Geographic Probability Algorithms with Security Force Applications

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Samuel H. Huddleston

August

2013

APPROVAL SHEET

The dissertation

is submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

AUTHOR

The dissertation has been read and approved by the examining committee:

Professor Donald Brown

Advisor

Professor Gerard Learmonth

Professor Matthew Gerber

Professor John Porter

Professor Robert Kewley

Accepted for the School of Engineering and Applied Science:

James H. Ay

Dean, School of Engineering and Applied Science

August

2013

Abstract

Every day, government executives, police officials, and military leaders must decide how to most efficiently and effectively employ their limited resources in an effort to secure the large and diverse populations they are charged to protect. Increasingly, these leaders rely on the analytic tools provided by the discipline of crime analysis. One of the most important tools in the discipline of crime analysis is the predictive hot-spot map, which is used to make tactical level decisions about the employment of resources. This dissertation develops methodological approaches for exploiting these predictive crime maps to improve the crime forecasts, geographic districting plans, intelligence assessments, and targeting plans that support military and police decision makers.

This research provides four multidisciplinary contributions. First, this dissertation provides a new method for forecasting noisy geographic time series that provides statistically significant performance improvements over the most-used forecasting methods while dramatically reducing modeling workload so long as several modeling assumptions are satisfied. Second, this new forecasting method is supported by the development of a statistical motivation that explains why weighted aggregate forecasts provide better forecasting performance for disaggregated event count time series than forecasts made using the observations from the many disaggregated event count time series themselves. Third, this dissertation documents a new method for geographically mapping the region where spatial choice behavior by one entity or group will dominate spatial choice behavior by all other considered groups. Finally, this dissertation documents the development of a new approach for Journey to Crime (JTC) analysis that adds to the existing literature by providing the ability to simultaneously model the effect of many environmental factors on the spatial choice behavior of the modeled agents (plants, animals, or criminals) while incorporating the distance-decay modeling used by existing JTC methods.

This dissertation demonstrates the practical application of these research contributions in four case studies. First, a new geographic forecasting method, Geographic Probability Forecasting (GPF), is applied to the problem of forecasting weekly burglary counts over a five-year period in Pittsburgh, Pennsylvania. The GPF method links the tactical and operational levels of planning, reduces modeling workload, and significantly improves forecasting performance for this problem. Second, the GPF method is leveraged to produce planning maps for Albemarle County, Virginia, that facilitate the development and evaluation of the districting plans that are used to define geographic areas of responsibility for patrolling units. Third, previous work in Criminal Site Selection (CSS) modeling is extended to develop a Sphere of Influence (SOI) analysis, improving the intelligence assessments for criminal gangs in Santa Ana, California. Finally, CSS models are leveraged to develop a new JTC analysis technique that outperforms the current best JTC method for predicting the geographic anchor points of criminal gangs in Santa Ana.

Acknowledgements

I would like to acknowledge all of the many professors, military professionals, police officers, and fellow students who have significantly contributed to my academic development over the last few years. Since that is not possible, I must mention the small handful of people who most directly impacted the research in this dissertation, beginning with those most influential in providing the opportunity: Colonel (Dr.) Robert Kewley and Brigadier General (Dr.) Timothy Trainor. I am forever indebted to both of them for the academic development, support, and example of officership they have provided over the last five years.

I must also acknowledge the significant role that my committee members from UVA played in the development of this research, especially my academic advisor Dr. Donald Brown. Dr. Brown recruited me to UVA for the M.S. program over the phone while I was deployed to Iraq, arguing that UVA would provide the opportunity to work on problems that directly affected deployed soldiers like myself on a daily basis. Over the last seven years of conducting research with Dr. Brown, I have appreciated the tremendous intellect, focus, and service he provides towards improving analytic capabilities across the Department of Defense. The other UVA members of my committee (Professor Matthew Gerber, Professor Gerard Learmonth, and Professor John Porter) all spent a considerable amount of time reviewing and contributing to my publications, serving as sounding boards for my off-the-wall ideas, and providing advice for problems both academic and otherwise.

I must also thank the researchers and police professionals who provided the data used in this dissertation. Professor James Meeker of the University of California and Professor Wilpen Gorr of Carnegie Mellon University both shared data sets they had developed over years of effort. Colonel (Chief) Steve Sellers of the Albemarle County Police Deparment (ACPD), Major John Parrente (ACPD), and Chief Paul Walters of the Santa Ana Police Department all agreed to have their sensitive law enforcement data sets used to further academic research. Lieutenants Pete Mainzer and Greg Jenkins of the ACPD spent a considerable amount of time helping me understand the practicalities of police needs and the real-world applications of this research.

Far too many of my fellow students made a significant impact on my research to list here. However, three fellow students have provided significant help to me over the course of many years. First, Anthony Fox, my former UVA office-mate, has served the role of professional critic (and emergency programming consultant) over the course of the last five years. My PhD office-mate Lieutenant Colonel (Dr.) Ed Teague spent countless hours critiquing my white board work (and listening to my rants) during our tour together at UVA. Lieutenant Colonel (Dr.) Jon Fox has been my primary academic sounding board, the most loyal friend, and the best example of officership that I have ever had the pleasure of working with.

Finally, I would like to recognize the love, patience, understanding and support of my wife, Kathy, and children, Caleb, Leah, and Molly, throughout my years of Army service and during my studies here at the University of Virginia. Thank you all for your help and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Every day, government executives, police officials, and military leaders must decide how to most efficiently and effectively employ their limited resources in an effort to secure the large and diverse populations they are charged to protect. Increasingly, these leaders rely on the discipline of crime analysis. While the field of crime analysis is very old, it has seen an explosion in methods and application since the advent of Geographic Information Systems (GIS). GIS technology has enabled police agencies and military organizations to develop sophisticated mapping and statistical analysis techniques to build models for the spatial prediction of crime incidents. These products produce a probability or intensity surface that predicts future concentrations of activity in space and time from spatial point patterns such as enemy Improvised Explosive Device (IED) attacks, criminal activity, or calls for emergency services. This dissertation develops methodological approaches for exploiting the information provided by these predictive crime maps to also improve the crime forecasts, geographic districting plans, intelligence assessments, and targeting plans used to support decision-making at the tactical and operational levels of war and policing.

1.1 Research Hypothesis

Figure 1.1 provides an illustration of the organizing principle for this research: geographic probability models have application to many different types of analyses that are used to support military and police security operations. Predictive geographic probability models such as kernel density estimation (Harris, 1999; Eck et al., 2005) and Criminal Site Selection (CSS) models (Huddleston and Brown, 2009; Huddleston et al., 2012; Smith and Brown, 2007; Xue and Brown, 2003) are traditionally used to predict the most likely locations for future criminal activity. This dissertation explores four additional applications for these geographic probability models that address existing gaps in the crime analysis literature. These applications are explored through the study of four hypotheses:

Hypothesis 1: Geographic probability models can improve the ability of military and police analysts to accurately forecast the noisy geographic time series produced by crime.

Hypothesis 2: Geographic probability models provide a simple and effective heuristic approach for patrol district design that is applicable to many police and military units.

Hypothesis 3: CSS geographic probability models can be used to accurately identify the regions in a city or area of interest where the probability of criminal behavior by one group dominates the probability of criminal behavior by any other group, a region we term the group's *sphere of influence*.

Hypothesis 4: CSS geographic probability models can improve upon existing journey to crime methods in identifying the anchor points for criminal groups, facilitating improved targeting of these criminal elements.

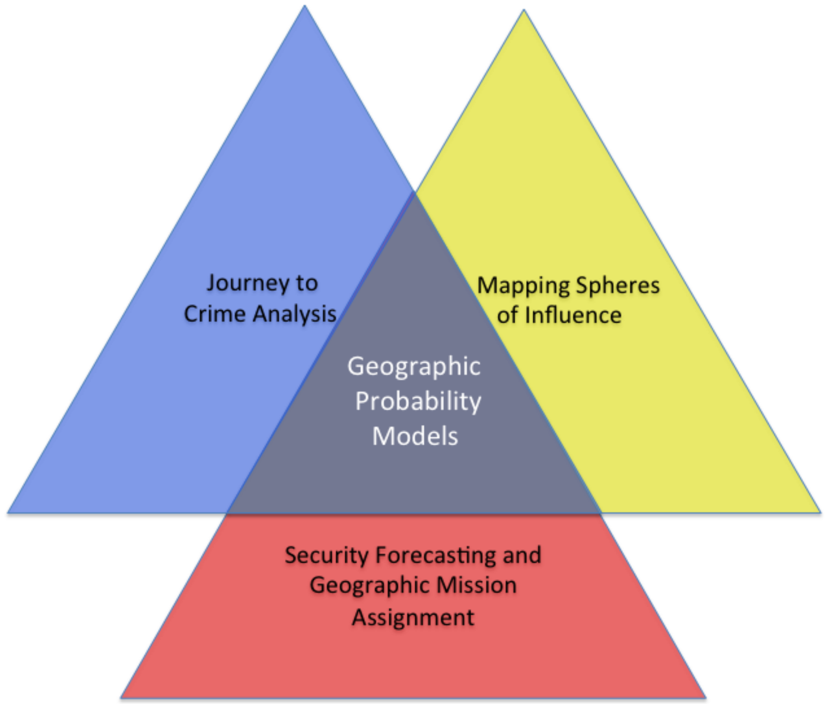


Figure 1.1: Illustration of the organizing principle of this dissertation: the information provided by geographic probability models can be used to improve the crime forecasts, geographic districting plans, intelligence assessments, and targeting plans used to support decision-making at the tactical and operational levels of war and policing.

1.2 Geographic Probablity Models

A geographic probability model is a two-dimensional probability or intensity map that predicts future concentrations of activity in space and time from spatial point patterns such as enemy Improvised Explosive Device (IED) attacks, criminal activity, or calls for emergency services. There are many approaches that can be used to develop hot-spot maps including manual methods (Boba, 2005; Eck et al., 2005), density methods (Harris, 1999; Eck et al., 2005), criminal site selection modeling (Huddleston and Brown, 2009; Huddleston et al., 2012; Smith and Brown, 2007; Xue and Brown, 2003), and increasingly machine learning or data mining techniques (McCue, 2007). Two methods for developing predictive crime maps have model structures that facilitate the research developments explored in this dissertation: kernel density estimation (KDE) and Criminal Site Selection (CSS) models. Both of these techniques generate mapped representations of future threat activity that have several distinct characteristics which make them suitable for use as geographic probability models:

(i) When mapped, the output of the model is a two-dimension probability density function.

(ii) This function can be mapped (indexed) at a very high resolution (i.e. 50 x 50 meters).

(iii) The two-dimensional probability distribution function developed by normalizing the mapped density function sums (integrates) to 1 over the location index because the probabilities are (conditionally) independent (on the feature set) at considered locations.

(iv) The mapped probability distribution function represents the following probabilistic statement: "Given that an event has occurred (or will occur) in the domain of interest, the probability that it occurred (or will occur) at location i is...".

1.3 Organization of the Dissertation

The remaining chapters of this dissertation explore how geographic probability models (hot-spot maps) can be used to address current gaps in the discipline of crime analysis. The taxonomy and literature review in Chapter 2 conducts an overview of the various taxonomies used for security planning, the theories of criminal behavior used to develop models and simulations of criminal behavior, and the crime analysis techniques currently used for tactical and operational level security planning. This literature review highlights several significant gaps in security planning capabilities at the tactical and operational level.

First, at the operational level of planning, there is a current need for a forecasting system to better support operational level planning in security applications such as policing and military counter-insurgency campaigns. Ideally, this forecasting system would provide the following:

(i) accurate forecasts for small geographic regions (precincts or car beats) for small temporal windows (days or weeks),

(ii) a link between tactical level planning (supported by the analytic discipline of prediction) and operational level planning (supported by the analytic discipline forecasting),

(iii) support for the operational level problem of geographic mission assignment (i.e. car beat and precinct boundaries) in applications such as geo-policing.

Chapters 3, 4, and 5 address this current gap. Chapter 3 documents the development of a new, simple forecasting technique for noisy geographic time series that consistently provides statistically significant performance improvement over the most commonly used methods for forecasting crime series. This forecasting method also dramatically reduces the modeling workload for developing recurring short-term forecasts in a geographic context. Chapter 3 also demonstrates that criminal hot-spot (geographic probability) maps provide accurate estimates for the geographic distribution of future crime counts within geographic regions. Chapter 4 applies this insight by leveraging these mapping products to develop a method for formally evaluating and comparing the performance of competing plans for geographic mission assignment for military and police units. Chapter 5 compares the performance of this closed form method for evaluating geographic districting plans to simulation methods to identify the appropriate context and limitations on using this approach.

Chapter 6 addresses an existing gap in operational level intelligence assessments: the ability to accurately map where different threat groups (criminal gangs or insurgent groups) present the greatest threat to the population. Multilevel modeling extensions of CSS models allows us to better answer this question by linking the incidence of gang crime to the spatial, demographic, and socio-economic features of specific locations. The output of these models can be used in a Geographic Information Systems (GIS) to develop a Sphere of Influence (SOI) analysis: mapping the region of dominance for each criminal group. These products have direct application in the intelligence assessments of criminal groups in both military and police security applications.

Chapter 7 develops an approach for improving the success rate of military cordon and search operations by leveraging Criminal Site Selection (CSS) models for Journey to Crime (JTC) analysis. JTC analysis is an investigative technique employed by police that uses the known locations of a crime series to determine a serial offender's most likely anchor point, usually a residence or workplace. This new modeling approach for JTC analysis provides statistically significant performance improvements over the current best method and provides geographic profiles (models of anchor point locations) that are often accurate enough to facilitate tactical success, with the modeled criminal group's anchor point falling within the search profile for military unit cordon and search operations. This CSS modeling approach also contributes to the JTC analysis literature by providing a method for modeling the effect of the journey to crime relationship after considering other environmental effects such as socio-economic conditions, crime generators, and crime attractors that might affect a criminal's decision-making process.

Chapter 8 concludes the dissertation by summarizing the multidisciplinary research contributions, domain-specific research contributions, and the most likely areas for future work motivated by the research in this dissertation.

Chapter 2

Literature Review

This chapter defines a taxonomy for crime analysis and security planning problems used throughout this dissertation and conducts a comprehensive literature review of current methods used to address these problems. The review begins by analyzing several taxonomies used in security planning and highlights the current need for the Geographic Probability Forecasting (GPF) method developed in the following chapter. The next two sections conduct an overview of the current theories of criminal behavior, discuss how those theories provide the basis for predictive models, and review how those theories are leveraged in simulation models used to study and validate predictive models. The review continues with overviews of the current literature on predictive models used to develop hot-spot maps (predictive threat surfaces), geographic profiling models, patrol district design methods, and forecasting models used for resource allocation and planning. These summaries include discussions of the performance assessments used to assess prediction, geographic profiles, and forecasting algorithms.

2.1 Crime Analysis and Security Planning

In his seminal text on policing, Orlando W. Wilson, Military Police Governor in postwar Berlin and Chicago Superintendent of Police, laid out the purpose of crime analysis in planning for the employment of police resources (Wilson, 1963):

> *The crime-analysis section studies daily reports of serious crimes in order to determine the location, time, special characteristics, similarities to other criminal attacks, and various significant facts that may help to identify either a criminal or the existence of a pattern of criminal activity. Such information is helpful in planning the operations of a division or district.*

Likewise, military doctrine specifically requires consideration of the effects of the terrain and civil considerations (environment), the threat (criminal organizations and their actions), and cultural/social events (temporal factors) in the planning of counter-insurgency operations, in which it has been engaged in for much of the past decade (U.S. Army, 2006). Rather than having a dedicated crime analysis section like that advocated for by O.W. Wilson, in military organizations this analytic function is performed by the military intelligence sections.

Most security organizations operate under a hierarchical command structure in which a higher level of the hierarchy allocates resources to designated subordinate commands/regions which employ them in time and space. However, there are several different taxonomies used to designate the various levels of planning. US military planning doctrine is based on a theory proposed by Carl von Clausewitz that asserts that the conduct and planning of war occurs at three levels: the tactical, the operational, and the strategic (von Clausewitz, 1984). The strategic level of war, defined as "the art and science of employing national power," is a clearly distinct class (Dunn, 1996). In contrast, the operational and tactical levels of war often overlap. In military applications, the tactical level of war is "characterized by the application of concentrated force and offensive action to gain objectives (Dunn, 1996)." The operational level of war is "implemented by assigning missions, tasks, and resources to tactical operations (Dunn, 1996)." Thus, in military doctrine, the planning at the tactical level of war and security planning is oriented on the spatial-temporal employment of resources (specific applications of resources in space and time) while operational level planning and decision-making is oriented on resource and mission allocation (allocating resources and missions between subordinate commanders who will employ them in space and time).

The most common taxonomy used to describe the various levels of security planning in crime analysis literature identifies four categories of crime analysis: tactical, operational, administrative, and strategic (Boba, 2005; Paulsen et al., 2010). Note that the strategic level of analysis and planning in policing literature does not correspond to the same level of security planning identified as strategic in military doctrine. This is because most policing is conducted at the city (or county) level or lower, which in military doctrine corresponds to an operational level of planning. In this taxonomy, strategic analysis is defined as being focused on long-term crime reduction and is oriented on "non-specific criminal activity problems (Paulsen et al., 2010)." Administrative analysis is "the study of police efficiency and effectiveness" through programs such as the CompStat, which "evaluates the performance of police districts through the study of performance statistics (Weisburd et al., 2004; Henry and Bratton, 2002)." The distinguishing feature of operational analysis is that it focuses on police procedures. It therefore overlaps with both the operational and tactical levels of military planning, depending upon the level of police procedure analyzed. Finally, the most specific type of crime analysis in this taxonomy is tactical crime analysis, which is "the comprehensive identification, evaluation, analysis, and resolution of specific criminal activity problems" such as emerging crime sprees (Paulsen et al., 2010). In the crime analysis literature, the difference between operational and tactical analyses is the subject of study. Tactical crime analysis focuses on studying crime occurrence (rates, locations, etc.) while operational analysis focuses on police resourcing and response to crime.

Because it provides the more general framework and categorizes planning processes in a way that aligns neatly with the analytic disciplines that support those processes, the military's taxonomy for security planning is used throughout this dissertation. Under this taxonomy, *planning at the tactical level of war/policing is concerned with how to employ resources in time and space* (locations/organizations/people to investigate, target, and observe) while *planning at the operational level of war/policing is concerned primarily with assigning missions for and allocating limited resources to subordinate regions and commanders who will employ them, as well as subsequently evaluating the performance of those units* (as in the CompStat program). Because the resources available at the operational level are almost always constrained in some way, the planning processes at the operational and tactical levels are inextricably linked.

Unfortunately, the analytic methods currently used for planning at the operational and tactical levels are not linked in the same way. Rather, they primarily leverage two different analytic disciplines: statistical assertions of multinomial counts, known in the crime analysis literature as *forecasts*, and statistical assertions of binomial probabilities, known in the crime analysis literature as *predictions*. In many disciplines, the terms forecasting and prediction are synonymous, but in the terminology of crime analysis "the terms forecasting and prediction have similar meanings, but they do not convey precisely the same idea (Paulsen et al., 2010)." The discipline of prediction is concerned with probabilistically foretelling *events* (with some probability) while the discipline of forecasting "attempts to estimate how much of something will occur in a given area over a given time period (Paulsen et al., 2010)." Thus, prediction is concerned with identifying the most likely times and places for criminal events to occur (so that resources can be concentrated in those areas) while forecasting is concerned with estimating *how much* criminal activity will occur in a given region over a given period of time (so that enough resources can be allocated to those commanders that need them most). As a recently published text in crime analysis notes, "forecasting is a strategic and administrative discipline while prediction is tactical and operational (Paulsen et al., 2010)." Under the more general military taxonomy, we can therefore state that *forecasting is an operational discipline while prediction is a tactical discipline*. These analytic disciplines are discussed at length in subsequent sections.

2.2 Theories of Criminal Behavior

In order to understand, build models of, and predict criminal behavior, it is necessary to understand the theories of criminal behavior that have been developed in the field of criminology. Environmental criminology seeks to to identify criminal patterns in the motivation, opportunities, and environments in which criminal events occur (Boba, 2005) and provides some key theories that are used to model and simulate criminal behavior. Three theories form the impetus for environmental criminology: rational choice theory, crime pattern theory, and routine activities theory. Rational choice theory is a fundamental micro-economic principle that states that people balance costs and benefits when making decisions about how to best meet their objectives (Coleman and Fararo, 1992). Becker (1976) first applied this economic theory to crime, asserting that criminals make choices about committing crimes based upon anticipated rewards and punishments. Routine activities theory suggests that the general patterns of behavior of both criminals and victims impact the incident of crime (Cohen and Felson, 1979). Crime pattern theory incorporates both rational choice theory and routine activities theory in asserting that criminal events occur most frequently where the activity space of offenders seeking to commit a crime overlaps with the routine activity of victims, generating crime patterns that cluster in space and time (Brantingham and Brantingham, 1993). Situational crime prevention applies these ideas in seeking to prevent crime by using statistical analysis to identify the features that make up criminals' preferences in the local environment and take specific actions to address those crime problems (Boba, 2005).

Another important principle in modeling and simulating crime is the principle of *journey to crime* (Rengert, 2004). Journey to crime maps criminal offender travel distance based on three factors (Paulsen et al., 2010). First, offender travel is heavily influenced by the presence of *crime attractors* and *crime generators*. Crime attractors are "places, areas, or neighborhoods where criminal opportunities are well known, and to which motivated offenders are subsequently attracted as a source for criminal activity (Paulsen et al., 2010; Brantingham and Brantingham, 1993)." Crime generators are places such as shopping malls, festivals, college campuses, and sports events that provide criminal opportunities due to the massive number of potential victims concentrated in one place at specific times (Brantingham and Brantingham, 1993). The second factor that influences offender travel is spatial attractiveness. Areas that are close to offenders' residences or other *anchor points* in their routine activities are more attractive because offenders prefer to travel less when possible. This means that criminals are less likely (distance-decay) to commit crimes far from their homes or other anchor points. Finally, an offender's journey to crime is influenced by the *target backcloth* created by victim's decisions (Paulsen et al., 2010). The term target backcloth encapsulates the idea that criminal opportunities are not uniformly distributed in geography. Criminals must move to locations where crime opportunities for the particular type of crime they wish to commit are available. For example, criminals wishing to commit a bank robbery are limited to selecting from already existing banking institutions, even if none of those exist near their anchor points or within areas conducive to crime opportunities.

Finally, the concepts of *mental maps* and *criminal awareness space* influence the specific locations individual criminals choose for their crimes (Paulsen et al., 2010; Brantingham and Brantingham, 1993). In brief, criminals have mental maps for areas with which they are familiar. These maps include a criminal's awareness space from which he or she selects their targets. This awareness space is for the most part defined by the criminal's anchor points and the paths between those anchor points.

2.3 Simulation Models of Criminal Behavior and Police Response

These theories from environmental criminology serve as a foundation for developing models and simulations of criminal behavior and police response. Several of the predictive models discussed in the next section specifically incorporate these theories and principles into the modeling process. Additionally, these theories are increasingly being used to develop agent-based simulation models of criminal behavior in efforts to better understand criminal behavior, test the veracity of criminological theories, optimize the efforts of security agencies, and validate predictive modeling approaches.

As Malleson (2011) notes, "Environmental criminology research tells us that the geographical patterning of crime rates is an emergent phenomenon, resulting from the interactions between individual people and objects in space." Emergence is defined as "displaying organization without a central organizing authority" and is the defining characteristic of *complex systems* (Ottino, 2004). Agent-based models are used to study complex systems because they model the complex interactions between independent, autonomous (and often competing) actors pursuing their own objectives.

In the context of studying the emergence of crime patterns, agent-based models provide the ability to simulate how micro-level decision processes give rise to patterns of crime (Liu and Eck, 2008). The agent-based modeling paradigm facilitates modeling the specific decision processes and behaviors of criminal offenders, their victims, and the security forces who try to protect the population. These models have been used to validate the theories of environmental criminology by showing how simulation models incorporating these theories produce crime patterns that mimic observed crime patterns. Liu et al. (2005) provided one of the first demonstrations of this application of agent-based modeling when they developed a cellular automata model based upon routine activities theory to produce spatial crime patterns that mimicked observed crime patterns in Cincinnati, Ohio.

Other researchers have developed agent-based models that validate additional theories from environmental criminology. For example, Wang (2005) extended the model of Liu et al. (2005) to mimic street-robberies, achieving similar results for a specific crime. Groff (2008) studied the spatial-temporal patterns of street robbery by leveraging routine activity theory and Bolstad (2008) develop an agent-based model that illustrates the development of crime patterns using crime pattern theory. Wang et al. (2008) demonstrate the use of all three environmental theories (rational choice theory, routine activities theory, and crime pattern theory) implemented within a GIS environment, which represents a significant advance in model complexity. However, Elffers and Baal (2008) argue that the use of real geographical backgrounds, while feasible, "may in fact be detrimental for the real reason of doing criminal simulation studies, which is understanding the underlying rules."

Agent based simulation models have also been recently used to determine more effective law-enforcement strategies. Examples in policing include optimizing patrol routing systems (Szakas et al., 2008), analysis of drug law enforcement efforts (Dray et al., 2008), and investigating the effects of situational crime prevention strategies on the incidence of burglary (Malleson, 2011). Huddleston et al. (2008) demonstrate the application of agent-based simulation models in a military application: determining the optimal spatial distribution of security outposts when countering an insurgency.

Finally, Fox and Brown (2012) and Fox et al. (2012) demonstrate a new use for agent-based models: using agent-based simulation models to validate and understand the properties of a predictive algorithm for spatio-temporal crime patterns. Fox and Brown (2012) adapts the insurgency model of Huddleston et al. (2008) to illustrate how a new predictive algorithm for the spatial-temporal prediction of crime in a small US city accurately captures the behavior of criminal actors. Fox et al. (2012) uses the same simulation model to demonstrate a method for capturing the effect of spatial-temporal pulse events on crime incidence. The agent-based simulation model provides the ability to assess the success of the predictive model performance because the operating rules of the agents are known. Thus, researchers can assess the effectiveness of predictive models in capturing the known behavior of the criminal agents in a controlled operating environment.

2.4 Crime Prediction Methods

Law enforcement analysts have developed and documented sophisticated statistical mapping techniques to build models for the spatial prediction of crime incidents. These efforts began with crime mapping: the use of geographic information systems to conduct spatial analysis of crime problems and other police-related issues. Harris (1999) documents the development of crime mapping within law enforcement circles from the time when maps consisting of pins pushed into wall mounted maps were used by the NYPD in the early 1900s through the 1990s, when the increasing availability of crime desktop GIS systems resulted in an explosion in the use of computer desktop crime mapping in law enforcement. GIS allow crime analysts to develop sophisticated maps that help police concentrate their resources in the locations most likely to need support through the use of predictive threat surfaces.

Predictive threat surfaces identify concentrations of criminal activity, or *hot-spots*. The National Institute of Justice defines a criminal hot-spot as “an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization (Eck et al., 2005).” Predictive threat surfaces are binomial classifiers, with the goal being to highlight specific locations likely to experience future criminal activity. As Fawcett (2004) notes, “A classifier need not produce accurate, calibrated probability estimates; it need only produce relative accurate scores that serve to discriminate positive and negative instances.”

There are many techniques for identifying criminal hot-spots but they for the most part fall into one of two categories: those techniques that treat the problem as a spatial point pattern vice those that treat the problem as a *marked* spatial point pattern. Point patterns are the type of spatial data that arise when the critical variable being analyzed is the location of *events* (Cressie, 1993). Most criminal incidents fall into this category of geographic analysis. A marked spatial point pattern is one in which the events in a point pattern are associated with measurements or categorical marks. In crime analysis, examples of “marks” would include identification of the type of crime, the responsible party (if known), and the environmental features associated with the location of the criminal event.

2.4.1 Spatial Point Pattern Methods

Spatial point pattern methods use the locations of previous crimes to predict concentrations of future criminal activity. Note that these approaches are not based in the theories from environmental criminology but rather assume that the best predictor of future criminal activity is past criminal activity nearby. The most common approaches to identifying criminal hot-spots rely on mapping techniques based upon kernel density estimation because these approaches are easily implemented in the Geographic Information Systems (GIS) most police agencies now employ (Eck et al., 2005;

Boba, 2005). These techniques do not leverage the additional "marked" information associated with a criminal event but leverage only location (Latitude-Longitude or X-Y) data to estimate the relative risk associated with each X-Y coordinate on the map.

Kernel density estimation uses a kernel smoothing function (Wand and Jones, 2004) to develop a probability density function of a random variable (crime occurrence) which is then represented as a continuous threat surface (hot-spot) map. Use of a kernel function involves two important decisions: the selection of a kernel function and the selection of the kernel bandwidth. Commonly used kernel functions include Gaussian, quartic, and Epanechnikov but the kernel function has relatively little effect on the density estimate. The bandwidth parameter does significantly effect the outcome and is an important consideration in the development of the model, with many studies that propose various methods for selecting this key model parameter including Sheather and Jones (1991), Jones et al. (1996), and Berman and Diggle (1989).

This modeling procedure has become nearly ubiquitous in crime analysis, but as a National Institute of Justice special report on hot-spot mapping notes, it is not always applied appropriately by police agencies (Eck et al., 2005):

> *The increased application of this type of continuous surface smoothing method is due largely to its more common availability and visual appeal. Continuous surface hot spot maps allow for easier interpretation of crime clusters and reflect more accurately the location and spatial distribution of crime hot spots. As their appeal has increased, however, few questions are being asked of the outputs generated. Many agencies often fail to question the validity or statistical robustness of the map produced, being caught instead in the visual lure of their sophisticated looking geo-graphic.*

Other point process methods are not often cited in texts on crime analysis and crime mapping (Gorr and Kurland, 2012; Boba, 2005; Paulsen et al., 2010; Chainey and Ratcliffe, 2005; Eck et al., 2005), but there are several relevant examples available. Liu and Brown (2003) develop a criminal incident prediction model based on a point-pattern density model discussed in more depth in the next section. Mohler et al. (2011) build on work developed in seismology for dealing with space-time clustering by using self-exciting point process models and demonstrate their use in modeling near-repeat burglaries in Los Angeles. Finally, Kerry et al. (2010) apply the geo-statistical technique of kriging to the problem of predicting car-related thefts.

2.4.2 Marked Spatial Point Pattern Methods

In recent years, there has been a growing body of literature in which researchers identify criminal hot-spots by using the *marks* associated with crimes in police databases to identify criminal hot-spots. Social researchers tend to use various regression techniques in the application of environmental criminology to link social, economic, or spatial features to the incidence of crime (Brantingham and Brantingham, 1981). Examples of sociological analysis include identification of factors important in the occurrence of residential burglaries (Bernasco and Nieuwbeerta, 2005), robberies in Chicago (Bernasco and Block, 2009), the link between drug street corners and crime (Ratcliffe and Taniguchi, 2008), and many studies of criminal gang activity (Tita et al., 2005; Tita and Ridgeway, 2007; Block, 2000). Cahill and Mulligan (2007) and Chainey and Ratcliffe (2005) identify that many of these factors have relationships with crime that vary by location and use geographically weighted regression to improve predictive performance over the more common standard least squares regression methods.

Other researchers have begun to apply newly developed data mining techniques to the problem of identifying the areas most likely to see a criminal incident. Data mining approaches to hot-spot identification include machine learning techniques such as neural networks (Olligschlaeger, 1997), fuzzy clustering (Grubesic, 2006), and support vector machines (Chang et al., 2005; Kianmehr and Alhajj, 2008). However, these techniques are not often applied by police analysts because they require an in-depth understanding of statistical methods or custom software to conduct the analysis.

Discrete Choice Modeling (DCM) bridges the gap between the regression techniques often developed by criminologists and predictive data mining algorithms. Criminal DCM models are based upon Daniel McFadden's development of discrete choice theory (McFadden, 1974). Several research groups have applied this approach in a spatial context for modeling criminal site selection preference. Xue and Brown (2003) introduce the use of McFadden's discrete choice theory in the context of modeling criminal choice behavior, leveraging the rational choice theory as applied to crime by Becker (1976). Liu and Brown (2003) incorporate the idea of using feature-space rather than geographic coordinates to represent the locations of crimes when conducting DCM. Feature-space is defined as the Euclidean distance to each of the features of interest such as various crime attractors and crime generators. Liu and Brown (2004) combine the feature-space methods with DCM modeling to develop a framework for the application of spatial choice in space and time. The efficacy of this Discrete Choice Modeling (DCM) approach has been proven in predicting criminal events as varied as burglaries (Liu and Brown, 2004; Bernasco and Nieuwbeerta, 2005), gang crimes (Huddleston and Brown, 2009), robberies (Bernasco and Block, 2009), and terrorist events (Brown et al., 2004) such as suicide bombings (Smith and Brown, 2007). Often DCM models will provide significantly improved performance over density methods because they identify potential high risk areas which have not yet been targeted, whereas the previously discussed density approaches only highlight areas that have previously seen criminal activity.

Recently these DCM models have been extended to more accurately predict crime by individual criminal groups or for spatio-temporal prediction. Huddleston and

Brown (2009) demonstrate that extending DCM with the application of multi-level (hierarchical) modeling can significantly improve predictive performance for individual criminal gangs. Fox and Brown (2012) demonstrate how this multi-level modeling approach can be leveraged to improve spatio-temporal prediction of criminal assaults while Wang and Brown (2012) uses generalized additive models (GAM) to conduct the same analysis for breaking and entering crimes.

2.4.3 Journey to Crime (Geographic Profiling) Models

Another predictive crime mapping technique often used in tactical level policing is Journey to Crime (JTC) analysis, more often called geographic profiling. JTC analysis is defined in crime science as identifying the anchor point for an unknown serial criminal offender based upon the locations of a linked crime series attributed to that offender (Mohler and Short, 2012). While geographic profiles (the mapped outputs of JTC analysis) are sometimes used to support the police investigative process by highlighting the most likely locations for the anchor points of an unknown serial criminal offender, they are more regularly used to narrow investigative efforts geographically to a subset of known offenders who have previously committed similar crimes and fall within the high-probability area of the geo-profile or the equivalent geographic subset of a longer list of possible suspects that the police have already identified (Rossmo et al., 2005). Effective solutions to this problem are in high demand because the most important function of the criminal investigative process is locating unknown offenders (Rossmo and Rombouts, 2008) .

There are several significant criticisms lobbied against geographic profiling and JTC analysis. The most significant criticism is that, despite several high-profile and widely publicized successes, geographic profiling models have not proven to be more accurate than simple centrographic techniques such as calculation of the Center of Minimum Distance (CMD), also known as the Fermat-Weber point (Levine and Block, 2011; Paulsen, 2006b). In several studies comparing the predictive performance of all of the available geographic profiling techniques, to include Canter's Dragnet (1993), Rossmo's Rigel (2000), Levine's JTC (2009a), and several centrographic methods, Paulsen (2006a; 2006b) found that the simple CMD method outperformed all others in identifying serial criminals' anchor points and the CMD method is widely considered to be the most accurate geographic profiling method (Levine and Block, 2011). Even the most recent geographic profiling techniques developed have only demonstrated the ability to match, not exceed, the performance of the CMD approach.

The second significant criticism of geographic profiling models is that all of the traditional geographic profiling methods do not account for the target backcloth. The need to incorporate environmental effects, crime attractors, and crime generators into geographic profiling models is well-documented (O'Leary, 2009), but the most-used geographic profiling methods (Rigel, Dragnet, and JTC) do not incorporate these predictive features. Instead, all of these methods make the assumption that criminal opportunities are uniformly distributed in geography. Recently, several different research groups have proposed Bayesian methods for incorporating predictive features into geographic profiling models (O'Leary, 2009; Levine and Block, 2011). However, these methods incorporate geographic representations of environmental features and known crime generators to filter (or screen out unlikely locations) through their use as prior probabilities in the Bayesian paradigm (rather than incorporating them as predictive features in the likelihood function). Thus, these models do not model how the journey to crime relationship changes in response to these predictive features or criminal opportunities. While the Bayesian modeling approach has facilitated consideration environmental effects into geographic profiling models, these models have not been able to provide better predictive performance than the much simpler CMD method (Levine and Block, 2011).

2.4.4 Prediction Model Performance Assessment

Because there are so many different approaches for developing predictive threat surfaces of criminal behavior (hot spot maps and geographic profiles), model selection and performance assessment are key considerations in employing these predictive models. The threat surface produced by the predictive algorithms can be thought of as a binary prediction of the probability of criminal incident at each individual location. Intensity measures produced by density approaches can be interpreted as a classification score or normalized into a 0 - 1 prediction probability rather than being regarded as an intensity rate. Thus, all threat surfaces can be evaluated using methods developed to assess the performance of binary classifiers.

Most of the statistics used to assess binary prediction model performance are based around the *confusion matrix*, which is a 2 x 2 contingency table which records counts of the four possible outcomes of a binary classification model: true positives, false positives, true negatives, and false negatives. Model performance measures derived from the confusion matrix include model sensitivity (also called hit rate, true positive rate, or recall), specificity, false positive rate, precision, accuracy, and F-measure (Fawcett, 2006). Recent years have seen one approach based on the confusion matrix become particularly popular for evaluating and comparing predictive algorithms in the machine learning community: the Receiver Operating Characteristic (ROC) curve (Fawcett, 2006, 2004; Spackman, 1989).

In founding the principles of signal detection theory, Birdsall and Fox (1954) advocated for assessing model performance using ROC curves. ROC curves illustrate the trade-offs between a model's *specificity* and *sensitivity* for all classification thresholds between 0 and 1. A model's specificity at a given threshold is the probability of the model asserting "false" when the actual state of the system is false. The model's *false positive rate* is calculated as 1-specificity and forms the horizontal axis of the ROC curve. A model's sensitivity at a given threshold is the probability of the model asserting "true" when the true state of the system is true. A model's sensitivity is also known as its *true positive rate*.

The ROC curve plots the cost-benefit trade-off for a classifier at all possible classification thresholds (Fawcett, 2004). The cost, plotted on the horizontal axis, is the model's false positive rate. The benefit, plotted on the vertical axis, is the model's true positive rate. Perfect prediction occurs when a model achieves 100% specificity and 100% sensitivity, which equates to a false positive rate of 0 and a true positive rate of 100%, plotted as the point (0,1) on the ROC curve.

ROC curves are a two-dimensional representation of classifier performance and often researchers would like to reduce performance to a single (scalar) statistic. The most common approach for summarizing ROC performance is to calculate the area under the ROC curve (denoted AUC) as a scalar value representing model classification performance (Fawcett, 2006; Hanley and McNeil, 1982). This statistic represents the probability that a randomly chosen positive incidence (in this application a randomly chosen location where a crime has occurred) will score higher than a randomly selected negative instance (i.e. a randomly selected location where a crime did not occur) (Fawcett, 2006). The AUC is equivalent to Wilcoxon test of ranks commonly used in categorical data analysis (Fawcett, 2006; Hanley and McNeil, 1982) and is also directly related to the Gini coefficient (Breiman et al., 1984). The AUC will always be a value between 0 and 1, and any AUC less than 0.5 indicates a model that performs more poorly than random guessing. The AUC is often considered to be the standard method to assess the accuracy of binomial classifiers, but as Lobo et al. (2007) records, it is much less reliable as a comparative measure of accuracy between models than a full ROC curve.

An alternative method for assessing threat surfaces specific to security applications is the *surveillance plot*, which defines the cost in the cost-benefit tradeoff in terms of the resources a security agency would need to expend in order to achieve a certain level of sensitivity (Huddleston and Brown, 2009). A surveillance plot records model sensitivity as a function of the surface area a security agency would need to monitor (surveil) and is an adaption of a model comparison metric introduced by Smith and Brown (2007). A model's trade-off in sensitivity and resources expended as shown in a surveillance plot is a very important characteristic for evaluating geospatial model performance because we desire a model that focuses security efforts in as small a geographic area as possible. The horizontal axis of the surveillance plot records the percentage of surface area, referenced in descending order of the threat surface (from highest probability/classification score to lowest probability/classification score), needed to observe the corresponding percentage of actual incidents in the dataset (the vertical axis).

The surveillance plot's definition of model cost provides a visualization of the efficiency of the model in allocating resources expended by a security agency attempting to observe or interdict the crimes as they occur. It is also easily extended for spatio-temporal modeling by rank-ordering the probability/score of spatio-temporal blocks throughout the spatial and temporal horizon of study (Fox and Brown, 2012). The best model is the one that achieves the highest percentage of observed incidents (sensitivity) for a given percentage of observed area required. The appropriate point of evaluation on the surveillance plot graph is contingent upon the resources available to the evaluating agency (Huddleston and Brown, 2009). For instance, if it is possible to monitor only 5% of the surface area in the study area with your available resources, then the model that provides the highest number of observed incidents in the test set in 5% or less of the high probability surface area observed is the best one for your organization. The use of surveillance plots in security applications is recommended because it couches ROC model performance assessment in terms of the resources security managers would need to expend in order to achieve a defined level of performance (percentage of crimes observed/interdicted), which are terms of reference much more familiar to security managers than statistical measures such as sensitivity and specificity. When applied to geographic profiling assessments, the surveillance plot is more commonly referred to as a plot of *effort rate* or *search cost*.

2.5 Crime Forecasting Methods

Gorr et al. (2003) notes the many applications of forecasting in policing:

> Municipal police would benefit greatly from accurate short-term forecasts of crime within small geographic areas, such as police precincts and patrol districts. Then it would be possible to target patrols to areas with forecasted crime increases, remove and redeploy special details in areas with forecasted crime decreases, schedule training and vacations in nonpeak periods, etc.

However, as Gorr and Harries (2003) note, crime forecasting, other than the naive methods used for CompStat programs, is not widely practiced by police, primarily because:

> ...the desired scale for observation is too small for reliable model estimation. For tactical purposes, police must pinpoint crime in areas as small as possible, at the patrol district level (geographic area of one officer or team) or smaller. We know that forecast errors increase as data aggregations become smaller but unfortunately, there has been very little systematic study of the effects of data scale on forecast accuracy.

Crime forecasting is also applied on a limited basis because, as recorded by Cohen and Gorr (2005), it is a relatively new discipline, with research in crime forecasting applications being driven primarily by research grants by the National Institute of Justice (NIJ) in the early 2000s. In one of those NIJ funded studies, Gorr et al.

(2003) establish a benchmark for forecasting performance in policing, stating that forecasting models become useful in police applications when forecasting errors are consistently less than 20% of the observed crime counts in the fixed observation units. They conclude in their exhaustive comparison of time series methods for forecasting at the car beat and precinct level in Pittsburgh that crime counts of 25 to 35 per fixed areal unit and time period were required to achieve this performance measure. Thus, they conclude that time series methods are primarily useful at the precinct level and higher because smaller fixed areal units do not provide enough crime counts for short-term crime forecasts for most crime types over the weekly or monthly horizons needed in policing.

In a summary of crime forecasting efforts from the early 2000s, Cohen and Gorr (2005) note that crime forecasts are in fact a time series: repeated measurements for a fixed observation unit (police precincts or car beats) and fixed time intervals (weekly or monthly). They categorize three types of time series methods used in crime forecasting: naive methods, univariate time series methods, and leading indicator models.

2.5.1 Naive Forecasting Methods

Naive forecasting methods are the most popular in police application, especially in the context of weekly or monthly CompStat meetings in which police administrators evaluate performance. The two most frequently used methods are the random walk approach and what Cohen and Gorr (2005) refer to as the "CompStat" method. The random walk forecast simply uses the previous time period's observed crime count as the forecast for the next period. Using a random walk forecast is appropriate in situations (such as the stock market) in which future behavior in the short term is unpredictable (random) and equally likely to move up and down (Malkiel, 2003). By definition, this is an inappropriate method for performance assessment of police activity in meetings such as CompStat, although it is widely used. The "CompStat" method is similar to the random walk approach, but uses the observed crime count from the same period one year earlier (Cohen and Gorr, 2005). Another naive approach not referenced in crime forecasting literature but sometimes applied in other forecasting applications is the use of the long-run average (Meade, 2000). None of these methods provides an informative forecast in policing applications.

2.5.2 Univariate Time Series Forecasting Methods

There are a plethora of time series methods available, although few of them have been applied to the problem of forecasting time series of crime over small geographic areas such as police precincts. Commonly employed time series methods include: time series regression, moving average models, exponential smoothing models (including extensions for seasonality and trend), Auto-Regressive Moving Average (ARMA) models, Auto-Regressive Integrated Moving Average (ARIMA) models, ARMA models with exogenous inputs (ARMAX), spectral time series models, state-space time series models, and Generalized Auto-Regressive Conditionally Heteroscedastic (GARCH) models (Shumway and Stoffer, 2006). Of these many methods, crime analysis and forecasting literature contains almost exclusively time series models from the nested exponential smoothing family as developed by Brown (1959; 1963), Holt (1957; 1960; 2004), and Winters (1960). Exceptions include several applications of ARIMA models used to forecast city-level crime rates (Chen et al., 2008b,a; Chamlin, 1988) and an unpublished effort leveraging time series regression (Pepper, 2007).

Gorr et al. (2003) demonstrate that univariate time series forecasting methods can significantly improve upon the naive forecasting approaches discussed above. This result is significant in that it establishes that crime counts in small geographic regions such as police precincts are not outcomes of a random walk, but rather can be predicted to a certain extent. As Cohen and Gorr (2005) note, police can therefore use crime forecasts developed from time series forecasting methods to evaluate police performance by leveraging these forecasts as "counterfactual" estimates of crime counts which can be compared to observed crime counts to assess police performance as well as employing them to proactively focus resources to needed areas in the next time period.

The nested exponential smoothing family of models (which includes the Holt-Winters method) and ARIMA models are the most popular time series methods in both business and crime forecasting for two reasons. First, both methods provide very flexible modeling frameworks that are robust to many types of time series patterns such as trends, seasonality, or unusual changes in the pattern such as the introduction of shocks (Hyndman and Khandakar, 2008). Second, even non-statisticians can easily automate Holt-Winters and ARIMA forecasting models using widely available software. For example, Microsoft Excel easily optimizes the parameters of a Holt-Winters smoothing model using Solver and freely available statistical software such as *R* provides algorithms for automatically fitting ARIMA models (Hyndman and Khandakar, 2008) .

Another important factor in the prevalence of exponential smoothing and ARIMA models is the lack of a successful identification procedure for determining the optimal time series method for a given forecasting problem prior to fitting the models. Researchers have proposed complicated expert systems for selecting appropriate time series models (Collopy and Armstrong, 1992; Arinze, 1994; Vovurka et al., 1996; Meade, 2000; Adya et al., 2001), but as Gardner (2006) records, these approaches produce mixed results at best. Additionally, fitting many of these models requires the use of statistical software often not available to crime analysts. Thus, naive, exponential smoothing, and ARIMA methods provide benchmarks for the performance of any new method for recurring short-term demand forecasts in many applications.

Several crime forecasting studies (Cohen and Gorr, 2005; Gorr, 2009) have recorded that Holt-Winters models with seasonality provide the best approach to routine forecasts for fixed geographic regions such as police precincts. However, these same studies also identify that another forecasting approach, multivariate leading indicator models, provides better performance in predicting exceptional conditions such as impending spikes in crime rates for specific geographic regions.

2.5.3 Leading Indicator and Time Series Monitoring Methods

Leading indicator crime models are designed to "make accurate forecasts of the relatively rare, large changes in crime (Cohen et al., 2007)." Note that in the terminology used in this dissertation, this is a *prediction* - an assertion of the probability that in the next time period there will be a significant spike (or drop) in crime. Leading indicator crime models use predictor variables such as selected lesser crimes to predict lagged changes in dependent variables such as violent crimes. Specifically, when the leading indicating variables (such as property crime) step up (or spike significantly) these models predict that violent crimes will increase (or spike) in the very near future (Gorr, 2009). These models often provide relatively poor performance in forecasting crime counts under normal conditions, but out-perform traditional forecasting methods in predicting "exceptional conditions," or spikes in violent crime (Gorr, 2009; Cohen et al., 2007).

These leading indicator models are used in crime analysis in *proactive* Management by Exception (MBE) (Gorr, 2009). MBE is a management principle developed by Taylor (1917, 1912) in which management devotes its time to those situations departing significantly from planned (or forecasted) results. This is ostensibly the purpose of weekly or monthly CompStat meetings often used in police management, although as previously discussed, the common use of naive forecasting methods significantly limits the utility of MBE when naive forecasting methods are applied. The leading indicator models developed by Cohen et al. (2007) provide a significant advantage in that they can be used to identify when crime spikes are likely to occur in specific police precincts or car beats, facilitating a shifting of resources to the affected region prior to the spike. However, these significant shifts in crime counts are relatively rare. For example, Gorr and McKay (2005) identify in a study of Pittsburgh crime that signifiant shifts in violent crime occur in high crime areas approximately twice every three years.

It is much more common for MBE to be applied retroactively (as in weekly or monthly CompStat meetings) to assess whether or not the observed crime count represents a significant departure from the forecasted crime count. Gorr (2009) terms this use *reactive* MBE. This analysis is easily conducted using methods from manufacturing such as Shewhart's control charts (Shewhart, 1939) or the tracking signals approach developed by Trigg (1964). Gorr and McKay (2005) provides an overview on the use of these models in retroactive performance assessment for crime applications while Huddleston et al. (2010) provide a similar discussion for military applications.

2.5.4 Forecasting Model Performance Assessment

Commonly used measures of forecast accuracy used in forecasting crime include Mean Absolute Deviation (MAD), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Median Symetric Absolute Percentage Error (mdsAPE) (Gorr, 2009). Several of these measures, specifically the MAPE and mdsAPE, are undefined for time series values of 0 (i.e. no crimes observed) during at least one of the time periods. This is because these measures use the observed crime count in the denominator, yielding infinite results when no observation occurs. Thus, in applications (such as weekly forecasts of crime counts at the precinct or car beat level) when it is possible that no crime counts occur, these measures of performance fail. Kolassa and Schutz (2007) introduce the MAD/MEAN (MAD divided by the historical mean) statistic as a measure equivalent to the MAPE that addresses this significant shortcoming.

As Hyndman and Koehler (2006) note, MSE and RMSE have been historically popular "largely because of their theoretical relevance in statistical modeling." The RMSE can also be used to develop uncertainty estimates (prediction intervals) for future forecasts when used as a forecasting model standard error. However, Armstrong and Collopy (1992) demonstrate that the RMSE is not a reliable accuracy measure for comparing forecasting performance across time series.

Hyndman and Koehler (2006) propose the MASE as a "generally applicable measurement of forecast accuracy without the problems seen in the other measurements." MASE provides an ideal statistic for assessing forecasting performance in many crime applications due to several properties. First, MASE is a scale-free measure of forecast error, which allows forecast accuracy comparisons between series. Secondly, MASE can be used on intermittent time series, which contain observed event counts of 0 in at least one time period. Event counts of 0 occur frequently in many crime forecasting and business demand forecasting applications. Finally, MASE scales the error by the Naive forecast. A MASE score less than one indicates a model that has smaller average error than the Naive method, with the reverse true when $MASE > 1$. Thus, the MASE statistic reports a scaled effect size performance improvement (or loss) as compared to the most commonly used forecasting method in security applications (i.e., the difference between the MASE statistic and 1 is the effect size for performance improvement or loss).

2.6 Patrol District Design

2.6.1 Police Patrol District Design and Geo-Policing

Police departments create geographic patrol districts (also called patrol sectors or car beats) as a standard management method to enhance the capabilities of the uniformed patrol force (Hale, 1980). In most police departments, one patrol unit is assigned to each patrol sector during each patrol shift (Gorr and Kurland, 2012). Although current crime analysis and crime mapping texts discuss the use of patrol district boundaries in the course of supporting police operations, they do not provide methodological approaches for designing patrol districts (Paulsen et al., 2010; Gorr and Kurland, 2012). Traditionally, the geographic patrol boundaries for a police department are drawn by hand based on a police department's knowledge, experience, and the available police resources (Mitchell, 1972; Taylor and Huxley, 1989). However, Curtin et al. (2010) note that given the complexities of the police districting plans, it is unlikely that an optimal districting plan will be chosen by relying only on the judgement and intuition of police planners.

2.6.2 Military Area of Operation (AO) Design

In US military doctrine, all geographic Areas of Responsibility (AOR) are divided into Areas of Operation (AO) at the tactical and operational levels for subordinate elements (U.S. Army, 2010). AOs define the areas that specific subordinate commanders are responsible for, and are therefore doctrinally assigned to commands from company level (approximately 100-200 soldiers) to division level (approximately 15,000 soldiers). It is a common practice for company level commanders to further divide the company AO into geographic sectors assigned to subordinate platoons or squads, but due to the higher intensity of operations in counter-insurgency and area security operations individual vehicles and soldiers rarely operate independently. Patrol sectors are instead patrolled by small military units such as squads (for foot patrols) or platoons (for vehicle patrols). While military doctrine contains a long list of key considerations in the development of unit AOs, it does not provide a methodological approach for applying those considerations in a practical situation (U.S. Army, 2008, 2009). Instead, military planners draw boundaries by hand in much the same way that police planners do. However, unlike police districting plans, military AOs change frequently due to changing force structures, requirements, and the evolving enemy situation.

2.6.3 Existing Methods for Security Force Geographic Mission Assignment

While there is no published research on developing military AO boundaries, researchers have proposed several different approaches for optimizing police patrol boundaries in a city. Most of these approaches rely on defining the problem mathematically as a set-covering or optimization problem. Considered approaches include using p-Median clustering to minimize the total weighted travel distance to service expected calls (Mitchell, 1972); solving a graph-partition problem with the constraints of contiguity and compactness, which has been shown to be NP-hard (Altman, 1997; Johnson, 1985); and heuristic methods for identifying "good" (locally optimal) districting plans including simulated annealing (D'Amico et al., 2002), maximal covering models (Curtin et al., 2010), genetic algorithms, or stochastic gradient ascent (Zhang and Brown, 2013).

An alternative approach is the use of simulation models to evaluate existing patrol district designs. Several researchers have developed and analyzed patrol districts by developing discrete event simulation models based on the Hypercube Queuing Model (HQM) (Zhang et al., 2013; Boyaci and Geroliminis, 2011). The HQM is a well-known descriptive model used to analyze emergency response systems as a spatially distributed queueing system (Larson, 1974). Others have developed agent-based simulation models to evaluate automated methods for generating police patrol district designs (Zhang et al., 2013; Zhang and Brown, 2012). However, all of the above approaches are unlikely to be implemented in most police departments and military units because the approaches are too complex for the average analyst to implement and require statistical or simulation modeling software that police and military units do not have. Therefore, a heuristic approach for police district design using methods and software military and police units have available to them is needed.

2.7 Summary of Taxonomy and Literature Review

As this taxonomy and literature review has shown, there are currently several significant gaps in security planning capabilities at the tactical and operational level. This section briefly summarizes the existing capability gaps identified in this literature review that motivate the methods developed in the subsequent chapters of this dissertation.

The first existing gap in the crime analysis literature is the current lack of effective methods for developing short-term crime forecasts within small geographic areas. While there have been relatively few studies in developing crime forecasts within small geographic regions, the existing crime forecasting methods require fitting models for every unique geographic region. In many military and police applications, this requires a good deal of modeling effort. Simple and accurate methods for developing regular forecasts at the precinct and car patrol sector level for time periods as short as weekly forecasts are needed to support police and military resource allocation and unit assessment efforts.

The second existing need in the crime analysis literature is the development of effective methods for addressing "the deployment, staffing, and redistricting of police beats or precincts (Boba, 2005)." There is currently no established procedure available for developing geo-policing units based on forecasted demand that is likely to be feasible for the vast majority of military and police units. The existing approaches are too complex for the average analyst to implement and require statistical modeling software that police and military units do not usually have.

Another potential area of contribution to the crime analysis literature includes the development of products that provide intelligence assessments of criminal groups. As noted, Huddleston and Brown (2009) develop a new method using multilevel Criminal Site Selection (CSS) models to improve the predictive performance of hot-spot maps for individual criminal gangs. They also provide a limited demonstration of the development of intelligence assessment products that map the spheres of influence for competing criminal groups. This dissertation extends this previous work to consider all of the criminal gangs in a geographic region and to develop additional products used in intelligence assessments of criminal groups.

Finally, as Huddleston et al. (2008) and Huddleston and Brown (2009) identify, the most important predictive variable in a CSS model for predicting crimes by criminal gangs is the distance-decay relationship from their gang headquarters location. This distance-decay relationship is leveraged in the JTC models used in criminal investigations to isolate the most likely anchor points (homes or work locations) for serial offenders. However, existing JTC models consider only the distance-decay effect, ignoring the effect of many other environmental factors. Existing JTC methods have not provided significant performance improvement over simple centrographic methods because even simple methods such as CMD can accurately capture this relationship. CSS models may provide the opportunity to estimate the effect of the distance-decay relationship *after* the effect of other environmental facts have been considered. This dissertation applies CSS models to the geographic profiling problem in an effort to improve performance over the current-best CMD method.

Chapter 3

Geographic Probability Forecasting

Every day, government executives, police officials, and military leaders must decide how to most efficiently and effectively employ their limited resources in an effort to secure the large and diverse populations they are charged to protect. Planning and decision-making processes for these leaders are often oriented around dividing limited resources across subordinate commands and geographic regions. Military and police leaders regularly rely on recurring forecasts of criminal events indexed by geographic region to support these processes. For example, Gorr et al. (2003) note the many applications of these forecasts for police use including: efficient resource allocation, geographic mission assignment planning, and unit performance assessment. Therefore, both military and police units stand to benefit from accurate models for producing regular forecasts of geographic time series.

Geographic time series are counts of events indexed over time by geographic region. Often, geographic time series are very noisy, with the observed counts by region varying considerably from period to period and region to region. This high variance is to be expected for time series of criminal events since these time series are the result of the superposition of many low-intensity point processes. The Palm-Khintchine theorem asserts that the superposition of many low-intensity independent renewal processes behaves asymptotically as a Poisson process (Heyman and Sobel, 1982). Thus, when the actions of many criminals acting independently in geography generate a time series, the time series should exhibit the noisy behavior of a Poisson process, in which the variance of event counts in a period is equal to the average count of a period. These noisy geographic time series are very difficult to forecast accurately.

This chapter documents the development of a new, simple method for crime forecasting and performance comparison of the new method to the three traditional methods most commonly used for crime forecasts: naive methods, Holt-Winters smoothing, and the ARIMA (Box-Jenkins) class of models. The new method, Geographic Probability Forecasting (GPF), simplifies the modeling process for Holt-Winters smoothing and ARIMA models by combining these univariate time series methods with predictive hot-spot maps, tools commonly used in law enforcement applications for identifying areas with a high probability of criminal activity. This new modeling approach significantly improves forecasting performance in both a motivating example and an in depth simulation study. The GPF method also significantly reduces the modeling workload by dramatically reducing the number of models and model parameters needed for producing recurring short-term demand forecasts. These results suggest that the GPF modeling approach provides a simple, robust, general purpose method for improving forecasts for noisy geographic time series.

3.1 Background

As Hyndman and Khandakar (2008) note, “Automatic forecasts of large numbers of univariate time series are often needed in business.” Commonly employed univariate time series methods include: time series regression, moving average models, exponential smoothing models (including extensions for seasonality and trend), Auto-Regressive Moving Average (ARMA) models, Auto-Regressive Integrated Moving Average (ARIMA) models, ARMA models with exogenous inputs (ARMAX), spectral time series models, state-space time series models, and Generalized Auto-Regressive Conditionally Heteroscedastic (GARCH) models (Shumway and Stoffer, 2006). Of the many available time series methods used, the most commonly used (Hyndman and Khandakar, 2008) are the ARIMA family of time series models developed by Box and Jenkins (1990) and the various exponential smoothing methods developed by Brown (1959; 1963), Holt (1957; 1960; 2004), and Winters (1960).

The most used forecasting methods in police applications are the naive approaches of referencing the crime count from the previous time period (i.e. previous week or month) or the crime count from the same period twelve months earlier (Gorr et al., 2003). Gorr et al. (2003) demonstrate that the Holt-Winters exponential smoothing approach can significantly improve upon the naive forecasting in crime forecasts, and the crime analysis and forecasting literature contains almost exclusively time series models from the nested exponential smoothing (Holt-Winters) family. The several exceptions to this rule include several applications of ARIMA models used to forecast city-level crime rates (Chen et al., 2008b,a; Chamlin, 1988) and several studies incorporating leading indicator models. Leading indicator crime models are designed to "make accurate forecasts of the relatively rare, large changes in crime (Cohen et al., 2007)." Although leading indicator models out-perform traditional forecasting methods in predicting exceptional conditions, such as spikes in violent crime, they often provide very poor performance in forecasting event counts under normal conditions, making them unsuitable for recurring short-term demand forecasts (Gorr, 2009; Cohen et al., 2007).

Exponential smoothing and ARIMA models are the most popular time series methods in both business and crime forecasting for two reasons (Hyndman and Khandakar, 2008). First, both methods provide very flexible modeling frameworks that are robust to many types of time series patterns such as trends, seasonality, or unusual changes in the pattern such as the introduction of shocks (Hyndman and Khandakar, 2008). Second, even non-statisticians can easily automate Holt-Winters and ARIMA forecasting models using widely available software. For example, Microsoft Excel easily optimizes the parameters of a Holt-Winters smoothing model using Solver and freely available statistical software such as R provides algorithms for automatically fitting ARIMA models (Hyndman and Khandakar, 2008). Thus, naive, exponential smoothing, and ARIMA methods provide benchmarks for the performance of any new method for recurring short-term demand forecasts in many applications.

3.2 Problem Definition

To formally define the forecasting problem considered in this chapter, let Y_t denote the number of events that occur within the domain of interest D during the time period t. The region of interest D contains sub-regions indexed by j: $\{D_1, D_2, ..., D_J\}$. The quantity of interest is Y_{jt}: the event count for each of the J sub-regions during each time period t. The problem considered is the regular production of one-step ahead forecasts for the noisy geographic time series indexed by Y_{jt}. This study uses a real-world crime data set to motivate this discussion and subsequently uses a simulation study to more formally compare the performance of the considered modeling approaches under many different time series patterns, including the introduction of trends, seasonality, and shocks.

3.3 Pittsburgh Burglary Data

Figures 3.1 and 3.2 provide illustrations of the data sets used as motivating examples in this chapter. These data sets contain records of burglary incidents in the city of Pittsburgh over two different time periods. Gorr and Kurland (2012) provide the first data set in a recently published GIS tutorial designed to "teach crime mapping and

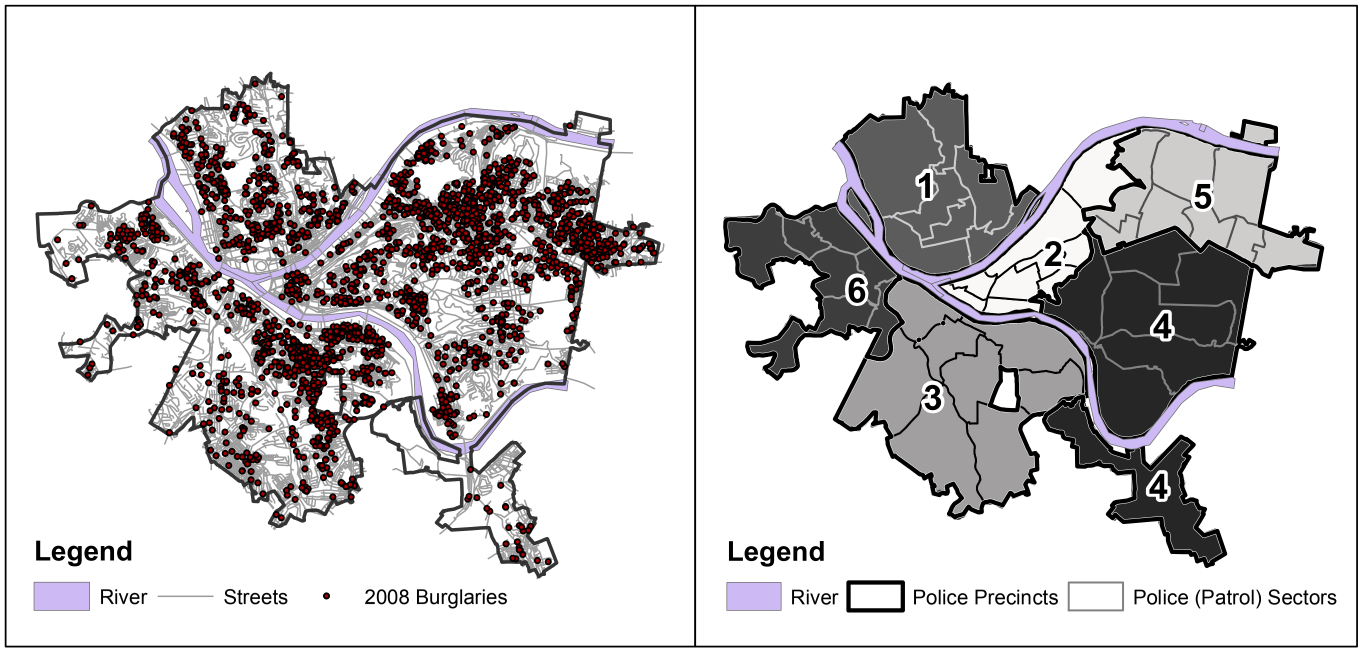


Figure 3.1: Illustration of the data set used for this analysis. The panel on the left plots the locations for the 3093 burglaries from the year 2008 in the city of Pittsburgh. The panel on the right illustrates the 6 police precincts and 46 car beats (patrol sectors) used by the Pittsburgh Police Department.

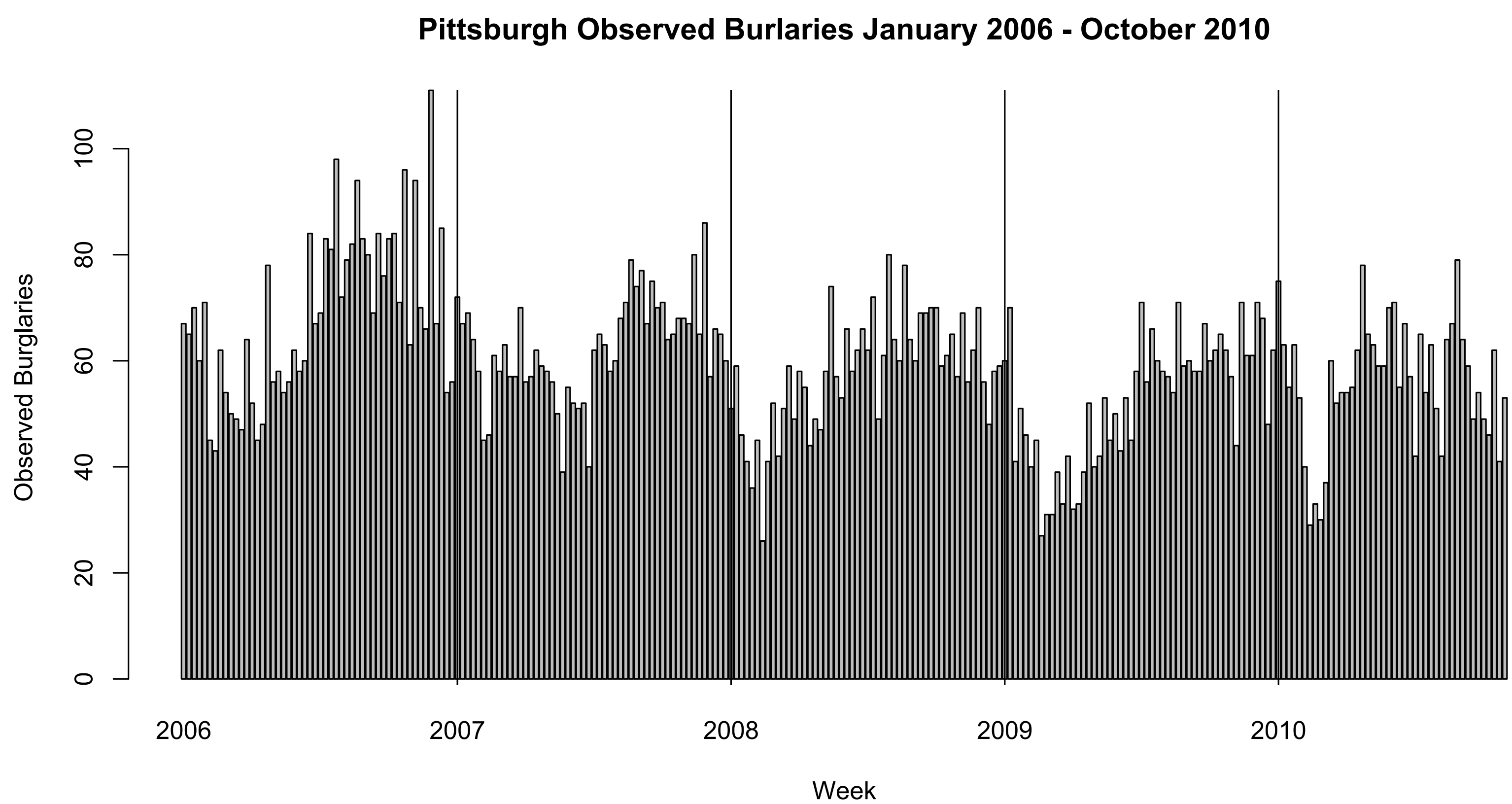


Figure 3.2: Illustration of Pittsburgh weekly burglary counts for the period from 1 January 2006 through 31 October 2010.

analysis skills using ArcGIS Desktop software." As the authors note, "...this book uses real crime data obtained from the Pittsburgh Police Bureau and the Allegheny County 911 Center in Pennsylvania." This publicly available data set contains the burglary records for the City of Pittsburgh during the year 2008. It also provides geographic data for Pittsburgh's 46 patrol car sectors that are made up of an average of three census tracts each. These patrol units are overseen by the six police precincts, with each precinct containing seven to nine car patrol sectors.

The second data set contains records of burglary incidents for the City of Pittsburgh from 1 January 2006 through 31 October 2010. This data is not publicly available but was obtained through a research agreement with Carnegie Mellon University, which works closely with the City of Pittsburgh Police. This multi-year data set allows performance comparison of the various forecasting methods over periods in which seasonality effects can be modeled. The motivating problem for both analyses is the weekly requirement to generate one-week-ahead burglary forecasts (a very resource-intensive crime type) for the 46 police sectors and six precincts in Pittsburgh. This requires a total of 52 one-step-ahead forecasts on a weekly basis.

3.4 Motivating Principle for Geographic Probability Forecasting

The motivating principle for the Geographic Probability Forecasting (GPF) methodology lies in exploiting the reduction in error variance gained by forecasting at an aggregated (i.e. city-wide) level and dividing this forecast according to a geographic probability. Examining the simple problem of estimating the mean of a process based upon a sample of observed counts in a geographic time series provides insight into how the GPF methodology improves forecasts. Consider a sequence of observations X_t that come from a Gaussian distribution $N(\mu, \sigma^2)$. The error for the sample mean estimated during time t is bounded for an arbitrary value ϵ as follows (Shumway and Stoffer, 2006):

$$P\{|\overline{x} - \mu| > \epsilon\} \leq E\left[(\overline{x}_t - \mu)^2\right] = \frac{\sigma^2}{t\epsilon^2} \tag{3.1}$$

Now consider that the sequence of events that make up X can be split between several geographic sub-regions. The count of observations X_t is made up of the sum of the counts in the several sub-regions. For sub-regions indexed by j:

$$X_t = \sum_{j=1}^{J} X_{jt} \tag{3.2}$$

If the mean of the distribution for sub-region j is some known fixed percentage w_j of the mean of the distribution for the region of interest, then one can use the sample mean of the higher distribution to develop an estimate of the mean for the sub-region. For sub-region j, assert that X_{jt} is Gaussian distributed $N(\mu_j, \sigma_j)$. Define that:

$$\mu_j = w_j \mu \tag{3.3}$$

where:

$$\mu = E[X] \tag{3.4}$$

Two approaches for estimating the mean for region j now exist. One can use the estimate $\overline{x}_{jt}$ for μ_j during time t or one can weight the region estimate such that:

$$\hat{\overline{x}}_{jt} = w_j \overline{x}_t \tag{3.5}$$

Using Equations 3.3 and 3.5, the error bound of the weighted estimate is:

$$P\{|w_j\overline{x}_t - \mu_j| > \epsilon\} \leq E\left[(w_j\overline{x}_t - \mu_j)^2\right] = w_j^2\frac{\sigma^2}{t\epsilon^2}$$

The error bound for the region sample mean based upon the observations (counts) in that regions is:

$$P\{|\overline{x}_{jt} - \mu_j| > \epsilon\} \leq E\left[(\overline{x}_{jt} - \mu_j)^2\right] = \frac{\sigma_j^2}{t\epsilon^2} \tag{3.6}$$

Therefore:

$$P\{|w_j\overline{x}_t - \mu_j| > \epsilon\} \leq P\{|\overline{x}_{jt} - \mu_j| > \epsilon\} \Leftrightarrow w_j^2 \leq \frac{\sigma_j^2}{\sigma^2} \tag{3.7}$$

For noisy geographic time series, this inequality holds true. A numerical example illustrates the dramatic reduction in estimation error that this weighting estimate can provide. If the counts for the region (i.e. city) come from a stationary (temporal component) homogenous (geographic component) Poisson process with mean and variance λ, then for a Poisson process with a rate greater than 10, the counts observed during each period are approximately Gaussian distributed as $N(\mu = \lambda, \sigma^2 = \lambda)$. If the region is divided into four equal geographic sub-regions, then for each sub-region, $\lambda_j = \frac{1}{4}\lambda$. In this example, by substitution:

$$P\{|w_j\overline{x} - \mu_j| > \epsilon\} = w_j^2\frac{\lambda}{t\epsilon^2} = \frac{1}{16}\frac{\lambda}{t\epsilon^2} < \frac{1}{4}\frac{\lambda}{t\epsilon^2} = w_j\frac{\lambda}{t\epsilon^2} = P\{|\overline{x}_j - \mu_j| > \epsilon\} \tag{3.8}$$

In the example above, both sample estimates of the mean ($\overline{x}$ and $\overline{x}_j$) are converging to their respective means (μ and μ_j). However the probability-weighted estimate $\hat{\overline{x}}_j = w_j\overline{x}$ converges to μ_j much faster than than $\overline{x}_j$ does. This result holds for any stationary Poisson process, regardless of the observed rates, because the relationship described in Equation 3.7 holds true for any Poisson process in which the region counts are made up of the sum of sub-region counts. The GPF methodology exploits this principle to improve the forecasts for noisy geographic time series by using kernel density estimation to estimate w_j.

3.5 Methodology

The GPF method consists of four modeling steps:

1. Develop a Spatial Geographic Probability Model (Hot-Spot Map)
2. Convert the Geographic Probability Model into a Region Event Probability
3. Develop a Domain Level Forecast
4. Spatially Weight the Domain Forecast Using the Geographic Probability Model

The following sections explain each of the modeling steps in detail.

3.5.1 Develop a Spatial Geographic Probability Model

The first GPF modeling step is to develop a spatial geographic probability model for event occurrence. These models are ubiquitous in crime analysis, where they are colloquially known as "hot-spot maps." The National Institute of Justice defines a criminal hot-spot as "an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization (Eck et al., 2005)." Police use two broad classes of methods to develop hot-spot maps. Methods in the first class statistically model the relationships between crime occurrence and environmental factors, geographic features, and other predictors (Huddleston et al., 2012; Huddleston and Brown, 2009; Smith and Brown, 2007). This approach requires sophisticated modeling software that is limited to only the largest police agencies. The second approach uses bivariate kernel density estimation to produce hot-spot maps and is widely used because the Geographic Information Systems (GIS) most police agencies now employ produce kernel density maps with ease. Figure 3.3 provides an example kernel density hot-spot map used for this problem.

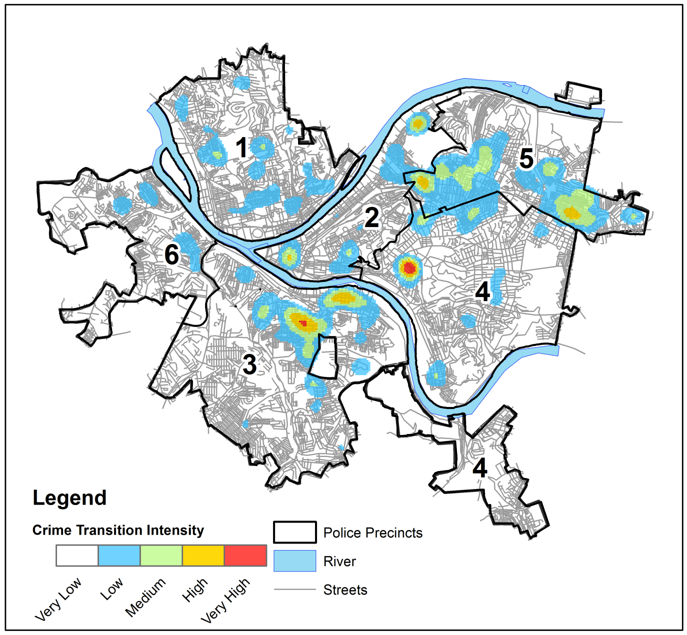


Figure 3.3: Illustration of the spatial probability model (hot-spot map) generated to predict Pittsburgh week 51 burglary activity using burglaries observed during weeks 1 through 50.

The use of the kernel density estimation requires some additional notation. Let b_i index two dimensional blocks within a spatial study region $D \subset \Re^2$. These two-dimensional spatial blocks denote unique locations created by laying a grid at a fine resolution across the study region: $\{b_1, b_2, ..., b_I\}$.

Note that:

$$\cup b_i = D \tag{3.9}$$

and

$$b_i \cap b_j = 0 \; \forall \; i, j \tag{3.10}$$

Let s_y denote the location in $\Re^2$ of event y and Y the total number of events occurring within D during the time period used to fit the model. The event intensity, $f(b_i)$, for each location is calculated using the kernel density function.

$$\hat{f}_h(b_i) = \frac{1}{hY} \sum_{y=1}^{Y} K\left(\frac{\|b_i - s_y\|}{h}\right) \tag{3.11}$$

In Equation 3.11, the notation $\|b_i - s_y\|$ denotes the Euclidean norm (distance) between location b_i and event s_y. Model fitting requires the selection of the kernel function K and the bandwidth parameter h. The choice of the kernel function K has relatively little effect on the kernel density model performance. The bandwidth parameter h does significantly affect model performance. Various statistical procedures automate the selection of the modeling parameter using plug-in estimates (Sheather and Jones, 1991; Jones et al., 1996; Berman and Diggle, 1989). The approach demonstrated here selects the plug-in estimate for bandwidth that minimizes the Mean Squared Error (MSE) of the hot-spot map over the previously observed time horizon using the procedure outlined by Berman and Diggle (1989).

3.5.2 Convert the Geographic Probability Model (Hot-Spot Map) into a Region Event Probability

The weighting parameter $\hat{w}_j$ represents the spatial probability-weight for geographic sub-region j derived from the kernel density hot-spot map.

$$\hat{w}_j = \frac{\sum_{b_i \in D_j} \hat{f}_h(b_i)}{\sum_{b_i \in D} \hat{f}_h(b_i)} \tag{3.12}$$

The weighting factor $\hat{w}_j$ captures the proportion of overall event probability across the region that falls within the geographic sub-region D_j. Note that this risk weighting both converts the kernel density estimate into a probability estimate over the region of interest and removes any of the risk probability that falls outside of the considered region (because the kernel density estimate will map onto a square grid whose boundaries extend beyond the considered region). GIS systems easily automate the calculation of the weighting factor $\hat{w}_j$ for any subset of the region (precincts, patrol sectors, etc). Table 3.1 provides the precinct risk weights for the Pittsburgh burglaries in week 51 of the year 2008 as calculated from the hot-spot map shown in Figure 3.3. Note that Table 3.1 neatly summarizes the difference in risk in these six precincts, with Precinct 6 the safest and Precinct 5 exhibiting the highest risk for burglary.

Precinct	1	2	3	4	5	6
$\hat{w}_j$	15%	11%	24%	19%	25%	6%

Table 3.1: Precinct estimated probability-weights for Week 51 forecast.

3.5.3 Develop a Domain Level Forecast

The third step in the GPF methodology is to develop a region (city-wide) forecast for event counts. As previously noted, Holt-Winters exponential smoothing and ARIMA models both provide easily automated, highly flexible modeling frameworks for generating recurring forecasts. Both methods are demonstrated in this example to mimic possible use in actual practice.

Fitting a Holt-Winters (HW) model requires optimizing the needed model parameters for mean, trend, and seasonality. For the Holt exponential smoothing model (employed on the 1-year Pittsburgh burglary data), this requires estimates for the two needed model parameters (smoothing and trend). Fitting Holt-Winters exponential smoothing models also requires estimates for each of the seasonal effects (i.e. 52 weekly effects over the course for the the Pittsburgh multi-year burglary study). We accomplished the needed model parameter estimation by using Nelder-Mead optimization to minimize the mean squared forecasting error over the previously observed time periods. Nelder-Mead optimization uses a simplex to estimate the direction of steepest descent, iteratively converging on the optimal point (Nelder and Mead, 1965). We implemented this model fitting procedure using the *stats* package in R software. Microsoft Excel spreadsheets using Solver also provide the ability to complete this model fitting procedure using brute-force optimization methods, a forecasting approach therefore available to virtually any analyst.

The automatic forecasting procedure for fitting ARIMA models is much more complex. In this modeling procedure, all appropriate ARIMA models are considered, with the best-fitting model selected according to Akaike Information Criteria (AIC) score. This procedure requires applying all appropriate ARIMA models to the training data (the time series from previous weeks), optimizing the parameters for those models (in the same manner as for the HW models above), and selecting the best model according to the AIC score calculated over the previous time periods (Hyndman and Khandakar, 2008; Akaike, 1974). We used the $forecast$ package in R software to execute this procedure model-fitting procedure (Hyndman and Khandakar, 2008). This approach mimics a modeling approach available to analysts with more formal statistical training and the ability to use statistical modeling software.

3.5.4 Spatially Weight the Domain Forecast Using the Geographic Probability Model

The final step of the GPF methodology is to break the region-level (aggregated) forecast across the geographic sub-regions using the probability weights $\hat{w}_j$. The forecast for region j in the next time period is a function of the region (aggregated) level forecast for that time period F_t and the estimated probability weight w_j:

$$F_{jt} = \hat{w}_j F_t \tag{3.13}$$

For example, if the region (aggregated) level burglary forecast for Week 51 is 51.4 crimes, then when that aggregated forecast is multiplied by the spatial risk-weights in Table 3.1, the precinct forecast for Precinct 1 is 7.71 crimes while the forecast for Precinct 2 is 5.65 crimes.

3.5.5 GPF Modeling Assumptions

The GPF modeling approach incorporates several modeling assumptions that should be addressed. First, the GPF model assumes that the spatial distribution of the events (as described by the kernel density hot-spot map) remains stationary over the period used to fit the model and forecast. The size of this modeling horizon is flexible (i.e., you can define how much previous history to use in estimating the spatial model: six months, one year, etc.). In the one-year Pittsburgh burglary study, the entire data horizon from Week 1 up to Week $t-1$ is used to forecast Week t. In the multi-year Pittsburgh burglary study and simulation study that follow in later sections, we examine the use of rolling time horizons to develop the kernel density estimates for the spatial distribution of crime events.

The second assumption is that any underlying trend or seasonality that affects event counts in one sub-region affects all sub-regions. Thus, any existing trend (or seasonal effect) applies to all precincts (and patrol sectors) simultaneously. The GPF model therefore models the crime counts as a separable space-time process in which the spatial probability density of events is fixed (or changes slowly) while the distribution of counts (rate) can change rapidly. This dissertation investigates the robustness of the GPF modeling approach to violations of these assumptions in later sections.

3.6 Comparison of Modeling Effort

The various forecasting methods considered require different levels of resourcing and modeling effort to fit. For each of the different methods outlined below, we assign a modeling effort score based on the number of forecasting models that must be fit, whether or not a geographic weighting model is needed, and the relative complexity of the forecasting and weighting models used. These scores are subjective and based upon the our familiarity with the capabilities of analysts who work in the considered problem domain. In other domains, the assigned modeling scores may differ based on the resources available to and experience of the forecasting analysts in that problem domain. Table 3.2 provides a summary of the calculation of the modeling effort scores, which are discussed in depth for each of the modeling methods below.

The naive forecast represents the simplest possible forecasting method and is the most-used method in law enforcement. The forecast for the next time period is simply the observed count from the previous time period:

$$Y_{jt} = Y_{j(t-1)} \tag{3.14}$$

Because this method requires no modeling, the naive method has a modeling effort score of 0.

The first univariate time series method studied is the use of Holt (for the one-year Pittsburgh burglary study) and Holt-Winters (in the multi-year Pittsburgh burglary study) forecasting models to develop disaggregated forecasts for each geographic region. Simple exponential smoothing models estimate only a smoothing parameter. Holt exponential smoothing models add a parameter for estimating trends. Holt-Winters exponential smoothing models estimate both trend and seasonality effects. We assign a model complexity score to the Holt/Holt-Winter (HW) procedure of 1. Using the HW method requires fitting a model for each of the six precincts and 46

Method	Forecasting Models	Forecast Model Complexity	Weighting Models	Weighting Model Complexity	Effort Score
Naive	0	0	0	0	0
Holt-Winters	52	1	0	0	52
GPF-HW	1	1	1	2.5	3.5
ARIMA	52	2	0	0	104
GPF-ARIMA	1	2	1	2.5	4.5

Table 3.2: Summary of forecasting model effort scores based on the number of forecasting models needed, whether or not a geographic weighting model is needed, and the relative complexity of the forecasting and weighting model used.

patrol sectors, for a total of 52 forecasting models on a weekly basis.

We assign a modeling effort score of 2 to the ARIMA model class. This assignment is based on the fact that, unlike the Holt-Winters method in which optimization of the needed parameters can be accomplished using spreadsheet software, this approach requires sophisticated statistical software and an analyst that knows how to use the software. As discussed in Section 3.5.3, this method is fairly complex, and requires significantly more statistical acumen. Using the ARIMA method also requires fitting 52 models for the motivating problem, one for each precinct and patrol sector.

Fitting the GPF models requires the use of the Geographic Information Systems (GIS) that most police agencies employ to automate the calculation of the kernel density map and weighting factor $\hat{w}_j$ for any subset of the domain (precincts, patrol sectors, etc). Virtually any crime analyst can complete this modeling procedure. The complexity of fitting the kernel density estimate corresponds closely to that of the automated ARIMA models discussed above. The procedure requires specific statistical software and analysts familiar with the modeling procedure. Therefore, we assign a model complexity score of 2 to the procedure in Equation 3.11 to reflect the similarity in modeling requirements to the ARIMA procedure. The requirement to subsequently calculate the weighting parameter in Equation 4.2 adds an additional 0.5 to the weighting model complexity score (for a total weighting model score of 2.5). This estimate yields overall model complexity scores of 3.5 (GPF-HW) and

4.5 (GPF-ARIMA). As with the simple exponential smoothing approach, different time horizons can be used to estimate the weighting parameters. For the one-year Pittsburgh burglary study using publicly available data, GPF modeling horizons are fit using the entire previously observed time horizon (i.e. for week 50, weeks 1 to 49 are used to estimate the spatial risk-weights). For the multi-year Pittsburgh Burglary study, GPF models are fit using the previously observed six month (GPF-HW-6 and GPF-ARIMA-6), 12 month (GPF-HW-12 and GPF-ARIMA-12), and total (GPF-HW and GPF-ARIMA) modeling horizons.

3.7 One-Year Pittsburgh Burglary Study Performance Comparison

In order to conduct a performance comparison of the GPF method against common benchmarks, we apply the three traditional approaches and the new GPF method (using both Holt exponential smoothing and ARIMA models) to the Pittsburgh data set. Holt exponential smoothing models are used for this data set as having only one year of data does not allow the estimation of the seasonality effects. For notational convenience, the notation HW describes both the Holt exponential smoothing models used for the analysis in this section and the Holt-Winters exponential smoothing models used in the next section as both methods employ the same modeling procedure.

The performance assessment horizon for the one-year Pittsburgh burglary study is for the 48 weeks from weeks 4 (to allow for model initialization) through week 51 (because week 52 is a partial week) during the year 2008. The naive forecast for each of the 52 sub-regions (six precincts and 46 patrol sectors) is the event count observed in the previous week for that sub-region. We fit the Holt (HW) and ARIMA methods as benchmarks for GPF performance by using the *stats* and $forecast$ packages in R software to fit a model for each precinct and patrol sector each week using all previous weekly event counts. This requires fitting 52 forecasting models each week with each method. We then used those models to provide weekly one-step-ahead forecasts for each of the 52 sub-regions. We also applied the GPF method to the same dataset on a weekly basis, which requires fitting one HW or ARIMA model and one kernel density estimate each week. The models are referenced as GPF-HW (when the HW method is used for the aggregated region forecast) and GPF-ARIMA.

Results are presented here using the Mean Absolute Scaled Error (MASE) as the performance assessment statistic here. Appendix A provides results in terms of Root Mean Squared Error (RMSE) at both the precinct and patrol sector level (which provides results in the scale of the data). With N denoting the number of forecasts made over the out-of-sample performance evaluation horizon, the MASE statistic over the time horizon $t = 1$ to N is calculated based on the observed counts Y_t and the forecasts F_t.

$$MASE = \frac{1}{N} \sum_{t=1}^{N} \left(\frac{|Y_{jt} - F_{jt}|}{\left|Y_{jt} - Y_{j(t-1)}\right|} \right) \tag{3.15}$$

Table 3.3 provides a performance summary using the MASE statistic for the Pittsburgh burglary example. Figure 3.4 plots the combined MASE performance from Table 3.3 against the modeling effort recorded in Table 3.2. As can be seen in Figure 3.4, the GPF method offers significant improvement in both modeling performance and modeling effort. Both GPF methods are Pareto-efficient, offering improved performance over the Holt exponential smoothing and ARIMA models for this case. In this example, the GPF method reduced the number of forecasting models from 52 weekly forecasting models (HW or ARIMA) to one forecasting model (HW or ARIMA at the city level) and one density estimate each week while improving forecasting performance in both the HW and ARIMA cases.

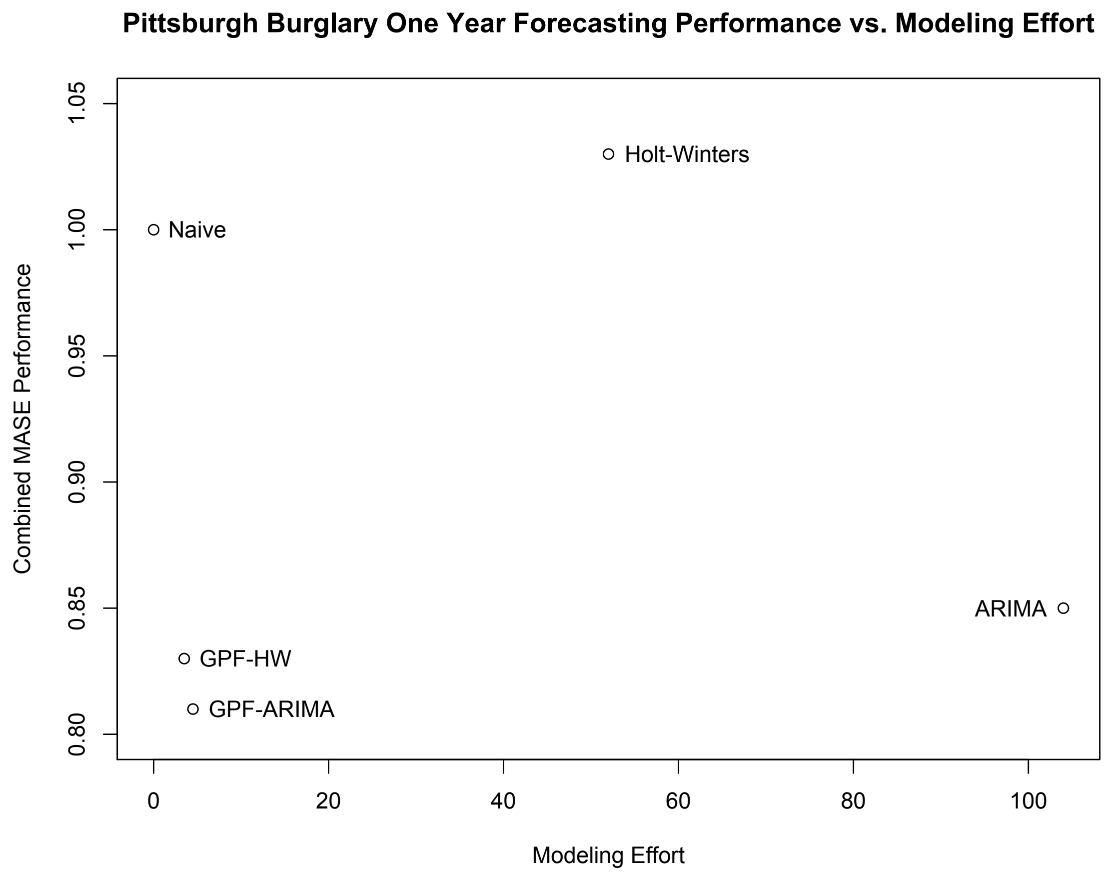


Figure 3.4: Plot of MASE forecasting performance vs. modeling effort for the Pittsburgh one year burglary analysis.

Region Average	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Precinct	1.00	1.12	0.96	0.92	0.88
Patrol Sector	1.00	1.01	0.81	0.84	0.80
Overall	1.00	1.03	0.83	0.85	0.81

Table 3.3: MASE statistic summary for the weekly one-step ahead burglary forecasts for Pittsburgh in 2008.

3.8 Multi-Year Pittsburgh Burglary Study Performance Comparison

We also extended the study comparing forecasting methods for weekly estimates of Pittsburgh burglaries to consider performance over a five-year period using the data obtained from Carnegie Mellon University. This allows us to consider forecasting models with seasonality effects. Modeling seasonal effects requires a minimum of three full seasonal cycles. Therefore, using the five-year Pittsburgh data set we implemented a rolling horizon forecast designs over two periods. For the first three years of the data, the forecasting model for each week is developed from all previously observed weeks. Thus, observations for weeks 1 to 3 are used to develop models to forecast week 4 burglary counts. Then, weeks 1 through 4 are used to forecast week 5, etc.. The first three years provide 153 periods over which to evaluate the model (three full years minus the three week initialization period). The second rolling design period considers the 96 weeks observed between 1 January 2008 and 31 October 2010. During this period, Holt-Winters exponential smoothing models and ARIMA models with estimated seasonality effects are fit using the first three years of observations to initialize the estimates for the seasonality effects. Table 3.4 provides a MASE performance summary for all of the considered forecasting models over the five-year period.

The GPF method continues to provide significant performance improvement while reducing modeling effort over the five-year study horizon. Figure 3.5 provides plots of MASE forecasting performance as recorded in Table 3.4 as a function of modeling effort. As Figure 3.5 illustrates, all the top-down forecasting methods significantly improve forecasting performance over the five year study when the appropriate time horizon is used to estimate the spatial distribution. These results correlate closely with the results seen in the one-year Pittsburgh burglary study and illustrated in

Method	**Precinct**			**Patrol Sector**			**Combined**
	Years 1-3 Non-Seasonal	Years 4 - 5 Seasonal	Years 1-5 Combined	Years 1-3 Non-Seasonal	Years 4 - 5 Seasonal	Years 1-5 Combined	
Naive	1.00	1.00	1.00	1.00	1.00	1.00	1.00
HW	0.89	0.90	0.89	0.95	0.95	0.95	0.95
HW-GPF	0.88	0.84	0.87	0.97	1.03	0.99	0.98
HW-GPF-6	0.78	0.83	0.80	0.82	0.83	0.83	0.82
HW-GPF-12	0.78	0.81	0.79	0.82	0.82	0.82	0.82
ARIMA	0.81	0.80	0.81	0.84	0.84	0.84	0.83
ARIMA-GPF	0.89	0.80	0.86	0.97	1.04	1.00	0.98
ARIMA-GPF-6	0.79	0.79	0.79	0.83	0.80	0.82	0.81
ARIMA-GPF-12	0.79	0.77	0.78	0.82	0.81	0.82	0.81

Table 3.4: MASE statistic performance summary for all forecasting models for Pittsburgh burglary counts at the precinct and sector level over the five year study horizon.

Figure 3.4. However, the GPF-HW and GPF-ARIMA full-horizon models do not perform well, apparently because the bandwidth for the kernel density estimate does not converge correctly when the number of events grows large (greater than 5000). Time series plots of the w_j estimates show sharp breaks and discontinuities in the w_j estimates as the number of criminal events grows larger than 5000, indicating that the automated bandwidth selection procedure as implemented in software is no longer converging smoothly. Thus, while the GPF method provides significant performance improvement when the plug-in kernel bandwidth estimator performs correctly, these results suggest that it is important to monitor the estimates for w_j for sharp breaks and discontinuities that may indicate a failure of the bandwidth selection procedure. When these discontinuities in the estimates for the w_j occur, it is necessary to shorten the time horizon used to estimate the kernel density surface.

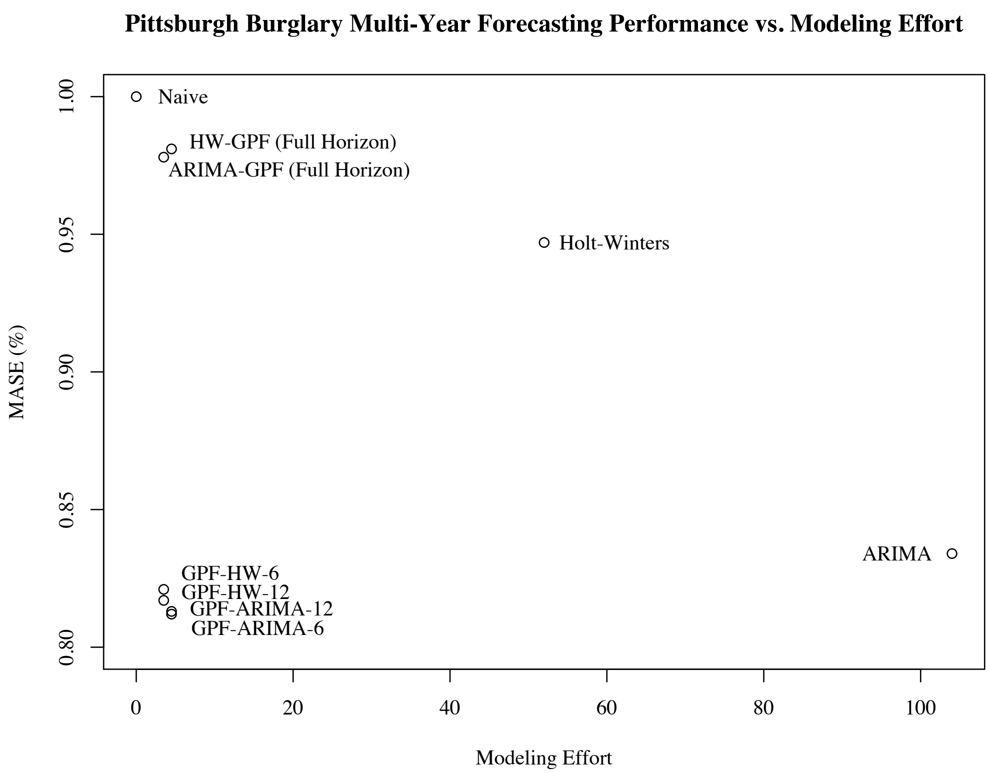


Figure 3.5: Plot of MASE forecasting performance vs. modeling effort for the Pittsburgh multi-year burglary analysis.

3.9 Model Performance Comparison Using Simulation

While a strong demonstration of the improvement that the GPF method can provide in practice, the Pittsburgh burglary study does not establish the general utility of the GPF approach because this study (conducted over several different time horizons) provides only one example of performance improvement. However, simulation models provide the opportunity to study the performance of the various forecasting approaches under many different conditions. Using a simulation model to study the properties of these forecasting methods offers three significant benefits. First, with a simulation model, one can easily vary the conditions of the simulation and observe the resulting effects on the performance of the methods. Within the simulation model, not only can one generate noisy geographic time series that include trends, seasonality, and shocks but one can vary the intensity of these effects at will. Second, in a simulation model, a *known* process generates the various time series. So, one can evaluate forecasting methods on how well they model a known process instead of conducting performance comparisons against observed counts in an observational setting in which the true spatial-temporal process is unknown. Removing the random noise from the evaluation measures is especially helpful when evaluating performance against exceptionally noisy processes such as Poisson event counts. Finally, simulations models replicate, repeatedly generating simulated outcomes from the same processes. This replication facilitates the study of the convergence properties of the forecasting methods.

Figure 3.6 provides a visualization of one of the simulation models developed for this analysis. This graphic illustrates the state of the simulation model in period 1 and after 50 periods of observation. The simulation environment contains a geographic extent (from -100 to 100 in x and y coordinates), a region of interest (from -60 to

60) and four smaller equal sized geographic sub-regions. The modeling problem is to accurately forecast the observed counts in each of the four spatial sub-regions during each time step. The simulation model provides the opportunity to vary many modeling parameters, including the number of spatial/temporal processes and the location, spatial distribution, and rate for each spatial process. The model in Figure 3.6 contains five spatial processes, each of which has a unique spatial distribution. Figure 3.7 shows the known process rate and resulting event counts due to the noisy process in Region 3 over the first 50 time steps of the simulation.

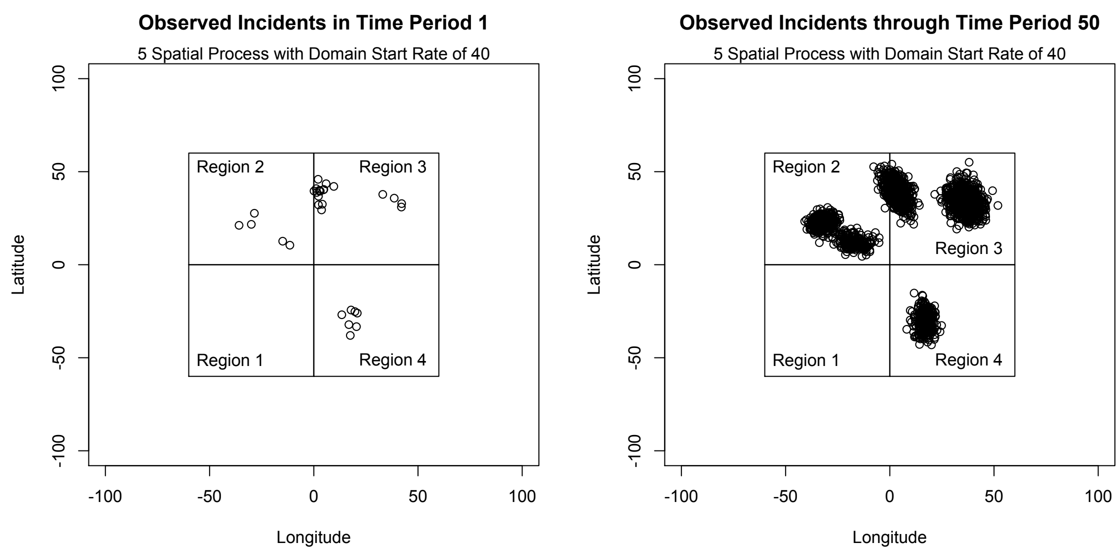


Figure 3.6: Graphical illustration of a non-homogenous point process model during time step 1 (left panel) and time step 50 (right panel)

Once the number of processes are defined for a given experiment, the simulation models randomizes the location and dispersion of those spatial processes by uniformly selecting parameters for the Gaussian spatial processes from the following intervals: $\mu_{x,y} = U(-60, 60)$, $\sigma_{x,y} = U(5, 30)$, $\rho = U(-.5, .5)$. In the notation above, μ denotes the location of the center of the spatial process, σ the spatial variance, ρ the covariance parameter, and $U(-, -)$ selection from the uniform distribution. The notation for the Gaussian spatial process is summarized as $N(\mu_p, \Sigma_p)$, where Σ_p denotes the

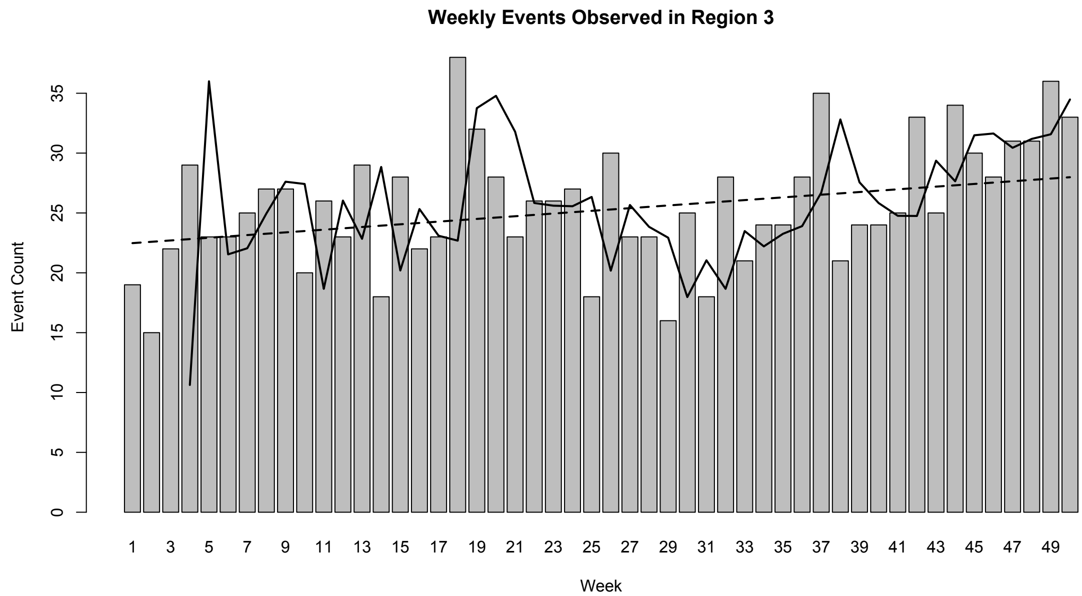


Figure 3.7: Observed event counts (barplot), known process mean (dashed line), and Holt-Winters forecast (solid line) for Region 3 for 1 replicate of the simulation illustrated in Figure 3.6.

covariance matrix for spatial process p containing ρ_p, σ_x, and σ_y. Note that given these distributions, some portion of the spatial process may overlap with the region boundary, so that some of the simulated incidents fall outside of the region of interest for the forecasting problem. The incidents that fall outside of the region of interest are treated in the same way that law-enforcement treats such criminal events: they are used in kernel density estimates (hot-spot maps) when known but since they do not occur in the regions or sub-regions of interest, they do not form part of the event counts for the forecasting models. The spatial distributions of the event processes remain fixed over the conduct of an experiment (although the simulation model generates many randomized replicates of the geographic time series within each experiment), so the spatial processes do not migrate or shift.

The rate of the spatial processes can be controlled dynamically, introducing trends, seasonality, or shocks into the process by adjusting the intensity (rate) parameter λ_p over time. Note Figure 3.7 shows a positive trend, although the actual observed counts fluctuate wildly, reflecting the noise of the Poisson process. During each model step t, the simulation model randomly draws an event count for each process from the Poisson distribution defined for that process by the rate parameter $\lambda_p(t)$ and randomly places each of those events within the model geography in accordance with the spatial distribution defined by the individual processes' spatial distribution model. These spatial distribution models are Gaussian: $N(\mu_p, \Sigma_p)$. As observed in Figures 3.6 and 3.7, these processes are noisy.

The use of square sub-regions and known process parameters facilitates direct calculation of the expected counts for each region D_j for each time step. For example, with P spatial processes taking place in the region of interest during time t, and each of these spatial processes Gaussian distributed in space with a Poisson arrival rate λ_p, the expected count for a given geographic sub-region during each temporal time period t is:

$$E\left[Y_{jt}\right] = \sum_{P} \lambda_p(t) \int_{D_j} N(\mu_p, \Sigma_p) \tag{3.16}$$

Note that due to trends and seasonality effects, $E\left[Y_{jt}\right]$ can change dynamically for each temporal block t.

Now each forecasting method can be evaluated on how well the method performs at modeling the known process taking place in each sub-region over many replicates for a given simulation experiment. Figure 3.7 provides an illustration of the difference between the process error (the difference between the known process rate and the forecast) and the observed error (the difference between the observed count and the forecast). The MASE statistic in Equation 3.15 is easily modified to reflect the process error and multiple replicates (indexed by r), yielding the following statistic calculated for each of the j sub-regions:

$$MASE_{PROCESS} = \frac{1}{RN} \sum_{r=1}^{R} \sum_{t=1}^{N} \left(\frac{|E[Y_{jt}] - F_{jrt}|}{|E[Y_{jt}] - Y_{jr(t-1)}|} \right) \tag{3.17}$$

This chapter provides results from testing forecasting performance over a wide variety of conditions, including the introduction of trends, seasonality, and shocks. Appendix B provides the full description of the design of experiments for those cases where the GPF assumptions discussed in Section 3.5.4 apply, specifically any situation in which trend or seasonality effects are global processes that affect all sub-regions equally. The first set of simulation experiments tested performance for a stationary homogenous point process, in which the observed counts in a sub-region are a direct function of the area of the sub-region. This type of spatial point process is the starting point for most spatial analysis because the stationary homogenous point process describes so many known spatial processes such as those "responsible for the location of things such as human settlements, store-types, plants and animals, and groups of plants and animals (Getis and Roots, 1978)." Stationary non-homogenous point processes have no trend or seasonality, but have non-homogenous spatial distributions such as those depicted in Figure 3.6, which exhibits spatial clustering (hot-spots) of events. Additional modeling scenarios simulated include the introduction of positive trends, negative trends, and seasonality effects.

Appendix C provides a full description of the design of experiments for those cases where the GPF assumptions in Section 3.5.4 do not apply. The first scenario includes the introduction of a shock process into one of the sub-regions in the region of interest (i.e., one precinct, patrol sector etc.). From that point on, that one region has a significantly different process than before, while the process in the remaining sub-regions remains the same. The second scenario includes the situation in which one sub-region (precinct, car patrol beat etc.) experiences a positive trend while the rest of the region of interest experiences a negative trend. The last considered scenario includes the situation in which every unique process in the region of interest has a unique trend (i.e. all trends are local not global).

These experimental designs replicate all scenarios under a variety of conditions such as varying the strength of the trends and the number of unique spatial processes in the region of interest. In all of the scenarios where the GPF assumptions do not apply, at least one of the sub-region level process changes uniquely over time, representing greater and greater departure from the GPF modeling assumptions. Thus, for each of the scenarios that consider violations of the GPF assumptions, we also fit the GPF method using rolling horizons of 20 time periods to develop the kernel density estimate (i.e., the kernel density estimate is fit using observations from periods $t-20$ to $t-1$ and then used to forecast period t). These rolling horizon models are designated GPF-HW-R and GPF-ARIMA-R.

3.10 Simulation Study Results

When any trends or seasonality effects are global, affecting all geographic sub-regions, the GPF method significantly improves predictive performance when applied using both HW and ARIMA methods. Tables 3.5 and 3.6 record the MASE statistic calculated using the observed and process error. Appendix E provides results for when over-dispersed Poisson distributions are used in the simulation (with similar results). As can be seen in the tables, the traditional forecasting methods provide performance improvement in keeping with increasing model complexity. The HW method reduces the observed scaled error by between 7 and 14% as compared to the Naive method. However, note that the HW method actually describes the known process much better than these results would suggest. In Table 3.6 the HW method provides up to a 33% improvement over the Naive method in process error. The performance improvement the ARIMA method provides over the HW method in these tables does not reflect that the ARIMA class is a better method than the HW method in general. As Hyndman and Khandakar (2008) note, they are overlapping model classes. Rather, the observed performance improvement is due to the more complex model fitting procedure employed to fit the ARIMA models in this study.

The most important finding is the improvement the GPF method provides. While the MASE performance improvements seen in Tables 3.5 are on the scale observed in the Pittsburgh example (see Table 3.3), the GPF method provides forecasts that much more accurately describe the *process*, as seen in Table 3.6. When the process noise is removed from the performance statistic, the GPF-HW method provides up to a 38% reduction in process error when compared to the HW method. The improvement the GPF-ARIMA method offers over the ARIMA method is more muted (5 - 18%), but still provides significant performance improvement for less effort.

Figures 3.8 - 3.10 provide graphical time series plots that visually illustrate the performance improvement provided by the GPF method for several of the consid-

Scenario	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Stationary Homogenous	1.00	0.91	0.77	0.74	0.72
Stationary Non-Homogenous	1.00	0.93	0.79	0.74	0.73
Trend	1.00	0.93	0.79	0.78	0.74
Seasonality	1.00	0.86	0.75	0.78	0.73
Season & Positive Trend	1.00	0.84	0.74	0.76	0.73
Season & Negative Trend	1.00	0.90	0.77	0.81	0.76

Table 3.5: MASE performance summary for the observed error for the five considered forecasting methods over the six scenarios in which GPF modeling assumptions apply.

Scenario	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Stationary Homogenous	1.00	0.80	0.42	0.25	0.18
Stationary Non-Homogenous	1.00	0.79	0.44	0.25	0.20
Trend	1.00	0.80	0.43	0.42	0.29
Seasonality	1.00	0.67	0.35	0.44	0.26
Season & Positive Trend	1.00	0.63	0.32	0.40	0.28
Season & Negative Trend	1.00	0.77	0.40	0.52	0.35

Table 3.6: MASE performance summary for the process error for the five considered forecasting methods over the six scenarios in which GPF modeling assumptions apply.

ered cases including homogenous stationary processes (Figure 3.8), non-homgenous processes with trends or seasonality effects (Figure 3.9), and the interaction between trend and seasonal effects (Figure 3.10). Figure 3.8 provides time series plots for the forecasts generated using ARIMA (at the region and sub-region level) and GPF-ARIMA for a stationary homogenous point process. As can be seen in the figure, the GPF-ARIMA forecast at the sub-region level is a scaled version of the forecast at the aggregate level (in this case $w_j = \frac{1}{4}$). The variance of the GPF-ARIMA forecast at the sub-region level is less than that for the ARIMA forecast. Note that both forecasts are converging (slowly) to the true process rate, but that the GPF-ARIMA method is converging faster as expected given the results of Equation 3.8. Figures 3.9 and 3.10 provide similar observations under the conditions of trends, seasonality, and the interaction of these effects.

The right panel of Figure 3.8 plots ratios of the ARIMA and GPF-ARIMA Mean Squared Error (MSE) to the aggregate level ARIMA model MSE for the scenario modeled in the left panel. This is the relationship explored in Equations 3.7 and 3.8 and the experimental results depicted in Figure 3.8 are for the example given in Equation 3.8. With a smoothly converging estimator (such as the sample mean) and a perfect estimate for w_k, we would expect these ratios to conform closely to the dashed lines on the graph. As can be seen, the MSE ratio for the ARIMA method noisily oscillates around the expected value, while the GPF-ARIMA MSE ratio does not fully achieve the performance gains expected given Equation 3.8 due to the modeling error of the kernel density estimate. In spite of this, the improvement provided by the GPF-ARIMA method is significant: in this case the GPF-ARIMA sub-region MSE averages $\frac{1}{10}$ of the aggregate ARIMA MSE while the ARIMA sub-region MSE averages $\frac{1}{4}$ of the aggregate ARIMA MSE. This pattern is repeated for all of the experiments recorded in Table B.1: the GPF method MSE ratio is a value slightly higher than (sometimes within one percent of) w_k^2 while the ARIMA or HW MSE

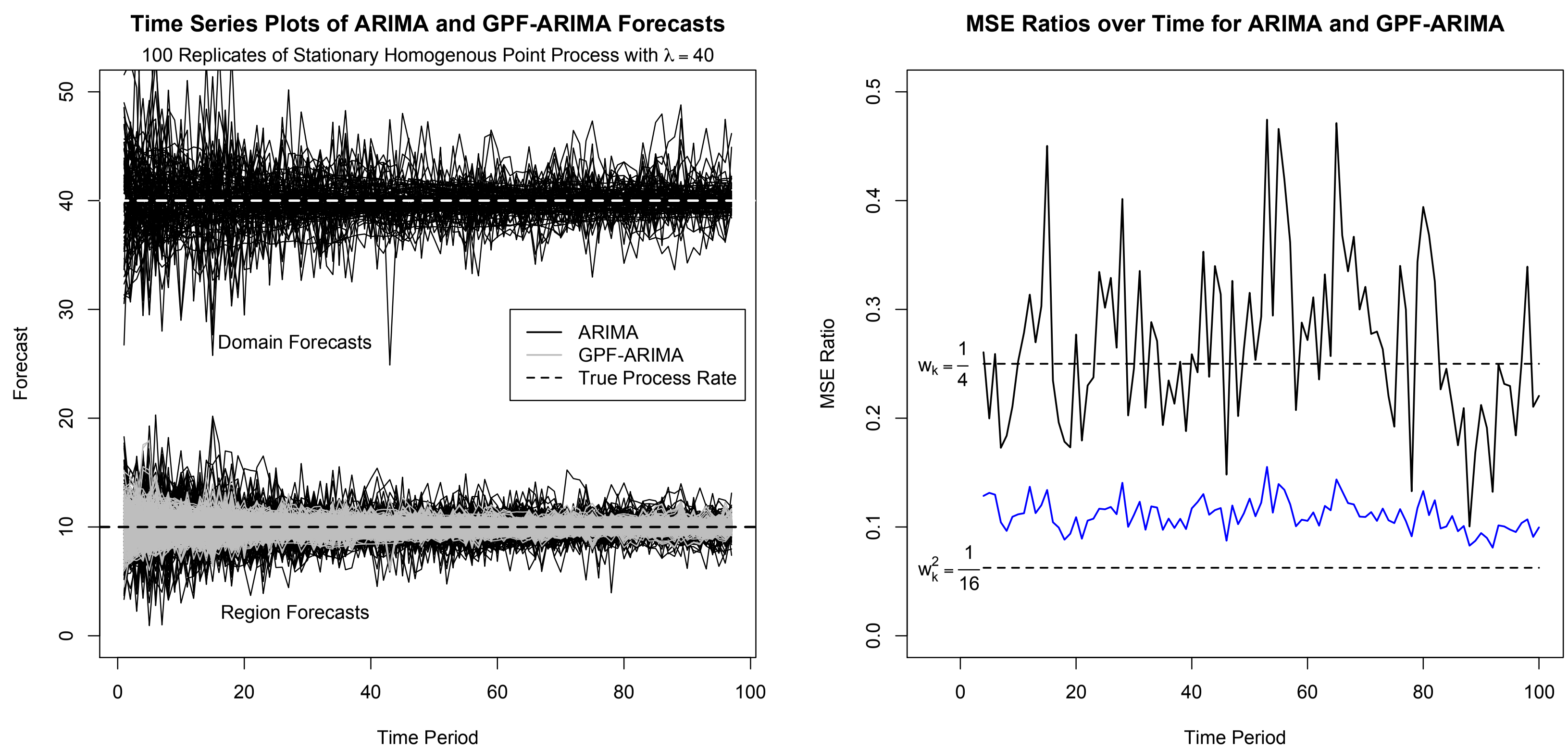


Figure 3.8: Left Panel: Time series plot comparisons of aggregate (top of the graphic) and sub-region (bottom of the graphic) forecasts using ARIMA (in black) and GPF-ARIMA (in gray) models. Right Panel: Time series plots of the resulting mean squared error (MSE) ratio for the Sub-Region MSE divided by the Domain MSE for the ARIMA method (top of the graph) and the GPF-ARIMA method (bottom of the graph) during each time step using the 50 experiment replicates. The dashed lines indicate the expected MSE ratios given Equation 3.8.

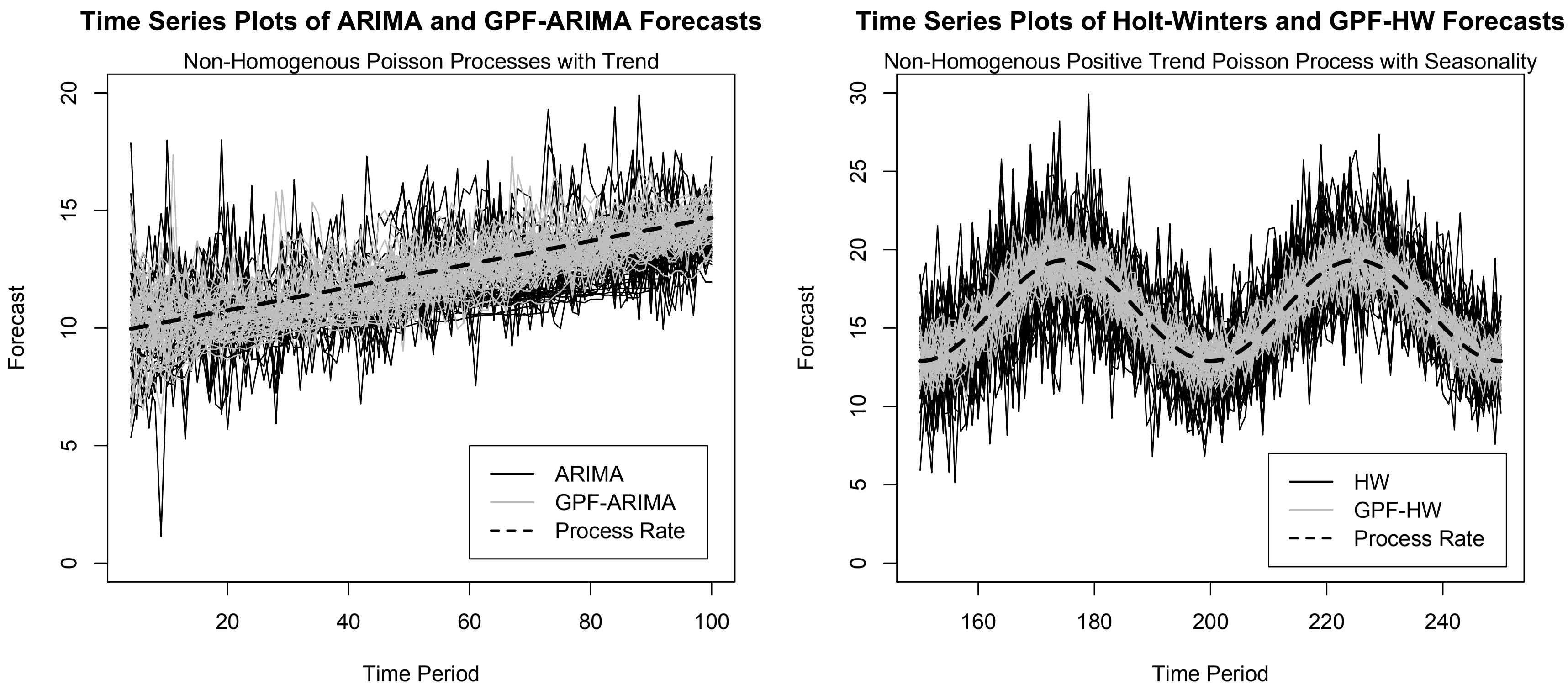


Figure 3.9: Example time series plots for forecast performance comparison of ARIMA and GPF-ARIMA models for an experimental precinct with a positive trend (left panel) and Holt-Winters and GPF-HW models for a precinct with seasonality (right panel).

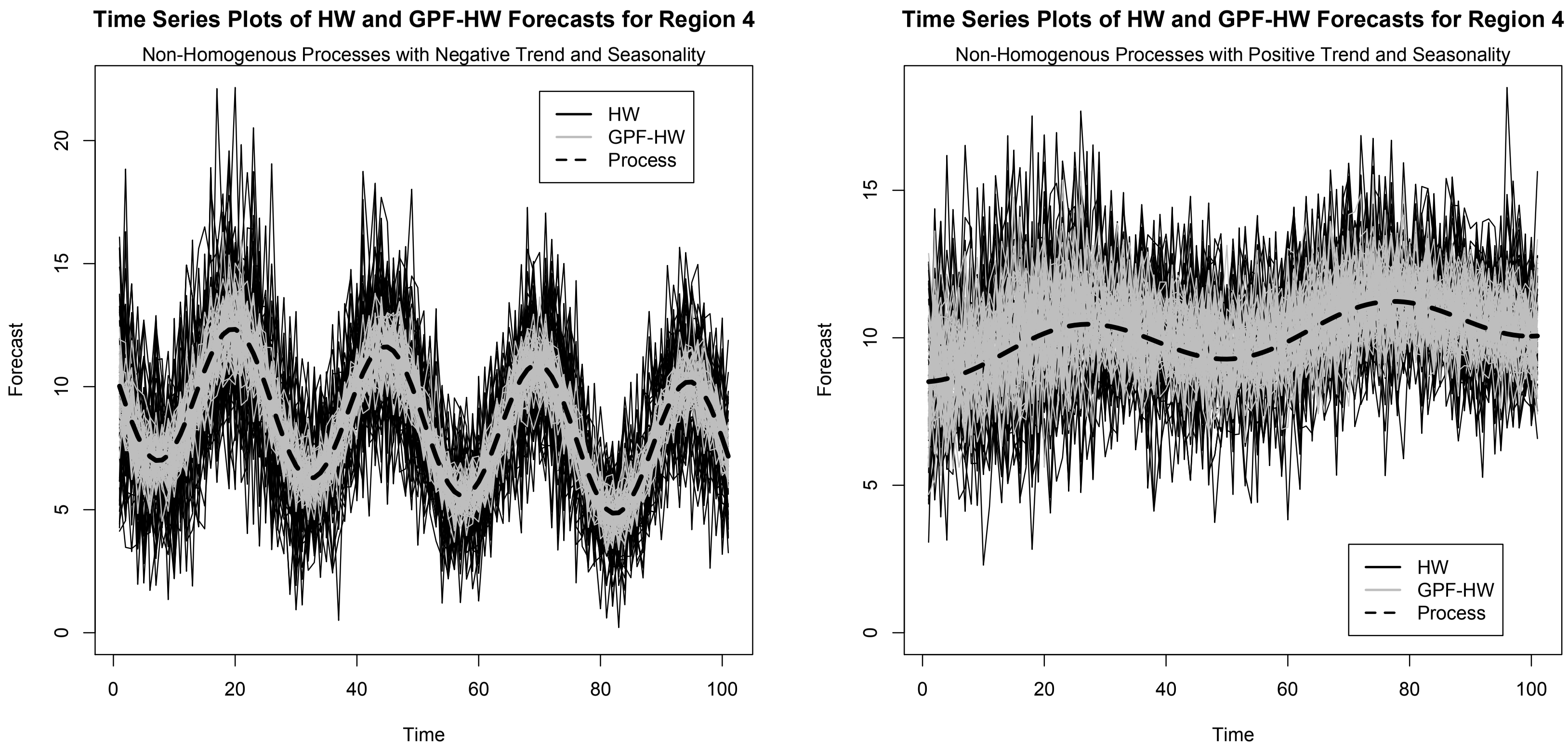


Figure 3.10: Time series plots of HW and GPF-HW forecasts in a scenario with negative trend and seasonality effects (left panel) and a positive trend and seasonality effects (right panel). The processes in the left panel are set at the high-frequency and high-amplitude setting in the simulation model as discussed in Appendix B. In the right panel, the processes are set at the low-frequency and low-amplitude settings.

ratio noisily oscillates around and averages w_k.

Tables 3.7 and 3.8 provide performance summaries for scenarios that violate the GPF modeling assumptions. Appendix E contains the same results for the studied simulation scenarios when over-dispersed Poisson distributions are used. In the scenarios recorded in Tables 3.7 - 3.8 and E.3 - E.4, the performance of the traditional methods continues to correlate with modeling effort as before, with more complex models providing better performance. As before, the GPF method improves forecasting when applied to the HW models. However, when shocks or competing trends exist, the ARIMA method provides better performance than the GPF-ARIMA method. The use of rolling time horizons to fit the GPF models (the models GPF-HW-R and GPF-ARIMA-R) does improve performance in the case of shocks and competing trends, but negatively affects performance under random trends. The use of rolling time horizons to fit GPF models therefore represents a trade-off. Using a shorter time horizon to fit a GPF model provides some insurance to the effects of gross violations of the modeling assumptions such as strong shocks or competing trends. However, the shorter the time horizon used, the greater the variance of the forecast. So, in the case of random trends, in which the effects of many competing trends average out for the most part across the sub-regions, the use of a rolling horizon negatively affects performance.

Based on these results of the Pittsburgh burglary study and the simulation study, the recommended procedure for selecting an appropriate time horizon for estimating the kernel density threat surface is to monitor (with time series plots) the weekly estimates of w_j. The time horizon should be both long and short enough that the estimates for w_j are stable (changing little from time period to time period). If the time horizon used is too long, the automated kernel density bandwidth selection procedure may provide unstable estimates and this will manifest itself as sharp discontinuities in the estimates for w_j. If the time horizon is too short, time series plots of the w_j

Scenario	Naive	HW	GPF-HW	GPF-HW-R	ARIMA	GPF-ARIMA	GPF-ARIMA-R
Competing Trends	1.00	0.91	0.79	0.73	0.69	0.73	0.65
Random Trends	1.00	0.92	0.79	0.79	0.76	0.74	0.74
Shocks	1.00	0.93	0.84	0.80	0.76	0.79	0.76

Table 3.7: MASE performance summary for the observed error for the seven considered forecasting methods for the considered scenarios where the GPF assumptions do not hold.

Scenario	Naive	HW	GPF-HW	GPF-HW-R	ARIMA	GPF-ARIMA	GPF-ARIMA-R
Competing Trends	1.00	0.82	0.58	0.49	0.41	0.46	0.34
Random Trends	1.00	0.78	0.45	0.46	0.35	0.28	0.30
Shocks	1.00	0.81	0.60	0.49	0.31	0.42	0.32

Table 3.8: MASE performance summary for the process error for the seven considered forecasting methods for the considered scenarios where the GPF assumptions do not hold.

will show estimates that are still converging (i.e. all estimates w_j will be changing during each time period, rather than arriving at a stable estimate).

Figure 3.11 illustrates how the time horizon used for estimating the spatial distribution affects the sensitivity of the GPF method to significant changes in the spatial distribution of events . This figure depicts time series plots for the 50 replicates of ARIMA, GPF-ARIMA, and GPF-ARIMA-R methods applied to the situation in which a new (shock) process is added to Region 3. This shock process increases the known process mean from 10 events to 22 events in each time period (although the actual observed counts fluctuate wildly). This situation represents a serious violation of the model assumptions because a significant change in the spatial distribution takes place in time period 50. As can be seen in the figure, the ARIMA method (and the HW and Naive methods as well) can quickly adjust to this change by adjusting the model parameters. However, the GPF method estimate for the w_j parameter in the left panel is significantly affected by the previous 50 observations. The rolling horizon GPF method (GPF-ARIMA-R in the right panel) is less affected by the previous history and adjusts the weighting parameter faster over time, resulting in faster recovery to accurate prediction following the shock. However, note that the variance of the forecasts for the rolling horizon method are larger.

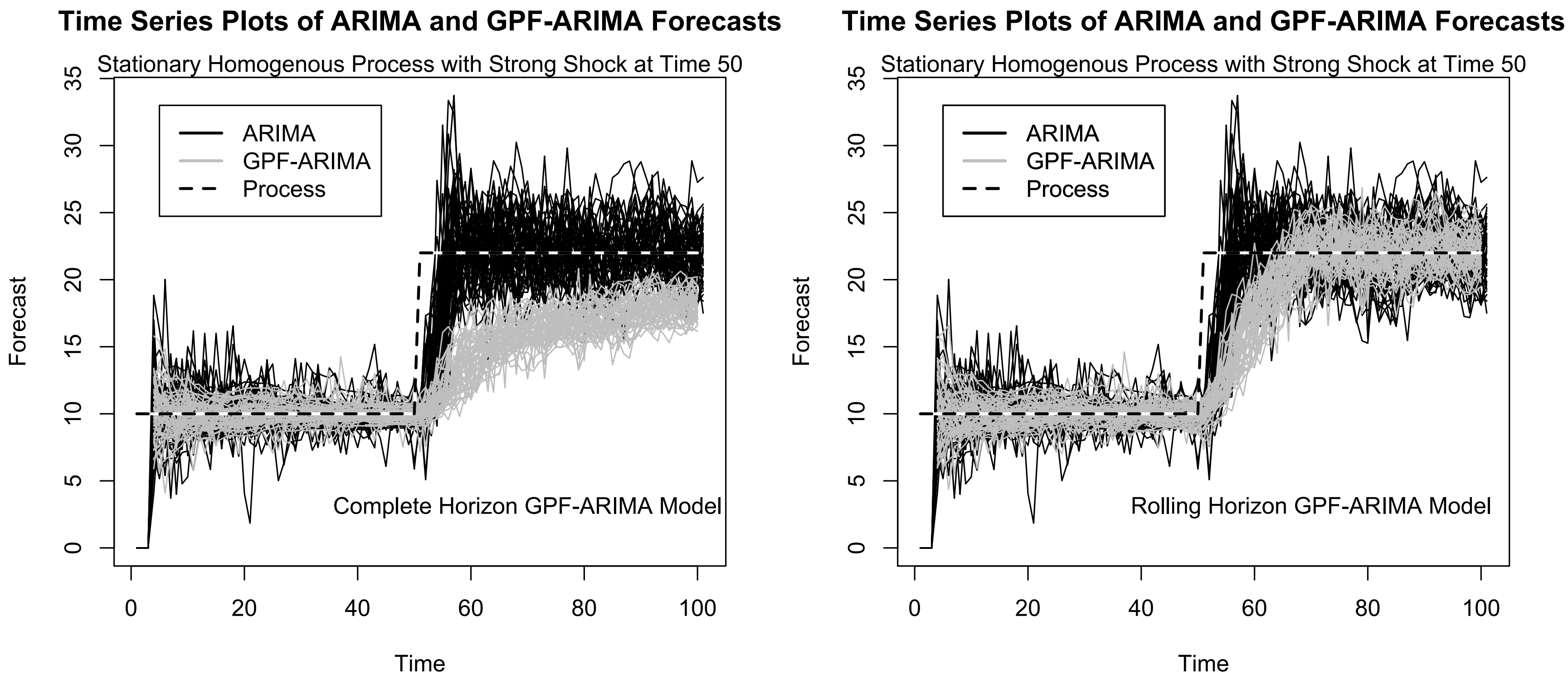


Figure 3.11: Time series plot comparisons for 50 replicates of ARIMA and GPF-ARIMA forecasts for a sub-region in which a shock process (innovation) is inserted during time period 50. The left panel depicts performance when the GPF-ARIMA method uses the complete time horizon of incidents and the right panel depicts performance when a rolling horizon of 20 periods is used.

Based upon these results, the GPF method does not seem well-suited to modeling situations in which strong shocks occur frequently in different sub-regions of the region of interest. In the case illustrated in Figure 3.11, only one shock occurs. Both the GPF-ARIMA and GPF-ARIMA-R methods recover much more slowly than the ARIMA, HW, and Naive methods in this case. In any situation in which large shock effects are frequent, the sensitivity of the HW and ARIMA methods will provide better performance than the GPF approach, although they do require significantly more modeling effort (as discussed in Section 3.6).

3.11 Conclusions

These results demonstrate that a relatively simple method that dramatically reduces the complexity and modeling workload for generating recurring forecasts of noisy geographic time series can also significantly improve forecasting performance so long as several important assumptions are satisfied. Three methods are commonly used to forecast noisy geographic time series: naive methods, Holt-Winters smoothing, and ARIMA models. While naive methods provide a simple, universally applicable approach, as demonstrated in this study, they also provide limited performance. The more complex univariate time series methods generally improved upon the naive method but required significantly more effort. In the motivating example of forecasting burglary counts in the City of Pittsburgh, using the two traditional univariate time series methods requires fitting a total of 52 separate models on a weekly basis. The GPF method requires fitting only one univariate time series model (at the city level) and one kernel density estimate for each period's forecast, a dramatic reduction in model complexity and modeling effort. This simple approach also significantly improved forecasting performance in the five-year Pittsburgh burglary study, providing a 19% reduction in forecast error (as measured by MASE) over the naive method, 14%

improvement over the Holt-Winters method, and 2% improvement over the ARIMA method (see Table 3.4).

The simulation study further demonstrates the robustness of the GPF method. When any existing trends or seasonality effects are global, the GPF method always improves upon the other methods because it provides a better estimate of the underlying process (see Tables 3.5 and 3.6). The use of the simulation model provides a better understanding of model performance by allowing performance comparison with the process noise removed from the performance assessment statistic. While the *observed* error performance improvements in these simulation study scenarios are of approximately the same size as the error observed in the burglary study, the GPF method actually reduces the *process* error much more significantly. For example, the GPF-ARIMA method improves upon the naive method MASE process error by between 71% and 82% and over the ARIMA method by 5% to 18% (see Table 3.6).

The simulation model also helps to study how robust the GPF method is to violations of the modeling assumptions (see Tables 3.7 and 3.8). As demonstrated, estimating the sub-region weighting factor w_j using rolling time horizons provides some protection to violations of the GPF model assumptions in the case of competing trends (cases where one sub-region has a positive trend while other sub-regions have a negative trend). However, the use of the rolling time horizons represents a trade-off, because shorter time horizons increase the variance of the GPF forecasts. Under the three conditions studied here (competing trends, random trends, and shocks), the GPF method always provides improved performance over Holt-Winters smoothing and the naive method. However, the ARIMA method did provide better performance than the GPF-ARIMA method in the case of competing trends and better performance than both GPF-ARIMA and GPF-ARIMA-R in the case of shocks.

In this simulation study, the GPF method provides reasonable overall performance in the case of shocks only because the region of interest experiences only one significant shock per experiment. In these cases, the improved forecasting performance over time periods *prior* to the shock balance the very poor performance the GPF method provides *after* the shock. Given these results, the GPF method does not seem well-suited to modeling situations in which strong shocks occur frequently in different sub-regions of the region of interest because it is so slow to adjust to any dramatic changes in one sub-region that do not affect the other sub-regions. In these cases, the additional work required for fitting many ARIMA models (one for each sub-region) seems warranted. In all other studied scenarios, the GPF method provides a robust, general purpose approach for improving forecasts for noisy geographic time series while greatly simplifying the modeling process.

These results provide three practical applications to security force planning. First, as demonstrated, the GPF method improves forecasting performance while also simplifying the modeling process. Second, this modeling approach directly links the analytic products used for operational level decision-making (region and sub-region forecasts) with the analytic products used for tactical level targeting and planning (threat surfaces or hot-spot maps), providing a common frame-work for tactical and operational level planners. Finally, this modeling approach suggests that threat surface maps can also improve the way that operational planners spatially assign areas of responsibility to subordinate elements by linking decisions about spatial areas of responsibility to forecasts of future activity.

Chapter 4

Geographic Probability Forecasting for Patrol District Design

Most security forces geographically divide their areas of responsibility into geographic sub-regions and assign subordinate elements responsibility for patrolling, securing, and responding to incidents within those regions. In policing, this is known the districting problem and nearly every police department creates geographic patrol districts (also called patrol sectors or car beats) as a standard management method to enhance the capabilities of the uniformed patrol force (Hale, 1980). Better districting plans lead to lower response times, officer's familiarization with their assigned area, more efficient use of personnel, more equal division of workload, a visible police presence, enhanced officer safety, officer accountability, and balanced security force response to calls (Hale, 1980). These same benefits generally apply to geographic mission assignment problems in many other security applications, including the assignment of military Areas of Operations (AOs) in the conduct of counter-insurgency campaigns and area security operations.

There are several different approaches currently used to solve the districting problem. A common approach requires military and police planners to draw district boundaries based on their knowledge, experience, and the constraints imposed upon them by the available patrolling resources (Mitchell, 1972; Taylor and Huxley, 1989). Researchers have proposed a variety of mathematical approaches to solving this complex problem that rely on optimization, set covering, N-P hard graph partitioning methods, genetic algorithms, or other advanced statistical approaches (see Section 2.6). Other researchers have demonstrated the use of agent-based and discrete event simulation models to evaluate proposed districting plans. However, of the existing approaches for district plan design, only the approach that relies exclusively on the judgement and intuition of the planner falls within the capabilities available to most military and police units, which explains its widespread use.

This chapter demonstrates a new method for developing and analyzing military and police patrol district designs that requires only tools available to most military and police planners. This new method leverages one of the key insights identified in the analysis of Geographic Probability Forecasting (GPF) in Chapter 3: domain-level forecasts that are weighted using geographic probability maps provide good estimates for future event counts for noisy geographic time series. The GPF district design method develops a planning surface for geographic mission assignment that estimates the cost (in terms of man-hours) for servicing Calls for Service (CFS) for every unique geographic location in the considered domain. This geographic cost estimate is calculated using manpower estimates for different CFS incidents and the GPF forecasting method. The resulting mapped planning surface will assist military and police planners in balancing the need for preventative patrols with the need to respond to critical incidents when they happen. The GPF district design method is demonstrated for re-designing a patrol sector plan for use by a county police agency in central Virginia.

4.1 Problem Definition

The problem definition for this analysis is to develop a generally applicable heuristic district design method that can be used to develop military Areas of Operation (AOs) or districting plans for police agencies employing a geo-policing strategy. As a motivating example, this chapter demonstrates the development of a districting plan for a county police department that provides a better officer workload balance between the patrol sectors used within the county.

4.2 Albemarle County Police Data

In order to demonstrate the proposed methodology, we develop a new patrol district plan for a police agency transitioning to a geo-policing approach, in which small groups of officers are assigned to exclusively patrol specific areas. This approach mimics one often used in military counter-insurgency campaigns in which small military units are assigned a specific and enduring Area of Operations (AO). The International Association of Police Chiefs has documented that this community policing method has produced significant crime reduction in several cities and formally endorses the approach (Welch and Bussiere, 2005).

The data for this analysis was provided by the Albemarle County Police Department (ACPD). Albemarle County is located in Virginia approximately 110 miles southwest of Washington, D.C. The county encompasses approximately 726 square miles and has approximately 100,000 county residents in the county. In the center of the county is the city of Charlottesville. The City of Charlottesville has a diameter of about 7 miles and a year-round population of about 40,000, which swells to about 66,00 during the academic year due to the presence of a major university. The City of Charlottesville is policed separately by the Charlottesville Police Department (CPD). Most of the county is rural and the county's population (and crime) is concentrated in the areas of the county immediately surrounding the City of Charlottesville. The county includes one incorporated town (Scottsville, located in the southeast corner of the county) and several unincorporated communities including: Barboursville (in the northeast corner of the county), Crozet (on the west side of the county), Earlysville (in the northwest corner of the county), and Ivy (east of Charlottesville).

The current patrol sector plan used by the ACPD is more than ten years old. In the intervening years since the plan's development, the county population has grown, changing the geographic distribution of crime. Additionally, the department's resources have expanded in such a way that the current patrolling plan does not align with the department's resources (for example, there are a minimum of 10 patrol cars assigned to each shift but there are eight patrol sectors in the current patrol sector plan). The ACPD is transitioning to a geo-policing model in which the officers geographically assigned to patrol districts are primarily responsible for servicing the calls within their districts (although officer support is provided across patrol sectors as needed) and exclusively responsible for patrolling within their districts (Richardson, 2012). As will be shown in later sections, the current districting plan does not effectively balance the workload between the different districts, with the result that there are signifiant differences in the CFS workloads for the different patrol districts and many rural regions of the county are very rarely patrolled.

The ACPD provided several forms of data used for this analysis. The first dataset provided was the PISTOL database, which contains records of all Call for Service (CFS) events for the years 2007 - 2012. This database contains the geographic location and call type for every CFS the department received over the five-year period. This data set was used to develop forecasts and geographic probability models for the various CFS types. The second data set provided was the department's service call database containing records of police response to calls for service (via radio reports) for the period from January 2009 through November 2012. This data set contained millions of records for officer status as they respond to incidents (i.e. the CFS time, the arrival time on scene, the departure time from scene, etc.). These records were used to establish the cost (in man-hours) for the various CFS. Finally, the ACPD provided geographic data sets for roads, county boundaries, and the currently used patrol sector plan used by the police department.

4.3 Methodology

The methodological approach demonstrated here leverages one of the key insights developed by Huddleston et al. (2013b): domain-level forecasts that are weighted using geographic probability maps provide good estimates for future event counts for noisy geographic time series. The methodological extension demonstrated here produces a planning surface for geographic mission assignment that estimates the cost (in terms of man-hours) for every geographic location in the domain (i.e., every unique 50 m x 50 m grid square in the county). The planning surface is estimated by generating event forecasts at a very fine resolution throughout the domain. These forecasts are then multiplied by an estimate of the cost (in terms of average man-hours) needed to support a call for service of every type to develop a planning surface. The planning surface provides a visualization tool that planners can use to divide the domain into geographic patrol regions and a method for evaluating how well proposed districting plans balance the patrolling and CFS workload between geographic patrol sectors. There are five steps to developing a districting plan using the GPF district design method:

1. Develop Geographic Probability Maps for Event Types
2. Develop Forecasts for Event Types
3. Estimate Event Costs

4. Generate a Geographic Planning Surface

5. Develop and Analyze Districting Plans Using the Geographic Planning Surface

Each of the five steps listed above is developed in depth below using the data provided by the ACPD as a motivating example. The methodological approach is simplified in this chapter to consider only two very resource intensive event types: breaking and entering (burglary) events and traffic accidents. The approach demonstrated here is easily extended to consider all of the different event types police and military units respond to on a regular basis.

4.3.1 Develop Geographic Probability Maps for Event Types

The first step in developing a geographic mission planning surface is to develop a geographic probability model for every event type. The geographic probability models used in this demonstration are kernel density estimates developed using the ACPD CFS records for breaking and entering (B&E) and traffic incidents. Let b_i index two dimensional blocks within a spatial study region $D \subset \Re^2$. These two-dimensional spatial blocks denote unique locations created by laying a grid at a fine resolution across the study domain: $\{b_1, b_2, ..., b_I\}$. Let s_{y_m} denote the location in $\Re^2$ of event y_m and Y_m the total number of events of type m occurring within D. The event intensity, $f_m(b_i)$, for each event type m and location i is calculated using the kernel density function K_h.

$$\hat{f}_m(b_i) = \frac{1}{hY_m} \sum_{y_m=1}^{Y_m} K\left(\frac{\|b_i - s_{y_m}\|}{h}\right) \tag{4.1}$$

The kernel density estimate above then needs to be converted into a probability estimate. This is easily accomplished by normalizing the density estimate to develop a geographic probability weight by event type for every unique location.

$$\hat{w}_{mi} = \frac{f_m(b_i)}{\sum_{b_i \in D} \hat{f}_m(b_i)} \tag{4.2}$$

Equation 4.2 converts the probability density function developed using the kernel density method in Equation 4.1 into a probability estimate at each geographic location. Thus, $\hat{w}_{mi}$ represents the probability of event type m occurring at location i.

As in Chapter 3, the kernel density estimate uses the plug-in estimate for bandwidth (h) that minimizes the Mean Squared Error (MSE) of the hot-spot map using the procedure outlined by Berman and Diggle (1989). Figure 4.1 provides the geographic probability (hot-spot) maps for Albemarle County B&E and traffic accident events. Note that although the two event types share many high-probability areas, the probability maps do contain some differences. Both event types are most probable in the areas immediately surrounding the city of Charlottesville (in the center of the map). However, traffic accidents are concentrated on the main transportation routes in the county while the burglaries are distributed more broadly throughout the major neighborhoods in the county.

4.3.2 Develop Forecasts for Event Types

The next step is to develop forecasts for the average count (at the domain-level) for each of the event types in future periods. We used time series decomposition to develop a current estimate for average number of monthly events in Albemarle county. Time series decomposition is a statistical method for deconstructing a time series into its trend, seasonal, cyclical, and noise components (Hyndman and Athanasopoulos, 2013). Classical time series decomposition, developed in the 1920s, decomposes a time series in three steps (Hyndman and Athanasopoulos, 2013). First, it fits a moving average model to estimate the trend component, which is then removed from

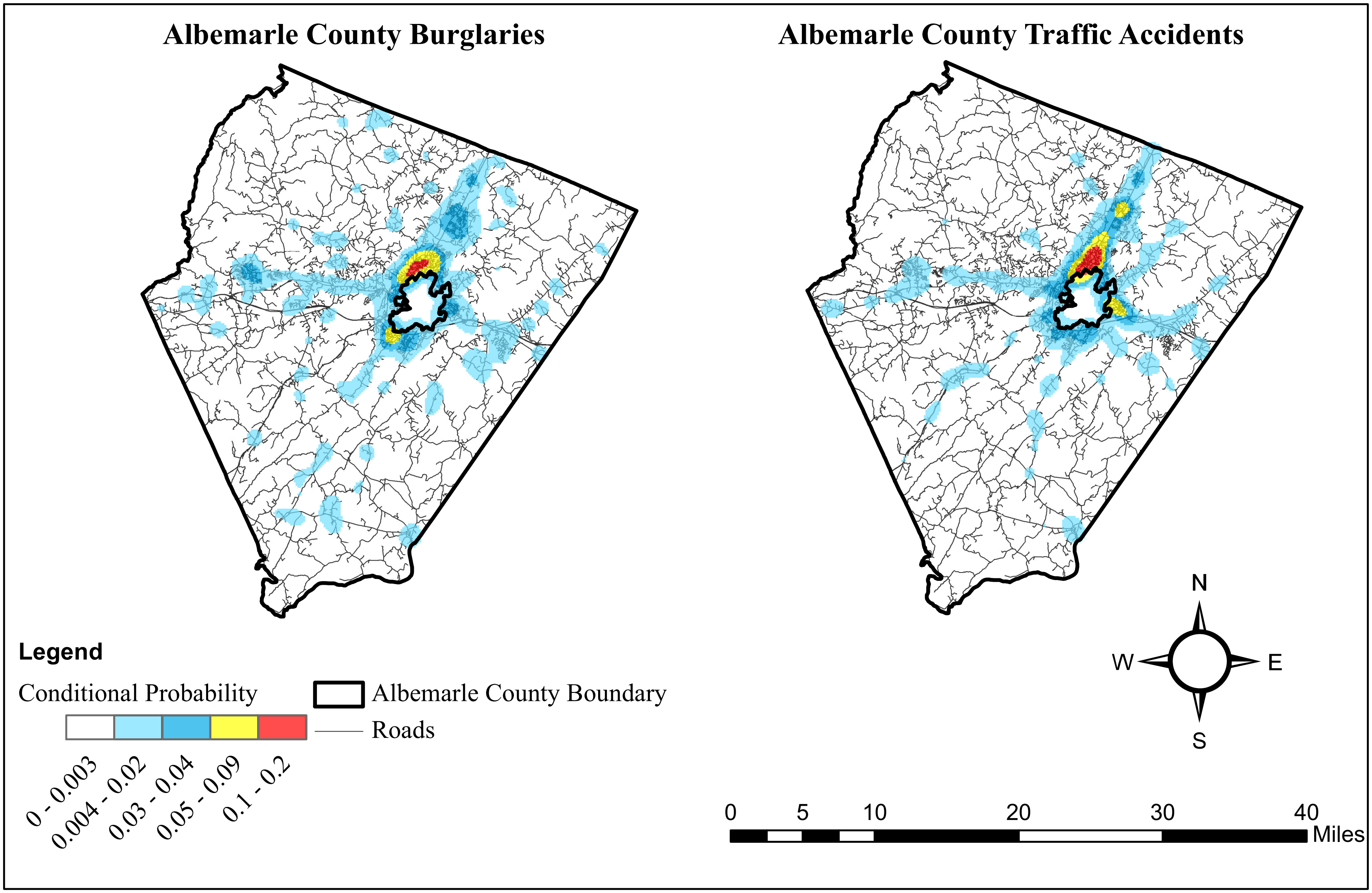


Figure 4.1: Geographic probability maps illustrating the conditional probability for B&E events (left panel) and traffic accidents (right panel).

the series. Then, seasonal effects are estimated by averaging over all unique time periods. Finally, the error component is estimated from the residuals that remain after the seasonal and trend components have been removed. We use the classical time series decomposition method provided by the *stats* package in R software for the example in this chapter. This time series decomposition approach can also be accomplished using the ubiquitous Microsoft Excel software using the procedure outlined by Lawrence et al. (2009). More robust time series decomposition methods available to security agencies with statistical software include X-11 and X-12 ARIMA decomposition (Hyndman and Athanasopoulos, 2013) and Seasonal Trend Decomposition based on Loess (STL) (Cleveland et al., 1990).

Figure 4.2 provides a graphic of the time series decomposition for Albemarle County monthly B&E and traffic accident events using classical time series decomposition. Note that the resulting error component is homoscedastic over the time period, indicating a well-fit time series decomposition. We used the current estimate for average monthly events as the best estimate for future monthly event counts. The notation for the domain forecast for the average event count at the domain-level for event type m is F_{mD}. The estimate provided by the time series decomposition of the ACPD dataset is 205 monthly traffic accidents and 21 B&E events.

4.3.3 Estimate Event Type Costs

The third step is to develop a cost estimate for each type of event. For this analysis, the cost estimate is based on the man-hours needed to meet the requirements of a CFS. The CFS database provided by the ACPD contains millions records for all service calls performed by responding officers. Many calls (especially traffic accidents) require more than one responding officer. The records in the database contain a unique index for each service call so that all radio reports for a given incident can be linked, unique identifier codes for each type of call (i.e., responding, arrival at CFS,

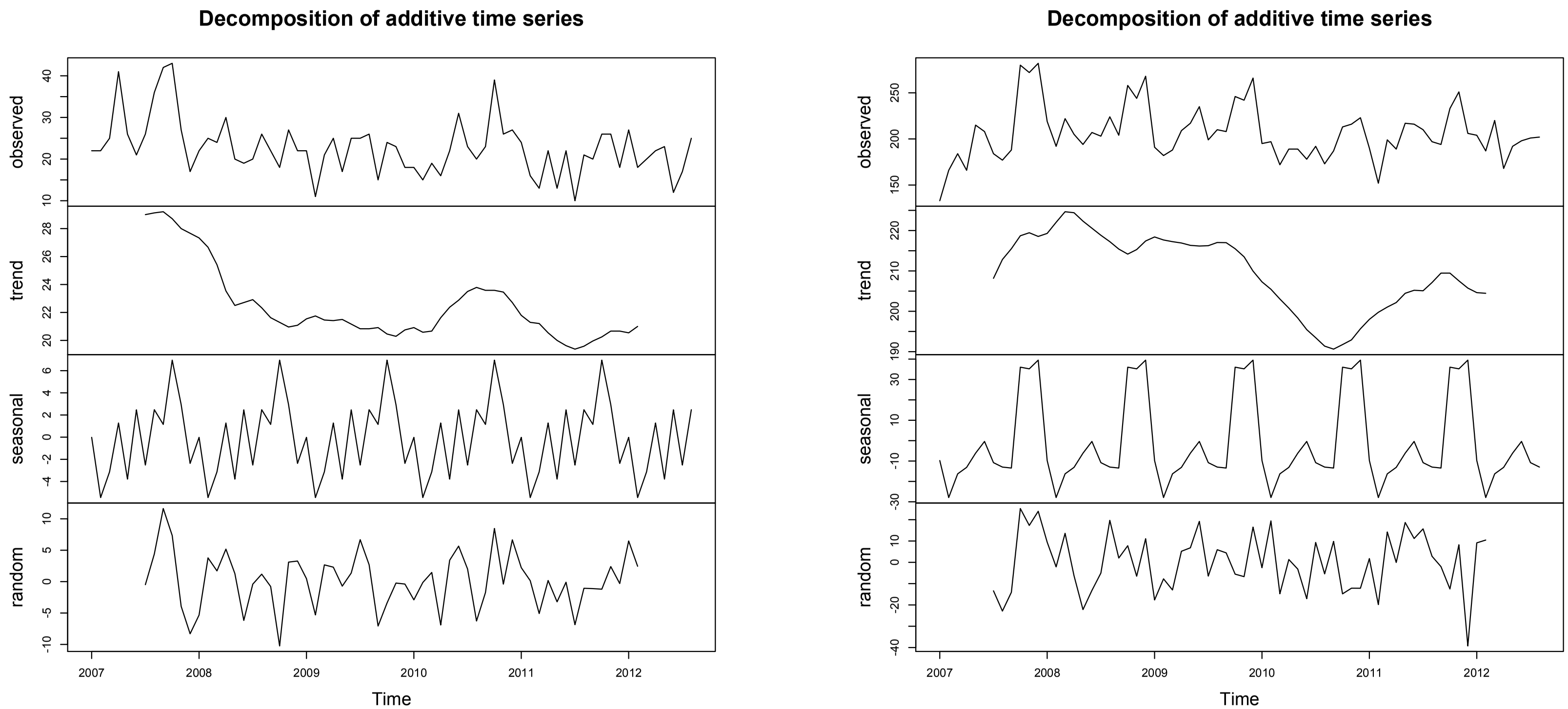


Figure 4.2: Time series decomposition of Albemarle County monthly breaking and entering (left panel) and traffic accident (right panel) events. The current estimate for average values are 21 burglaries and 205 traffic accidents per month.

departure from CFS, etc.), and time stamps. These records were parsed to produce estimates for the total man-hours required to service each type of call.

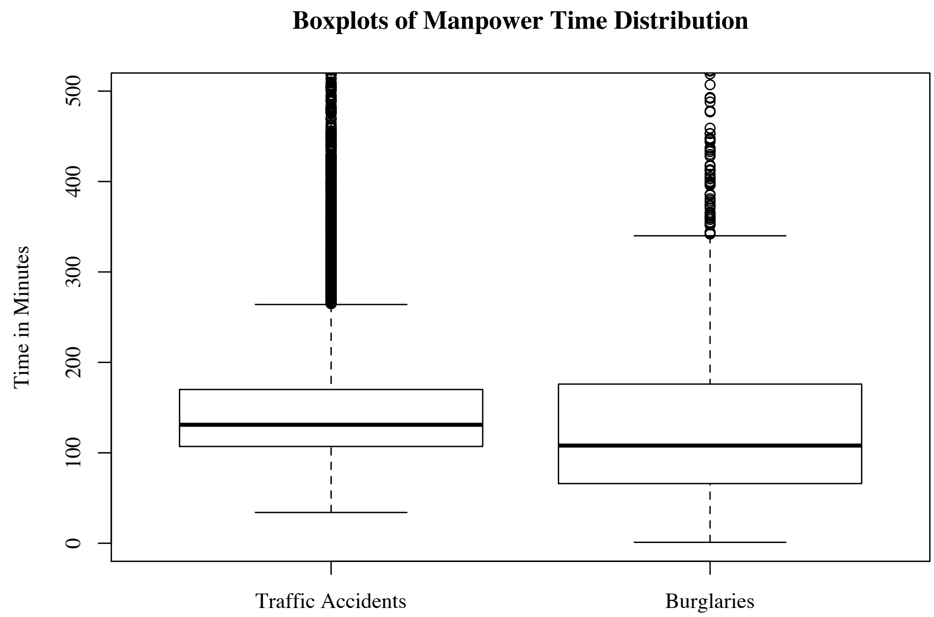


Figure 4.3: Boxplots of manpower time distribution for B&E and traffic accident events in Albemarle County.

Figure 4.3 provides box plots for the man-hours needed to service B&E and traffic accident incidents. As can be seen, these incidents are very time intensive and the probability distribution has high variance. However, under the assumption that the probability distribution for man-hours is independent of location, then the average man-hours needed to service each event type in a given patrol sector should converge to the average of the distribution over the long term. Thus, in this analysis, we used the average as the estimate for the cost of each event type C_m.

4.3.4 Generate a Geographic Planning Surface

The fourth step of the analysis is to develop a mapped geographic planning surface that maps the expected manpower required to service every unique location in the domain (see Figure 4.4). This estimate is not a forecast of the expected manpower needed during the next time period (i.e., it is not analogous to a short-term manpower forecast). Rather, it is an estimate of the long-term average manpower needed for a given unique location i. The planning factor is calculated for each unique location i as:

$$P_i = \sum_{m=1}^{M} C_m \hat{w}_{mi} F_{mD} \tag{4.3}$$

This planning factor is easily calculated and mapped using raster overlay operations in a GIS (Bolstad, 2008) or using statistical software that supports geographic analysis such as *R*. This calculation in Equation 4.3 is a simple sum-weight calculation, with the GIS system used to match and overlay the summation for the different incident types for each location. Figure 4.4 provides an illustration of the mapped P_i for Albemarle county estimated using B&E and traffic accident events. Note that there are very small regions of the county that generate many man-hours of activity while there are vast regions of the county that are expected to generate very little activity. The map in Figure 4.4 illustrates where CFS man-hours are expected to be concentrated in the future.

4.3.5 Develop and Analyze Districting Plans Using the Geographic Planning Surface

The last step of the analysis is to develop and analyze potential districting plans using the geographic planning surface. The planning surface provides a visualization of the average manpower demand expected at every unique location. This estimate is easily aggregated for geographic districts, providing a fast method for analyzing the long-term manpower requirements for a given patrol sector. With the notation for the regions (patrol sectors) within the domain indexed by j: $\{D_1, D_2, ..., D_J\}$, the

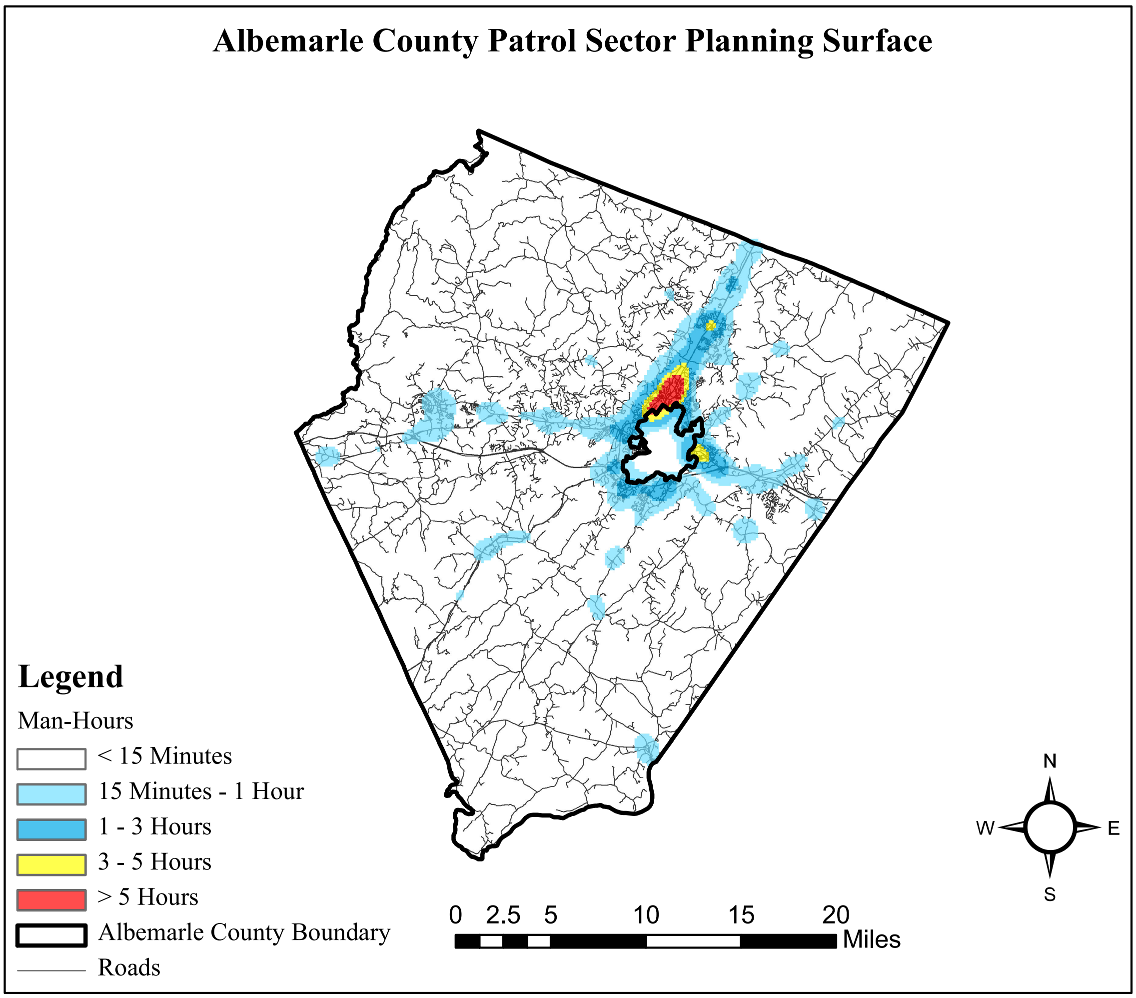


Figure 4.4: Planning surface for Albemarle County developed from B&E and traffic accident events

planning estimate for patrol sector j is the sum of the planning factor estimates that fall within it:

$$P_j = \sum_{i \in D_j} P_i = \sum_{m=1}^{M} C_m \hat{w}_{mj} F_{mD} \tag{4.4}$$

where:

$$\hat{w}_{mj} = \sum_{i \in D_j} \hat{w}_{mi} \tag{4.5}$$

Note that Equation 4.4 employs the Geographic Probability Forecasting (GPF) method, which has been validated in both empirical (Huddleston et al., 2013b) and simulation (Huddleston and Brown, 2013) studies. The only innovation to the planning surface used for geographic mission assignment is multiplying the forecast by the estimated cost C_m and summing over the various event types.

The planning approach employed for Albemarle county was to begin with the currently employed districting plan and then modify that plan to better balance the workload across the patrol sectors. Figure 4.5 provides maps for the original patrol sector plan employed by the county and the patrol sector plan recommended after this analysis. Table 4.1 documents the planning surface calculations used in the analysis. Figure 4.6 provides a graphed version of the information presented in Table 4.1, which is useful for articulating the planning approach.

After analyzing the current districting plan used by the ACPD with the planning surface, the workload imbalance experienced by patrol officers becomes very clear. As can be seen in the left panel of Figure 4.6, under the current plan, Patrol Sectors 3, 4, and 5 have both very large geographic areas to patrol and a large workload generated due to calls for service. As can be seen in Figure 4.5, most of the CFS for these districts are generated by the area immediately surrounding Charlottesville. Thus, the officers assigned to patrol these sectors are constantly responding to CFS

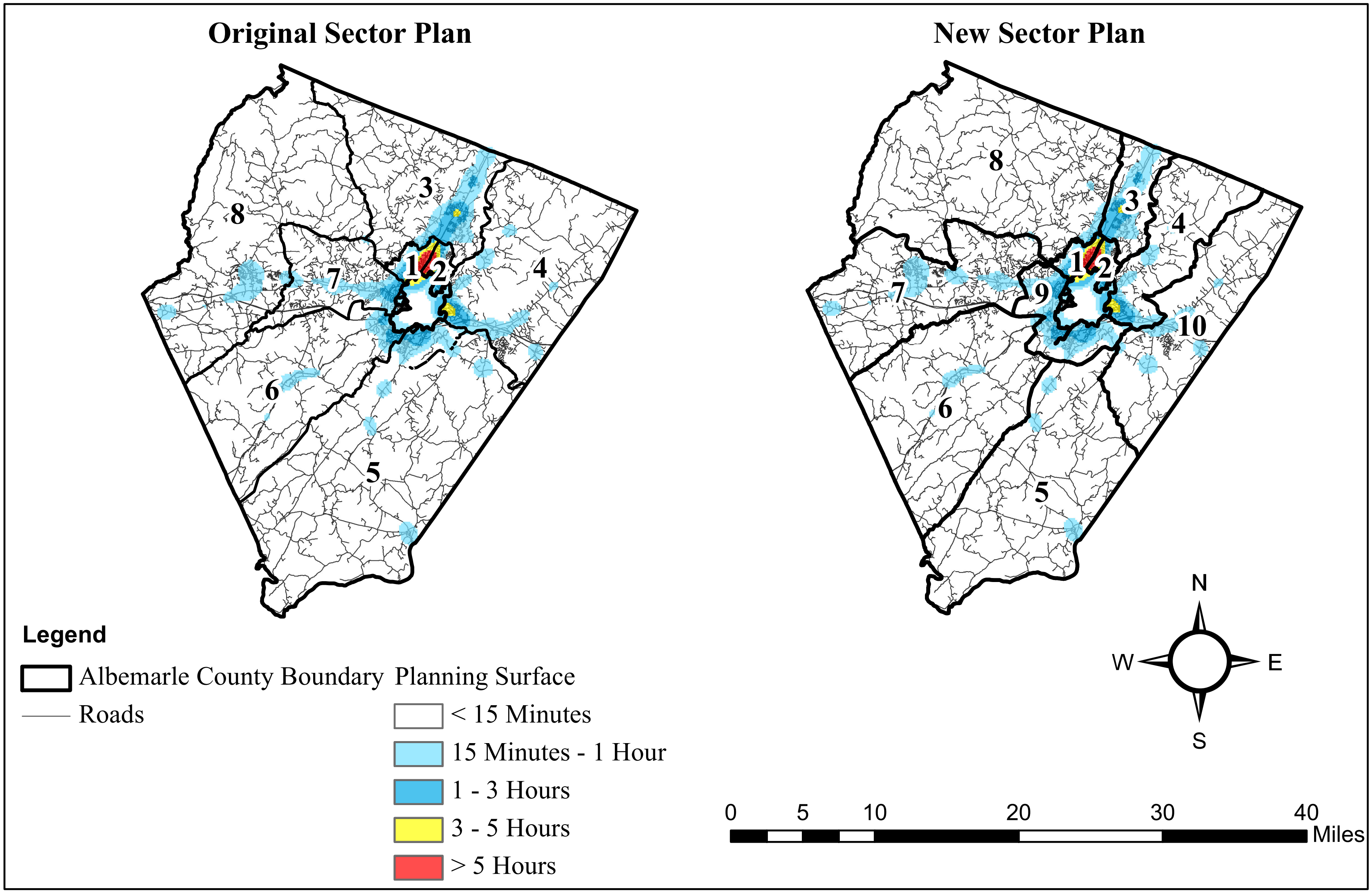


Figure 4.5: Map of Albemarle County Police Department's currently employed districting plan (left panel) and recommended plan (right panel) overlain over the geographic planning surface.

Table 4.1: Summary of patrol sector analysis for Albemarle County B&E and traffic accident events.

Original Patrol Sector Plan

Event Type	**Event Forecast**	**Event Cost**	**Patrol Sector 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
			Geographic Probability Weight ($\mathbf{w}_j$)							
Traffic Accident	205	2.65	16.0%	10.9%	17.6%	15.2%	15.9%	8.9%	8.0%	7.4%
B&E	21	2.28	14.6%	7.9%	16.4%	12.3%	18.9%	10.9%	8.9%	10.2%
Calls-for-Service Man-Hours			93.7	63.2	103.7	88.7	95.4	53.4	47.7	45.3
Geographic Patrolling Area (Sq. Mi.)			5.7	4.8	85.8	114.4	206.6	100.5	40.8	166.2

New Patrol Sector Plan

Event Type	**Event Forecast**	**Event Cost**	**Patrol Sector 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
			Geographic Probability Weight ($\mathbf{w}_j$)									
Traffic Accident	205	2.65	16.3%	10.9%	13.2%	11.2%	6.7%	6.2%	9.1%	6.4%	12.9%	7.1%
B&E	21	2.28	14.6%	7.9%	10.8%	7.2%	8.8%	7.9%	12.9%	7.8%	14.1%	8.2%
Calls-for-Service Man-Hours			95.4	63.2	76.9	64.3	40.7	37.5	55.3	38.3	77.1	42.6
Geographic Patrolling Area (Sq. Mi.)			5.5	4.6	23.8	53.7	137.1	123.6	91.5	169.7	25.1	90.4

near the center of the county, and are therefore rarely available to patrol the large districts they are assigned. These officers have a much higher workload than average for both CFS response and patrolling requirements.

The approach taken to develop a new patrol sector plan was to use the planning surface to break up the very large districts into smaller districts to better balance the workload between the competing objectives of preventive patrols and response to CFS. Because ACPD now provides a minimum of 10 patrol cars for each shift, the opportunity exists to add additional patrol sectors. The heuristic approach taken to develop the new patrols sectors relies upon the following design rules:

1. Use current districting boundaries when possible

2. Balance geographic patrol area requirements against CFS workload such that large patrol districts have low CFS man-hours

3. Design patrol districts such that officers assigned very intense CFS areas have minimal patrolling requirements

4. Do not divide responsibility for existing townships or neighborhoods between patrol officers (i.e. follow geo-policing principles)

5. Design patrol districts around high speed routes that allow officers to move quickly throughout their patrol sectors (i.e. E-W or N-W orientation for patrol districts around high-speed routes)

The modeling approach relied on manually updating the performance metrics used in Table 4.1 after each boundary change using summary table queries in ESRI ArcGIS software. These performance metrics are then mapped as in Figure 4.6, which provides a visualization of which sectors are most out of balance. Sector boundaries for patrol sectors that do not provide a good trade-off between balancing patrolling and CFS response workloads are then further adjusted. The precinct design process using this method is therefore very iterative, with each boundary change requiring the production of a table, investigation of the trade-offs of the in-progress plan using a graphical interface as demonstrated in Figure 4.6, and then further adjustment of the districts most out of balance. The planning surface provides a feedback mechanism that forecasts the average workloads over future time periods under any programmed boundary shift.

4.4 Results

Figure 4.6 illustrates how the use of the geographic planning surface allows us to develop an improved patrol sector plan. As Figure 4.6 shows, the new plan better balances the competing requirements on patrol officers. In the left panel, there is no correlation between an officer's patrol area and the man-hours needed to service CFS in an officer's district ($R^2 = 0.005$). In the right panel, there is a strong correlation ($R^2 = 0.776$) in the relationship between an officer's patrolling requirements and an officer's expected CFS workload. Officers with large districts that require more time to patrol should now spend less time responding to CFS requirements. The new patrol sector plan provides a much better trade-off in the workload demands upon officers and should ensure that all regions of the county are more regularly patrolled.

4.5 Conclusions

This chapter demonstrates an additional practical application of Geographic Probability Forecasting (GPF). GPF forecasting weights univariate time series forecasts using geographic probability maps. The GPF method for police district design demonstrated in this chapter first decomposes a domain-level forecast to obtain the current estimate for the average event count and then divides this forecast estimate probabilistically among the geographic patrol sectors. The geographic planning surface

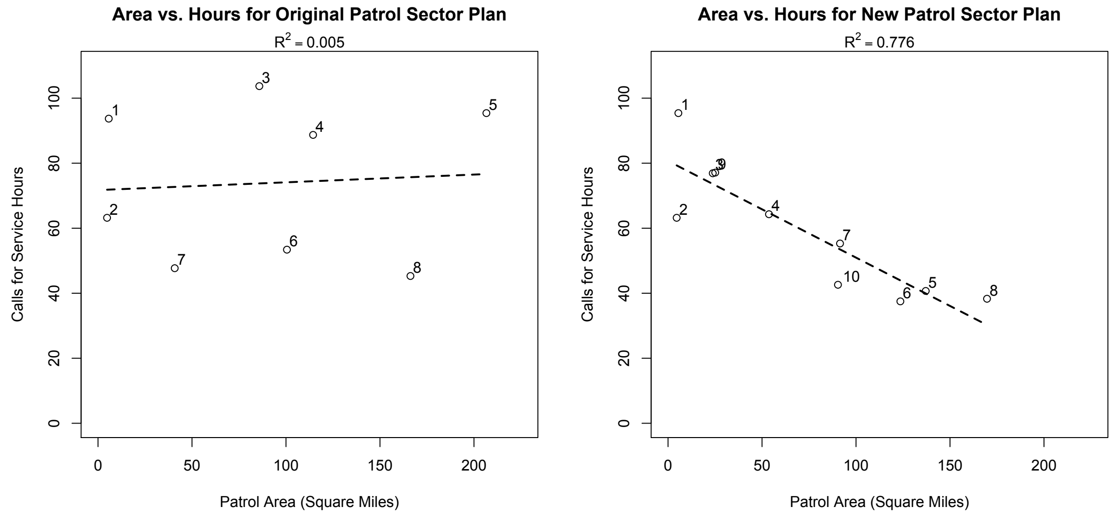


Figure 4.6: Workload balance between patrol sector area and CFS man-hours for the existing (left panel) and recommended (right panel) ACPD patrol sectors. The recommended plan better balances the competing workload requirements for officers between preventive patrols and CFS man-hours.

produced by multiplying the estimate for the long-term average event count at each unique location by a cost estimate for each event type produces an intuitive, interactive planning interface that allows planners to estimate the impacts of moving patrol sector boundaries. Because the planning surface is mapped, it allows planners to interact simultaneously with many heuristic planning rules, taking into consideration topology, geography, the "human terrain," the size of the patrol sectors, and available patrol units. This would seem to provide an opportunity for the development of patrol planning software that could be incorporated as a module into ArcGIS or crime mapping software, providing an automatically updated, interactive interface for police and military planners to use as they design patrol districts.

Chapter 5

Comparing Evaluation Methods for Police Patrol District Design

Nearly every police department creates geographic patrol districts (also called patrol sectors or car beats) as a standard management method to enhance the capabilities of the uniformed patrol force (Hale, 1980). As noted in Chapter 4, better police districting plans lead to lower response times, officer's familiarization with their assigned area, more efficient use of personnel, more equal division off workload, a visible police presence, enhanced officer safety, officer accountability, and balanced police response to calls (Hale, 1980). Traditionally, these geographic patrol boundaries are drawn by hand based on a police department's knowledge, experience, and the available police resources (Mitchell, 1972; Taylor and Huxley, 1989). Most police departments also lack a formal method for formally evaluating and comparing the performance of competing district plans, instead relying on the judgement and intuition of police planners. However, given the complexities of the police districting plan, it is unlikely that an optimal districting plan will be chosen by chance using this method (Curtin et al., 2010).

As discussed in Section 2.6 and the introduction to Chapter 4, researchers have proposed a variety of mathematical approaches to solving this complex problem that rely on optimization, set covering, N-P hard graph partitioning methods, genetic algorithms, or other advanced statistical approaches that are well beyond the capabilities of the vast majority of police departments. Instead, we propose that a more reasonable approach would be to provide police departments with the ability to evaluate the performance of districting plans they produce themselves and/or automated methods to generate a large number of possible districting plans. This approach would allow police departments the ability to find the optimal plans within any defined set. While these plans might not be globally optimal, they are likely to provide good performance, especially when compared to plans drawn by hand that consider the resulting performance trade-offs only in the minds of the planners.

For many police departments, police patrol district design presents a multi-objective optimization problem with two goals: minimizing workload variation between patrol districts and minimizing the response time for officers responding to calls for service. Fast response to citizen Calls for Service (CFS) improves the chances of arresting offenders, increases the chances of identifying and locating witnesses, provides immediate gathering of physical evidence, provides immediate life-saving aid, enhances the reputation of the police department, and increases citizen satisfaction with police (D'Amico et al., 2002; Hancock and Simpson, 2009). Therefore, virtually every police department seeks to minimize their average response time to CFS. Workload variation between districts arises because crime (and other CFS) tends to cluster in "hot-spots" rather than being uniformly distributed in the city. Workload variation between districts is often high, with some officers/districts experiencing much higher CFS volume than others. When small districts are created around very "hot" zones, the remaining districts can be quite large, resulting in slow response times for many citizens. Thus, minimizing workload and reducing response times are often competing objectives, requiring police to select a comfortable trade-off point between the competing objectives.

In this chapter, we compare three different methods for evaluating the performance of patrol district designs in this trade-off space: a closed form probability based approach, a discrete-event simulation based on hypercube models for spatial queuing systems, and an agent-based simulation model. We use the selection of a new patrol districting plan for the City of Charlottesville, Virginia, as a motivating example to compare and contrast the different methods for choosing a districting plan. We find that although all three methods provide similar evaluations of the districting plans when the emergency response system is not stressed, the agent-based simulation model more accurately represents the system dynamics when the system is highly stressed and also yields important insights into the system dynamics that the other two methods do not provide.

5.1 Charlottesvile Police Department Data

The Charlottesville Police Department provided the data used as the case study for this analysis. The City of Charlottesville is a mid-size city centrally located in the state of Virginia, USA. The city has a diameter of about 7 miles and a year-round population of about 40,000, which swells to about 66,00 during the academic year due to the presence of a major university. The current districting plan used by the CPD is about 20 years old. The city uses eight city patrol districts, with one car routinely assigned to each patrol district during each patrol shift. The police department operates three shifts a day: morning, evening, and overnight. As is the policy in many police departments, the CPD always dispatches the nearest available car to the scene of a CFS in an effort to minimize response time, rather than relying on each police car to respond to all calls within its district.

Figure 5.1 provides an illustration of how the demand for police assets varies over the 24 hour period. This graph references 330,000 CFS incidents observed over a four year period. During the night shift, the inter-arrival time for Calls-For-Service (CFS) is high, meaning that the CFS intensity is low. During the day and evening hours, CFS intensity is high, placing greater demands on the police patrols. As Figure 5.1 illustrates, the response time for police responding to CFS is highly correlated with traffic volume in the city. At night, police can respond relatively quickly to CFS because there is little traffic. During the morning and evening rush hour periods, it takes much longer for police to navigate traffic to the scene of calls for service. The time on the scene for a CFS remains relatively stable over the 24 hour period. The dashed lines in Figure 5.1 correspond to the three modeling scenarios used to study the performance of the police patrol district designs:

1. Low-Intensity Demand: 5 AM (Night Shift)

2. Medium-Intensity Demand: 7 PM (Evening Shift)

3. High-Intensity Demand: 9 AM (Day Shift)

The low-intensity scenario represents when the system is most idle. The medium-intensity demand scenario represents a time period when all system parameters are near their average values. The high-intensity demand scenario represents the time period when the system is most stressed - during the morning rush hour. Examining these scenarios allows us to evaluate patrol district designs under minimum, maximum, and average conditions.

Figure 5.2 provides one of the considered patrol district designs plotted over a map of the city of Charlottesville. As can be seen in Figure 5.2, we organized the city into 323 atoms (locations) for assignment. Each of these atoms must be assigned to one of the eight police districts. We generated 150 possible police district designs for consideration by the CPD using the procedure outline by (Zhang and Brown, 2013). This procedure develops districting plans that are both contiguous and compact. By

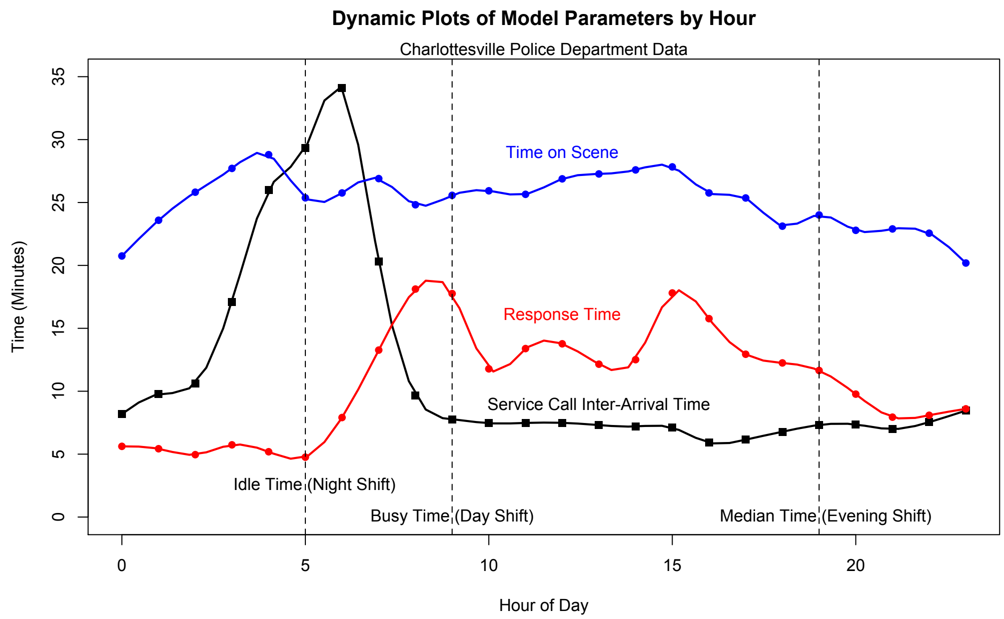


Figure 5.1: A plot showing how the service call inter-arrival time, response time, and time on service model parameters vary over time in the Charlottesville Police Department (CPD) data set. The dashed lines represent time periods selected for study in the simulation models: a low intensity (idle) period, a high intensity (busy) period, and a period representing the median situation.

definition, the patrol district must be contiguous so that one patrol car can patrol that area without departing. Police desire compact districts to provide shorter travel times within the patrol district. This procedure converts the NP-hard graph partitioning problem into a much more tractable problem: choosing the best of a defined set of options. The limitation of this approach is that the generated choice set is not guaranteed to contain the optimal solution. Rather, the approach we have taken provides a set of reasonable solutions from which we would like to choose the best available.

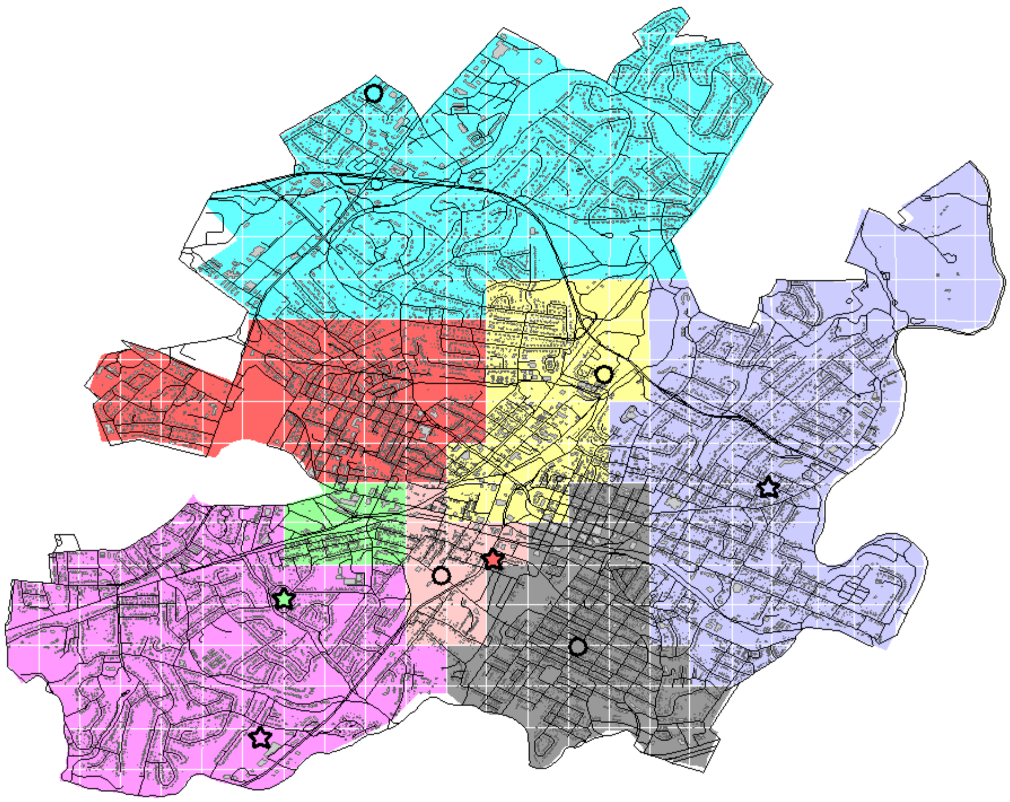


Figure 5.2: Visualization of District Plan 21 under evaluation in the agent-based simulation. The eight districts are color-coded. In-use patrol cars are labeled as stars while patrolling cars are labeled with circles. Note that the green and red patrol cars are responding to incidents out of sector because they were the closest available car to the incident at the time of the CFS.

5.2 Comparison of Police Patrol District Design Evaluation Methods

Given a patrol district design, there are several different approaches available for assessing the utility of that design. One approach would be to try the different districting plans by asking police patrols to change their patrol sectors every few weeks and assess how well the various districting plans worked. However, the number of possible districting plans that a police agency could try would be very limited and it would take a long time to test even a limited few competing plans. Efficient methods for evaluating patrol district designs without actually testing them in practice are therefore in high demand.

In this chapter, we compare three different methods for scoring district designs: a closed form probability based approach, a discrete-event simulation based on hypercube models for spatial queuing systems, and an agent-based simulation model. The closed form probability-based approach greatly simplifies the modeling problem and requires only geographic data in order to make an estimate. It is very simple, fast, and can be applied by virtually any police analyst with access to a GIS system. The discrete-event model takes longer to develop and makes some simplifying assumptions about the problem. However, once developed, it can evaluate districting plans quickly and it is relatively simple to adapt the model to different cities and scenarios. The agent-based approach takes the most development time, is harder to adapt to different cities and environments, and takes longer to evaluate competing plans. However, the agent-based modeling approach provides the most high-fidelity representation of the system and the most flexible modeling framework. We expound on each the three evaluation methods in greater detail below.

5.2.1 Closed Form Evaluation Method

The closed-form evaluation method relies on the relationship between location event CFS probability and the observed CFS counts over a geographic area. (Huddleston et al., 2013a) demonstrate that criminal hot-spot (probability) maps can be used to accurately forecast future crime counts within police patrol districts. These criminal hot-spot maps are two dimensional probability density functions that can be estimated using kernel density estimation (Harris, 1999), predictive crime models (Smith and Brown, 2007; Huddleston and Brown, 2009), or by binning historical crime counts by atom (Zhang and Brown, 2013). We use the binning approach in this chapter. These criminal hot-spot maps provide estimates for the probability of crime occurrence within each atom, with the notation π_i. For district j, the workload score w_j is estimated as the sum of atom event probabilities π_i within the district.

$$w_j = \sum_{i \in j} \pi_i \tag{5.1}$$

The district workload score w_j represents the proportion of work each district patrol is expected to perform. Since the objective is to provide equal workloads across the districts, district plans are scored using the sample standard deviation of the district workload scores σ_{w_j}. Lower workload standard deviation scores equate to better performance.

Criminal hot-spot maps can also be used to estimate the response time for officers to service calls within their districts. The response time score R is calculated as the sum of the probability weighted distances between each district centroid C_j and each atom location i.

$$R = \sum_{j=1}^{J} \left[\sum_{i \in j} \left(\pi_i ||C_j - i|| \right) \right] \tag{5.2}$$

In the formula above, the notation $||C_j - i||$ denotes the norm (distance) between the district centroid and atom i. Depending on the situation, Euclidean, Manhattan, or travel (road) distance can be used to estimate the travel cost. In this application, we used the Euclidean distance. Lower response time scores equate to better district plan performance. This method assumes that there will be very limited cross-boundary service by the patrols within the sectors.

5.2.2 Discrete-Event Simulation Model Method

The discrete event simulation model is based on the Hypercube Queuing Model (HQM), a well-known descriptive model used to analyze emergency response systems as spatially distributed queueing systems (Larson, 1974). In the HQM model, each server (patrol car, fire engine, ambulance, etc.) has two states: idle (0) and busy (1). The state of the whole system is represented as a binary sequence of server statuses. When the number of servers exceeds three, all possible system states form a hypercube.

Historical CFS incidents data and traffic information provide estimates for the arrival rates of the servers into each geographical atom. As long as the aggregate service rate of the system exceeds the total arrival rates of CFS incidents (i.e., supply exceeds the demand), calculating the steady-state probability of the resulting Markov chain provides the probability of being in each possible system state in the hypercube. System performance metrics such as average response time and workload variation are calculated from the hypercube probabilities. While the basic HQM provides a very flexible framework for modeling emergency response systems, the size of the problem grows exponentially with the number of servers. Solving each instance requires solving a linear system with an exponential number of variables (Boyaci and Geroliminis, 2011).

Boyaci and Geroliminis (2011) demonstrate that Monte Carlo discrete-event simulations based on HQM converge to the steady-state probabilities estimated by HQM very quickly. Therefore, discrete event simulations provide an alternative method for solving for the HQM steady-state probabilities. The discrete-event model can more easily be extended to simulate complex situations, such as multiple cars responding, different priorities of CFS incidents, different CFS arrival rates at different times of day, as well as various patrol and dispatch rules. We developed the simulation model for the CPD districts in Java 1.6 SE using pseudocode provided in Boyaci and Geroliminis (2011). Sacks (2003) provides the method we used to calculate the expected locations of CFS and patrol cars in the city using the Charlottesville data.

The inputs for the discrete event simulation model are:

- CFS Inter-Arrival Time
- Service Time (Time on Scene)

- CFS Probability for Each Atom
- Geographical Information (District Plan, District Plan Centroids, and Atom Centroids)
- Responding Speed

The simulation model generates CFS using the exponential distribution model defined by the inter-arrival rate parameter. The CFS incidents are spatially distributed within the city according to the geographic probability model generated from historical data as previously discussed (i.e., the CFS probabilities for each atom are the same as those used for the closed form evaluation method). CFS service times are randomly selected from the exponential model defined by the service time parameter. The parameters of the discrete event simulation model are then calibrated such that the queuing parameters (total arrival rate and total service rate) of the system match the historical data set. The discrete-event simulation model tracks the occurrence of four types of events:

- Calls For Service (CFS)
- Patrol Car Arrival at CFS
- Patrol Car Departure from CFS
- Patrol Car Arrival at Base (Idle Position)

When a CFS occurs in the simulation model, the nearest idle patrol car is "dispatched" by changing the server availability status from idle (0) to busy (1). The patrol car (server) status returns to idle once the car returns to its base location within the patrol sector after each event. To simplify the problem, the discrete-event model assumes zero line capacity; if all servers are busy when incident happens, the incident is "dropped" or considered as being responded to by units outside the modeled system. The historical Charlottesville Police Department data illustrated in Figure 5.1 provides estimates for the travel time, service time, and CFS inter-arrival time for the simulation model for each of the three considered scenarios. The simulation model dynamically tracks the average response time and workload standard deviation measures and stops when these measures converge.

5.2.3 Agent-Based Simulation Model Method

Agent-based simulation models are increasingly used to model the complex dynamics of resource allocation problems in security applications. Examples include optimizing the location of combat outposts in counter-insurgency (Huddleston et al., 2008), examining the use of unmanned surface vehicles for securing Navy ships (Cioppa et al., 2004), and emergency management in disaster response (Wu et al., 2008). Agent-based modeling provides the ability to accurately represent the behaviors of these complex systems by modeling the interactions of the agents of the system.

Zhang and Brown (2013) provide an agent-based model that captures the behaviors of police patrols in a city through the use of use of model parameters and decision rules. This model is based on RepastCity from Malleson (2010), which implements agent movement along roads in an urban GIS environment and provides a flexible framework for adapting the simulation by changing the associated GIS layers. The simulation model operates by having the agents and environment interact through the use of simple rules. The rules for this agent-based simulation model are:

1. The simulation model generates CFS using the exponential distribution model defined by the inter-arrival rate parameter.

2. CFS incidents are spatially distributed within the city according to the geographic probability model generated from historical data as previously discussed.

3. CFS service times are randomly selected from the exponential model defined by the service time parameter.

4. Police cars randomly patrol the road network within their defined district when not in service.

5. The nearest available patrol car (regardless of district) responds to a CFS at emergency speed.

6. The responding car takes the shortest (road network) path to the location of the CFS.

7. Upon completion of the CFS, if the patrol car is out of its district, it returns to its district moving at the speed limit and begins patrolling.

8. Upon completion of the CFS, if the patrol car is within its district, it begins randomly patrolling from its current location.

The input parameters for the agent-based simulation include the CFS inter-arrival time, CFS service time, the emergency speed, and geographic information (the road network, road network speed limits, and the district plan). We calibrate the agent-based simulation model by tuning the model until simulating the currently employed districting plan with the simulation model produces the historically observed average response time in all three modeled scenarios. The simulation model dynamically tracks the average response time for all cars and the workload (time in service) proportion for all cars. We run the simulation for a district plan until the average response time and workload proportion for all cars converge to a steady state.

5.3 District Plan Selection

The three district plan evaluation methods above each provide an approach for scoring district plan response time and workload variation performance. The goal is to identify the district plans that provide good performance in both objectives. For multi-objective problems such as this one, there usually does not exist a single solution that simultaneously optimizes both objectives. Instead, there exists a (possibly infinite) set of Pareto-efficient solutions. A solution is Pareto-efficient (also called non-dominated or Pareto-optimal) for two objectives if one cannot improve performance in one performance measure by selecting a different alternative without sacrificing performance in another. Graphing the performance of the solutions provides a simple way to identify the Pareto-efficient frontier (Gass and Saaty, 1955). Figure 5.3 illustrates the trade-off space and resulting Pareto-efficient solutions identified by the agent-based model in both the low-intensity and high-intensity settings. The agent-based model identifies five non-dominated solutions for the low-intensity scenario and two non-dominated solutions for the high-intensity setting. We used the same approach to identify the non-dominated solution set in each scenario for the closed-form and discrete-event simulations. Figure 5.3 also identifies these solutions as well as the current district plan used by the CPD.

5.4 Results

As Figure 5.3 illustrates, there is some disagreement between the three different methodologies about which plans are best. Note that in the low-intensity scenario, the agent-based simulation model scores all of the non-dominated solutions by the other two methods relatively highly (they are all clustered in the lower-left hand corner). There is also some agreement on plans that are Pareto-optimal, with some districting plans on the Pareto frontiers of all three methods in the low-intensity scenario. However, in the high-intensity scenario, the Pareto-efficient plans identified by the closed-form and discrete-event methods tend to be rated as relatively average in at least one measure by the agent-based simulation.

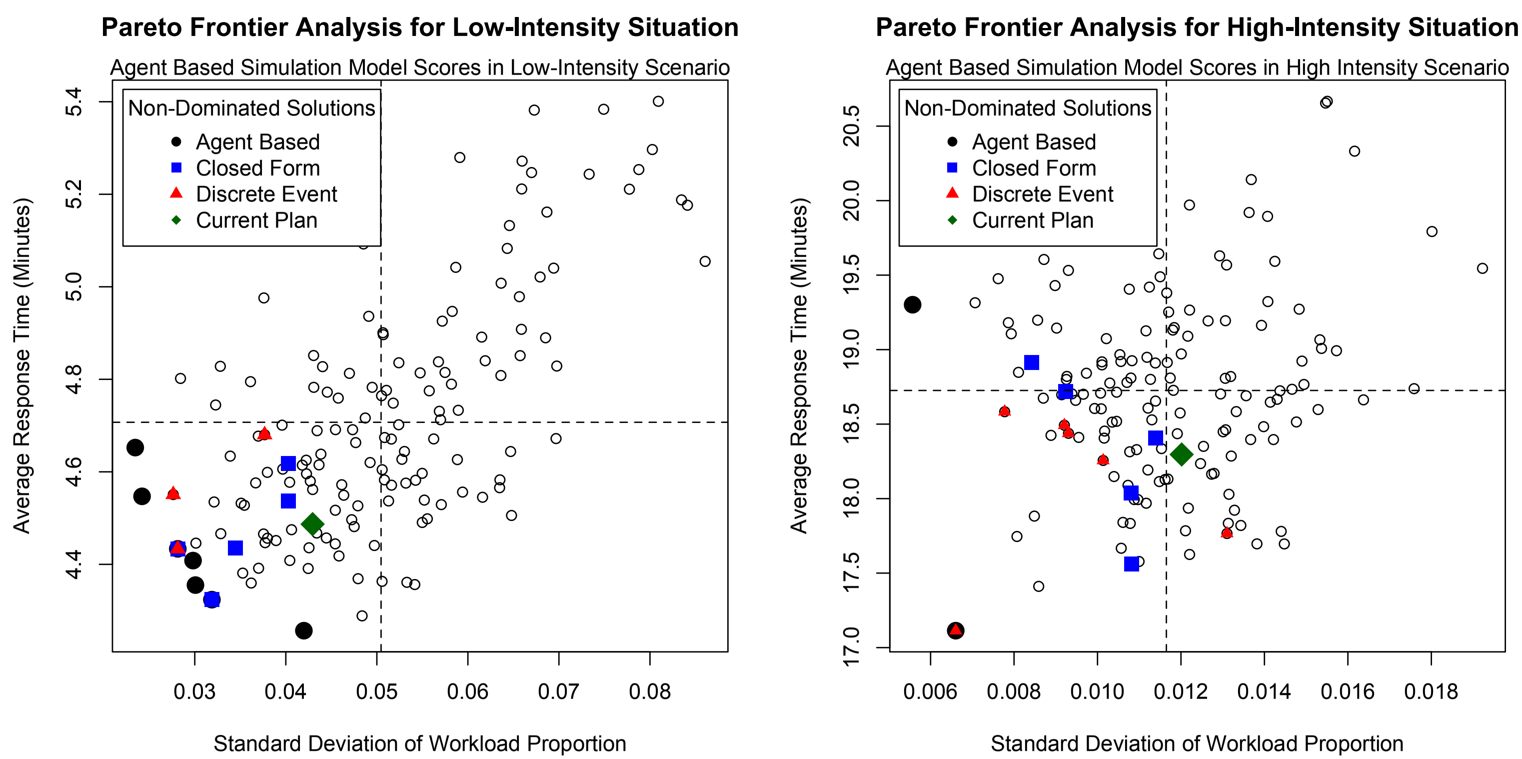


Figure 5.3: Pareto analysis using the scores from the agent-based simulation in the low-intensity scenario (left panel) and the high-intensity scenario (right panel). Black circles represent non-dominated solutions identified by the agent-based model while colored points represent the non-dominated solutions identified by the other approaches. The dashed lines provide a visual reference for average performance in each evaluation measure as defined by the agent-based model. Note that all three methods provide highly scored non-dominated solutions in the low-intensity scenario but that solutions recommended by the CF and DE methods are rated as relatively average by the agent-based method in the high-intensity scenario. Also note that the average response times on these graphs for the current district plan correlate closely to the observed historical response times for the two scenarios in Figure 5.1, indicating a well-calibrated simulation model.

Table 5.1 and Figure 5.4 provide an explanation for these differences. Table 5.1 provides the coefficient of determination (R^2) statistic comparing Closed Form (CF), Discrete Event (DE) and Agent-Based (AB) scores for workload variation and response time under low, medium, and high event intensity conditions. Figure 5.4 provides a pair-wise scatterplot for the most correlated (workload variation in the low-intensity scenario) and least correlated (workload variation in the high-intensity scenario) situations in this table for a visual reference. All methods provide highly correlated workload variation scores in the low intensity scenario but highly uncorrelated scores for workload variation in the high-intensity scenario. The response time scores are less correlated than workload variation scores in the low-intensity scenario but more correlated than workload variation in the high-intensity scenario. For both performance measures, correlation between methods decreases as event intensity increases.

Table 5.2 further explains the results observed in Table 5.1. Table 5.2 provides the

Event	Cross-	Workload Variation			Response Time		
Intensity	Sector %	CF- AB	CF - DE	DE - AB	CF - AB	CF - DE	DE - AB
Low	42%	0.65	0.70	0.81	0.45	0.51	0.74
Medium	70%	0.33	0.44	0.47	0.30	0.44	0.72
High	75%	0.13	0.35	0.19	0.21	0.43	0.52

Table 5.1: Table of pairwise R^2 statistics comparing Closed Form (CF), Discrete Event (DE) and Agent-Based (AB) scores for Workload Variation and Response Time under low, medium, and high event intensity conditions.

Pairwise	Workload Variation			Response Time		
Comparison	CF	DE	AB	CF	DE	AB
Low-Med	1	0.69	0.45	1	0.92	0.55
Low - High	1	0.60	0.24	1	0.91	0.41
Med - High	1	0.91	0.11	1	0.98	0.46

Table 5.2: Table of within-method pairwise R^2 statistics showing how the Workload Variation and Response Time evaluation scores correlate within methods across the three scenarios. The Closed Form (CF) method provides the exact same evaluation scores for every scenario. The Discrete-Event (DE) method provides similar scores across the three scenarios. The Agent-Based (AB) method provides very different scores across the three scenarios, especially for Workload Variation.

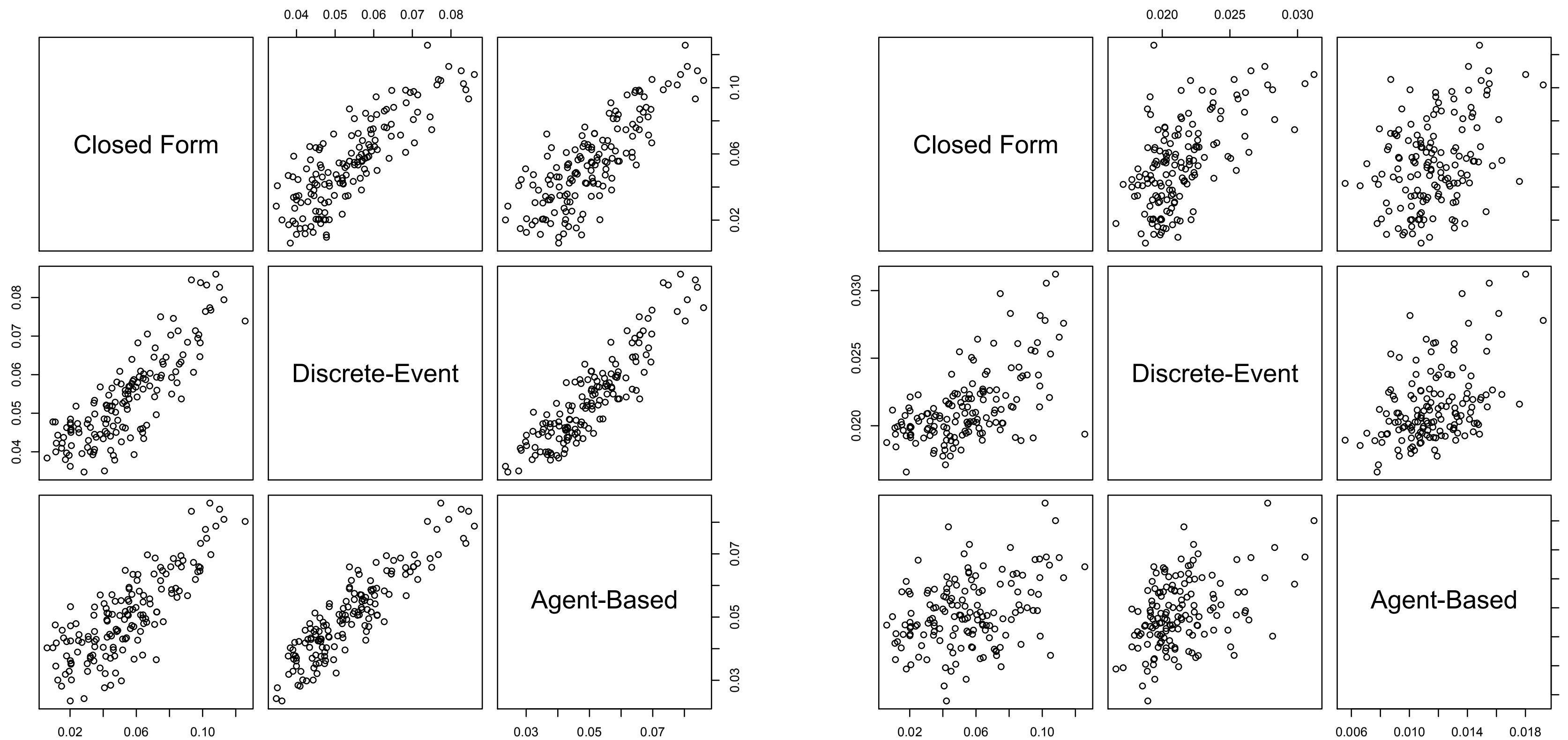


Figure 5.4: Pairs plots showing pairwise correlation between workload variation scores when event intensity is low (left panel) and when event intensity is high (right panel). All three methods score district plans similarly when event intensity is low but differently when event intensity is high.

within-method coefficient of determination (R^2) across the three scenarios. As can be seen, the closed form method provides the exact same scores for every scenario, the discrete event scenarios are highly correlated across scenarios, but the agent-based scores change significantly.

These results prompted further analysis to understand why the agent-based simulation model scores change so significantly in the high-intensity scenario. We identified two dynamics within the system that cause the agent-based simulation model to significantly alter the scores as CFS intensity increases. The first insight the agent-based model provides concerns the effect of the patrolling behavior of the police cars. When the police cars randomly patrol within their districts, they are often far from the patrol district centroid (as can be seen in the snapshot of the agent-based simulation model in Figure 5.2). The CPD always dispatches the nearest available police car to the scene of a CFS. Thus, cross-boundary support is quite frequent. In the low-intensity scenario, cross-boundary response averages about 42%. However, this cross-boundary support rises to 70% in the medium intensity scenario and 75% in the high-intensity scenario. These rates roughly correspond to the rates observed in a 1971 New York City study that found that cross-boundary support accounted for more than half of police dispatches (Larson, 1971).

The second significant system dynamic is the effect CFS intensity and slow response times due to traffic have on the workload variation during the busy periods of the day. The difference in workload variation (standard deviation of the workload proportion) among districting plans during the high-intensity period around the morning rush hour is very low (note the difference in scales on the horizontal axis in Figure 5.3). During the high-demand period, all police cars experience a high workload due to the high CFS intensity and slow response speeds due to traffic. Thus, the districting plan has little to do with the workload officers experience during this busy time; for the most part, the police cars are all responding to CFS. This observation yields an important insight for the CPD. Counter-intuitively, the districting plan becomes most relevant when CFS intensity is low and less important when CFS intensity is high. This is because when CFS intensity is low, the officers spend most of their time patrolling, but when CFS intensity is high, all officers are responding to calls rather than patrolling (on average, 80% of available officer man-hours are employed responding to calls during this period). During the peak rush-hour periods, it may be possible to significantly reduce the average CFS response time by positioning police cars throughout the city near those locations most likely to need CFS during this busy time instead of having officers attempt to both patrol throughout the districts and respond to calls, especially since officers spend relatively little time patrolling the districts they are assigned. In discussions with the CPD, they verified this effect and commented that the system dynamics observed in the agent-based model seemed to correspond closely to that experienced by their officers. In this case, the agent-based simulation model reveals complexities in behavior and applicable insights that the other two evaluation methods do not provide.

5.5 Conclusions

Our results indicate that all three evaluation methods produce very similar scores for workload variation when CFS intensity is low enough that the car patrols can meet the demand in their own sectors. However, when the in-district demand exceeds in-district supply, police patrols begin crossing boundaries to meet demand in other police sectors at a very high frequency. This scenario produces a level of complexity that the closed form and discrete event approaches are not well-equipped to handle. Only the agent-based simulation model accurately represents the resulting complexities and significantly changes the workload variation scores to reflect the behavior of the system. The significant insight the agent-based model provides is that, because call volume is so high, officers rarely patrol their sectors in this period, instead spending most (on average about 80%) of their time responding to calls both in and out of sector. The visualization of the system's complexities the agent-based model provides was also helpful in validating the performance of the simulation with the CPD client.

The scores the three methods provide for response time were less correlated with each other in the low intensity setting than they were for workload variation (ranging between 0.45 and 0.74). However, the correlations between response time scores for the three methods were less sensitive to changes in intensity than the workload variation scores, and the discrete event and agent-based simulations maintained relatively high correlation with each other throughout all three scenarios. The closed form approach did not seem to provide good estimates as it did not have high correlation with either of the other two methods in any of the scenarios. This is probably due to the fact that this method does not account for cross boundary support, and therefore underestimates the effect out of sector CFS have on the average response time.

Future work for this study includes extending the discrete event and agent-based simulation models to dynamically change the modeling parameters for response speed, service time, and inter-arrival time over the 24 hour cycle to correspond with the rates seen in Figure 5.1. Using this approach will provide an estimate for how well the various districting plans perform over a 24 hour period in actual practice. Planned extensions to the current simulation models include more complex response rules such as call prioritization and multiple car response for certain types of calls. Planned extensions to the closed form approach include performance comparisons using other distance measures (i.e., road-network distance, Manhattan distance, etc.) and development of methods for estimating the effect of cross-boundary support on the average response time performance measure.

Chapter 6

Mapping Gang Spheres of Influence

Many urban environments have criminal gangs competing for control of available resources and territory. Intelligence analysis of these groups requires not only the prediction of future attack locations but also answers to questions such as: Who is the most likely perpetrator of a criminal incident at a given location? What is the most likely course of action for a given criminal group? What makes one location more likely to experience a gang incident over another? With limited resources, how can I best employ those resources most efficiently to engage specific criminal elements? Multilevel modeling of criminal site selection preference allows us to better answer these questions by linking the incidence of gang crime to the spatial, demographic, and socio-economic features of specific locations.

There is a rich body of literature that links gang activity spaces, socioeconomic conditions, and other factors to the incidence of crime committed by criminal gangs (Taniguchi et al., 2011; Tita and Ridgeway, 2007; Tita and Cohen, 2004; Block, 2000). Many previous studies of gangs note that criminal gangs seize, control, and defend home territories from rival gangs (Thrasher, 1927; Whyte, 1937; Ley and Cybriwsky, 1974; Bernasco and Block, 2009). Other studies explore the effects that gang formation have on crime patterns and rates in the local communities where they form (Tita and Ridgeway, 2007). For instance, Tita and Cohen (2004) demonstrate that areas where gangs congregate are highly correlated with high levels of gun violence and Ratcliffe and Taniguchi (2008) show that drug-gang street corners are highly correlated to the general incidence of crime. Several other studies also find that there exists a strong relationship between the number of gangs that are active in an area and the general level of criminal activity (Block, 2000) and that gang set spaces serve as crime attractors and crime generators (Tita and Ridgeway, 2007). However, analysis of at least one well-known gang data set demonstrates that many gangs commit as many as half of their crimes outside of their own controlled gang territories (Meeker et al., 2002). An analysis of the six most active gangs in Santa Ana, California for the two year period from 1999-2000 reveals that these gangs committed 30% of their crimes outside any known gang territory and an additional 19% of their crimes in areas claimed by more than one gang, which we term gang conflict territories. This indicates that an analysis of gang activity beyond their home territories is warranted. Multilevel modeling of criminal site selection provides a means for structuring this analysis and Geographic Information Systems (GIS) provide a means for communicating the results of this analysis in a format employable by police agencies.

6.1 Background

This research chapter merges research on criminal activity spaces with a multilevel modeling extension to previous criminal site selection methods and describes why the approach better predicts the actions of specific criminal groups and facilitates identification of gang *criminal spheres of influence* - the geographic regions in which each criminal group presents the greatest threat. This methodological approach rests on several foundations of research: data mining, criminal hot-spot prediction, criminal site selection modeling, and multilevel (hierarchical) modeling.

6.1.1 Criminal Hot-Spot Prediction

As noted above, many researchers have linked gang *activity spaces, drug corners, set spaces*, or *turfs* to concentrations of criminal activity, or *hot-spots*. The National Institute of Justice defines a criminal hot-spot as "an area that has a greater than average number of criminal or disorder events, or an area where people have a higher than average risk of victimization (Eck et al., 2005)." There are many techniques for identifying criminal hot-spots but they for the most part fall into one of two classes: those techniques that leverage only historical location data in the analysis (treating the problem as a spatial point pattern) versus those that treat the problem as a *marked* spatial point pattern. Point patterns are the type of spatial data that arise when the critical variable being analyzed is the location of events (Cressie, 1993). Most criminal incidents fall into this category of geographic analysis. A *marked* spatial point pattern is one in which the events in a point pattern are associated with features, measurements, or categorical marks. In crime analysis, examples include identification of the type of crime, the responsible party (if known), and the geographic features associated with the location of the criminal event.

There are several techniques that have been developed to identify criminal hot-spots using only the spatial point patterns generated by past observations. The most common approaches to identifying criminal hot-spots rely on kernel density estimation because these approaches are easily implemented in the Geographic Information Systems (GIS) most police agencies now employ (Eck et al., 2005; Boba, 2005). These techniques do not leverage the additional "marked" information associated with a criminal event but leverage only location (Latitude-Longitude or X-Y) data to estimate the relative risk associated with each X-Y coordinate on the map. Recently, an approach using self-exciting point process models has been shown to improve upon kernel density methods for predicting burglaries in Los Angeles in both space and time (Mohler et al., 2011) and the same approach has been applied to modeling civilian deaths in Iraq (Lewis et al., 2012).

In recent years, there has been a growing body of literature in which researchers identify criminal hot-spots by using the *marks* associated with crimes in police databases to identify criminal hot-spots. Social researchers tend to use various regression techniques in the application of environmental criminology to link social, economic, or spatial features to the incidence of crime (Brantingham and Brantingham, 1981). Examples of sociological analysis include identification of factors important in the occurrence of residential burglaries (Bernasco and Nieuwbeerta, 2005), robberies in Chicago (Bernasco and Block, 2009), the link between drug street corners and crime (Ratcliffe and Taniguchi, 2008), and several of the previously mentioned studies in gang activity (Tita et al., 2005; Tita and Ridgeway, 2007; Block, 2000). Other researchers have begun to apply newly developed data mining techniques to the problem of identifying the areas most likely to see a criminal incident. Data mining approaches to hot-spot identification include machine learning techniques such as neural networks (Olligschlaeger, 1997), fuzzy clustering (Grubesic, 2006), and support vector machines (Chang et al., 2005; Kianmehr and Alhajj, 2008).

6.1.2 Criminal Site Selection Models

There is one modeling approach that provides both insight into the environmental processes that generate crime (the focus of sociological inquiries) and improved predictive performance: spatial choice modeling. Spatial choice models are based upon the work of Daniel McFadden's development of discrete choice theory (McFadden, 1974). In McFadden's formulation, actors, indexed by j, evaluate the utility, U, that they would derive from choosing an alternative based upon the features or attributes of that alternative:

$$U_{ij} = \beta X_{ij} + \epsilon_{ij} \tag{6.1}$$

In the above formulation, X denotes the vector of features or attributes for alternative i. The ϵ term captures the error associated with each pair of actors and alternatives while β records the regression coefficients of the model. McFadden established the theoretical foundation for the use of conditional logistic regression to model choice from a discrete set of alternatives. When actors are choosing from a discrete set of alternatives, then their probability of selecting alternative i, $P(y = i)$, can be modeled using the well-known logistic regression equation:

$$P(y = i) = \frac{e^{\beta X_{ij}}}{\sum_{i=1}^{N} e^{\beta X_{ij}}} \tag{6.2}$$

Several groups of researchers have applied this approach in a spatial context for modeling criminal site selection preference. Several examples of the direct application of McFadden's discrete choice theory to crime include an analysis of the target selection by burglars in The Hague, Netherlands (Bernasco and Nieuwbeerta, 2005) and several studies of robberies in Chicago (Bernasco and Block, 2009; Bernasco et al., 2012). Xue and Brown (2006) develop criminal site selection models that adapt the spatial choice modeling approach for conditions in which the individual discrete choices (crimes) cannot be attributed to individual criminals, which is the case for most of the crime data available to police for use in predictive policing. Their work provides an extensive discussion of the assumptions involved in this model adaption but, in brief, their approach relies upon assuming that both the choice set and the decision-making preferences of all of the modeled actors (criminals) in the study domain are similar, and the model therefore describes what is generally true about the criminal preferences in a geographic region.

Xue and Brown (2006) also incorporate the idea of using feature-space rather than geographic coordinates to represent the locations of crimes. Feature-space is defined as the Euclidean distance to each of the features of interest such as various crime attractors and crime generators (Liu and Brown, 2004). Their research group has shown that various forms of these criminal site selection models significantly improve predictive performance over the traditional kernel density method in predicting burglaries (Liu and Brown, 2004; Xue and Brown, 2006) and terrorist events (Brown et al., 2004) such as suicide bombings (Smith and Brown, 2007). On noted reason for this performance improvement is that these criminal site selection models can highlight high risk areas (those very likely to observe a future criminal incident based upon the features of that location) that kernel density approaches do not highlight because they are far from previously observed crimes (Liu and Brown, 2004). Huddleston and Brown (2009) extend these criminal site selection models using multilevel modeling to further improve performance for predicting the locations of crimes by specific criminal street gangs.

6.1.3 Multilevel Modeling

Multilevel models, sometimes called heirarchical models, extend traditional regression models by allowing regression coefficients to vary from group to group (Gelman and Hill, 2007). The two common alternatives to multilevel models are pooled models, in which all groups are pooled together and treated as one, and no-pooling models, which build a separate model for each group. No-pooling models often suffer by not considering generalities of behavior that are captured only when all incidents are included in the model development. On the other hand, modeling spatial behavior with a pooled model can apply generalities to groups to whom they may not apply (Fotheringham et al., 2000). Multilevel modeling addresses both of these shortcomings simultaneously by partially pooling the results of both analyses (Gelman and Hill, 2007).

6.2 Santa Ana Gang Data

Data for this analysis came from three sources, which are illustrated in Figure 6.1. First, the authors used the Gang Incident Tracking System (GITS) database to evaluate the performance of this new methodological approach. The GITS project was introduced in 1993 to help law enforcement officials in Orange County, California "make more informed decisions to counter gang activity, which had been on the rise in recent years (Meeker et al., 2002)." The authors used a subset of the data particular to the city of Santa Ana, California, for the period from 1994 through 2000. Incidents from the period 1994-1998 were placed into a training set and data for the period 1999-2000 was held out to serve as a model performance test data set. This approach was taken to mimic the approach that would be taken by law enforcement agencies in using statistical software to predict gang activity in their jurisdictions. In order for an incident to be classified as gang activity and entered into the database, it went through a rigorous verification process described by (Meeker et al., 2002). Information about each gang incident included information on the responsible gang (if known), the specific crime (one of 21 different crimes such as felonious assault, homicide, burglary, etc.), criminal event type (violent, weapons, property, drug, or vandalism), and geographic information about the location of the crime. Table 6.1 provides a summary of the various crime counts by gang and crime type for the training and test data sets.

(Meeker et al., 2002) extensively document the three step verification process used to add gang-related incidents into the data base. An incident was added to the database if it met one of four criteria for establishing a gang-related incident:

Table 6.1: Gang-related criminal incidents for the training (1994-1998) and testing (1999-2000) data sets.

Gang ID	**Crime Type**												**Total**
	Violent		**Property**		**Drugs**		**Weapon**		**Vandalism**		**Total**		
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	
1	29	9	4	2	5	1	13	7	7	3	58	22	80
2	25	9	2	4	5	2	6	10	2	1	40	26	66
3	31	1	7	-	5	-	16	7	7	1	66	9	75
4	21	7	4	1	2	-	6	3	4	1	37	12	49
5	37	11	5	-	2	-	13	3	8	3	65	17	82
6	20	11	3	-	-	3	7	4	-	6	30	24	54
7	13	4	1	1	-	1	9	5	4	3	27	14	41
8	19	3	6	2	3	1	4	2	4	1	36	9	45
9	26	4	3	-	2	-	11	1	2	1	44	6	50
10	9	7	1	-	3	-	6	3	2	1	21	11	32
11	15	3	4	1	1	-	3	-	2	2	25	6	31
12	18	3	3	2	1	-	2	1	-	-	24	6	30
13	8	3	2	-	1	1	11	1	-	-	22	5	27
14	8	8	5	-	-	-	3	4	3	-	19	12	31
15	9	7	-	-	-	-	4	2	-	-	13	9	22
Other Gangs	193	71	23	7	16	10	95	38	42	25	369	151	520
Unknown	2615	1204	72	10	19	3	209	31	63	16	2978	1264	4242
Total	3096	1365	145	30	65	22	418	122	150	64	3874	1603	5477

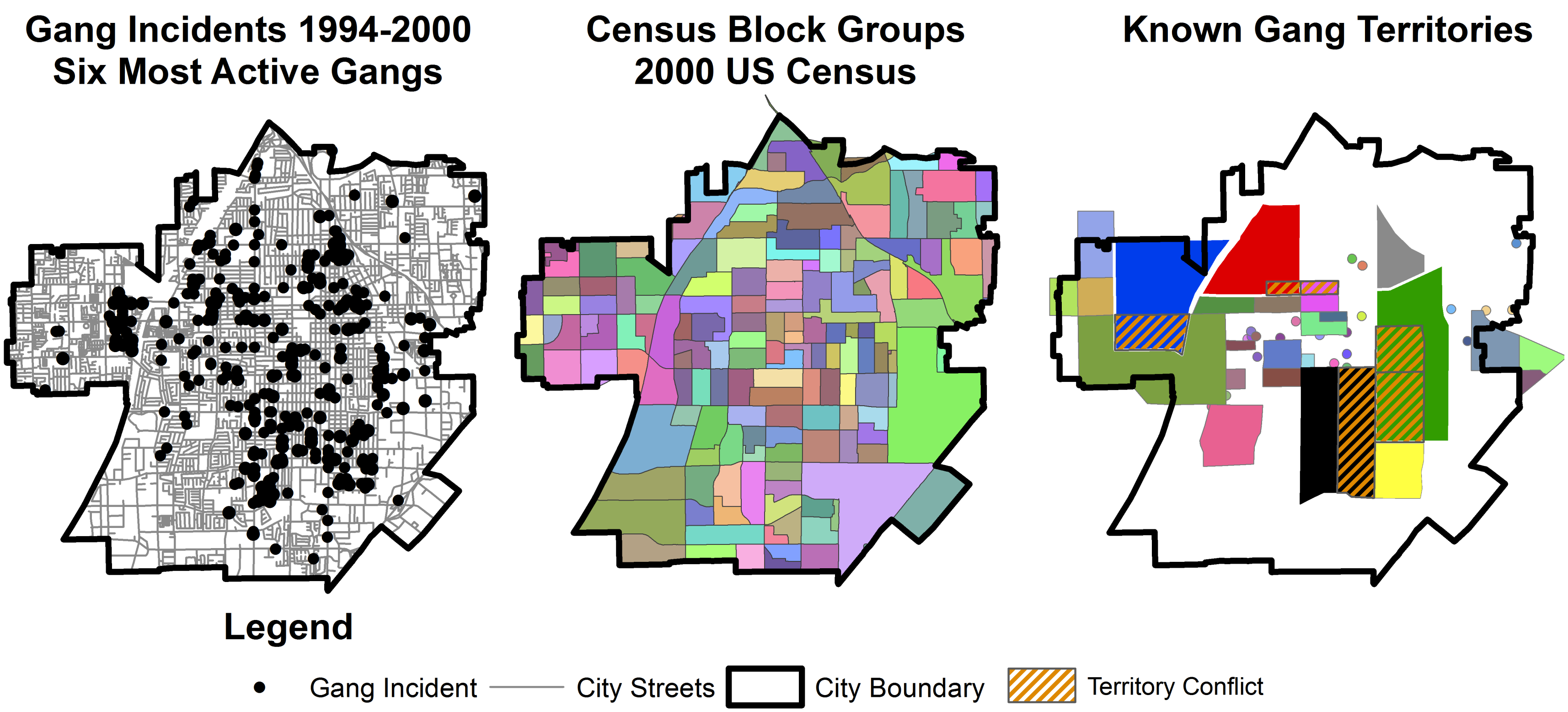


Figure 6.1: Data sources used for the analysis of gang spheres of influence. The left panel illustrates the locations of gang crimes by the six most active gangs in the city during the period 1994 - 2000. The center panel illustrates the 180 census block groups from the 2000 US Census containing census demographic and socio-economic data. The panel on the right is a digitized form of a gang intelligence map provided by the City of Santa Ana Police Department. The gang territories in the right panel overlap the boundaries of the city, but are illustrated as even gang territories outside the boundaries of the city serve as predictive features for activity within the boundaries of Santa Ana.

1. A suspect or suspects are identified as gang members or admit membership in a gang

2. A person becomes a victim due to his or her gang association

3. A reliable informant identifies an incident as gang activity

4. An informant of previously untested reliability identifies an incident as gang activity, and this identification is corroborated by other independent information

More than 75% of the incidents in the data set have no identified gang affiliation for the perpetrator. While there are more than 132 unique gangs identified in the database, the vast majority of them are attributed very few crimes. For instance, there are 59 gangs in the dataset who are attributed only one crime over the eight year period and 97 of the gangs commit an average of less than one crime a year.

Results are presented in the next section for the analysis of three different data sets. First, the Sphere of Influence (SOI) analysis is demonstrated on an analysis of the six largest and most active gangs in the city. These six gangs account for 33% of the crimes in the data set for which the perpetrator's gang was identified. All six of these gangs have large gang territories mapped on the police intelligence map discussed below. The second data set extends the first SOI analysis to mapping the spheres of influence for the 15 most active gangs in the city. Incidents by these 15 gangs comprise 58% of the crimes in the data base for which the perpetrator's gang is known and all of these gangs have either a gang territory or gang point location (address) identified for them on the police gang intelligence map discussed below. The last SOI analysis discussed uses all of the incidents in the dataset.

The US 2000 Census provided the second data source, used for demographic information. Demographic information from the 2000 census was represented as an irregular surface with discrete demographic values recorded at the census block group level. Each criminal incident that fell into a given census block was given the socio-demographic information of the census block group it fell within. Socio-economic and demographic features found to be relevant included: median income, property, and rental values; racial demographics; the percentage of males in the population; the percentage of the population on public assistance; and the percentage of residents who own their homes.

The authors also obtained a gang intelligence map from the Santa Ana Police Department, which was dated in the year 1998. The gang intelligence map detailed the claimed gang territories of the city's largest gangs and point locations (addresses) for many of the smaller gangs in the city. They also showed regions of the city claimed as gang territory by more than one gang - areas we termed "conflict territories."

The data available presents some constraints and limitations on the analysis. First, as seen in Table 6.1, the crime counts for the property, drug, weapon, and vandalism crime counts are very limited, with null observations for many crime type and gang combinations. In a study of gang street crime in Chicago, (Block and Block, 1993) found that the spatial distribution of drug crimes and gang turf-motivated violent crimes differed. It may be true that the spatial pattern of the various crime types differ in Santa Ana, but the limited number of crimes by type, especially in the test data set, prevent this consideration. Due to the limited number of observations, crime types were pooled together by gang.

The second significant limitation was that the only gang intelligence product still available for this time period was a gang intelligence map for the city of Santa Ana from the year 1998. Thus, we made the assumption that the gang territories remained static throughout the study period (1994-2000). Some authors have noted that gang boundaries in other cities tend to shift relatively frequently (Block, 2000), so this assumption may not be valid but was necessary for the purposes of this study. More accurate intelligence products that incorporate shifting gang boundaries (if neces-

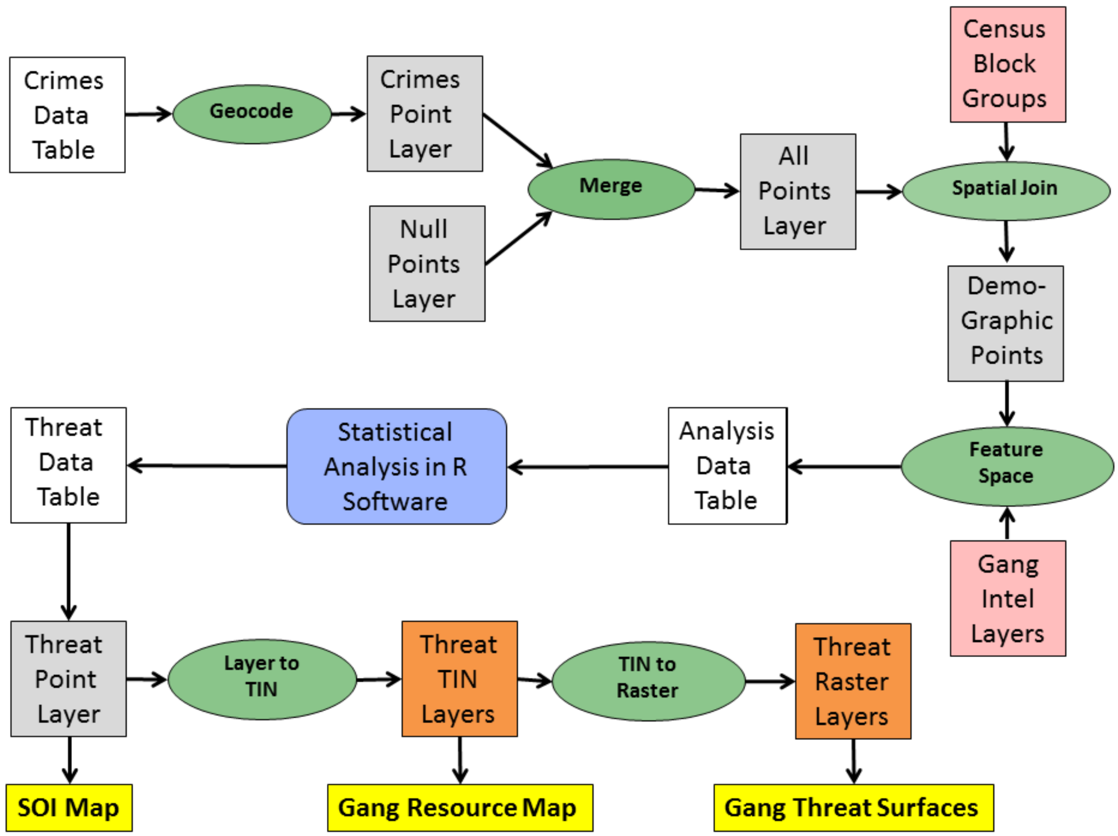


Figure 6.2: Workflow for producing a gang criminal SOI analysis.

sary) should improve the accuracy of this approach in police applications because the statistical models used map the relationship between the probability of crime by a particular gang and distance to their gang territory or known address. Therefore, the models demonstrated here can adapt the model prediction (and the predicted sphere of influence for a criminal gang) if gang territory boundaries change over time.

6.3 Methodology

Developing the Sphere of Influence (SOI) analysis for criminal gangs consists of three steps: data-set preparation, statistical modeling of gang criminal preference, and communication of results in a GIS system. Figure 6.2 provides a work-flow diagram for completing this analysis.

Developing the dataset for the statistical model is not possible without the use of a GIS system to conduct data preparation. During this step, inputs from many different data sources are fused together into a single data table that is exported into a statistical software package. The final product of the data preparation stage is a GIS point layer containing descriptive information about both the locations where the crime occurred and a "null grid."

In order to provide representation for all of the locations *not* chosen by the criminal gangs, we incorporated a null grid by laying a point grid spaced at 200 feet over the study area. This use of a null grid converts the irregular census block data surface into a regular lattice of point sites with discrete variables (Besag, 1974). This approach allows us to fit the regression model by providing null occurrence observations in geographic space in very rough approximation to the proportion of the surface area of the city that did not observe criminal incidents. It also allows us to develop the predicted continuous threat surfaces illustrated in Figure 6.5 by mapping the predictions over the null grid surface.

Additional GIS data preparation processes include: geocoding the data set of criminal incidences, developing a point grid layer to represent locations where a crime did not occur as previously described, merging these two point layers together, conducting a spatial join to give the point layer socio-demographic features from census blocks, and calculating the Euclidean distance from each individual point to important spatial features (sometimes referred to as feature space modeling) in the environment such as gang territories or addresses identified in police intelligence products. The final product of this step is a data table produced from a GIS point layer that contains the following information for each point: binary incident marker (0=null, 1=incident); Census 2000 socio-economic and demographic data; and the Euclidean distance to the nearest: gang territory, gang territory boundary, known gang address, and gang territory conflict. In the case that a crime occurred in one of the geographic regions such as a gang territory, then the feature-space distance to that predictor was recorded as 0.

Once the analysis data table is built in a GIS, the results are exported into the statistical software package R for statistical modeling. Using a multilevel generalized linear model, we examine not only the relationships between the features of interest and the incidence of crime, but we also identify how those relationships vary across the modeled groups. Thus, the multilevel modeling approach allows one to identify the generalities of gang criminal behavior while still modeling the uniqueness of how each gang interacts with the environment. Capturing this trade-off between what is generally true of all gangs and what is unique to each gang significantly improves our ability to predict the criminal behavior of each individual gang in the future. This stage of the analysis includes developing models of the criminal site selection behavior of the various criminal gangs including: feature selection, performance assessment of the various models (on both training and test data sets), and using the multilevel modeling to develop predictions for future behavior by the criminal gangs. These statistical models answer one of the research questions posed in the introduction: What makes one location more likely to experience a gang incident than another location? The final output of the statistical modeling stage is a "threat data table" which records the risk of criminal behavior by each criminal gang at every point in the original data table. This threat data table is then exported back into a GIS system and appended to the original data point layer.

The last stage consists of using a GIS system to develop products for use by law-enforcement agencies. First, the threat data table can be used to plot "threat surfaces" for each of the criminal gang in the GIS system (see Figure 6.5). These threat surfaces show areas where each criminal gang presents the greatest threat and are analogous to the "hot-spot maps" commonly developed using kernel density estimation. The point layer can also be used to map the regions where each of the criminal gangs presents the greatest threat (see Figure 6.5), a product we term the "sphere of influence" map. This product illustrates where each criminal gang is the most likely to commit a crime. Finally, the threat surfaces can be used to develop products useful for allocating police resources to specific high threat areas (see Figure 6.6). This product shows the highest probability areas in the city to observe a crime (for example the highest 5 percent risk areas) and the criminal gang most likely to commit a crime there.

6.3.1 Multilevel Modeling of Gang Criminal Site Selection

The following notation captures the results of the data preparation stage:

N = the number of observations in the data set

J = the number of groups represented in the data set

K = the number of predictors

X = the predictor matrix

Y = the response vector

$$y_i = \left\{ \begin{array}{c} \text{1 if an incident occurred at the location} \\ \text{0 if an incident did not occur at the location} \end{array} \right\}$$

for $i = 1,...,N$

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NK} \end{bmatrix}$$

This takes into account that criminal actors can select more than one location for their crimes. Thus, each actor can make several selections out of the set of available locations, each of which will be coded with the dummy variable 1, while all locations not selected by that actor would be coded with a 0. Note that in this application, we are modeling each criminal gang as an "actor." Thus, all individuals within the gang are assumed to have the same criminal preference set and the same choice set. These criminal preferences and choice sets are assumed to vary between gangs.

As documented by Brown et al. (2004), generalized linear models using a logit link function can account for the feature space distances to key features as well as categorical variables. Applying a logistic regression to model the criminal preferences of the studied group allows us to incorporate all different types of available data discussed above: feature-space data, categorical data (such as presence in a gang territory), and socio-demographic information from the census. This methodological approach is also beneficial in that it allows us to incorporate the idea of a criminal's *journey to crime*, a theory that assumes that the likelihood of an offender's target selection decreases with the distance to the target from his home (Bernasco and Nieuwbeerta, 2005; Rengert, 2004). In this case, the perpetrator's "home" is represented by the gang's home territory or known point site (address).

A logistic regression also provides a closed form solution to modeling the criminal preference by returning a value between 0 and 1 indicating the conditional likelihood of an event occurring at a given location. The conditional likelihood expresses the following idea: *given that a criminal event has occurred*, the probability that it occurred at this location is [a number between 0 and 1]. Although the geographic spatial choice methodology developed by Xue and Brown (2003) can incorporate temporal considerations, we have ignored temporal variations and constraints as is often done in criminological research (Bernasco and Block, 2009; Ratcliffe, 2006). Fox and Brown (2012) provide a methodology for incorporating temporal considerations using a multilevel modeling approach.

6.3.1.1 The (No-Pooling) Group Specific Criminal Site Selection Model

First, as an illustration of how spatial choice models have previously been used to predict attacks by specific criminal groups, we define the group specific spatial choice model. We therefore define $\P_j(y_i = 1)$ as the conditional probability of a criminal incident by criminal gang j. Thus, we get J models of the form:

$$P_j(y_i = 1|X) = logit^{-1}(\alpha_{j[i]} + \beta_{j[i]}X_i) \tag{6.3}$$

The notation $j[i]$ indicates that the observations for each individual group j are indexed during model fit by the known group for the actual event observation i. In other words, when the model is fit using actual incidents, model $P_j(y_i = 1|X)$ is developed using only incidents identified in the training data set as having been committed by group j. In this manner, the no-pooling model fits the group specific criminal preference. This model represents the method currently used to build criminal preference models for specific groups in military and police crime analysis software that employ the spatial choice method. This model approach often fails to provide good performance when there is a small number of observations for each criminal group. Accordingly, there is often not enough information to build a good predictive model for a criminal group based only upon the number of incidents they are known to have committed. Statistically derived models often benefit greatly from the addition of observations.

6.3.1.2 The Multilevel Criminal Site Selection Model

The multilevel model is of the same logistic regression form as the above model. The key difference in the multilevel model is that an additional constraint is placed on the coefficients for each of the different groups. We require that the coefficients for each feature of interest across the modeled criminal groups come from a common distribution that is estimated at the time of model fit. This defines the two levels of the multilevel model. We model both the relationship between the features of interest and the incidence of crime and how those relationships vary across the modeled groups. This two part structure can be seen in the multilevel model below:

$$P_j(y_{[i]} = 1|X) = N\left(logit^{-1}\left[\alpha_{j[i]} + \beta_{j[i]}X_i\right], \sigma_y^2\right) \tag{6.4a}$$

$$\begin{pmatrix}\alpha_j \\ \beta_j\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_\alpha \\ \mu_\beta\end{pmatrix}, \begin{pmatrix}\sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2\end{pmatrix}\right) \tag{6.4b}$$

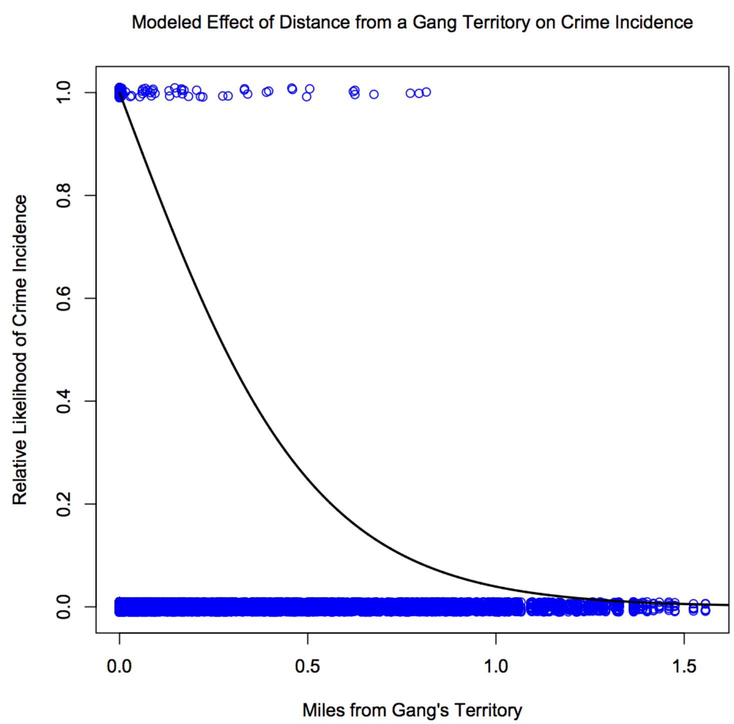
Thus, the probability of an event by group j at location i, $P_j(y_i = 1)$, is Gaussian distributed with a mean determined via the logistic regression and variance σ_y^2. Additionally, we have a model for how the different gangs interact with each feature of interest, with the assumption that there are differences in how the various gangs interact with each feature of interest and that those differences can be modeled with a Gaussian distribution. As an example, every gang will have an intercept term in the logistic regression equation which we model as the term α. We model that all α_j come from a Gaussian distribution with estimated mean μ_α, variance σ_α^2, and covariance parameters to the other features of interest. These coefficients that vary by group are sometimes referred to as *random effects*, referring to the randomness in the probability model for the group level coefficients (Gelman and Hill, 2007).

The requirement for each coefficient for the groups to come from a common distribution generates a tradeoff between the two extremes of the general model and group-specific model. It incorporates what is known about the general criminal preferences but is weighted towards the group specific model in rough proportion to each groups' contribution to the general model. This approach allows us to identify the generalities of gang behavior while still modeling the uniqueness of each gang.

6.3.2 Model Fitting and Feature Selection

To fit the models for this analysis, we used a popular and freely available statistical package in the statistical software R (Cahill and Mulligan, 2007), which provides a computing package for fitting multilevel models based on the algorithm developed by Bates and Pinheiro (1998). Detailed information on algorithms for fitting the multilevel models using a maximum likelihood approach can be found in Huddleston and Brown (2009), Jiang (2007), and Bates and Pinheiro (1998). We also unsuccessfully sought an automated process for feature selection. In fitting traditional regression models, stepwise regression is a popular choice for automated selection of features from a set of possible predictors (Hastie et al., 2001). During each "step", the algorithm adds or drops one predictor and calculates both the statistical significance of all the predictors and a performance statistic such as model deviance or Akaike Information Criterion (AIC) (Akaike, 1974). The subset of features that provides the best performance for a given performance statistic is the subset selected for use. Feature selection for a multilevel model is more difficult because we are simultaneously modeling several (or many) groups at a time. A predictive feature which is important for the performance of one group may be unimportant for others, and insignificant in the pooled model from which the multilevel models are extended. The solution we used to address this problem was to first conduct a step-wise regression of all available features for each of the gangs. Then, we iteratively conducted stepwise regression on the entire subset of features to identify the model that provided the best performance on the training data set of all analyzed gangs (holding out the test data set for later performance evaluation of the selected model). The model fitting package provided for R did not support step-wise regression of a multi-level model and the inability to conduct automated feature selection of the multilevel model remains a hindrance in the automated application of this approach to law-enforcement software

For each of the features selected, we then have a mathematically defined relation-



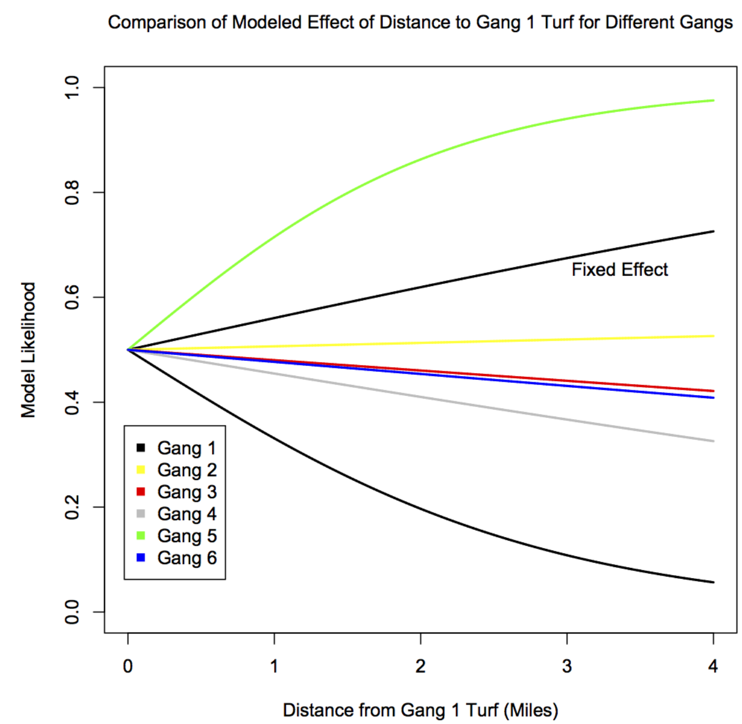


Figure 6.3: The figure in the left panel illustrates the modeled effect of gang territories on the incidence of gang crime by any of the groups. The plotted points indicate the locations where a gang crime did (the top of the graphic) or did not (the bottom of the graphic) occur. The solid line graphs the modeled change in probability of observing a crime at a given location as distance to a gang territory increases. As can be seen in the figure, there are thousands of locations where a crime did not occur. The exponential drop in probability (which models the effect of gang territories after all other factor have been considered) mathematically models the fact that the vast majority of crimes occur within 1/2 mile of a gang territory. The figure in the right panel illustrates how the log-odds change for the incidence of crime by specific criminal gangs with respect to distance from the gang territory of Gang 1 after all other predictive features (including gang territories in general) have been considered.

ship for how that feature is related to the incidence of crime. For example, Figure 6.3 illustrates the effect that gang territories have on the incidence of gang crime. This relationship is modeled as an exponential drop in probability as you move away from gang territories and confirms previous research on the effect of gang territories on the incidence of crime (Ratcliffe and Taniguchi, 2008; Tita and Ridgeway, 2007; Tita and Cohen, 2004). The advantage of the multilevel model is that we can also observe the differences in how various gang respond to individual features. The right panel of Figure 6.3 illustrates the differing effects of Gang 1's territory on the incidence of crime by various criminal groups. As one would expect, Gang 1's territory is correlated with incidences of crime by that group, after all other factors have been considered. The overall effect of Gang 1's territory on the incidence of crime is minimal (this is the "fixed effect"). The multilevel model does identify that increasing distance from Gang 1's territory increases the log-odds for a crime by Gang 5.

The relationships that the multilevel model identifies in the fixed effect and in the incidence of crime by Gang 5 with respect to Gang 1's territory merits additional discussion. These two territories overlap and therefore there is some co-linearity between these two predictive features. Likewise, each individual gang's territory has some co-linearity with the predictive feature "Distance from Gang Territory." The most important predictive feature for each gang in the multilevel model is the distance to its own gang territory, so each gang's territory must appear in the model to preserve model predictive performance for that gang. As a result, we must be very careful not to draw general conclusions about the effects of the individual gang territories predictors, especially those with known co-linearity to other territories. Because we are concerned primarily with predictive performance, and predictive performance suffers greatly with the removal of these features, we have continued to simultaneously use the feature-space distances to *each* of the considered gang territories and the feature-space distance to the *nearest* gang territory as a general predictive variable in our analysis. This use facilitates predictive performance but limits our ability to draw strong conclusions about the significance and importance of the various features. Because of the intended application for building predictive software, this is acceptable. It would not be acceptable for a sociological examination of features important to the incidence of crime.

6.3.3 Kernel Density Mapping of Gang Criminal Site Selection

As previously noted, kernel density estimation is the most frequently used approach for developing predictive threat surfaces in police applications. The kernel density approach is briefly presented here to provide a point of reference for performance assessment of the criminal site selection methods presented in the previous section. The kernel density method estimates the crime intensity at each location for gang j. Let s_{jn} denote the location of the nth crime by group j and Nj the total number of crimes in the training data set attributed to gang j. The crime intensity for gang j at location i, λ_{ij}, is calculated using a kernel smoothing function.

$$\lambda_{ij} = \hat{f}_{jh}(y_i) = \frac{1}{N_j h} \sum_{n=1}^{N_j} K\left(\frac{\|y_i - s_{jn}\|}{h}\right) \tag{6.5}$$

In Equation 6.5, the notation $\|y_i - s_{jn}\|$ denotes the Euclidean norm (distance) between locations y_i and s_{jk}. Model fitting requires the selection of the kernel function K and the bandwidth parameter h. In this application, we used the quartic kernel function and selected the optimal bandwidth for each gang using maximum likelihood estimation on the training data set as implemented in R software by the `splancs` package (Rowlingson et al., 2012; Bivand et al., 2008).

6.4 Results

Once the models have been fit, we build predictive threat surfaces for each of the modeled groups by predicting the conditional probability for every point on the null grid. The null grid, coded with the conditional probabilities at each grid point, is then exported to a GIS system as a point shapefile. This point layer is converted into a raster threat surface by interpolating between points. The final product of the predictive models is a threat surface created in a GIS system such as those in Figure 6.5. The ability of the multilevel model to provide distinct threat surfaces for each of the gangs is evident in Figure 6.5. Note that many of the gangs share high-probability areas. Note also that many of the high-threat areas for these gangs lie outside of their claimed gang territories. These threat surfaces address the most likely courses of action for each of the studied criminal groups, facilitating targeting, interdiction, and observation of their members.

A threat surface produced by a predictive algorithm can be thought of as a binary prediction of the probability of criminal incident at each individual location. Thus, threat surfaces can be evaluated using methods developed to assess the performance of binary classifiers. Recent years have seen the Receiver Operating Characteristic (ROC) curve become particularly popular for evaluating and comparing predictive algorithms in the machine learning community (Fawcett, 2006, 2004; Spackman, 1989). The ROC curve plots the cost-benefit trade-off for a classifier at all possible classification thresholds (Fawcett, 2004). The cost, plotted on the horizontal axis, is the model's false positive rate. The benefit, plotted on the vertical axis, is the model's true positive rate (also called model sensitivity or hit rate).

ROC curves are a two-dimensional representation of classifier performance and often researchers would like to reduce performance to a single (scalar) statistic. The most common approach for summarizing ROC performance is to calculate the area under the ROC curve (denoted AUC) as a scalar value representing model classifica-

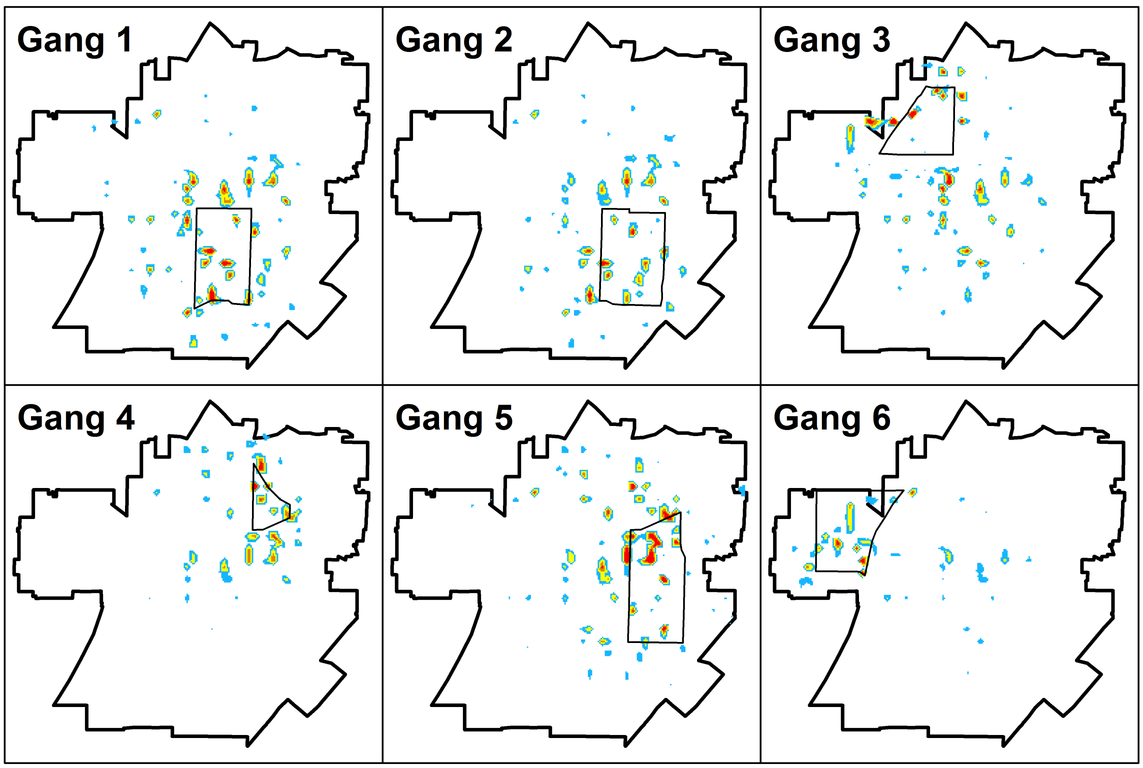


Figure 6.4: Threat surfaces for each criminal gang illustrated with the corresponding gang territory.

tion performance (Fawcett, 2006; Hanley and McNeil, 1982). This statistic represents the probability that a randomly chosen positive incidence (in this application a randomly chosen location where a crime has occurred) will score higher than a randomly selected negative instance (i.e. a randomly selected location from the “null grid” created to represent non-incidents) (Fawcett, 2006). The AUC is equivalent to the Wilcoxon test of ranks commonly used in categorical data analysis (Fawcett, 2006; Hanley and McNeil, 1982) and is also directly related to the Gini coefficient (Breiman et al., 1984). The AUC is often considered to be the standard method to assess the accuracy of binomial classifiers. Table 6.2 provides a performance comparison of the predictive performance of the three previously discussed modeling approaches on the test data set using the Area Under the Curve (AUC) statistic derived from the Receiver Operating Characteristic (ROC) curve.

As can be seen in Table 6.2, the multilevel modeling approach provides the most accurate predictive threat surfaces. AUC scores of 0.5 provide no discriminatory value. An AUC score above 0.75 is considered to provide enough discriminatory power to be clinically useful in the medical community, and AUC scores above 0.97 are considered to provide excellent discriminatory power (Fan et al., 2006). The kernel density method provides moderately useful discriminatory power and this performance combined with its ease of use explains its wide-spread use in policing applications. Note that both criminal site selection models significantly improve on the often-used kernel density approach, confirming similar results obtained in other studies (Liu and Brown, 2004; Brown et al., 2004; Xue and Brown, 2006; Smith and Brown, 2007). The performance improvement between the no-pooling and multilevel modeling approach is primarily due to the ability of the multilevel model to leverage the additional information provided by considering the preferences of the other gangs and ability of the model to find the optimal tradeoff point between individual gang models and a model that pools all gang information. The performance improvement

Gang	Multi-Level Model	No-Pooling Model	Kernel Density Model
1	0.97	0.97	0.82
2	0.98	0.98	0.76
3	0.99	0.96	0.88
4	0.99	0.94	0.78
5	0.98	0.97	0.72
6	0.98	0.92	0.85

Table 6.2: Predictive performance summarized using the Area Under the Curve (AUC) statistic.

between the no-pooling model and the multilevel model is relatively small, but the multilevel modeling provides an additional benefit: the ability to conduct a Sphere of Influence (SOI) analysis.

6.5 Sphere of Criminal Influence

The Encarta World English Dictionary (Soukhanov, 1999) defines a sphere of influence as "a region of dominance; a geographic region or area of activity in which a state, organization, or person is dominant." We define the predicted *criminal sphere of influence* for a group as the geographical area where a given model predicts that each group is the most likely to commit a crime (i.e. the geographic regions where one gang's threat surface is higher than all other gang's threat surfaces). This sphere of influence is created by mapping the calculated $P_j(y_{[i]} = 1)$ for all j groups and all grid points, y_i. By comparing each group level threat surface, we determine where each of the J groups is dominant. Figure 6.5 illustrates the results when sphere of influence maps are built with the three modeling approaches discussed in this paper.

As can be seen in Figure 6.5, neither the no-pooling or kernel density models provide coherent sphere of influence maps. The issue with both of these approaches is that the probabilities (or intensities in the kernel density approach) have been calculated only considering the effect of one gang at a time (i.e. there are six different models built in each case). When the threat surfaces for the different gangs are com-

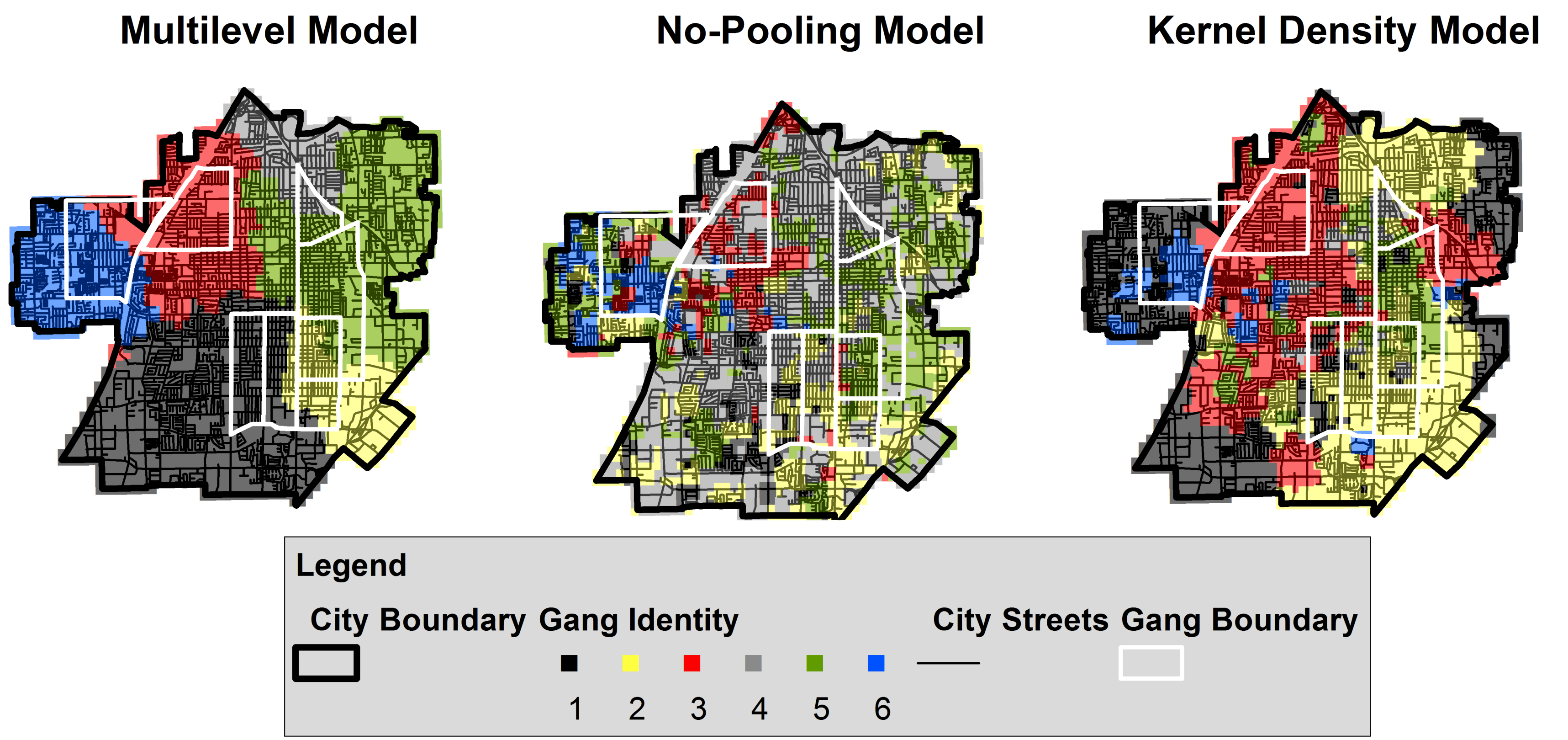


Figure 6.5: Gang Sphere of Influence (SOI) maps comparing the spheres of influence for an analysis of the six largest gangs in Santa Ana using the multilevel model (left panel), group specific (no-pooling) modeling approach (center panel) and kernel density estimation (right panel).

pared with these models, the resulting sphere of influence maps are non-contiguous representations due to the fact that the different rates of criminal activity were not considered in the developed models. In contrast, the multilevel modeling approach simultaneously models all gangs. The resulting sphere of influence map plots a contiguous sphere of influence for each of the six gangs centered around their claimed gang territories. In some cases, a dominant gang's sphere of influence encroaches upon a neighboring gang's territory and all gang "territory conflicts" have been divided into competing spheres of influence.

Table 6.3 displays the results obtained when we investigate the veracity of the gang SOI analysis by examining the predictive performance of the SOI analysis against a test set. Table 6.3 records the percentage of incidents during the test period (1999-2000) which were committed in each predicted criminal sphere of influence by the various gangs. As can be seen in the table, the multilevel model accurately predicts the gang most likely to commit a crime for each sphere of influence.

Table 6.4 expands this analysis by comparing the performance of the sphere of influence analysis conducted with the three models. While the gang spheres of influence from all three approaches often correctly identify a geographic area in which the predicted group is the most likely to commit a crime, the performance of the multilevel model is much better. The spheres of influence predicted by the multilevel model contain a higher percentage of incidents committed by the predicted group in every case and provides significantly better overall performance, providing a more accurate prediction for where each gang is dominant.

The no-pooling model and KDE models perform as well as they do on this analysis only because most of the crimes by these gangs fall into areas with high predictions for the respective gangs (i.e. both models' predictive performance is pretty good). These areas represent *hot-spots* for each group and the peak risk areas for each gang

Sphere of Influence	Incidents Committed by Each Gang in Indicated SOI					
	Gang 1	Gang 2	Gang 3	Gang 4	Gang 5	Gang 6
Gang 1 SOI	**61**%	19%	3%	3%	10%	3%
Gang 2 SOI	7%	**80**%	0%	0%	7%	7%
Gang 3 SOI	0%	21%	**50**%	14%	0%	14%
Gang 4 SOI	0%	0%	0%	**78**%	11%	11%
Gang 5 SOI	9%	23%	5%	9%	**55**%	0%
Gang 6 SOI	0%	0%	0%	0%	0%	**100**%

Table 6.3: Multilevel Model SOI predictive performance.

Sphere of Influence	Multi-Level Model	No-Pooling Model	Kernel Density Model
Gang 1	61%	48%	56%
Gang 2	80%	78%	55%
Gang 3	50%	25%	35%
Gang 4	78%	58%	50%
Gang 5	55%	45%	53%
Gang 6	100%	85%	94%
Overall Performance	69%	51%	57%

Table 6.4: Percentage of test incidents committed by each group in predicted SOI.

are still accurately mapped. The multilevel model better sorts out who presents the greater threat in the regions that aren't hot-spots for the different gangs.

The predicted criminal sphere of influence provides important insight for an analyst that is not available to analysts who currently only have access to incident hot spots and known gang areas. This is because criminal gangs often commit crimes outside of their known territories or in geographic areas contested between gangs. For example, in the two year test data set period (1999-2000) of this study, nineteen percent of the crimes attributed to the six most active gangs were committed in areas claimed by more than one gang, which we term "territory conflicts." Thirty percent of the crimes committed by these six gangs occur outside any of their known territories. The sphere of influence prediction maps which of the groups is most likely to commit

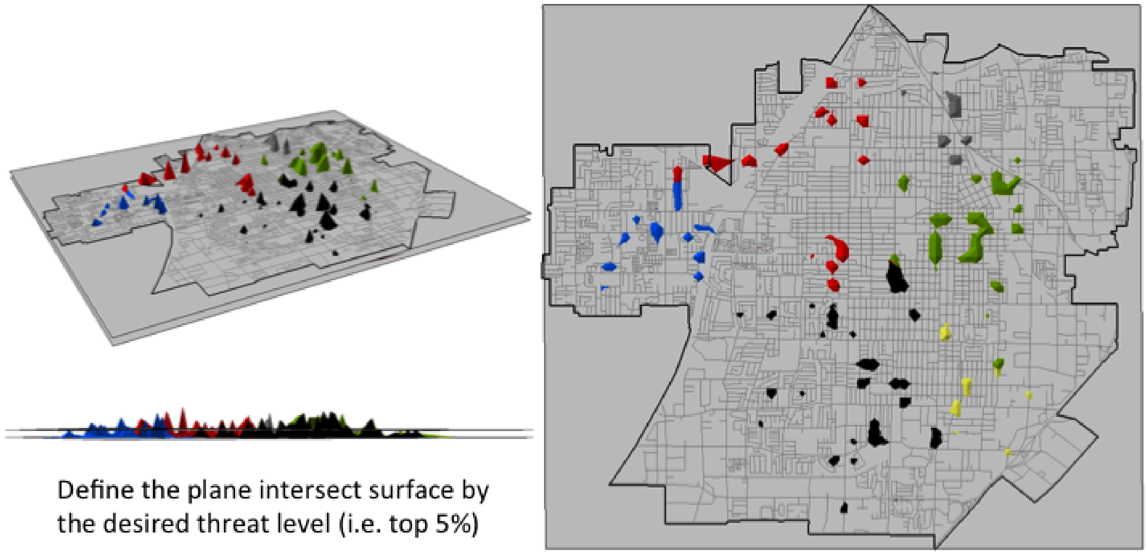


Figure 6.6: Gang Resource Map showing the highest probability areas for gang activity in the city and the gang most likely to commit a crime at that location.

a crime in these geographic areas which lay away from the areas other intelligence assets indicate is their home territory. Figure 6.6 provides an illustration of how this information can be leveraged. It provides a "gang resource map" that illustrates the highest five percent risk area for gang activity in the city and identifies the gang most likely to commit a crime in those areas. This map can be developed either by directly identifying the highest probability risk areas and responsible party in the statistical package and exporting that information in the threat data table or by leveraging the three-dimensional modeling capabilities of GIS systems that use the "heights" of the threat surfaces and an intersecting plane to produce this reference product (the authors developed this illustration using the ArcGIS ArcScene software).

6.6 Extending the Analysis to More Gangs

We then extended the SOI analysis for Santa Ana to consider more than the six largest and most active gangs. First, we extended the analysis of the gangs in Santa Ana to include all incidents from the 15 most active gangs in Santa Ana. These 15

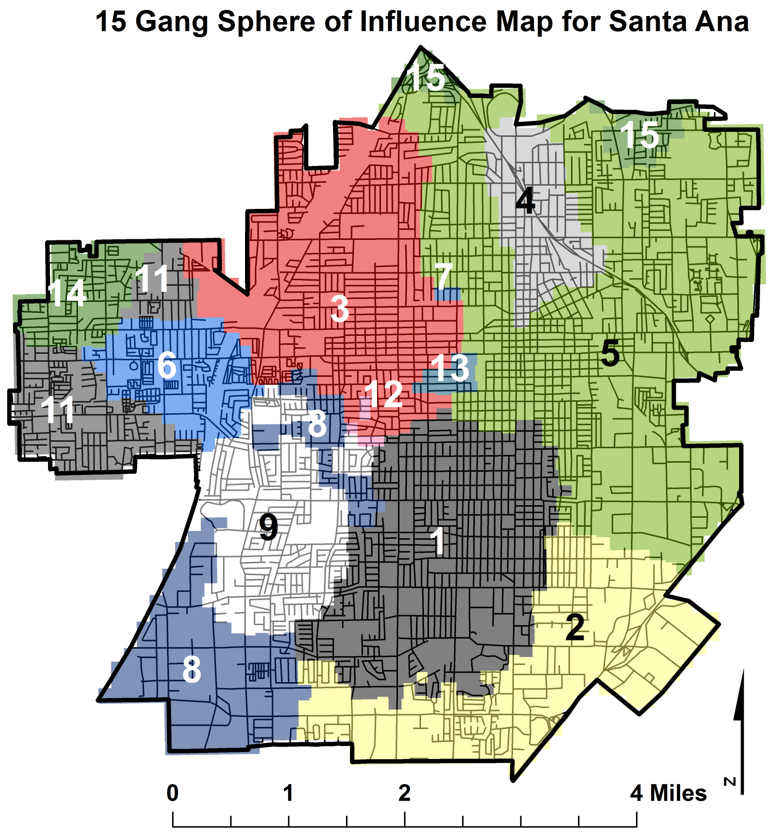


Figure 6.7: Sphere of Influence Map for an SOI analysis of the 15 most criminally active gangs in Santa Ana.

gangs commit 58% of the crimes in the data base for which the gang affiliation of the perpetrator was known. The sphere of influence continues to be a very accurate predictor of where each gang is the most likely offender for a gang related crime incident, even as more gangs are added to the analysis. Extending this analysis is important because, while the analysis of the six largest gangs serves to illustrate the effectiveness of the multilevel model over the other approaches, to accurately assess the situation in a given geographic region we would have to consider all of the criminal elements.

The 15 gang sphere of influence map in Figure 6.7 identifies where in the city each of the 15 most active gangs presents the dominant threat. Although there are more gangs active within the city of Santa Ana, the remaining gangs all commit less than five crimes per year, and thus have comparatively little influence. The map above serves as a reference document for a law enforcement analyst, identifying the dominant threats in each area of the city. There are some edge effects visible when comparing the change from the six gang SOI to the fifteen gang SOI. Work continues on how to reduce the edge effects near the boundaries of the city that are most easily visible in the changes to gang boundaries across the southern border of the city. Note that Gang 10 has no visible sphere of influence in Figure 6.7. This is because it is dominated by larger gangs nearby.

Table 6.5 records the predictive accuracy of the 15 gang sphere of influence analysis. The multilevel model sphere of influence accurately predicts the gang most likely to commit a crime in the given sphere of influence for 10 out of 12 spheres of influence in which a crime occurred during the test period. In Gang 9's predicted sphere of influence both Gangs 8 and 9 had an equal number of crimes. In Gang 8's sphere of influence, Gang 10 (which has no predicted sphere of influence) committed the majority (three out of four) of the crimes. Given the limited number of crimes in these two smaller spheres of influence, it is difficult to determine whether the SOI analysis did poorly for these two regions due to changing conditions during the test period (1999 - 2000) or because it simply did not model the SOI for these two gangs well. Overall, the SOI analysis appears to provide a fairly accurate map of the regions of dominance for these criminal gangs.

The last analysis conducted used all incidents in the data set, including the incidents in which the offending gang was not identified. These incidents were classified as belonging to the "unknown" gang. This analysis revealed several important insights into conducting an SOI analysis. First, the sphere of influence for the "unknown gang," which is attributed more than 75% of the crimes in the data set, extends to cover the entire city. Thus, the sphere of influence analysis developed from this scenario is uninformative. Second, the predictive performance for the known gangs already modeled suffered significantly when the “unknown gang” incidents were included in the training data set. This almost certainly happens because the model structure assumes that all members of a gang operate from a similar preference set and that the model coefficients for all gangs come from a common distribution (see Equation 6.4). Because of the large amount of incidents attributed to the “unknown” gang, this gang’s incidents provide most of the information for the model fitting, resulting in a model skewed towards providing a good fit for a gang that does not exist. This insight generates the recommendation that when developing predictive models or a sphere of influence analysis using multilevel models, the appropriate dataset should contain only incidents in which the discrete choices (for example crimes) can be attributed to individual groups (or individual actors).

The second insight developed from analyzing the entire data set concerns the need for location data and sufficient sample sizes for the modeled criminal groups. The gang addresses or territories for most of the “other gangs” weren’t known and most of these gangs commit very few crimes. These gangs did not have enough observations during the testing period to conduct performance assessment of the generated threat surfaces. They also have no mapped sphere of influence because their threat surfaces were dominated by at least one of the 15 most active gangs in the city. Note that multilevel modeling can be conducted even in the case in which there are as few as one or two observations for individual groups so long as there is a sufficient number of groups but this exponentially increases the computational burden for fitting the models (Gelman and Hill, 2007). In this case, including the remaining gangs into the analysis generated no more insights for a police agency. Thus, we recommend the SOI analysis be limited to active gangs for which specific location data (addresses or mapped gang territories) is available.

6.7 Conclusions

Although this new approach does provide a way of automatically developing a threat assessment product for use by law enforcement or intelligence agencies, there are some important caveats to the use of this methodological approach. The first issue concerns identification of features of interest for use in the multilevel models. The approach relies on manual step-wise regression because the statistical package used did not provide a convenient way to automate this. This would be a key feature needed to provide maximum benefit in the application to crime analysis software. It took a great deal of time to find appropriate models for increasing numbers of criminal gangs. This process was very iterative and required some understanding of statistical significance and the ability to fit and interpret mathematical models. We were not able to find an approach to easily automate this process, limiting immediate application in law enforcement software. We also did not have access to all of the features that we would have liked to include in the analysis. There were many crime generators and crime attractors (Brantingham and Brantingham, 1981) we did not include in the analysis that could significantly improve predictive performance.

Another concern with the sphere of influence approach relates to a potential use by law enforcement personnel. Given a sphere of influence map, many law enforcement officials are likely to want to use the map to identify the most likely perpetrators of criminal events *after the fact*. However, when including all of the incident data from the GIT database, the "unknown" gang becomes the most dominant sphere of influence throughout the geographic region. This result is not unexpected since about 75% of the incidents in the GIT database are attributed to an unknown party. Since the "unknown" gang commits 75% of the crimes and these crimes are spread throughout the spatial region of study, the "unknown" gang's sphere of influence is modeled as most dangerous throughout the city. This underscores the fact that in spite of the precise boundaries established between the groups by the sphere of influence map, there is still a good deal of uncertainty inherent in the models - the "unknown" gang's effect is not mapped.

The sphere of influence map serves as *a priori* predictors of criminal activity but should not be used as a *posterior* identifier of criminal responsibility. Instead, the sphere of influence map should be a useful tool for resource allocation decisions, assigning areas of responsibility as part of counter-gang initiatives, or prioritizing investigative efforts after the occurrence of a crime. One scenario might be that, given a suspected gang-related event at a specific location, we can identify which of the gangs are the most likely to have perpetrated the crime and allocate investigative resources to the most likely culprit(s). The multi-level model can be used to calculate a confidence interval for the stated conditional probability. Especially in areas near the sphere of influence boundaries, these confidence intervals often overlap for two or more groups. The correct approach would be to iteratively assign investigative resources in a manner consistent with the probabilities proposed by the model. When no other information about an incident is known, start with the most probable culprits and work your way through the list of predicted suspects in decreasing order of the probability of activity at that location.

This methodological approach also has several obvious applications in other domains. The sphere of influence analysis should apply in any domain in which we want to compare some spatial choice behavior (criminal, consumer, etc.) across individuals or groups and map the probability that a group's or individual's target selection behavior will dominate all others at a given geographic point. Future research efforts in this area include identifying improved approaches for automated feature selection, the modeling of temporal choice behavior, and the application of multilevel spatial choice modeling to the spatial choice behavior of insurgent/terror groups, retail customers, and corporate real estate.

SOI	Percentage of Test Data Set Incidents Committed by Each Gang Which Occur in the Selected Sphere of Influence														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	**38.3**%	27.7%	0.0%	2.1%	6.4%	4.3%	4.3%	2.1%	2.1%	4.3%	0.0%	2.1%	0.0%	2.1%	4.3%
2	0.0%	**71.4**%	0.0%	0.0%	14.3%	0.0%	14.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	8.7%	**30.4**%	8.7%	0.0%	8.7%	0.0%	8.7%	0.0%	13.0%	0.0%	0.0%	0.0%	13.0%	8.7%
4	0.0%	0.0%	0.0%	**63.6**%	9.1%	9.1%	9.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	9.1%	0.0%
5	8.5%	12.8%	2.1%	4.3%	**25.5**%	0.0%	21.3%	4.3%	2.1%	2.1%	0.0%	4.3%	4.3%	0.0%	8.5%
6	0.0%	0.0%	0.0%	0.0%	0.0%	**89.5**%	0.0%	0.0%	5.3%	0.0%	0.0%	0.0%	0.0%	5.3%	0.0%
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**25.0**%	0.0%	75.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	25.0%	**25.0**%	16.7%	16.7%	8.3%	0.0%	0.0%	0.0%
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	0.0%	0.0%	0.0%	0.0%	0.0%	28.6%	0.0%	0.0%	0.0%	0.0%	**57.1**%	0.0%	0.0%	14.3%	0.0%
12	0.0%	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**66.7**%	0.0%	0.0%	0.0%
13	0.0%	25.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**75.0**%	0.0%	0.0%
14	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0**%	0.0%
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 6.5: SOI predictive performance for 15 gang analysis.

Chapter 7

Journey to Crime Analysis for Military Cordon and Search

Military forces engaged in counter-insurgency campaigns perform many of the same security and crime suppression activities as domestic police forces. However, they are much more focused on capturing or suppressing groups of offenders than capturing and prosecuting individual serial criminals. These criminal groups, organized into "cells" or small paramilitary elements, cooperate to coordinate attacks on the local population, the government, and security forces in an effort to destabilize the government. Cordon and search operations are one of the most frequently employed techniques military forces use in targeting these criminal groups (U.S. Army, 2009). Military forces perform cordon and search operations by establishing an impermeable outer security perimeter (the cordon) and then systematically searching the target area in an effort to locate enemy combatants or equipment. Effective Journey to Crime (JTC) techniques provide a method for determining high probability search zones for military cordon and search.

Journey to Crime analysis, also called geographic profiling, is an investigative technique employed by police that uses the known locations of a crime series to determine a serial offender's anchor point, usually a residence or workplace. Despite several high-profile successes, JTC methods (Mohler and Short, 2012; Rossmo and Rombouts, 2008; Paulsen et al., 2010; Levine and Block, 2011) have not yet been developed that are demonstrably more accurate than simple centro-graphic techniques such as calculation of the Fermat-Weber point, more commonly known as the Center of Minimum Distance (CMD) (Levine and Block, 2011; Paulsen et al., 2010; Paulsen, 2006b). Effective solutions to this problem are in high demand because a critical function in the criminal investigative process is locating unknown serial offenders (Rossmo and Rombouts, 2008).

This chapter demonstrates the use of two new JTC techniques in support of military cordon and search operations against the anchor points (addresses) for the criminal activities of known criminal groups. These methods leverage Criminal Site Selection (CSS) modeling, a modeling technique previously used to develop hot-spot maps that predict future criminal activity. CSS models have been shown to significantly improve predictive performance over traditional kernel density hot-spot methods for predicting crimes such as burglaries (Liu and Brown, 2004; Xue and Brown, 2006), terrorist events (Brown et al., 2004), suicide bombings (Smith and Brown, 2007), and criminal activity by street gangs (Huddleston, 2008; Huddleston and Brown, 2009; Huddleston et al., 2012). CSS models provide an important extension to existing JTC methods because the model structure can incorporate the effect of many crime generators and attractors (Brantingham and Brantingham, 1981) into the geographic profile (the mapped probability map for the anchor point) while incorporating the distance-decay modeling used by existing JTC methods.

This research provides several important research contributions. First, one of the approaches developed in this chapter is more accurate than the current best method for JTC analysis, calculation of the Fermat-Weber point. Second, the results provided demonstrate that the geographic profiles developed for these criminal groups are often accurate enough to facilitate tactical success, with the modeled criminal group's anchor point falling within the search profile for military unit cordon and search operations. Third, the proposed methodology provides a framework for incorporating environmental effects, crime attractors, and crime generators as predictive features in a JTC model. Finally, whereas previous JTC methods have been focused on individual serial criminals, this chapter documents the first use of JTC methods for identifying the anchor point for criminal groups.

7.1 Background

JTC analysis (geographic profiling) is defined as estimating the anchor point of a criminal offender given a set of observed spatial locations for crimes assumed to have been committed by the individual (Mohler and Short, 2012). It is applied by police to the investigation of crimes such as serial murder, rape, arson, robbery, and bombing (Rossmo, 2000). In police practice, JTC is more often used to focus investigative resources on a subset of an existing list of suspects rather than to identify a point location for the conduct of a search. For example, in the course of investigating a serial rape crime series, a mapped geographic profile might be used to focus investigative resources on the subset of previous sexual offenders granted a high "hit score" based on a JTC model. This might reduce the list of working suspects from hundreds or thousands of offenders (in a mid-size city) to less than a dozen. Police might also use the geographic profile to select an region over which to canvass area residents with a sketch of the suspected perpetrator etc.

The first JTC models were developed in the late 1990s by Dr. Kim Rossmo of the Vancouver Police Department. The first generation of JTC models relied on mathematical scoring methods that use various mathematical functions to estimate the distance-decay relationship between an offender's anchor point and the probability that they commit a crime. Subsequently, other researchers applied known centrographic techniques to this problem and demonstrated that simple centrographic statistics such as CMD often provided better predictive performance than the more complex scoring approaches. Recently, there has been a surge in research on JTC methods resulting in several new modeling approaches that employ Bayesian methods to develop probabilistic estimates for offender anchor points. Several of these new approaches provide predictive performance equivalent to the simple centrographic techniques while providing some additional insight.

While JTC models have been shown to provide some value in the police investigative process, they currently suffer from several shortcomings that make them of limited use for application in military cordon and search operations. First, no JTC method has been shown to improve upon simple centrographic methods such as CMD in terms of predictive accuracy (Levine, 2009a; Snook et al., 2005; Paulsen, 2006b; Levine and Block, 2011). Second, although the need to incorporate environmental effects, crime attractors, and crime generators into JTC models is well-documented (O'Leary, 2009), current JTC methods are very limited in their ability to incorporate these predictive features. Criminal Site Selection (CSS) models provide a way to address these existing shortcomings in JTC models. Details on these modeling methods are provided in the sections below.

7.1.1 Mathematical Scoring Methods

Scoring methods are based upon the idea that the probability of a crime by a given criminal decreases with distance from the criminal's anchor point (i.e., distance-decay). These approaches use a scoring (decay) function, f, and a distance function, d, to identify the likely locations for a serial offender's residence for an index of possible locations i based upon a set of observed crimes by an offender $\{s_1, s_2, ...s_N\}$:

$$L(i) = \sum_{n=1}^{N} f(d(s_n, z)) \tag{7.1}$$

The distance function, $d(s_n, z)$, provides the distance between an observed crime, s_n, and the criminal's anchor point, z.

Researchers have explored many different methods for both distance and scoring functions. Reasonable choices for the distance function include "Euclidean distance, Manhattan distance, the total street distance following the local road network, or the total time to make the trip while following the local road network (O'Leary, 2009)." The first JTC method, Rossmo's Rigel method, uses a Manhattan distance function and a truncated negative exponential distance decay function (Rossmo, 2000; Paulsen, 2006b). This decay function includes a buffer around the anchor point under the assumption that criminals will not commit crimes in very close proximity to their anchor points. The approach developed by Canter et al. (2000) similarly uses a buffer or plateau around the anchor point, employing a truncated negative exponential distance decay function (Paulsen, 2006b). However, Canter et al. (2000) employ a Euclidean rather than Manhattan distance function. Levine (2009a) develops a more general framework for scoring methods known as the Journey-to-Crime (JTC) algorithm, which is implemented in the software CrimeStat. The CrimeStat software program (Levine, 2009c) offers a plethora of scoring (decay) functions including linear, negative exponential, normal (Gaussian), log-normal, and truncated negative exponential functions. The parameters of the distance decay functions are either fixed based upon previous research (for Rossmo's Rigel method) or derived from a training data set (in the case of JTC/CrimeStat). As O'Leary (2009) notes, "Though each of these approaches are distinct, they share the same underlying mathematical structure; they vary only in the choice of decay function and the choice of distance metric."

7.1.2 Centrographic Methods

Centrography is the study of descriptive statistics that describe geographic measures of central tendency. Several centrographic statistics have been proposed for predicting anchor points for serial offenders including: center of minimum distance (CMD), also known as the Fermat-Weber point; median center; mean center; directional mean; triangulated mean; geometric mean; and harmonic mean. Levine (2009a), Snook et al. (2005), and Paulsen (2006b) conducted studies comparing the performance of simple centrographic statistics to the more complex mathematical scoring function approaches. Levine (2009a), the developer of the CrimeStat software program, compared the performance of all scoring methods in the CrimStat software to simple centrographic measures found that, "simple centrographic measures, especially the centre of minimum distance (CMD), were, on average, the most accurate measures of where the offender lived." Snook et al. (2005) found similar results on a different data set. Paulsen (2006b) compared the performance of four different JTC software packages, including Rigel and CrimeStat, to simple centrographic measures and also found that the CMD method consistently outperformed the other methods. The CMD method is therefore considered the "current best method" in terms of accuracy (minimizing the error distance). However, centro-graphic methods have one significant short-coming as compared to the scoring methods in that they provide only a point estimate for the location of the criminal's anchor point. Scoring methods provide a "jeopardy surface" that can be used to prioritize search zones (see Figure 1). A full jeopardy surface gives police an ordered list of high-probability areas to search rather than a single point.

7.1.3 Bayesian Methods

Several research groups have proposed Bayesian methods to extend existing mathematical scoring methods in hopes of improving predictive performance. These models improve distance decay scoring methods by "adding information about where offenders who commit crimes in particular locations tend to be based (Canter, 2009)" or by modeling environmental effects on criminal target selection. Levine (2009b) and Levine and Block (2011) develop a framework for using the Bayesian paradigm to incorporate the geographic distribution of other offenders who committed similar crimes as a prior probability for the anchor point. Levine and Block (2011) notes that this approach, "was more accurate than existing journey-to-crime methods and was as accurate as the center of minimum distance, the current best method."

O'Leary (2009) proposes employing the Bayesian framework for addressing what he asserts to be the most significant shortcoming of existing methods: the inability to "account for geographic features that influence the selection of a crime site and geographic features that influence the potential anchor points of offenders." In the parlance of environmental criminology, these features are known as crime generators and attractors (Brantingham and Brantingham, 1981). O'Leary (2010) provides a demonstration of this approach for a single crime series. However, rather than modeling the effect of crime attractors and generators, he incorporates a kernel density surface of previous criminal offenses as a proxy for these features and provides no model performance assessments or comparisons.

Mohler and Short (2012) incorporate kinetic models of criminal motion into the Bayesian framework for JTC analysis of serial burglars in Los Angeles. In this modeling approach, kinetic models of motion replace the distance-decay function used in the scoring methods. Additionally, they incorporate mapped housing density as a "target attractiveness" predictive feature within the model. In a predictive performance comparison with the Rigel method, they find that performance is similar to the Rigel method for JTC analysis of marauding offenders but offers some performance improvement in providing geographic profiles for commuting offenders. They define marauding offenders as those whose anchor points fall within the smallest circle containing the two most widely separated crimes in the series. Marauding offenders are also assumed to commit crimes isotropically around their anchor point (Canter and Larkin, 1993). Commuting offenders are those that travel (sometimes long distance) to a target area to commit there crimes rather than fanning out from their criminal anchor point.

7.1.4 Criminal Site Selection Models

Criminal Site Selection (CSS) models are used to develop predictive threat surfaces (hot-spot maps) for future criminal activity in a region of interest. They are based upon the work of Daniel McFadden's development of discrete choice theory (McFadden, 1974). In McFadden's formulation, actors, indexed by j, evaluate the utility, U, that they would derive from choosing an alternative based upon the features or attributes of that alternative:

$$U_{ij} = B^T X_{ij} + \epsilon_{ij} \tag{7.2}$$

In the above formulation, X denotes the vector of features or attributes for alternative i. The ϵ term captures the error associated with each pair of actors and alternatives while B records the regression coefficients of the model. McFadden established the theoretical foundation for the use of conditional logistic regression to model choice from a discrete set of alternatives. When actors, indexed by j, are choosing from a discrete set of alternatives, then their probability of selecting alternative i, $P_j(y = i)$, can be modeled using the well-known logistic regression equation:

$$P_j(y = i) = \frac{e^{B^T X_{ij}}}{\sum_{i=1}^{N} e^{B^T X_{ij}}} \tag{7.3}$$

Several research groups have applied this approach in a spatial context for modeling criminal site selection preference. Several examples of the direct application of McFadden's discrete choice theory to crime include an analysis of the target selection by burglars in The Hague, Netherlands (Bernasco and Nieuwbeerta, 2005) and several studies of robberies in Chicago (Bernasco and Block, 2009; Bernasco et al., 2012). Xue and Brown (2006) develop criminal site selection models that adapt the spatial choice modeling approach for conditions in which the individual discrete choices (crimes) cannot be attributed to individual criminals, which is the case for most of the crime data available to police for use in predictive policing. Their work provides an extensive discussion of the assumptions involved in this model adaption. In brief, their approach relies upon assuming that both the choice set and the decision-making preferences of all of the modeled actors (criminals) in the study domain are similar, and the model therefore describes what is generally true about the criminal preferences in a geographic region.

Xue and Brown (2006) incorporate the idea of using feature-space rather than geographic coordinates to represent the locations of crimes. Feature-space is defined as the Euclidean distance to each of the features of interest such as various crime attractors and crime generators (Liu and Brown, 2004). Their research group has shown that various forms of these criminal site selection models significantly improve predictive performance over the traditional kernel density methods for predicting burglaries (Liu and Brown, 2004; Xue and Brown, 2006) and terrorist events (Brown et al., 2004) such as suicide bombings (Smith and Brown, 2007). One noted reason for this performance improvement is that these criminal site selection models can highlight high risk areas (those very likely to observe a future criminal incident based upon the features of that location) that kernel density approaches do not highlight because they are far from previously observed crimes (Liu and Brown, 2004). Huddleston and Brown (2009) and Huddleston et al. (2012) extend these criminal site selection models by using multilevel modeling to further improve performance for predicting the locations of crimes by specific criminal street gangs and by using gang anchor points as predictive features. Based on this research, Huddleston (2008) notes that the strong distance-decay effect observed when gang anchor points are used as predictive features indicate that it should be possible to adapt CSS models for JTC analysis. The remainder of this chapter documents the adaption of CSS models for use in JTC analysis using a well-known gang data set as an example application.

7.2 Santa Ana Gang Data

Many researchers have noted the similarities between criminal street gangs and insurgent groups operating in urban environments (Freeman and Rothstein, 2011; Arnold et al., 2010; Manwaring, 2005). In this chapter, we use crime data for criminal street gangs in Santa Ana, California, as a substitute for sensitive datasets from ongoing military operations. The data comes from three sources: the Gang Incident Tracking System (GITS) crime dataset for the city of Santa Ana (Meeker et al., 2002), the 2000 US Census, and a gang intelligence map provided by the Santa Ana Police Department. The gang intelligence map details known gang territories and point locations (addresses) for many of the criminal gangs active in the city during the study period. The data set contains crime series containing more than 3 criminal events from the GITS database for 17 of the gangs for which there was a defined anchor point (address) in the gang intelligence map. The US Census provided socio-economic and demographic information at the census block group level for Santa Ana. This gang data set has been used in several previous studies employing CSS models (Huddleston, 2008; Huddleston and Brown, 2009; Huddleston et al., 2012, 2013a).

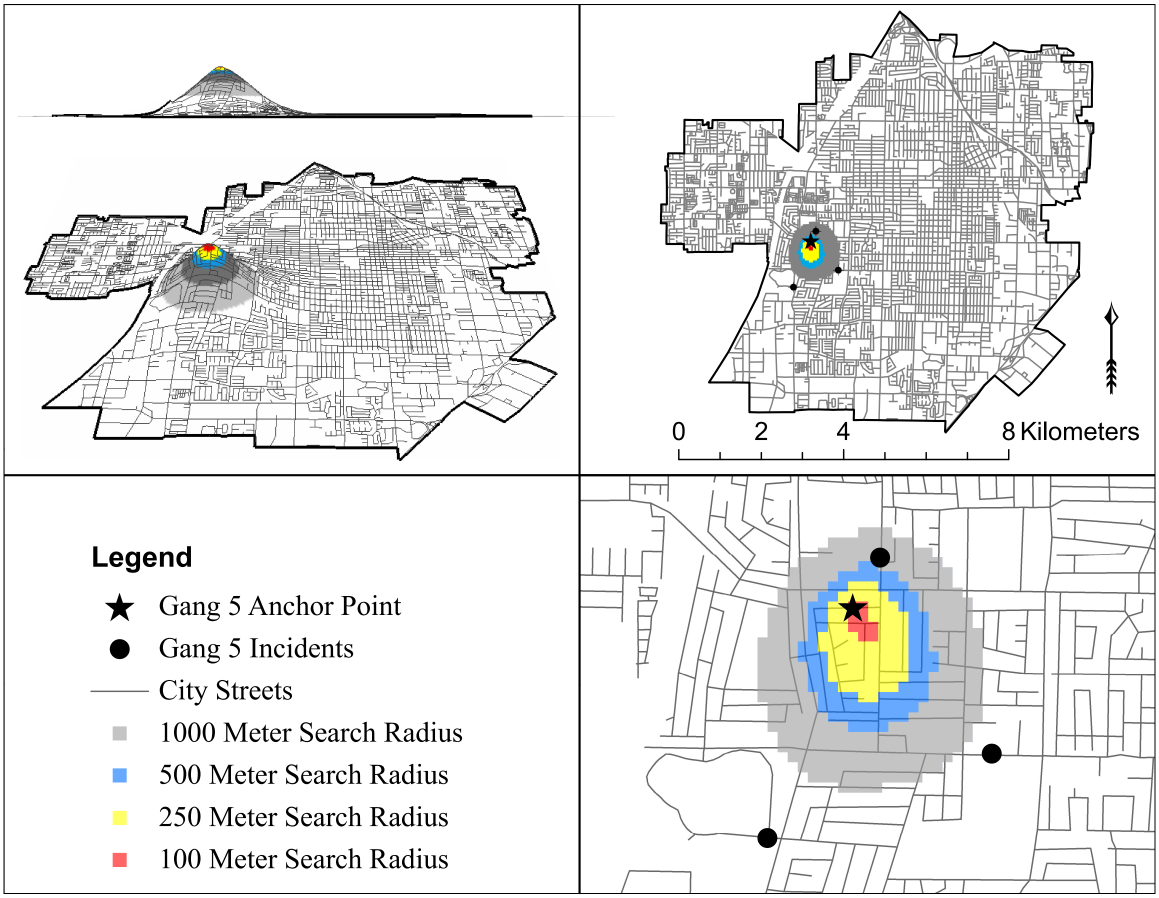


Figure 7.1: A geographic profile produced using the CSSB approach for Gang 5's anchor point in 3-D (top left), at the city scale (top right), and at the scale appropriate for tactical level planning (bottom right)

7.3 Methodology

The formal definition of this problem is to identify the anchor point $z_j \in \Re^2$ for criminal group j from a crime series of size N_j committed by that group at locations $S_j = \{s_{j1}, s_{j2}, ..., s_{jN_j}\}$. This anchor point will be probabilistically assigned to a grid cell of possible locations $i \in \Re^2$ where i indexes a series of 50 meter x 50 meter grid cells in the domain of interest. The variable I represents the total count of approximately 30,000 grid cells mapped within the city limits of Santa Ana, California. Figure 7.1 provides an example of a geographic profile (the mapped model outputs over the index i) for a criminal anchor point using one of the methods developed below.

7.3.1 Data Set Preparation

Developing a CSS JTC model requires a training dataset containing (solved) crime series linked to serial offenders and their known anchor points. There are several data preparation steps necessary to develop a CSS model of criminal behavior. First, the geographic locations of crimes in the training dataset are converted into a marked spatial point pattern by using a Geographic Information System (GIS) to attach socio-economic features from the US census as well as calculating the Euclidean distance to feature-space predictors (Liu and Brown, 2004) such as the responsible party's anchor point, the nearest gang territory, the nearest gang address and other crime generators or attractors (Brantingham and Brantingham, 1981). Note that the distance between the crimes and anchor points, denoted $d(s_n, z)$ in Equation 7.1, is only one of many geographic distances considered in the model structure. The response variable y for all of the actual crime locations is recorded as $y_i = 1$. A *null grid* is laid over the study domain at 50-meter intervals and marked with predictive features in the same way. The null grid provides observations for locations where criminals chose not to commit crimes (i.e. the response variable at these locations is recorded as $y_i = 0$). This grid is also used to map probabilistic estimates for gang anchor points by plotting the model predictions over the grid as a raster image, as seen in Figure 7.1. The training data set therefore contains a response vector Y indexed by i and a predictor matrix X, with each row of X corresponding to a location i and each column of X corresponding to a predictive feature. For notational convenience, the distance between group j's known anchor point z_j and location i is held out of the predictor matrix X and referenced in the notation below as $||i - z_j||$.

7.3.2 Criminal Site Selection (CSS) Modeling of Group Behavior

Criminal Site Selection (CSS) models calculate the probability that a crime by group j occurs at location i using a logistic regression equation.

$$P_j(y_i = 1|X_i, z_j) = \frac{exp\,(AX_i + B\|i - z_j\|)}{1 + exp\,(AX_i + B\|i - z_j\|)} \tag{7.4}$$

As noted above, the vector Y records the binary response ($y_i = 1$ for crime incidence and $y_i = 0$ for no crime) observed for all unique locations in the study domain. A defines a coefficient vector that specifies the relationship to the environmental factors associated with a location which are recorded in vector X_i. The notation $||i - z_j||$ denotes the Euclidean norm (distance) between the known gang anchor point z_j and location y_i. The B coefficient captures the distance-decay relationship for the criminal's journey to crime. This model structure improves current JTC analysis techniques in that it models the effect of the journey to crime relationship after considering other environmental effects such as socio-economic conditions, crime generators, and crime attractors that might affect a criminal's decision-making process.

We fit the logistic regression models in this paper using the `stats` package in R software, using step-wise regression to automate feature selection. Predictors consistently selected using step-wise regression for the CSS models included distance to the gang's anchor point, distance to other gang addresses (known crime generators and attractors), median home values, the percentage of homes that were owner occupied, the percentage of the population 18-30 years old, the percentage of the population on public assistance, and racial demographics.

7.3.3 Developing a Geographic Profile from a New Crime Series

The CSS model is a descriptive model of what is generally true about how criminals in a specific geographic area respond to environmental factors, crime attractors, crime generators, and their own geographic anchor points. Once this model is built for a specific geographic region, it can be used to find the geographic anchor points for a newly observed crime series. Three methods for generating a predictive geographic profile are outlined below. The first method (CSSM) uses the CSS model as a mathematical scoring function, similar to the approach taken in Equation 7.1. The second approach (CSSB) uses the CSS model within the Bayesian framework. The last approach uses the CMD algorithm as a mathematical scoring function. The CMD method is included to enable performance comparison of the CSS modeling approach for JTC analysis to what is widely considered to be the current best method (Levine, 2009a; Snook et al., 2005; Paulsen, 2006b; Levine and Block, 2011).

The input data used to generate the geographic profile for a new group is a crime series by that group. This crime series defines the set of crime locations $S_j = \{s_{j1}, s_{j2}, ...s_{jN_j}\}$. In order to employ the CSS model, we must develop the predictive matrix X used in the CSS model for each of the incidents in S_j, as discussed in Section 7.3.1 above . This forms the predictive matrix X_j, which contains N_j rows, with each row indexing crime n. Finally, the Euclidean distance $\|s_{jn} - i\|$ is calculated for all i and n.

7.3.3.1 A Mathematical Scoring (CSSM) Method

The CSS mathematical scoring method (CSSM) calculates the score L for location i as:

$$L(i) = \sum_{n=1}^{N_j} P\left(s_{jn}|X_{jn}, z_j = i\right) = \sum_{n=1}^{N_j} \left[\frac{exp\left(AX_{jn} + B\|s_{jn} - i\|\right)}{1 + exp\left(AX_{jn} + B\|s_{jn} - i\|\right)}\right] \tag{7.5}$$

Note that this modeling approach is of the same form as Equation 7.1. The Euclidean distance function $\|s_{jn} - i\|$ fulfills the role of $d(s_n, z_j)$ in Equation 7.1. The CSS model serves as the scoring function f. The significant difference between this approach and previous mathematical modeling approaches is the ability of the CSS model to incorporate additional information about environmental factors such as socio-economic conditions, crime generators, and crime attractors.

7.3.3.2 A Bayesian (CSSB) Method

The Bayesian approach incorporates the CSS model into the Bayesian modeling paradigm. In Bayes Theorem, the probability of interest $P\left(A|B\right)$, called the *posterior probability*, is calculated from the *prior probability* of A, $P\left(A\right)$; the *likelihood function* $P\left(B|A\right)$; and the *data probability* $P\left(B\right)$. Thus, the probability of A being true given that we have observed data B is a function of the likelihood of B occurring given A (calculated from the likelihood function), the unconditional probability of A occurring (the prior probability), and the unconditional probability of B occurring (the data probability).

$$P\left(A|B\right) = \frac{P\left(B|A\right) P\left(A\right)}{P\left(B\right)} \tag{7.6}$$

For this problem, the posterior probability of interest is the probability that the gang's anchor point z_j is located at location i given criminal event s_{jn} and the predictive features at the location of the criminal event described by the vector X_{jn}. Using Bayes Theorem:

$$P\left(z_j = i|X_{jn}, s_{jn}\right) = \frac{P\left(s_{jn}|X_{jn}, z_j = i\right) P\left(z_j = i|X_{jn}\right)}{P\left(s_{jn}|X_{jn}\right)} \tag{7.7}$$

The CSS model shown in Equation 7.4 provides the likelihood function. As previously discussed, the CSS model describes the probability that a crime occurs at location s_{jn} as a function of the gang's geographic anchor point and the features of the environment at the location of interest. Using back-substitution, we can calculate the likelihood function in the numerator of Equation 7.7 as:

$$P\left(s_{jn}|X_{jn}, z_j = i\right) = \frac{exp\left(AX_{jn} + B\|s_{jn} - i\|\right)}{1 + exp\left(AX_{jn} + B\|s_{jn} - i\|\right)} \tag{7.8}$$

Here, we use a non-informative uniform distribution to define the prior distribution $P\left(z_j = i|X_{jn}\right)$. This is equivalent to assuming that a group's anchor point is equally likely to be located in every possible grid cell in the domain (i.e. a group's anchor point could be anywhere within the geographic limits of Santa Ana). Thus, the prior probability for every grid point is a constant, $1/I$. One possible improvement to the demonstrated model would be to assert an informative prior probability for a group's anchor point based upon housing densities or other other predictive features within the domain. Alternatively, as discussed in the conclusions section, additional intelligence information obtained (such as a list of suspected locations and their relative probabilities) could be used to provide the prior probabilities for locations. Levine (2009b) use the jeopardy surface provided by JTC crime modeling approach to develop a prior probability for their Bayesian JTC model.

Estimating the unconditional probability of the data, i.e. the events observed in the crime series, has presented some significant challenges in other geographic modeling approaches employing the Bayesian paradigm. As Levine (2009b) notes, "there is no simple way of estimating the probability of obtaining the information [data] under all possible scenarios." He proposes using the (geographic) probability distribution of all offender residences as a rough approximation for the denominator term in the posterior probability equation. CSS modeling provides one way of addressing this gap. We can simply fit Equation 7.4 without considering the distance to a gang's anchor point as a predictive feature. Then, we can use the resulting CSS model, which has coefficient vector C, to estimate $P(s_{jn}|X_jn)$.

$$P(s_{jn}|X_{jn}) = \frac{exp\,(CX_{jn})}{1 + exp\,(CX_{jn})} \tag{7.9}$$

This requires fitting an additional model. An alternative approach, employed here, is to not estimate the prior probability of the data as it isn't necessary to obtain a good estimate for the posterior probability density. Note that in the way we have developed the problem, the unconditional prior probability of the events in the crime series depends only on the feature set of that location. Since this unconditional probability is by definition independent of the location of the anchor point, we can assert that for all possible anchor points we evaluate with the posterior probability function, this probability will be the same (unknown) constant. The posterior probability can now be defined as the product of some unknown constant D and the likelihood function defined in Equation 7.8.

$$P\left(z_j = i|X_{jn}, s_{jn}\right) = \frac{P\left(s_{jn}|X_{jn}, z_j = i\right)}{IP\left(s_{jn}|X_{jn}\right)} = \frac{1}{ID_{jn}} P\left(s_{jn}|X_{jn}, z_j = i\right) \tag{7.10}$$

Therefore:

$$P\left(z_j = y_i|X_{jn}, s_{jn}\right) \propto P\left(s_{jn}|X_{jn}, z_j = i\right) \tag{7.11}$$

Finally, if we assume that the crime series is a set of independent observations:

$$P\left(z_j = i | s_{j1}, s_{j2}, .., s_{jN_j}\right) = \prod_{n=1}^{N_j} \left[\frac{1}{ID_{jn}} P\left(s_{jn} | X_{jn}, z_j = i\right)\right] \tag{7.12}$$

The product term $\prod_{n=1}^{N_j}\left[1/\left(ID_{jn}\right)\right]$ represents some unknown constant. We can estimate a probability density for the anchor point for unique location, $f(y_i)$ by dropping this constant term. The resulting density estimate at location y_i is the product of the conditionally independent probabilities for the entire crime series and is proportional to the joint posterior probability in Equation 7.13:

$$f(y_i) = \prod_{n=1}^{N_j}\left[P\left(s_{jn}|X_{jn}, z_j = i\right)\right] \propto P\left(z_j = i|s_{j1}, s_{j2}, .., s_{jN_j}\right) \tag{7.13}$$

Figure 7.1 illustrates the resulting mapped probability density surface for one of the criminal gangs. This density surface is sufficient for planning cordon and search operations in an effort to locate the anchor point for a criminal group. To obtain an estimate of the posterior probabilities at the various locations, one can normalize the probability density surface $f(Y)$ to sum to one, producing a probability surface. The resulting mapped joint posterior probability surface is indistinguishable from the mapped density surface shown in Figure 7.1.

7.3.3.3 The Center of Minimum Distance (CMD) Method

The center of minimum distance (CMD) is calculated as:

$$CMD = \underset{i \in \Re^2}{\operatorname{argmin}} \sum_{n=1}^{N_j} \|s_{jn} - i\| \tag{7.14}$$

Many researchers have noted that this statistic provides the most accurate point estimate for the location of a serial offender's anchor point (Levine and Block, 2011; Paulsen et al., 2010; Paulsen, 2006b). This significant drawback to the use of this statistic has been that it provides only a point estimate. However, with some adjustment, the algorithm used to calculate CMD can also be leveraged as a simple heuristic approach for developing a geographic profile:

$$L(i) = \frac{1}{\sum_{n=1}^{N_j} \|s_{jn} - i\|} \tag{7.15}$$

The simple heuristic in Equation 7.15 uses the inverse function to reverse the minimization function and maps this calculation for all i. This produces a geographic profile (jeopardy surface) similar to that shown in Figure 7.1.

7.4 Results

To conduct a performance comparison of the three modeling approaches, we used cross-validation to develop these results by iteratively using the crime series and anchor points for 16 of the gangs to develop a CSS model and then applied that CSS model to predict the "unknown" anchor point of the gang held out of the data set. We assess JTC model performance using three metrics commonly used for geographical profiling models: error distance, search cost, and profile accuracy (Rich and Shively, 2004). Error distance is the Euclidean distance between the point location predicted for the anchor point (the i with the highest geographic profile score) and the actual address for the criminal gang. Search cost is the number of 50 x 50 m grid squares that would have to be searched in order to find the gang anchor point. Table 7.1 provides error distance and search cost performance for each of the gangs.

The CSSM method provides the best overall performance on these metrics. In pairwise comparison, the performance improvements the CSSM method provides over the CMD method are not statistically significant for the the error distance metric but are significant for search cost performance ($p = 0.484$ and 0.032 by Wilcoxon Signed Rank Test). The CSSM approach provides statistically significant performance im-

Gang	Crime Count	Error Distance			Search Cost		
		CMD	CSSB	CSSM	CMD	CSSB	CSSM
1	8	110	710	184	24	444	83
2	9	572	535	742	435	334	607
3	10	100	100	100	14	13	14
4	8	1535	1535	1564	2418	2761	1794
5	4	69	69	69	10	7	9
6	18	118	273	100	14	90	18
7	22	1010	1112	1067	1307	1653	1047
8	10	2189	2334	1462	6479	6920	4404
9	15	50	20	50	2	1	2
10	5	414	534	387	199	339	195
11	14	1944	1957	1955	4099	4421	2640
12	11	1681	1807	1343	5083	5564	2985
13	14	1517	1632	1490	4012	4219	3394
14	7	459	462	588	289	327	376
15	32	31	73	31	1	3	1
16	15	287	519	183	101	330	23
17	18	216	174	216	46	44	37
Average	13	724	814	678	1443	1616	1037

Table 7.1: Crime counts, error distance (in meters), and search cost (in count of 50 meter x 50 meter grid cells) by gang for the three JTC methods

Search Diameter	Search Blocks	CMD	CSSB	CSSM
100 M	4	12%	12%	12%
250 M	25	35%	24%	35%
500 M	100	41%	35%	47%
1000 M	400	59%	59%	59%

Table 7.2: Profile accuracy comparison for various search profiles

provement over the CSSB approach for both error distance and search cost ($p = 0.032$ and 0.005). The CMD method likewise provides statistically significant performance improvement in both measures over the CSSB method ($p = 0.006$ and 0.004).

Table 7.2 summarizes the profile accuracy performance for the three JTC methods for profile areas that can be used to define the cordon limits for increasing echelons of military units conducting cordon and search operations. Profile accuracy measures the percentage of criminal gang anchor points that would be found by conducting a cordon and search for a specifically defined region. For example, the geographic search profile for the smallest echelon of military unit that could conduct a cordon and search operation is a diameter of about 100 meters laid over the target location, or the cordon of a neighborhood region containing four 50 x 50 meter search blocks. The CSSM approach again provides the best overall performance, although for three of the four search profile zones, the CMD method provides equivalent performance.

Figure 7.2 visually summarizes the information in Tables 7.1 and 7.2 . It provides a plot of search cost efficiency: the percentage of gang anchor points in the dataset that are identified when a cordon and search operation of a defined search cost is conducted. As can be seen in the left panel of Figure 7.2, the CSSM provides better overall performance in search cost efficiency. The right panel illustrates the search cost efficiency over the region applicable to military cordon and search operations. As can be seen in this graphic, the performance of the CSSM approach and CMD approach are very similar in this trade-off space.

7.5 Conclusions

Overall, the CSSM model provided the best performance. However, it requires significantly more data than the CMD approach. The CMD method requires only a crime series in excess of three crimes while both of the CSS methods require mapped information about environmental influences (known in military parlance as the *human geography*), a training data set containing crime series linked to their known anchor points, and the ability to fit the CSS statistical model. Therefore, the CMD method provides a simple method that provides good performance for the small search profiles applicable to military cordon and search operations.

A significant shortcoming of all of the modeling approaches demonstrated here

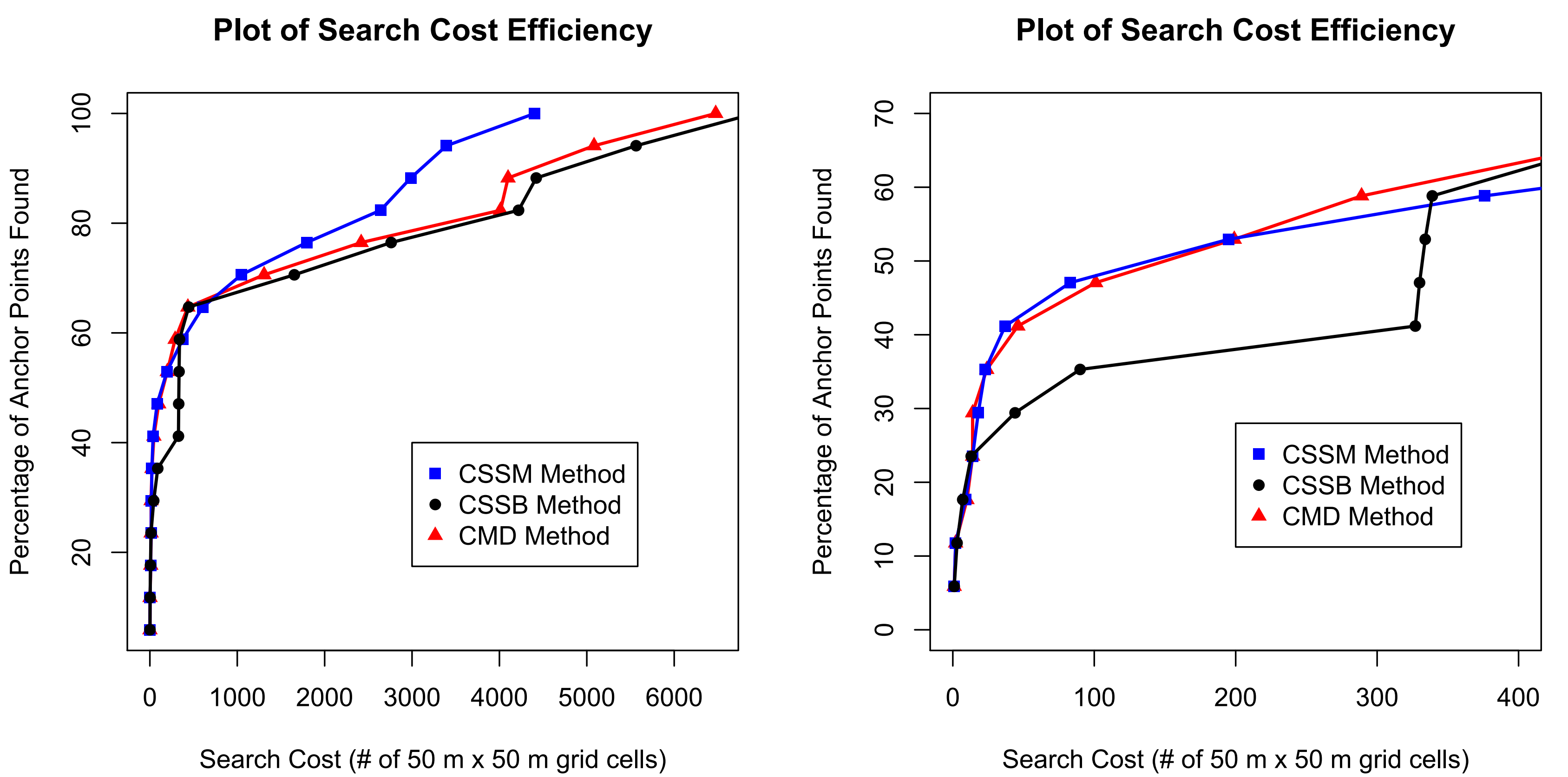


Figure 7.2: A graphical search cost efficiency comparison throughout the trade-off space (left panel) and scaled to highlight the region applicable to military and cordon and search operations (right panel)

(and JTC models in general) is that they cannot accurately identify an anchor point that is not encircled by a crime series (Gangs 8, 10, 12, and 13) and tend to perform poorly when the anchor point is very close to the edge of the crime series (Gangs 7 and 11). Thus, these models are inappropriate for *commuter offenders*, who travel to target areas away from their anchor points to commit their crimes, but are applicable to *marauding offenders* who fan out from a central anchor point in search of criminal opportunities.

While the CSSB method provided the worst performance in this case, a data source important to the CSSB approach was unavailable. One of the strengths of the CSSB approach is the ability to leverage additional information such as an informative distribution for the prior probability for the criminal anchor point for a criminal group. These informative distributions for the prior probability for anchor points could be developed by incorporating data received from additional intelligence sources such as human and signals intelligence (HUMINT/SIGINT). However, the data available for this study did not contain information that could be leveraged in this way.

Both CSS modeling approaches contribute to the JTC analysis literature by providing a method for modeling the effect of the journey to crime relationship after considering other environmental effects such as socio-economic conditions, crime generators, and crime attractors that might affect a criminal's decision-making process. The CSSM approach developed in this chapter is more accurate than the current best method for JTC analysis, calculation of the Fermat-Weber point. The results provided also demonstrate that the geographic profiles developed for these criminal groups are often accurate enough to facilitate tactical success in military application, with the modeled criminal group's anchor point falling within the search profile for military unit cordon and search operations. The CMD method developed here provides a simple heuristic for developing geographic profiles that performs well in comparison to more complex methods in the region of the cost vs. benefit trade-off space applicable to military cordon and search operations. Finally, whereas previous JTC analysis methods have been focused on individual serial criminals, this chapter documents the first use of JTC methods for identifying the criminal anchor point for criminal groups.

Chapter 8

Conclusion

This dissertation develops methodological approaches for exploiting the information provided by predictive crime maps to improve the crime forecasts, geographic districting plans, intelligence assessments, and targeting plans that support decision-making in military and police units. The following sections document the broader implications of this research in other domains, the contributions made in specific research domains, and future work that can be undertaken to extend the results recorded in this dissertation.

8.1 Multidisciplinary Research Contributions

Although this research focuses on addressing existing gaps in the crime analysis literature, this dissertation provides several contributions that are broadly applicable in other research domains. The most broadly applicable contribution provided by this dissertation is the statistical motivation developed to validate the top-down Geographic Probability Forecasting (GPF) approach. The applicability of a top-down forecasting approach, in which weighted aggregate forecasts are used for disaggregated regions or product lines, is highly debated in many disciplines, including sales forecasting, product demand forecasting, and econometrics (Widiarta et al., 2007).

There have been many arguments made using empirical studies that argue both sides of the issue, with some studies advocating for the top-down forecasting approach (Gross and Sohl, 1990; Fogarty et al., 1991; DeLurgio, 1998; Kahn, 1998; Ballou, 1999) and other studies asserting that disaggregated forecasts should provide similar or better results (Dunn et al., 1971; Shlifer and Wolff, 1979; Dangerfield and Morris, 1992; Gordon et al., 1997; Diebold, 1998). The empirical results in many of these studies are largely unexplained (Birmingham and D'Agostino, 2011). As a result, the general rule of thumb in forecasting is to comprehensively fit both top-down and bottom-up forecasting models and use the results of each method to better inform the other models (Allen, 2001; Lapide, 2006). This requires fitting and analyzing a very large number of models in practical application.

The simple proof developed to motivate the GPF method suggests that when all disaggregated regions, product lines, or sales locations are equally affected by trend and seasonality effects (i.e. time series correlated), and the same estimator (forecasting method) is used to forecast at the aggregated and disaggregated levels, then the weighted aggregate forecast should always outperform the disaggregated forecasts for the various regions, product lines, or sales location. The results in this dissertation demonstrate that, especially when the time series are noisy, the top-down forecasting approach can offer significant performance improvement.

The simulation study conducted in this dissertation validated the results suggested by the statistical motivation for cases where the modeling assumptions hold and found that the method is also fairly robust to some violations of the model assumptions. However, when the modeling assumptions are violated due to the introduction of large shocks or step-changes in the process in one of the regions, then the disaggregated forecasts provide better performance. Since in many real-world applications, such as crime forecasting, the GPF method also dramatically reduces the modeling workload, the GPF method provides a simple, robust, general purpose method for improving forecasts for noisy geographic time series. These results also indicate that top-down forecasting methods may be far more robust and applicable than the current literature in sales/demand forecasting and econometrics suggest.

The Sphere of Influence (SOI) analysis developed in this dissertation should apply in any research domain in which we want to compare some spatial choice behavior (criminal, consumer, etc.) across individuals or groups and map the probability that a group's or individual's spatial choice behavior will dominate all others at a given geographic point. This research has additional applications in commercial real estate development (Waddell and Moore, 2008), transportation and travel demand analysis (Ben-Akiva and Lerman, 1985; Timmermans and Golledge, 1990), and marketing (Timmermans and Golledge, 1990). In all of these applications, the SOI analysis provides the opportunity to answer the question, "Who is my most likely customer and where do they live/work?".

Finally, this dissertation documents the development of a new method for geographic profiling models. This approach models the effect of distance decay relationships on the spatial choice behavior of agents originating from geographic anchor points after the considering the effects of many environmental factors such as socio-economic conditions and geographic attractors and then leverages those relationships to find the most likely origin for new agents exhibiting similar choice behavior within the domain of interest. While geographic profiling models were originally developed to assist police units solve serial criminal cases, they are increasingly being applied in ecological models that study animal foraging behavior (Comber et al., 2006). These geographic profiling models have been applied to ecological studies of sharks (Martin et al., 2009), bumble-bees (Raine et al., 2009), and for identifying the source populations for invasive species (Stevenson et al., 2012). The geographic profiling models used in these ecological studies consider only the distance-decay variable in the spatial choice behavior of the studied animals/plants, ignoring the effect of any environmental factors. The CSS models used in this dissertation incorporate both distance decay relationships and the effect of other environmental factors. The results demonstrated in this dissertation suggest that incorporating the CSS model structure into these ecological studies may provide significant benefits.

8.2 Domain Specific Research Contributions

This dissertation also made several research contributions specific to the crime analysis and simulation communities. These research contributions are summarized for each of the problem domains below.

8.2.1 Crime Forecasting

The research into crime forecasting in this dissertation has three practical applications to security force planning. First, as demonstrated, the GPF method improves forecasting performance while also simplifying the modeling process. Second, this modeling approach directly links the analytic products used for operational level decision-making (region forecasts) with the analytic products used for tactical level targeting and planning (threat surfaces or hot-spot maps), providing a common frame-work for tactical and operational level planners. Finally, this modeling approach suggests that threat surface maps can also improve the way that operational planners spatially assign areas of responsibility to subordinate elements by linking decisions about spatial areas of responsibility to forecasts of future activity.

8.2.2 Patrol District Design

The geographic planning surface produced by multiplying the estimate for the long-term average event count at each unique location by a cost estimate for each event type produces an intuitive, interactive planning interface that allows planners to estimate the impacts of moving patrol sector boundaries. Because the planning surface is mapped, it allows planners to interact simultaneously with many heuristic planning rules, taking into consideration topology, geography, the "human terrain," the size of the patrol sectors, and available patrol units. This would seem to provide an opportunity for the development of patrol planning software that could be incorporated as a module into ArcGIS or crime mapping software, providing an automatically updated, interactive interface for police and military planners to use as they design patrol districts.

This dissertation also explored the limitations of applying this approach. The GPF district design method applies when police or military units apply geo-policing principles, with units within sector providing support to calls-for-service and incidents within their district. However, in many policing applications, when the in-district demand exceeds in-district supply, police patrols begin crossing boundaries to meet demand in other police sectors at a very high frequency. This scenario produces a level of complexity that the GPF district design method is not well-equipped to handle. In the Charlottesville Police Department study, in which a geo-policing strategy is not employed and cross-boundary support is frequent, only the agent-based simulation model accurately represents the resulting complexities and significantly changes the workload variation scores to reflect the behavior of the system.

8.2.3 Criminal Group Intelligence Assessment

Previous work in Criminal Site Selection (CSS) modeling is extended to develop a Sphere of Influence (SOI) analysis, which provides new products for the intelligence assessment of criminal groups. These intelligence products accurately map where in geography different threat groups present the dominant threat. CSS models can also be used to generate products such as the "gang resource map" that illustrates the highest five percent risk area for gang activity in a city and identifies the gang most likely to commit a crime in those areas. These results can also be extended to consider the most likely locations for various crime types (i.e., car bombs vs. roadside IEDs or gang drug sales vs. criminal assaults) or any other categorical or hierarchical structure within crime data (Huddleston, 2008; Huddleston and Brown, 2009).

8.2.4 Geographic Profiling

Criminal Site Selection (CSS) models are leveraged to develop a new geographic profiling technique that out-performs the current best method in predicting the geographic anchor points of criminal gangs committing crimes in a dense urban environment. The CSS modeling approaches developed in this dissertation contribute to the geographic profiling model literature by providing a method for modeling the effect of the journey to crime relationship after considering other environmental effects such as socio-economic conditions, crime generators, and crime attractors that might affect a criminal's decision-making process. Current geographic profiling methods rely exclusively on exploiting the distance-decay relationship or can incorporate environmental considerations only as prior probabilities using a Bayesian modeling paradigm.

The extension of the existing Center of Minimum Distance (CMD) method developed in this dissertation provides a simple approach for developing geographic profiles that performs well in comparison to more complex methods in the region of the cost vs. benefit trade-off space applicable to military cordon and search operations. This method offers a fast, generally applicable method for generating a geographic profile surface that can be applied by virtually any analyst. Finally, whereas previous geographic profiling methods have been focused on individual serial criminals, this dissertation documents the first use of geographic profiling methods for identifying the anchor point for criminal groups.

8.2.5 Simulation Modeling

This research also demonstrates a significant role for simulation models in the study of forecasting methods. Using a simulation model to study the properties of these forecasting methods offers three significant benefits. First, with a simulation model, one can easily vary the conditions of the simulation and observe the resulting effects on the performance of the methods. Within the simulation model, not only can one generate noisy geographic time series that include trends, seasonality, and shocks but one can vary the intensity of these effects at will. Second, in a simulation model, a *known* process generates the various time series. So, one can evaluate forecasting methods on how well they model a known process instead of conducting performance comparisons against observed counts in an observational setting for which the true spatial-temporal process is unknown. Removing the random noise from the evaluation measures is especially helpful when evaluating performance against exceptionally noisy processes such as Poisson event counts. Finally, simulation models replicate, repeatedly generating simulated outcomes from the same processes.

This replication provides the opportunity to study the convergence properties of the estimators. Because of the nature of the problem, closed form proofs can only be developed for specific and limiting cases, such a previous study comparing top-down AR(1) forecasts to AR(1) forecasts made for disaggregated regions (Widiarta et al., 2007). It would be a very large task to develop closed form proofs for every limiting case. This dissertation demonstrates that we can quickly study the practical benefit of a method and confirm that results converge to those suggested by the statistical motivation, without needing to study the convergence properties of every model case with closed form proofs. The simulation model in this dissertation confirms the statistical properties suggested by the very simple proof developed to motivate the top-down GPF method and reveals the significant benefits of the method in practical application.

8.3 Future Work

There are improvements and research extensions that could be made to the models presented here. These improvements and research extensions are outlined for each of the problem domains below:

8.3.1 Forecasting

There are several improvements and research extensions to the GPF forecasting method that should be investigated in future work.

(i) Previous research has shown that predictive modeling techniques such as CSS modeling can provide significantly improved predictive performance over the more often used kernel density method for predicting future criminal activity. Since CSS models have a modeling structure that allows spatial integration through summation (due to the conditional independence of probability at all of the considered locations), CSS threat surfaces (hot-spot maps) could be used in the GPF method instead of kernel density hot-spot maps. While these models require a good deal more modeling effort, they may provide performance improvements that would make the extra modeling requirements worth the effort.

(ii) Another research extension would be to investigate other (simpler) approaches for estimating the sub-region weighting parameters w_k. One simple method used in sales forecasting is to use the percentage of the total counts observed in each region up to the current time period.

$$w_{kt} = \frac{\sum_{i \in D_k} \sum_{t=1}^{N-1} y_{it}}{\sum_{i \in D} \sum_{t=1}^{N-1} y_{it}} \tag{8.1}$$

This approach ignores the potential diffusion effects of crime over small geographic regions and does not provide the visual planning surface of the GPF district design method. However, if the simpler approach provides equivalent or better performance than the kernel density method, the simpler calculations might help expand the use of the top-down GPF forecasting method to security agencies that do not have access to Geographic Information Systems (GIS).

(iii) Previous research has established that the spatial distribution of crime does change over time. Therefore, another research extension for the GPF method would be to use exponential smoothing (or another forecasting method) to adjust the estimates for the weighting parameters such that.

$$w_{kt} = w_{k(t-1)} + \alpha \left(\frac{y_{kt}}{y_t} - w_{k(t-1)} \right) \tag{8.2}$$

The α parameter in the above equation could be selected using exponential smoothing methods to minimize the sum of squared error. When the errors in the weighting parameters are significantly auto-correlated (such as immediately after the introduction of a shock), then the approach above should significantly improve the forecasts of the top-down forecasting method.

(iv) One shortcoming of the GPF method is that it provides only a point estimate for a forecast. Holt-Winters and ARIMA models can be implemented within a state-space modeling structure to provide prediction intervals for the forecasts. Research extensions could include developing good estimates for prediction intervals using state-space methods, leveraging error estimates provided by RMSE, or by employing Bayesian methods to develop posterior distributions for the forecasts.

8.3.2 Patrol District Design

There are two research extensions suggested to further develop the applications of the GPF method to patrol district design.

(i) The GPF district design method developed in Chapter 4 used manpower hours as the cost function for the patrol district design. Future research could consider the use of alternative cost functions such as the perceived risk in applications such as military patrol district design. For example, car bombs might present a higher risk or cost function for military units than attacks such as small arms fire. Thus, the patrol district design method outlined in Chapter 4 could be modified to balance the risk that various military units experienced within their districts in the same manner that manpower requirements were balanced in the Chapter 4 example.

(ii) The methodological approach demonstrated in Chapter 4 relies on the use of kernel density estimation for providing the geographic probability (hot-spot) maps. As previously discussed, CSS models often provide significantly improved performance over density methods because they identify potential high-risk areas that have not yet been targeted, whereas the previously discussed density approaches only highlight areas that have previously seen criminal activity. Therefore, CSS or other improved crime mapping methods may provide more accurate representation of future workload requirements for geographic patrol areas.

8.3.3 Mapping Spheres of Influence

There are three research extensions suggested to further develop the SOI analysis.

(i) The multilevel regression modeling approach relies on manual step-wise feature selection across each of the modeled groups because the statistical package used did not provide a convenient way to automate this procedure. This would be a key feature needed to provide maximum benefit in the application to crime analysis software. The development of methods for automated model selection procedures to determine the best model for the entire population of considered groups would greatly expand the opportunities for applying this modeling approach.

(ii) The CSS models used for this analysis ignored temporal considerations. Future research could include temporal features and map how spheres of influence change over time in response to various environmental changes.

(iii) As noted, additional applications of the SOI analysis include modeling the spatial choice behavior of insurgent/terror groups, retail customers, corporate real-estate, and public transportation customers. Future research for the SOI analysis could investigate how well an SOI analysis performs in these research domains.

8.3.4 Geographic Profiling Models

Finally, there are two research extensions suggested by the results obtained in examining the use of CSS models for geographic profiling.

(i) One of the strengths of the Criminal Site Selection - Bayesian (CSSB) approach to geographic profiling is the ability to leverage informative prior probability distributions for the criminal anchor point for a criminal group. In this analysis, only non-informative prior distributions were used. One significant research extension would be be to leverage other available data from sources such as human and signals intelligence (HUMINT/SIGINT) to develop informative prior distributions for the geographic anchor points of criminal groups.

(ii) The CSS geographic profiling models used in this dissertation incorporate both distance decay relationships and the effect of other environmental factors. As discussed, distance-decay based geographic profiling models have recently been applied to ecological studies of animal foraging behavior and the spread of invasive species. The results demonstrated in this suggest that incorporating the CSS model structure into these ecological studies may provide significant benefits.

8.4 Finale

The results presented in this dissertation demonstrate how maps of geographic probability can significantly improve the decision-making tools that government executives, police officials, and military leaders use every day to employ their limited resources in an effort to secure the large and diverse populations they are charged to protect. These results also suggest that these maps of geographic probability, and analytic methods that leverage the insight they provide, can provide greater insight in many other disciplines such as business forecasting, marketing, and ecology. While further work remains in extending the applications of these models and in further refining modeling procedures, this dissertation provides practical models that can be employed immediately by many analysts to provide better support to their decision-makers and clients in a variety of applications.

Appendix A

Pittsburgh Burglary Study Results Reported in RMSE

Tables A.1, A.2, and A.3 provide a model performance comparison using Root Mean Squared Error (RMSE) for weekly burglary forecasts in Pittsburgh. The performance assessment horizon is for the 48 weeks from weeks 4 (to allow for model initialization) through week 51 (because week 52 is a partial week) during the year 2008. Based on model performance recorded in these three tables, the order of model preference (in increasing order) is Naive, Holt-Winters, GPF-HW, ARIMA, and GPF-ARIMA. The Naive method out-performs the Holt-Winters method in aggregate over the precincts, while the Holt-Winters method provides better performance on more patrol sectors. The most significant finding from this example is that the GPF method can improve forecasts for noisy geographic time series in real-world application. In this example, it both simplified weekly forecasting (by reducing the number of forecasting models from 52 weekly forecasting models to one forecasting model and one density estimate each week) and improved the forecasting performance for both the Holt-Winters and ARIMA methods.

Region	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Precinct 1	4.83	4.99	4.03	3.99	**3.83**
Precinct 2	4.03	3.31	3.15	3.23	**2.81**
Precinct 3	4.83	5.32	4.20	4.37	**3.93**
Precinct 4	4.03	5.19	4.14	4.08	**3.84**
Precinct 5	4.73	5.37	4.72	4.31	**3.93**
Precinct 6	1.84	2.23	1.98	**1.73**	1.87
Precinct Aggregate	4.18	4.56	3.81	3.73	**3.46**
Sector Aggregate	1.62	1.62	1.22	1.30	**1.19**

Table A.1: RMSE summary for the weekly one-step ahead burglary forecasts for Pittsburgh in 2008. Table A.3 provides full results for the patrol sector (car beat) level.

	Naive	HW	ARIMA	GPF-HW	GPF-ARIMA
Naive	-	56	92	98	98
HW		-	87	98	98
ARIMA			-	88	98
GPF-HW				-	73
GPF-ARIMA					-

Table A.2: Percentage of precincts and patrol sectors (combined) for which the RMSE of the method on the column is better than the method in the row (i.e., the HW forecasting method improves upon Naive method for 56% of the studied time series).

Sector	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
1	1.23	0.99	0.93	1.05	0.92
2	2.45	2.55	1.77	1.82	1.68
3	1.53	1.22	1.14	1.18	1.16
4	1.18	0.88	0.89	0.99	0.88
5	1.27	1.02	0.95	1.02	0.94
6	1.15	1.21	0.97	1.03	0.96
7	1.57	1.18	1.13	1.21	1.10
8	1.02	0.71	0.66	0.68	0.66
9	1.31	1.04	0.92	0.98	0.87
10	1.40	1.52	0.98	1.05	0.98
11	1.32	1.28	0.87	0.98	0.91
12	0.79	0.66	0.57	0.62	0.57
13	0.41	0.57	0.33	0.35	0.33
14	0.89	0.90	0.60	0.64	0.59
15	0.88	0.64	0.61	0.60	0.60
16	2.11	1.84	1.56	1.50	1.47
17	2.66	3.17	2.19	2.27	2.10
18	2.40	2.13	1.79	1.90	1.73
19	1.41	1.20	0.92	1.08	0.94
20	2.24	2.16	1.61	1.62	1.56
21	1.49	1.28	1.05	1.17	1.06
22	1.21	1.21	0.95	0.99	0.94
23	2.20	2.33	1.76	1.84	1.70
24	1.93	2.34	1.54	2.10	1.57
25	1.74	2.47	1.57	1.60	1.53
26	2.40	2.12	1.86	2.12	1.85
27	1.57	1.74	1.31	1.51	1.32
28	1.46	1.46	1.14	1.19	1.08
29	1.72	1.72	1.23	1.38	1.25
30	1.78	1.49	1.32	1.40	1.31
31	0.95	1.35	0.73	0.82	0.71
32	1.94	2.19	1.61	1.52	1.55
33	1.67	1.34	1.31	1.37	1.26
34	2.25	2.04	1.66	1.58	1.54
35	1.63	1.79	1.23	1.34	1.23
36	2.39	2.15	1.56	1.65	1.57
37	1.90	2.22	1.29	1.41	1.26
38	1.81	2.11	1.42	1.51	1.42
39	2.25	2.19	1.56	1.65	1.54
40	0.87	0.76	0.66	0.75	0.67
41	0.76	0.68	0.59	0.67	0.59
42	1.04	1.02	0.85	0.89	0.83
43	0.98	0.90	0.68	0.74	0.68
44	1.01	0.98	0.92	0.94	0.90
45	1.03	1.16	0.78	0.84	0.78
46	0.58	0.53	0.45	0.58	0.45
Aggregate	1.62	1.62	1.22	1.30	1.19

Table A.3: RMSE summary for the weekly one-step ahead burglary forecasts at the patrol sector level for Pittsburgh in 2008.

Appendix B

Design of Experiments for when GPF Modeling Assumptions Apply

Table B.1 contains the DOE for study cases where the GFP modeling assumptions apply (i.e., where any trends or seasonality effects affect all regions). This section briefly outlines the various simulation settings used for these experiments. Each experimental block in Table B.1 contains the number of unique experiments performed, with the number of replicates within each experiment recorded in parenthesis.

As the DOE table depicts, experimental blocks include three different domain rates (20, 40, and 80 events per period) and five different numbers of spatial processes (5, 20, 40, 80, and 100). This blocking enables performance comparison under a wide spectrum of conditions, from those in which region counts are high and processes are intense (i.e., a domain rate of 80 and number of processes is 5) to those when region counts are low and each process is very intermittent. For each experiment, the simulation randomly samples for the spatial location (spatial mean), spatial distribution, and domain rate for the defined number of spatial processes, scaling the rates of the individual processes so that they add up to the rate defined for the experimental block. The simulation replicates this process 50 times with the same experimental

Stationary Homogenous Point Process

Domain Rate	20	40	80	100
Experiments (Replicates)	1 (100)	1(100)	1(100)	1(100)

Stationary Non-Homogenous Point Processes

	Number of Spatial Processes			
Domain Rate	5	20	40	100
20	10 (50)	10 (50)	10 (50)	10 (50)
40	10 (50)	10 (50)	10 (50)	10 (50)
80	10 (50)	10 (50)	10 (50)	10 (50)

Non-Stationary Non-Homogenous Point Processes with Trend

	Positive Trend (100%)		Negative Trend (50%)	
Domain Start Rate	Number of Spatial Processes			
	20	80	40	100
20	10 (50)	10 (50)	10 (50)	10 (50)
40	10 (50)	10 (50)	10 (50)	10 (50)
80	10 (50)	10 (50)	10 (50)	10 (50)

Non-Stat. Non-Homog. Point Processes with Seasonality

	Cycle Freq. = 1/50		Cycle Freq. = 1/25	
Domain Start Rate	Number of Spatial Processes (Amplitude)			
	20 (L)	80 (H)	40 (H)	100 (L)
20	1 (50)	1 (50)	1 (50)	1 (50)
40	1 (50)	1 (50)	1 (50)	1 (50)
80	1 (50)	1 (50)	1 (50)	1 (50)

Non-Stat. Non-Homog. with Positive Trend and Seasonality

	Cycle Freq. = 1/50		Cycle Freq. = 1/25	
Domain Start Rate	Number of Spatial Processes (Amplitude)			
	20 (L)	80 (H)	40 (H)	100 (L)
20	1 (50)	1 (50)	1 (50)	1 (50)
40	1 (50)	1 (50)	1 (50)	1 (50)
80	1 (50)	1 (50)	1 (50)	1 (50)

Non-Stat. Non-Homog. with Negative Trend and Seasonality

	Cycle Freq. = 1/25		Cycle Freq. = 1/50	
Domain Start Rate	Number of Spatial Processes (Amplitude)			
	20 (H)	80 (L)	40 (L)	100 (H)
20	1 (50)	1 (50)	1 (50)	1 (50)
40	1 (50)	1 (50)	1 (50)	1 (50)
80	1 (50)	1 (50)	1 (50)	1 (50)

Table B.1: Design of Experiments (DOE) for simulation study cases showing the number of experiments and replicates per experiment (in parenthesis) where GPF modeling assumptions apply (i.e. where any trends and seasonality are global effects).

settings over 100 time periods, which provides 97 time periods per replicate over which to evaluate one-step ahead forecasts.

Subsequent experimental blocks include the addition of trends and seasonality. Positive trends are scaled so that the expected event count of the process increases by 100% (i.e., the event count in period 100 is double that during period 1). Negative trends are scaled such that the expected event count of the process decreases by 50% over the time horizon, so the expected count in period 100 is $\frac{1}{2}$ that expected in period 1.

The seasonality effects are sinusoidal functions, with two different amplitudes and cycle frequencies considered. The high-amplitude setting [identified using the notation (H) in the DOE table] uses an amplitude equivalent to $\frac{1}{2}$ of the process mean while the low-amplitude setting (L) uses an amplitude of $\frac{1}{4}$ of the process mean. The $\frac{1}{50}$ cycle frequency corresponds to completing one cycle every 50 periods (which mimics the annual temperature cycle which has been shown to affect crime rates), with two full cycles over the evaluation period. The $\frac{1}{25}$ cycle completes four cycles over the evaluation period. Simulation experiments containing seasonality are evaluated from time steps 151 through 250 (100 periods) to provide enough observations for estimating seasonality effects using the Holt-Winters method.

Appendix C

Design of Experiments for when GPF Assumptions Do Not Apply

Table C.1 provides the DOE for simulation study cases that violate the GPF assumptions. The first set of experiments depicted in Table C.1 investigates the effect of shocks on model performance. These experiments contain a stationary homogenous process throughout the domain (with a Poisson rate of 40 events per time period throughout the domain). At time period 50, a new non-homogenous spatial process is introduced in the domain centered in Region 3 and scaled so that the new spatial process only affects that region. Shock processes rates vary from 5% of the domain rate (i.e., a rate of 2 events per time period) to 30% of the domain rate (i.e., the rate within Region 3 more than doubles from 10 to 22). Note that with the introduction of the shock process, both of the model assumptions of the GPF method are violated: the spatial distribution of crimes is no longer fixed over the time horizon used to fit and forecast the model and a phase change has occurred in one geographic region that does not affect the other regions.

The second set of experiments in Table C.1 test model performance under the situation in which competing trends exist in the various regions. These experiments

Shocks

Size of Shock	5%	10%	20%	30%
Experiments (Replicates)	1 (50)	1 (50)	1 (50)	1 (50)

Competing Trends

	Unique Process Rate as % of the Global Rate			
Strength of	10		25	
Negative Trend	Strength of Positive Trend			
	100%	200%	100%	200%
25%	10 (50)	10 (50)	10 (50)	10 (50)
50%	10 (50)	10 (50)	10 (50)	10 (50)

Random Trends

	Number of Spatial Processes			
Domain Rate	20	40	80	100
20	10 (50)	10 (50)	10 (50)	10 (50)
40	10 (50)	10 (50)	10 (50)	10 (50)
80	10 (50)	10 (50)	10 (50)	10 (50)

Table C.1: Design of Experiments (DOE) for simulation study cases showing the number of experiments and replicates per experiment (in parenthesis) where GPF modeling assumptions do not apply (i.e., where trends are different across sub-regions).

contain a non-stationary homogenous process in the domain (with a Poisson rate of 40 events per time period at the model start time). This Poisson process has a negative trend that reduces the Poisson rate by either 25% or 50% over the study time horizon. At the same time, a non-homogenous Poisson process with a positive trend is introduced into Region 3. Thus, Region 3 has a trend that is moving in the opposite direction of the rest of the domain. Several different rates and strengths of trend for the process in the unique region are studied, as depicted in the table.

The third set of experiments in Table C.1 randomly assign trends to every unique process in the domain. The number of spatial processes varies from 20 to 100 and the overall domain rates (at the start of the simulation) vary from 20 to 80. The simulation model randomly selects a unique trend for each spatial process from $U(-50\%, 100\%)$. Thus, a spatial process is equally likely to double in intensity or halve in intensity over the simulation time horizon. Each spatial process is therefore a non-stationary, non-homogenous Poisson process, with the spatial location and spatial distribution parameters randomly chosen for each spatial process at the beginning of the experiment. Thus, each region also has a unique trend, and the overall trend in the domain can be positive or negative but on average should be slightly positive.

Appendix D

Statistical Significance Test Comparisons for Forecasting Methods

Table D.1 provides pairwise performance comparisons between each of the five studied methods for the simulation cases where the GPF method assumptions hold. Each block of the table reports the percentage of experimental precincts for which the method in the column outperforms the method in the row in RMSE over the evaluation time horizon. For example, in Table B.1 there are 4 experiments conducted for Stationary Homogenous Point Processes. Each of those experiments has four precincts. So, in Table D.1, the performance evaluation takes place over 16 experimental precincts in the Stationary Homogenous Point Process case. The HW method outperforms the Naive method in 94% (15/16) of these experimental precincts.

Table D.1 also records in parenthesis the percentage of precinct time periods for which the difference in MSE performance is statistically significant. In the experiments conducted for Stationary Homogenous Point Processes, there are 97 evaluation time periods per experiment, four experiments conducted, and four precincts per experiment. Thus, there are 1552 time periods over which to test the statistical significance of performance differences. Figure 3.8 illustrates why it is necessary to conduct the statistical tests over the 1552 precinct time periods. Figure 3.8 provides time series plots for the 100 replicates of experiment conducted using a Stationary Homogenous Point Process with a domain rate of 40 events per time period. As can be seen in this plot, the errors for both the ARIMA and GPF-ARIMA methods are heteroscedastic, with both methods providing increasingly accurate forecasts. Thus, statistical testing of performance differences cannot be conducted over multiple time periods.

The Wilcoxon Signed Rank Test (WSRT) is used to test the hypothesis that the median residual forecast error for the method in the column is less than the median residual forecast error for the method in the row. The error residuals within most precinct time periods are non-Gaussian (thus requiring a non-parametric statistical test) and each replicate provides a paired sample across the methods because the same observations from each replicate are used to produce forecasts with all five methods. This pairwise statistical test is conducted for each precinct time period and the percentage of time that the p-value from the WSRT is smaller than 0.05 (i.e., provides statistical significance at the 95% level) is recorded in parenthesis in Table D.1. Thus, the Holt-Winters (HW) method provides statistically significant performance improvement over the Naive method for 61% of the 1552 precinct time periods in the simulation model experiments conducted using stationary homogenous point processes. It is possible for a method to provide statistically significant performance improvement for some percentage of time periods without providing better overall RMSE performance for any precinct.

Table D.1 organizes the methods in order of increasing performance from left to right across the top of the table (and from top to bottom in the left column). Thus, for stationary time series and time series with trends, the (increasing) order of preference for the methods is Naive, HW, GPF-HW, ARIMA, and GPF-ARIMA. For time series containing seasonality, the preference order remains the same with the exception that the GPF-HW method provides better performance than the ARIMA method in this case.

Stationary Homogenous Point Process

	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Naive	-	94 (61)	100 (97)	100 (100)	100 (100)
HW	6 (9)	-	100 (100)	100 (100)	100 (100)
GPF-HW	0 (1)	0 (0)	-	100 (100)	100 (100)
ARIMA	0 (0)	0 (0)	0 (0)	-	100 (81)
GPF-ARIMA	0 (0)	0 (0)	0 (0)	0 (0)	-

Stationary Non-Homogenous Point Process

	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Naive	-	98 (0)	98 (93)	98 (97)	98 (97)
HW	0 (8)	-	98 (93)	98 (98)	98 (98)
GPF-HW	0 (1)	0 (0)	-	95 (84)	98 (95)
ARIMA	0 (0)	0 (0)	0 (1)	-	98 (26)
GPF-ARIMA	0 (0)	0 (0)	0 (0)	0 (2)	-

Non-Stationary Non-Homogenous Point Process with Trend

	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Naive	-	99 (0)	100 (96)	100 (100)	100 (100)
HW	1 (8)	-	100 (97)	100 (92)	100 (100)
GPF-HW	0 (1)	0 (0)	-	77 (17)	100 (73)
ARIMA	0 (0)	0 (0)	23 (25)	-	100 (73)
GPF-ARIMA	0 (0)	0 (0)	0 (0)	0 (1)	-

Non-Stationary Non-Homogenous Point Process with Seasonality

	Naive	HW	ARIMA	GPF-HW	GPF-ARIMA
Naive	-	100 (57)	98 (97)	100 (99)	100 (100)
HW	0 (0)	-	98 (66)	100 (94)	100 (98)
ARIMA	2 (0)	2 (0)	-	94 (39)	100 (77)
GPF-HW	0 (0)	0 (0)	6 (6)	-	100 (44)
GPF-ARIMA	0 (0)	0 (0)	0 (0)	0 (1)	-

Table D.1: Pairwise performance comparison for simulation cases where the GPF assumptions hold that shows the percentage of experimental precincts over which the method in the column outperforms the method in the row and the percentage of experimental precinct time periods (in parenthesis) in which the method in the column provides statistically significant performance improvement over the method in the row at the 95% confidence level by Wilcox Signed Rank Test (WSRT).

Table D.2 provides pairwise performance comparisons for the simulation cases where the GPF method assumptions do not hold. There are several important insights gained in examining these cases. Using rolling horizons to fit the GPF models provides some insurance against the effect of shocks or changing spatial distributions (as in the case of competing trends) but it comes with the cost of increased variance (and error) for those periods in which the spatial process does not change. The GPF methods do not perform well in periods immediately following shocks. When the spatial distribution of events changes over time, it is necessary to use a rolling horizon model to capture these changes. However, when the distributions do not change significantly, using a rolling horizon negatively affects performance. Thus, Table D.2 records mixed results. The overall preference order for situations in which there are shocks or competing trends is Naive, HW, GPF-HW, GPF-HW-R, GPF-ARIMA, ARIMA, and GPF-ARIMA-R. However, if there are many shocks that take place in a domain (in comparison to just inserting one as in this case), the ARIMA and HW methods will improve over the GPF methods because there will not be opportunities for the estimates of w_k to recover (it takes time). Thus, while the overall performance of the rolling horizon methods were better in this case, they were better only because there was time after the shock for the estimates of w_k to be accurately represented. In the case of random trends, the fact that many independent and uncorrelated trends are competing produces fairly stable results across the different domains. Thus, the w_k for the precincts are fairly stable and the rolling horizon GPF methods do not improve upon those using all previous time periods.

Shocks

	Naive	HW	GPF-HW	GPF-HW-R	GPF-ARIMA	ARIMA	GPF-ARIMA-R
Naive	-	100 (66)	94 (83)	100 (93)	94 (90)	100 (98)	100 (98)
HW	0 (8)	-	88 (70)	100 (89)	88 (79)	100 (98)	100 (97)
GPF-HW	6 (4)	13 (13)	-	81 (34)	88 (83)	100 (87)	100 (94)
GPF-HW-R	0 (1)	0 (2)	19 (12)	-	81 (70)	100 (82)	94 (93)
GPF-ARIMA	6 (4)	13 (10)	13 (4)	19 (4)	-	56 (31)	81 (14)
ARIMA	0 (1)	0 (1)	0 (1)	0 (1)	44 (27)	-	63 (30)
GPF-ARIMA-R	0 (1)	0 (2)	0 (1)	6 (1)	38 (34)	19 (32)	-

Competing Trends

	Naive	HW	GPF-HW	GPF-HW-R	GPF-ARIMA	ARIMA	GPF-ARIMA-R
Naive	-	100 (69)	100 (84)	100 (95)	100 (90)	100 (99)	100 (100)
HW	0 (8)	-	100 (62)	100 (87)	100 (68)	100 (91)	100 (99)
GPF-HW	0 (2)	0 (10)	-	100 (46)	100 (44)	100 (51)	100 (80)
GPF-HW-R	0 (1)	0 (0)	0 (3)	-	63 (34)	88 (34)	100 (59)
GPF-ARIMA	0 (1)	0 (9)	0 (8)	38 (8)	-	63 (34)	100 (55)
ARIMA	0 (0)	0 (0)	0 (6)	13 (6)	38 (23)	-	100 (58)
GPF-ARIMA-R	0 (0)	0 (0)	0 (0)	0 (0)	0 (5)	0 (7)	-

Random Trends

	Naive	HW	GPF-HW-R	GPF-HW	ARIMA	GPF-ARIMA-R	GPF-ARIMA
Naive	-	99 (72)	100 (96)	100 (96)	100 (100)	100 (100)	100 (100)
HW	0 (7)	-	100 (94)	100 (93)	100 (98)	100 (100)	100 (98)
GPF-HW-R	0 (1)	0 (0)	-	86 (25)	97 (55)	100 (81)	98 (84)
GPF-HW	0 (1)	0 (0)	12 (6)	-	95 (48)	99 (66)	99 (81)
ARIMA	0 (0)	0 (0)	2 (5)	4 (5)	-	97 (29)	93 (46)
GPF-ARIMA-R	0 (0)	0 (0)	0 (0)	0 (0)	1 (8)	-	82 (41)
GPF-ARIMA	0 (0)	0 (0)	1 (1)	0 (1)	5 (5)	16 (9)	-

Table D.2: Pairwise performance comparison for simulation cases where the GPF do not hold that shows the percentage of experimental precincts over which the method in the column outperforms the method in the row and the percentage of experimental precinct time periods (in parenthesis) in which the method in the column provides statistically significant performance improvement over the method in the row at the 0.05 level by Wilcox Signed Rank Test (WSRT).

Appendix E

MASE Performance Summaries for Over-Dispersed Scenarios

This appendix provides MASE performance summaries for when the DOE Tables recorded in Appendices B and C are conducted using overdispersed Poisson distributions to generate the event counts. The Poisson distributions used for the results in this appendix are specified such that the variance of the process is 125% of the mean of the process. Thus, the event counts in the various regions are much noisier for the experiments recorded below than for those discussed in Chapter 3. The results in these tables are directly comparable to Tables 3.5 - 3.8 in Chapter 3. The results recorded below in Tables E.1 - E.4 are very close to those recorded in Tables 3.5 - 3.8, indicating that the overdispersed event counts do not significantly affect the results recorded in Chapter 3. When the modeling assumptions of the GPF method apply, the GPF method improves performance when applied to both Holt-Winters and ARIMA models. The ARIMA method still provides better performance than the GPF-ARIMA method in the case of competing trends and shocks and provides performance equivalent to the GPF-ARIMA-R method in the case of shocks.

Scenario	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Stationary Homogenous	1.00	0.92	0.78	0.74	0.72
Stationary Non-Homogenous	1.00	0.92	0.79	0.74	0.73
Trend	1.00	0.93	0.98	1.01	0.97
Seasonality	1.00	0.86	0.75	0.76	0.73
Season & Positive Trend	1.00	0.84	0.75	0.75	0.73
Season & Negative Trend	1.00	0.90	0.77	0.80	0.75

Table E.1: MASE performance summary for the observed error for the five considered forecasting methods over the six scenarios in which GPF modeling assumptions apply for scenarios in which over-dispersed Poisson distributions are used to generate event counts.

Scenario	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Stationary Homogenous	1.00	0.80	0.45	0.25	0.18
Stationary Non-Homogenous	1.00	0.79	0.45	0.25	0.20
Trend	1.00	0.80	0.44	0.42	0.29
Seasonality	1.00	0.67	0.35	0.42	0.26
Season & Positive Trend	1.00	0.62	0.32	0.39	0.27
Season & Negative Trend	1.00	0.77	0.40	0.50	0.33

Table E.2: MASE performance summary for the process error for the five considered forecasting methods over the six scenarios in which GPF modeling assumptions apply for scenarios in which over-dispersed Poisson distributions are used to generate event counts.

Scenario	Naive	HW	GPF-HW	GPF-HW-R	ARIMA	GPF-ARIMA	GPF-ARIMA-R
Competing Trends	1.00	0.93	0.83	0.80	0.78	0.79	0.76
Random Trends	1.00	0.92	0.79	0.79	0.75	0.74	0.74
Shocks	1.00	0.92	0.84	0.80	0.76	0.79	0.75

Table E.3: MASE performance summary for the observed error for the seven considered forecasting methods for the considered scenarios where the GPF assumptions do not hold for scenarios in which over-dispersed Poisson distributions are used to generate event counts.

Scenario	Naive	HW	GPF-HW	GPF-HW-R	ARIMA	GPF-ARIMA	GPF-ARIMA-R
Competing Trends	1.00	0.80	0.57	0.49	0.39	0.44	0.31
Random Trends	1.00	0.78	0.45	0.46	0.33	0.28	0.30
Shocks	1.00	0.79	0.61	0.50	0.31	0.42	0.31

Table E.4: MASE performance summary for the process error for the seven considered forecasting methods for the considered scenarios where the GPF assumptions do not hold for scenarios in which over-dispersed Poisson distributions are used to generate event counts.

Bibliography

M. Adya, F. Collopy, J.S. Armstrong, and M. Kennedy. Automatic identification of time series features for rule-based forecasting. *International Journal of Forecasting*, 17, 2001.

Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

P. Geoffrey Allen. Econometric forecasting. In J. Scott Armstrong, editor, *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, Norwell, MA, 2001.

M. Altman. The computational complexity of automated redistricting: Is automation the answer? *Rutgers Computer Technology Law Journal*, 23:81–142, 1997.

B. Arinze. Selecting appropriate forecasting models using rule induction. *Omega*, 22, 1994.

J. Scott Armstrong and Fred Collopy. Error measures for generalizing about forecasting methods: Emperical considerations. *International Journal of Forecasting*, 8, 1992.

Laurence H. Arnold, Christopher W. O'Gwin, and Jeremy S. Vickers. Small town insurgency: The struggle for information dominance to reduce gang violence. Technical report, Monterey, CA, 2010.

R. H. Ballou. *Business Logistics Management*. Prentice-Hall, New Jersey, 1999.

Douglas M. Bates and Jose C. Pinheiro. Computational methods for multilevel modelling. Technical report, Bell Labs, 1998.

G.S. Becker. *The Economic Approach to Human Behavior*. University of Chicago Press, Chicago, 1976.

Mosh Ben-Akiva and Steven R. Lerman. *Discrete Choice Analysis: Theory and Applications to Travel Demand*. MIT Press, Cambridge, MA, 1985.

Mark Berman and Peter Diggle. Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(1), 1989.

Wim Bernasco and Richard Block. Where offenders choose to attack: A discrete choice model of robberies in chicago. *Criminology*, 47(1):93–130, 2009.

Wim Bernasco and Paul Nieuwbeerta. How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *The British Journal of Criminology*, 45(3):296–316, 2005.

Wim Bernasco, Richard Block, and Stijn Ruiter. Go where the money is: modeling street robbers location choices. *Journal of Economic Geography*, 2012.

Julien Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, 1974.

T. Birdsall and W. Fox. The theory of signal detectability. *IEEE Transactions on Information Theory*, 4(4), 1954.

Coline Birmingham and Antonello D'Agostino. Understanding and forecasting aggregate and disaggregate price dynamics. Technical report, European Central Bank, 2011.

Roger S. Bivand, Edzer J. Pebesma, and Virgilio Gomez-Rubio. *Applied Spatial Data Analysis with R.* Springer, 2008.

C.R. Block and R. Block. Street gang crime in chicago. Technical report, A Research Bulletin of the National Institute of Justice, 1993.

Richard Block. Gang activity and overall levels of crime: A new mapping tool for defining areas of gang activity using police records. *Journal of Quantitative Criminology*, 16(3):369–383, 2000.

Rachel Boba. *Crime Analysis and Crime Mapping*. Sage Publications, London, 2005.

Paul Bolstad. *GIS Fundamenals: A First Text on Geographic Information Systems.* Eider Press, White Beark Lake, Minnesota, 3 edition, 2008.

George Edward Pelham Box and Gwilym Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, Incorporated, 1990. ISBN 0816211043.

Burak Boyaci and Nikolas Geroliminis. Extended hypercube models for large scale spatial queuing systems. Technical report, Ecole Polytechnique Federale de Lausanne, 2011.

Patricia L. Brantingham and Paul J. Brantingham. Environment, routine, and situation: Toward a pattern theory of crime. In Ronald V. Clarke and Marcus Felson, editors, *Routine Activity and Rational Choice: Advances in Criminological Theory.* Transaction Publishers, 1993.

Paul J. Brantingham and Patricia L. Brantingham. *Environmental Criminology*. Sage Publications, Thousand Oaks, 1981.

L. Breiman, J. Fiedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth International Group, Belmont, California, 1984.

Donald E. Brown, Jason Dalton, and Heidi Hoyle. Spatial forecast methods for terrorist events in urban environments. In *Second Symposium on Intelligence and Security Informatics*, June 2004.

R.G. Brown. *Statistical Forecasting for Inventory Control*. McGraw-Hill, New York, 1959.

R.G. Brown. *Smoothing, Forecasting, and Prediction of Discrete Time Series*. Prentice-Hall, Englewood Cliffs, NJ, 1963.

Meagan Cahill and Gordon Mulligan. Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, 25(2), 2007.

David Canter. Developments in geographical offender profiling: Commentary on bayesian journey-to-crime modelling. *Journal of Investigative Psychology and Offender Profiling*, 6:161–166, 2009.

David Canter and Paul Larkin. The environmental range of serial rapists. *Journal of Environmental Psychology*, 13(1):63 – 69, 1993.

David Canter, Toby Coffey, Malcolm Huntley, and Christopher Missen. Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, 16:457–478, 2000.

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

Mitchell B. Chamlin. Crime and arrests: An autoregressive integrated moving average (ARIMA) approach. *Journal of Quantitative Criminology*, 4(3), 1988.

Wei Chang, Daniel Zeng, and Hsinchun Chen. Prospective spatio-temporal data analysis for security informatics. In *Proceedings of the 8th International IEEE Conference on Intelligent Tranportation Systems*, Vienna, Austria, 2005.

P. Chen, H.Y. Yuan, and X.M. Shu. Making weekly/daily forecasting of property crime using ARIMA model. *Progress in Safety Science and Technology Vol Vii Pts a and B*, 7:531–535, 2008a.

Peng Chen, Hongyong Yuan, and Xueming Shu. Forecasting crime using the ARIMA model. In *Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery*. IEEE Computer Society, 2008b.

Thomas M. Cioppa, Thomas W. Lucas, and Susan M. Sanchez. In *Proceedings of the 2004 Winter Simulation Conference*, 2004.

Robert B. Cleveland, William S. Cleveland, Jean E. McRae, and Irma Terpenning. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 1990.

Jacqueline Cohen and Wilpen L. Gorr. Development of crime forecasting and mapping systems for use by police. Technical report, U.S. Department of Justice, 2005.

Jacqueline Cohen, Wilpen L. Gorr, and Andreas M. Olligschlaeger. Leading indicators and spatial interactions: A crime forecasting model for proactive police deployment. *Geographical Analysis*, 39(1), 2007.

Lawrence Cohen and Marcus Felson. Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44(4):588–608, 1979.

James S. Coleman and Thomas J. Fararo, editors. *Rational Choice Theory: Advocacy and Critique*. Sage Publications, London, 1992.

F. Collopy and J.S. Armstrong. Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, 38, 1992.

Steven C. Le Comber, Barry Nicholls, D. Kim Rossmo, and Paul A. Racey. Geographic profiling and animal foraging. *Journal of Theoretical Biology*, 240(2):233 – 240, 2006.

Noel A. Cressie. *Statistics for Spatial Data*. John Wiley and Sons, New York, 1993.

Kevin M. Curtin, Karen Hayslett-McCall, and Fang Qiu. Determining optimal police patrol areas with maximal covering and backup covering location models. *Networks and Spatial Economics*, 10:125–145, 2010.

Steven J. D'Amico, Shoou-Jiun Wang, Rajan Batta, and Christopher M. Rump. A simulated annealing approach to police district design. *Computers & Operations Research*, 29(6), 2002.

Byron J. Dangerfield and John S. Morris. Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting*, 8(2):233 – 241, 1992.

S. A. DeLurgio. *Forecasting Principles and Applications*. Irwin/McGraw-Hill, Missouri, 1998.

F. X. Diebold. *Elements of Forecasting*. International Thomson Publishing, Ohio, 1998.

Anne Dray, Lorraine Mazerolle, Pascal Perez, and Alison Ritter. Drug law enforcement in an agent-based model: Simulating the disruption to street-level drug markets. In Lin Liu and John Eck, editors, *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*. Information Science Reference, New York, 2008.

Douglas M. Dunn, William H. Williams, and W. Allen Spivey. Analysis and prediction of telephone demand in local geographical areas. *The Bell Journal of Economics and Management Science*, 2(2), 1971.

Martin Dunn. Levels of war: Just a set of labels. *Research and Analysis: Newsletter of the Directorate of Army Research and Analysis*, 1996.

John E. Eck, Spencer Chainey, James G. Cameron, Michael Leitner, and Ronald E. Wilson. Mapping crime: Understanding hot spots. Technical report, National Institute of Justice, 2005.

Henk Elffers and Pieter Van Baal. Realistic spatial backcloth is not that important in agent based simulation research: An illustration from simulating perceptual deterrence. In Lin Liu and John Eck, editors, *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*. Information Science Reference, New York, 2008.

J. Fan, S. Upadhye, and A. Worster. Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, 8(1), 2006.

Tom Fawcett. ROC graphs: Notes and practical considerations for researchers. Technical report, 2004.

Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 2006.

D. W. Fogarty, T. R. Hoffman, and Jr. J. H. Blackstone. *Production and Inventory Management*. South-Western Publishing, Cincinnati, 1991.

A. Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Sage Publications, London, 2000.

Jon Fox and Donald Brown. Hierarchical feature-space modeling of temporal patterns in criminal site selection problems. *Crime Mapping: A Journal of Research and Practice*, 4(2), 2012.

Jon Fox, Gerard Learmonth, and Donald Brown. Simulating spatial-temporal pulse events in criminal site selection problems. In *Proceedings of SpringSim 2012*, Orlando, 2012.

Michael Freeman and Hy Rothstein. Gangs and guerillas. Technical report, Monterey, CA, 2011.

Everette S. Gardner, Jr. Exponential smoothing: The state of the art - Part II. *International Journal of Forecasting*, 22, 2006.

Saul Gass and Thomas Saaty. The computation algorithm for the parametric objective function. *Naval Research Logistics Quarterly*, 2(1), 1955.

Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, 2007.

Arthur Getis and Barry Roots. *Models of Spatial Processes: An Approach to the Study of Point, Line, and Area Patterns*. Cambridge University Press, 1978.

T. P. Gordon, J. S. Morris, and B. J. Dangerfield. Top-down or bottom-up: Which is the best approach to forecasting? *Journal of Business Forecasting*, 16(3), 1997.

Wilpen Gorr and Richard Harries. Introduction to crime forecasting. *International Journal of Forecasting*, 19, 2003.

Wilpen Gorr, Andreas Olligshlaeger, and Yvonne Thompson. Short-term forecasting of crime. *International Journal of Forecasting*, 19, 2003.

Wilpen L. Gorr. Foreast accuracy measures for exception reporting: Using receiver operating characteristic curves. *International Journal of Forecasting*, 25(1), 2009.

Wilpen L. Gorr and Kristen S. Kurland. *GIS Tutorial for Crime Analysis*. Esri Press, Redlands, California, 2012.

W.L. Gorr and S.A. McKay. Application of tracking signals to detect series pattern changes in crime mapping systems. In F. Wang, editor, *Crime Mapping and Beyond: GIS Applications in Crime Studies*. Idea Group Publishing, Hershey, PA, 2005.

Elizabeth Groff. Characterizing the spatio-temporal aspects of routine activities and the geographic distribution of street robbery. In Lin Liu and John Eck, editors, *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*. Information Science Reference, New York, 2008.

Charles W. Gross and Jeffrey E. Sohl. Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*, 9(3):233–254, 1990.

Tony H. Grubesic. On the application of fuzzy clustering for crime hot spot detection. *Journal of Quantitative Criminology*, 22(1), March 2006.

C. D. Hale. *Police Patrol, Operations, and Management*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1980.

P. G. Hancock and N. C. Simpson. Fifty years of operational research and emergency response. *Journal of the Operational Research Society*, 60(51):126–139, 2009.

J.A. Hanley and B.J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 43, 1982.

Keith Harris. Mapping crime: Principle and practice. Technical report, National Institute of Justice, 1999.

Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2001.

V.E. Henry and W.J. Bratton. *The COMPSTAT Paradigm: Management Accountability in Policing, Business, and the Public Sector*. Looseleaf Law Publications, Flushing, NY, 2002.

Daniel P. Heyman and Matthew J. Sobel. *Stochastic Models in Operations Research: Volume I*. McGraw-Hill, New York, 1982.

C.C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. In *ONR Memorandum, vol. 52*, Pittsburgh, PA, 1957. Carnegie Institute of Technology.

C.C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20, 2004.

C.C. Holt, F. Modigliani, J.F. Muth, and H.A. Simon. *Planning Production, Inventories, and Work Force*. Prentice-Hall, Englewood Cliffs, NJ, 1960.

Samuel H. Huddleston. The statistical threat assessment. Technical report, Charlottesville, VA, 2008.

Samuel H. Huddleston and Donald E. Brown. A statistical threat assessment. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 39(6), 2009.

Samuel H. Huddleston and Donald E. Brown. Using discrete event simulation to evaluate time series forecasting methods for security applications. In *Proceedings of the 2013 Winter Simulation Conference*, 2013.

Samuel H. Huddleston, Gerard P. Learmonth Sr., and Jon Fox. Changing knives into spoons. In *Proceedings of the 2008 IEEE Systems and Information Engineering Design Symposium*, Charlottesville, VA, 2008.

Samuel H. Huddleston, Clark Heidelbaugh, and Don Brock. The warfighter's guide to counter-IED analysis. Technical report, Alexandria, VA, 2010.

Samuel H. Huddleston, Jon Fox, and Donald E. Brown. Mapping gang spheres of influence. *Crime Mapping: A Journal of Research and Practice*, 4(2), 2012.

Samuel H. Huddleston, Matthew Gerber, and Donald E. Brown. Geographic profiling of criminal groups for military cordon and search. In *Proceedings of the 2013 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 2013a.

Samuel H. Huddleston, John H. Porter, and Donald E. Brown. Improving forecasts for noisy geographic time series. *The Journal of Business Research*, 2013b.

Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. 2013.

Rob J. Hyndman and Yeasmin Khandakar. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 2008.

Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679 – 688, 2006.

Jiming Jiang. *Linear and Generalized Linear Mixed Models and Their Applications*. Springer, New York, 2007.

David S. Johnson. The np-completeness column: An ongoing guide. *Journal of Algorithms*, 6:434–451, 1985.

M.C. Jones, J.S. Marron, and S.J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433), 1996.

Kenneth B Kahn. Revisiting top-down versus bottom-up forecasting. *Journal of Business Forecasting Methods and Systems*, 17:14–19, 1998.

Ruth Kerry, Pierre Goovaerts, Robert P. Haining, and Vania Ceccato. Applying geostatistical analysis to crime data: Car-related thefts in the baltic states. *Geographical Analysis*, 42, 2010.

Keivan Kianmehr and Reda Alhajj. Effectiveness of support vector machine for crime hot-spot prediction. *Applications of Artificial Intelligence*, 22(5), 2008.

S. Kolassa and W. Schutz. Advantages of the MAD/MEAN ratio over the MAPE. *Foresight: The International Journal of Applied Forecasting*, 6, 2007.

Larry Lapide. Top-down and bottom-up forecasting in S&OP. *The Journal of Business Forecasting*, 25(2), 2006.

R. C. Larson. Measuring the response patterns of new york city police patrol cars. Technical Report R-673-NYC/HUD, RAND Institute, New York City, 1971.

Richard C. Larson. A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, 1(1), 1974.

Kenneth D. Lawrence, Ronald K. Kimberg, and Sheila M. Lawrence. *Fundamentals of Forecasting Using Excel*. Industrial Press, Inc., New York, 2009.

Ned Levine. Crimestat: A spatial statistics program for the analysis of crime incident locations (version 1.1). Technical report, Ned Levine and Associates, 2009a.

Ned Levine. Introduction to the special issue on bayesian journey-to-crime modelling. *Journal of Investigative Psychology and Offender Profiling*, 6, 2009b.

Ned Levine. Crimestat: A spatial statistics program for the analysis of crime incident locations (version 3.2). Technical report, Ned Levine and Associates, 2009c.

Ned Levine and Richard Block. Bayesian journey-to-crime estimation: An improvement in geographic profiling methodology. *The Professional Geographer*, 63(2), 2011.

E. Lewis, G. Mohler, P.J. Brantingham, and A. Bertozzi. Self-exciting point process models of civilian deaths in iraq. *Security Journal*, 25(3), 2012.

David Ley and Roman Cybriwsky. Urban graffiti as territorial markers. *Annals of the Association of American Geographers*, 64(4):491–505, 1974.

Hua Liu and Donald E. Brown. Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, 19, 2003.

Hua Liu and Donald E. Brown. A new point process transition density model for space-time event prediction. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34:310–324, 2004.

Lin Liu and John Eck. An overview of crime simulation. In Lin Liu and John Eck, editors, *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*. Information Science Reference, New York, 2008.

Lin Liu, Xuguang Wang, John Eck, and Jun Liang. Simulating crime events and crime patterns in a RA/CA model. In Fahui Wang, editor, *Geographic Information Systems and Crime Analysis*. Idea Group Publishing, 2005.

Jorge M. Lobo, Alberto Jimenez-Valverde, and Raimundo Real. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 2007.

Burton G. Malkiel. *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing*. W.W. Norton, New York, 2003.

N. Malleson. *Agent-Based Modelling of Burglary*. Ph.D. thesis, School of Geography, University of Leeds, 2010.

Nick Malleson. Using agent-based models to simulate crime. In Alison J. Heppenstall, editor, *Agent-Based Models of Geographical Systems*. Springer, 2011.

Max G. Manwaring. Street gangs: The new urban insurgency. Technical report, Carlisle, PA, 2005.

R. A. Martin, D. K. Rossmo, and N. Hammerschlag. Hunting patterns and geographic profiling of white shark predation. *Journal of Zoology*, 279(2):111–118, 2009.

Colleen McCue. *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*. Elsevier, Amsterdam Heidelburg, 2007.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*. Academic Press, New York, 1974.

Nigel Meade. Evidence for the selection of forecasting methods. *Journal of Forecasting*, 19, 2000.

James W. Meeker, Bryan Vila, and Katie J.B. Parsons. Developing a GIS-based regional gang incident tracking system. In *Responding to Gangs: Evaluation and Research*. National Institute of Justice, Washington, 2002.

P. S. Mitchell. Optimal selection of police patrol beats. *The Journal of Criminal Law*, 63(4):577–584, 1972.

George Mohler and Martin B. Short. Geographic profiling from kinetic models of criminal behavior. *SIAM Journal of Applied Mathematics*, 2012.

G.O. Mohler, M.B. Short, P.J. Brantingham, F.P. Schoenberg, and G.E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 2011.

John A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7, 1965.

Mike O'Leary. The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, 6(3):253–265, 2009.

Mike O'Leary. Implementing a bayesian approach to criminal geographic profiling. In *Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application*, pages 59:1–59:8, New York, NY, USA, 2010. ACM.

Andreas M. Olligschlaeger. Artificial neural networks and crime mapping. In D. Wisburd and J.T. McEwen, editors, *Crime Mapping and Crime Prevention*, pages 313–347. Criminal Justice Press, Monsey, NY, 1997.

J. M. Ottino. Engineering complex systems. *Nature*, (427), 2004.

Derek Paulsen. Human versus machine: A comparison of the accuracy of geographic profiling methods. *Journal of Investigative Psychology and Offender Profiling*, 3, 2006a.

Derek Paulsen. Connecting the dots: Assessing the accuracy of geographic profiling software. *Policing: An International Journal of Police Strategies & Management*, 29(2), 2006b.

Derek J. Paulsen, Sean Bair, and Dan Helms. *Tactical Crime Analysis: Research and Investigation*. CRC Press, Boca Raton, Florida, 2010.

John V. Pepper. Forecasting crime: A city level analysis. Sep 2007.

Nigel E. Raine, D. Kim Rossmo, and Steven C Le Comber. Geographic profiling applied to testing models of bumble-bee foraging. *Journal of the Royal Statistical Society*, 6(2), 2009.

Jerry Ratcliffe. A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquincy*, 43:261–291, 2006.

Jerry H. Ratcliffe and Travis A. Taniguchi. Is crime higher around drug-gang street corners? Two spatial approaches to the relationship between gang set spaces and local crime levels. *Crime Patterns and Analysis*, 1(1):23–45, 2008.

George Rengert. The journey to crime. In *Punishment, Places, and Perpetrators: Developments in Criminology and Criminal Justice Research*. Willan Publishing, Cullompton, UK, 2004.

Tom Rich and Michael Shively. A methodology for evaluating geographic profiling software. Technical report, Cambridge, MA, 2004.

Aaron Richardson. Albemarle moving to a geo-policing model. *The Daily Progress*, June 2012.

D. Kim Rossmo. *Geographic Profiling*. CRC Press, 2000.

D. Kim Rossmo and Sacha Rombouts. Geographic profiling. In *Environmental Criminology and Crime Analysis*. Willan Publishing, New York, 2008.

Kim Rossmo, Ian Laverty, and Brad Moore. Geographic profiling for serial crime investigation. In *Geographic Information Systems and Crime Analysis*. Idea Group Publishing, Hershey, PA, 2005.

Barry Rowlingson, Peter Diggle, adapted, packaged for R by Roger Bivand, pcp functions by Giovanni Petris, and goodness of fit by Stephen Eglen. *splancs: Spatial and*

Space-Time Point Pattern Analysis, 2012. URL `http://CRAN.R-project.org/package=splancs`. R package version 2.01-31.

Stephen Sacks. Evaluation of police patrol patterns. Technical report, Economics Working Papers, University of Connecticut, 2003.

S.J. Sheather and M.C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, 53(3), 1991.

Walter Andrew Shewhart. *Statistical Method from the Viewpoint of Quality Control*. Courier Dover Publications, 1939.

E. Shlifer and R. W. Wolff. Aggregation and proration in forecasting. *Management Science*, 25(6):pp. 594–603, 1979.

Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications*. Springer, New York, 2006.

Michael A. Smith and Donald E. Brown. Application of discrete choice analysis to attack point patterns. *Information Systems and e-Business Management*, 5(3), June 2007.

Brent Snook, Michele Zito, Craig Bennell, and Paul J. Taylor. On the Complexity and Accuracy of Geographic Profiling Strategies. *Journal of Quantitative Criminology*, 21:1–26, 2005.

Anne Soukhanov. *Encarta World English Dictionary*. St. Martin's Press, 1999.

K.A. Spackman. Signal detection theory: Valuable tools for evaluating inductive learning. In *Proceedings of the Sixth International Workshop on Machine Learning*, San Mateo, CA, 1989. Morgan Kaufman.

Mark D. Stevenson, D. Kim Rossmo, Robert J. Knell, and Steven C. Le Comber. Geographic profiling as a novel spatial tool for targeting the control of invasive species. *Ecography*, 35(8):704–715, 2012.

Joseph Szakas, Christian Trefftz, Raul Ramirez, and Eric Jefferis. Development of an intelligent patrol routing system using GIS and computer simulations. In Lin Liu and John Eck, editors, *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*. Information Science Reference, New York, 2008.

Travis A. Taniguchi, Jerry H. Ratcliffe, and Ralph B. Taylor. Gang set space, drug markets, and crime: Violent and property crimes around drug corners in camden. *Journal of Research in Crime and Delinquency*, 2011.

Frederick Winslow Taylor. *Shop Management*. Harper and Brothers Publishers, New York, 1912.

Frederick Winslow Taylor. *The Principles of Scientific Management*. Harper and Brothers Publishers, 1917.

P. E. Taylor and S. J. Huxley. A break from tradition for the San Francisco Police: Patrol officer scheduling using an optimization-based decision support system. *Interfaces*, 19(1):4–24, 1989.

Frederic Thrasher. *The Gang*. University of Chicago Press, Chicago, 1927.

Harry Timmermans and Reginald G Golledge. Applications of behavioural research on spatial problems II: preference and choice. *Progress in human geography*, 14(3): 311–354, 1990.

George Tita and Greg Ridgeway. The impact of gang formation on local patterns of crime. *Journal of Research in Crime and Delinquincy*, 44(2), 2007.

George E. Tita and Jacqueline Cohen. Measuring spatial diffusion of shots fired activity across city neighborhoods. In *Spatially Integrated Social Science*. Oxford Press, New York, 2004.

George E. Tita, Jacqueline Cohen, and John Engberg. An ecological study of the location of gang set space. *Social Problems*, 52:272–299, 2005.

D.W. Trigg. Monitoring a forecasting system. *Operational Research Quarterly*, 15, 1964.

U.S. Army. *Field Manual 3-24: Counterinsurgency*. 2006.

U.S. Army. *Field Manual 3-0: Operations*. Department of the Army, Washington, D.C., 2008.

U.S. Army. *Field Manual 3-24.2: Tactics in Counterinsurgency*. Department of the Army, Washington, D.C., 2009.

U.S. Army. *Field Manual 1-02: Operational Terms and Graphics*. Department of the Army, Washington, D.C., 2010.

Carl von Clausewitz. *On War*. Princeton University Press, Princeton, 1984.

R.J. Vovurka, B.E. Flores, and S.L. Pearce. Automatic feature identification and graphical support in rule-based forecasting: A comparison. *International Journal of Forecasting*, 12, 1996.

Paul Waddell and Terry Moore. Forecasting demand for urban land. In JohnM. Marzluff, Eric Shulenberger, Wilfried Endlicher, Marina Alberti, Gordon Bradley, Clare Ryan, Ute Simon, and Craig ZumBrunnen, editors, *Urban Ecology*, pages 493–518. Springer US, 2008.

M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman and Hall, 2004.

X. Wang. Spatial adaptive crime event simulation with the RA/CA/ABM computational laboratory. Technical report, University of Cincinnati, 2005.

Xiaofeng Wang and Donald E. Brown. The spatio-temporal modeling for criminal incidents. *Security Informatics*, 2012.

Xuguang Wang, Lin Liu, and John Eck. Crime simulation using GIS and artificial intelligent agents. In Lin Liu and John Eck, editors, *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*. Information Science Reference, New York, 2008.

David Weisburd, Stephen D. Mastrofski, Rosann Greenspan, and James J. Willis. The growth of Compstat in American policing. Technical report, Police Foundation, 2004.

William E. Welch and Michael J. Bussiere. Innovations in policing: Geo beat policing in Lewiston. *The Police Chief*, January 2005.

William Foote Whyte. *Street Corner Society*. University of Chicago Press, Chicago, 1937.

Handik Widiarta, S. Viswanathan, and Rajesh Piplani. On the effectiveness of top-down strategy for forecasting autoregressive demands. *Naval Research Logistics (NRL)*, 54(2):176–188, 2007.

Orlando W. Wilson. *Police Administration*. McGraw-Hill, New York, 1963.

P.R. Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 1960.

Shengnan Wu, Larry Shuman, and Bopaya Bidanda. Agent-based discrete event simulation modeling for disaster response. In *Proceedings of the 2008 Industrial Engineering Research Conference*, 2008.

Yifei Xue and Donald E. Brown. A decision model for spatial site selection by criminals: A foundation for law enforcement decision support. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 33(1):78–85, 2003.

Yifei Xue and Donald E. Brown. Spatial analysis with preference specification of latent decision makers for criminal event prediction. *Decision Support Systems*, 41 (3), 2006.

Yue Zhang and Donald E. Brown. Police patrol district design using agent-based simulation and GIS. In *Proceedings of the 2012 International Conference on Intelligence and Security Informatics*. Institute of Electrical and Electronics Engineers, Inc., 2012.

Yue Zhang and Donald E. Brown. Police patrol districting method and simulation evaluation using agent-based model and GIS. *Security Informatics*, 2(7), 2013.

Yue Zhang, Samuel H. Huddleston, Donald E. Brown, and Gerard P. Learmonth. A comparison of evaluation methods for police patrol district designs. In *Proceedings of the 2013 Winter Simulation Conference*, 2013.