

2006

Shoeprint analysis: A GIS application in forensic evidence

Michael Anthony Walnoha Jr.
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Walnoha, Michael Anthony Jr., "Shoeprint analysis: A GIS application in forensic evidence" (2006).
Graduate Theses, Dissertations, and Problem Reports. 4278.
<https://researchrepository.wvu.edu/etd/4278>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Shoeprint Analysis:
A GIS Application in Forensic Evidence

Michael Anthony Walnoha Jr.

Thesis submitted to the
Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Arts
in
Geography

Ge Lin, Ph.D. Chair
Gregory A. Elmes, Ph.D.
Kenneth C. Martis, Ph.D.

Department of Geology and Geography

Morgantown, West Virginia
2006

Keywords: Forensic evidence, crime mapping, GIS applications

Abstract

The overall intent of this study is to illustrate how GIS and crime mapping methods can be applied to forensic evidence to better understand and comprehend spatial patterns that exist in these data. This study bridges common crime mapping principles such as hot spot mapping, exploratory data analysis, and spatial statistics to spatial forensic evidence investigation. In particular, forensic shoeprint evidence is examined and spatial relationships are analyzed using both exploratory and confirmatory statistical analysis. It is found that crime mapping principles can be indirectly related to shoeprint evidence mapping. Exploratory spatial data analysis is extremely helpful in breaking up large sets of shoeprint evidence into smaller and manageable sets for spatial forensic analysis. This work is one of few studies to incorporate shoeprint evidence in a crime mapping context. With that in mind the author hopes that this study has shed some light on this subject to advance these methods in this field.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Figures	iv
Chapter 1: Introduction	1
Chapter 2: Literature Review	5
2.1 Geographic Traditions	5
2.2 Crime Mapping	9
2.3 Shoeprint and Spatial Implications	12
2.4 Literature Synthesis and Objectives	15
Chapter 3: Methodology	17
3.1 Data and Study Setting	17
3.2 Methods	19
Chapter 4: Exploratory Data Analysis	22
4.1 Forensic Evidence Exploration	23
4.2 Exploratory Data Analysis	28
4.3 Time-Space Criteria	32
4.4 Case Study: E 46	37
4.5 Case Study: Rectangle	39
Chapter 5: Hot Spot Analysis	44
5.1 Case Study: N/K	48
5.2 Case Study: N 846	51
Chapter 6: Conclusion	55
Appendix A	60
References	63

List of Figures

Figure 1 Example Shoeprints Images	18
Figure 2 Kernel Density Map	24
Figure 3 Residential Recovery Rate Map	25
Figure 4 Shoeprint Frequency Distribution Histogram	26
Figure 5 Total Count of Shoeprints in London Metropolitan Area	29
Figure 6 Adidas 12 Matched Shoeprints	30
Figure 7 Close up of Adidas 21	31
Figure 8 Self-Exclusion Algorithm Graphic	32
Figure 9 N 175 Shoeprints After Self Exclusion Algorithm	33
Figure 10 Result Matrix from Time-Space Algorithm	34
Figure 11 Classification of Hot spots Map	36
Figure 12 Matched E 46 shoeprints at the 15 day time constraint	38
Figure 13 Matched Rectangle shoeprints at the 15 day time constraint	40
Figure 14 Moran's <i>I</i> Scatter Plot	47
Figure 15 Example Cluster Map	48
Figure 16 N/K Moran's <i>I</i> Scatter Plot	49
Figure 17 N/K Cluster Map	50
Figure 18 N 846 Moran's <i>I</i> Scatter Plot	51
Figure 19 N 846 Cluster Map	52

Chapter 1

Introduction

The importance of geographic information science (GIS) applications in various fields cannot be denied. Many academic and industry fields have benefited from the implementation of GIS. Although early applications were in resource management and urban planning, recent GIS applications have been integrated into many practical fields including; criminal justice, natural disaster management, health care delivery, ecological and environment studies, and various fields in social sciences. In the field of criminal justice in particular, the importance of GIS was recognized by the National Institute of Justice and established by the Crime Mapping Research Center in 1997. One of the center's primary goals was to enhance the ability of researchers and practitioners to analyze complex spatial crime patterns (Harries, 1999).

The use of GIS in the criminal justice field has its roots in an earlier generation of police crime mapping. Recent improvements in GIS technology have pushed this use to new heights. In the past, law enforcement agencies used 'pin mapping' to pinpoint crime locations on a hardcopy map (Harries, 1999). While these crime maps were useful for showing geographic patterns, as crime events became denser these maps became difficult to read. An additional weakness of 'pin mapping' was the limited amount of spatial analysis that could be performed on the data. However, with the implementation of GIS, functions such as, hot spot analysis, time-series mapping, and pattern detection have become an integral part of crime pattern analysis. Additionally, the increased availability of digital data is opening new opportunities for detailed spatial analysis of criminals' social behaviors, which in turn, can direct resources to where they are most needed

(Craglia, Haining, & Wiles, 2000). One policy area in which the use of geographic information systems has made significant progress is in crime analysis and mapping at city and neighborhood levels, where neighborhood characteristics and socioeconomic conditions have also be linked to crime hot spots.

While there is a clear advantage to using a GIS for crime mapping, potential GIS application areas in forensic investigation and crime scene analysis have rarely been explored. Forensic science is an interdisciplinary study that applies scientific knowledge to evidentiary material, and provides crucial information for the courts of law. Forensic evidence is often collected at a crime scene or from a suspect. While there are many potential GIS applications, such as crime scene digital mapping, archiving, and profiling, this research assess the utility of spatial aspects of impression evidence, such as shoeprint markings or shoeprints in crime investigation. Shoeprints are just one of many types of forensic evidence that are recovered from a crime scene.

While there are many different types of forensic evidence, two common types are trace evidence and biological evidence. Trace evidence is often found at a crime scene in small amounts, examples include hairs, glass shards, clothing fibers, soils, gunshot and explosives residue. Biological evidence, on the other hand, usually consists of evidence for which DNA analysis can be performed. Typical examples are blood, saliva, or other body fluids, and even trace evidence such as hair can be processed for DNA analysis.

Even though fingerprints are the most prevalent impression evidence, shoeprints, tool marks, and tire tracks are also routinely collected at a crime scene. Shoeprint evidence is sometimes overlooked and misinterpreted. Lead latent print examiner for the Scottsdale Police Crime Lab, Dwayne S. Hilderbrand, has written a book on this

overlooked evidenced titled, “Footwear, The Missed Evidence (1999)”. In his work, he stresses the importance of shoeprint evidence and discusses why this type of evidence is frequently overlooked.

According to Hilderbrand, criminals must enter and exit crime scene areas, therefore, it should be reasonably assumed that they might leave shoeprint impressions. Criminals have become smarter and wiser by wearing gloves to avoid leaving fingerprints, and masks to avoid identification. However, they are rarely aware of, or make little attempt to conceal footwear (Hilderbrand, 1999). Because of the blatant disregard for concealing footwear, investigators have been able to use shoeprint evidence in the apprehension of criminals. A case in point is from the Ohio Bureau Criminal Identification and Investigation annual report 03-04:

March 2, 2003

“A footwear impression found at a crime scene can establish not only the presence of the wearer, but in certain instances, the time frame of the deposit of the impression. Such was the case in a Greene County house burglary processed by the Trace Evidence Unit. Examination of the kicked-in door at the point of entry revealed three partial footwear impressions in the dust. Comparison of a suspect’s shoes revealed a matching wear pattern. Thus, not only was the wearer at the scene, he was at the scene when the door was forced in. Faced with the evidence and with no explanation as to how his shoe impressions got on the door, the suspect pleaded guilty to the crime” (p.31, 2004).

Unfortunately, sometimes impression evidence can be overlooked, contaminated, or destroyed when a crime scene is improperly secured or is disorganized. Shoeprints can be a crucial piece of evidence if it is properly collected, preserved, and examined. This evidence can become an important part of proving or disproving a suspect’s presence at the crime scene. However, forensic evidence such as fingerprints and DNA are routinely collected, their recovery rates (percent of crime in which evidence is recovered) tend to be low (Leist, 2005). Therefore, integrating under-utilized forensic

evidence, such as shoeprints, bears potential for solving cases that have little DNA and fingerprint evidence.

In conclusion, the primary goal of this study is to bridge the apparent knowledge gap that exists between crime mapping and forensic investigation so that both fields can benefit from each other through GIS and other information technologies.

Chapter 2

Literature Review

2.1 Geographic Traditions

The core of this research lies within the spatial science tradition within Geography. Its basic framework was constructed in the 1950s with the quantitative revolution. During this time, geographers supplemented traditional descriptive statistics with the application of inferential statistics (Cloke, 1991). Many geographers of this time used a positivist approach to their research, where knowledge was gained through experience, but it also argued that knowledge must be verifiable and replicable (Johnson, 1983). Many geographers who conducted this type of research studied regions and became interested in explaining their differences based on space. To explain the spatial arrangements of people, places and activities spatial scientists used techniques such as modeling, mathematics, and statistics. This work, in particular, makes use of locational analysis, an inductive approach which attempts to explain spatial patterns of shoeprint evidence. Additionally, this research utilizes exploratory data analysis and spatial statistics to discern spatial patterns within a given phenomenon. A brief history of spatial science is presented in this literature review.

In the 1960s the spatial science approach to studying geography was becoming more prevalent within the discipline. This tradition projected the discipline into a new direction and critiqued the past approaches of geography. Spatial science distanced itself from idiographic approaches, dismissed the empiricist philosophy, and introduced a nomothetic approach which was concerned with the universal and the general, to counteract the idiographic approach.

Spatial science became very attractive to geographers that were looking for a scientific approach to geography. “The expression of research results in mathematical or statistical form, in a way that implied precision, replicability, and certainty” (Johnson, 1983). Many using quantitative analysis felt they had ‘hard evidence’ to support their work and some geographers called themselves ‘statistical geographers’. Until spatial science became a prominent geographic approach, the academic world questioned the validity of geography because of the lack of theory and emphasis on description. Many geographers adopted this approach and hoped to get better recognition from the academic community. “Science was academically and socially respectable, and so was social science; it was useful, and geographers perceived that they, by becoming more scientific, could be useful too...” (Johnson, 1983). By integrating hypotheses, theories, and methods to the discipline, scholars felt geography was becoming equally recognized with disciplines like psychology, mathematics, and sociology.

Within the positivist framework spatial scientists became interested in studying regions based on space and distance. Spatial scientists used quantitative methods, mathematical techniques, theorems, and proofs to explain spatial distributions of people and activities. Spatial science proposed the idea that space was the key factor that shapes behavior. The most important addition to the spatial science tradition was the scientific method.

The scientific method, as known in spatial science, has two distinct methods, the inductive and deductive approach. Most spatial scientist employed the inductive approach. The inductive approach begins with perceptual experiences of the world from the researcher. Next, the research collects unordered facts about the research. In the next

step, the researcher begins working with actual data, and explores facts to identify generalized behavior, using quantitative methods. Now, the research has ordered and classified facts which he or she can now make generalizations. After a generalization was made, the researcher constructs a hypothesis, and from that, theories and laws are formed. The last step in the inductive process is the explanation, which is usually geographic knowledge produced as a paper or a journal (Johnson, 1979).

The deductive approach is not dissimilar from the inductive, as the research begins with predetermined ideas based on experience. The next step is to develop a theory, or image of real world structure. Next, the researcher must form a quantitative model, a formal representation of the image. After the model is complete, the researcher must form a hypothesis and then proceed to experimental design, which entails measurement, definition, and classification. Now, the researcher must collect data, and then use quantitative methods to verify the hypothesis. At this point, the process could return to the theory stage if the results are not supported, or if successful, the supported theory becomes a law. Lastly, the research can provide an explanation of his research through journals or other academic mediums (Johnson, 1979).

While spatial science was a significant improvement within the geographic realm, still, many in the discipline felt it had some shortcomings. Some criticized it for its use of mathematics, and its geometric languages filtered out any social or ethical questions. Others thought spatial science 'failed to see beyond the map' (Cloke et al. 1991). With some distrust for spatial science within the discipline, some turned to new ideas that would inevitably spawn behavior and humanist geography.

Recently, geography has recognized and accepted applied geography as a ‘tradition’ within the field. Applied geography is the application of geographical knowledge and skills to the solution or resolution of problems within society (Johnson et al., 2000). Applied geography is a ‘different’ geography when comparing it with previous traditions. Applied geography does not have a set philosophy, meaning there is no standard epistemology, ontology, or methodology like the other traditions.

Geography has been applied, and has been solving problems since the late 1800s, and is now practiced in both the private and public sectors. The private sector consists of geographers who work for businesses. Private engineering firms could potentially employ geographers to perform GIS analysis and related tasks. The public sector has two main areas that use applied geography; planning is one, and the other is natural resources. As relevant to this work, crime analysts and forensic investigators can utilize GIS and geography in an applied environment to solve problems within society.

This research is firmly rooted in the spatial science tradition because of its emphasis on distance, space, and statistical methods by using the inductive approach. This study draws on both location theory and locational analysis to help explain spatial patterns and how phenomena are related to each other based on space. With the utilization of GIS in geography in the past 30 years geographers have applied geography in ways that can directly help solve society’s problems. Applied geography is evident within urban planning, crime analysis, and watershed analysis, all of which apply geography in a manner which helps society.

In order to apply the aforementioned tradition, this research relates two separate entities within the field of criminal justice. In an attempt to link common practices in

crime mapping with forensic evidence, shoeprint evidence is especially intriguing because few researchers have documented the spatial relationship of forensic evidence in a crime mapping context. Much of the crime mapping field is devoted to detecting areas of high crimes, or clusters of events into so-called hot spots, and confirming the existence of hot spots. Forensic impression evidence, on the other hand, is collected and usually stored in an image database, without explicit spatial reference. To help connect these two distinct fields, the following literature review will first concentrate on crime mapping and secondly forensic shoeprint evidence. Finally, these two areas will be synthesized in order to identify the major the objectives for the research project.

2.2 Crime Mapping

Crime is not uniformly spread across space and it is often clustered in some areas and absent in others. People are aware of this phenomenon and they avoid certain areas and gravitate toward others. Their choices of neighborhoods, stores, and recreation are dictated, to a certain degree, by the understanding that their chances of being a victim of crime is greater in some places rather than in others (Eck, 2005). Police are aware of the fact that crime is not randomly distributed in space. More police resources tend to be allocated hot spots or areas where crime is more concentrated. In contrast, few resources tend to be distributed in low crime areas (Craglia, Haining, & Wiles, 2000). Hence, the study of geographic distribution of crime is not only relevant to a person's daily life, but is also relevant to public safety management.

From a crime mapping context, hot spot mapping is the spatial representation of areas with high densities of crime. One hot spot mapping method is to draw contour lines around areas of high concentration of crime. The maximum concentration contour forms

a polygon, which can be used within GIS to calculate the values of the polygon with respect to other attribute layers (Eck, 2005). For instance, one might use crime incident data to define a contour hot spot, then, use GIS software to calculate correlation such as income, population, or other variables. “Comparing hot spot contour polygons and their correlates across urban areas could provide insight into factors associated with crime incidence above a researcher-determined threshold” (Garson, 2001, 473). The researcher must be aware the search for correlates is laden with potential difficulties emerging from the ecological fallacy. Alternatively, one can use a density map, such as various kernel methods to provide a continued crime density surface for a study area (Gatrell, et. al 1996).

Crimes occur in different types of place and space, as well as at different scales. As a corollary, hot spots occur at different scales, too. Hot spots exist as streets, neighborhoods, regions, and cities. While some hot spot characteristics exist generally across scales, the relative importance of factors that influence hot spots at the street scale differ from those that influence hot spot regions, or hot spot cities. Additionally, the actions that must take place to intervene in and reduce these varying hot spots must be different. The level at which crime or hot spots are examined is dictated by the way the question, or problem is presented. These approaches differ on the level of analysis, or the size of the geographic area of crime about which one is concerned (Eck, 2005).

Identifying the appropriate scale of geographical analysis is critical to understanding the nature of crime and determining what action to take. There are a number of theories that exist to explain hot spots and they are useful in crime mapping because they aid in the interpretation of crime data (Eck, 1998).

Once the scale of analysis is determined, hot spot theories can be integrated into the study to help further explain why crime happens at certain places. Corresponding to spatial scale or geographic coverage, there are a number of hot spot theories: place, street, and neighborhood theories. Place theory explains why crime events occur at specific locations. It deals with crimes that occur at the lowest level of analysis, specific places. This theory attempts to answer questions like why are burglaries occurring at one home and not another? Street theory deals with crimes that occur at a slightly higher level than specific places, usually streets or blocks. This theory attempts to answer questions such as why are drugs being sold on these streets, but not another street? One theory of crime attempts to explain neighborhood differences. At a higher level than place or street, neighborhood theory deals with larger areas. It is generally concerned with why are gangs present in a particular area, and absent in others? Finally, repeat victimization theory asks, why are certain victims targeted continually? It should be noted that this theory can operate at any of the scales discussed, and can be represented geographically as a point, line, or polygon.

In addition to the theories and levels of analysis different types of hot spots exist as well. “The most basic form of a hot spot is a place that has many crimes” (Eck, 2005). An area that has repeated crime incidents are known as repeat place hot spots. The underlying cause of this type of hot spot is explained in the routine activity theory. “The routine activity perspective studies the processes and patterns associated with the path of the offender and the victim in time and space within an environment suitable for criminal activity” (Rossmo, 1999). “Structural changes in routine activity patterns can influence crime rates by affecting the convergence in space and time” (Cohen & Felson, 1979,

p.598). Repeat place hot spots are not alone in this category, repeat street hot spots and repeat victimization hot spots exist as well. When considering the aim of this study, repeat place hot spots are most suited for mapping forensic shoeprint data because the majority of crimes are committed by repeat offenders.

The development and implementation of crime mapping has been beneficial to law enforcement agencies in the past, and has given crime analysts various methods to visualize crime. However, what crime mapping has not done is evidence based crime solving. Investigators are vocal about the fact that evidence solves crime, whether it is circumstantial, conclusive, or trace evidence. Mapping forensic evidence has the potential to solve crime by using spatial relationships based on existing hot spot methods.

2.3 Shoeprints and Spatial Implications

Forensic crimes scenes are usually documented in two distinct ways (Gardner, 2005). While crime scene sketching and crime scene mapping may seem the similar, they are actually quite different. A crime scene sketch is usually a rough drawing that concentrates on measurements and evidence placement. Crime scene mapping, on the other hand, is concerned with documenting the size of the scene, elements within the scene, and 'fixing' evidence using some common mapping methods. The most common methods include: triangulation, rectangular, and polar coordinates (Gardner, 2005). Both of these methods are excellent ways to document a crime scene, yet neither of them incorporates spatial relationships between evidence.

Several types of forensic evidence can be retrieved from a crime scene (Gardner, 2005). According to the Metropolitan Police in London, 2004 evidence recovery rates are as follows: DNA 9%, fingerprints 25%, shoe marks 12.4% and tool marks 6.2%.

Clearly, shoeprints are a valuable evidence resource and are recovered from a substantial amount of crime scenes. According to Girod (1996) crime scene examiners in the region of Neuchatel in Switzerland find usable shoeprints in almost 30% of the burglaries investigated. The majority of these burglaries are committed by repeat offenders living within the region. When forensic evidence is collected it is usually photographed and entered into an image database with little or no spatial reference. At the most, street address data might exist. This limits the potential value of the evidence because forensic evidence is intrinsically spatial, and spatially enhanced information has the potential to increase the probability of solving a crime.

In forensic science the importance of image databases has been known for a long time (Geradts, 2002). Common impressions that can become forensic evidence are fingerprints, toolmarks, shoeprints, and tire tracks. There are many successful cases where impression evidence at the crime scene has been used to match available evidence in databases. Partially because of its importance and notoriety, many criminals tend to avoid leaving fingerprints at crime scenes, but they normally cannot avoid leaving other marks. Hence, other databases (e.g., shoeprints, tool marks, handwriting, cartridge cases, and bullets) are also important for casework (Geradts, 1995). As acknowledged by Hamm (1989), “shoe impressions or more correctly track impressions have been around for a very long time, well before the identification of fingerprints”.

Shoe impressions and fingerprints are some types of evidence a criminal may leave behind at a crime scene. Such impressions or other evidence have great potential to the investigator. The fine scratches in outsoles can be used to identify a shoeprint. A database of shoes of suspects in comparison with a database of shoeprints can be valuable

for solving crimes. Often the outsole design has to be classified for the shapes that are visible in the design. Not only can impressions lead to the identification of the make and style of a particular shoe, they can also provide approximate sizing of that shoe and the direction of gait, and other actions such as dragging or scuffing (Ashley, 1996).

Currently, shoeprint databases are rarely used. However, some police regions use these databases as first-generation databases. Searching in the databases on content is still a difficult task, since shoeprints are often blurred (Geradts, 2002).

Girod (1995) lists three common practices that use a shoeprint database:

1. determine the brand and type of shoe that left the shoeprint at the crime scene,
2. a comparison of these shoeprints with a suspects', and
3. demonstrate that a particular impounded shoe left the shoeprint at the scene.

As noted previously, shoeprint impressions have great potential to the crime scene investigators. However, some crime investigators neglect to see their potential.

According to Hilderbrand (1999), shoeprints are sometimes overlooked or under-evaluated because of a lack of training and education in the proper search for, collection, and preservation of evidence. The failure to properly collect this type of evidence revolves around the aforementioned two reasons, but the lack of success in finding this evidence is often due to: 1. a result of the investigator not believing that the impressions can be found at the scene after people have walked over the scene, 2. incomplete searches of the crime scene, 3. prevailing weather conditions, and 4. the impression has been intentionally destroyed (Hilderbrand, 1999). If the potential of shoeprint evidence was more widely appreciated, there is a greater likelihood that investigators would collect it.

Once the investigators collect it, it is more likely that it would be stored in databases and its potential could be recognized.

2.4 Literature Synthesis and Objectives

In order to synthesize these two topics, one must borrow elements from each field and intertwine them to expand and enhance them. This synthesis is comprised of two objectives. First, the research attempts to expand the crime mapping field by integrating forensic evidence into set practices, such as hot spot mapping. Although this study only focuses on one aspect of forensic evidence (shoeprints) any of the aforementioned evidence could be substituted in its place (e.g. DNA or fingerprint). The second objective is to enhance forensic evidence analysis by explicitly integrating spatial information, such as integrating geographic location with forensic shoeprint evidence. Investigations are conducted on a local scale, which will attempt to enhance shoeprint matching rates by incorporating geographic and crime matching principles from a few locations or crime scenes. This study applies to journey-to-crime and distance decay models that will provide insight into criminal's spatial decisions.

In summary, the intent of this research is to combine the effectiveness of hot spot crime analysis with forensic shoeprint evidence found at crime scenes. While hot spot mapping and shoeprint evidence classification has been noted previously rarely, has a study combined both aspects in the field of geography. This study makes use of exploratory spatial data analysis to help bring crime mapping concepts into a forensic mapping context. In addition, results from some exploratory data analyses can help to conceptualize and generate hypotheses for forensic investigation. The potential of this

study is intriguing because the coalescing of crime mapping (hot spot) and forensic evidence is a fairly new concept in both geography and forensic investigation.

Chapter 3

Methodology

3.1 Data and Study Setting

The Bigfoot database is maintained by the Metropolitan Police Department in London, England. They have maintained this database since 1997. The department retrieves shoeprints from approximately 10% of all crime sites and more than 10,000 shoeprints are retrieved per year. For the year 2004 the department recovered 10,096 recovered shoeprints from approximately 100,000 burglaries.

Crime and police data are typically sensitive information when the general public is concerned. The nature of this data is no different. Sensitive data, while interesting and unique, can sometimes prove difficult to obtain. Such is the case with this forensic evidence. Initially, a verbal sharing agreement was established in the fall of 2005 with WVU researchers Michael Walnoha and Ge Lin and the London Metropolitan Police. A written contractual agreement was signed later which specified the terms in which the London Metropolitan Police could share this data. Personal information about the victim(s) and criminal(s), address locations, and shoeprint images were omitted. Sensitivity issues ultimately turned into time issues, and the data was received later than expected, in February 2006. This issue is particularly important when the researcher does not collect their own data.

Due to the confidentiality and sensitivity issues, there were some limitations placed on the data. As previously mentioned personal information and address locations were omitted, but this did not hinder the study. Shoeprint images were not present either. This data would have been useful for determining how the London Metropolitan Police

confirmed ‘matched’ prints. Since the researchers did not collect the data, there was some uncertainty about retrieval methods and how the match was determined. The database was limited in this aspect because the only indication of a match was a coded letter and number representing a ‘matched’ shoeprint (e.g. A 23). Additional research was conducted as to how shoeprint data was collected and integrated into a database.

A variety of methods can be employed to collect impression evidence from crime scenes. The common methods to retrieve shoeprints are electrostatic lifting devices, plaster casts, and adhesive lifters (Gardner, 2005). Once the prints are recovered they are scanned into an image database where experts can perform matching analysis. The researchers did not perform, nor were they qualified to perform this step in the forensic process therefore, some uncertainty was generated as to how a match is determined.

After consulting with forensic experts they assured us a ‘match’ was based on many like characteristics of the shoeprint. Size, brand, and tread design (etc.) were all considered when determining a match. Forensic experts use special image software to verify a match and Figure 1 illustrates the basic criteria for a match. Images similar to Figure 1 are example prints that forensic experts use to compare unique and intricate patterns when they are deciding whether a shoeprint is matched or not. The shoeprints on the left are obviously the same print declaring a match therefore, the prints would be coded the same (e.g. A 15). However, the prints on the right are not. Each of these prints would be given a separate code and could be matched later.



The shoeprint database is for data collected in 2004 and has 10,096 total shoeprints. Both point (shoeprint) and polygon (London boroughs) data were projected on the British National Grid. The shoeprints are georeferenced to the x y coordinates according to this gridsystem. Besides coordinates and coded shoeprints, additional attribute data about the crime consists of; date, time, borough in which the offence occurred, type of offence, and modus operandi. All of this data lies within the London Metropolitan region and was recorded and entered by the London Metropolitan police.

The study area extends 36 miles from east to west, and 27 miles from north to south. The London Metropolitan region consists of 32 boroughs, however, burglary data from the central London (City of London) was suppressed since the metropolitan police do not have jurisdiction for this area. The analyses are conducted at both the point and borough levels for 32 boroughs.

3.2 Methods

The field of forensic evidence mapping lacks a standardized methodology. The following methods of spatial analysis will be among the first implemented in this setting. Normally, when one has little prior knowledge, a good starting point is exploratory data analysis. Here, exploratory spatial data analysis is featured throughout the study by using the ArcGIS and GeoDa. The emphasis is not on investigating a particular crime or a set of crime incidents, but on how GIS and spatial analysis can be meaningfully implemented to map shoeprint evidence and to aid in forensic investigation. The following sections describe the approaches to meet the intents that were introduced in the literature review.

Objective 1 methods. A variety of existing approaches are used to map shoeprint evidence and show geographic patterns. The first approach uses exploratory data

analysis to categorize and display shoeprint evidence. For example, recovery rate of forensic evidence can be represented and displayed spatially. The recovery rate is defined by dividing the total number of recovered shoeprints by the total number of crimes. The geographic distribution of the recovery rate will reveal the spatial relationships between the total number of recovered shoeprints the total number of crimes. If recovery rates across boroughs are fairly evenly distributed, it implies that the spatial distribution of matched shoeprints cannot be attributable to potentially biased evidence collection from one borough to another borough.

Even though an area has a high recovery rate, it does not necessarily mean that there is a high concentration of matched shoeprints. Therefore, the second approach is to examine the spatial relationship between available and matched shoeprint evidence, so that the certainty of matched shoeprint evidence attributed to one or few suspects can be enhanced. Borrowing the concept of journey-to-crime, a time-space constraint can be set out to narrow the scope of matched evidence. If a large geographical distance exists between two sets of potentially matched shoeprints, the level of uncertainty is much higher than if they were in close proximity. Existing geographical models, such as distance decay, kernel density and nearest neighbor methods are used to supplement time-space inquiries. Once a particular matched shoeprint does not meet time-space constraint, it should be excluded from the analysis. For this reason, a self-exclusion algorithm is developed and discussed in detail in the next chapter.

Objective 2 methods. In contrast to point-based analysis in the previous method section, the primary focus of this method is on rate by areal units or boroughs. Here the rate is based on matched shoeprints divided the recovered shoeprints at the borough level.

In order to classify or group a large number of the matched shoeprints that meet the time-space criteria, area based exploratory spatial data analysis methods, such as local Moran I_i can be used. The global Moran's I tests for spatial independence. If a set of matched shoeprints are deemed spatially dependent or correlated, local Moran I_i can be used to classify the detected pattern into hot spots, cool spots, and positively and negatively correlated patterns.

Chapter 4

Exploratory Spatial Data Analysis

Given the lack of a standardized methodology and available literature concerning forensic evidence mapping, it is necessary to determine the extent of forensic evidence mapping, the concept of a hot spot within this context, and how to utilize them for forensic investigation. The purpose of this chapter is to document the characteristics and distribution of shoeprint data by using exploratory data analysis. The following sections briefly sketch the main ideas behind EDA and to determine time-space criteria of matched shoeprints for hot-cool spot analysis. Finally, the results are presented with a number of case studies.

Exploratory data analysis (EDA) is an extremely valuable starting point for this investigation. When there is little knowledge about a phenomenon, investigation is often started with exploratory data analysis using its extensive graphic and descriptive statistic capabilities. The intent is to discover inherent patterns in the data and to generate hypotheses by imposing as little prior structure as possible (Tukey, 1977). For example, a histogram can be used to determine the frequency distribution of matched shoeprints, so that we can visually identify 'natural breaks' to decide if a set of shoeprints can be analyzed by either EDA or statistical methods.

When an EDA method is furnished with spatial or map analysis, it becomes exploratory spatial data analysis (ESDA). For instance, we do not know, as a priori, what would be considered geographic closeness of shoeprints. ESDA is likely to provide clues because we can visually inspect a number of matched shoeprints that are considered to be near each other and measure their distance. In addition to pattern and knowledge

discovery by describing and visualizing spatial distributions for various variables, ESDA can be used to identify atypical locations or spatial outliers. In the context of shoeprints, two or three offenders may happen to wear an identical brand and type of shoe, and thus leave nearly identical shoeprints over a wide geographic area. However, if the geographic locations, dates, and times are inspected, there may be shoeprints that are highly related to other isolated shoeprints. In this way, spatial clusters, hot spots, and spatial patterns can be identified (Anselin, 1998).

4.1 Forensic Evidence Exploration

As previously noted, very little literature exists concerning forensic evidence mapping from a geographical perspective. To establish 'operational' definitions of clusters and hot spots from a forensic evidence perspective, it is necessary to make several distinctions between forensic evidence and crime mapping. From a crime mapping perspective, a hot spot is conventionally defined by a spatial representation of areas with high incidence, rates or densities of crime. Various methods of spatial representation of hotspots exist, point mapping being the most common (Jefferies, 1999). Additional hot spot mapping methods include: thematic boundary mapping, interpolation, and spatial ellipses (Chainey and Ratcliffe, 2005). Figure 2 represents an example hotspot using kernel density, a method that uses a grid function to determine areas with a high concentration of points.

While a hot spot is somewhat similar in forensic evidence mapping, one substantial distinction should be noted. An area that has a high concentration of crime does not necessarily mean that it has a high concentration of forensic evidence. A large number of crimes may yield little or no forensic evidence, while a small number of

crimes can yield a large amount of evidence. This depends on how much evidence remains at the crime scene and how much effort is devoted to recover the evidence.

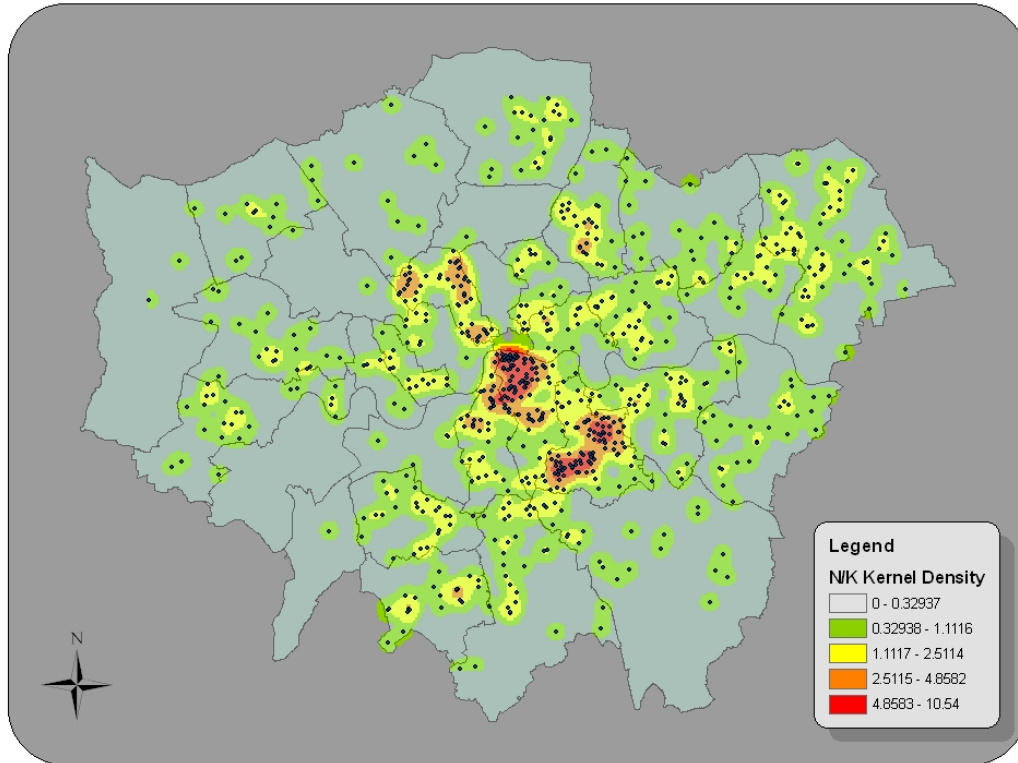


Figure 2: Kernel Density Map

Although a direct relationship between hot spots of crime and forensic evidence have yet to be established, Figure 3 shows the spatial relationship between recovered evidence and total crime at the borough level. In this case, the recovery rate is calculated from shoeprints recovered from residential burglaries divided by total residential burglaries. From this map, it appears the recovery rate is generally lower in the center of the metropolitan area and higher on the periphery. However, the global Moran's I suggests that there is no spatial clustering of the recovery rate (p -value: 0.060). The statistical results suggest that the recovery rate distribution is spatially independent.

Therefore there is no direct relationship and or bias toward a particular borough when evidence collection was performed.

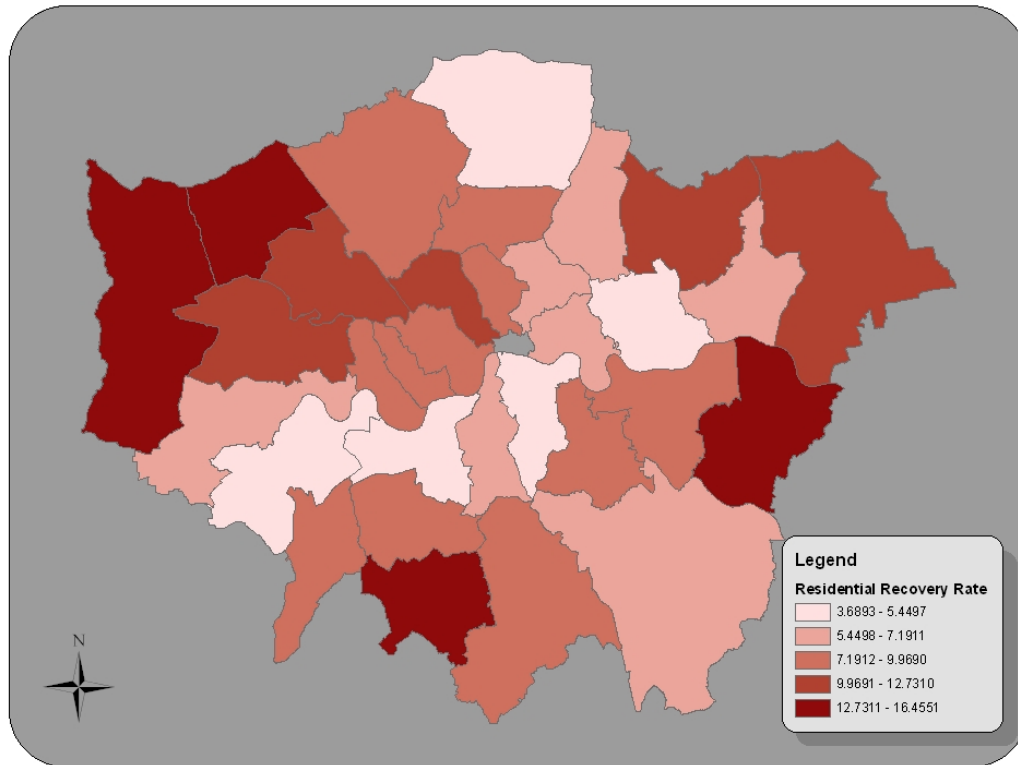
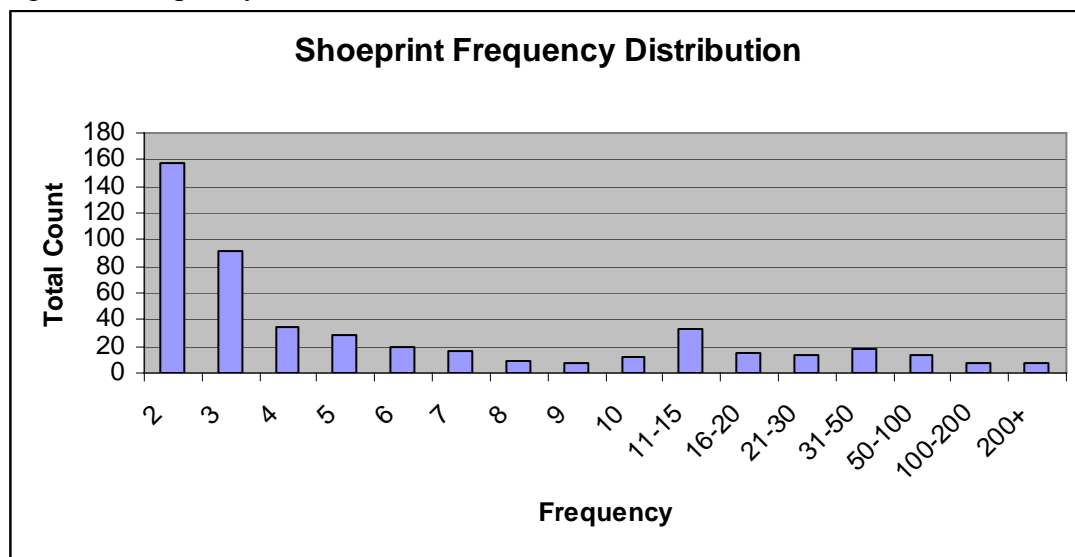


Figure 3: Residential Recovery Rate

The second distinction between shoeprint evidence and crime is that we cannot match crime without matching evidence at the same time. A crime hot spot may be attributed to many suspects, but a matched pair of shoeprints may point to a single suspect, especially if the prints are rare enough. Although shoeprints are far less unique than comparing fingerprints, they are analytically similar. Like a set of matched fingerprints point to a single suspect just as a set of matched rare shoeprints can point to a single or a few suspects. However, some shoeprints may not be rare. Although we can infer that two or three pairs of matched shoeprints are highly likely to be from one suspect, as the number of matched shoeprints increases, it becomes less reasonable to

assume they are from one or two suspects. Figure 4 shows the frequency distribution of the total number of shoeprints after deleting 886 unknown shoeprints. The histogram does not show the frequencies of 900 unique shoeprints (one occurrence) because they would suppress visual effects of other matched frequencies. There are 91 sets of matched shoeprints with only 3 matches. For example, C15 shoeprints were recovered only three times, and there are a total of 91 similar cases. On the other extreme, we find that 8 sets of matched shoeprints have at least 200 ‘identical’ shoeprints corresponding to their burglary incidents. It is relatively straightforward to suggest that three matched shoeprints corresponding to three burglary incidents are highly likely to be committed by one individual. It is unreasonable to suggest that 300 matched shoeprints in 300 burglary incidents are likely to have been committed by one individual. To reduce uncertainty as to whether or not shoeprints are associated with one another, journey-to-crime theories were applied when considering our criteria.

Figure 4: Frequency Distribution



In order to infer a set of matched shoeprints belong to one or, at most, a few suspects, we can borrow the concept of journey-to-crime to set up time-space constraints, which presents the final distinction between crime and forensic incidents. Journey-to-crime literature suggests that criminals tend to commit crime in geographic locales that are familiar to them (Rossmo, 1999). Two likely locales are neighborhoods near their residential and employment locations. For instance, we might assume that a burglar may only burglarize a house within one mile of his residential location. However, if he burglarized twice, one toward the east, and one toward the west of his residential location, we would need two miles to cover two matched shoeprint incidents. In spatial journey-to-crime profiling, suspect location is a reference point to a crime location; in spatial forensic profiling, an evidence location is a reference point to another evidence location.

In order to infer from one shoeprint evidence location to another, we rely on: 1) exploratory data analysis, and 2) expert opinion by interviewing forensic investigators. Exploratory data analysis suggests that it is relatively easy to make spatial inference from matched shoeprint data with less than 6 crime incidents. Given there are only a few matched cases, most of these shoeprints consist of one or two clusters, or have one cluster and a few unseemingly related incidents far from the cluster. It was suggested from an interview with forensic investigators, that a set of matched shoeprints corresponding to more than 7 crime incidents would be hard to infer without some space and time constraint. The remaining sections discuss how to reduce highly related shoeprints beyond 7 matches

4.2 Exploratory Data Analysis

The beginning stages of the analysis presented a seemingly overwhelming task to explore a large amount of data (figure 5). The dataset consisted of a total of 10,096 recovered shoeprints from burglary crime scenes in the London Metropolitan Area for 2004. To properly assess this data, EDA was used in order to narrow the focus or micromanage the data. While exploring this data, a set of basic criteria was determined to help classify shoeprints so the number of prints in each category could become manageable. The first step was to identify apparent geographic clusters. Once a small cluster was identified the date duration and distance attributes were considered among each point(s). If any points were related by small distance proximity and date duration, offence types were considered to identify additional relationships. Finally, time of day and modus operandi (MO) were secondary criteria at this stage to help confirm any common links.

The first step in the exploratory stage was to investigate a small set of observations that had an apparent geographic cluster. This was achieved by selecting a set of matched shoeprints from attribute data. There was no previous inclination on how to first select these points. A random set of prints were selected based on a number of low occurrences. Once a cluster was identified, distances were measured from each point to gather an estimation of what distances existed between recovered shoeprints. For example, it was found that shoeprints R18 had a cluster of 5 shoeprints within 0.9 mile in rectilinear distance, among them, 4 were within 16 days. Shoeprints A21 had a cluster of 4 shoeprints that cover a geographic distance of 3 miles in rectilinear distance. If there



Figure 5: Total number (10,096) of shoeprints in London Metropolitan Area

was any suspicion that these points were related, modus operandi and type of offense (eg. residential or non residential burglary) were included in this exploratory analysis. The following specific examples show that an informal, exploratory case study would help to identify patterns and generate hypotheses.

A small set of observations of matched shoeprints are much more manageable at this stage of the analysis. Initially, a set of matched Adidas shoeprints were explored. These shoeprints, coded AD 21 have a total of 8 observations (Figure 6). This particular shoeprint is interesting for two reasons: 1) it appears only 8 times out of 10,096 observations, deeming it a 'rare' mark, meaning the probability of it coming from the

same offender is high, and 2) within these 8 marks there is an apparent geographic cluster of 4 marks within 2 miles (Figure 7).

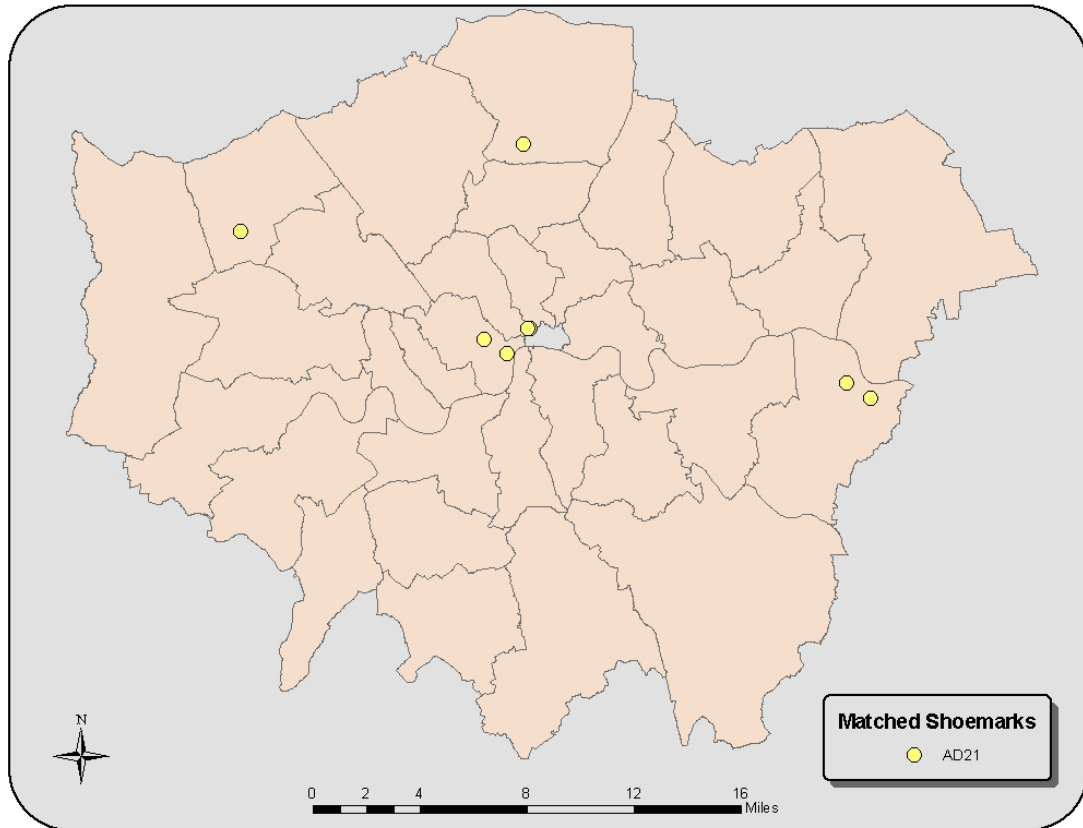


Figure 6: Adidas 21 matched

Upon further review of this cluster, it was speculated that the likelihood of these marks being left by the same offender was quite high. This speculation was attributed to the exploration of the aforementioned criteria. Once the cluster was identified, the attributes of these marks were examined closely. The cluster of 4 marks (Figure 7) were within 2 miles of each other. Secondly, the duration of time in which the evidence was recovered from the scene was considered; three of the four marks were within 12 days of each other. Lastly, the type of offense was considered. All marks were recovered from residential burglaries. These 4 steps aided to help produce a set of primary criteria that

could be used to develop a ‘profile’ to base our analysis upon. A secondary criterion was also considered which consisted of the time of day in which the crime was committed and the modus operandi (MO). In this case the MO is useful only as a secondary characteristic because only a short, vague description of the crime is present, such as: “suspects kicked in front door of flat and entered premises and searched through occupier’s property”. Specific details of the crime(s) are not released due to confidentiality issues.

This set of criteria works well with a small set of observations, but it is not very conducive to analyzing a set of data with 50 or more observations. It is very difficult to explore a set of matched shoeprints with a large number of observations, therefore, the previous criteria had to be amended for this purpose.

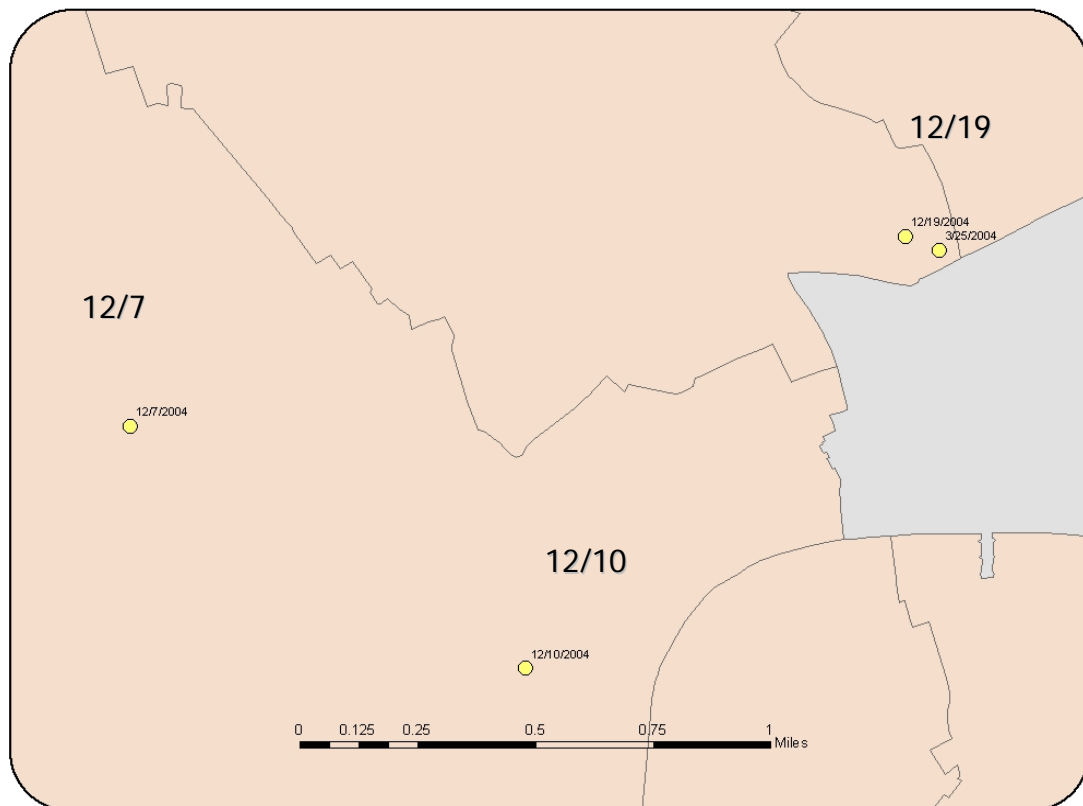


Figure 7: Close up of Adidas 21

4.3 Time-Space Criteria

In order to assess a large set of observations a time-space constraint was implemented to narrow the number of observations. To establish the criteria, two general assumptions were based on the characteristics of matched shoeprints: 1) matched shoeprints were likely left behind by the same offender(s), and 2) within a small time-space constraint a set of clustered matched shoeprints could implicate a single offender. Based on these assumptions a self-exclusion algorithm was constructed to eliminate points that did not fit the time-space criteria. A time constraint of 15, 30, and 45 days, and a distance constraint of 1, 3, and 5 miles were provided for the algorithm. This process was replicated nine times.

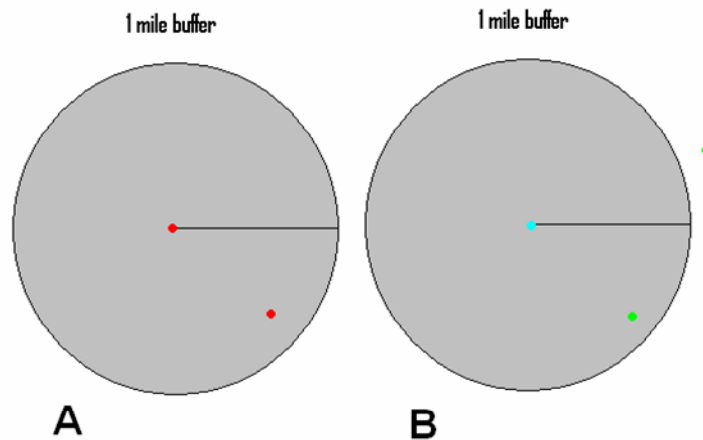


Figure 8: Self-exclusion algorithm graphic

Figure 8 illustrates how the self-exclusion algorithm eliminates points based on the criteria. The center point in Figure 8a represents the point of interest. The center point will remain only if there is a point within the respective distance buffer (1 mile) and if the dates are within the specified time frame. Figure 8b illustrates the opposite, where the center point is excluded because there are no points within the specified time frame,

even though a point lies within the specified buffer. In order for a point to remain it must satisfy both criteria. The end result of this algorithm eliminates points that have low probability of clustering.

Figure 9 shows an example of the self-exclusion algorithm as it relates to the set of matched N 175 shoeprints. This set of points is based on the 15 day 1 mile time-space constraint. The white triangles represent the total count of matched shoeprints (201), and the points represent the remaining number of prints (38) based on the aforementioned exclusion criteria.

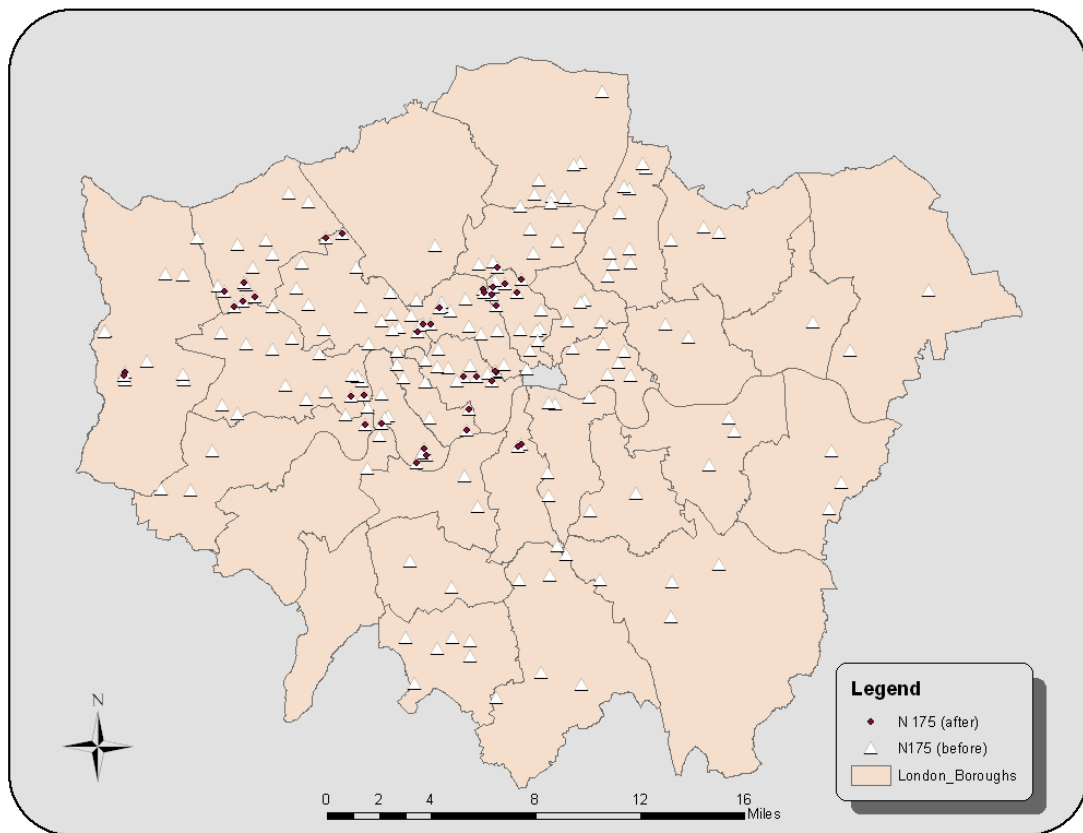


Figure 9: N 175 Shoeprints after algorithm

Figure 10: Results from the space-time algorithm			
	1 mile	3 miles	5 miles
45 days	5427(41.1%)	6214(32.2%)	6578(28.6%)
30 days	4373(52.5%)	5122(44.4%)	5693 (38.2%)
15 days	2003(78.3%)	2621(71.5%)	3077(66.6%)

Figure 10 illustrates the number of shoeprints that remained after the algorithm processed the total number of shoeprints based on each set of criteria. As expected, the greatest reduction resulted from the smallest time-space constraint of 15 days and 1 mile (lower left) from the original of 9,210 to 2,003 or 78.3% reduction from the original sample. On the other extreme, the widest space-time constraint yielded the largest number of observations with only 28.6% sample reduction.

Upon reviewing the results, it was concluded that the 45 day 5 mile time-space constraint was the least useful due to its tendency to produce points that did not generally cluster together. This time-space constraint had the lowest reduction percentage and the average nearest neighbor distance was calculated to confirm this. There was an average distance of 1,152 meters for shoeprints N846. It should be noted that there is no time constraint on this distance, unlike the self-exclusion algorithm, only distance is considered. Several other matches were also calculated for average nearest neighbor distances and their range was from 0.5 m to 1 mile. From this analysis it is determined that most neighboring shoeprints lie within a mile radius. This was the basis of the distance constraint. The nearest neighbor is not necessarily linked to the same type of crime/shoeprint burglary within a particular time frame, only the distance between the

events were linked. From this analysis it was decided that the smaller time-space constraints were likely to yield related clusters and they were examined first.

Given the largest time-space constraint(45 days, 5miles) it was obvious that it would yield the most prints, but not necessarily yield the most favorable spatial clustering. Therefore, the results of the 15 day 1 mile constraint were the first to be examined. From this exploratory stage a hotspot cluster classification scheme was implemented. After reviewing the observations, three classes of hotspot would be assessed in the analysis, as depicted in Figure 11. First, Co-location - a set of matched shoeprints retrieved from the same crime scene. From a 'zoomed-out' view these clusters appear to lie on top of each other; the marks are clustered very tightly together and are almost certainly left behind by the same offender(s). Second, Borough Specific Hotspot - a set of matched shoeprints that exist and lie within by a borough boundary. These hotspots are apparent after the algorithm was run. A later case study demonstrates the likelihood of a small cluster of prints in one borough after the self-exclusion are likely related. These types of hotspots exist in a particular borough or two and are not found uniformly or sporadically across the metropolitan area. Third, Conventional Hotspot - is a hotspot of matched shoeprints that exists throughout the metropolitan region. This type is best suited for hotspot statistical analysis so that significant clusters can be identified. A number of case studies are presented to expand on these types of clusters. These case studies examine shoeprints based on the 15 day time criteria.

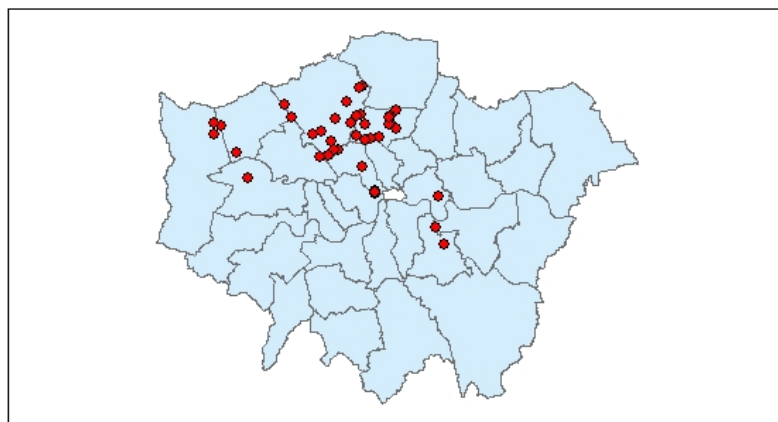


Figure 11: Illustrates 3 different classes of hotspots from 3 different sets of matched shoeprints.

Top to bottom: Co-location, Borough Specific, and Conventional Hotspots.

4.4 Case Study: E 46

A total of 23 matched E 46 shoeprints existed before the algorithm eliminated points at each time-space constraint. While the original 23 matched shoeprints are spaced out fairly evenly over the metropolitan area, the set of matched E 46 shoeprints exist mostly as a borough specific hotspot after the algorithm was implemented. This is an interesting hotspot because of the concentrated location of the shoeprints in Ealing Borough. As shown in Figure 12, the first three algorithm runs at 15 days yielded 10 marks at 1 mile, 11 marks at 3 miles, and 13 marks at 5 miles. The E 46 marks are a highly concentrated hotspot because the largest time-space constraint of 45 days with a 5 mile buffer only yielded 14 shoeprints, only 5 more prints than the smallest time-space constraint. This particular print is rare because it occurs at a very low frequency and probability of the prints are left behind by multiple offenders is unlikely. At this point the rest of the primary and secondary criteria are considered. The offence pattern is very consistent, 12 of the 13 prints are present at residential burglaries; providing a means of association when developing a 'profile'. Additionally, 9 of the 13 of the crimes were committed in the early or late evening hours, only a few were committed in the morning. Lastly, the MO is considered to reinforce the possibility of regularity within the criminals' mode of operation (e.g. suspect may always enter through an open window, or break a basement window). As previously noted, the MO provided in the dataset is vague and is only considered as an indication of a possible secondary source for links with the crime description. The MO description of this particular set of matched shoeprints has the term 'window' appear in 10 of the 13 and the terms 'rear' or 'back' in

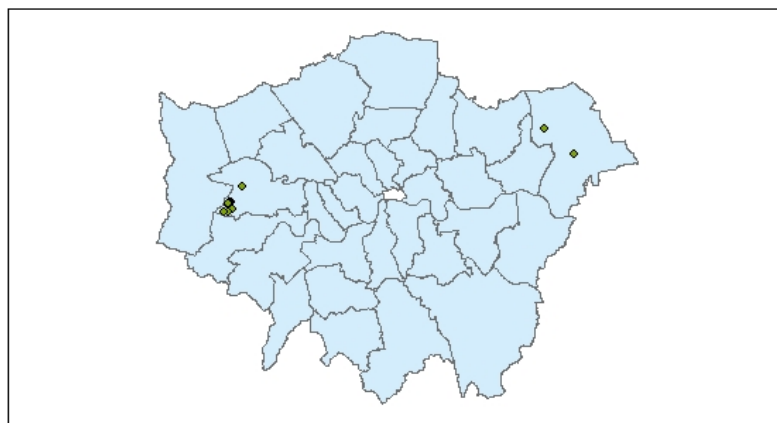


Figure 12: Matched E 46 shoeprints each at the 15 day time constraint.
From top to bottom:
1m-10 points, 3m-11 points, and 5m-13points

8 of the 13 descriptions, indicating a possible consistent point of entry as a rear or back window. Based on the primary and secondary criteria it is highly likely that the matched E 46 shoeprints were left behind by the same offender. If an additional piece of matched forensic evidence (e.g. fingerprints) were recovered from any of the locations with matched shoeprints it is much more likely that a criminal investigation could be performed. This case study shows that self-exclusion algorithm is effective for these types of hot spots.

4.5 Case Study: Rectangle

The code Rectangle refers to a specific geometric pattern, in this case a matched rectangle, not a specific matched shoeprint like E 46. There is a total of 34 matched Rectangle shoeprints before the algorithm eliminated points at each respective time-space constraint. Unlike the set of E 46 shoeprints, Rectangle shoeprints are primarily scattered in the north-west region of the metropolitan area, with a few in the south. Much like the set of E 46 shoeprints, the set of Rectangle prints exist as a borough specific hot spot, yet there is one large distinction between the two sets. This type of shoeprint forms an important hot spot because of what can be inferred by the lack of a 'match' in the traditional sense. While this shoeprint is not matched as a specific type of shoe, it is matched to a specific design or shape. The uncertainty that surrounds the match is interesting because of its spatial pattern. One would expect that this set of prints would have no spatial pattern because they are only a partial print and they are not matched to a specific brand of shoe. However, at the highest time-space constraint they are concentrated mainly in 3 boroughs; Haringey, Harrow, and Islington. As shown in Figure 13, the first three algorithm runs at 15 days yielded 13 marks at 1 mile, 27 marks

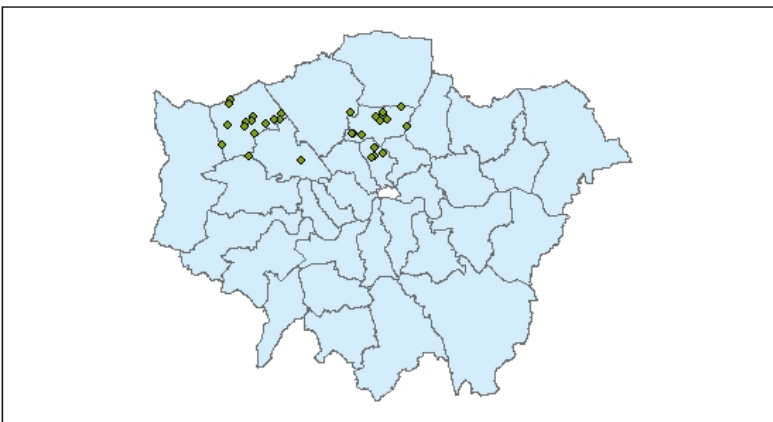
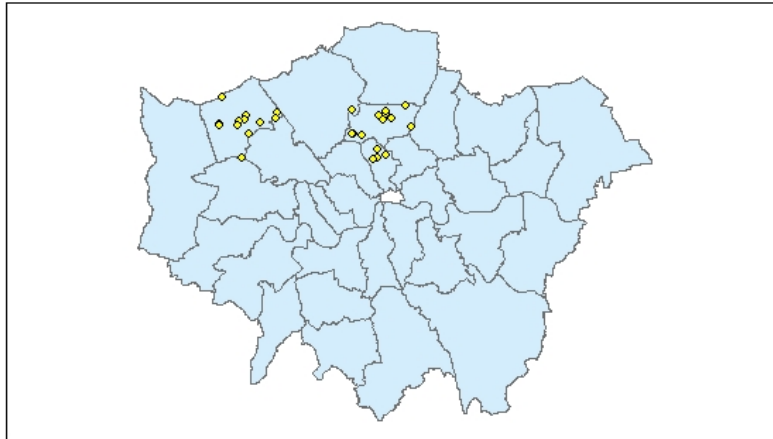
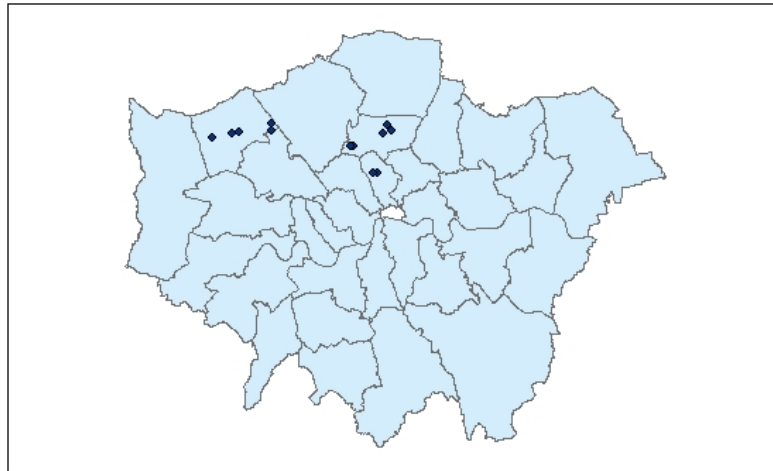


Figure 13: Matched Rectangle shoeprints at the 15 day time constraint
From top to bottom: 1m-13 points, 2m -27 points, and 3m -31 points

at 3 miles, and 31 at 5 miles. The 3 mile space constraint was selected because of its manageability and clustering potential. As previously noted, it was decided that the 1 and 3 mile(s) buffer yielded the best results. The results are clustered in three boroughs, but the majority of the marks are in two boroughs. Based on exploratory data analysis two assumptions can be made: 1) the concentration of the these shoeprints are from one offender whose home is in one borough and works in the other, or 2) the prints are left behind by two separate offenders in their respective boroughs. Of the 27 total shoeprints that remain from the 3 mile buffer, 23 of them are residential burglaries. In this particular set of burglaries, the term 'smashed' or 'forced' appears in 15 of the 27 MO descriptions. It appears that the suspect(s) favored smashing windows (window appears 16 times) or forcing open windows or doors as a point of entry. Additionally, about half of the burglaries are committed in the evening or early morning hours, but there is no convincing cluster in a specific borough. However, when specific clusters in each borough are observed there are some interesting results. A cluster of residential burglaries exists in the borough of Harrow and 7 of the 11 residential burglaries were committed during the workday hours (7:30 am-5:00 pm) and are tightly clustered, revealing some consistency within that particular borough. It can be speculated that there is more than one offender(s) in this borough. However, it can be assumed, due to that consistency, one offender has left behind a majority of rectangle shoeprints in that borough. Additionally, the Borough of Haringey has a similar cluster that exists within the smallest time-space constraint, but is not as convincing as the borough of Harrow. Four of the 5 nonresidential burglaries occur in this borough and all 4 occur in the evening or early morning hours, opposite of the usual workday hours. However, these 4

points are not very tightly clustered and the dates of the offences are not within a convincing time interval. Overall, there is some uncertainty with the set of matched shoeprints generally, but there is one cluster that exists in the Borough of Harrow that warrants further investigation. Most of the residential burglaries that exist in that borough have some consistency within the set criteria and present an above average chance of implementing one offender to those specific crimes. If another piece of matched forensic evidence (e.g. fingerprints) was recovered from any of the residential burglaries within the Borough of Harrow it is likely that there would be a match. The case study further implies that when shoeprints are relatively concentrated in few boroughs, the space-time algorithm may not be the best method to break them up into smaller and consistent clusters.

In conclusion, crime mapping studies are abundant but, conceptual frameworks and hypotheses derived from them cannot be directly applied to forensic mapping. This is attributed to the said distinctions between crime and forensic evidence mapping. To extend them to forensic mapping, both EDA and ESDA have proved useful for generating hypotheses and discovering patterns that have not been reported previously. In this study, it was found that the most valuable matched shoeprint clusters were likely to occur within 1 mile in rectilinear or Manhattan distance, although a 3 mile-range would also cover some clusters where burglary incidents were within 15 or 30 days of each other. Based on these exploratory analyses, a time-space algorithm was developed to systematically capture and break down these spatial patterns. Even though the algorithm was very effective, some shoeprints from popular brands may still have hundreds of occurrences that were deemed to be unmanageable using the current method.

The problem of ubiquitous prints will be further addressed in the next chapter using hot spot analysis.

Chapter 5

Hot Spot Analysis

Chapter five expands on methods to identify significant spatial clusters using local indicators of spatial autocorrelation (LISA). The previous chapter used ESDA to identify significant clusters by exploring selected spatial and attribute data. Those methods were effective for smaller sets of observations, but were impractical for a set of matched shoeprints with a large number of observations. To help better understand and identify spatial hot spots, LISA analysis was executed for matched shoeprint data that have a large number of observations. Two case studies based on 1 mile and 15 day criteria were furnished, and the LISA statistic was implemented in hopes to find a spatial relationship between the total number of shoeprints verses the total number of recovered matched shoeprints by establishing a ratio or rate of recovery.

Based on the notion that most burglaries are committed by repeated offenders in the same region (Girod, 1996) boroughs were used as the aeral unit of analysis. It can be assumed that the ratio of the matched and recovered shoeprints in each borough can provide some clues about a large set of matched shoeprints across the Greater London area. A high ratio in a borough suggests that the set of matched shoeprints is more likely to be concentrated in this borough; a low ratio, on the other hand, suggests that the set of matched shoeprints is less likely to be concentrated in this borough. The LISA statistic is implemented in hopes to find a spatial relationship between the number of matched shoeprints versus the total number of recovered shoeprints in a spatial context. In this way, a relatively large set of matched shoeprints in the study area can be further broken down into smaller and more manageable sets for spatial forensic investigation.

In order to calculate the univariate Moran's I statistic a rate or percentage value must be established. According to Anselin (1995),

Global Moran's I can be specified as:

$$I = \frac{\sum_{i=1}^m \sum_{j=1, j \neq i}^m w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{[\sum_{i=1}^m \sum_{j=1, j \neq i}^m w_{ij}] [\sum_{i=1}^m (X_i - \bar{X})^2 / m]}$$

Local Moran's I_i can be specified as:

$$I_i = \frac{(X_i - \bar{X}) \sum_{j=1, j \neq i}^m w_{ij} (X_j - \bar{X})}{\sum_{i=1}^m (X_i - \bar{X})^2 / m},$$

where X_i is the variable of interest in Borough unit i ($i=1, \dots, m$) in London;

$\bar{X} = \sum X_i / m$, and w_{ij} is the (i,j) -the element of a spatial weight matrix W . Commonly, $w_{ij} = 1$ if boroughs i and j are adjacent and $w_{ij} = 0$ if the two boroughs do not share a common boundary.

Significant and positive Moran's I typically indicates positive autocorrelation, which is attributed to high-high or low-low value clustering. Conversely, a significant and negative Moran's I indicates negative autocorrelation when high values are adjacent to low values. When Moran's I is close to 0, it represents no spatial autocorrelation. (Anselin, 1995; Kitchen and Tate, 2000).

Local indicators of spatial association test a null hypothesis of local spatial randomness by comparing the values (shoeprint occurrences) in each borough with values in neighboring boroughs. Several LISA statistics can be considered, but a local version of Moran's I is particularly useful because it allows for the breakdown of the

patterns of spatial association into four distinct categories. These categories can also be represented with four quadrants in the Moran Scatterplot (Anselin, 1995).

Two of these categories imply positive spatial association. When an above average value in a location is surrounded by neighbors whose values are above average (high-high) or when a below average value is surrounded by neighbors with below average values (low-low).

By contrast, negative spatial association is implied when a high value is surrounded by low neighbors and vice versa (high-low or low-high). Both of these instances are labeled spatial outliers when the matching LISA statistics are significant. Each of the quadrants matches a different color in the LISA map, a map that shows both the locations with significant LISA statistics (i.e., a rejection of the null hypothesis of spatial randomness) as well as the category of spatial association (Anselin, 1995).

In this case, the total number of shoeprints (matched or unmatched) that occurred in each London Borough was summarized by a spatial join. Once this total count was established, another count of a particular matched shoeprint was summarized by each borough. These two counts were divided (matched/total shoeprint count) by borough to establish a percentage or rate of occurrence. Once this is established a spatial weight must be created. For this study, a queen-based contiguity was chosen because it assesses its neighbor's values that share common borders and/or corners. The queen-based contiguity assesses the values of each borough and its neighbors to determine whether there is a cluster of similar values in the surrounding area (Anselin, 2004).

Figure 14:
example of
Moran's I
scatter plot.

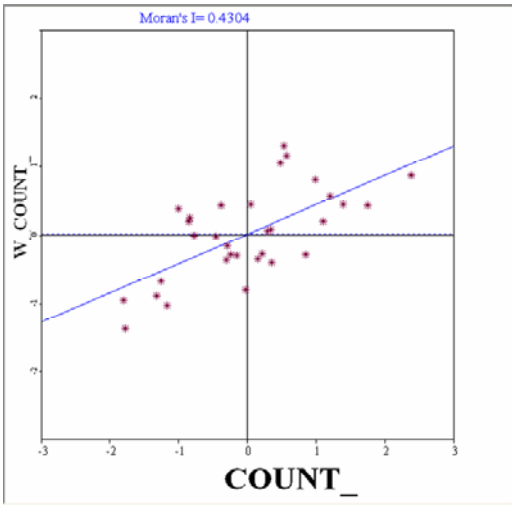


Figure 14 represents an example Moran's I scatter plot. The x axis represents the occurrences of shoeprints per borough and the y axis represents the spatial weight. In this particular case, Moran's I finds that there is positive spatial autocorrelation which can be confirmed by a positive slope and significant p value. The scatter plot illustrates the number of observations that lie within each category. The scatter plot can be read as follows: quadrant 1: high-high, quadrant 2: high-low, quadrant 3: low-low and quadrant 4: low-high.

According to local Moran's I , an area that has positive spatial autocorrelation is an area that possesses high values as does its neighbors. This area can be considered a hot spot, because it has a high concentration of high values in a specific location. Conversely, an area can still be positively correlated even if its values are low, and its surrounding areas are low. This type of correlation would be considered a cool spot, or an area with a low concentration of observations.

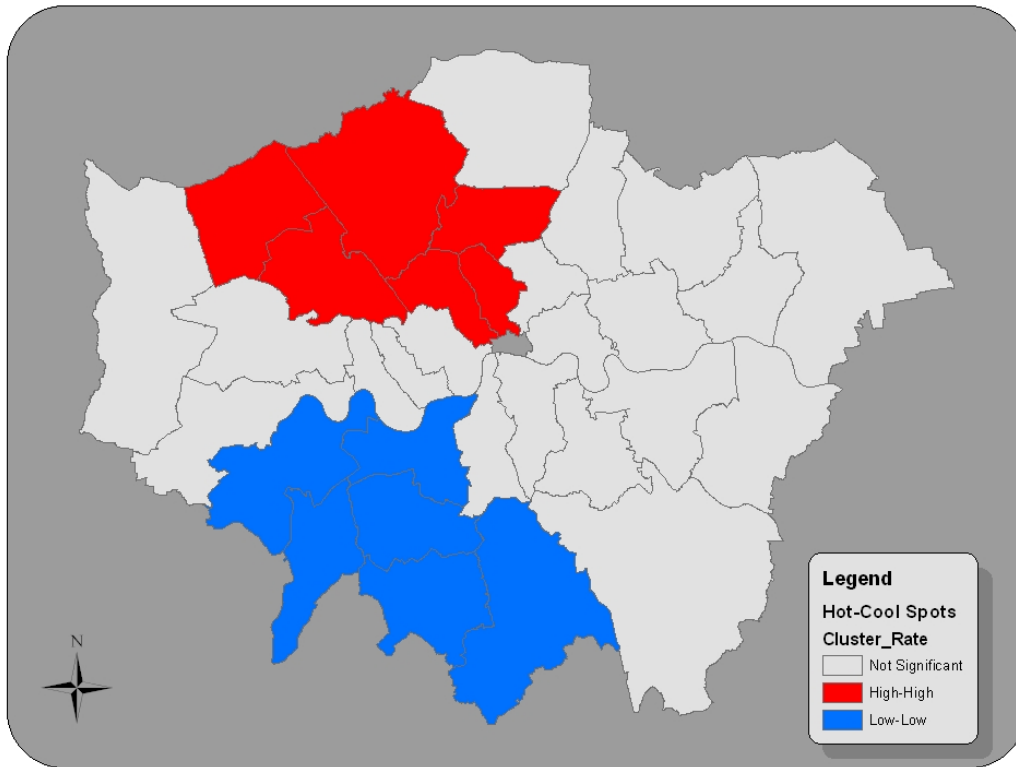


Figure 15: Example Cluster Map

Figure 15 shows positive spatial autocorrelation. This particular map shows the total number of shoeprints (count) per London borough. This map illustrates which boroughs have a high count of recovered shoeprints when compared with other boroughs. This map is useful because it depicts where the most total shoeprints were recovered in 2004. Whether or not those areas are correlated to crime incidents has yet to be seen. A number of case studies are presented to illustrate positive and negative spatial autocorrelation and show relationships between matched shoeprints.

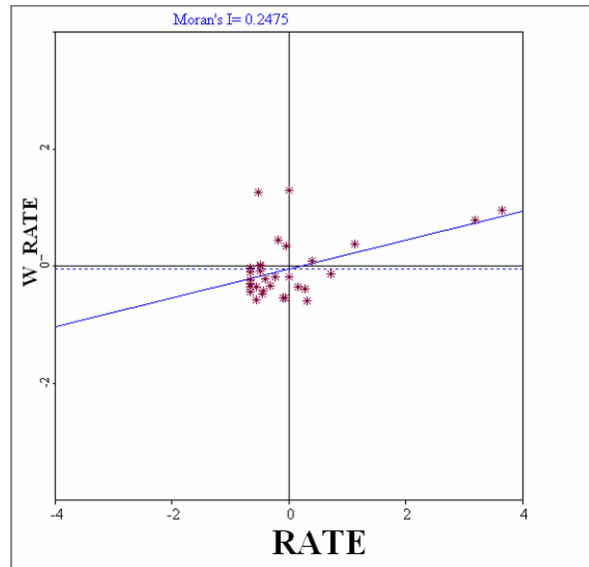
5.1 Case Study: N/K

Once faced with the problem of how to break down a large set of matched shoeprints, a test for spatial autocorrelation using Moran’s *I* is a good starting point. N/K shoeprints were selected because they have a large number of observations (347) even

after the self-exclusion algorithm. The spatial pattern of this set of shoeprints is of interest because they are coded N/K, which indicates a pattern cannot be found. Given the uncertainty surrounding these prints, what kind of spatial pattern would be revealed, and if there is a pattern, would it be positively or negatively correlated?

Results:

Figure 16:
N/K
Moran's *I*
scatter plot



Moran's *I*: 0.2475 p-value: 0.0140

A permutation test based on 499 runs was conducted to test the results for spatial randomness. The results show the Moran's *I* value of 0.2475 with a significant p-value, which reject the null hypothesis of spatial independence. In other words, there is less than 5 percent chance that this pattern was spatially random. This particular set of shoeprints, although having an unknown source possess positive spatial autocorrelation. This means a high percentage of matched shoeprints are close to each other, while a low percentage of matched shoeprints are near each other. Figure 17 displays the local Moran's *I* N/K cluster map. The map shows that hot spots exist in the south and east of central London, while the cool spots exist to the northwest of central London. Even

though high-low and low-high patterns are detected, they are not statically significant according to the global Moran's *I* test.

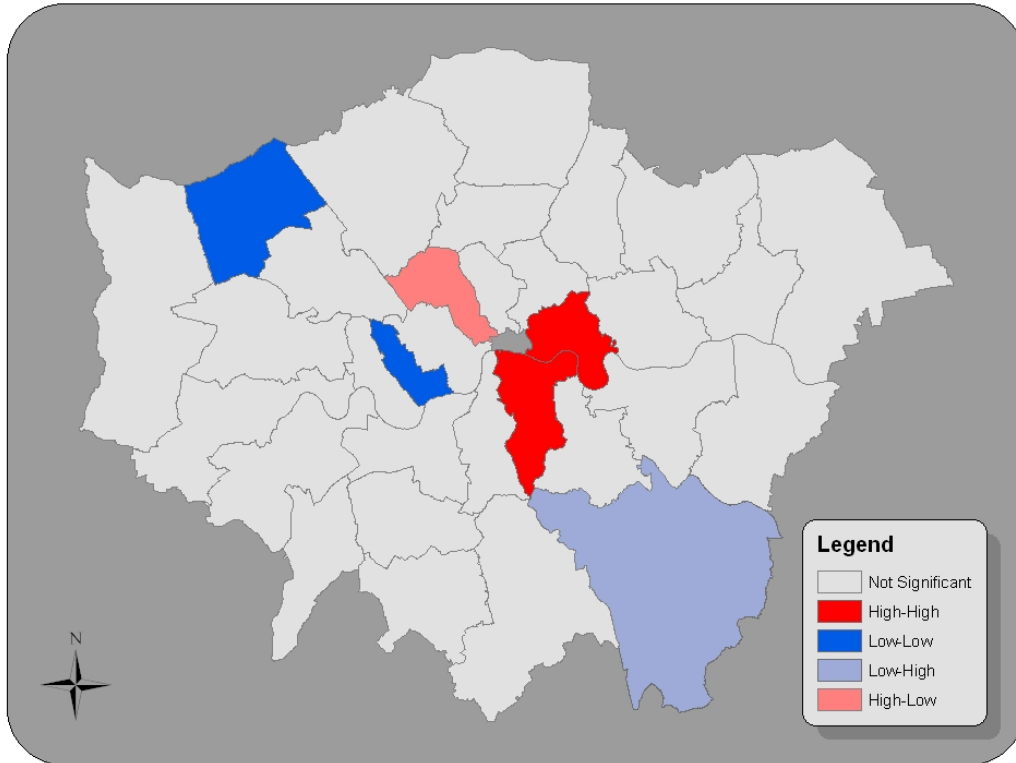


Figure 17: N/K Cluster Map

The patterns displayed in this cluster map can be used for specific targeted forensic spatial analysis. For example, the cool spot suggests there are very few matched shoeprints centered in that region. It can be speculated that the prints recovered in that region are from a few suspects due to the low percentage of occurrences. It should be noted that prints that exist at this stage have already gone through self-exclusion algorithm, and are related by a small distance and time frame. A low percentage (cool spot) means that each set of matched shoeprints are much less likely to happen by chance given its ratio to the total number of recovered shoeprints in that region. However, the

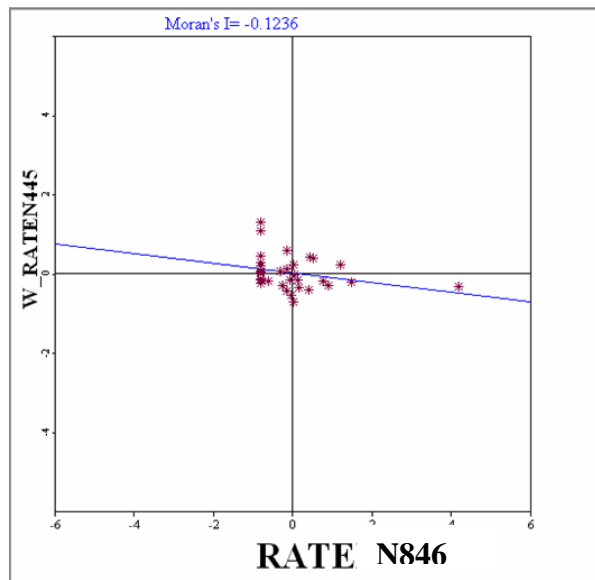
hot spot suggests there is a more than expected number of matched shoeprints clustered in that region. Given this region has a high number of matched shoeprints, it may be less appropriate to implicate a few suspects but still renders further review by forensic investigators.

5.2 Case Study: N 846

The set of N 846 shoeprints has a smaller number of observations (74) after the self-exclusion algorithm. Even though there are significantly fewer shoeprints than the previous case study, 74 prints are still unmanageable when considering the previous EDA and EDSA methods. Additionally, there is no obvious or apparent spatial pattern from the point data. This set of shoeprints can be further investigated by using Moran's I to identify significant clusters or other spatial relationships.

Results:

Figure 18:
N 846 Moran's
 I scatter plot



Moran's I : -0.1236 P value: 0.1460

A permutation test based on 499 runs was conducted to test the results for spatial randomness. The value of Moran's I is -0.1236, and the p-value of 0.146 is not

significant; however the Moran's I scatter plot (Figure 18) suggests negative spatial autocorrelation. According to Anselin (1995) even though the global Moran's I is not significant a local cluster can still be significant. In this case, the set of shoeprints is spatially random and possess negative spatial autocorrelation. Figure 19 illustrates the local Moran's I N 846 cluster map. Clustering is not apparent within the set of matched shoeprints therefore, no significant hot or cool spots exist.

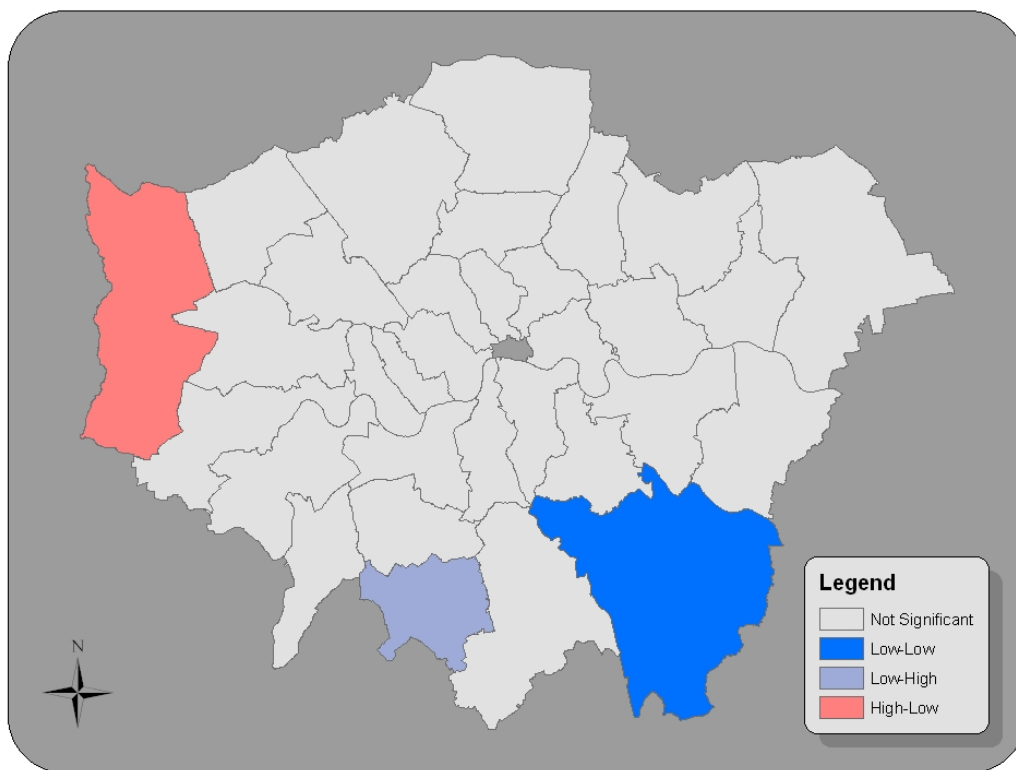


Figure 19: N 846 Cluster Map

The implication of the cool spot scenario has been discussed in the previous case study, yet the significance of high-low and low-high spatial relationships may also shed some light on forensic spatial analysis. In this particular case, negative autocorrelation is obvious. As the high-low relationship suggests, a large ratio of matched shoeprints is

surrounded by no or very few matches. To some extent, the results may be interpreted in terms of the edge effect. All identified clusters are located on the boundary of Metropolitan London. Therefore boroughs with observations are bordered by regions with no observations; and they could skew the results because they are outside the study area. The isolation of a high percentage of matched shoeprints further suggests that shoeprints could be left behind by one suspect within a local area because there are no matched shoeprints of the same kind in the surrounding boroughs. Because of the edge effect, this suggestion is probably an over identification of the results.

Additional concerns with spatial autocorrelation and hot spot clustering can be assessed with other methods. Some methods that might reinforce or supplement this study could be implemented in further studies. (e.g. grid and kernel mapping) Issues of scale and how it relates to correlation could also be assessed as correlation at the borough level in somewhat ambiguous. This research simply implies the possibility of correlation in a conceptual context. Given the lack of forensic investigative knowledge, the author's aim of this was not to identify these clusters in order to solve crime in this environment, but rather to identify and lay groundwork for further investigations.

In conclusion, Moran's I , scatter plots, and cluster maps prove to be useful when further classifying forensic shoeprint evidence after the self-exclusion algorithm was imposed. The four types of local associations may provide clues to different types of shoeprint clustering. The previous case studies showed how both positively and negatively autocorrelated patterns can be used to identify spatial cluster in forensic evidence investigation, but care must be taken not to infer patterns that result purely from the spatial arrangement of the areal units and the inevitable effect of the boundary of the

study area. Finally, spatial autocorrelation is a measure of the similarity or independence of an object with surrounding objects (Kitchen and Tate, 2000). As these prints exist presently, it is reasonable to assume a relationship between matched shoeprints is apparent, but to what extent, it is not known. Further investigations could further determine which spatial unit is most appropriate to assess significant spatial autocorrelation.

Chapter 6

Conclusion

As explained in this thesis, forensic evidence mapping is a relatively new concept on the boundary between the fields of geography and forensic investigation. While both fields exist as well-founded disciplines, rarely have the two been combined previously to form a spatial or geographical perspective on forensic science. This initial study has shown that the combination of these two fields has considerable potential for joint investigation and for future research.

Using exploratory spatial data analysis methods, this study has extended common principles of crime mapping and analysis, such as hot spot mapping, to forensic evidence mapping. In particular, forensic shoeprint evidence from Greater London was examined and spatial relationships were analyzed using exploratory and confirmatory statistics. To assist in crime solving, local spatial autocorrelation statistics, and consideration of models such as distance decay and journey-to-crime, were employed to better understand the spatial relationships that exist in forensic evidence.

Based on these exploratory tools and crime mapping principles, the study has presented an approach that systematically breaks down a large set of matched shoeprints to a set of spatially clustered shoeprints, either in time-space, or in terms of spatial adjacency within a given spatial unit. A smaller and clustered set of matched shoeprints is not only much more manageable than unsorted shoeprints, but also more likely to reveal and is able to profile and prioritize evidence to improve the efficiency of investigations.

As journey-to-crime and distance decay theories suggest, crimes are more likely to be committed within a short distance of the offender's residence. This study examines these relationships from a reverse perspective, similar to that of geographic profiling. If a number of matched shoeprints exist within a short distance of each other they may be said to form a cluster, it can then be inferred that the offender lives relatively close to the crime scene. From relationships inferred from both ESDA and the LISA statistic, investigators could consider these relationships, along with other information, when they are establishing a priority in the search for suspects.

The study has also showed that, although there are plentiful and varied spatial examples, the conceptual frameworks and hypotheses underlying crime mapping studies cannot be applied directly to forensic mapping. Crime mapping principles such as distance decay have not been tested thoroughly enough in a forensic evidence perspective to apply the methods routinely. As with crime mapping, both EDA and ESDA have proved useful for generating hypotheses and discovering patterns in forensic evidence that have not been reported previously. This study implemented a time-space constraint to filter massive amounts of evidence so that forensic investigators can better manage large databases of matched shoeprints for cluster and other pattern analyses. It was found that matched clusters were most likely to be within 1 mile in rectilinear or Manhattan distance. The 3 mile-range also uncovered some significant clusters. Clustered revealed under the larger time constraints were rendered insignificant because of the high number of shoeprint occurrences.

Although the algorithm was very effective in reducing the number of records and placing focus on subsets most likely to be related, some shoeprints from popular brands

still have hundreds of occurrences. These large subsets were deemed to be unmanageable using the proposed method and were assessed using the LISA statistic. It is stressed that the LISA method was investigated as a proof-of-concept rather than an empirical test.

Scatter plots and cluster maps proved to be extremely useful when further classifying forensic shoeprint evidence after the self-exclusion algorithm was imposed. The four types of local associations (high-high, high-low, low-low, and low-high) can provide clues to different types of shoeprint clustering. The case studies showed how both positively and negatively autocorrelated patterns can be used to identify spatial clusters, or lack thereof, in forensic evidence investigation.

Areas that are positively correlated have a high percentage of a specific matched shoeprint when compared to the overall number of shoeprints in that area. In many cases, areas that are positively autocorrelated have hot spots. They provide a focus for further forensic investigation. Negatively correlated areas suggest that the shoeprint in question does not occur at a high percentage when compared with the total number of prints in that area. There are usually no hot or cool spots associated with negative correlation. However, as stated this study was performed on a conceptual level, therefore it is difficult to discern to what extent an investigator will find it useful in practical terms.

Reflecting back on the study, there are a number of beneficial suggestions that could be used to improve this type of analysis. If this study was to be conducted based on what has been learnt, a similar form of analysis would be performed, but with modified techniques and data. For example, it is possible that in being of irregular size and shape, borough boundaries are not the best areal units, even though some studies suggest repeat regional offenders. Census units and police jurisdiction boundaries could

be used to see if the hot spot relationships persist with different sets of boundaries. The appropriate spatial resolution or scale of areal units requires further study.

Additionally, the use of borough boundaries may have affected the LISA statistics due to the fact that evidence was slightly denser in the center of the Greater London study area and the inner boroughs are significantly smaller than peripheral ones. A grid was considered to provide regular areal units, but cells with no evidence counts render the LISA statistic unusable and at a small spatial resolution many cells had no data. An alternative would be to experiment with grids of several spatial resolutions to ensure that all cells had values. Some investigators have overcome the problem of sparse data by adding one observation to all zones.

The Modifiable Areal Unit Problem (MAUP) is another issue that could be considered if the analysis was to be repeated. If the nature and degree of spatial relationships between variables change with the choice of different areal units, the reliability of the results is called into question (Chainey and Ratcliffe, 2005). As noted before, the borough boundaries may not be best suited for this analysis and other boundaries would likely provide different results.

Finally, looking back on this analysis, its basis is reminiscent of geographic profiling. Geographic profiling connects series of crime locations to determine the most likely area of offender residence (Rossmo, 1999). The present analysis is similar by seeking to link common evidence and speculating that the offender could live within a close proximity to the matched evidence. This research is conceptual, but hopes to establish a foundation and an interest in forensic evidence so it can be used from a spatial perspective to aid in crime solving.

In conclusion, the field of geography is a multidisciplinary study. Many facets of geography are utilized in other disciplines. GIS, for example is one technology that has been integrated into many private and public applications. This study is an example of how geography can be utilized in other fields. Police and law enforcement officials have benefited from the implementation of GIS and its applications. The field of forensic investigation can employ the same methods in a different context. In a similar manner to crime mapping, forensic investigators can apply principles such as hot spot mapping into their analysis. A method including the forensic recovery rate, which is similar in theory to a crime rate, was shown to illustrate spatial relationships between total evidence and matched evidence. Hot spot analysis of shoeprint data was used to identify matched shoeprint clusters that were related by a set of criteria. Finally, tests for spatial autocorrelation using the LISA statistic were performed to reveal spatial relationships of shoeprints in boroughs when compared with their neighbors. As noted previously, although conceptual, this is one of the first studies to implement these principles within forensic evidence and the author hopes this work will establish a foundation for later studies

Appendix A

Time-Space Algorithm

```
Private Sub TimeSpace_Click()
    *****
    *****
    ***** Part I: Preparation of the function*****
    *****
    *****

    'Get the focus map
    Dim pDoc As IMxDocument
    Set pDoc = ThisDocument
    Dim pMap As IMap
    Set pMap = pDoc.FocusMap

    ' Get the selected layer
    Dim pLayer As IGeoFeatureLayer
    Set pLayer = pDoc.SelectedLayer

    ' Make sure a layer was selected
    If pLayer Is Nothing Then
        MsgBox "You must select the destination Layer"
        Exit Sub
    End If

    ' Get the selected features from the layer
    Dim pActiveView As IActiveView
    Dim pFeatureLayer As IFeatureLayer
    Dim pFeatureSelection As IFeatureSelection
    Dim pQueryFilter As IQueryFilter

    Set pActiveView = pMap

    'For simplicity sake let's use the first layer in the map
    If Not TypeOf pMap.Layer(0) Is IFeatureLayer Then Exit Sub
    Set pFeatureLayer = pMap.Layer(0)
    Set pFeatureSelection = pFeatureLayer 'QI

    Dim pFSel As IFeatureSelection
    Set pFSel = pLayer

    Dim pSelSet As ISelectionSet
    Dim pFCur1 As IFeatureCursor, pFCur2 As IFeatureCursor

    Dim pFeat1 As IFeature, pFeat2 As IFeature
    Dim pPnt1 As IPoint, pPnt2 As IPoint, pTemp As IPoint

    Dim dDate1 As Date, dDate2 As Date
    Dim difDate As Integer

    Dim strSql As String
    Dim SelectedTotal As Integer
    Dim i As Integer, j As Integer
    Dim Dbl_Distance As Double
    Dim delOrNot As Boolean          ' Mark the point will be deleted or not

    delOrNot = False

```

```

*****
*****
**** Part II: Delete points based on Criteria*****
**** The predefined field for footprint is: SHOEMARK****
**** The distance and time period can be customized ****
*****
*****
Dim pData As esriGeoDatabase.IDataStatistics
Dim pCursor As esriGeoDatabase.ICursor, pStatResults As esriSystem.IStatisticsResults

Dim num As Integer
num = 0

Set pCursor = pLayer.Search(Nothing, False)

Set pData = New esriGeoDatabase.DataStatistics

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
$$$$$$$ modify the field name to it should be $$$$$$$$$$
$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
pData.Field = "SHOEMARKS"
Set pData.Cursor = pCursor

Dim pEnumVar As esriSystem.IEnumVariantSimple, value As Variant
Set pEnumVar = pData.UniqueValues
value = pEnumVar.Next

' This loop is to pick the points with the same shoemaker value
Do Until IsEmpty(value)
'Create the query filter
Set pQueryFilter = New QueryFilter

*****
**** The single quote in the field can be handled here****
*****
pQueryFilter.WhereClause = "SHOEMARKS = " & Replace(value, "'", "''") & ""

    p ActiveView.PartialRefresh esriViewGeoSelection, Nothing, Nothing 'Invalidate only the selection cache.
    Flag the original selection
pFeatureSelection.SelectFeatures pQueryFilter, esriSelectionResultNew, False ' Perform the selection
pActiveView.PartialRefresh esriViewGeoSelection, Nothing, Nothing 'Flag the new selection

Set pSelSet = pFSel.SelectionSet
pSelSet.Search Nothing, False, pFCur1

SelectedTotal = pSelSet.Count
'MsgBox SelectedTotal
num = num + pSelSet.Count

Debug.Print pQueryFilter.WhereClause & " " & pSelSet.Count & " " & num

Set pFeat1 = pFCur1.NextFeature

' If there is only one point for one shoemaker, this point will be deleted
If SelectedTotal = 1 Then
    pFeat1.Delete
    pSelSet.Search Nothing, False, pFCur1
End If

'If there is more than one points for one shoemaker, delete the points based on the criteria.

```

```

If SelectedTotal > 1 Then

  For i = 0 To SelectedTotal
    Set pPnt1 = New Point
    Set pPnt1 = pFeat1.Shape
    Set pTemp = pFeat1.Shape
    dDate1 = pFeat1.value(5)

    pSelSet.Search Nothing, False, pFCur2

    Set pFeat2 = pFCur2.NextFeature
    Do Until pFeat2 Is Nothing
      Set pPnt2 = New Point
      Set pPnt2 = pFeat2.Shape
      dDate2 = pFeat2.value(5)
      difDate = Abs(DateDiff("d", dDate1, dDate2))

      '$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
      '$$$$$$$ modify 15 to correct time period in day $$$$$$$$
      '$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$

      If difDate < 15 Then
        If pPnt1.X <> pPnt2.X And pPnt1.Y <> pPnt2.Y Then
          'Dbl_Distance = Sqr((pPnt1.X - pPnt2.X) ^ 2 + (pPnt1.Y - pPnt2.Y) ^ 2)
          Dbl_Distance = (Abs(pPnt1.X - pPnt2.X) + Abs(pPnt1.Y - pPnt2.Y)) / 1609.3
          'MsgBox Dbl_Distance

          '$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
          '$$$$$$$ modify 1 to correct distance in mile $$$$$$$$
          '$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$

          If Dbl_Distance < 1 Then
            delOrNot = False
            Exit Do
          End If
          delOrNot = True
        End If
      Else
        delOrNot = True
      End If
      Set pFeat2 = pFCur2.NextFeature
    Loop

    If delOrNot Then
      pFeat1.Delete

      pSelSet.Search Nothing, False, pFCur1
      SelectedTotal = pSelSet.Count
      i = 0
    End If

    Set pFeat1 = pFCur1.NextFeature

    If pFeat1 Is Nothing Then Exit For
  Next i
End If

value = pEnumVar.Next
Loop

End Sub

```

References

- Anselin, L., Exploring Spatial Data with GeoDa: A Workbook. Spatial Analysis Laboratory and Center for Spatially Integrated Social Sciences (CSISS), Department of Agriculture and Consumer Economics, University of Illinois, Urbana-Champaign. 2004.
- Anselin, L., Local Indicators of Spatial Association-LISA. *Geographical Analysis*, 27 (2), 1995, 93-115.
- Anselin, L., Exploratory spatial data analysis in a geocomputational environment *Geocomputation, a Primer*, Wiley, New York:,1998, 77-94
- Ashley, W., What shoe was that? The use of computerized image database to assist in identification, *Forensic Science International*, 82 (1), 1996, 7-20.
- Chainey, S., & Ratcliffe, J., GIS and Crime Mapping. West Sussex, England: John Wiley and Sons; 2005.
- Cloke, P., Philo, C., & Sadler, D., Approaching Human Geography: An introduction to contemporary theoretical debates. New York, NY: Gilford Press. 1991.
- Cohen, L., & Felson, M., Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44, 1979, 588-608.
- Craglia, M., Haining, R., & Wiles, P., A comparative evaluation of approaches to urban crime pattern analysis, *Urban Studies*, 37 (4), 2000, 711-730.
- Eck, J.E. What Do Those Dots Mean? Mapping Theories With Data. In D. Weisburd and T. McEwen (eds.), *Crime Mapping and Crime Prevention* (8), 1998, 379-406. Monsey, NY: Criminal Justice Press.
- Eck, J.E., Mapping Crime: Understanding Hot Spots. Washington, DC: U.S. Department of Justice, National Institute of Justice. 2005, 1-13.
- Gardner, R. M., Practical Crime Scene Processing And Investigation. FL: CRC Press. 2005.
- Garson, G. D., & Vann, I. B., Crime Mapping and Its Extension to Social Science Analysis, *Social Science Computer Review*, 19 (4), 2001, 471-479.
- Gatrell, A.C., Bailey, T.C., Diggle, P.J. Rowlingson, B.S., Spatial Point Pattern Analysis and Its Application in Geographic Epidemiology, *Transactions of the Institute of British Geographers*, New Series, 21 (1), 1996, 256-274.
- Geradts, Z., Bijhold J., Content based information in forensic image databases, *Journal of Forensic Science*, 47(2), 2002, 285-292.

Geradts, Z., Keijzer, J., The image-database REBEZO for shoeprints with developments on automatic classification of shoe outsole designs, *Forensic Science International*, 82 (1), 1996, 21-31.

Girod, A., Computerized classification of the shoeprints of burglars' shoes, *Forensic Science International*, 82(1), 1996, 59-65.

Hamm E.D., Track identification: an historical overview, *Journal of Forensic Identification*, 39 (6), 1989, 333-338.

Harries, K., Mapping Crime: Principle and Practice. Washington, DC: U.S. Department of Justice, National Institute of Justice. (1999),1-206.

Hilderbrand, D., Footwear, the Missed Evidence. Wildomar, CA: Staggs Publishing; 1999

Johnston, R., Gregory, D., Pratt, G., & Watts, M., The Dictionary of Human Geography. Malden, MA: Blackwell Publishing, 2000.

Johnston, R., Philosophy and Human Geography: an introduction to contemporary approaches. London: E. Arnold, 1983.

Kitchen, R., & Tate, N., Conducting Research into Human Geography. Essex, England: Pearson Educated Limited; 2000

Leist, C. & Milne, B., The Mapping Of Forensic Data, PowerPoint presentation, Metropolitan Police, London, U.K. 2005.

Messner, S., & Anselin, L., Spatial Analyses of Homicide with Areal Data. Spatially Integrated Social Science. Oxford University Press, NY: 2004 127-144.

Messner, S., & Anselin, L., et al. The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis. *Journal of Quantitative Criminology*, 15 (4), 1999.

Rossmo, K., Geographic Profiling. Boca Raton, FL: CRC Press. 2000.

Report: Ohio Bureau of Criminal Identification and Investigation, 2003-2004

Tukey, J. W., Exploratory Data Analysis, Addison-Wesley, Reading, MA, 1977.