics to detect heteroskedasticity, (i.e., a non-constant error variance). Both the Breusch-Pagan and Koenker-Bassett tests are implemented as tests on random coefficients, which assumes a specific functional form for the heteroskedasticity. Specifically, the heteroskedasticity is a function of the squares of the explanatory variables (Anselin, 2005). The Koenker-Bassett test is essentially the same as the Breusch-Pagan test, except that the residuals are studentized, i.e., they are made robust to non-normality. Both test statistics indicate no problems with heteroskedasticity in OLS specifications.

The final step of the model diagnostics consists of tests for spatial autocorrelation. The tests are all computed for the same weights matrix. We use the Lagrange Multiplier $LM_\rho$ for testing the Spatial Lag Dependence and $LM_\lambda$ for testing the Spatial Error Dependence. From Table 4.3 below, we can see that: the LM-Lag ($LM_\rho$) statistic is (nearly) significant (with $p < 0.087$), while the LM-Error ($LM_\lambda$) statistic is not at all (with $p = 0.51$). This indicates that a spatial lag specification should be estimated for the Freeman-Tukey transformed data of nonresidential B&E 2004.
**Diagnostics for Spatial Dependence**

For Weight Matrix: (row-standardized weights)

<table>
<thead>
<tr>
<th>Test</th>
<th>MI/DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moran's I (error)</td>
<td>-0.097757</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Lagrange Multiplier (lag)</td>
<td>1</td>
<td>2.9313436</td>
<td>0.0868752</td>
</tr>
<tr>
<td>Robust LM (lag)</td>
<td>1</td>
<td>2.7514314</td>
<td>0.0971674</td>
</tr>
<tr>
<td>Lagrange Multiplier (error)</td>
<td>1</td>
<td>0.4294664</td>
<td>0.5122509</td>
</tr>
<tr>
<td>Robust LM (error)</td>
<td>1</td>
<td>0.2495542</td>
<td>0.6173891</td>
</tr>
<tr>
<td>Lagrange Multiplier (SARMA)</td>
<td>2</td>
<td>3.1808978</td>
<td>0.2038341</td>
</tr>
</tbody>
</table>

---

Table 4.3. Spatial autocorrelation diagnostics – OLS model.

The estimation by means of maximum likelihood of a spatial regression model that includes a spatially lagged dependent variable (ZI04) appears below in Table 4.4.

**Regression Summary of Output: Spatial Lag Model – Maximum Likelihood Estimation**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Wards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Weight</td>
<td></td>
</tr>
<tr>
<td>Dependent Variable</td>
<td>ZI04</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>21</td>
</tr>
<tr>
<td>Mean dependent var</td>
<td>1.83623</td>
</tr>
<tr>
<td>Number of Variables</td>
<td>7</td>
</tr>
<tr>
<td>S.D. dependent var</td>
<td>1.03591</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>14</td>
</tr>
<tr>
<td>Lag coeff. (Rho)</td>
<td>-0.468377</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.708081</td>
</tr>
<tr>
<td>Sq. Correlation</td>
<td>-</td>
</tr>
<tr>
<td>Sigma-square</td>
<td>0.313865</td>
</tr>
<tr>
<td>S.E of regression</td>
<td>0.560236</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-17.8209</td>
</tr>
<tr>
<td>Akaike info criterion</td>
<td>49.6419</td>
</tr>
<tr>
<td>Schwarz criterion</td>
<td>56.9535</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std.Error</th>
<th>z-value</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>W ZI04</td>
<td>-0.4683772</td>
<td>0.1711053</td>
<td>-2.737362</td>
<td>0.0061935</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>-1.543531</td>
<td>0.8819647</td>
<td>-1.750105</td>
<td>0.0801001</td>
</tr>
<tr>
<td>RO4TR</td>
<td>0.2952175</td>
<td>0.0946753</td>
<td>3.118216</td>
<td>0.0018197</td>
</tr>
<tr>
<td>AREA</td>
<td>0.0113695</td>
<td>0.004929051</td>
<td>2.895452</td>
<td>0.0238552</td>
</tr>
<tr>
<td>PERIMETER</td>
<td>0.05860588</td>
<td>0.03049428</td>
<td>-1.823468</td>
<td>0.0682298</td>
</tr>
<tr>
<td>POP2004</td>
<td>6.0158626e-005</td>
<td>1.323515e-005</td>
<td>4.582845</td>
<td>0.0010278</td>
</tr>
<tr>
<td>HHG0SQ</td>
<td>-0.0001946763</td>
<td>0.0003121742</td>
<td>-0.623614</td>
<td>0.5328809</td>
</tr>
</tbody>
</table>

---

Table 4.4. ML estimation results, spatial lag model, ZI04.

The estimates and measures of fit are listed in Table 4.4 above. It is not appropriate in a spatial regression model to focus on usual measures, such as the $R^2$. The value listed in the spatial lag output is not a real $R^2$, but a pseudo-$R^2$, which is not directly comparable with the measure given for OLS results. More correct measures of fit are the log-likelihood, AIC ($AIC = -2L + 2K$), where L
is the log-likelihood and $K$ is the number of parameters in the model) or SC (the
$SC = -2L + K \ln(N)$, where "ln" is the natural logarithm) (Anselin, 2005). If we compare the values
in Table 4.4 to those for OLS in Table 4.1, we notice an increase in the log-likelihood from -19.9
(for OLS) to -17.8 showing the payback for the improved fit with the added variable (the spatially
lagged dependent variable). The AIC (from 51.8 to 49.6) and SC (from 58.1 to 56.9) both decrease
relative to OLS, again suggesting an improvement in fit by using the spatial lag specification. The
spatial autoregressive coefficient $\rho$ is estimated as -0.46, and is very significant ($p < 0.006$). There
are some differences in the significance of the other regression coefficients between the spatial lag
model and the classic specification: the constant is less significant than before ($p < 0.08$), but, more
importantly, the significance of $R04TR$ changes from $p < 0.06$ to $p < 0.0018$, and the significance
of $POP2004$ changes from $p < 0.091$ to $p < 0.001$. The magnitude of all the estimated coefficients is
also affected, all of them showing a decrease in absolute value.

<table>
<thead>
<tr>
<th>REGRESSION DIAGNOSTICS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DIAGNOSTICS FOR HETEROSEDASTICITY</td>
<td></td>
</tr>
<tr>
<td>RANDOM COEFFICIENTS</td>
<td></td>
</tr>
<tr>
<td>TEST</td>
<td>DF</td>
</tr>
<tr>
<td>Breusch-Pagan test</td>
<td>5</td>
</tr>
</tbody>
</table>

| DIAGNOSTICS FOR SPATIAL DEPENDENCE | |
| SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : | |
| TEST | DF | VALUE | PROB |
| Likelihood Ratio Test | 1 | 4.207377 | 0.0402485 |

Table 4.5. Diagnostics, spatial lag model, ZI04.

Table 4.5 gives some diagnostics associated with the ML lag estimation. First is a
Breusch-Pagan test for heteroskedasticity in the error terms. The highly insignificant value of 0.718
suggests that there is no problem with the heteroskedasticity.

The Likelihood Ratio Test is one of the three classic specification tests comparing the null
model (the classic regression specification) to the alternative (spatial lag model). The value of 4.2 with $p = 0.04$ confirms the significance of the spatial autoregressive coefficient for the 2004 data.

The other two classic tests are the Wald test, i.e., the square of the asymptotic $z$-value, and the LM-lag test based on OLS residuals. The three classic tests are asymptotically equivalent, but in finite samples should follow the ordering: $W > LR > LM$ (Anselin, 2005). In our case, the Wald test is $2.7^2 = 7.29$, the LR test is 4.2 and the LM-Lag test was 2.93, which is compatible with the expected order.

### 4.7.2 For 2001 data

ZI01 is the Freeman-Tukey transformed data of nonresidential B&E as the dependent variable, and

R01TR is the Freeman-Tukey transformed data of residential B&E,

AREA is the area of each ward,

PERIMETER is the perimeter of each ward,

POP2001 is population of each ward, and

HH01SQ is the number of households per square kilometer in each ward

are the independent variables, the OLS output is
The regression output shows an adjusted $R^2$ of 0.31, and $\text{Prob}(F\text{-statistic}) = 0.05$. A plot of the residuals against the predicted values appears below in Figure 4.5.

The plot of the residuals vs the predicted value suggests no pattern.
REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 32.13285
TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 0.3535914 0.8379509

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 5 5.348019 0.3748990
Koenker-Bassett test 5 6.561263 0.2553667

Table 4.7. Regression diagnostics – OLS model for 2001 data

The output of diagnostics provided in the above regression output window suggests that the
multicollinearity condition number is around 32 which is not bad. The Jarque-Bera test on normality
of the errors gives a strong suggestion of normality of the errors. The heteroskedasticity is also not
significant.

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : nb1.GWT (row-standardized weights)
TEST MI/DF VALUE PROB
Moran's I (error) -0.123704 N/A N/A
Lagrange Multiplier (lag) 1 1.3151691 0.2514609
Robust LM (lag) 1 0.6692115 0.4133266
Lagrange Multiplier (error) 1 0.5877132 0.4069433
Robust LM (error) 1 0.0417557 0.8380862
Lagrange Multiplier (SARMA) 2 1.3569248 0.5073966

Table 4.8. Spatial autocorrelation diagnostics – OLS model.

Table 4.8 shows that neither $LM_p$ nor $LM_1$ are significant, so OLS is used as the final
specification. Thus for the 2001 data we do not find a spatial relationship like we did for the 2004
data.
Chapter 5: Spatial Point Pattern (SPP) Analysis

Now, adopting the use of stochastic processes, we address the question of whether there is spatial structure or complete randomness in our data. If there is structure, we will explore the nature of the spatial point pattern by examining the pattern of location of events in a particular region. We will look at measures of the intensity of a spatial point pattern and at ways to test whether or not there is clustering. In the event of clustering, we present ways of identifying the clustering patterns. To illustrate, we will take Breaking and Entering (B & E) incidents occurring in central east area of Ottawa in the year 2004.

5.1 Introduction

The realization of a point process \( \{Y(A), A \subseteq R\} \) consists of a collection of points (events) randomly occurring in sub-region \( A \) of the total study region \( R \), where \( Y(A) \) is number of events occurring in \( A \). These points are expressed as the events of the point process, to distinguish them from arbitrary points \( x \) in the plane. If the events are partially observed, the recorded pattern is called a *sampled* point pattern. When all events of the realization are recorded the point pattern is said to be *mapped*. Since \( A \) is random, the observed random events can be viewed as a random draw of locations in \( R \) at which events are observed, and all mapped point pattern realizations (outcomes) can be distinguished as one of random, clustered, or regular. A point pattern is completely random if it follows the following conditions:

i. the average number of events per unit area (the intensity \( \lambda(s) \)) is homogeneous throughout \( A \)

ii. the number of events in two non-overlapping subregions are independent.
iii. the number of events in any subregion is Poisson distributed.

The mathematical expression of complete spatial randomness (CSR) is analogous to the homogeneous Poisson process. This process serves as the null hypothesis; if not rejected, it leads to the conclusion that there is no spatial structure, i.e. clustering. Rejection of the CSR hypothesis leads to further exploration of the nature of the spatial point pattern (Schabenberger and Gotway, 2005).
5.2 First- and second-order properties

The behaviour of a general spatial stochastic process might be described in terms of its so-called first-order and second-order properties. The first-order properties explain the manner in which the expected value $E(Y(A))$ of the process varies throughout the space, whereas second-order properties express the covariance $COV(Y(A_i), Y(A_j))$ between values of the process at different regions in space. In order to detect a 'pattern' in spatial data, it is essential to consider that this might happen either from region-wide 'trends' (first-order variation) or from correlation structures (second-order variation), or from a mixture of both. First-order properties are expressed in terms of the intensity, $\lambda(s)$, of the process, which is the average number of events per unit area at the point $s$, defined as the mathematical limit:

$$\lambda(s) = \lim_{ds \to 0} \left\{ \frac{E(Y(ds))}{ds} \right\}$$

where $ds$ is a small region around the point $s$, $E(\bullet)$ is the expectation operator and $ds$ is the area of this region, and $Y(ds)$ is the number of events in this small region. Second-order properties of a spatial point process involve the relationship between numbers of events in pairs of subregions within $R$. This is defined in terms of a limit, as the second-order intensity of the process:

$$\gamma(s_i, s_j) = \lim_{ds_i, ds_j \to 0} \left\{ \frac{E(Y(ds_i)Y(ds_j))}{ds_i ds_j} \right\}$$

where the notation is similar to that described above. If the intensity is constant over $R$, the point process is stationary, hence $\lambda(s) = \lambda$ and, in addition, $\gamma(s_i, s_j) = \gamma(s_i - s_j) = \gamma(d)$. This implies that the second-order intensity depends only on the vector difference, $d$ (direction and distance), between
and not on their absolute locations. The process is isotropic if such dependence is a function only of the length, \( d \), of this vector \( \mathbf{d} \) and not its orientation (Gatrill et al 1996; Bailey & Gatrill 1995).

5.3 Visualizing Spatial Point Patterns

A dot map of a spatial point pattern is one way to visualize the data. The map gives a preliminary idea about where events occurred in the region of interest; in addition, it shows whether there is any obvious pattern present among the events or not. To illustrate, we will take Breaking and Entering (B&E) incidents occurring in central east area of Ottawa in the year 2004 (this is just one third of the data provided by the Ottawa Police, as we mentioned before). The dataset we use here consists of the locations of 61 B&Es, and they are distributed in a region of 1356 X 1251 square meters (approximately 1.7 km\(^2\)). To have an idea about the area, the following Figure 5.1 demonstrates the locations of these B&Es events.
The plot gives an impression that the events are randomly distributed, but if we count the dots
we find that they are less than 61, which means that some events are overlapped. In addition we are
looking at a relatively large area (approximately 1.7 km$^2$) with a small number of events, and we can
see some events are close together, so we need more information to judge the behaviour of the data.

5.4 Exploring Spatial Point Patterns

This is done by deriving summary statistics or plots from the observed distribution of events to
investigate first-order effects, or spatial dependence (second-order effects) in the process. We use
them to examine hypotheses of interest or to suggest possible models.

5.4.1 Quadrat Methods

To examine the pattern of location of events in a particular region R, we partition R into
sub-regions of equal area called quadrats. Then counts of the number of events in each quadrat are
used to summarize the spatial pattern. This simply will create a 2-dimensional histogram which shows how intensity changes over $R$. For our data we partitioned the area $R$ into $7 \times 7$ sub-regions of equal area (quadrats). The following Figure 5.2 shows the quadrats:

![Figure 5.2 Partitioning into quadrats](image)

and Table 5.1 below summarizes the counts of the number of events in each quadrat.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.1. Counts by quadrat

To summarize the spatial pattern in order to detect any pattern, we can see the frequencies of
quadrats that contain number of events is as follows:

<table>
<thead>
<tr>
<th>Number of events</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td># of quadrats</td>
<td>21</td>
<td>12</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.2. Frequency distribution of quadrat counts re Fig.5.2

We note that 21 quadrats do not contain any events and 12 quadrats contain only 1 event, but also we note 2 clusters of 5 events, 2 clusters of 4 events, 7 clusters of 3 events, and 5 clusters of 2 events. Although we note these clusters, we cannot give a formal conclusion of clustering unless we test them. By obtaining a set of quadrat counts, the original point pattern has been transformed into a set of area data (or a lattice) so we can also use the tools of previous chapters. The problem with the quadrat counts is that many spatial details of the initially observed pattern are neglected. As the quadrats are made smaller to get more spatial details as in the following Figure 5.3 where we partitioned \( R \) into 20 \( \times \) 20 quadrats

![Figure 5.3. Finer partitioning into quadrats](image-url)
with the following frequencies:

<table>
<thead>
<tr>
<th>Number of events</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td># of quadrats</td>
<td>353</td>
<td>35</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.3. Frequency distribution of quadrat counts re Fig.5.3

we get high variability in quadrat counts and many empty cells or quadrats, making a significant interpretation impossible.

5.4.2 Kernel Estimation

Kernel estimation is used to further investigate the first-order properties of the events distribution. The kernel estimate provides a global measure of the intensity of a spatial point pattern. Intensity can be thought of as a summary estimate of the number of events per unit area across a plane. Kernel estimation was initially developed to obtain a smooth estimate of a univariate or multivariate probability density function (pdf) from an observed sample of observations, i.e., a smooth histogram. Estimating the intensity of a spatial point pattern is very similar to estimating a bivariate probability density function. Therefore, bivariate kernel estimation can be adapted to give an estimate of the intensity. If \( s \) is an arbitrary location in \( \mathbb{R} \), and \( s_1, \ldots, s_n \) are the locations of the \( n \) observed events, then intensity \( \lambda(s) \) at location \( s \) can be estimated by

\[
\hat{\lambda}_r(s) = \frac{1}{\delta_r(s)} \sum_{i=1}^{n} \frac{1}{\tau^2} k\left( \frac{(s - s_i)}{\tau} \right)
\]

where \( k(\cdot) \) is the chosen bivariate pdf, called the kernel, which is symmetric about the origin. The parameter \( \tau > 0 \) is the bandwidth that determines the amount of smoothing (essentially the radius of a disc centered on \( s \) within which points \( s_i \) will contribute significantly to \( \hat{\lambda}_r(s) \)). The factor

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[ \delta_r(s) = \int_R \frac{1}{\tau^2} k \left( \frac{s - u}{\tau} \right) du \]

is the edge correction, that is, the volume under the scaled kernel centered on \( s \) which lies inside \( \mathbb{R} \).

For most reasonable choices of pdf \( k(\bullet) \), the kernel estimate \( \hat{\lambda}_r(s) \) will be very similar for a given bandwidth \( \tau \). The quartic kernel is frequently used for \( k(\bullet) \), where

\[ k(u) = \begin{cases} \frac{3}{\pi} (1-u^T u)^2 & \text{for } u^T u \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

Ignoring the edge correction, the estimated \( \hat{\lambda}_r(s) \) becomes

\[ \hat{\lambda}_r(s) = \sum_{h_i \leq \tau} \frac{3}{\pi \tau^2} \left( 1 - \frac{h_i^2}{\tau^2} \right)^2 \]

where \( h_i \) is the distance between the point \( s \) and event location \( s_i \). Summation is only over values of \( h_i \) which do not exceed \( \tau \). The region of influence in which observed events contribute to \( \hat{\lambda}_r(s) \) is a circle of radius \( \tau \) centered on \( s \). Choice of \( \tau \) impacts the "stretch" of the region around \( s \). When \( \tau \) is too large, \( \hat{\lambda}_r(s) \) will be flat, and local features will be obscured; if \( \tau \) is too small, then \( \hat{\lambda}_r(s) \) is often a collection of spikes centered on the \( s_i \)'s. A rough recommendation for \( \tau \) is to set it equal to \( 0.68n^{-0.2} \) when \( \mathbb{R} \) is the unit square and \( n \) is the number of observed events in \( \mathbb{R} \). The recommendation could be scaled up to fit the dimensions of other studies. In practice one can experiment with different values of \( \tau \) at different points in \( \mathbb{R} \). To explore the surface \( \hat{\lambda}_r(s) \) for B & E data, we can use different degrees of smoothing depending on different values of the bandwidth \( \tau \) in order to look at the variation in \( \lambda(s) \) at different scales. Hence, any number of kernel estimate maps could be produced.

Taking \( \tau = 60 \), \( \tau = 120 \), and \( \tau = 180 \) meters will produce the following kernel estimation maps:
Figure 5.4. Kernel estimate map, $\tau = 60$

Figure 5.5. Kernel estimate map, $\tau = 120$
At $\tau = 60$, the kernel smoothing reveals a collection of spikes centered on the locations of events, which gives the impression of a spiky spatial distribution, while at $\tau = 180$ we see that the kernel smoothing is excessive and we lose information from the kernel smoothing map. At $\tau = 120$ we have a reasonable representation of the regional distribution, confirming the results of quadrat counts about clustering.

5.4.3 Nearest Neighbour Distances

Nearest neighbour distances are exploratory methods which are designed to investigate second-order properties using the distances between observed events in the study area. Two such kinds of distances are considered. The first one is the Nearest Neighbour Event-Event Distance ($W$),
which takes the distance between a randomly chosen event and the nearest neighbouring event. The second one is the Nearest Neighbour Point-Event distance \( X \), which takes the distance between a randomly selected point in \( R \) and the nearest neighbouring observed event. The second one is generally most useful in sampling from point patterns. Both provide useful information about inter-event interactions at a small physical scale because, by definition, they use small inter-event distances. If there is large scale variation in the intensity of a point pattern over \( R \), these measures are not very important. To summarize nearest neighbour distances \( W \) and \( X \), we estimate them using the empirical cumulative distribution function (c.d.f.) \( \hat{G}(w) \) of \( W \) and \( \hat{F}(x) \) of \( X \) from the data as

\[
\hat{G}(w) = \frac{\#(w_i \leq w)}{n}
\]

and

\[
\hat{F}(x) = \frac{\#(x_i \leq x)}{m}
\]

where \# means the number of, \( n \) is number of events in \( R \), and \( m \) is number of random points sampled. Plotting \( \hat{G}(w) \) against distance \( w \) or \( \hat{F}(x) \) against distance \( x \) will give a useful interpretation. For instance if the c.d.f. is initially steep and then flattens, it indicates clustering due to inter-event attraction. While if the c.d.f. is first flat, then steep, it suggests inter-event repulsion or regularity. In Figure 5.7 the distribution function climbs steeply indicating that much of the distribution is comprised of event-event nearest neighbour distances. This suggests the presence of spatial clustering.
In Figure 5.8, there is an observed high probability of point-event distances which also suggests clustering in the event distributions for B & E in the central east area of Ottawa in 2004.
Another idea is to plot $\hat{G}(w)$ against $\hat{F}(x)$ to compare the two. If there is no interaction (i.e. a random pattern), then $\hat{G}(w)$ and $\hat{F}(x)$ are similar and we get a straight line. If there is attraction between points (i.e. clustering), then $\hat{G}(w) > \hat{F}(x)$; in the case of repulsion between points (regularity), $\hat{G}(w)$ would be less than $\hat{F}(x)$. 

Figure 5.8. Plot of $\hat{F}(x)$ against distance $x$
Clearly we can see from Figure 5.9 for our Ottawa 2004 B & E data that $\hat{G}(w) > \hat{F}(x)$, i.e. there is attraction between points indicating a positive interaction or clustering.

5.4.4 The $K$ Function

The problem with nearest neighbour distances is that they only consider the smallest scales of the pattern and information on larger scales of the pattern is ignored. Hence, it would be better to use an estimate of the reduced second moment measure or $K$ function of the observed process. The $K$ function relates to the second-order properties of an isotropic process, so if there are large scale first-order effects, then any spatial dependence indicated by the $K$ function might be due to first-order effects rather than to interaction effects. The definition of the $K$ function is

$$\lambda K(h) = E(\# \text{events within distance } h \text{ of an arbitrary event})$$
where \# means the number of, \( E() \) is the normal expectation, and \( \lambda \) is the intensity (assumed to be constant over \( R \)). If \( R \) is the area of \( R \), then the expected number of events in \( R \) is \( \lambda R \). Consequently, from the definition of the \( K \) function, the expected number of \textit{ordered} pairs of events at distance at most \( h \) apart, with the first event in \( R \), is \( \lambda R \cdot \lambda K(h) = \lambda^2 R K(h) \). If \( d_{ij} \) is the distance between \( i \)th and \( j \)th observed events in \( R \), and \( I_h(d_{ij}) \) is an indicator function which equals 1 if \( d_{ij} \leq h \) and equals 0 otherwise, then the observed number of ordered pairs of events at distance at most \( h \) apart is

\[
\sum_i \sum_{j \neq i} I_h(d_{ij}).
\]

Hence the estimator of \( K(h) \) is given by

\[
\hat{K}(h) = \frac{1}{\lambda^2 R} \sum_i \sum_{j \neq i} I_h(d_{ij}).
\]

There is a problem with points that are close to the edge, where the summation excludes pairs of events for which the second event is outside \( R \) and hence unobservable. To solve this problem we need an edge correction. Consider a circle centered on event \( i \), passing through the point \( j \) and let \( w_{ij} \) be the proportion of the circumference of the circle which lies within \( R \). Then \( w_{ij} \) is the conditional probability that an event is observed in \( R \), given that it is within distance \( d_{ij} \) from the \( i \)th event. Then the edge-corrected estimator for \( K(h) \) is

\[
\hat{K}(h) = \frac{1}{\lambda^2 R} \sum_i \sum_{j \neq i} \frac{I_h(d_{ij})}{w_{ij}}.
\]

Since intensity \( \lambda \) is unknown, we replace it by \( \hat{\lambda} = n / R \). We then get

\[
\hat{K}(h) = \frac{R}{n^2} \sum_i \sum_{j \neq i} \frac{I_h(d_{ij})}{w_{ij}}.
\]

The calculation of the weights \( w_{ij} \) for arbitrarily shaped \( R \) can be computationally intensive,
however it is simple for rectangular or circular shaped R. When using edge-correction algorithms, $h$ should be relatively small compared to R. In practice, it is not a serious problem because we do not try to explore second-order effects which operate on the same physical scale as the dimensions of R. It is useful to plot $\hat{K}(h)$ against $h$ in order to examine any suggestion of spatial dependence in the point pattern at different values of $h$. For a random spatial point process, we suppose that the probability of the occurrence of an event at any point in R is independent of the other events that have occurred, and the probability of the occurrence of an event is constant over the whole R. Therefore, for such a random process, $E(\# \text{ of events within distance } h \text{ of a randomly chosen event})$ is $\lambda \pi h^2$. This will lead to $K(h) = \pi h^2$ for a homogeneous process with no spatial dependence. So, under regularity, $K(h) < \pi h^2$ and under clustering, $K(h) > \pi h^2$. This suggests an obvious approach to exploratory use of the $K$ function, where we compare $\hat{K}(h)$, estimated from the observed data, with $\pi h^2$. The following Figure 5.10 is the plot of $\hat{K}(h)$ vs $\pi h^2$.
Another useful way is to use $\hat{L}(h)$ where

$$\hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h$$

and plot $\hat{L}(h)$ against $h$. In this plot, positive peaks show clustering for distances of $h$, and negative peaks show regularity for distances of $h$. The following Figure 5.11 is a plot $\hat{L}(h)$ against $h$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
As exploratory tools the plots of $\hat{K}(h)$ and $\pi h^2$ against the distance $h$, and also the plot of $\hat{L}(h)$ against $h$ show that for distances less than about 1 km there is clustering, while for distances of more than 1 km there is regularity. We will test this later.

5.5 Modeling Spatial Point Patterns

The purpose of modeling spatial point patterns is to test various hypotheses in order to construct specific models to explain observed point patterns, or to carry out statistical comparison of summary measures calculated from observed events under various theoretical models. The standard against which spatial point patterns are compared is a Completely Spatially Random (CSR) point process.
5.5.1 Complete Spatial Randomness (CSR)

The typical model for CSR is that events follow a homogeneous Poisson process over the study region. For the set of random variables \( \{ Y(A), A \subseteq R \} \), it implies that \( Y(A_i) \) and \( Y(A_j) \) are independent for any choices of subregions \( A_i \) and \( A_j \); in addition, the probability distribution of \( Y(A) \) is a Poisson distribution with mean value \( \lambda A \), where \( A \) is the area of \( A \), i.e., the pdf of \( Y(A) \) is

\[
f_{Y(A)}(y) = \frac{(\lambda A)^y}{y!} e^{-\lambda A}
\]

where \( \lambda \) is a constant intensity and equals the average number of events per unit area. Conditional on \( n \) (the total number of events in \( R \)), events are independently and uniformly distributed over \( R \). First, any event has an equal probability of occurring at any position in \( R \) and second, the position of any event is independent of the position of any other event, i.e., events do not interact with one another.

We simulate of \( n \) events from a homogeneous Poisson process by enclosing \( R \) in a rectangle \( \{(x,y), x_1 \leq x \leq x_2, y_1 \leq y \leq y_2\} \); after that we generate events with \( x \)-coordinates from a uniform distribution on \((x_1,x_2)\) and \( y \)-coordinates from a uniform distribution on \((y_1,y_2)\), rejecting any event with coordinates \((x,y)\) that do not lie in \( R \), and continuing until \( n \) events have been obtained in \( R \).

The CSR represents the null hypothesis against which to assess whether an observed pattern is regular, clustered or random.

5.5.2 Simple Quadrat Tests for CSR

One way to test for CSR is to use the quadrat counts by implementing an Index of Dispersion Test. Let \((x_1,\ldots,x_m)\) be the counts of the number of events in \( m \) quadrats, either randomly scattered in \( R \), or forming a regular grid covering the whole of \( R \). Based on the idea that if the counts follow a
Poisson distribution, it is expected that the mean and variance of these counts are equal, the test statistic then is

\[ X^2 = \frac{(m-1)s^2}{\bar{x}} = \frac{\sum_{i=1}^{m} (x_i - \bar{x})^2}{\bar{x}} \]

where \( \bar{x} \) and \( s^2 \) are the mean and variance of the observed counts respectively. Under CSR, the theoretical probability distribution of \( X^2 \) can be approximated by a \( \chi^2_{m-1} \) distribution, given that \( m > 6 \) and \( \bar{x} > 1 \). This offers the foundation of the significance test, which compares the observed value of \( X^2 \) with percentage points of the \( \chi^2_{m-1} \) distribution. Significantly large or small values of \( X^2 \) indicate clustered or regular departures from CSR, respectively.

Then \( s^2/\bar{x} \) is called the index of dispersion, and \( (s^2/\bar{x}) - 1 \) is called the index of cluster size (ICS). For CSR, \( E(ICS) = 0 \), and \( ICS > 0 \) implies clustering (a number of extra events) and \( ICS < 0 \) implies regularity (a deficiency of events). Similarly, if we assume intensity \( \lambda \) is a constant, and CSR holds, then taking \((x_1, \ldots, x_m)\) based on \( m \) quadrats that are randomly scattered in \( R \) to estimate \( \lambda \) as

\[ \hat{\lambda} = \frac{\bar{x}}{Q} \]

(where \( Q \) is the area of each quadrat and should be constant) and given that \( VAR(\hat{\lambda}) \) is approximately \( \hat{\lambda}/mQ \), then an approximate 95% confidence interval for \( \hat{\lambda} \) would be

\[ \hat{\lambda} \pm 1.96 \sqrt{\frac{\hat{\lambda}}{mQ}} \]

where \( mQ \) is the total area.

Note that if \( m \) is large, then the randomly scattered quadrats are likely to overlap and the \( x_i \) counts will not be independent. Also overlaps with the edge of \( R \) can be a problem. For the B&E
incidents we have number of events \( n = 61 \), and we partitioned the area \( R \) into \( 7 \times 7 \) quadrats, (i.e. \( m = 49 \)), so that the mean quadrat count \( \bar{x} = 61/49 = 1.244898 \). Counts for each quadrat are indicated previously in the plot and the table of the quadrat method section. These quadrat counts are compiled in Table 5.4 below from which we calculate the observed variance \( s^2 \) to be

\[
101.06/49 = 2.062.
\]

<table>
<thead>
<tr>
<th>No</th>
<th>#events ((x_i))</th>
<th>#quadrats ((m_i))</th>
<th>((x_i - \bar{x}))</th>
<th>((x_i - \bar{x})^2)</th>
<th>(m \times (x_i - \bar{x})^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>21</td>
<td>-1.244898</td>
<td>1.54977093</td>
<td>32.5451895</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>12</td>
<td>-0.244898</td>
<td>0.05997501</td>
<td>0.7197001</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td>0.755102</td>
<td>0.57017909</td>
<td>2.8508955</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>7</td>
<td>1.755102</td>
<td>3.08038317</td>
<td>21.5626822</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2.755102</td>
<td>7.59058726</td>
<td>15.1811745</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>2</td>
<td>3.755102</td>
<td>14.10079134</td>
<td>28.2015827</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4. Quadrat frequency distribution and computations

The index of dispersion \( s^2/\bar{x} = 2.062/1.244898 = 1.656 \) and index of cluster size \( ICS \)

\[
(s^2/\bar{x}) - 1 = 0.656 \quad (> 0)
\]

which implies clustering. In fact, we can do more, since we observe

\[
X^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{\bar{x}} = \frac{101.0612245}{1.244898} = 81.18
\]

which follows a \( \chi^2 \) \(_{m-1}\) distribution. The value \( X^2 = 81.18 \) is greater than \( \chi^2 \) value 48 degrees of (49 quadrats -1) at the 0.5% level \( (\chi^2_{48} = 76.9688) \), and the resulting \( p - value = 0.00195 \). This leads us to reject the null hypothesis that the observed distribution of B&E is produced by CSR, since we would expect CSR to produce the observed quadrats by chance in less than 0.5% of the realizations.

5.5.3 Nearest Neighbour Tests for CSR

Recall the nearest neighbour event-event distance \( W \) and nearest neighbour point-event
distance \( X \) with cdfs \( G(w) \) and \( F(x) \) respectively. Unfortunately, theoretical distribution functions \( G(w) \) of \( W \) and \( F(x) \) of \( X \) under CSR are hard to determine since they depend on the shape of \( R \) due to edge effects. However, theoretical distributions of \( W \) and \( X \) can be obtained when edge effects are ignored. Consider a circle of radius \( x \) around a randomly chosen point, with area of \( \pi x^2 \), and let the mean intensity of events per unit area be \( \lambda \). Then under CSR, the probability that no events fall in \( \pi x^2 \) is \( e^{-\lambda \pi x^2} \). Hence, under CSR, the distribution function \( F(x) \) of nearest neighbour point-event distance \( X \) is

\[
F(x) = Pr(X \leq x) = 1 - e^{-\lambda \pi x^2}, x \geq 0.
\]

Under CSR, it implies that \( \pi X^2 \sim \text{Exp}(\lambda) \), or equivalently \( 2\pi \lambda X^2 \sim \chi_2^2 \), therefore

\[
E(X) = \frac{1}{2\sqrt{\lambda}}
\]

and

\[
\text{VAR}(X) = \frac{(4 - \pi)}{4\lambda \pi}.
\]

Also, if \( X_1, \ldots, X_n \) are independent nearest neighbor point-event distances, then \( 2\pi \lambda \sum X_i^2 \sim \chi_{2n}^2 \).

Under CSR, we get similar results for the nearest neighbour event-event distance \( W \). Explicitly, the distribution function \( G(w) \) is given by

\[
G(w) = Pr(W \leq w) = 1 - e^{-\lambda \pi w^2}, w \geq 0
\]

and \( E(W), \text{VAR}(W) \) are the same as \( E(X), \text{VAR}(X) \). Knowledge of the theoretical distributions of \( W \) and \( X \) under CSR allows us to derive approximate sampling distributions under CSR of major
summary statistics of observed nearest neighbour distances. The distribution theory requires that nearest neighbour distances are independently sampled from study region R and have not been biased by edge effects. We must be careful with a mapped point pattern, that is, when all events in R have been observed, and usually one cannot use all $n$ nearest neighbor distances. If all nearest neighbour distances are taken into account, these are not independent. (Think of an extreme case where there are only two events, so both nearest neighbour distances are identical.) Generally we use a sample of size $m$ nearest neighbour distances, where $m \leq 0.1n$ and $n$ is the total number of events in R. Also, nearest neighbour distances of points located near the edges are biased. To fix this problem we use an edge correction, (say a toroidal correction), or alternatively use

$$\hat{G}(w) = \frac{\#(b_i > w \geq w_i)}{\#(b_i > w)}$$

where $b_i$ is the distance from event $i$ to the nearest point on the boundary of R. (The same holds for $\hat{F}(x)$.)

After performing the edge correction, we can plot $\hat{G}(w)$ or $\hat{F}(x)$ against their theoretical equivalents under CSR, (i.e., $G(w) = 1 - e^{-\lambda \pi w^2}$ or $F(x) = 1 - e^{-\lambda \pi x^2}$).

At least we are comparing with known theoretical forms under CSR. As yet we do not have any formal way to assess the significance of differences in the plot. Where significant differences in the plot suggest departures from CSR, to determine this significance we will use simulation.

Simulating a CSR process of $n$ points (where $n$ is the number of events in the observed process) $m$ times in the same region R results in $\hat{G}_i(w), i = 1, ..., m$ or $(\hat{F}_i(x); i = 1, ..., m)$, i.e., $m$ empirical c.d.f.'s. Here we do not apply the edge correction when calculating these empirical c.d.f.'s.
Simulation average estimates of $G(w)$ and $F(x)$ are

$$\tilde{G}(w) = \frac{1}{m} \sum_{i=1}^{m} \tilde{G}_i(w) \quad \text{and} \quad \tilde{F}(x) = \frac{1}{m} \sum_{i=1}^{m} X_i \tilde{F}_i(x).$$

We define upper and lower simulation envelopes:

$$U(w) = \max_{i=1,...,m} \{\tilde{G}_i(w)\},$$

$$L(w) = \min_{i=1,...,m} \{\tilde{G}_i(w)\}.$$

(The envelopes are applied similarly for $U(x)$ and $L(x)$).

We now plot $\tilde{G}(w)$ (estimated without edge correction) against $\tilde{G}(w)$ and add $U(w)$ and $L(w)$ (similarly this is also done for $\tilde{F}(x)$ and its corresponding $U(x)$ and $L(x)$). Under CSR, the plot should be roughly linear and at 45 degrees, and in the case of clustering, the plot of $\tilde{G}(w)$ against $\tilde{G}(w)$ will lie above the 45 degree line, while in the case of regularity, the plot of $\tilde{G}(w)$ against $\tilde{G}(w)$ will lie below the 45 degree line. $U(w)$ and $L(w)$ help to assess the significance of departures from the 45 degree line in the plot since they have the property that

$$Pr(\tilde{G}(w) > U(w)) = Pr(\tilde{G}(w) < L(w)) = \frac{1}{(m + 1)}.$$

This also shows what value we should use for $m$, i.e. how many simulations we need to run, in order to be able to detect departures at a specific significance level. For the 2004 B & E data the test for CSR based on $G(w)$ appears below as Figure 5.12.
Figure 5.12. Test for CSR based on $\hat{G}(w)$

Since the plot of $\hat{G}(w)$ against the simulated $\bar{G}(w)$ lies above the 45 degree line, modeling the event distributions of B & E occurring in 2004 in the central east area of Ottawa against CSR provides evidence against randomness and suggests clustering or spatial dependence. So the hypothesis that the spatial distribution of 2004 B & E events in central east Ottawa is CSR is rejected and we conclude there is evidence of clustering.

5.5.4 K Function Tests for CSR

As we saw earlier, under CSR,

$$K(h) = \pi h^2$$

and
\[ \hat{L}(h) = \sqrt{\frac{\hat{K}(h)}{\pi}} - h. \]

Since it is difficult to establish the theoretical sampling distribution of \( \hat{L}(h) \) and hence \( \hat{K}(h) \) when considering the problem of constructing an appropriate edge correction, again simulation will be used. Like the simulation of \( \hat{G}(w) \) and \( \hat{F}(x) \) described in the previous section, we can simulate a CSR process of \( n \) points \( m \) times, and construct the simulation envelopes

\[ \begin{align*}
U(h) &= \max_{i=1,\ldots,m} \{ \hat{L}_i(h) \} \\
L(h) &= \min_{i=1,\ldots,m} \{ \hat{L}_i(h) \}.
\end{align*} \]

We include these envelopes in plotting the actual observed \( \hat{L}(h) \) against \( h \). The significance of peaks or troughs is assessed on the basis that

\[ Pr(\hat{L}(h) > U(h)) = Pr(\hat{L}(h) < L(h)) = \frac{1}{(m+1)}. \]

As before, this also explains what value we should use for \( m \) (i.e. how many simulations we need to run) in order to be able to detect departure at a specific significance level.

If we take \( m = 99 \), it follows that

\[ Pr(\hat{L}(h) > U(h)) = Pr(\hat{L}(h) < L(h)) = \frac{1}{(99+1)} = 0.01. \]

That is, we are constructing a one-sided test at the 1% significance level that rejects the null hypothesis, \( H_0: \) CSR in favour of the alternative hypothesis \( H_A: \) Clustering at distance \( h \), when \( \hat{L}(h) > U(h) \) or in favour of \( H_A: \) regularity at distance \( h \) when \( \hat{L}(h) < L(h) \) (Bailey and Gatrell 1995). For B & E cases, to get further evidence for or against complete spatial randomness(CSR) we
plot the estimate of $L(h)$ against the upper and lower simulation envelopes of estimates of $L(h)$ derived for $m = 99$ simulations of a set of points meeting the requirements of complete spatial randomness. Figure 5.13 below summarizes the results:

![Graph showing the estimate of $\hat{L}(h)$ and its simulation envelope for B & E data.](image)

Figure 5.13. $\hat{L}(h)$ and its simulation envelope for B & E data

The estimate of $L(h)$ for the 2004 B & E events distribution is above the upper simulation envelope for distances less than 1000 meters (except for a few distances less than 200 meters) indicating that there is some randomness in the events distribution at this scale and, for distances more than 1000 meters, the estimate of $L(h)$ for the B&E events distribution is below the lower simulation envelope. To conclude, this means that the clustering is among neighbourhood blocks where the distances are less than 1000 meters, as we can note from Figure 5.14 below:
Figure 5.14. B & E incidents in Central East Area of Ottawa, 2004

and, for distances of more than 1000 meters, the street pattern affects the distribution of the B&E

events, which is shown in the Figure 5.15 below:
Figure 5.15. B & E incidents in Central East Area of Ottawa, 2004
Chapter 6: Conclusion

6.1 The main contribution

The main contribution of this thesis is that it brings together theory for several tools from GIS, spatial statistics, and data mining to present a statistical analysis framework for the analysis of criminal incidents; it provides an illustrative example by examining Ottawa Police Services data for 2001 and 2004.

6.2 Data source:

6.2.1 Source of Tabular Data:

1. Crime data was provided by Ottawa Police Services, who gave 30% of their data from 1/1/2000 to 12/31/2004. The City of Ottawa provided information about city wards, population and number of households from the year 2000 to the year 2004.

2. Geographic data (Maps and other geographic features)

Ottawa Police Services provided the maps of districts, neighbourhoods, and zones. A variety of geographic data was obtained through the City of Ottawa website http://www.city.ottawa.on.ca

6.2.2 Limitations:

1. The data represents a 30% random sample of data from the Ottawa Police Service Records Management System for data collected between 2000-01-01 and 2004-12-31. Records were limited to incidents considered to be Crimes Against Persons, Crimes Against Property or other Criminal Code Violations according to the UCR 2.1 Violation Coding Structure (Canadian Center for Justice Statistics/Statistics Canada). So the results of the analysis are illustrative only of the suggested analysis framework.
2. Privacy issues and geo-referencing data prevent one from getting data about individuals. Hence, there is no capability of analyzing beyond the spatial analysis without special permissions.

3. Abu Dhabi police data is not complete nor it does not contain any x-y coordinates hence it could not be used to illustrate analytical procedures outlined in this thesis; hence analysis has just been done for Ottawa Police data augmented by the x-y coordinates of the crimes. This is for illustrative purposes only.

6.3 Future directions re analysis of Ottawa crime data

There is extra work still needed regarding the post-processing stage and the interpretation of results. For illustration purposes we have focused analysis on B & E incidents. There is much additional data that we have not analysed in this thesis.

For the lattice data analysis of Chapter 3 we could collect information about smaller divisions of Ottawa such as census tracts or use the breakdowns made by Ottawa Police which divides the city into districts, neighbourhoods, and zones. We could then analyze the data at this finer level to examine spatial dependence between these divisions and then model according to the spatial regression techniques in order to get more sophisticated results.

For spatial point patterns illustrated in Chapter 5, the 'discovered' clusters of crimes would require further research to explain such clustering. The street pattern that appears in the data might be natural or it may be the result of bad geo-coding.

We did not explore the effect of time on the spatial data. It would be useful to implement spatio-temporal analysis of the point pattern process to see whether or not there is a space-time
clustering.

6.4 Implementation for Abu Dhabi Police General Headquarters crime analysis:

For Abu Dhabi Police General Headquarters crime analysis, in order to implement the analyses outlined in this thesis, there are two crucial types of data that must be available.

6.4.1 Tabular data

As found previously, the tabular data consists of information about a criminal activity accompanied with some addresses. The address is then geo-coded into $x-y$ coordinates. Abu Dhabi Police General Headquarters already uses The Uniform Criminal Information System (UCIS) which provides the data, but it needs to be modified in order to insert the x-y coordinates corresponding to each criminal record in the system. The Global Positioning System (GPS) can very efficiently do this job. Moreover, GPS is included in the TETRA technology which Abu Dhabi Police General Headquarters have started to use, so it would be easy and advantageous to obtain coordinates provided through this system.

6.4.2 Geographic data

Geographic data simply is the layers that are used by the GIS system, which includes roads, land parcels, block clusters, neighbourhood boundaries, district boundaries, metropolitan boundaries, and other such things. The Department of Information Technology in Abu Dhabi Police Headquarters has the ability to get (or even to design) these kinds of data. Moreover, Abu Dhabi
Police General Headquarters may cooperate with other governmental establishments that use such types of data, like the Abu Dhabi municipality, the Department of Planning & Economy - Abu Dhabi, and the Ministry of Economy & Planning.

Finally, it would be important to use sophisticated GIS software which integrates with the relational database system that exists in Abu Dhabi Police Headquarters.
References

Anselin, L. (2005), *Exploring Spatial Data with GeoDa™: A Workbook*, Spatial Analysis Laboratory (SAL), Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL. [https://www.geoda.uiuc.edu/documentation.php](https://www.geoda.uiuc.edu/documentation.php)


City of Ottawa, January 2006 [http://www.city.ottawa.on.ca](http://www.city.ottawa.on.ca)


127

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Bibliography


limits of partially ordered Markov models", Methodology and Computing in Applied


## Appendix A

<table>
<thead>
<tr>
<th>No</th>
<th>TYPE OF CRIME</th>
<th>DATE</th>
<th>DAY</th>
<th>TIME</th>
<th>FOLLOW_UP</th>
<th>LOCATION DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>7/5/2004</td>
<td>Mon</td>
<td>0</td>
<td>FOUNDED NOT CLEARED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>2</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>7/9/2004</td>
<td>Fri</td>
<td>450</td>
<td>CHARGED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>3</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>7/25/2004</td>
<td>Sun</td>
<td>2215</td>
<td>FOUNDED NOT CLEARED</td>
<td>Single Home/House</td>
</tr>
<tr>
<td>4</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>8/4/2004</td>
<td>Wed</td>
<td>640</td>
<td>FOUNDED NOT CLEARED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>5</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>8/4/2004</td>
<td>Wed</td>
<td>1940</td>
<td>CHARGED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>6</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>8/6/2004</td>
<td>Fri</td>
<td>2230</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>7</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>8/12/2004</td>
<td>Thu</td>
<td>1315</td>
<td>FOUNDED NOT CLEARED</td>
<td>Single Home/House</td>
</tr>
<tr>
<td>8</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>8/16/2004</td>
<td>Mon</td>
<td>845</td>
<td>FOUNDED NOT CLEARED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>9</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>8/17/2004</td>
<td>Tue</td>
<td>1000</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>10</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>8/23/2004</td>
<td>Mon</td>
<td>114</td>
<td>CHARGED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>11</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>8/26/2004</td>
<td>Thu</td>
<td>0</td>
<td>FOUNDED NOT CLEARED</td>
<td>Private Property</td>
</tr>
<tr>
<td>12</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>8/26/2004</td>
<td>Thu</td>
<td>45</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>13</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>8/26/2004</td>
<td>Thu</td>
<td>53</td>
<td>FOUNDED NOT CLEARED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>14</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>9/3/2004</td>
<td>Fri</td>
<td>1700</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>15</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>9/4/2004</td>
<td>Sat</td>
<td>2240</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>16</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>9/4/2004</td>
<td>Sat</td>
<td>1742</td>
<td>FOUNDED NOT CLEARED</td>
<td>Single Home/House</td>
</tr>
<tr>
<td>17</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>9/5/2004</td>
<td>Sun</td>
<td>1930</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>18</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>9/5/2004</td>
<td>Sun</td>
<td>902</td>
<td>FOUNDED NOT CLEARED</td>
<td>Single Home/House</td>
</tr>
<tr>
<td>19</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>9/9/2004</td>
<td>Thu</td>
<td>1230</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>20</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>9/10/2004</td>
<td>Fri</td>
<td>1001</td>
<td>FOUNDED NOT CLEARED</td>
<td>Single Home/House</td>
</tr>
<tr>
<td>21</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>9/11/2004</td>
<td>Sat</td>
<td>900</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>22</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>9/16/2004</td>
<td>Thu</td>
<td>244</td>
<td>CHARGED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>23</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>9/28/2004</td>
<td>Tue</td>
<td>1900</td>
<td>FOUNDED NOT CLEARED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>24</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>10/4/2004</td>
<td>Mon</td>
<td>654</td>
<td>FOUNDED NOT CLEARED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>25</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>10/23/2004</td>
<td>Sat</td>
<td>418</td>
<td>CHARGED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>26</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>11/4/2004</td>
<td>Thu</td>
<td>1910</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>27</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>11/28/2004</td>
<td>Sun</td>
<td>2021</td>
<td>CHARGED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>28</td>
<td>B&amp;E NON-RES.PROP/COMMERCIAL</td>
<td>11/28/2004</td>
<td>Sun</td>
<td>2021</td>
<td>CHARGED</td>
<td>Other Commercial/Corpo</td>
</tr>
<tr>
<td>29</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>12/17/2004</td>
<td>Fri</td>
<td>400</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
<tr>
<td>30</td>
<td>B&amp;E RESIDENTIAL PROPERTY</td>
<td>12/18/2004</td>
<td>Sat</td>
<td>1000</td>
<td>FOUNDED NOT CLEARED</td>
<td>Dwelling Unit</td>
</tr>
</tbody>
</table>

Table A.1. Break and enter incidents in the neighbourhood of Dalhousie Street

**Data Source:** Ottawa Police. This data represents a 30% random sample of data from the Ottawa Police Service Records Management System for data collected between 2000-01-01 and 2004-12-31.