

**THE EFFECTIVENESS OF CALIBRATED VS. DEFAULT
DISTANCE DECAY FUNCTIONS FOR
GEOGRAPHIC PROFILING**

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree

Masters of Arts

by

Karla Emeno

Department of Psychology
Carleton University

September 2008

©2008 Karla Emeno



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-43457-4
Our file Notre référence
ISBN: 978-0-494-43457-4

NOTICE:

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

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

AVIS:

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

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

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

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

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

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

■ ■ ■
Canada

Abstract

Geographic profiling systems use distance decay functions to predict where serial offenders live. These functions assume that the likelihood of an offender residing at a particular location decreases with increasing distance from the offender's crime sites. Currently, each system relies on a default function, which was validated on data that are unrelated to the crimes being submitted to these systems during investigations. This occurs despite the possibility that a single decay function cannot be used with equal effectiveness across data collected under varying conditions. This study determined whether a decay function calibrated for a particular crime type or geographic region resulted in greater predictive power than an uncalibrated, default function. Decay functions were calibrated for three different types of serial crime (residential burglary, theft, and auto theft) collected from the same geographic region (Glendale, AZ) and for serial burglary collected from three different geographic regions (Glendale, AZ, Baltimore, MD, and Dorset, UK). The two default functions (truncated negative exponential and negative exponential) used for comparison purposes came from *CrimeStat* (v 3.1), a computerized geographic profiling system. The hypothesis that calibrated functions would possess more predictive power than the two default functions, as measured by error distance and hit percentage, was not supported to the extent that was expected, with the majority of analyses finding non-significant differences across the various functions within each data file. Potential explanations for these findings are provided, implications are discussed, and directions for future research are presented.

Acknowledgements

First and foremost I would like to thank my thesis supervisor, Craig Bennell, for all of his guidance and encouragement (and patience) throughout these past two years. I know without a doubt that this thesis would not have been possible without Craig's dedicated involvement. Craig was never anything but supportive throughout this entire process (even when yet another deadline had to be "adjusted") and for that I am grateful.

My gratitude also goes out to all of my committee members. I truly do appreciate the time and effort you have all contributed. My thanks go out as well to Michael Leitner, Phil Canter of the Baltimore County Police Department, Ned Levine, Simon Merry of the Dorset Police Service, and Bryan Hill of the Glendale Police Department for providing me with the necessary crime data. I would also like to thank Ned Levine for all of his help with procedural issues.

I would also like to thank all of the other graduate students in Craig's lab, Shevaun, Alyssa, Tamara, and Rebecca. I knew that I could always count on all of you to be there for me. You will never know how much your support has meant to me. I would especially like to thank Tamara who began the programme with me. Your presence in the lab meant that there was always someone nearby who understood exactly what I was going through. Without you I know this process would have been much more overwhelming.

Finally, I would like to thank my parents, my sister, Jill, and my friends back home whose support extends much farther back than just over these past two years. I know that I will always be able to count on all of you to be there when I need you most. Mom and Dad, your unconditional love and support have made me who I am today.

Table of Contents

Abstract.....ii

Acknowledgements.....iii

List of Tables.vii

List of Figures.....viii

List of Appendices.....ix

The Effectiveness of Calibrated vs. Default Distance Decay Functions for Geographic Profiling.1

Conducting GP.....2

 Spatial Distribution Strategies.2

 Probability Strategies.....3

 Distance Decay.6

 Marauding Behaviour.8

Explanations for Serial Offender Spatial Behaviour.10

 Routine Activity Theory.11

 Rational Choice Theory.....13

Accuracy of GP Systems.14

 Measures of Accuracy.....14

 Error Distance.14

 Hit Percentage.....15

 Evaluation Research.....16

Improving the Accuracy of GP.....19

 Offender Characteristics.21

Offence Characteristics.....	21
Crime Type.....	22
Environmental Factors.....	22
The Current Study.....	23
Method.....	24
Data.....	24
Glendale, AZ.....	24
Baltimore County, MD.....	25
Dorset, UK.....	25
Procedure.....	25
Stage 1: Constructing the Calibrated Distance Decay Functions.....	26
Stage 2: Estimating the Parameters for the Calibrated Functions.....	27
Stage 3: Validating the Calibrated Functions.....	30
Stage 4: The Impact of Calibrated and Uncalibrated Functions on GP	
Accuracy.....	31
Results.....	37
Analysis 1: Crime Type.....	37
Descriptive Analysis.....	37
Calibrated Functions.....	38
Graphic Representations.....	38
Parameters.....	41
Split-Half Validation.....	42
Error Distance and Hit Percentage Results.....	43

Glendale: Burglary.....	46
Glendale: Auto Theft.	47
Glendale: Theft.	47
Analysis 2: Geographic Region.	49
Descriptive Analysis.	49
Calibrated Functions.	49
Graphic Representations.	49
Parameters.....	51
Split-Half Validation.....	51
Error Distance and Hit Percentage Results.....	52
Glendale, AZ: Burglary.	55
Baltimore, MD: Burglary.....	55
Dorset, UK: Burglary.....	56
Discussion.....	57
The Lack of a Calibration Effect.	60
The Lack of any Difference between the Default Functions.	64
Conclusion.....	66
References.....	68
Appendices.....	75

List of Tables

Table 1. A schematic of the comparisons that will be made using the calibrated and uncalibrated functions for Analysis 1 (comparisons across crime type).35

Table 2. A schematic of the comparisons that will be made using the calibrated and uncalibrated functions for Analysis 2 (comparisons across geographic region).....35

Table 3. Descriptive analysis of JTC distances (in miles) for each data file in Analysis 1.....38

Table 4. Parameters and R^2 values for the calibrated functions in Analysis 1.....42

Table 5. Split-half validation results (mean error distance in miles) for Analysis 143

Table 6. Minimum, maximum, and mean error distances (in miles) for Analysis 144

Table 7. Minimum, maximum, and mean hit percentages for Analysis 145

Table 8. Descriptive analyses of JTC distances (in miles) for each data file in Analysis 2.....49

Table 9. Parameters and R^2 values for the calibrated functions in Analysis 2.....51

Table 10. Split-half validation results (mean error distance in miles) for Analysis 252

Table 11. Minimum, maximum, and mean error distances (in miles) for Analysis 253

Table 12. Minimum, maximum, and mean hit percentages for Analysis 254

List of Figures

Figure 1.	Centroid calculation.....	3
Figure 2.	Example of a GP probability surface, which was derived from the application of a negative exponential distance decay function.....	5
Figure 3.	The classic distance decay function with a buffer zone	7
Figure 4.	Common distance decay functions used in the context of GP	8
Figure 5.	Models of marauders and commuters' spatial behaviours	9
Figure 6.	The routine activity space of an offender and a potential victim	12
Figure 7.	Calibrated function for Glendale: Burglary.....	39
Figure 8.	Calibrated function for Glendale: Auto Theft	39
Figure 9.	Calibrated function for Glendale: Theft	40
Figure 10.	Calibrated function for Glendale: All Crime Types.....	40
Figure 11.	Calibrated function for Baltimore: Burglary	50
Figure 12.	Calibrated function for Dorset: Burglary	50

List of Appendices

Appendix A.	Probability Distance Strategies	75
Appendix B.	<i>CrimeStat</i> 's Default Negative Exponential and Truncated Negative Exponential Functions	77

The Effectiveness of Calibrated vs. Default Distance Decay Functions for
Geographic Profiling

Rossmo (2000) describes *geographic profiling* (GP) as “an investigative methodology that uses the locations of a connected series of crimes to determine the most probable area of offender residence” (p. 1). GP is most commonly used in cases of serial murder and rape, but is also used in cases of serial arson, robbery, and bombing. Investigators often rely on GP as a tool for prioritizing potential suspects, with those suspects living closest to the predicted home location being focused on first (Rossmo, 2000). Although there are different strategies for conducting GP, they all rely on the same underlying assumptions that: (1) most serial offenders do not travel far from their home location to commit their crimes and (2) most serial offenders live within the area covered by their criminal activity. When these assumptions are valid, it should be possible to accurately predict where an offender resides based on the locations of his¹ crimes (Rossmo, 2000).

Currently, there are several different computer systems available to conduct GP (e.g., *CrimeStat*, *Dragnet*, *Rigel*). All of these GP systems rely on *default distance decay functions* in order to predict the serial offender’s home location (i.e., functions that mathematically represent the assumption that offenders are more likely to live close to their crimes, as opposed to further away) (Canter, Coffey, Huntley, & Missen, 2000; Levine & Associates, 2007; Rossmo, 2000). Default functions are developed on one specific sample of serial offenders and are then applied in the same way to all cases,

¹ Throughout this thesis, offenders will be referred to using the male pronoun due to the fact that the vast majority of serial offenders are male (e.g., Bourget & Bradford, 1989; Kraemar, Lord, & Heilbrun, 2004). This is not meant to indicate that no female serial offenders exist, only that their numbers pale in comparison to the number of male serial offenders.

regardless of potentially important between-case differences in offender characteristics, offence characteristics, and/or environmental factors. This thesis will assess whether the accuracy of GP systems is improved when distance decay functions are calibrated for the crime type or geographic region they are being applied to; essentially, are GP systems more accurate when they rely on calibrated or default distance decay functions?

Conducting GP

When conducting GP, there are numerous strategies available which, according to Snook, Zito, Bennell, and Taylor (2005) can be broadly classified as either: (1) spatial distribution strategies or (2) probability distance strategies.

Spatial distribution strategies involve using the distribution of crime site locations to calculate a central point, which serves as the offender's predicted residence. Examples of common spatial distribution strategies include the calculation of: (1) the centroid, (2) the centre of the circle, and (3) the centre of minimum distance. For illustrative purposes, Figure 1 provides an example of a centroid calculation, perhaps the most commonly used spatial distribution strategy. In this case, the x - y coordinates for a linked series of crimes are averaged to identify the likely location of the offender's residence.

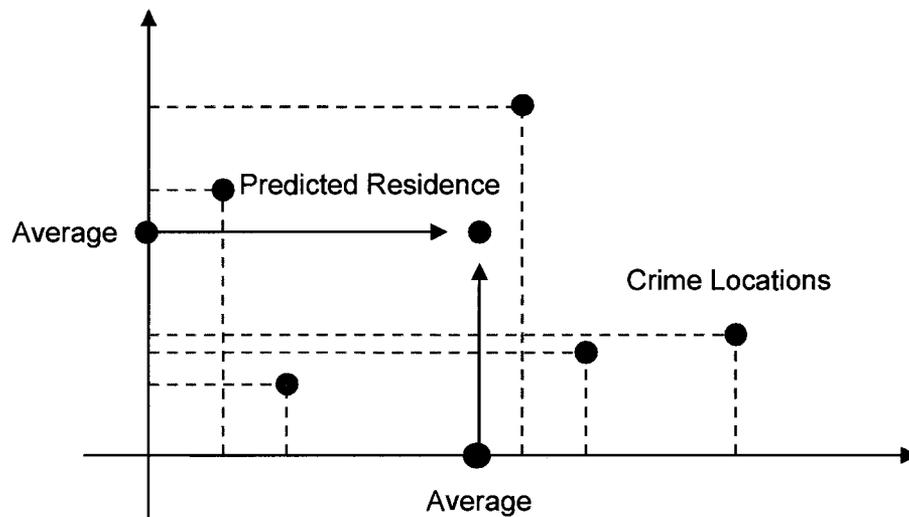


Figure 1. Centroid calculation.

Probability distance strategies, which are used more frequently than spatial distribution strategies, operate on the assumption that crime site locations define an offender's criminal activity space and that their home location is located within this space. In order to predict the location of an offender's residence, some type of mathematical function is applied to each of the crime sites. This mathematical function assigns a small positive number to the areas (i.e., cells) surrounding the crime sites. Once the mathematical function has been applied to all crime site locations, the values within each cell are summed and converted into probabilities in order to produce a *probability surface* that specifies how likely it is that the offender resides in each of the possible areas within their activity space (i.e., each overall cell value indicates the likelihood that the offender resides within that particular cell) (Snook, Zito, et al., 2005). The cells within the probability surface are converted to various colours, or varying shades of the same colour, to produce a sort of temperature map. More specifically, cells of darker

shades indicate a higher likelihood that the offender resides within that area, whereas lighter shades indicate a lower likelihood. This surface can then be searched in a systematic fashion for the offender's residence (i.e., starting at the highest point of probability or the darkest coloured cells and working outward). Examples of common probability distance functions include: (1) linear, (2) lognormal, (3) negative exponential, (4) normal, and (5) truncated negative exponential functions. These functions are defined in Appendix A, along with their mathematical formulae. For illustrative purposes, Figure 2 provides an example of a probability surface produced by the application of a negative exponential distance decay function, which is described in more detail below.

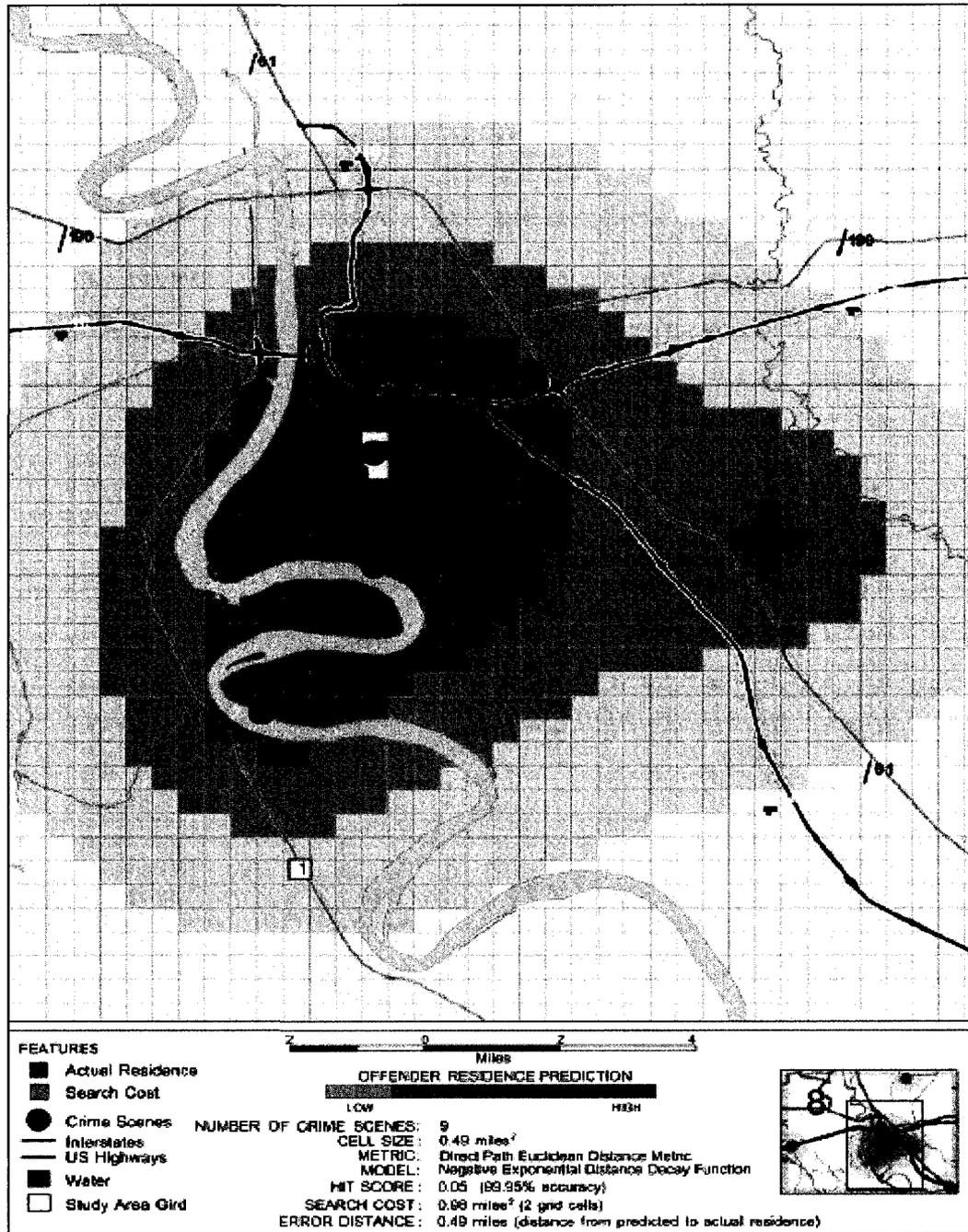


Figure 2. Example of a GP probability surface, which was derived from the application of a negative exponential distance decay function (Source: Kent, Leitner, & Curtis, 2006).

Distance decay. Virtually all of the probability distance strategies currently in use today rely on a similar class of mathematical functions, which are referred to as *distance decay functions* (Rengert, Piquero, & Jones, 1999). Historically, distance decay functions have been used to reflect the fact that an offender tends to commit their crimes closer, rather than further away, from their home location. This tendency results in the frequency of crimes committed by an offender decreasing as the distance from their residence increases (e.g., Canter & Larkin, 1993; Kocsis & Irwin, 1997; Sarangi & Youngs, 2006).

While the distance decay phenomenon is generally accepted as true (see, however, Van Koppen & De Keijser, 1997), the way in which crime decays as an offender moves away from their home is still a matter of much debate. For example, some researchers argue that the likelihood of crimes being committed is highest at the home location and decreases at an exponential rate as the offender travels away from home (reflected in a negative exponential distance decay function) (Canter et al., 2000). However, as can be seen in Figure 3, others argue that there is likely to be an area around the offender's home location where crime is less likely to occur (i.e., a *buffer zone*), and at some optimal distance away from their home crime occurrences begin to decrease in an exponential fashion (reflected in a truncated negative exponential function) (Rossmo, 2000).

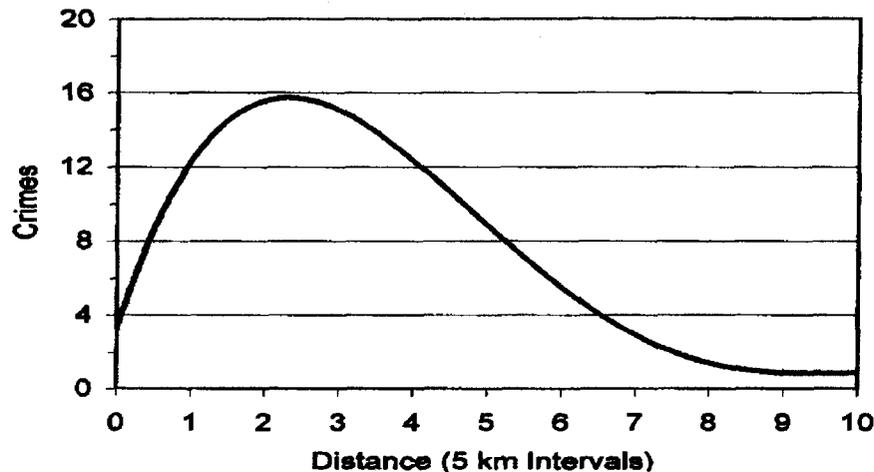


Figure 3. The classic distance decay function with a buffer zone (Source: Rossmo, 2000).

The real breakthrough in the area of geographic profiling came when Rossmo (2000) realized that one can predict where a serial offender lives by simply inverting the traditional distance decay function. Thus, instead of the x -axis reflecting distances from a home location and the y -axis reflecting the likelihood of crimes being committed at those various distances, the x -axis in GP reflects increasing distance from a crime location and the y -axis reflects the likelihood of the offender residing at those various distances. As is the case when considering traditional distance decay functions, the functions are expected to take on different forms depending on the assumptions that are made about an offender's spatial behaviour (Canter & Hammond, 2006). The most common functions used in the context of GP, including the negative exponential function and the truncated negative exponential function that were described above, are presented in Figure 4.

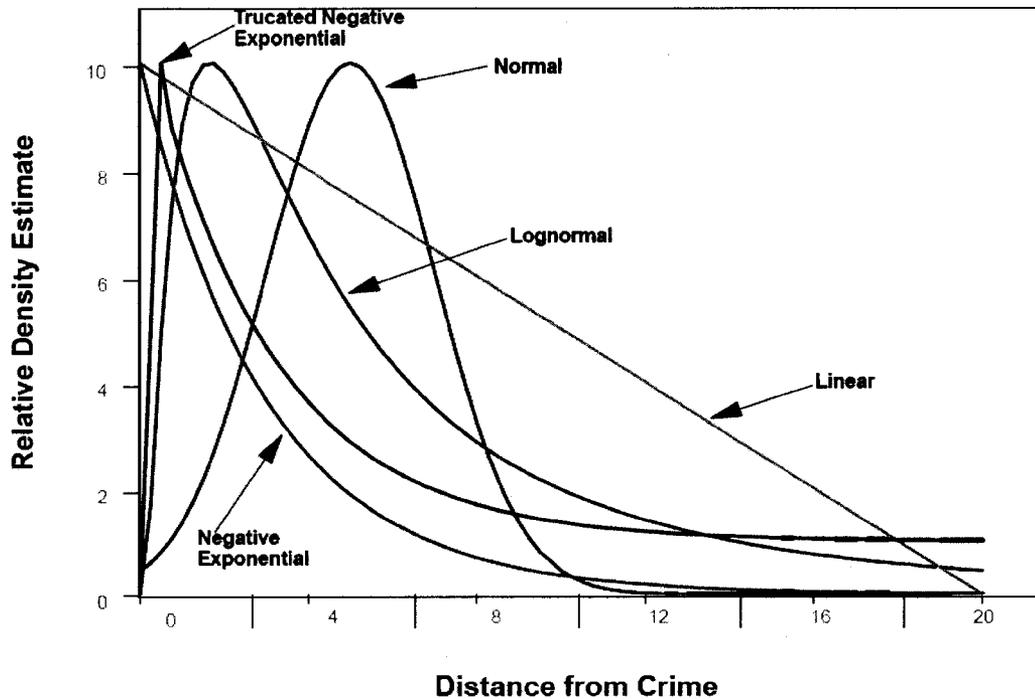


Figure 4. Common distance decay functions used in the context of GP. Note that each type of distance decay function can also vary with respect to the steepness of decay and, where relevant, the size of the buffer zone. (Source: Levine & Associates, 2007).

Marauding behaviour. When any of the mathematical functions in Figure 4 are applied to a series of crimes, and the resulting cell values are summed to produce likelihood estimates, the areas of highest likelihood end up occurring within an offender's criminal activity space. Thus, for this method of GP to result in accurate predictions, offenders must reside within their areas of criminal activity. Research that has examined this issue finds that this is often the case, though the proportion of offenders living central to their crimes varies depending on crime type.

For example, in one of the first studies to examine this issue, Canter and Larkin (1993) examined the proportion of serial rapists in the UK that could be classified as

marauders or *commuters* (see Figure 5). Marauders were described as offenders who move outward from their home base to commit their crimes and then return. Next time these offenders commit a crime, they travel out from their home base in a different direction than the previous crime. In this case, a circle can be drawn around the offender's crimes (with the diameter being defined by the two crimes that are furthest away from each other) and the offender's home would be located in the circle.

Commuters, on the other hand, were described as offenders who travelled away from their home in one particular direction in order to commit their crimes. In this case, the home location of offenders would not be included in a circle that encompasses their crimes.

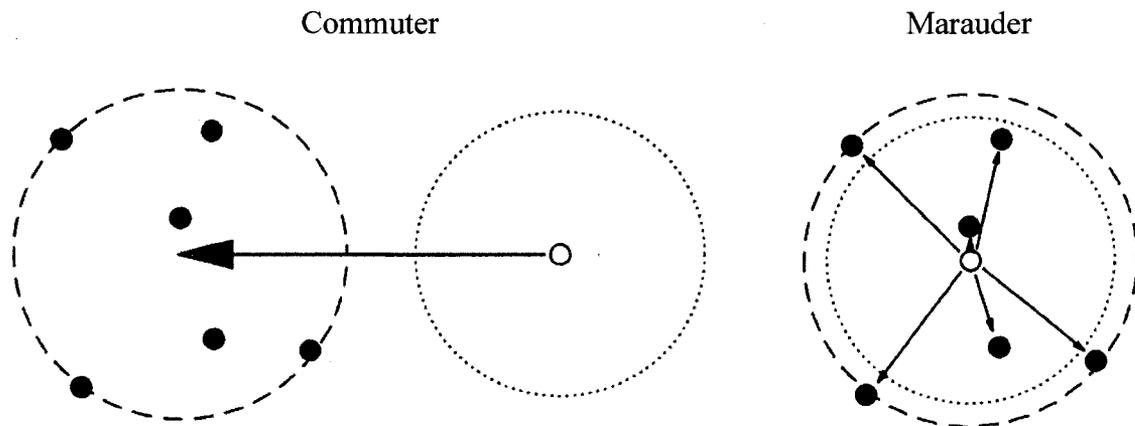


Figure 5. Models of marauders' and commuters' spatial behaviour (Source: Canter & Larkin, 1993). ○ home location; ● crime site locations; home range; ----- criminal range.

Canter and Larkin (1993) found that when circles were drawn around the crimes of 45 serial rapists, 87% were found to fit a marauding pattern, while the remaining 13% were commuters. Other researchers have found similar findings for offenders committing interpersonal crimes (e.g., Meaney, 2004; Santtila, Laukkanen, & Zappala, 2007). In contrast to these findings, when property offenders are examined, a smaller proportion of offenders are found to maraud. For example, Kocsis and Irwin (1997) examined the spatial behaviour of serial criminals and found that only 48% of serial burglars could be classified as marauders. The researchers suggest that motivation for the crime could explain these differences in spatial behaviour across crime types. Specifically, Kocsis and Irwin propose that rapists are more likely to be marauders because they are motivated largely by interpersonal violence with opportunities to commit that violence existing in every direction. Burglars, on the other hand, are equally likely to be marauders or commuters because they are motivated by incentives of profit. Some burglars may see attractive targets in every direction, thus exhibiting marauder behaviour, whereas other burglars may be more selective when choosing a target and will travel farther distances (and in particular directions) for greater expected rewards, thus exhibiting commuter behaviour.

Explanations for Serial Offender Spatial Behaviour

There are several theories available for explaining offender spatial behaviour, specifically the distance decay and marauding phenomenon, and it is these theories that provide the underlying basis for GP strategies. Two of the most commonly cited theories are routine activity theory and rational choice theory. Each of these will now be briefly discussed.

Routine activity theory. First proposed by Cohen and Felson (1979), *routine activity theory* (RAT) argues that, within an environment considered appropriate for criminal activity, crime occurs when a motivated offender's *routine activity space* intersects, in time and space, with a desirable target's routine activity space in the absence of appropriate guardians (e.g., watchful neighbours, police patrols, security devices) (see Figure 6). An offender's/target's activity space consists of the area where the offender/target performs their routine daily activities (e.g., live, work, shop, etc.). This space is found within the offender's/target's *awareness space*, which includes those areas that the offender/target habitually travels to (i.e., their routine activity space) as well as those areas that the offender/target has above minimal knowledge of, even if the offender/target has not visited those areas in the past (Brantingham & Brantingham, 1984).

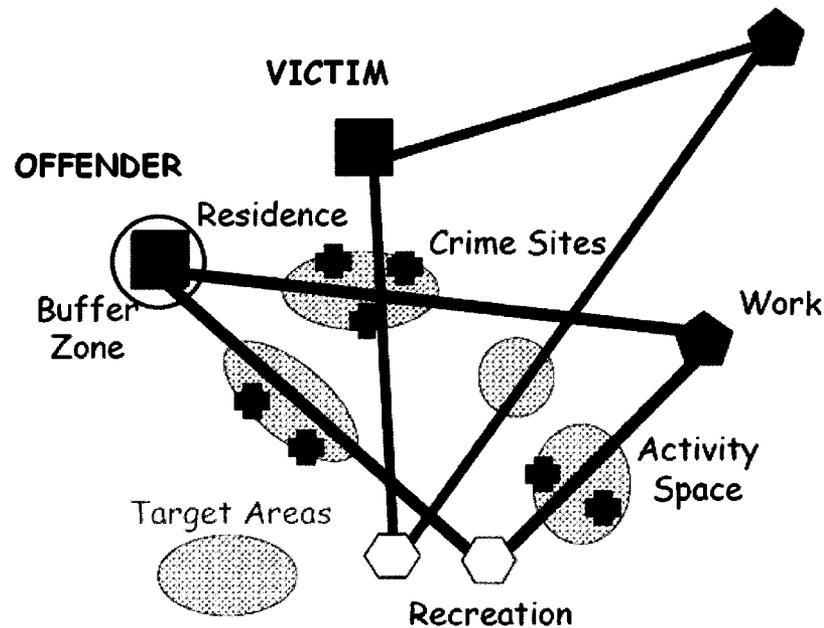


Figure 6. The routine activity space of an offender and a potential victim. According to RAT, crime is likely to occur in locations where the routine activities of offenders and victims overlap when there is an absence of appropriate guardians.

Research on RAT indicates that the awareness space of offenders and targets tends to be anchored to their home locations (Brantingham & Brantingham, 1984; Canter & Larkin, 1993; Mustaine & Tewksbury, 1999; Roncek & Maier, 1991). Given the important role that the home plays in all of our lives, this should not be all that surprising. Nor should it be a surprise, if one takes an RAT perspective, that crimes are often spread out around an offender's home (i.e., marauding behaviour) and at a relatively short distance from their home (i.e., distance decay). In fact, when combined with other common principles of human activity, these findings almost become inevitable. For example, the commonly cited *least effort principle*, which states that humans will choose

actions requiring the least amount of effort when presented with various options of equal reward (Reber, 1985), leads to the expectation that offenders will select victims/targets that are closest to them geographically when presented with multiple victims/targets of equal desirability who are distributed around the offender's home location.

Rational choice theory. According to *rational choice theory* (RCT), criminal behaviours, including their spatial behaviour, is the outcome of choices that an offender makes based on the rational consideration of the efforts, rewards, and costs involved in the criminal behaviour (Cornish & Clarke, 1986). By rational, it is not being implied that offenders are *fully rational* (i.e., processing and weighing all available information), or necessarily even intelligent or sophisticated (Rossmo, 2000). Instead, offenders are expected to go through a process of *bounded rationality* when making decisions (Simon, 1957). Specifically:

[RCT assumes] that offenders seek to benefit themselves by their criminal behaviour; that this involves the making of decisions and of choices, however rudimentary on occasion these processes might be; and that these processes exhibit a measure of rationality, albeit constrained by limits of time and ability and the availability of relevant information. (Cornish & Clarke, 1986, p. 1)

With respect to the decision of where to commit crimes, RCT can be seen to converge with RAT (Canter & Hodge, 1997). For example, in general, offenders will commit crimes in an area that is anchored to their home location, which is predicted from RAT, but the specific location of crime sites within this general area seems to be governed by a decision-making process, as predicted by RCT. For example, as indicated above in the discussion of the truncated negative exponential function, several

researchers have identified the existence of buffer zones when modelling offender spatial behaviour, which refers to a relatively small area centred around the offender's home location where he is considered less likely to commit an offence (e.g., Canter & Hodge, 1997; Canter & Larkin, 1993; Rossmo, 2000). Buffer zones are thought to exist because the victims/targets in these areas are perceived as less desirable due to a *rational consideration*, on the offender's part, of the perceived likelihood that the offender will be recognized by neighbours during the commission of his crime (Rossmo, 2000).

Accuracy of GP Systems

Currently, the most common way of conducting GP is to use computerized systems that rely on one or more of the probability distance strategies outlined above (see Figure 4). Given that these systems are firmly based in the principles of RAT and RCT there has always been the expectation that they would be able to make accurate predictions of where offenders reside based on the locations of their crime sites. However, only recently have serious efforts been made to empirically examine the accuracy of GP systems.

Measures of accuracy. The first step in conducting research that assesses the accuracy of GP systems is determining specifically how accuracy should be measured. Although there are numerous methods for measuring GP accuracy (Rich & Shively, 2004), the two most common ones are error distance and hit percentage. There are different advantages and disadvantages associated with each of these measures.

Error distance, one of the less demanding means of assessing the accuracy of GP systems, involves measuring the distance between the offender's predicted home location (which is produced by the GP system) and his actual residence. The error distance, being

fairly simplistic in nature, can be easily calculated and readily applied to all methods of GP (i.e., both spatial distribution strategies and probability distance strategies). As a result of these advantages, a number of researchers have opted to use this measure in their studies of GP accuracy (e.g., Bennell, Snook, Taylor, Corey, & Keyton, 2007; Paulsen, 2006a; Snook, Canter, & Bennell, 2002; Snook, Taylor, & Bennell, 2004; Snook, Zito, et al. 2005) and it has been recommended for use in a recent National Institute of Justice (NIJ) funded proposal of how to evaluate GP systems (Rich & Shively, 2004).

However, error distance is also associated with a number of potentially important disadvantages and it has recently been criticized (e.g., Rossmo, 2005a). Perhaps the most common and valid criticism is that error distance focuses on one single location or point of prediction, rather than on an area that can be searched, as is produced by GP systems. As Rossmo (2005a) states:

[GP systems] produce probability surfaces that look like coloured topographic maps. These outline the optimal search strategy. Simplifying this to a single point from which error distances are measured is not valid as the search function is rarely uniformly concentric. In effect, the use of error distance undermines the very mechanics of how the geographic profiling process works. (p. 653)

Another common criticism is that error distance is rarely standardized across studies, making comparisons impossible (Rossmo, 2005b). For example, sometimes error distance is measured in mm (e.g., Snook et al., 2002) and sometimes in km (e.g., Snook, Zito, et al., 2005).

Hit percentage refers to the percentage of the total prioritized search area (produced by the GP system) that has to be searched (when working from the highest to

the lowest probability point) before the offender's home base is located. Thus, GP systems are more accurate at predicting the offender's residence as the hit percentage decreases. Hit percentage is generally considered the best measure of GP accuracy because it indicates how much, on average, the geographic profile reduces the area that needs to be searched by the police before finding the offender's residence. Because of these advantages, this measure has been used in several evaluations of GP accuracy (e.g., Canter et al., 2000; Rossmo, 2000). However, this measure is limited because computer software is required for its calculation and it cannot be easily used to evaluate spatial distribution strategies that do not produce search areas (i.e., strategies that result in a single prediction point) (Snook, Zito, et al., 2005).

Evaluation research. A range of research has been conducted to assess the accuracy of various GP approaches. Some of this research has examined the accuracy of GP systems using various types of probability distance strategies. For example, Rossmo (2000) used information from selected FBI serial murder cases to evaluate *Rigel*, which relies on a truncated negative exponential function.² He found that *Rigel* achieved a mean hit percentage of 6% (i.e., on average, only 6% of the total prioritized search area had to be searched before the offender's home location was found). The hit percentages across all of the serial murder cases he examined ranged from 1.1% to 17.8%, with a median of 4.2%.

In a similar study to Rossmo's (2000), Canter et al. (2000) examined the GP system *Dragnet*. Unlike Rossmo, Canter et al. empirically examined a range of distance decay functions to determine which produced the most accurate GP predictions. Using

² The exact parameters for the truncated negative exponential function used in *Rigel* are not publicly available.

the body disposal locations of 79 American serial killers who had each committed a series of 2 to 24 crimes, they found that several negative exponential functions were capable of producing accurate results (more accurate than the truncated negative exponential function) with the most accurate parameters resulting in an average hit percentage across the sample of 11% (i.e., on average, 11% of the total prioritized search area needed to be searched in order to locate the cell containing the offender's home location).³ The slightly higher average hit percentage in this study, as compared to Rossmo's study, can be attributed to the fact that Canter et al. used a random sample of crimes, whereas Rossmo specifically selected crimes that were more suitable for GP (i.e., commuters were excluded from Rossmo's analysis).

More recently, several attempts have been made to compare the simpler spatial distribution strategies to the more complex probability distance strategies. For example, Snook, Zito, et al. (2005) rated 11 different GP strategies (six spatial distribution and five probability distance strategies) in terms of their complexity (i.e., how many steps were required to perform each method) and assessed their accuracy (using error distance) on GP tasks involving 16 UK burglars who had each committed a minimum of 10 crimes. They found that strategy complexity was not positively correlated with accuracy, which suggests that complex GP strategies are not necessarily better than simpler strategies. In fact, the most accurate GP strategy out of the 11 that were examined was the calculation of the centre of the circle.

In a similar study, Paulsen (2006a) compared several GP systems (*Rigel*, *Dragnet*, *CrimeStat*) to spatial distribution strategies using crime series of various types. In

³ Like *Rigel*, the exact parameters for the negative exponential function currently used in *Dragnet* are not publicly available.

contrast to the study by Snook, Zito, et al. (2005), Paulsen tested four different accuracy measures: (1) dichotomous profile accuracy, (2) simple error measurement, (3) profile error distance, and (4) relative size of the top profile area.⁴ Across all four accuracy measures, Paulsen found that the more complex probability strategies were not substantially more accurate, and were sometimes even less accurate, than the simple spatial distribution strategies. Type of crime was found to influence the accuracy of the GP profile, regardless of what GP profiling strategy was used, with certain crime types (auto theft, street robbery, and residential burglary) yielding substantially more accurate results than other crime types (commercial robbery and larceny).

Finally, a series of studies have been conducted to examine whether clinical (i.e., human-based) forms of GP are as effective as more complex (i.e., computer-based) forms of GP (e.g., Bennell, Snook, et al., 2007; Bennell, Taylor, & Snook, 2007; Paulsen, 2006b; Snook et al., 2002; Snook et al., 2004). Snook et al. (2002), for example, presented participants with 10 maps containing the crime site locations (five crimes per map) of different serial offenders. Each participant was asked to predict where they thought each serial offender lived by marking an “X” on each map. After making their predictions, the experimental group was presented with two heuristics: (1) the distance-decay heuristic and (2) the circle heuristic. The distance-decay heuristic stated that “the majority of offenders commit their offences close to home,” whereas the circle heuristic stated that “the majority of offenders’ home locations are contained within a circle that

⁴ The measures used by Paulsen (2006a) can be defined as follows: (1) dichotomous profile accuracy: a dichotomous determination (i.e., yes or no) of whether the top profile area included the offender’s home location (since spatial distribution strategies produce only a single point, rather than an area, this measure assessed whether or not the offender’s residence was within a one mile diameter circle around that point), (2) simple error measurement: the direct (i.e., crows-flight) distance between the predicted and actual home location of the offender, (3) profile error distance: the direct distance between the offender’s actual home location and the nearest part of the top profile area, and (4) relative size of the top profile area: the relative size of the top profile area in relation to the total offence domain.

has a diameter defined by the offender's two furthest crime sites." The control group was simply provided with a break. Both groups were then asked to make 10 more GP predictions. The results indicated that the accuracy of the experimental group, but not the control group, improved following training and their accuracy was not significantly different from that of *Dragnet*. Similar results have been found in more recent studies (e.g., Bennell, Snook, et al., 2007; Paulsen, 2006b; Snook et al., 2004).

To summarize, research into the effectiveness of GP strategies indicates that simple strategies, such as human-based judgments and spatial distribution methods, are just as accurate as more complex probability distance strategies. Given that simple GP methods are much more cost-effective and easy to implement than complex methods, this has led some to question the need for GP systems (Bennell, Taylor, et al., 2007). However, it is possible that modifications to GP systems might make them more effective, thus increasing their usefulness. A number of potential strategies for accomplishing this goal will now briefly be discussed, with an emphasis on the development of calibrated distance decay functions.

Improving the Accuracy of GP

One particular strategy for improving the accuracy of GP is to identify the conditions under which GP will be most effective and then only apply GP in those circumstances. This strategy is beginning to be explored, but the results are inconclusive at present. Perhaps the most commonly discussed factor is whether offenders are marauders or commuters (e.g., Paulsen, 2007). As discussed above, GP will be more effective when applied to marauders versus commuters because the implementation of probability distance strategies in GP systems will always lead to the prediction that serial

offenders live roughly in the centre of their crimes. Given this fact, it might be useful to restrict the use of GP to cases of marauders. Unfortunately, there is currently no reliable method in place for determining whether an offender is a marauder or a commuter at the time of the investigation (Snook, Taylor, et al., 2005). Indeed, this problem has proved problematic recently in cases where GP has been unsuccessfully applied to crimes committed by commuting offenders (e.g., the Washington Snipers; Canter, 2003).

A different solution, and the one that will be focused on in the current thesis, is to develop calibrated distance decay functions for use in GP. As highlighted previously, GP systems currently rely on default distance decay functions in order to predict an offender's home location. In the majority of cases, these functions have been developed on samples of US serial killers (Canter et al., 2000; Rossmo, 2000). Rarely, however, is the use of these functions restricted to the construction of geographic profiles in cases involving US serial killers. Indeed, these functions have been applied in countries around the world, in cases involving virtually every type of crime (e.g., Canter, 2003; Rossmo, 2000; Santtila, Zappala, Laukkanen, & Picozzi, 2003; Trainor, 2006).

As indicated above, the use of these functions assumes that an offender is more likely to commit crimes closer, rather than further away, from his residence. While this assumption may be generally valid, the degree to which the assumption generalizes across situations is currently unknown. In fact, there are several factors that have the potential to influence these functions (e.g., with respect to their shape and steepness). By incorporating some of these factors into geographic profiles (e.g., by creating distance decay functions that are calibrated to different situations) it may be possible to improve their accuracy.

One set of factors that may be important to consider relate to *offender characteristics*. Indeed, a wide range of characteristics, including the offender's age, race, and gender, have all been shown to affect the spatial behaviour of serial offenders (e.g., Baldwin & Bottoms, 1976; Gabor & Gottheil, 1984; Warren, Reboussin, & Hazelwood, 1995).⁵ In one recent study, for example, Snook (2004) found that the age of burglars in St. John's, Newfoundland influenced journey-to-crime (JTC) distances, with younger burglars travelling significantly shorter distances than older burglars. This was attributed primarily to the fact that younger offenders in their sample had more limited access to a vehicle compared to older offenders. In another study by Snook, Cullen, Mokros, and Harbort (2005) that focused on German serial killers, the IQ of offenders was found to correlate negatively with JTC distances.

Offence characteristics also appear to play a role in determining how far offenders are likely to travel to commit their crimes. For example, a common finding in the research literature is that the monetary amount of valuables stolen in property crimes relates to JTC distances, with larger distances being associated with higher values (Baldwin & Bottoms, 1976; Gabor & Gottheil, 1984; Laukkanen, Santtila, Jern, & Sandnabba, 2008; Snook, 2004). JTC distances have also been found to differ depending on the method that offenders use to approach their victims in cases of rape (LeBeau, 1987). Even the venue where crimes occur can influence JTC distances. Canter and Gregory (1994), for instance, found that serial rapists who committed their crimes in outside venues travelled 2.7 times further, on average, than rapists who chose inside venues.

⁵ This research indicates that older offenders generally travel farther distances than younger offenders, white offenders generally travel farther distances than black offenders, and male offenders generally travel farther distances than female offenders.

As alluded to already, research indicates that offenders committing different *types of crimes* also travel in different fashions (e.g., Baldwin & Bottoms, 1976; Gabor & Gottheil, 1984; LeBeau, 1987). For example, Rhodes and Conly (1981) found significant differences in JTC distances between robbers, burglars, and rapists in their study of offenders from Washington, DC. In a similar study, Rand (1986) reported (the now common finding) that offenders committing interpersonal crimes travelled significantly shorter distances than offenders committing property crimes. Most recently, Santtila et al. (2007) found that, even within crime types, different sub-types of offenders can vary with respect to their JTC distances. Specifically, they found that, when homicides and rapes were examined, offenders travelled longer distances when committing crimes that expressed instrumental aggression (i.e., aggression for ulterior gains) compared to crimes demonstrating expressive aggression (i.e., aggression to harm the victim).

Finally, a range of *environmental factors*, or geographical features, have been found to influence the spatial behaviour of serial offenders. For example, when areas surrounding offenders consisted of small versus large businesses, Rhodes and Conly (1981) found that burglars and robbers generally tended to travel further from their home locations to commit their crimes. In another study, Bennell and Stein (2008) found a significant negative correlation between JTC distance and population density in cases of serial homicide within the US, presumably because potential victims are more readily available at locations that are close to offenders' homes in densely populated areas. Lastly, Capone and Nichols (1976) found that burglars tended to travel further from their home locations when committing crimes in closed spaces rather than when committing crimes in open spaces (e.g., parking lots, streets). The researchers propose that closed

spaces vary greatly in their level of expected rewards, thus burglars are willing to travel farther distances to commit their crimes when the expected rewards are greater. Open space burglaries, on the other hand, are committed based on opportunity and convenience. Since opportunities to commit open space crimes are found everywhere, there is little to attract the burglar to travel farther distances to commit open space crimes.

The Current Study

At present, the most common procedure for constructing geographic profiles is to use default (i.e., uncalibrated) distance decay functions that are incorporated into computerized GP systems. These default functions were developed using samples of crimes/criminals, usually from a particular geographic region, that often bear little resemblance to the situations where the GP systems are ultimately applied. For example, the GP system *Rigel*, which relies on a truncated negative exponential function derived from US serial killer data (Rossmo, 2000), is currently the standard GP system used to profile all crime types (e.g., murder, rape, and burglary) in the UK (Trainor, 2006).

Research from the fields of environmental criminology and psychology clearly suggests that a wide range of factors may limit the extent to which default distance decay functions generalize beyond the samples they were developed from (e.g., offender characteristics, offence characteristics, crime type, and/or environmental features). Thus, the purpose of the current thesis is to test the hypothesis that distance decay functions that are calibrated on, and then applied to, similar types of data will produce more accurate profiles than default functions. In particular, the current study will examine whether decay functions that are calibrated for *crime type* or *geographic region* outperform several types of default functions.

Method

Data

A number of different serial crime databases will be used in this study in order to conduct two different sets of analyses. First, data from Glendale, AZ will be used to examine the accuracy of calibrated versus uncalibrated distance decay functions across three different crime types: (1) residential burglary, (2) auto theft, and (3) theft. Second, data from Glendale, AZ, Baltimore, MD, and Dorset, UK will be used to examine the accuracy of calibrated versus uncalibrated distance decay functions for residential burglaries committed across different geographic regions.

Glendale, Arizona. The Glendale data was provided by the Glendale Police Department for the purpose of conducting this research. The data consists of geo-coded *x-y* coordinates for offence and offender home locations⁶ across three different types of serial crimes: (1) 77 residential burglaries committed by 16 offenders, (2) 53 auto thefts committed by 15 offenders, and (3) 585 thefts committed by 131 offenders. For all three crime types, each offender committed a minimum of three offences, which is the currently accepted definition of a serial crime (Holmes & DeBurger, 1988). All of the crimes in this dataset took place between January 1, 1995 and January 31, 2003.

For the purpose of this thesis, residential burglary is defined as a crime that occurs when an offender unlawfully enters a residential dwelling with the intent to commit a theft or felony (Federal Bureau of Investigation; FBI, 2006). Theft occurs when an offender unlawfully takes or attempts to take property away from another individual, but

⁶ It is now standard practice in most police forces around the world to obtain geo-coded *x-y* coordinates for all offence and offender home locations. This is typically done by the first responder when they visit the scenes through the use of a Global Positioning System (GPS). GPS provides an extremely accurate estimate of location (at the scale of +/- feet usually).

the offender has not unlawfully entered a dwelling in order to commit the theft (e.g., shoplifting) (FBI, 2006). The theft of motor vehicles is excluded from this category (FBI, 2006). Auto theft is defined as a crime that occurs when an offender steals or attempts to steal an automobile, which includes any self-propelled motor vehicle that runs on land, but not on rails (i.e., the definition excludes trains, boats, and airplanes) (FBI, 2006).

Baltimore County, Maryland. The Baltimore data was originally provided to Levine and Associates (1999) by the Baltimore County Police Department for the purpose of developing *CrimeStat*, a software package for analyzing the spatial behaviour of criminals. The data consists of geo-coded *x-y* coordinates for offence and offender home location across 260 residential burglaries (defined as above) committed by 61 offenders. Each offender committed a minimum of three offences. The crimes took place between 1992 and 1997.

Dorset, United Kingdom. The Dorset data was originally collected by Bennell (2002) from the Dorset Police Service for the purpose of carrying out research on linkage analysis (i.e., linking serial crimes committed by the same offender). The data consists of geo-coded *x-y* coordinates for offence and offender home location across 494 residential burglaries (defined as above) committed by 38 offenders. Each offender committed a minimum of five offences (the crime series were capped at five offences per offender in the original research project). The crimes took place between May 1997 and June 1999.

Procedure

Both sets of analyses will follow the same general procedure, which involves four stages: (1) constructing the calibrated distance decay functions, (2) estimating the parameters of those functions, (3) validating the calibrated functions, and (4) assessing

the impact of the calibrated (and uncalibrated) functions on GP accuracy. Each of these stages will now be briefly discussed.

Stage 1: Constructing the calibrated distance decay functions. For both sets of analyses (i.e., comparing distance decay functions across crime types and geographic regions) the first stage of the procedure involved the construction of calibrated distance decay functions. *CrimeStat* (v 3.1) was used to complete this stage of the procedure (Levine & Associates, 2007). As briefly highlighted above, *CrimeStat* is a software program that was developed to analyze the spatial behaviour of criminals. *CrimeStat* allows the user to accomplish a wide range of analytical tasks. However, for this part of the thesis, only the *spatial modelling routine* in *CrimeStat* was used. This routine allows the user to calibrate distance decay functions based on specific datasets. The process of developing calibrated functions involved several steps. These are outlined by Levine and Associates (2007) as follows:

1. Using *Microsoft Excel*, where the data is stored, the Glendale data was sorted into three sub-samples of crimes (residential burglary, theft, and auto theft) and saved as separate files (these files were used for the first set of analyses examining crime type). The other datasets were left intact for the second set of analyses, which examined geographic region.
2. Each data file was examined to ensure that each record included an *x*- and *y*-coordinate for both the offender's home location and the offence location.
3. The offenders in each data file were randomly divided into two separate data files to form a development sample and a test sample. The *development sample* was used to develop a calibrated distance decay function for a

particular crime type or geographic region and the *test sample* was used to validate the calibrated function.

4. The development sample data files, which consisted of *x-y* coordinates for each offender's home and their entire offence series, were read into *CrimeStat*, where a calibrated function was produced. One function was calibrated for each of the three crime types in the Glendale data, and one function was calibrated for each geographic region (Glendale [residential burglary], Baltimore, and Dorset). In addition, the development samples from all three crime types for Glendale were pooled together to form one aggregate development sample and a function was then calibrated for this data file as well.⁷

Stage 2: Estimating the parameters for the calibrated functions. For both sets of analyses, the second stage of the procedure involved estimating the parameters for each calibrated function. Although *CrimeStat* can be used for this purpose, the *nonlinear regression routine* in *SPSS* (v 16.0) is more user friendly than *CrimeStat* and provides more information (e.g., parameter estimates and graphs of predicted values). For these reasons, *SPSS* was relied on for estimating the parameters using the following steps:

1. In order to determine the parameters of the calibrated functions, the data from the development samples were used to create frequency distributions. Specifically, for each development sample, the direct distance (in miles) between the offender's home location and the offence location was calculated

⁷ The three Glendale development samples were pooled and an aggregate function was developed in order to ensure that all possible functions were investigated. Unfortunately, a function could not be constructed by aggregating the development samples of all three geographic regions due to the fact that the search grid (Step 2 in the profile production stage, which is described below) would have encompassed crimes committed across distant areas (i.e., Dorset, Glendale, and Baltimore).

for each record (using *Microsoft Excel*). This was done by applying the following formula to the data:

$$d_{AB} = \sqrt{[(x_A - x_B)^2 + (y_A - y_B)^2]}$$

where d is the distance between the offender's home (A) and offence location (B), x_A and x_B are the x -coordinates of the home and offence location, respectively, and y_A and y_B are the y -coordinates of the home and offence location, respectively.

2. For each development sample, these distances were grouped into distance intervals, or bins. Based on previous research, the size of each of these bins was set at .25 miles (Levine & Associates, 2007). The grouping was accomplished by sorting the distances within each file into ascending order and then assigning a grouping variable to each distance (e.g., a value of 1 was assigned to all distances between 0 miles and .25 miles, a value of 2 was assigned to all distances between .25 miles and .50 miles, etc.).
3. For each development sample, a new file was created that included only the frequency distribution of the distances (i.e., the number of journey-to-crime [JTC] distances belonging in each bin).
4. In order to compare the frequency distributions across the different types of crimes in the Glendale data, and across the different geographic regions represented by the Glendale, Baltimore, and Dorset data files, two new variables needed to be created. First, the frequency of distances in each bin

was converted to a percentage in order to calculate relative frequencies (i.e., the percentage of crimes in a particular dataset that fall within each bin). This was accomplished by dividing the frequency of distances in a particular bin by the total number of offences in the dataset, and multiplying by 100. Second, the distance intervals were adjusted to the mid-point of each bin by adding .125 (half the bin width) to the starting distance of each bin.

5. Based on the frequency distributions, all of the calibrated functions were classified as being either truncated negative exponential or negative exponential. Specifically, if the first bin contained the highest percentage of JTC distances, then the calibrated function was classified as a negative exponential function (no buffer zone). If, however, the bin containing the highest percentage of JTC distances was not the first bin, then the calibrated function was classified as a truncated negative exponential function. The formulae for the negative exponential and truncated negative exponential functions are presented in Appendix A.
6. Within *SPSS*, the *nonlinear regression routine* was then used to determine parameters for each calibrated function. The parameters used in *CrimeStat*'s default truncated negative exponential and negative exponential functions were used as the starting values when running the nonlinear regressions in *SPSS* (starting values are defined as estimates for the value of each parameter; thus, the default parameters from *CrimeStat* were considered to be appropriate starting values). R^2 values were provided as part of the *SPSS* output for each calibrated function as were graphs representing the actual frequency

distributions and the predicted frequency distributions (using the estimated parameters).

Stage 3: Validating the calibrated functions. Once the distance decay functions were calibrated for each development dataset, and the parameters were calculated, the next stage of the procedure involved validating the functions using their corresponding test samples, a method referred to as split-half validation (Efron, 1982).⁸ Specifically, the calibrated functions were applied to data in each of the relevant development samples, and the relevant test samples, and an error distance and hit percentage were calculated for each offender's crime series. While it should be expected that calibrated functions will result in more accurate predictions (smaller error distances and hit percentages) when applied to data in the development samples, significant increases in mean error distance and mean hit percentage when moving from the development sample to the test sample would indicate problems with generalizability.

Not only will this validation procedure indicate whether the function generalizes to crimes that it was not based on, it also provides a way of interpreting the results that will emerge in this thesis when the calibrated functions are applied to totally different samples. Indeed, without carrying out split-half validation, it would not be possible to determine if the results of any tests that are conducted across samples are due to validity

⁸ Another option for validating the various calibrated functions was the leave-one-out procedure (Efron, 1982). For each of the development samples, this validation technique would have involved: (1) removing one offender from the sample, (2) developing a calibrated function based on the remaining offenders, and (3) applying the calibrated function to the removed offender and determining error distance and hit percentage. This procedure would be repeated for each offender in the sample and if the error distances and hit percentages did not vary significantly across the offenders, then a function would be calibrated based upon all of the offenders in the development sample (and this final aggregate function would be considered validated). However, the leave-one-out cross-validation procedure has many drawbacks including: (1) it is time consuming, (2) it is used infrequently, and (3) the final aggregate model has never been validated (i.e., it has never been applied to a sample beyond the one that was used to develop it). Thus, although the split-half validation procedure can be considered inefficient (i.e., development and test samples may end up being fairly small), the costs of this approach were deemed more minimal than the costs associated with the leave-one-out procedure.

issues or the differing characteristics of the samples in question. For example, if the calibrated Baltimore function generalizes to crimes that were not used to develop the function (i.e., the Baltimore test sample), and a significantly higher error distance is found when applying the calibrated Baltimore function to a dataset from another geographic region (i.e., Dorset), then that difference in error distance is likely an indication of the importance of geographic region rather than an indication of over-fitting. If the split-half validation is successful (i.e., no significant differences are found between the two samples in terms of mean error distance and mean hit percentage) then the development and test samples will be pooled for all analyses in Stage 4.

Stage 4: The impact of calibrated and uncalibrated functions on GP accuracy.

The real test of the usefulness of calibrated distance decay functions is to determine whether they ultimately result in the development of more accurate geographic profiles (compared to uncalibrated functions). To examine this issue, *CrimeStat's spatial modelling routine* was used in the fourth stage of the procedure. Using this routine, it was possible to examine the differences in GP accuracy that occurred when using calibrated versus uncalibrated distance decay functions. Again, this process involved several steps, as outlined by Levine and Associates (2007):

1. Data files consisting of x - y coordinates for all of the offenders' homes and crime site locations from each crime type and geographic region were read separately into *CrimeStat* and a rectangular grid was overlaid on top of the area of criminal activity. The size of the grid was specific to each data file. Specifically, for each sample, the x - y coordinates of the most south-west offence location served as the coordinates for the bottom-left corner of the

rectangle, whereas the x - y coordinates of the most north-east offence location served as the coordinates for the rectangle's upper-right corner.

2. *CrimeStat*'s interpolation routine was then run in order to develop a grid for each data file.
3. The interpolation grid could have been saved in a variety of formats. In the current study, the outputs were saved as shape (.shp) files, which can be read by *ArcView* (v 9.2), a desktop geographic information system (GIS). An extension of *ArcView*, known as *ArcMap*, was then used to open the .shp files and calculate the x - y coordinates of the centroid of each cell.
4. The .shp files (including the centroids) were then used within *CrimeStat* to predict the cell in the grid where the offender is most likely to reside based on his crime locations. To accomplish this, *CrimeStat*'s *spatial modelling routine* was used to make predictions for each offender within all three crime type data files and all three geographic region data files. This routine first calculates the distance between each offence location committed by the serial offender and each cell in the grid.⁹ A distance decay function is then applied to each grid cell-offence pair.
5. The user can choose to use a default distance decay function that is already incorporated into *CrimeStat* or they can manually input parameters from an empirically derived function. In the current study, parameters related to all of the empirically derived functions were used, in addition to two default

⁹ Within *CrimeStat*, both direct (i.e., crows-flight) and indirect (i.e., Manhattan) distances can be used. Manhattan distances are only useful if the geographic area where the system is being applied consists of road networks that are based on uniform grids and if offender travel routes are known. In most instances direct distances are more suitable so these will be used in the current thesis.

functions used in *CrimeStat*: (1) the truncated negative exponential function and (2) the negative exponential function. These default functions were selected for inclusion in the current study because they are the most commonly used functions in GP systems. See Appendix B for a summary of *CrimeStat*'s default negative exponential and truncated negative exponential functions.

6. When a distance decay function is applied to an offender's crime series, a positive real number is assigned to the cells or area surrounding each crime site location. These numbers are then added up to produce an overall value for each grid cell. The summing of these numbers results in a probability surface that indicates the likelihood of the offender residing within each particular cell.
7. *CrimeStat* then provided, as output, the error distance (in miles) and hit percentage for each offender in the data file. Error distance was calculated by measuring the direct distance (in miles) between the offender's actual home location and the centroid of the cell predicted by *CrimeStat* to contain the offender's home location (i.e., the grid cell that has the highest likelihood value). Hit percentage refers to the proportion of the entire prioritized search area (defined by the superimposed grid) that needs to be searched before finding the offender's home location. Specifically, *CrimeStat* calculates hit percentage by first ranking the grid cells from highest to lowest likelihood of containing the offender's home location. The cell with the offender's actual home location was then identified and the number of cells that have a

probability equal to or higher than that cell were added up and divided by the total number of cells.

8. The calibrated functions were crossed (i.e., applied to all other crime types and geographic regions) to determine if calibrated functions produce more accurate geographic profiles than uncalibrated functions. A schematic representing the specific function applications is provided in Tables 1 and 2 for Analyses 1 and 2, respectively. For the purpose of this thesis, uncalibrated functions include: (1) functions that have been calibrated, but not using the data to which they are being applied (e.g., applying a function that was calibrated on burglary in Glendale to theft in Glendale, or applying a function that was calibrated on burglary in Glendale to burglary in Dorset) and (2) the two most commonly used default distance decay functions within *CrimeStat* (i.e., truncated negative exponential and negative exponential functions).

Table 1. A schematic of the comparisons that will be made using the calibrated and uncalibrated functions for Analysis 1 (comparisons across crime type).

	Function applied to:		
Function developed using:	Glendale: Burglary	Glendale: Auto	Glendale: Theft
Default: Truncated Neg. Exp.	Uncalibrated	Uncalibrated	Uncalibrated
Default: Neg. Exp.	Uncalibrated	Uncalibrated	Uncalibrated
Glendale: All	Uncalibrated	Uncalibrated	Uncalibrated
Glendale: Burglary	Calibrated	Uncalibrated	Uncalibrated
Glendale: Auto	Uncalibrated	Calibrated	Uncalibrated
Glendale: Theft	Uncalibrated	Uncalibrated	Calibrated

Note. For each cell, the following data was gathered: minimum, maximum, and mean error distances and hit percentages.

Table 2. A schematic of the comparisons that will be made using the calibrated and uncalibrated functions for Analysis 2 (comparisons across geographic region).

	Function applied to:		
Function developed using:	Glendale: Burglary	Baltimore: Burglary	Dorset: Burglary
Default: Truncated Neg. Exp.	Uncalibrated	Uncalibrated	Uncalibrated
Default: Neg. Exp.	Uncalibrated	Uncalibrated	Uncalibrated
Glendale: Burglary	Calibrated	Uncalibrated	Uncalibrated
Baltimore: Burglary	Uncalibrated	Calibrated	Uncalibrated
Dorset: Burglary	Uncalibrated	Uncalibrated	Calibrated

Note. For each cell, the following data was gathered: minimum, maximum, and mean error distances and hit percentages.

9. For each set of analyses (i.e., Analyses 1 and 2), minimum, maximum, and mean error distances and hit percentages were calculated for all of the functions applied to the data file.
10. Repeated measures ANOVAs were conducted in order to determine whether the type of function applied to the data files (represented by different rows in Tables 1 and 2) had an overall effect on mean error distance or mean hit percentage. Thus, error distance and hit percentage served as the dependent variables, whereas the function applied to the data file served as the independent variable. In cases where the overall ANOVA was significant, post hoc analyses were run. It was expected that error distances and hit percentages would be significantly smaller when calibrated functions were used to construct the profiles. A significance level of .05 (with a Bonferroni correction) was adopted for the post hoc follow-up comparisons.
11. A priori contrasts were run that compared the mean error distance and mean hit percentage of the calibrated functions to those of the negative exponential and truncated negative exponential functions. These comparisons were determined to be of the most practical importance because both of these default functions are currently being used in naturalistic settings, and calibration of functions is being proposed as a means of improving on those functions. Again, it was expected that error distances and hit percentages would be significantly smaller when calibrated distance decay functions were used to construct geographic profiles. A significance level of .05 (with a

Bonferroni correction) was also adopted for the planned follow-up comparisons.

12. Finally, given that there is ongoing debate about which default function produces more accurate profiles (e.g., Canter et al., 2002; Rossmo, 2000), the mean error distance and mean hit percentage of the default truncated negative exponential function were compared to those of the default negative exponential function. It is unclear at this stage which function will produce smaller error distances and hit percentages. A significance level of .05 (with a Bonferroni correction) was also adopted for this planned follow-up comparison.

Results

Analysis 1: Crime Type

Descriptive Analysis

To begin, a descriptive analysis was conducted on the three Glendale data files (burglary, auto theft, and theft). For each data file, the number of offenders and crimes was calculated, along with descriptive statistics relating to the JTC distances for each sample. This data is presented in Table 3.

Table 3. Descriptive analysis of JTC distances (in miles) for each data file in Analysis 1.

Sample	Number of Offenders	Number of Crimes	JTC Distances			
			Min	Max	Median	Mean (<i>SD</i>)
Glendale: Burglary	16	77	0	24.35	1.29	2.94 (5.49)
Glendale: Auto Theft	15	53	0	34.36	4.63	9.36 (9.69)
Glendale: Theft	131	585	0	91.10	4.58	8.57 (12.71)

Calibrated Functions

Graphic representations. Figures 7 through 10 are graphic representations of the distance decay functions calibrated for each development sample from Glendale (burglary, auto theft, theft, and all three crime types). The solid line in each figure represents the actual observed relative frequencies for the development sample, whereas the dashed line represents the relative frequencies predicted by the calibrated function (discussed in more detail below). It is important to note that the observed frequencies (represented by the solid line) are only for the data file's development sample (i.e., the sample used to develop the calibrated function).

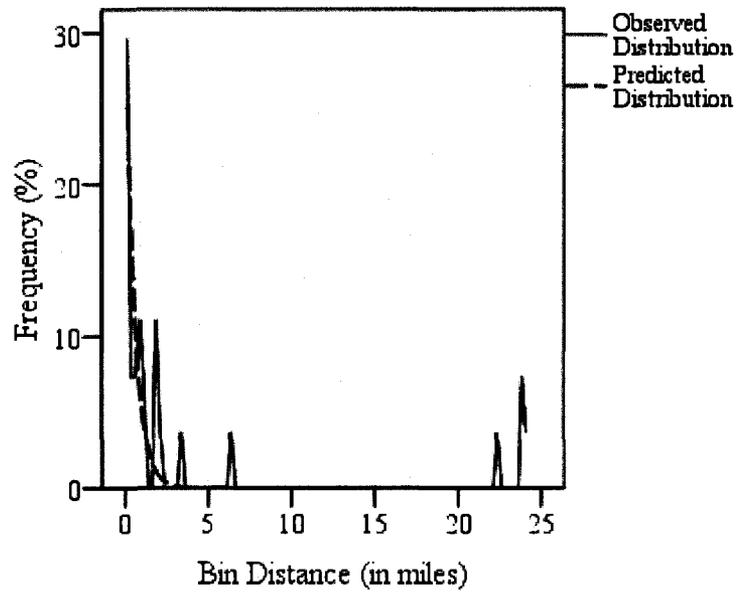


Figure 7. Calibrated function for Glendale: Burglary.

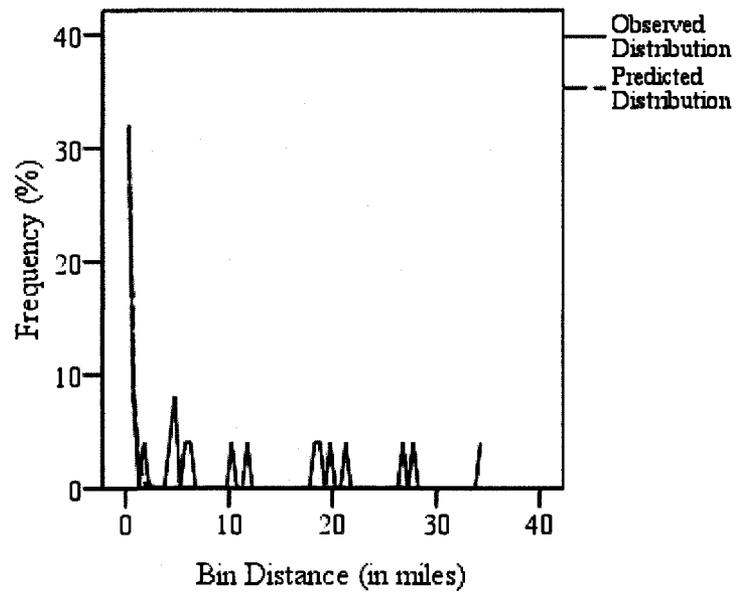


Figure 8. Calibrated function for Glendale: Auto Theft.

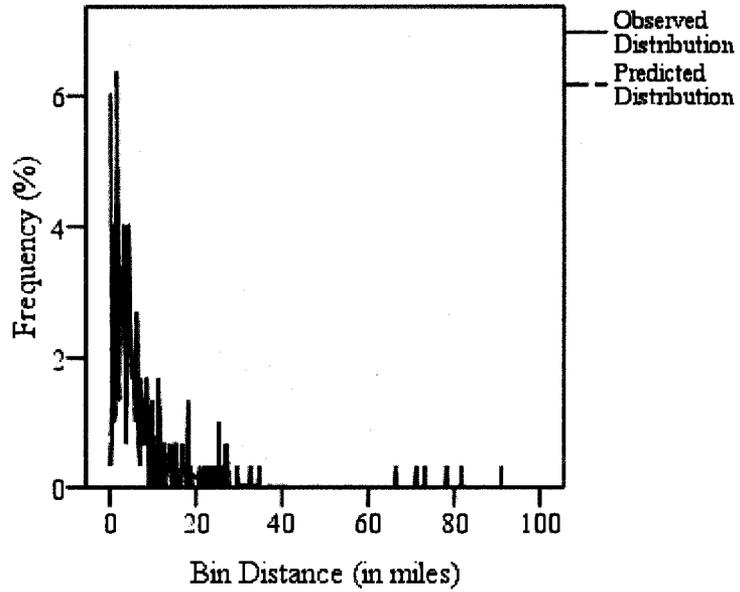


Figure 9. Calibrated function for Glendale: Theft.

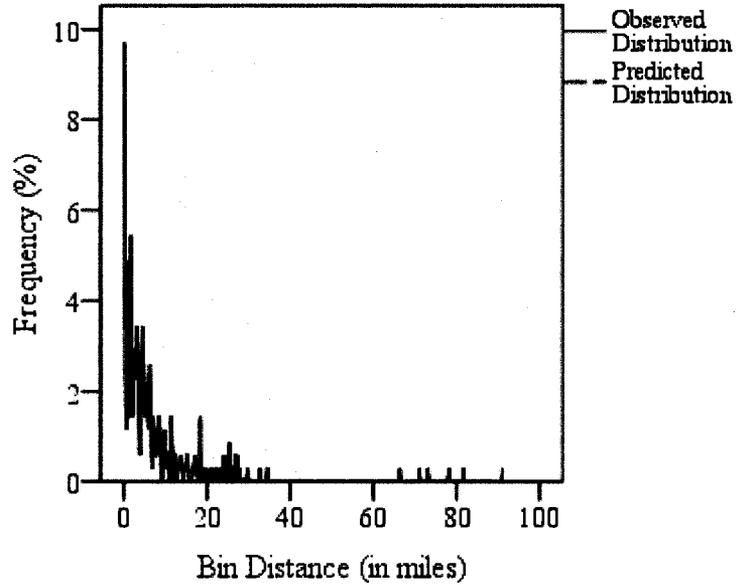


Figure 10. Calibrated function for Glendale: All Crime Types.

Parameters. As previously mentioned, all of the calibrated functions were classified as being either a truncated negative exponential function or a negative exponential function based on the distribution of frequencies across the JTC distance bins. As a reminder, the negative exponential function is:

$$Pct_i = A * e^{-B * d_i}$$

where Pct_i = the percentage of all JTC distances in the data file that fall into interval, i ; e = base of the natural logarithm, d_i = the distance for interval, i ; A = coefficient, and B = exponent. The truncated negative exponential function combines a linear function:

$$Pct_i = C * d_i, \text{ for } 0 \leq d_i \leq d_p$$

and a negative exponential function:

$$Pct_i = A * e^{-B * d_i}, \text{ for } d_i > d_p$$

where Pct_i = the percentage of all JTC distances in the data file that fall into interval, i ; d_i = the distance for interval, i ; C = slope, A = coefficient, B = exponent, and d_p = distance at which peak frequency occurs.

Based on the distribution of frequencies across the JTC distance bins, the data files for burglary, auto theft, and all crime types are best represented by negative exponential functions, whereas the data file for theft is best represented by a truncated negative exponential function. Parameters for each calibrated function and their corresponding R^2 values are presented in Table 4, and the predicted frequencies are presented in Figures 7 through 10 (represented by the dashed line). Based on the R^2 values, all of the calibrated functions fit their corresponding development sample moderately well (R^2 values range from .64 to .78).

Table 4. Parameters and R^2 values for the calibrated functions in Analysis 1.

Dataset	Function	Parameters			Distance of Peak Freq. (in miles)	R^2
		A	B	C		
Glen: All	Neg Exp	4.37	.19	--	--	.68
Glen: Burglary	Neg Exp	31.68	1.85	--	--	.72
Glen: Auto Theft*	Neg Exp	65.19	2.85	--	--	.78
Glen: Theft	Trunc Neg Exp	4.31	.17	2.67	1.625	.64

Note. The parameter notations (A, B, and C) in this table refer to the same parameter notations used in the equations on the previous page. * indicates that the parameters in this row are for .50 mile bins instead of .25 mile bins due to the low number of crimes in the development sample for this data file.

Split-Half Validation

A series of independent samples *t*-tests were run to compare the mean error distances and mean hit percentages from the development samples to the mean error distances and mean hit percentages from the test samples when the calibrated functions (that were developed using the development samples) were applied to both samples. This was done for the three functions calibrated for crime type (i.e., burglary, auto theft, and theft) and the aggregate calibrated function (i.e., all crime types).

The *t*-tests revealed that the mean error distances and mean hit percentages associated with offenders in the development samples were not significantly different from those of the test samples; this was the case for all of the calibrated functions (see Table 5). The non-significant results obtained in the split-half validation indicate that the calibrated functions can be used with sufficient accuracy when being applied to data beyond that which was used to derive them (but still from the same crime type). Since the

split-half validation was successful, the development and test samples were pooled for all further analyses.

Table 5. Split-half validation results (mean error distance in miles) for Analysis 1.¹⁰

Dataset	Development Sample		Test Sample		<i>t</i>	df	Sig
	N/#	Error Distance Mean (<i>SD</i>)	N/#	Error Distance Mean (<i>SD</i>)			
Glendale: All	80/350	6.99 (12.90)	82/365	6.76 (11.57)	.12	160	.91
Glendale: Burglary	8/27	3.59 (8.26)	8/50	2.82 (4.69)	.23	14	.82
Glendale: Auto Theft	7/25	7.69 (10.97)	7/28	7.13 (9.10)	.11	13	.92
Glendale: Theft	65/298	7.38 (13.45)	66/287	8.29 (12.21)	-.41	129	.69

Note. *N* represents the number of offenders in the sample and # represents the number of crimes in the sample.

Error Distance and Hit Percentage Results

Minimum, maximum, and mean error distances (in miles) and hit percentages for Analysis 1 are presented in Tables 6 and 7. One-way repeated measures ANOVAs were conducted in order to determine whether the function had an overall effect on error distance or hit percentage. Error distance and hit percentage served as the dependent variables, whereas the function applied to the data file served as the independent variable.

¹⁰ This table only presents results from the error distance analysis. Results from the hit percentage analysis are not presented here as they were also non-significant.

Table 6. Minimum, maximum, and mean error distances (in miles) for Analysis 1.

Function developed using:		Function applied to:		
		Glendale: Burglary	Glendale: Auto	Glendale: Theft
Trunc. Neg. Exp.	Min.	.03	.40	.32
	Max.	23.47	27.35	81.35
	Mean (SD)	3.03 (6.39)	8.33 (9.42)	7.18 (13.13)
Neg. Exp.	Min.	.05	.09	.04
	Max.	23.83	27.57	81.67
	Mean (SD)	3.10 (6.52)	8.32 (9.50)	7.14 (13.10)
Glendale: All	Min.	.05	.09	.04
	Max.	23.92	27.85	80.59
	Mean (SD)	3.29 (6.47)	7.49 (9.26)	7.24 (12.99)
Glendale: Burglary	Min.	.21	.24	.07
	Max.	24.00	27.98	77.99
	Mean (SD)	3.20 (6.51)	9.33 (9.25)	7.25 (12.94)
Glendale: Auto	Min.	.05	.02	.04
	Max.	23.83	27.71	90.98
	Mean (SD)	3.22 (6.49)	7.39 (9.64)	7.76 (13.73)
Glendale: Theft	Min.	1.09	1.84	.40
	Max.	22.44	28.34	80.59
	Mean (SD)	3.68 (5.37)	9.14 (8.57)	7.84 (12.80)

Note. Shaded cells refer to the results of the calibrated functions.

Table 7. Minimum, maximum, and mean hit percentages for Analysis 1.

Function developed using:		Function applied to:		
		Glendale: Burglary	Glendale: Auto	Glendale: Theft
Trunc. Neg. Exp.	Min.	.004	1.38	.03
	Max.	55.03	81.01	100.00
	Mean (<i>SD</i>)	15.05 (17.71)	26.08 (27.43)	7.74 (16.65)
Neg. Exp.	Min.	.004	.01	.01
	Max.	55.04	80.65	99.61
	Mean (<i>SD</i>)	6.56 (16.61)	17.93 (27.27)	7.29 (17.00)
Glendale: All	Min.	.004	.007	.01
	Max.	48.67	81.77	100.00
	Mean (<i>SD</i>)	6.17 (14.69)	20.03 (31.37)	7.51 (17.52)
Glendale: Burglary	Min.	.05	.04	.01
	Max.	100.00	100.00	100.00
	Mean (<i>SD</i>)	10.83 (28.86)	24.89 (36.67)	19.98 (29.87)
Glendale: Auto	Min.	.004	.004	.01
	Max.	69.52	90.02	100.00
	Mean (<i>SD</i>)	7.74 (19.86)	23.30 (32.56)	18.13 (25.79)
Glendale: Theft	Min.	1.95	.35	.05
	Max.	49.30	78.72	100.00
	Mean (<i>SD</i>)	11.68 (12.96)	22.25 (27.91)	8.23 (17.77)

Note. Shaded cells refer to the results of the calibrated functions.

Glendale: Burglary. The one-way ANOVA revealed that the function applied to the Glendale: Burglary data file did not have a significant overall effect on mean error distance, $F(5, 75) = 1.46, p = .21, \eta^2 = .09$. The planned comparisons also found non-significant differences in mean error distance when applying the calibrated function versus applying either the default negative exponential function, $t(15) = .61, p = .55$, or the default truncated negative exponential function, $t(15) = 1.25, p = .23$. In addition, the mean error distance was not significantly different when applying the negative exponential function than when applying the truncated negative exponential function, $t(15) = .46, p = .65$.¹¹

A one-way ANOVA did, however, reveal a significant overall effect of function applied to the Glendale: Burglary data file on mean hit percentage, $F(5, 75) = 2.96, p = .02, \eta^2 = .17$. Post hoc analyses revealed that the mean hit percentage was significantly higher when applying the Glendale: Theft function to the Glendale: Burglary data file than when applying either the Glendale: All crime types function ($p < .001$) or the default negative exponential function ($p = .001$).¹² All other post hoc comparisons were not significant.

The planned comparisons found non-significant differences in hit percentage when applying the calibrated function compared to applying either the default negative exponential function, $t(15) = 1.38, p = .19$, or the default truncated negative exponential function, $t(15) = -.84, p = .41$. However, applying the negative exponential function to

¹¹ A Bonferroni correction was used to calculate a significance level of .02 for the three planned comparisons ($\alpha_{\text{planned}} = \alpha_{\text{per contrast}}/\text{number of planned contrasts} = .05/3 = .02$).

¹² A Bonferroni correction of .004 was used for the 12 post hoc comparisons ($\alpha_{\text{post hoc}} = \alpha_{\text{per contrast}}/\text{number of post hoc contrasts} = .05/12 = .004$).

the Glendale: Burglary data file did result in a significantly lower mean hit percentage than applying the truncated negative exponential function, $t(15) = -2.63, p = .02$.

Glendale: Auto theft. The one-way ANOVA revealed that the function applied to the Glendale: Auto theft data file did not have a significant overall effect on either mean error distance, $F(5, 70) = 2.10, p = .08, \eta^2 = .13$, or mean hit percentage, $F(5, 70) = .84, p = .52, \eta^2 = .06$. The planned comparisons also found non-significant differences when applying the calibrated function versus applying either the default negative exponential function (error distance: $t(14) = -1.01, p = .33$; hit percentage: $t(14) = 1.01, p = .33$) or the default truncated negative exponential function (error distance: $t(14) = -.99, p = .34$; hit percentage: $t(14) = -.43, p = .68$). In addition, the mean error distance was not significantly different when applying the negative exponential function than when applying the truncated negative exponential function, $t(14) = -.06, p = .95$. However, the mean hit percentage was significantly lower when applying the negative exponential function than when applying the truncated negative exponential function, $t(14) = -3.07, p = .01$.

Glendale: Theft. The one-way ANOVA revealed that the function applied to the Glendale: Theft data file had a significant overall effect on mean error distance, $F(5, 650) = 3.07, p = .01, \eta^2 = .02$. Post hoc analyses, however, indicated that these differences in mean error distance were not significant at the Bonferroni corrected significance level of .004. The planned comparisons did find significant differences in mean error distance between applying the calibrated function and applying either the default negative exponential function, $t(130) = 2.67, p = .01$, or the default truncated negative exponential function, $t(130) = 2.59, p = .01$, with both of the default functions resulting in mean error

distances that were lower than that of the calibrated function. Finally, the mean error distance was not significantly different when applying the negative exponential function than when applying the truncated negative exponential function, $t(130) = -.47, p = .64$.

A one-way ANOVA revealed a significant overall effect of function applied to the Glendale: Theft data file on mean hit percentage, $F(5, 650) = 20.44, p < .001, \eta^2 = .14$. Post hoc analyses revealed that the mean hit percentage was significantly higher when applying the Glendale: Burglary function to the Glendale: Theft data file than when applying the Glendale: All crime types ($p < .001$), negative exponential ($p < .001$), or truncated negative exponential functions ($p < .001$). Similarly, the mean hit percentage was significantly higher when applying the Glendale: Auto Theft function to the Glendale: Theft data file than when applying the Glendale: All crime types ($p < .001$), negative exponential ($p = .001$), or truncated negative exponential functions ($p = .002$). In addition, the calibrated function resulted in a significantly lower mean hit percentage than either the Glendale: Burglary ($p < .001$) or Glendale: Auto theft functions ($p < .001$). All other post hoc comparisons were not significant.

The planned comparisons found non-significant differences in mean hit percentage when applying the calibrated function compared to applying either the default negative exponential function, $t(130) = 2.25, p = .03$, or the default truncated negative exponential function, $t(130) = 1.41, p = .16$. In addition, the mean hit percentage was not significantly different when applying the negative exponential function than when applying the truncated negative exponential function, $t(130) = -1.96, p = .05$.

*Analysis 2: Geographic Region**Descriptive Analysis*

A descriptive analysis was conducted on the burglary data files from the various geographic regions (Glendale, Baltimore, and Dorset). For each data file, the number of offenders and crimes was calculated, along with descriptive statistics relating to the JTC distances. This data is presented in Table 8.

Table 8. Descriptive analysis of JTC distances (in miles) for each data file in Analysis 2.

Sample	Number of Offenders	Number of Crimes	JTC Distances			
			Min	Max	Median	Mean (<i>SD</i>)
Glendale: Burglary	16	77	0	24.35	1.29	2.94 (5.49)
Baltimore: Burglary	61	260	0	36.89	.97	4.12 (6.50)
Dorset: Burglary	38	494	.04	36.00	1.76	8.14 (11.66)

Calibrated Functions

Graphic representations. Figures 7 (on page 39), 11 and 12 are graphic representations of the distance decay functions calibrated for the Glendale, Baltimore, and Dorset data files, respectively. It is important to note that the observed frequencies (represented by the solid line) are only for the data file's development sample (i.e., the sample used to develop the calibrated function).

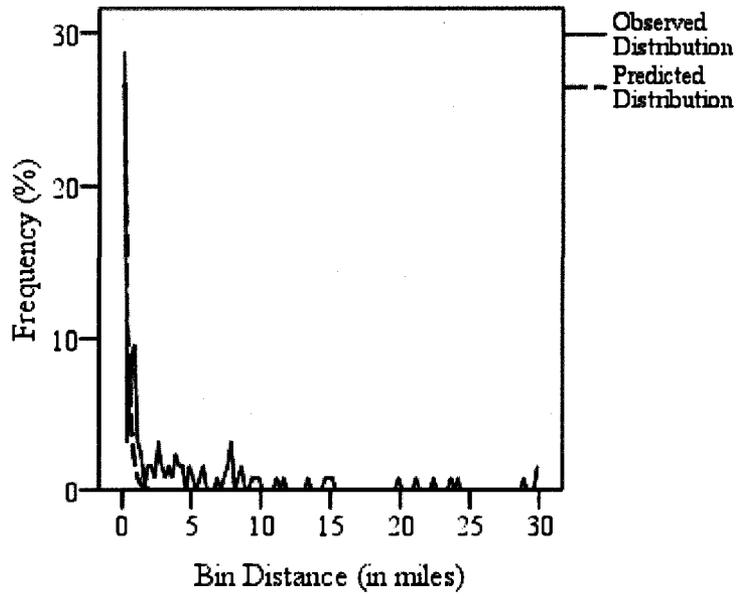


Figure 11. Calibrated function for Baltimore: Burglary.

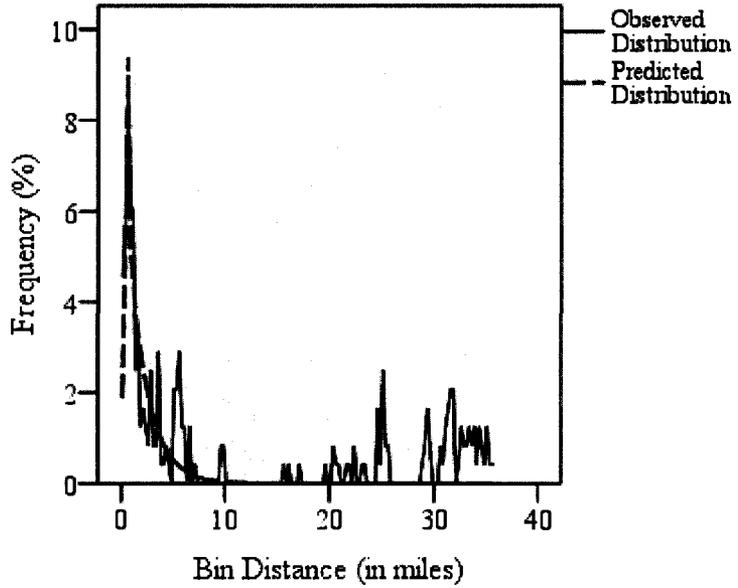


Figure 12. Calibrated function for Dorset: Burglary.

Parameters. As was done in Analysis 1, the frequency distributions were used to classify the functions calibrated for geographic region as being either a truncated negative exponential or a negative exponential function. Based on the distributions, the Glendale: Burglary and Baltimore: Burglary functions were classified as negative exponential, whereas the Dorset: Burglary function was classified as truncated negative exponential. Parameters for each calibrated function and their corresponding R^2 values are presented in Table 9, and the predicted frequencies are presented in Figures 7, 11, and 12 (represented by the dashed line). The R^2 values indicate that the calibrated functions fit their corresponding development samples relatively well (R^2 values range from .66 to .78).

Table 9. Parameters and R^2 values for the calibrated functions in Analysis 2.

Dataset	Function	Parameters			Distance of Peak Freq. (in miles)	R^2
		A	B	C		
Glen: Burglary	Neg Exp	31.68	1.85	--	--	.72
Balt: Burglary	Neg Exp	41.69	3.54	--	--	.78
Dorset: Burglary	Trunc Neg Exp	8.09	.55	14.99	.625	.66

Note. The parameter notations (A, B, and C) in this table refer to the same parameter notations used in the equations on page 40.

Split-Half Validation

Following the same validation procedure as the one used for the analysis of crime type, the functions calibrated for geographic region were applied to both the development samples and the test samples and t -tests were run to compare the mean error distances

and mean hit percentages. The *t*-tests revealed that the mean error distances and mean hit percentages associated with offenders in the development samples were not significantly different from those of the test samples; this was the case for all of the calibrated functions (see Table 10). The non-significant results obtained in the split-half validation indicate that the calibrated functions can be used with sufficient accuracy when being applied to data beyond that which was used to derive them. Since the split-half validation was successful, the development and test samples were pooled for all further analyses.

Table 10. Split-half validation results (mean error distance in miles) for Analysis 2.¹³

Dataset	Development Sample		Test Sample		<i>t</i>	df	Sig
	N/#	Error Distance Mean (<i>SD</i>)	N/#	Error Distance Mean (<i>SD</i>)			
Glendale: Burglary	8/27	3.59 (8.26)	8/50	2.82 (4.69)	.23	14	.82
Baltimore: Burglary	31/125	4.16 (7.45)	30/135	4.29 (6.96)	-.07	59	.94
Dorset: Burglary	19/241	8.96 (11.52)	19/253	6.35 (10.81)	.72	36	.48

Note. *N* represents the number of offenders in the sample and # represents the number of crimes in the sample.

Error Distance and Hit Percentage Results

Minimum, maximum, and mean error distances (in miles) and hit percentages for Analysis 2 are presented in Tables 11 and 12. Repeated measures ANOVAs were conducted in order to determine whether the function applied to the data file had an overall effect on error distance or hit percentage.

¹³ This table only presents results from the error distance analysis. Results from the hit percentage analysis are not presented here as they were also non-significant

Table 11. Minimum, maximum, and mean error distances (in miles) for Analysis 2.

Function developed using:		Function applied to:		
		Glendale: Burglary	Baltimore: Burglary	Dorset: Burglary
Trunc. Neg. Exp.	Min.	.03	.13	.39
	Max.	23.47	34.90	33.67
	Mean (SD)	3.03 (6.39)	4.38 (7.14)	7.69 (11.08)
Neg. Exp.	Min.	.05	.01	.08
	Max.	23.83	34.93	33.96
	Mean (SD)	3.10 (6.52)	4.29 (7.16)	7.64 (11.23)
Glendale: Burglary	Min.	.21	.05	.08
	Max.	24.00	34.73	33.96
	Mean (SD)	3.20 (6.51)	4.40 (7.25)	7.65 (11.20)
Baltimore: Burglary	Min.	.15	.09	.15
	Max.	24.09	34.73	33.96
	Mean (SD)	3.42 (6.58)	4.22 (7.15)	7.73 (11.18)
Dorset: Burglary	Min.	.38	.36	.35
	Max.	23.22	34.18	33.67
	Mean (SD)	3.20 (6.24)	4.50 (7.10)	7.66 (11.11)

Note. Shaded cells refer to the results of the calibrated functions.

Table 12. Minimum, maximum, and mean hit percentages for Analysis 2.

		Function applied to:		
Function developed using:		Glendale: Burglary	Baltimore: Burglary	Dorset: Burglary
Trunc. Neg. Exp.	Min.	.004	.05	.04
	Max.	55.03	97.16	90.71
	Mean (<i>SD</i>)	15.05 (17.71)	14.41 (21.47)	18.13 (32.78)
Neg. Exp.	Min.	.004	.01	.01
	Max.	55.04	96.73	90.60
	Mean (<i>SD</i>)	6.56 (16.61)	10.15 (20.91)	17.97 (32.98)
Glendale: Burglary	Min.	.05	.01	.01
	Max.	100.00	100.00	86.40
	Mean (<i>SD</i>)	10.83 (28.86)	12.04 (25.11)	11.77 (22.21)
Baltimore: Burglary	Min.	.02	.01	.01
	Max.	91.30	96.03	93.34
	Mean (<i>SD</i>)	9.00 (24.78)	7.10 (17.51)	14.06 (25.92)
Dorset: Burglary	Min.	.10	.15	.02
	Max.	64.39	70.95	77.57
	Mean (<i>SD</i>)	7.80 (17.67)	9.66 (17.75)	8.54 (19.62)

Note. Shaded cells refer to the results of the calibrated functions.

Glendale: Burglary. The one-way ANOVA revealed that the function applied to the Glendale: Burglary data file did not have a significant overall effect on either mean error distance, $F(4, 60) = 1.68, p = .17, \eta^2 = .10$, or mean hit percentage, $F(4, 60) = 2.38, p = .06, \eta^2 = .14$. The planned comparisons also found non-significant differences in mean error distance when applying the calibrated function versus when applying either the default negative exponential function, $t(15) = .61, p = .55$, or the default truncated negative exponential function, $t(15) = 1.25, p = .23$. In addition, the mean error distance was not significantly different when applying the negative exponential function than when applying the truncated negative exponential function, $t(15) = .46, p = .65$.

The planned comparisons also found non-significant differences between the mean hit percentage when the calibrated function was applied to the Glendale: Burglary data file versus when the negative exponential, $t(15) = 1.38, p = .19$, or truncated negative exponential functions, $t(15) = -.84, p = .41$, were applied. The planned comparisons did, however, find a significant difference between the mean hit percentage when applying the two default functions, with the negative exponential function outperforming the truncated negative exponential function, $t(15) = -2.63, p = .02$.

Baltimore: Burglary. The one-way ANOVA revealed that the function applied to the Baltimore: Burglary data file did not have a significant overall effect on mean error distance, $F(4, 240) = .77, p = .54, \eta^2 = .01$. The planned comparisons also found non-significant differences between applying the calibrated function and applying either the default negative exponential function, $t(60) = -.29, p = .77$, or the default truncated negative exponential function, $t(60) = -.64, p = .52$. However, the mean error distance

was significantly lower when applying the negative exponential function than when applying the truncated negative exponential function, $t(60) = -2.45, p = .02$.

Although the one-way ANOVA revealed a significant overall effect of function applied to the Baltimore: Burglary data file on mean hit percentage, $F(4, 240) = 2.90, p = .02, \eta^2 = .05$, the post hoc analyses indicated that the differences in mean hit percentage when applying the various functions were not significant at the Bonferroni corrected significance level of .007.¹⁴ The planned comparisons, however, found that the mean hit percentage was significantly lower when applying the calibrated function than when applying either the default negative exponential, $t(60) = -2.66, p = .01$, or the default truncated negative exponential functions, $t(60) = -4.58, p < .001$. In addition, the mean hit percentage of the default negative exponential function was significantly lower than that of the default truncated negative exponential function, $t(60) = -3.65, p = .001$.

Dorset: Burglary. The one-way ANOVA revealed that the function applied to the Dorset: Burglary data file did not have a significant overall effect on mean error distance, $F(4, 148) = .26, p = .91, \eta^2 = .01$. The planned comparisons also found non-significant differences in mean error distance when applying the calibrated function compared to when either the default negative exponential function, $t(37) = .23, p = .82$, or the default truncated negative exponential function, $t(37) = -.25, p = .80$, was applied. In addition, the mean error distance was not significantly different when applying the negative exponential function than when applying the truncated negative exponential function, $t(37) = -.51, p = .61$.

¹⁴ A Bonferroni correction of .007 was used for the seven post hoc comparisons ($\alpha_{\text{post hoc}} = \alpha_{\text{per contrast}}/\text{number of post hoc contrasts} = .05/7 = .007$).

Although, the one-way ANOVA revealed that the function applied to the Dorset: Burglary data file had a significant overall effect on mean hit percentage, the post hoc analyses indicated that none of the differences between the functions were significant at the Bonferroni corrected significance level of .007. The planned comparisons also found non-significant differences in mean hit percentage between the calibrated function and the two default functions (negative exponential: $t(37) = -2.17, p = .04$; truncated negative exponential: $t(37) = -2.22, p = .03$). Finally, the mean hit percentage did not differ significantly between the default negative exponential and truncated negative exponential functions, $t(37) = -1.18, p = .24$.

Discussion

At present, the most common procedure for constructing geographic profiles is to use default (i.e., uncalibrated) distance decay functions that are incorporated into computerized GP systems. Given that research from the fields of environmental criminology and psychology suggests that crime type and geographic region may limit the extent to which default distance decay functions generalize beyond the samples they were developed from, the primary purpose of the current thesis was to determine whether decay functions calibrated for crime type or geographic region produce more accurate profiles than default functions. In addition to examining the impact of calibrated functions, this thesis also compared the degree of GP accuracy that could be achieved using different types of default functions, namely the negative exponential and truncated negative exponential function. Given that these two functions are the most commonly used in naturalistic settings, this comparison was considered to have serious practical implications. Before moving on to discuss the results from these analyses, it is important

to highlight a number of findings which emerged from this thesis that are of a more general nature.

Two findings in particular warrant discussion. First, it is important to mention the fact that, across each of the samples examined in this thesis, the JTC frequency distributions were found to follow a distance decay pattern, specifically a negative exponential or truncated negative exponential function. While not unexpected, this finding is also not inevitable and it lends support to a large body of literature, which shows that offenders are more likely to live close to their crimes, as opposed to further away. The fact that this was found in the current study, where property offences were focused on, is particularly important, since much of the previous research in this area has examined interpersonal crimes. This finding also provides further support for routine activity theory and rational choice theory. However, a much more detailed examination of offender spatial behaviour, including interviews with the offenders, would be needed to make any direct links between the findings reported here and these theories.

Second, the findings from this thesis suggest that probability distance strategies can be used to accurately profile offender residences. For example, an examination of Tables 7 and 12 indicate that the mean hit percentages ranged from 6.17 to 26.63 when applying the various distance decay functions. Even the default functions that are currently used in *CrimeStat* produced accurate profiles. Thus, even at their worst, the decay functions examined in this thesis were still able to limit the total area that would need to be searched by the police by almost 75%. Although a variety of factors would influence the value of this finding (e.g., the population density of the area under

consideration), reductions in search area of this magnitude would likely be very useful from an investigative standpoint.

Despite the fact that the aforementioned results emerged from the current study, each of which is clearly important, the major hypotheses that were being tested in this thesis were not supported to the extent that was expected. Specifically, in almost all cases, regardless of whether GP accuracy was assessed using error distance or hit percentage, the results from this thesis revealed that functions calibrated for crime type or geographic region did not significantly improve GP accuracy. In addition, except for the Baltimore: Burglary file where the calibrated function resulted in a smaller hit percentage than the two default functions, the a priori comparisons indicated that calibrated functions did not result in significantly smaller error distances or hit percentages than the default functions. Interestingly, across four of the five data files, the two default functions also did not differ significantly from one another with respect to GP accuracy as assessed by error distance. The negative exponential function did, however, result in a significantly smaller mean hit percentage than the truncated negative exponential function across three of the data files.

Overall then, the results from this thesis provide fairly strong evidence indicating that calibrating for crime type or geographic region does not result in significantly better distance decay functions for use in GP systems, at least not consistently across data sets. In addition, and in line with previous research (e.g., Canter & Hammond, 2006; Paulsen, 2006a; Taylor, Bennell, & Snook, 2002), the mostly non-significant differences between the default truncated negative exponential function and the default negative exponential function suggests that neither default function is significantly more accurate than the

other in terms of GP accuracy when measured by error distance. The results do, however, suggest that the negative exponential function may lead to more accurate geographic profiles in terms of hit percentage, but these results were not consistent across all of the data files. Overall, these results lead to certain questions that require answers. Most obviously, why did these results emerge? Why did calibration not have its intended effect, and why were no consistent differences found between the default functions?

The Lack of a Calibration Effect

The fact that calibrating for crime type and geographic region did not, for the most part, result in more accurate geographic profiles is somewhat surprising given that a wealth of research indicates that a wide variety of crime- and environment-related factors influence the spatial behaviour of offenders (e.g., Baldwin & Bottoms, 1976; Capone & Nichols, 1976; Gabor & Gottheil, 1984; LeBeau, 1987; Rhodes & Conly, 1981). It was assumed that some of these factors were being tapped into in the design of the current study. However, one of the most likely explanations for why the results emerged in the way they did is that potentially important between-sample differences were not being exploited in the way that was initially intended.

For example, it was predicted that calibrating for crime type would result in more accurate profiles because previous research has clearly shown that the spatial behaviour of offenders differ depending on the type of crime being committed. For instance, Rhodes and Conly (1981) found differences in JTC distances between robbers, burglars, and rapists in their study of offenders from Washington, DC, with robbers and burglars travelling farther distances to commit their crimes than rapists. Similarly, Rand (1986) found that offenders committing crimes against another individual (e.g., criminal

homicide, aggravated assault) travelled shorter distances than offenders committing crimes against property (e.g., burglary, vehicle theft, larceny). Thus, the findings from these two studies suggest that differences in JTC distances may exist between offenders committing different types of crimes. However, in the current study, while crime type varied across the Glendale sub-samples, each of the sub-samples represented only one broad class of crimes – property crimes. Therefore, a potential explanation for why the crime type calibration did not have the anticipated effect was because the spatial behaviour of different types of property offenders is relatively similar (an explanation which is supported to some extent by the results in Table 3, with the possible exception of the burglary dataset). If the crime types had varied more dramatically in the current study, for example by including samples of interpersonal offenders *and* property offenders, the results of the study may have been different. This will be an important avenue for future research.

A similar argument can be made for the analysis of geographic regions. Initially, the idea behind this analysis was that, because various environmental factors are known to influence offender spatial behaviour (e.g., Bennell & Stein, 2008; Capone & Nichols, 1976; Rhodes & Conly, 1981), calibrating functions for different geographic regions, which presumably differ with respect to environmental factors, would result in more accurate geographic profiles. In hindsight, it seems possible that the sampling procedure used in this study was unsuccessful at capturing important between-sample environmental differences that might actually exist across the various geographic regions that were targeted.

For example, the geographic regions targeted in this study each covered a relatively large area that could most certainly have been broken down into smaller, more homogeneous sub-regions. Indeed, in hindsight there is every reason to believe that important environmental differences would exist between these sub-regions within the same sample. Take Dorset, for example. One of the reasons why a calibration effect was expected in the analysis of Dorset offenders was that that population density and road networks in Dorset are likely to differ from the US samples, given that Dorset is largely a rural county in the UK, with an overall population density of 158 persons per square kilometre (Dorset Databook, 2007), and the road networks there do not follow the standard grid pattern that is found in many US cities (e.g., Levine & Associates, 2007; Rhodes & Conly, 1981). Having said this, Dorset is a large county, consisting of urban sub-regions that are highly populated and potentially more similar to Glendale and Baltimore than was originally thought. If a large portion of the Dorset serial offenders examined in this thesis committed crimes within these regions, rather than the very rural areas of Dorset, there would likely be no calibration effect for the Dorset data file. Unfortunately, it was not possible given the data available for the current study to examine this issue in more depth. Again, this will be a useful line of future research and will require much more detailed environmental data from each of the regions that were examined (e.g., micro-level population density data).

In line with this argument, it is also important to highlight the fact that some of the data files spanned a wide temporal period. For example, the Glendale data file included crimes that occurred over a span of eight years. It is highly probable that significant environmental changes, such as new residential development, the construction

of new roadways, and changes in population density, occurred during those eight years and this would likely have an effect on the spatial behaviour of offenders operating in this area. Indeed, as previously mentioned, if an offender's spatial behaviour is determined in part by their routine activities, which most people accept (e.g., Cohen & Cantor, 1980; Sherman, Gartin, & Buerger, 2006), such environmental changes would likely have a large impact on the behaviour of offenders committing crimes at different times in the same geographic area. Thus, the fact that each data file may consist of different sub-samples based on temporal period of offending could have masked many of the environmental factors that were trying to be exploited in the current study. Unfortunately, the data files could not be restricted to shorter time periods due to the limited data available in some of the data files. If larger samples can be collected in the future, then such an analysis should be conducted.

Other possible explanations also obviously exist to explain the lack of a calibration effect. One of the most obvious has to do with the relatively small sample sizes that were available for analysis in the current study. While small sample sizes are not uncommon in research of this type (e.g., Snook et al., 2002; Snook et al., 2004; Snook, Zito, et al., 2005), due largely to difficulties in obtaining very sensitive data from the police, the sample sizes are cause for concern, especially in the analysis of crime types. Clearly, collecting larger samples for future research must be a high priority. Furthermore, it could be that the factors targeted in the current study are simply not the factors that need to be considered when calibrating distance decay functions. For example, instead of crime type or geographic region, calibrating functions based on offender characteristics might be more useful since a variety of characteristics are known

to strongly influence spatial behaviour (e.g., Baldwin & Bottoms, 1976; Gabor & Gottheil, 1984; Snook, 2004; Warren, Reboussin, & Hazelwood, 1995). Unfortunately, even if this were successful, the police would not know which function to apply in an investigation since offender characteristics are often not known at that time. Thus, research that examines methods of predicting offender characteristics at the time of the investigation would also need to be conducted.

Regardless of the explanation for the lack of a calibration effect, this result, if replicated in future studies, has important implications. Most obviously, the results indicate that it may be unnecessary to calibrate distance decay functions. Instead, default functions that have been developed on samples of crimes and criminals that are unrelated to the cases where the GP system is being applied could be used with the same degree of accuracy in most instances. This is an important piece of practical information because it means that valuable time, effort, and money, which would have to be spent in order to carry out the calibration exercise, might be saved.

The Lack of Any Difference between the Default Functions

In accordance with previous research (e.g., Canter & Hammond, 2006; Paulsen, 2006a; Snook, Zito, et al., 2005; Taylor et al., 2002), the results of the current study also suggest that, if using a default function to construct a geographic profile, it does not really matter, at least in terms of error distance, whether the profiler opts for a negative exponential function or a truncated negative exponential function. However, the results of the current study suggest that the negative exponential function may, on occasion, result in more accurate geographic profiles in terms of hit percentage, which makes some sense

given that four of the six calibrated functions were found to follow a negative exponential function.

Of course, future research that deals with the study limitations discussed above might result in more obvious differences between the performance of the two functions that were tested in this thesis, and it might be the case that other types of decay functions prove to be more effective than the two default functions that were tested. However, based on the current results, there seems to be little consistent difference between the two functions, with both resulting in relatively accurate profiles for most (but not all) offenders (see Tables 6, 7, 11, and 12). Given that these functions clearly differ from one another (i.e., one has a buffer zone and the other does not) the obvious question to ask is why, for the most part, no differences were found between them (especially in relation to error distance). There are in fact a number of potential explanations for this finding.

One likely explanation is that the precision of the distance intervals used to measure JTC distances may have contributed to the non-significant findings. For example, it is possible that the bin size adopted did not accurately capture the probability that offender's commit crimes very close to home, which is where the most important differences between the two default functions lie. Indeed, if the distance intervals were too large, the existence of a buffer zone might have been totally masked. The reliance in the current study on relatively small sample sizes, and relatively short crime series, might also have contributed to this problem, given that increases to either would have a disproportionate effect on short JTC distances. Perhaps if shorter distance intervals were used in future research, with larger samples sizes and series lengths, larger differences in GP accuracy would emerge across different functions. It must also be pointed out that

the use of aggregate distance decay functions may also reduce the chance of finding differences between functions, as they will minimize the impact of offenders who show very distinct patterns of behaviour (e.g., very clear buffer zones) (Van Koppen & De Keijser, 1997).

Thus, future research will need to carefully examine the issue of distance intervals, and to think through the impact of using aggregate decay functions to profile individual offenders. However, having said this, it seems clear that, if any *significant* improvements are going to be made to the predictive accuracy of GP systems it will require that we move beyond a strict reliance on distance decay models. Indeed, even if a procedure were found to improve the degree of GP accuracy via the use of decay functions, for example by incorporating information about the cognitive processes of offenders (Canter & Hammond, 2006), the effect would likely be constant across the various types of functions. Currently, new procedures are being explored, such as the Bayesian routine by Levine and Associates (2007), which uses ones knowledge of how homes and offences cluster in relevant samples of solved crimes in order to improve the accuracy of calibrated decay functions. Approaches like these still remain largely untested and require further research.

Conclusion

The results from the current thesis suggest that serial offenders from multiple samples demonstrate the common distance decay phenomenon when their JTC distances are mapped, suggesting that there is some logic to their crime site selection choices and some support for RAT and RCT. The distance decay functions that emerged from the development samples could also be used to profile the residential locations of the

offenders in a relatively accurate fashion. However, in contrast to what was expected, calibrating distance decay functions for crime type or geographic region was, in general terms, not a consistent and effective way of improving GP accuracy, as assessed by either error distance or hit percentage. In only one data file was the calibrated function found to significantly outperform the types of default functions that are currently utilized computerized GP systems, and this finding only applied to hit percentage and not error distance. In addition, while the default negative exponential function occasionally outperformed the truncated negative exponential function with respect to GP accuracy, this finding was not consistent across all data files.

Having reached these conclusions, there are obviously a number of limitations to the current study that need to be addressed in future research. This research may lead to findings that require revisions to be made to these conclusions. However, at the present time, it appears that it is not worthwhile for police departments to calibrate GP functions for crime type or geographic region as this will not lead to consistent and significant improvements in overall GP accuracy. Also, whether the default negative exponential or truncated negative exponential function is used should not be a major concern to police departments given that, overall, major differences in GP accuracy were not found between these functions.

References

- Baldwin, J., & Bottoms, A. E. (1976). *The urban criminal: A study in Sheffield*. London: Tavistock Publications.
- Bennell, C. (2002). *Behavioural consistency and discrimination in serial burglary*. Unpublished doctoral dissertation, University of Liverpool, Liverpool, UK.
- Bennell, C., Snook, B., Taylor, P. J., Corey, S., & Keyton, J. (2007). It's no riddle, choose the middle: The effect of number of crimes and topographical detail on police officer predictions of serial burglars' home locations. *Criminal Justice and Behavior, 34*, 119-132.
- Bennell, C., & Stein, K. (2008). *Relationship between population density and journey-to-crime in serial homicide*. Unpublished manuscript, The University of Liverpool, Liverpool, UK.
- Bennell, C., Taylor, P. J., & Snook, B. (2007). Clinical versus actuarial geographic profiling strategies: A review of the research. *Police Practice and Research, 8*, 335-345.
- Bourget, D., & Bradford, J. M. W. (1989). Female arsonists: A clinical study. *Bulletin of the American Academy of Psychiatry and Law, 17*, 293-300.
- Brantingham, P., & Brantingham, P. (1984). *Patterns in crime*. New York: Macmillan Publishing Company.
- Canter, D. V. (2003). *Mapping murder: The secrets of geographical profiling*. London: Virgin Books.

- Canter, D. V., Coffey, T., Huntley, M., & Missen, C. (2000). Predicting serial killers' home base using a decision support system. *Journal of Quantitative Criminology*, *16*, 457-477.
- Canter, D. V., & Gregory, A. (1994). Identifying the residual location of rapists. *Journal of the Forensic Science Society*, *34*, 169-175.
- Canter, D. V., & Hammond, L. (2006). A comparison of the efficacy of different decay functions in geographical profiling for a sample of US serial killers. *Journal of Investigative Psychology and Offender Profiling*, *3*, 91-103.
- Canter, D. V., & Hodge, S. (1997). *Predatory patterns of serial murder*. Unpublished manuscript, The University of Liverpool, Liverpool, UK.
- Canter, D. V., & Larkin, P. (1993). The environmental range of serial rapists. *Journal of Environmental Psychology*, *13*, 63-69.
- Capone, D. L., & Nichols, W. W., Jr. (1976). *Urban structure and criminal mobility*. *American Behavioral Scientist*, *20*, 199-213.
- Cohen, L. E., & Cantor, D. (1980). The determinants of larceny: An empirical and theoretical study. *Journal of Research in Crime and Delinquency*, *17*, 140-159.
- Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Psychological Review*, *44*, 588-608.
- Cornish, D. B., & Clarke, R. V. (1986). *The reasoning criminal: Rational choice perspectives on offending*. New York: Springer-Verlag.
- Dorset Databook. (2007). *Statistics and census information*. Retrieved August 10, 2008, from <http://www.dorsetforyou.com/index.jsp?articleid=333003>.

- Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- Federal Bureau of Investigation. (2006). *Crime in the United States*. Retrieved January 2, 2008, from http://www.fbi.gov/ucr/cius2006/about/offense_definitions.html.
- Gabor, T., & Gottheil, E. (1984). Offender characteristics and spatial mobility: An empirical study and some policy implications. *Canadian Journal of Criminology*, 26, 267-281.
- Holmes, R. M., & De Burger, J. (1988). *Serial murder*. Beverly Hills: Sage.
- Kent, J., Leitner, M., & Curtis, A. (2006). Evaluating the usefulness of functional distance measures when calibrating journey-to-crime distance decay functions. *Computers, Environment and Urban Systems*, 30, 181-200.
- Kocsis, R. N., & Irwin, H. J. (1997). An analysis of spatial patterns in serial rape, arson, and burglary: The utility of the Circle Theory of environmental range for psychological profiling. *Psychiatry, Psychology and Law*, 4, 195-206.
- Kraemar, G. W., Lord, W. D., & Heilbrun, K. (2004). Comparing single and serial homicide offenses. *Behavioral Sciences & the Law*, 22, 325-343.
- Laukkanen, M., Santtila, P., Jern, P., & Sandnabba, K. (2008). Predicting offender home location in urban burglary series. *Forensic Science International*, 176, 224-235.
- LeBeau, J. L. (1987). The journey to rape: Geographic distance and the rapist's method of approaching the victim. *Journal of Police Science and Administration*, 15, 129-136.
- Levine, N., & Associates. (1999). *CrimeStat: A spatial statistics program for the analysis of crime incident locations* (v 1.0). National Institute of Justice, Washington, DC.

- Levine, N., & Associates. (2007). *CrimeStat: A spatial statistics program for the analysis of crime incident locations* (v 3.1). National Institute of Justice, Washington, DC.
- Meaney, R. (2004). Commuters and marauders: An examination of the spatial behaviour of serial criminals. *Journal of Investigative Psychology and Offender Profiling, 1*, 121-137.
- Mustaine, E. E., & Tewksbury, R. (1999). A routine activity theory explanation for women's stalking victimization. *Violence Against Women, 5*, 43-62.
- Pal, T. (2007). *Accuracy assessment of individually calibrated journey-to-crime geographic profiling models*. Unpublished master's thesis, Louisiana State University, Baton Rouge, Louisiana, USA.
- Paulsen, D. J. (2006a). Connecting the dots: Assessing the accuracy of geographic profiling software. *Policing: An International Journal of Police Strategies & Management, 29*, 306-334.
- Paulsen, D. J. (2006b). Human versus machine: A comparison of the accuracy of geographic profiling methods. *Journal of Investigative Psychology and Offender Profiling, 3*, 77-89.
- Paulsen, D. J. (2007). Improving geographic profiling through commuter/marauder prediction. *Police Practice and Research, 8*, 347-357.
- Rand, A. (1986). Mobility triangles. In R. M. Figlio, S. Hakim, & G. F. Rengert (Eds.), *Metropolitan crime patterns* (pp. 117-126). Monsey: Criminal Justice Press.
- Reber, A. S. (1985). *The penguin dictionary of psychology*. Harmondsworth: Penguin.
- Rengert, G. F., Piquero, A. R., & Jones, P. R. (1999). Distance decay reexamined. *Criminology, 37*, 427-445.

- Rhodes, W. M., & Conly, C. (1981). Crime and mobility: An empirical study. In P. J. Brantingham & P. L. Brantingham (Eds.), *Environmental criminology* (pp. 167-188). Beverly Hills: Sage.
- Rich, T., & Shively, M. (2004). *A methodology for evaluating geographic profiling software*. Retrieved December 18, 2007, from <http://www.ojp.usdoj.gov/nij/maps/gp.html>.
- Roncek, D. W., & Maier, P. A. (1991). Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of "hot spots." *Criminology*, 29, 725-753.
- Rossmo, D. K. (2000). *Geographic profiling*. Boca Raton: CRC Press.
- Rossmo, D. K. (2005a). Geographic heuristics or shortcuts to failure? Response to Snook et al. *Applied Cognitive Psychology*, 19, 651-654.
- Rossmo, D. K. (2005b). *An evaluation of NIJ's evaluation methodology for geographic profiling software*. Retrieved January 5, 2008, from <http://www.txstate.edu/gii/documents/Response%20to%20NIJ%20GP%20Evaluation%20Methodology.doc>.
- Santtila, P., Laukkanen, M., & Zappala, A. (2007). Crime behaviours and distance travelled in homicides and rapes. *Journal of Investigative Psychology and Offender Profiling*, 4, 1-15.
- Santtila, P., Zappala, A., Laukkanen, M., & Picozzi, M. (2003). Testing the utility of a geographical profiling approach in three rape series of a single offender: A case study. *Forensic Science International*, 131, 42-52.

- Sarangi, S., & Youngs, D. (2006). Spatial patterns of Indian serial burglars with relevance to geographical profiling. *Journal of Investigative Psychology and Offender Profiling*, 3, 105-115.
- Sherman, L. W., Gartin, P. R., & Buerger, M. E. (2006). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27, 27-56.
- Simon, H. A. (1957). *Models of man*. New York: Wiley.
- Snook, B. (2004). Individual differences in distance travelled by serial burglars. *Journal of Investigative Psychology and Offender Profiling*, 1, 53-66.
- Snook, B., Canter, D. V., & Bennell, C. (2002). Predicting the home location of serial offenders: A preliminary comparison of the accuracy of human judges with a geographic profiling system. *Behavioral Sciences and the Law*, 20, 109-118.
- Snook, B., Cullen, R. M., Mokros, A., & Harbort, S. (2005). Serial murderers' spatial decisions: Factors that influence crime location choice. *Journal of Investigative Psychology and Offender Profiling*, 2, 147-164.
- Snook, B., Taylor, P. J., & Bennell, C. (2004). Geographic profiling: The fast, frugal and accurate way. *Applied Cognitive Psychology*, 18, 105-121.
- Snook, B., Zito, M., Bennell, C., & Taylor, P. J. (2005). On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology*, 21, 1-26.
- Taylor, P. J., Bennell, C., & Snook, B. (2002). *Problems of classification in investigative psychology*. Proceedings of the 8th Conference of the International Federation of Classification Societies, Krakow, Poland.
- Trainor, N. (2006). *Geographic profiling: Theory versus practical application*. Paper presented at the 4th National Crime Mapping Conference, London, UK.

Van Koppen, P. J., & De Keijser, J. W. (1997). Desisting distance decay: On the aggregation of individual crime trips. *Criminology*, 35, 505-515.

Warren, J. I., Reboussin, R., & Hazelwood, R. R. (1995). *The geographic and temporal sequencing of serial rape*. Final report submitted to the National Institute of Justice, July 15.

APPENDIX A¹⁵

Probability Distance Strategies

Please note that for all functions, $f(d_{ij})$ is the likelihood that the offender will commit a crime at a particular location, i ; d_{ij} is the distance between the offender's residence location, j , and crime location, i .

Linear

According to this function, the likelihood of an offender having committed a particular crime decreases, in a linear fashion, as the distance between the crime site and the offender's residence increases.

$$f(d_{ij}) = A + B * d_{ij}$$

where: A = y-intercept; B = slope coefficient (expected to have a negative sign since, according to this function, the likelihood should decline as distance increases).

Negative Exponential

According to this function, the likelihood of an offender having committed a particular crime decreases, in an exponential fashion, as the distance between the crime site and the offender's residence increases.

$$f(d_{ij}) = A * e^{-B * d_{ij}}$$

where: e = base of natural logarithm; A = coefficient; B = exponent.

Truncated Negative Exponential

This function relies on both the linear and the negative exponential functions. For crime sites close to the offender's residence, the positive linear function is used; for crime sites further away from the offender's residence, the negative exponential function is

¹⁵ The references for this appendix are Levine and Associates (2007) and Pal (2007).

used. The function switches from a linear to a negative exponential one at the distance where peak likelihood occurs.

$$\text{Linear:} \quad f(d_{ij}) = Cd_{ij} \quad \text{for } 0 \leq d_{ij} \leq d_p$$

$$\text{Negative Exponential:} \quad f(d_{ij}) = A * e^{-B * d_{ij}} \quad \text{for } d_{ij} > d_p$$

where: C = slope; e = base of natural logarithm; A = coefficient; B = exponent;

d_p = distance at which peak likelihood occurs.

Normal

According to this function, the likelihood of an offender having committed a particular crime is at its highest (or peak) at an optimal distance from the offender's residence. Thus, the likelihood increases until that optimal distance is reached and then decreases thereafter.

$$f(d_{ij}) = A * \text{Normal}(Z_i)$$

where: A = y-intercept and is estimated by the regression coefficient; Z_i = standardized variable.

Lognormal

This function is identical to the normal function except that it is either positively or negatively skewed.

$$f(d_{ij}) = A * \text{Lognormal}(d_i)$$

where: A = y-intercept and is estimated by the regression coefficient; Z_i = standardized variable.

APPENDIX B¹⁶

CrimeStat's Default Negative Exponential and Truncated Negative
Exponential Functions

The two default functions examined in this thesis include the negative exponential and truncated negative exponential functions currently used in *CrimeStat*. The following is the default negative exponential function used in *CrimeStat*:

$$f(d_{ij}) = A * e^{-B * d_{ij}}$$

where: $f(d_{ij})$ = the likelihood that the offender will commit a crime at a particular location, i (defined here as the centroid of a grid cell), e = base of natural logarithm, d_{ij} = the distance between the offender's residence location, j , and crime location, i ; $A = 1.89$, and $B = .06$.

The default truncated negative exponential function in *CrimeStat*, unlike the default negative exponential function, assumes that the offender's home location will be surrounded by a buffer zone. Thus, the function is a linear one up until peak likelihood is reached, whereupon it switches to a negative exponential function. The following is the default truncated negative exponential function used in *CrimeStat*:

$$\text{Linear:} \quad f(d_{ij}) = C * d_{ij} \quad \text{for } 0 \leq d_{ij} \leq d_p$$

$$\text{Negative Exponential:} \quad f(d_{ij}) = A * e^{-B * d_{ij}} \quad \text{for } d_{ij} > d_p$$

where: $f(d_{ij})$ = the likelihood that the offender will commit a crime at a particular location, i (defined here as the centroid of a grid cell), d_{ij} = the distance between the offender's residence location, j , and crime location, i ; d_p (distance of peak likelihood) = .4 miles, $C = 34.5$, $A = 14.22$, and $B = .2$.

¹⁶ The reference for this appendix is Levine and Associates (2007).