Geographic Information Systems (GIS) and Predictive Modeling of Body Disposal Sites

A Thesis Presented for the
Master of Arts
Degree
The University of Tennessee, Knoxville

Annemarie Catherine Gundel
August 2017

Copyright © 2017 by Annemarie Gundel
All rights reserved.

DEDICATION

For Jared and Ashe – I couldn't have done it without you.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Amy Mundorff for her support and encouragement, Dr. Michael Kenyhercz for his help in learning GIS and for his patience in answering my questions along the way, and Dr. Dawnie Steadman for her helpful comments during the writing process.

I am grateful to Dr. Derek Congram for meeting with me to discuss predictive modeling and for his comments on my proposal. Thank you also to Dr. Nicholas Nagle for his patience in answering my many stats questions.

I would also like to thank Dr. Kristen Hartnett and Dr. James Gill for allowing me to collect data at the Office of the Chief Medical Examiner, CT. Thank you also to the committee for the William M. Bass Endowment, whose decision to fund this project made my thesis possible.

Thank you to my friends and family, who encouraged me and listened to me constantly talk about this project. A special thanks to Jared for the helpful GIS hints, for the constant and unending support, and especially for the Chinese food delivery during the final writing hours of this thesis.

ABSTRACT

Homicide victims are often discovered by accident or located through witness testimony, both of which are unreliable methods. Moving a victim's body from the scene of the crime to a secondary disposal site may further complicate their discovery, delaying recovery, identification, and evidence collection. Homicides are exponentially more difficult to investigate, solve, and prosecute without a body. In the medicolegal context, a body disposal site prediction model is an alternative to relying on luck or witness testimony. Predictive models were created using body disposal data collected from the Office of the Chief Medical Examiner (OCME), Connecticut, to explore the feasibility of predicting body disposal sites. Three models were created: one inductive model using non-sites mimicking complete spatial randomness (CSR), one inductive model using non-sites mimicking population density (PD), and one deductive model using non-sites mimicking complete spatial randomness (CSR). Spatial statistical analyses confirm that body disposal locations are inhomogeneously distributed across Connecticut. Both the inductive and deductive CSR models generated the most optimal predictive models. Results indicate predictive models of body disposal locations are 49 – 59% more likely to predict body disposal sites in Connecticut than random chance. At present, the models are most successful at predicting body disposal sites in urban areas. The results indicate that predictive models of body disposal location have a real possibility of narrowing search areas and maximizing resources for law enforcement when searching for missing victims. Future modeling efforts should address predicting rural body disposal site location.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION……1
Introduction……..1
Context……2
Objectives……4
Thesis Outline……...5
CHAPTER 2: LITERATURE REVIEW……....7
GIS and Spatial Analysis……7
GIS……...7
Spatial Analysis……10
Spatial Analysis and Predictive Modeling in Anthropology……13
History of Predictive Modeling……13
Inductive v. Deductive Models……..17
Common Statistical Methods of Predictive Modeling……...18
Theory……19
Spatial Analysis, Predictive Models, and Forensic Anthropology……20
Geography and Crime Studies……...25
Introduction……....25
Environmental Criminology……..26
Routine Activity Theory……....26
Rational Choice Theory……27
Crime Pattern Theory……27
Geographic Profiling……..28
Journey to Crime Research……29
Disposal Site Characteristic Research……...29
Spatial Analysis and Predictive Modeling of Body Recovery Site Location……29
CHAPTER 3: DATA COLLECTION AND METHODS……33
Data Sources and Collection……..33
Body recovery location……..33
Random non-sites...……...35
Dependent variable……36

Predictive variables……37
Methods……..44
Exploratory Spatial Data Analysis……44
Inductive Model……46
Deductive Model……....49
Model Validation……...53
Summary……54
CHAPTER 4: RESULTS……...56
Exploratory Spatial Data Analysis……56
Moran's I……56
Kernel Density Analysis……57
Ripley's *K*-function……....58
Univariate and Bivariate Tests of Variables……..60
Summary Statistics……60
Univariate Tests……62
Bivariate Tests……...63
Inductive Model……67
Complete Spatial Randomness (CSR) Model……67
Population Density (PD) Model……72
Deductive Model……77
Results……77
Model Validation……...80
Summary of Results……...80
CHAPTER 5: DISCUSSION……82
Spatial Statistics…….....……82
Summary Statistics……83
Univariate and Bivariate Tests of Significance……88
Inductive Model……88
CSR Model……88
PD Model……..93
Deductive Model……99

Model Comparison……...101
Examples of Use……..103
Predictive Modeling of Body Disposal Sites: Lessons Learned……..104
Context and Evaluation……106
Data Access……..……109
CHAPTER 6: CONCLUSION……111
REFERENCES CITED…….....114
APPENDIX……..123
VITA……132

LIST OF TABLES

Table 1.1 Distance from victim's home to body disposal location (Lundrigan 2016) 30
Table 1.2. Distance to road from body disposal location 30
Table 1.3. Percentage of bodies deposited in water.....30
Table 1.4. Percentage of bodies buried.....31
Table 1.5. Body disposal sites by land use 31
Table 2.1. Original land use categories (USGS).....40
Table 2.2. Aggregated land use categories 41
Table 2.3. Deductive model independent variables and categories 51
Table 3.1. Results for Univariate Moran's I Test of Spatial Autocorrelation.....58
Table 3.2. Descriptive statistics for binary independent variables 61
Table 3.3. Descriptive statistics for continuous independent variables.....61
Table 3.4. Descriptive statistics for body disposal site location (continuous variables) 61
Table 3.5. Univariate tests of significance for independent variables between site and non-site (Complete Spatial Randomness).....62
Table 3.6. Univariate tests of significance for independent variables between site and non-site (Population Density).....63
Table 3.7. Correlation of variables to urban areas.....64
Table 3.8. Correlation of variables to open space.....64
Table 3.9. Correlation of variables to primary and secondary roads.....65
Table 3.10. Correlation of variables to tertiary roads.....65
Table 3.11. Correlation of variables to slope.....66
Table 3.12. Correlation of variables to population density.....66
Table 3.13. Logistic regression results (Complete spatial randomness model).....68
Table 3.14. Forest*Slope interaction odds ratios (Complete spatial randomness model).....68
Table 3.15. Classification matrix of correctly and incorrectly predicted sites and non-sites (Complete spatial randomness model).....72
Table 3.16. Logistic regression results (Population density model).....73
Table 3.17. Forest*Slope interaction odds ratio (Population density model).....74
Table 3.18. Classification matrix of correctly and incorrectly predicted sites and non-sites (Population density model).....77

Table 3.19. Deductive model map values, coverage, and proportion of sites……...78
Table 4.1. Summary statistics for predicted sites (Complete spatial randomness model)……..…92
Table 4.2. Summary statistics for predicted sites outside study area (Complete spatial randomness model)……...94
Table 4.3. Summary statistics for predicted sites (Population density model)……..98
Table 4.4. Summary statistics for predicted sites outside study area (Population density model)…….100
Table 4.5. Comparison of model validation results……...……..102

LIST OF FIGURES

Figure 3.1. Kernel density analysis results for known sites……...57
Figure 3.2. Ripley's *K*-function results for known site clustering and dispersion…….59
Figure 3.3. Complete spatial randomness predictive model, full range…….70
Figure 3.4. Complete spatial randomness predictive model, classified…….71
Figure 3.5. Population density predictive model, full range……..75
Figure 3.6. Population density predictive model, classified……..76
Figure 3.7. Deductive predictive model…….79
Figure 4.1. Example of known sites clustering in town vicinities…….84
Figure 4.2. Known site relationship to population density……85
Figure 4.3. Complete spatial randomness model predicted sites……...91
Figure 4.4. Histogram of site population density values……...95
Figure 4.5. Population density model predicted sites……97
Figure 4.6. Example of complete spatial randomness model applied within an established search area (#1)……..105
Figure 4.7. Example of complete spatial randomness model applied within an established search area (#2)..……106

CHAPTER ONE
INTRODUCTION

I. Introduction

While the majority of missing persons are recovered safely, there are instances in which the missing person is the victim of homicide (Lundrigan 2016). In suspected homicide cases, locating and recovering the body is crucial to law enforcement's investigation (Keppel and Weis 1994; Rossmo 2000; Lundrigan 2016). Many victims are found at the primary crime scene but some perpetrators will move the body to a secondary location. Moving a body to a secondary location separates it from the original crime scene, which may further delay body recovery, hides or destroys evidence, and complicates the investigation (Rossmo 2000). Forensic site prediction models can aid law enforcement in locating a victim's body by narrowing search areas and helping to maximize resources.

Keppel and Weis (1994) report the clearance rate of homicide cases drops from 84% to 42% when a victim's body is not recovered within the first 24 hours. DiBiase (2014) notes that only approximately five "no body" cases go to trial each year in the United States (Lundrigan 2016). In most cases, the body is either accidentally found (e.g., a hiker in the forest) or located through witness testimony, both of which are unreliable, rendering locating a victim's body a difficult process (Fruzzetti et al. 1992). Forensic site prediction models can augment relying solely on witness testimony or luck. Research in archaeology, environmental criminology, and geographic profiling suggests that human interaction with the environment is predictable and non-random (Brandt et al. 1992; Verhagen and Whitley 2012; Lundrigan and Canter 2001; Rossmo 2000). It follows that if offender interaction with the environment is patterned, offender

rationale behind the disposal and concealment of a body will also be patterned and predictable (Congram 2010; 2013).

To date, predictive modeling has not been applied in a U.S. medicolegal context. However, the growing body of research on patterned human interaction with the environment suggests that predictive modeling can be applied in a U.S. medicolegal context. This thesis explores the feasibility of predicting body disposal and recovery site locations in four Connecticut counties using data collected from Connecticut's Office of the Chief Medical Examiner.

II. Context

Body disposal sites occur at a position in space that can be recorded and understood in relation to other data elsewhere in space; consequently, body disposal site locations are categorized as spatial data (Haining 2003; Bivand et al. 2008). Spatial data present a problem for traditional statistics, as observations in geographic space are not independent. Using traditional statistical techniques on spatial data results in a reduction of statistical precision (Haining 2003; Waller and Gotway 2004). Spatial statistics are just beginning to be utilized in forensic anthropology (Congram et al. 2016). However, while predictive modeling is a nascent field in forensic anthropology, archaeologists have successfully been utilizing predictive models for years (Verhagen 2007; Verhagen and Whitley 2012; Kvamme 2006).

To conduct analyses on spatial data, Geographic Information Systems (GIS) is used. GIS allows researchers to associate data based on shared geographic locations. In its most basic form, GIS can be thought of as a spatially referenced database, but it can also calculate statistics and conduct spatial analyses (Maschner 1996; Ebert 2004; Fischer and Getis 2010).

Predictive modeling is a technique that attempts to predict the location of sites (e.g., archaeological sites or body disposal sites) in a region, given a sample of the region or a theoretical explanation of human behavior (Verhagen 2007; Kohler and Parker 1986). Predictive models project known patterns and relationships into unknown spaces. There are two main types of predictive models: inductive and deductive. Inductive models are "data-driven" models, where samples of known sites are linked with predictive variables and statistical methods are used to determine the relationship between the two. This relationship is then extrapolated to the rest of the study area, most frequently using logistic regression (Kvamme 1990; Wheatley and Gillings 2002). Deductive models are "theory-driven" models, which begin with a hypothesis of how predictive variables influence site location based on theory and previous experience. Selected variables are assigned weights based on theoretical importance, and then maps of each variable are overlaid on top of each other (Kohler and Parker 1986; Verhagen and Whitley 2012).

The theoretical basis for modeling body disposal sites is shared with archaeological justifications for predictive modelling: 1) human behavior is patterned with respect to the natural and social environment, 2) we can know or learn something about how people interacted with the environment by observing the relationship between the environment and what is left behind (e.g., artifacts or bodies), and 3) "GIS provides a tool for mapping what we know" (Kvamme 2006:4). The notion that human interaction with the environment is patterned, nonrandom, and predictable is also supported by criminological theory. Environmental criminology studies the interaction of humans with the environment, and how this interaction influences an offender's decision to commit a crime (Brantingham and Brantingham 1981; Rossmo 2000). Routine activity theory, rational choice theory, and crime pattern theory each provide insight into how human interaction

with the environment structures and influences an offender's actions when committing a crime (Chainey and Ratcliffe 2005; Rossmo 2000).

Research in spatial statistics and predictive modeling in forensic anthropology is just beginning. At this point in time, only two researchers have investigated the feasibility of creating predictive models of body disposal location. Orengo (2006) created a predictive model of body disposal location in the United Kingdom medicolegal context. Congram (2010) created predictive models of clandestine mass grave location in the context of armed conflict in Spain. Manhein et al. (2006) and Kolpan (2015) investigated patterns of body disposal in the United States medicolegal context, but neither study included predictive claims about its results.

III. Objectives

The main purpose of this thesis is to investigate the feasibility of creating predictive models of body disposal sites in the United States medicolegal context. This study examines the body disposal locations of 173 homicide cases from the state of Connecticut with the intention of answering three research questions:

1) Are body disposal sites in Connecticut patterned and nonrandom?
2) Can predictive models be created for body disposal locations in the Connecticut medicolegal context?
3) If predictive models can be created, which type of predictive modeling is most appropriate for predicting body disposal site location in the Connecticut medicolegal context?

The first objective will be achieved by utilizing spatial statistical tests. If the results indicate that body disposal sites are spatially autocorrelated, are inhomogenously distributed

across the landscape, and/or are significantly clustered, then it can be confirmed that body disposal sites in Connecticut are patterned and nonrandom.

The second and third objectives will be achieved by creating predictive models of body disposal sites using 173 known cases from the Office of Chief Medical Examiner (OCME), Connecticut, and an additional 173 simulated random non-sites for comparison.

The third objective of determining which predictive model is best suited for creating predictive models of body disposal sites in Connecticut will be achieved by assessing each model using model validation techniques. Kvamme's Gain statistic tests the site prediction values on known sites not included in the creation of the model. A Gain statistic of +0.86 is ideal (Gibson 2005), but a model with a Gain statistic over +0.5 is considered successful (Ducke and Munch 2005). Additionally, most models will predict 60 – 70% of known sites (Ebert 2000). Other forms of model validation include the positive predictability and negative predictability of the model. The most ideal form of model validation is testing the model on new data. For this reason, each model will be tested on eight sites outside of the study area, but still within the state of Connecticut.

IV. Thesis Outline

Following this introduction, Chapter 2 provides some context and the theoretical framework within which this study was conceived and designed. Spatial statistics and Geographic Information Systems (GIS) are introduced. A discussion of archaeological predictive modeling is presented, as well as a summary of current research on spatial statistics and predictive modeling in forensic anthropology. Criminological literature is reviewed to support the premise that human interaction with the environment is patterned and nonrandom.

This chapter presents a summary of background information necessary for the development of predictive models, hypotheses, statistical tests of variables, results, and discussion that come in subsequent chapters.

In Chapter 3 – Data Collection and Methods – the steps for creating predictive are outlined. Data collection for sites, non-sites, and predictive variables is explained. Two different data sets are included in the study: the first uses non-sites mimicking complete spatial randomness, and the second uses non-sites mimicking the population density of the study area. The spatial statistical methods used to test for spatial autocorrelation, site inhomogeneity, and site clustering are introduced. A series of analyses starting with univariate and bivariate tests between body disposal location and different predictive variables are outlined. Three predictive models are generated: one inductive model created using the data from 173 known body disposal sites and 173 random non-sites mimicking complete spatial randomness; one inductive model created using the data from 173 known body disposal sites and 173 random non-sites mimicking population density in the study area; and one deductive model, derived from observations of the site data and current theory of offender behavior in forensic anthropology and criminology.

Chapter 4 is a presentation of the results of the analyses and predictive models. Chapter 5 is a discussion of the significance and usefulness of the findings and predictive models. Recommendations for future predictive modeling efforts are included. Chapter 6 briefly reviews the results and implications of this thesis.

CHAPTER TWO
LITERATURE REVIEW

This chapter provides a brief introduction to Geographic Information Systems (GIS) and spatial analyses. It presents a short history of the development of predictive modeling in archaeology, and a detailed review of spatial statistics and predictive modeling research in forensic anthropology. Criminological theory is discussed to highlight environmental variables that influence offender decision to commit a crime. The chapter concludes with a discussion of how GIS, spatial statistics, predictive modeling, and criminology relates to the objective of creating predictive models of body disposal location in the United States medicolegal context.

I. GIS and Spatial Analysis

A. GIS

a. History of Geographic Information Systems

Geographic Information Systems (GIS) is a way of associating data based on shared geographic locations. It is a computer system that provides a means of storing, manipulating, analyzing, and presenting geographic data (Longley et al. 2011; Wheatley and Gillings 2002; Kvamme 1989). In simple terms, GIS can be thought of as a spatially referenced database (Maschner 1996; Ebert 2004). At a more complex level, GIS can calculate statistics and conduct spatial analyses (Fischer and Getis 2010; Longley et al. 2011).

GIS has its origins in the Canadian department of natural resources. In the early 1960s, the Canadian government needed to develop a plan to manage the renewable and non-renewable resources within the country. This involved taking inventory of the available resources, as well as evaluating how these resources were being utilized at that time and how they would be

utilized in the future (Wheatley and Gillings 2002). The Regional Planning Information Systems Division created the first GIS in an effort to address these complex and dynamic issues of cultural and natural resource management (Wheatley and Gillings 2002; Peuquet 1977; Tomlinson 1982; Longley et al. 2011). At its earliest stages, GIS was primarily used to automate mapping via computer. The technology was improved with advances in satellite imagery and military development of the Global Positioning System (GPS) (Longley et al. 2011; Chainey and Ratcliffe 2005).

Widespread use of GIS did not occur until the reduction of computer hardware costs in the 1980s. At present, GIS is ubiquitous and used for everything from in-car navigation to utilities management and military planning to crime mapping (Chainey and Ratcliffe 2005; Fischer and Getis 2010).

b. GIS Data

Spatial data are data with a geographic reference. Data occur at a position in space that can be recorded and understood in relation to other data elsewhere in space (Longley et al. 2011). There are two different types of spatial data in GIS – vector and raster. Vector data is discrete, occurring at a specific point in space. Examples of vector data include buildings, roads and designated spaces such as parks. Vector data is stored as points, lines, or polygons. Raster data is continuous data that can be measured at any point within a given study area. Examples of raster data include temperature and elevation data (Longley et al. 2011). Raster data is stored in square pixels that vary in size depending on the resolution of the data. Additional information can be included with the spatial location of data. This information is known as the attributes of

the data. Both vector and raster data can have associated attributes (Waller and Gotway 2004; Longley et al. 2011).

Working with spatial data requires a way to reference spatial location, as well as measure distances between locations and a means of describing complex shapes and their properties (Waller and Gotway 2004). One solution to this problem is the use of coordinate systems. Coordinate systems create a grid around the earth (which is not a perfect sphere, nor a perfect ellipsoid, which complicates calculations) from which to reference a point in space on earth. Longitude grid lines travel north-south around the earth, passing through the north and south poles. Latitude grid lines are perpendicular to the lines of longitude, traveling east-west around the earth. Longitude and latitude provide a means of describing the location of a point on earth by the coordinate pair (Waller and Gotway 2004; Chainey and Ratcliffe 2005). The system of longitude and latitude is called geographic coordinates (Longley et al. 2011). The most common geographic coordinate system is the World Geodetic System of 1984 (WGS84). Another geographic coordinate system commonly used in North America is the North American Datum of 1983 (NAD83). The difference between these geographic coordinate systems is based on how each treats the complications presented by earth's complex shape (Longley et al. 2011).

Three-dimensional (3D) representations of the earth are difficult to use in spatial analysis and GIS. Two-dimensional (2D) surfaces are easier and more commonly used. Representing a 3D surface on a 2D surface requires a mathematical transformation, known as a map projection. Each map projection has some level of distortion, whether it be in area, shape, distance, or direction. This is because a curved 3D surface does not naturally lie on a flat 2D surface; some distortion must be introduced to make it fit (Waller and Gotway 2004; Chainey and Ratcliffe 2005). Conformal projections preserve shape, but distort area at a large scale. Equal-area

projections preserve area, while distorting shape. The choice of map projection will depend on the project at hand. Conformal projections are best for navigational use. If quantitative assessment of area is necessary, then an equal-area projection is better suited for the task (Waller and Gotway 2004; Longley et al. 2011).

B. Spatial Analysis

a. Spatial Data

Spatial data are data with a geographic reference. Data occur at a position in space that can be recorded and understood in relation to other data elsewhere in space. Spatial data present a problem for traditional statistical analyses, as observations in geographic space are not independent. Tobler's 1st Law of Geography states that objects close together in space are more similar than objects further away in space (also known as spatial autocorrelation) (Tobler 1970; Haining 2003; Bivand et al. 2008). Dependent data provide less information than independent data due to information redundancy; spatial data carry "duplicate" information, making the effective sample size less than the number of sample collected (Haining 2003:41; Waller and Gotway 2004). Consequently, traditional statistical techniques result in a reduction of statistical precision of spatial data (Congram et al. 2016; Waller and Gotway 2004). Recognition of these limitations has led to recent developments in spatial analysis, including techniques and models that explicitly take the spatial referencing associated with data into account (Haining 2003; Bivand et al. 2008; Fischer and Getis 2010).

b. Spatial Analysis

The first step in spatial analysis is determining if the data are spatial. This requires testing for spatial autocorrelation. There are two types of spatial autocorrelation: positive and negative. Positive spatial autocorrelation visually looks like a clustering of points. It follows Tobler's Law that everything is related to everything else, but near things are more related than others (Schwarz and Mount 2006). Negative spatial autocorrelation is the clustering of dissimilar values in space (Waller and Gotway 2004). As discussed above, spatial autocorrelation is a major problem for traditional statistics and the presence of spatial autocorrelation requires spatial statistics. Moran's I is used to identify spatial autocorrelation and will be discussed in more detail in Chapter 3, section II (Schwarz and Mount 2006; Waller and Gotway 2004).

There are three types of spatial data: 1) continuous; 2) lattice models; and 3) point pattern. Continuous field data are data with observations that can be measured anywhere in space, such as temperature data. Lattice models, also known as network models, have one fixed observation. These models are often aggregated data contained within a set area of space, such as income data for a United States county. The third type of data is point pattern data. Point pattern data include observations that are irregularly or randomly distributed in space, such as the location of trees (Illian et al. 2008; Zuur et al. 2007). This thesis will focus specifically on point pattern data, since body disposal sites are point patterns.

After establishing that data are spatial in nature, the next step in analyzing point pattern data is to determine if the points deviate from complete spatial randomness (CSR). With CSR, the observed points are equally as likely to occur in one location as they are in another, given the established study area. The points are independent of each other. In other words, complete

spatial randomness occurs when the location of one point on the landscape does not influence the location of other points on the landscape. Complete spatial randomness is the baseline for understanding spatial point patterns that are more clustered than random, and spatial point patterns that are more regular than random (Waller and Gotway 2004). When points deviate from CSR, they are no longer independent. Clustered points indicate that some locations are more likely than others, so the presence of one point may attract other points to the same vicinity. Dispersed points are the opposite; the presence of one point may make other points less likely to appear in the same vicinity (Longley et al. 2011).

There are two processes to consider when studying point patterns – first and second order. First-order processes investigate point density. Point density refers to how the number of points are distributed across the study area, but does not address point independence. Second-order processes investigate the interaction between points. Both first- and second-order processes identify clustering, but they do so in different ways; first-order processes identify patterns, while second-order processes identify patterns and begin to explain their existence (Longley et al. 2011). One common method of first-order analysis is kernel density mapping. Kernel density mapping examines the intensity λ of points across the study area, which is defined as the number of expected points per unit area at a given location (Waller and Gotway 2004). The λ of points is calculated using overlapping bins called kernels. Kernel density mapping is dependent on the parameters of the kernel; if the width of the kernel changes, then the results of the density map will change (Longley et al. 2011; Chainey and Ratcliffe 2005). Ripley's *K* statistic measures second-order processes. It compares the expected average point intensity at set intervals within a study area to the observed point intensity. The *K* statistic identifies clustering of points within the study area. Like kernel density mapping, the *K* statistic is dependent on the

established parameters (Schwarz and Mount 2006; Fischer and Getis 2010). In summary, kernel density mapping identifies site patterning, and the Ripley's *K*-statistics identifies the relationship between sites that is causing the observed pattern.

It is important to note that the results of a spatial analysis change depending on the scale of the analysis and how the boundaries of the study area are drawn, which is known as the modifiable areal unit problem (Longley et al. 2011; Waller and Gotway 2004; Congram et al. 2016). The modifiable areal unit problem is inescapable; every spatial analysis is an abstract representation of the real world, and is therefore always a subsample of a larger data set. The modifiable areal unit problem is also present when representing aggregated data. The spatial analysis results will provide different answers depending on the boundaries used to aggregate the data (Longley et al. 2011). The importance of scale and the boundaries of the study area are also important to first- and second-order analyses; the results will differ depending on the size of the kernels and the distance set for the *K*-statistics (Waller and Gotway 2004).

II. Spatial Analysis and Predictive Modeling in Anthropology

A. History of Predictive Modeling

Archaeological data is inherently spatial. Sites, artifacts, features, and structures all occur at a position in space. These remains are spatially related to each other, and to other features of the environment (Verhagen and Whitley 2012; Wheatley and Gillings 2002; Wescott 2000). Although conceptions of space have evolved over time, archaeologists record and analyze the spatial relationship of archaeological sites, artifacts, and features. Detailed site maps and records of the three-dimensional location of artifacts date back to the 18th century (Wheatley and Gillings 2002). Prior to the mid-20th century, archaeologists mapped artifacts by traits to visually

establish the extent of "culture-areas" (Wheatley and Gillings 2002). By the mid-20th century, archaeologists took interest in the environmental factors influencing the observed spatial relationships among archaeological sites. Settlement pattern studies played a role in this interest, which can be traced back to prominent studies by Steward (1938) and Willey (1953) on settlement patterns and their relationship to environmental features (Verhagen 2007; Kvamme 2006; Kohler and Parker 1986). Early studies utilized distribution maps to visualize site patterns. More intensive quantitative approaches to spatial analysis evolved in the 1960s, with the advent of New Archaeology. Rather than describing the observed patterns of artifact and site distribution, archaeologists sought to explain the patterns. It is from this combination of an ecological approach to settlement analysis and the growing interest in quantitative analysis that led to the birth of predictive modeling in archaeology (Verhagen and Whitley 2012; Verhagen 2007; Wheatley and Gillings 2002).

Predictive modeling is defined as a technique that attempts to predict the location of archaeological sites in a region, given a sample of the region or a theoretical explanation of human behavior (Verhagen 2007; Kohler and Parker 1986). Predictive models project known patterns and relationships into unknown spaces (Warren and Asch 2010). The majority of predictive modeling, especially in its nascent stages, assumes that archaeological sites are related to certain characteristics of the environment (Verhagen 2007; Warren and Asch 2010).

Archaeological predictive models emerged in the late 1970s, but it was not until the introduction of GIS that they truly expanded. The computing power of GIS allowed for the development of more complex predictive models. Using GIS, Kvamme (1983; 1984; 1988) developed statistical and spatial analysis techniques that formed the basis of predictive modeling in the United States (Verhagen 2007).

Predictive modeling initially held the most appeal to those working in Cultural Resource Management (CRM). The National Historic Preservation Act (1966) states that federal agencies must "identify historic properties on their lands… and to record such properties when they must be destroyed" (King 1984; Verhagen and Whitley 2012; Verhagen 2007). CRM agencies needed techniques to predict where sites occur on the landscape, in an effort to be able to protect and/or record them. Predictive models have the potential to maximize available resources by defining the regions of high archaeological probability, and therefore, aiding CRM firms in deciding where to spend their limited funding (Kohler and Parker 1986; Dore and Wandsnider 2006). GIS continued its popularity within the CRM realm into the 1990s. Verhagen (2007) suggests that this popularity was due to the ease with which GIS appeared to simplify complex statistics and present results that were easy to understand. The ease of using GIS is viewed as both a positive and a negative. Utilizing GIS as a "black box" in which data is input and results are produced without understanding the underlying processes can lead to oversimplified, or just plain incorrect results (Wheatley and Gillings 2002; Verhagen 2007). As the use of GIS and predictive modeling grew in the 1990s, so did the critiques of their use.

Unlike their CRM counterparts, academic archaeologists were wary of incorporating GIS into their research. By the 1990s, many academic archaeologists were divided into two camps: those that accepted GIS and those that heavily critiqued it (Verhagen 2007). Part of the academic hesitation was due to a theoretical paradigm change in the discipline. The focus on quantitative methods of analysis in archaeology diminished as New Archaeology was replaced by post-processual archaeology. Post-processual archaeologists suggest a dialectic relationship exists between social action and space. Space is shaped and constructed by the social actions that occur within it; at the same time, these social actions are shaped and constructed by the

space in which they take place (Wheatley and Gillings 2002; Lock and Harris 2006). The primary concern of post-processualists regarding predictive modeling is the environmental determinism inherent in including only environmental factors in the models (Ebert 2004; Gaffney and van Leusen 1995).

The environmental determinism critique continues to be one of the strongest critiques against archaeological predictive modeling (Verhagen and Whitley 2012; Kvamme 2006). Ebert (2004) argues that too much emphasis is placed on the physical aspects of the sites, and not enough attention is paid to the social aspects. Following this argument, some suggest that it is not possible to know the social aspects influencing site location in the past (Ebert 2004; van Leusen 2002; Kvamme 2006). A similar argument is that human behavior is too unpredictable to model (Kvamme 2006). Kvamme (2006) reviews other concerns, including the critique that archaeological site distribution cannot be modeled because "site" is a meaningless concept and human behavior is continuous over the landscape. Furthermore, the site use is often unknown and therefore the sites cannot be modeled because the locational differences between different types of sites cannot be addressed (Kvamme 1988; Dore and Wandsnider 2006). Another concern is that many archaeological sites are buried and remain unknown to researchers; this missing data creates a biased model. A similar bias exists in the known sites, as the sites are frequently not systematically surveyed and recorded (Carrer 2013; Wheatley 2004; Dore and Wandsnider 2006; Kvamme 2006). There is also concern that the designation of "low archaeological potential" leads to less intensive site survey, which culminates in sites being overlooked and lost forever (Verhagen and Whitley 2012:53).

Efforts to address critiques of environmental determinism have been slow to develop, but some research on incorporating cultural and social variables does exist (Verhagen et al. 2007;

Ridges 2006; Carrer 2013; Verhagen and Whitley 2012). One example of this comes from Carrer (2013), who incorporated theory from ethnoarchaeology into his predictive model of site location in Italy. As for the critique of biased data, Verhagen et al. (2007) note that the lack of appropriate data sets is not purely an archaeological predictive modeling issue; biased data sets are characteristic of how most archaeological data is collected. Verhagen (2007) also argues that the failure to survey areas of "low archaeological potential" is not just the result of a map that marks "high" and "low" areas. CRM archaeologists are forced to make decisions about what areas are surveyed and which are not based on their limited time and monetary resources. Predictive model or no predictive model, some areas will receive the "low archaeological potential" designation. Using predictive models provides a quantitative basis for making those designations (Verhagen 2007; Verhagen and Whitley 2012).

B. Inductive v. Deductive Models

Predictive model methodology is most frequently discussed as a dichotomy between inductive and deductive modeling. Inductive models, also known as "data-driven" models, are a quantitative approach to modeling (Kvamme 1990; Wheatley and Gillings 2002; Verhagen 2007). Samples of known archaeological sites are linked with environmental factors. Statistical methods are used to determine the relationship between the sites and the factors. Once the relationship between the sites and the environment is known, the relationship can be extrapolated to the entire study region to provide probabilities of site location (Kohler and Parker 1986; Verhagen and Whitley 2012). Inductive models are the most commonly used in United States CRM. Deductive models, also known as "theory-driven" models, begin with a hypothesis on factors influencing site location based on previous research and theory based on human activity

in the area (Wheatley and Gillings 2002; Verhagen 2007). Selected variables are weighted given their theoretical importance. Maps of each variable are overlaid on top of each other. The total score of the weighted variables determines the areas of high and low site probability (Kohler and Parker 1986; Verhagen and Whitley 2012).

In practice, the dichotomy between inductive and deductive models is a false one. Most models involve a mix of inductive and deductive methods (Verhagen and Whitley 2012; Verhagen 2007; Kvamme 2006). For example, deciding which variables to include for statistical analysis in an inductive model requires some previous knowledge of which variables may be important in the region. Alternatively, statistical methods can be used to determine which variables are significant, and then the variables can be weighted based on theory to create the final model.

C. Common methods of predictive modeling

Logistic regression is a useful statistical tool for inductive modeling. It is an appropriate fit for modeling site location, as it analyzes binary dependent variables (e.g. site v. non-site). Furthermore, logistic regression is flexible, does not require normality of data, and it works with categorical variables. Most inductive archaeological predictive models have used logistic regression (Warren and Asch 2000; Conolly and Lake 2006).

The 'map algebra', or weight map-layer approach, is the most common method of deductive modeling. As mentioned above, each variable map layer gets an assigned value based on theoretical importance and the predictive model is created by reclassifying each layer, assigning each layer a weight, overlaying the layers and adding them together (Brandt et al. 1992; Kohler and Parker 1986).

D. Theory

Kvamme (2006:4) simplifies the theoretical basis for predictive modeling into three points:

> "1. Human behavior is patterned with respect to the natural environment and to social environments created by humanity itself.
> 2. We know or can learn something about how people interacted with these environments by observing relationships between human residues (i.e. the archaeological record) and environmental features.
> 3. GIS provides a tool for mapping what we know."

Support for Kvamme's (2006) suggestion that human behavior is patterned with respect to the natural environment stems from quantitative ecology and niches. A species niche is "the total range of conditions in the environment under which a population lives and replaces itself" (Pianka 1974: Kvamme 2006). Kvamme (2006) argues that in theory, humans can inhabit all environments on earth; yet, in reality, humans are limited by unfavorable conditions (e.g. steep cliff faces, bodies of water, glaciers, etc.). Furthermore, some of these human niches are only accessible to those with the technology and resources to reach them. In addition to the limited area for human inhabitance, certain activities are linked to certain environmental characteristics (e.g. fishing occurs in lakes or streams). Consequently, archaeological predictive modeling is justified given the observed characteristics of niches; analysis of favorable physical and social characteristics in the environment provides an opportunity to identify favorable and unfavorable regions for archaeological site location (Kvamme 2006).

Kvamme's (2006) argument that we can know or learn something about people's interaction with the environment by studying the archaeological record is supported by site formation theory (Schiffer 1983). Schiffer (1983) notes that formation processes create the archaeological record. Artifacts are ingrained in human society; they reflect behavior and

organization, actively structure social interaction, and participate in most activities (Hodder 1982; Rathje and Schiffer 1982; Rathje 1979; Schiffer 1983). To understand the patterns of artifact deposition, archaeologists must take formation processes (both cultural and non-cultural) into account. In practice, this means incorporating knowledge of the "past agencies" into the interpretation of the observed pattern of artifacts at a site (Schiffer 1983:697). What this means for predictive modeling is that human behavior can be inferred (to a degree) by studying the relationship between artifacts and environmental and social features of the landscape.

If patterns exist between regional variables and known site locations, a model can be constructed through exploitation of these relationships (Brandt et al. 1992). This is the final link in Kvamme (2006)'s argument in favor of predictive modeling. Once the associations of environmental and social variables are made, these associations can be exploited using the tools within GIS to create a predictive model of site location.

E. Spatial Analysis, Predictive Models, and Forensic Anthropology

Very little research has been published on spatial analysis and predictive modeling within the field of forensic anthropology. The following is a chronological detailed review of the limited available research.

Reddick (2005) reviews the benefits of remote sensing, geophysics, and GIS to search for mass graves. Her use of GIS does not focus on predictive modeling, but on utilizing GIS mapping capabilities to record crime scenes in detail. Computerized mapping has been used to investigate and record mass graves in Former Yugoslavia and Cambodia. Reddick (2005) argues that while the outcome of these maps was beneficial, GIS was not used to its full capacity as a reconstruction tool. GIS can be utilized at almost all levels: in search, planning and logistics,

excavation, analysis, as well as data and information control and dissemination. Reddick (2005) uses GIS to analyze the location of mass grave location in relation to the landscape to identify search parameters for possible clandestine mass graves in the Former Yugoslavia, to analyze potential vehicle routes between primary and linked secondary sites, and to analyze the relationship between minimum number of individuals (MNI) and mass grave location. The variables included in her analysis include distance to road, land use, satellite image spectral analysis, population, and slope.

Manhein et al. (2006) utilize GIS to determine if there is a selective bias in where bodies are deposited. The study has a sample size of 175 forensic anthropology cases from the state of Louisiana. Descriptive statistics for the cases reveal that approximately three-quarters of the cases were dumped in a rural area, and approximately half of these cases were dumped in a wooded environment. A very low percentage of bodies were recovered from water environments (6%). Body recovery sites are typically within half a mile of a road. The results of the study indicate that if an individual is not recovered within 12 hours after being reported missing, it is more likely that the individual has been dumped in a rural area. Manhein et al. (2006) utilize spatial analysis to examine the relationship between the postmortem interval and dispersion of skeletal remains, but they do not conduct spatial analysis or make predictive claims about body recovery location in their study.

Orengo (2006) applies GIS to the analysis of nonurban body recovery sites in the United Kingdom. He utilizes the deductive technique to develop a predictive model depicting the probabilities of body recovery site location in Leicestershire, United Kingdom. The theoretical basis for his predictive model is that body dump sites and archaeological sites are subject to the same social and environmental influences. Orengo (2006) first establishes the theory behind his

model. His basic assumptions are that a search following a road network will have greater success than a normal search (Streed 1989; Killam 2004); that a body will be dragged an average of 50 feet from a road or trail on even terrain (McLaughlin 1974; Morse et al. 1983; Rossmo 2000); that terrain slope will influence where a body is left and whether or not it is buried; that bodies are often found on the right passenger side of the road headed outbound from a city or town (Killam 2004); that most disposal sites are 30-45 minutes from where the body was picked up (Streed 1989; Killam 2004; Rossmo 2000); that lakes, deep rivers, and canals are likely points of body disposal (Killam 2004; Rossmo 2000); that offenders are drawn to pre-constructed hiding places, such as mine shafts, wells, and other pre-existing holes (Killam 2004; Levine et al. 1984); that it is possible to predict where graves are feasibly dug based on soil profiles, land use, and underlying geography (Hunter 1996); that the body recovery site is accessible in the dark (Killam 2004); and that the offender is heading from a high density area to a low density area (Orengo 2006:186). Orengo (2006) collected relevant maps (e.g. Digital Elevation Models (DEM) for slope) and created buffers around certain features (e.g. roads and cities). Each of these maps was assigned a gradient of values based on the above theoretical outline. As is typical of researchers using the deductive method, Orengo (2006) overlaid the maps using GIS. The combined values from each map are the predicted values representing the final predictive model of body recovery location. While Orengo (2006) utilizes relevant criminological theory to create his predictive map, he was denied access to body recovery locations in his study area, preventing him from conducting any model validation studies. Despite his inability to test the model, Orengo's (2006) study is valuable as the only example of a predictive model of body recovery location in the medico-legal context.

Tuller et al. (2008) utilize spatial analysis and GIS to analyze spatial relationships between disarticulated and commingled remains within a mass grave. Spatial location data were obtained for different parts of the body and recorded as points to be entered into GIS. A spatial analysis program was used to calculate the distance between bodies missing elements and potential matching body parts. The output of this analysis is a list of every possible match within the grave, ordered by distance. Tuller et al. (2008) utilize spatial analysis to examine dispersion of skeletal remains within a mass grave, but do not look at relationships between grave sites or make predictive claims regarding site location in their study.

Congram (2010) utilizes spatial analysis, predictive modeling, and GIS to analyze perpetrator behavior during the Spanish Civil War. He uses spatial analysis and GIS to investigate the hypothesis that the location of clandestine graves is patterned and nonrandom. Congram (2010:20) cites Kvamme (2006:6) for the theoretical basis of his research, noting that "human behavior is patterned with respect to the natural environment and to social environments created by humanity itself", and that "we know or can learn something about how people interacted with these environments by observing relationships between human residues and environmental features". The variables chosen for analysis were number of victims in the grave, distance travelled between origin and kill site, land use/cover at kill site, road side, road type, soil type, slope, distance to last town, distance to next town, distance from road to grave, number of females among victims, and population density change. These variables were linked to the 47 known sites and 100 non-sites, and statistically analyzed for significance. The sites were also tested for spatial autocorrelation using Univariate Moran's I. Congram (2010) created three models: one inductive model and two deductive models. The inductive model was built using logistic regression. The two deductive models were built using the weighted intersection

method. The final models were produced in ArcMap (ESRI 2009). All models were validated using Kvamme's Gain statistic. Congram's (2010) results suggest that perpetrator behavior when creating clandestine graves is predictable and nonrandom. The consistent pattern of distance from kill site to nearest town, and from grave location to nearest town, suggests that killers wanted to bring their victims to a remote location before killing them. The tendency to move to a remote location is also reflected in the decrease in population density between origin and kill sites. Congram's (2010) research is the first example of using known site locations to produce predictive models of mass grave site location.

Kolpan (2015) presented a paper on utilizing GIS to analyze geographic and demographic patterns related to forensic case recovery locations in Florida. She entered 92 known forensic body recovery site locations between the years 2007 and 2012 into ArcGIS (ESRI 2014). These known sites were linked with geospatial variables, including physical land cover, population density and crime rate. Kolpan (2015) found that 77% of cases were recovered in urban areas, but most cases were found in areas of low population density and very low to medium crime rates. While Kolpan (2015) utilized GIS to identify patterns in the data, she did not conduct spatial analysis or make predictive claims about forensic body recovery location.

Congram et al. (2016) discuss the application of spatial statistics to forensic anthropology. They review spatial statistical methods, such as Moran's I, Ripley's K, Geographically Weighted Regression, Kriging, and Bayesian analysis. The authors present a case study from Bosnia and Herzegovina in which spatial statistics are applied to known clandestine burial sites. The variables included in their study were vegetation type, presence and size of human settlements, road types, and several other environmental factors. The authors used Average Nearest Neighbor (ANN) to test if the observed mean distance between sites was more

or less than the expected mean distance for sites in the study area. Ripley's K was also used to test for clustering in the study area. Based on the results of these statistical tests, Congram et al. (2016) argue that clandestine graves in their study area of Bosnia and Herzegovina are spatially clustered and typically within 8 – 9.5 km of each other.

III. Geography and Crime Studies

A. Introduction

Geography is a key component to contemporary crime studies. Like archaeological data, crime is inherently spatial. Crime occurs at a given location with a geographic reference point (Chainey and Ratcliffe 2005). The use of geographic techniques in crime analysis began as early as the mid-19th century, when two men mapped violent and property offenses with poverty levels to see if there was a spatial relationship (Rossmo 2000; Chainey and Ratcliffe 2005). Interest in the spatial component of crime continued to evolve in the 20th century through research conducted by the Chicago School, a group of researchers and a set of sociological theories originating from the School of Sociology at the University of Chicago (Chainey and Ratcliffe 2005). Burgess (1925) examined the relationship of juvenile delinquency, boys' gangs, crime, poverty, and single-parent families in relation to location in Chicago (Chainey and Ratcliffe 2005:83). At present, geography is a primary component of crime studies, including environmental criminology, geographic profiling, and forensic spatial analysis (Rossmo 2000; Chainey and Ratcliffe 2005). As with archaeological predictive modeling, the spatial analysis of crime evolved following improved technological advances and a decrease in cost of these technological advances. Law enforcement now implements GIS in various scenarios, including: identifying crime hotspots for crime reduction responses, mapping and helping to understand

crime distribution through pattern analysis, and using maps to present public crime statistics to the general public (Chainey and Ratcliffe 2005).

B. Environmental Criminology

Environmental criminology is a branch of criminology spearheaded by Brantingham and Brantingham (1981; 1984; 1998) that focuses on the interactions between people and the environment, and how these interactions influence an offender's decision to commit a crime (Rossmo 2000; Lundrigan 2016). The primary concern in environmental criminology is the crime setting, or the crime place. The environmental focus ranges in scale from micro-environments to macro-environments. Consequently, there are various theoretical approaches that spawned from environmental criminology, including the consequence model, human ecology, pattern theory, rational choice theory, routine activity theory, and strategic analysis (Brantingham and Brantingham 1998; Rossmo 2000). Routine activity theory, rational choice theory, and crime pattern theory are the three theoretical approaches that are most relevant to understanding the spatial behavior of offenders (Lundrigan 2016; Rossmo 2000).

C. Routine Activity Theory

Routine activity theory suggests there are three components to a crime: a potential offender, the presence of a suitable target, and the lack of a capable guardian. These three ingredients must come together at the right place and time for a crime to occur. The underlying basis for routine activity theory is the suggestion that the risk of a crime occurring will change over the course of a day, depending on the daily routine of both the offender and the potential target (Chainey and Ratcliffe 2005; Lundrigan 2016; Rossmo 2000).

D. Rational Choice Theory

The basis for rational choice theory is the belief that offenders decide to commit a crime after weighing the pros and cons of the situation. This viewpoint suggests that crime is a rational decision fueled by a certain goal. These decisions can be long-term (criminal involvement decision), or short-term and more immediate (criminal event decision) (Chainey and Ratcliffe 2005). Rational choice theory focuses on understanding the decision-making process related to a crime. In the context of body disposal site location, the offender must weigh the pros (well-hidden body) with the cons (discovery and apprehension) when deciding where to dispose of a body (Lundrigan 2016).

E. Crime Pattern Theory

Crime pattern theory is a convergence of routine activity theory and rational choice theory (Chainey and Ratcliffe 2005). Crime pattern theory focuses on the relationship between the crime and the offender's routine activities. Everyday activities and routines create a 'cognitive map' of places, routes and associations with areas. The more established the 'cognitive map' is, the more well-known an area is and the more comfortable an individual is in that area. In other words, an offender's cognitive map indicates the offender's level of familiarity with the area. Brantingham and Brantingham (1984) emphasize that the cognitive map includes not just the physical environment, but also the social and economic environments. The places that individuals travel throughout the day (e.g. school, work, home) and the routes they take to get there become part of an 'awareness space' (Brantingham and Brantingham 1981; Chainey and Ratcliffe 2005; Lundrigan 2016).

Offenders are more likely to commit a crime within their 'awareness space' because they are comfortable in the environment and are cognizant of the escape routes. The 'awareness space' is also defined by the offender's access to resources. For example, if an offender does not have access to a car, the 'awareness space' will be much more limited than the awareness space of an offender with a car (Chainey and Ratcliffe 2005).

Even when an offender's 'awareness space' is large, research indicates that crimes still occur within a radius of the offender's home (Chainey and Ratcliffe 2005; Rossmo 2000; Lundrigan 2016). Zipf's principle of least effort, known as Zipf's Law, states that offenders will typically travel the shortest distance possible from an anchor point (e.g. home or work) to commit a crime (Rossmo 2000; Chainey and Ratcliffe 2005). Crime offense decreases with increased distance from an offender's home in a pattern known as distance decay effect (Rengert et al. 1999; Rossmo 2000; Chainey and Ratcliffe 2005). Committing crimes close to home is convenient, but also places the offender at higher risk of discovery. Brantingham and Brantingham (1981) suggest there is a buffer zone directly around the home base in which no crimes are committed.

F. Geographic Profiling

Geographic profiling utilizes theory on offender behavior to predict the probable area of offender residence. It is most frequently applied to serial murders or rapes, but has also been used in single crimes and bombings (Rossmo 2000). Geographic profiling incorporates both inductive and deductive techniques. The inductive portion is the analysis of geographic sites related to a crime, such as victim encounter, attack, murder, and body deposition sites, using spatial statistics. The deductive portion is the reconstruction and interpretation of the offender's

"awareness space" (Lundrigan 2016; Rossmo 2000). The two are combined into a "strategic information management system" (Rossmo 2000:211). The primary version of GIS used to produce final predictive maps of offender residence is called Rigel (Rossmo 2000).

G. Journey to Crime Research

Journey to crime research focuses on different summary statistics related to a crime. These studies have primarily examined the distance from offender residence to body recovery site, but recent studies have begun to include other measures, such as victim's residence, murder site, and location the victim was last seen (Lundrigan 2016; Hakkanen et al. 2007). Journey to crime studies typically indicate that the distance between the offender's home and the crime is short (Chainey and Ratcliffe 2005; Lundrigan 2016). See Table 1.1 for summary statistics on distance from victim residence to body disposal site.

F. Disposal Site Characteristic Research

Body disposal site characteristics are frequently left out of journey to crime research, but knowledge of underlying patterns in recovery location may aid in developing search strategies (Lundrigan 2016). See Tables 1.2 – 1.5 for summary statistics on body recovery locations.

IV. Spatial Analysis and Predictive Modeling of Body Recovery Site Location

This thesis proposes that forensic body disposal and recovery site locations are predictable and nonrandom. While little research has been conducted on the spatial relationships of body deposition sites, the research that does exist supports this assertion (Manhein et al. 2006;

Table 1.1. Distance from victim's home to body disposal location (Lundrigan 2016).

Source	Median Distance	Mean Distance	Distance in Percentage of Cases	Details
Fowler (1994)			9 km	Majority of cases
Hanfland et al. (1997)			< 2.4 km	37% of child victim cases
Safarik et al. (2000)			Within 6 blocks	56% elderly female victims
Nethery (2004)		10 km		Child victims
		30 km		Adult victims
Hakkanen et al. (2007)	19.8 km	56.5 km		Adult victims
Van Patten & Delhauer (2007)	7.4 km	91.6 km		Sexual homicide victims

Table 1.2. Distance to road from body disposal location.

Source	Distance from Road	Details
Keppel and Birnes (1995)	Within 45 m	Child victims
	Within 15 m	Adult victims
Burton (1998)	88% within 43 m	Child victims
	97% within 91 m	Child victims
Rossmo (2000)	88% within 50 yards	Child victims
	97% within 100 yards	Child victims
	100% within 150 yards	Child victims
Hakkanen et al. (2007)	26.8% < 200 m	Adult victims
	43% < 500 m	Adult victims
	29.2% > 1000 m	Adult victims

Table 1.3. Percentage of bodies deposited in water.

Source	Percentage in Water	Details
James (1991)	21%	Female victims
Rossmo (2000)	20%	Child victims
Nethery (2004)	12%	Child victims
Hakkanen et al. (2007)	27%	Adult victims

Table 1.4. Percentage of bodies buried.

Source	Percentage Buried	Details
James (1991)	29%	Female victims
Rossmo (2000)	12%	Child victims
Hakkanen et al. (2007)	43%	Victims recovered in woods.

Table 1.5. Body disposal sites by land use.

Source	Percentage	Land Use
Rossmo (2000)	45.2%	Residential
	3.8%	Commercial
	5.8%	Industrial
	1%	Institutional
	5.8%	Park
	12.5%	Rural or agriculture
	21.2%	Wilderness or uninhabited
	4.8%	Unknown
Nethery (2004)	38%	Forested
	14%	Beside side-road or path
	12%	Water
	10%	Farm land
Hakkanen et al. (2007)	73.2%	Woods

of body deposition sites, the research that does exist supports this assertion (Manhein et al. 2006; Congram 2010; Kolpan 2015; Congram et al. 2016). The theoretical basis for predictive models of forensic body disposal and recovery site location is similar to Kvamme's (2006) simple list in support of archaeological predictive modeling. Archaeological research (Schiffer 1983; Verhagen 2012; Carrer 2013) indicates that past human behavior is patterned with respect to the natural and social environment. Criminological research (Rossmo 2000; Lundrigan and Canter 2001; Lundrigan 2016) suggests that present human behavior is also patterned with respect to the natural and social environment. Consciously or not, humans make decisions about how to move across the landscape based on considerations of the available resources and the constraints posed by the land (Lundrigan 2016). Something about the decision-making process can be understood by studying what remains on the landscape, whether that is archaeological remains or a human body (Kvamme 2006).

GIS, spatial statistics, and predictive modeling are powerful tools that can be utilized to understand the relationship between forensic body disposal and recovery site location and the surrounding physical and social environment. Predictive models of forensic body disposal and recovery location sites can provide law enforcement with the tools to maximize their resources and narrow search locations during search and recovery efforts. Such models are not intended to replace the current methods for search and recovery; predictive modeling is yet another tool to add to the tool box.

CHAPTER THREE
DATA COLLECTION AND METHODS

This chapter conducts spatial analyses of various aspects of body disposal location in Connecticut and develops three predictive models: two inductive and one deductive. The objective is to prove that body disposal locations in Connecticut are patterned, nonrandom, and predictable. Two data sets are used in creating the models: one data set with random non-sites simulated to mimic complete spatial randomness, and one data set with random non-sites simulated to mimic the population density of the study area. The population density data set tests the assumption that body disposal sites are distributed inhomogenously across the landscape. Logistic regression is used to evaluate the relationship between the dependent variable (site/non-site) and predictive variables in the inductive models. The deductive model variable weights are derived through a combination of logical reasoning and theoretical assumptions about offender behavior in forensic anthropology and criminology.

This chapter will begin by discussing the data used in the study. Following data collection, this chapter reviews the spatial statistical analyses used. The methods utilized for inductive model and deductive model creation are explained, and the methods of model evaluation are discussed.

I. Data Sources and Collection

A. Body recovery location

The study area includes four counties (Hartford, New Haven, New London, and Fairfield) in the state of Connecticut (CT). In the 2000 Census, Connecticut's population fell in the middle of the overall state range (29 out of 50), with 3,405,565 people (FBI 2016). The state's

population also fell in the middle range of northeastern states (5 out of 9), as reported in the 2014 FBI Crime Statistics. The homicide rate per 100,000 people was 2.9 in 2000 and 3.3 in 2012, which is similar to the overall homicide rate per 100,000 people in the northeast United States (FBI 2016). While no state will accurately represent the entire United States, Connecticut is at least representative of the more populated states in the Northeast.

Known geographic location of body disposal sites was collected on-site at the Office of the Chief Medical Examiner, Connecticut (OCME). A total of 173 cases were selected based on the following criteria: 1) the manner of death is homicide, 2) the body disposal site is not within a residential structure, and 3) the body recovery occurred between 1990 and 2015. A search engine within the OCME records database identified potential cases based on the key words "homicide" and "[year]". Cases were excluded when the individual was left within a residential yard and no attempt was made to hide the body. Cases were also excluded if the victim was found in a moving body of water. Data were not limited to anthropological cases. All data recorded were anonymized and de-identified.

Case locations were recorded by address, cross streets, or a general description of the area. The longitude and latitude were obtained for each location using either Google Maps (maps.google.com) or freegeocoder.com, which also uses the Google Maps API to search for the supplied address. Known body recovery site locations were entered into ArcGIS 10.3 (ESRI 2014) and a shapefile of body recovery location points was created. Shapefiles are a data file commonly used in GIS that store location and attribute information of spatial data.

B. Random non-sites

It is necessary to include simulated random non-site locations in the model for comparison with the known body disposal site locations (Carrer 2013; Brandt et al. 1992). The inductive model uses logistic regression, in which the dependent variable is a binary measure (e.g., site or non-site). The deductive model uses random simulated non-sites for comparison in the univariate tests of independence. In both the inductive and deductive models, random simulated non-sites are compared to the known sites. If the sites and simulated non-sites are significantly different, it indicates that the sites are nonrandomly distributed in the study area.

Non-site locations (n=173) were generated in the statistical program "R" (R Core Team 2015). The sites were exported as a shapefile and imported into ArcGIS 10.3 (ESRI 2014) for analysis. This study assumes body disposal locations are not homogeneously distributed across the landscape and that body disposal locations are more common near areas of higher population. Consequently, the random non-site locations were generated using a Poisson distribution to simulate points that mimic population density (PD) in the study area. The benefit of generating points using a heterogeneous Poisson distribution is that the method considers that populations are not distributed homogeneously across the landscape; generating points with complete spatial randomness (CSR) assumes homogeneity (Bivand et al. 2008; Waller and Gotway 2004). Adjusting the density of points to match population density controls for effects of population on cluster analyses. If clusters are detected, the clusters will represent clusters of body recovery site locations and not just clusters of people on the landscape. This is a technique commonly used to identify disease clusters in public health research (Waller and Gotway 2004).

A second set of non-site locations (n=173) were generated in ArcGIS 10.3 (ESRI 2014) using the Create Random Points tool. This set of non-site locations represents complete spatial

randomness (CSR), and does not take population density into account. The CSR non-site points will be run in a separate model from the population density (PD) non-sites.

C. Dependent Variable

The dependent variable is a dichotomous variable recording site presence or absence. The known site (n=173) and random PD non-site data (n=173) were combined to form the dataset (n=346). The data were split into a training set and a validation set, with 80% of the data included in the training set and 20% reserved for the model validation. Reserving the validation set data allows for cross-validation of the model using data that was not included in the model (Kohler and Parker 1986). A subset of 138 known sites and 138 non-sites were randomly chosen for the training set. The remaining 35 known sites and 35 non-sites were withheld as the validation set to test the predictive model.

The same data organization methods were applied to the known site (n=173) and random CSR non-site data (n=173) to form a second dataset (n=346). A subset of 138 sites and 138 CSR non-sites was randomly chosen for the training set, and the remaining 35 known sites and 35 CSR non-sites were withheld as the validation set.

D. Predictive Variables

Variables were derived from geospatial datasets freely available online (Connecticut Department of Energy and Environmental Protection; University of Connecticut MAGIC; United States Geographical Survey). All shapefiles were converted into raster data using the Polygon to Raster tool (ESRI 2014). Due to the limited amount of freely accessible geospatial datasets, not all the shapefiles are the same map scale and not all the geospatial data fall within the 1990 –

2015 timeframe. Map scale indicates how detailed a map can be in representing the location and shape of geographic features. Larger map scales include more details; as map scales decrease, feature boundaries are smoothed, simplified, or not shown at all (ESRI 2016a). To provide an example, a map scale of 1:20,000 will include more detail than a map scale of 1:100,000. Map scale is linked to data resolution. The larger the map scale, the better the data resolution (ESRI 2016a).

a. Population Density

Rossmo (2000) notes that crime frequently occurs in remote or out of the way locations. One measure of the remoteness of an area is population density. There is an inverse relationship between population density and remoteness; a higher population density corresponds to a lower measure of remoteness. Traditional representations of population density are based on census data. Census data from 2010 were accessed through Topologically Integrated Geographic Encoding and Referencing (TIGER) at a scale of 1:100,000. Area (km^2) was calculated for each census block in the study area. Population density was calculated by dividing the population count for each block by the area. This measure of population density takes residential populations into account and represents the location of populations at night. Another interpretation of population density is ambient population, which takes diurnal movements and collective travel habits into account to provide a 24-hour average of a population in a given location (ORNL 2016). Ambient population data is available from Oak Ridge National Laboratory for the year 2010. The ambient population raster cell values are normalized population counts. The raster is in the WGS 84 geographical coordinate system and cannot be projected into a different coordinate system without re-sampling the data and producing

inaccurate population counts; therefore, the ambient population raster was not included in this analysis (ORNL 2016). The data for this analysis is in the NAD 83 geographical coordinate system.

Urban area was also included as a measure of population density. Urban area can be defined two ways. The first is to define a county as metropolitan (metro) or non-metropolitan (non-metro). The Office of Management and Budget (OMB) defines metro counties as counties with a core of high population density, surrounded by less dense areas of industry and commuter residence. Non-metro counties are those that do not meet these standards (USDA 2017). This standard of measurement does not work in the given study area, as every Connecticut county except Litchfield is assigned metro standing (USCB 2017). Another measure of urban v. rural is assignment by population density by census block. The U.S. Census's urban-rural classification system defines rural as areas with fewer than 2,500 residents. Urban areas are the more densely populated areas surrounding the rural areas (USDA 2017). The U.S. Census definition of urban and rural was used for this analysis. The data was accessed via TIGER for the year 2010 at a map scale of 1:100,000.

b. Land Use

Environmental criminology and geographic profiling research indicates that certain areas are more "attractive" to perpetrators due to possible targets or environmental location (Brantingham and Brantingham 1998; Rossmo 2000). Rossmo (2000) cites parking lots, dark city streets, university campuses, school playgrounds, and rural roads as examples of areas with higher crime rates. There is also evidence indicating that water environments are favorable for body deposition. Rossmo (2000), James (2001) and Hakkanen et al. (2007) all reported bodies

recovered in water in at least 20% of cases. To take different types of areas into account, land use is included as a predictive variable. Land use data were obtained for the years 1970 – 1985 at a map scale of 1:250,000 from the U.S. Geological Survey.

The original file included thirty different subsets of land use (Table 2.1). For this analysis, the data were aggregated into seven land use files based on similar characteristics (Table 2.2). The four subcategories of agricultural land were combined with the seven subcategories of barren land, and the three categories of rangeland. Agricultural land, barren land, and range land each represent areas of open space and similar land cover. The subcategories of forest land were similarly aggregated together. The four subcategories of water were combined with the two subcategories of wetland due to the shared presence of water.

The seven subcategories of urban or built up land were not as broadly aggregated. Visual assessment of the data suggests that many body disposal site locations are in urban areas. Furthermore, the subcategories of urban or built up land use have different levels of "attractiveness" for crime, as discussed in Rossmo (2000). For this reason, it was important to draw a distinction between the different types of land use. Residential land use stands alone as a category, as do commercial and services land use. Industry is combined with industrial and commercial complexes due to shared characteristics of building type. Mixed urban or built up land use, other urban or built up land use, and transportation, communication, and utilities were also combined as a category.

The land use data does not explicitly take parks and other public open space into account, so municipal and private open space was also included. The open space data set from the University of Connecticut dates to 1997, and includes parks, cemeteries, and other green spaces. The map scale is 1:24,000.

Table 2.1. Original land use categories (USGS).

Land Use Category	Land Use Subcategory
Urban or Built Up Land	Residential
	Commercial and services
	Industrial
	Transportation, communication and services
	Industrial and commercial complexes
	Mixed urban or built up land
	Other urban or built up land
Agricultural Land	Cropland pasture
	Orchards, groves, vineyards, nurseries, ornamental
	Confined feeding operations
	Other agricultural land
Rangeland	Herbaceous
	Shrub and brush rangeland
	Mixed rangeland
Forest land	Deciduous
	Evergreen
	Mixed forest land
Water	Streams and canals
	Lakes
	Reservoirs
	Bays and estuaries
Wetland	Forested wetland
	Non-forested wetland
Barren land	Dry salt flats
	Beaches
	Sandy area not beach
	Bare exposed rock
	Strip mines, quarries, gravel pits
	Transitional areas
	Mixed barren land

Table 2.2. Aggregated land use categories.

Aggregated Land Use Category	Subsumed Land Use Category
Residential	Residential
Industrial	Industrial
	Industrial and commercial complexes
Commercial and services	Commercial and services
Mixed and other urban or built up land, transportation communication, and services	Mixed urban or built up land
	Other urban or built up land
	Transportation, communication, and services
Agricultural, range and barren land	Cropland pasture
	Orchards, groves, vineyards, nurseries, and ornamental
	Confined feeding areas
	Other agricultural land
	Herbaceous
	Shrub and brush rangeland
	Mixed rangeland
	Dry salt flats
	Beaches
	Sandy area not beach
	Bare exposed rock
	Strip mines, quarries, gravel pits
	Transitional areas
	Mixed barren land
Forest land	Deciduous
	Evergreen
	Mixed forest land
Water and Wetland	Streams and canals
	Lakes
	Reservoir
	Bays and estuaries
	Forested wetland
	Non-forested wetland

c. Slope

Perpetrators moving a victim's body are subject to time, distance, speed, and effort constraints. These constraints will influence where a body is deposited (Lundrigan 2016; Rossmo 2000). Naismith's rule is a technique to estimate travel time to a body deposition site location based on distance, degree of travel difficulty, elevation, and load (Rossmo 2000). The present model includes slope instead of elevation. Travel can occur at almost any elevation; it is the change in elevation over the distance travelled that will constrain body movement efforts. Elevation data was obtained from the USDA Natural Resources Conservation Service, and is part of the National Elevation Data Set (10 Meter). The National Elevation Data Set is a mosaic of different Light Detection and Ranging (LIDAR) datasets collected between 2004 and 2014. Slope was calculated for the study area using the Slope tool in the Spatial Analyst Extension in ArcGIS 10.3 (ESRI 2014).

d. Distance to Road

This study combines primary and secondary roads into one group for analysis. TIGER (2010) defines primary roads as divided highways that are included in the interstate highway system or under State management. Primary roads are distinguished from secondary roads by the presence of interchanges, are accessible by ramps, and may include highway tolls. Secondary roads are "main arteries" with one or more lanes of traffic that frequently intersect with other roads and driveways (TIGER 2010). Secondary roads are usually included in the U.S. Highway, State Highway, and/or County Highway system and have both a local name and route number (TIGER 2010). Primary and secondary roads are included in this analysis as they are considered more "porous" and provide the quickest path of exit from areas of high population

density (Brantingham and Brantingham 1995; Rossmo 2000). Furthermore, roads are an important aspect of an offender's mental map. Roads represent both "paths", or routes of travel, and "nodes", or intense areas of activity (Lynch 1960; Rossmo 2000).

Primary and secondary road data were accessed through the University of Connecticut from the year 2010 at a map scale of 1:100,000. A buffer was created using the Buffer tool in ArcGIS 10.3 (ESRI 2014) to determine distance from road. Previous predictive modeling studies (Orengo 2006; Congram 2010) emphasize the importance of distance to road in locating body disposal sites. For this reason, the distance of each known site in the dataset (n=174) to the nearest primary and secondary road was calculated. The 95th percentile for the dataset was calculated in "R" by bootstrapping the data by the median, as the data are not normally distributed and are skewed (R Core Team 2015). This value (1865 ft.) was used as the assigned buffer for this study. The assigned buffer is similar to previous crime research. While Keppel and Birnes (1995) and Burton (1998) suggest that victims are found within 200 ft. of the road, Hakkanen et al. (2007) found this distance extended even further to 1640 ft. from the road.

Tertiary roads were also included in the analysis. Tertiary roads are lower volume roads that typically connect smaller settlements in rural areas. These roads can also connect minor streets to more major (primary and secondary) roads in more populated areas. Tertiary roads generally have lower speed limits than primary and secondary roads, and may be unpaved in some areas (TIGER 2010). Although tertiary roads are not as "porous" as primary and secondary roads (Rossmo 2000), they are still an important aspect of an offender's mental map (Brantingham and Brantingham 1995; Rossmo 2000). Since tertiary roads are typically low volume and occur in lower population density areas, they can be more remote than some primary and secondary roads.

Tertiary road data was accessed through the University of Connecticut from the year 2010 at a map scale of 1:100,000. A buffer was created using the Buffer tool in ArcGIS 10.3 (ESRI 2014) to determine distance from road. As with the primary and secondary roads, the distance of each known site (n=173) to the nearest tertiary road was calculated. The 95th percentile for the dataset was calculated in "R" by bootstrapping the data by the median, as the data are not normally distributed and are skewed (R Core Team 2015). The value (267 ft.) was used as the assigned buffer for this study.

II. Methods

A. Exploratory Spatial Data Analysis

Exploratory spatial data analysis (ESDA) includes a range of spatial statistical tests used to visualize and analyze spatial data. ESDA can identify spatial patterns not immediately visible by simply looking at the data (Congram et al. 2016). The following ESDA techniques used in this thesis are primarily spatial clustering algorithms, used to determine whether there are spatial clusters of data variables in the data set.

a. Moran's I

Moran's I is a measure of the spatial autocorrelation, or the degree to which known sites tend to occur near each other in the study area (Waller and Gotway 2004). The Moran's I was calculated for each variable using the Incremental Spatial Autocorrelation tool in ArcGIS 10.3 (ESRI 2014). The Incremental Spatial Autocorrelation tool was used rather than the Spatial Autocorrelation (Moran's I) tool because the appropriate distance band for each variable was unknown. The benefit of Incremental Spatial Autocorrelation is that the tool calculates the

spatial autocorrelation at different distances and calculates z-scores indicating the intensity of spatial clustering. The distance band chosen for each variable corresponded to the first peak in z-scores, where the clustering was most pronounced (ESRI 2016b).

b. Kernel Density Analysis

Kernel Density Analysis calculates the density of features within a study area. The kernel density was calculated using the Kernel Density tool in ArcMap 10.3 (ESRI 2014). The area was calculated in ft^2 in the Connecticut State Plane projection of the NAD 83 geographic coordinate system to preserve distance. The default search radius was used for this analysis.

c. Ripley's K-Function

Ripley's *K*-function is a multiscalar approach to identifying relative aggregation and segregation of point data (Ripley 1976; 1977; 1981). It compares the expected average point intensity at set intervals within a study area to the observed point intensity. Significance intervals are generated by Monte Carlo simulation of random point distributions. Ripley's *K*-function provides a robust measure of cluster size and cluster distance in the data set (Bevan and Conolly 2006; Ducke 2015; Sayer and Weinhold 2013). The ExpectedK and ObservedK values were calculated using the Multi-Distance Spatial Cluster Analysis (Ripley's *K*-Function) in ArcGIS 10.3 (ESRI 2014) using 100 distance bands, with distance increments of 1,640 ft. (0.3 mi.), and confidence envelopes generated by 99 Monte Carlo simulations.

B. Inductive Model

Two different inductive models will be created. One will use the population density non-site data set (PD model), and one will use the complete spatial randomness data set (CSR model). The same statistical methods will be applied to each data set.

a. Univariate Tests

Statistical assessment of the variables is necessary to see if site values are significantly different from non-site values. Continuous variables are tested for independence using Welch's Two-Sample T-Tests. Nominal presence/absence variables are tested for independence using Chi-Square Tests. Both the Welch's Two-Sample T-Test and the Chi-Square tests are available in the baseline program of "R" (R Core Team 2015). Significant differences between site and non-site values for a variable indicate that the variable has a significant impact on site location.

b. Bivariate Tests

Statistical assessment of the variables is necessary to see if there is a significant correlation between site variables. Tetrachoric correlation tests are used to assess if the nominal presence/absence variables are significantly correlated. The tetrachoric correlation test is available in the "psych" package for "R" (Revelle 2017). Point-biserial correlation tests are used to assess if the nominal and continuous variables are significantly correlated with each other. The point-biserial correlation tests are available in the package "ltm" in "R" (Rizopoulos 2015). Pearson correlation tests are used to assess if the continuous variables are significantly correlated with each other. The Pearson correlation tests are available in the baseline package of "R" (R

Core Team 2015). Significantly correlated variables may result in co-variation that must be considered in the final logistic regression model.

c. Logistic Regression

Logistic regression is a statistical method that predicts the probability of an event occurring (Kleinbaum 1994; Hosmer and Lemeshow 2000). It identifies the relationship between one dependent variable and several independent variables. Logistic regression produces a positive binomial outcome in the form $P = e^x/(1+e^x)$ where P is the probability of site presence and x is $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i$, with β_i as the regression coefficients, x_i as the values for the independent variables, and $_i$ as the number of variables (Rupert et al. 2008). Another benefit of logistic regression is that it does not assume normality of the variables and categorical variables can be included (Vaughn and Crawford 2009).

Predictive modeling utilizes logistic regression to compute probability values for the study area. To determine the final logistic regression equation for this analysis, the independent variable values were linked to the dependent variable using the Extract Values to Points tool in ArcGIS 10.3 (ESRI 2014). The values were exported to an Excel file using the Table to Excel tool (ESRI 2014; Microsoft 2015). The data values were then entered into the statistical program "R" (R Core Team 2015). The model was built by overfitting the logistic regression with every variable and removing non-significant variables with each iteration using stepwise selection. The final model was chosen for its low Akaike information criterion (AIC) value.

The resulting coefficients of the logistic regression correspond to the change in log odds. The results are discussed as odds because logistic regression predicts the probability, or odds, of an event occurring. The exponentiated coefficient is the odds ratio, which is a relative measure

of effect. It measures the odds that an outcome will occur given the independent variables affecting the outcome. The exponentiated coefficient is the odds ratio associated with a one-unit increase in exposure to the predictive variable (Szumilas 2010). The odds ratio can be weighted by one standard deviation to provide perspective on what the change in odds means on a larger scale. This is useful when comparing odds ratios of presence/absence data with the odds ratios of continuous data. Within this study, the odds ratio indicates the odds of a body disposal site being present or absent as compared to the odds of having no site at all, in relation to the given variable (e.g., the odds of finding a body disposal site are 20% higher than the odds of not finding a body disposal site when forest is present). The odds ratio for an interaction in logistic regression cannot be discussed in the same way as the other single variables, as there is not a single odds ratio estimate for it since the odds ratio is representing the odds of the interaction (SAS 2017).

d. Predictive Model

A predictive model is created using the logistic regression equation and the given dataset of dependent and predictive variables to generate a predictive surface. Using the Spatial Analyst extension in ArcGIS 10.3 (ESRI 2014), the Calculate Raster tool utilizes map algebra to compute a probability value for each grid cell in the study area based on the given logistic regression formula. The resulting output is a continuous raster surface of predictive values, ranging from 0 to 1, that represent the probability of body recovery location. For this analysis, high probability was defined as ≥ 0.5, and low probability was defined as < 0.5. The value of 0.5 is the cutoff for low and high probability because values over 0.5 indicate probabilities of body recovery site location greater than random chance.

C. Deductive Model

The deductive model will be created using the weighted map-layer approach. The model will only include variables where site values are significantly different from non-site values. Each variable has its own raster which is categorized into classes. Nominal presence/absence variables have two categories: presence and absence. Continuous variables are classified into groups of values, e.g. one slope category is 0 – 10 degrees, the next is 11 – 20 degrees, and so on. Using the Reclassify tool in ArcGIS 10.3 (ESRI 2014), each variable raster will receive a value assigned based on theory and prior knowledge of the topic. This is the most subjective aspect of deductive modeling. The results will differ if different values are selected for the variables (Brandt et al. 1992). Low values are typically associated with less favorable conditions, while high values are typically associated with more favorable conditions. When calculating the predictive model, each raster layer also receives a weight. The weight is determined by theory, prior knowledge of the topic, and/or is informed by the dataset. The final predictive model is created using the Weighted Sum tool in ArcGIS 10.3 (ESRI 2014). The weighted rasters are summed together to achieve the final product (Brandt et al. 1992; Congram 2010).

The final deductive model equation is as follows:

= (Slope of 0.0 – 1.0) * 7 + (Slope of 1.1 – 3.0) * 5 + (Slope of 3.1 – 10.0) * 3 + (Slope of 10.1 – 90.0) * -5 + (PopDens of 0.0 – 1,000.0) * 8 + (PopDens of 1,000.1 – 5,000.0) * 5 + (PopDens of 5,000.1 – 10,000) * 3 + (PopDens of 10,000.1 – 22,041.2) * -5 + Urban * 9 + Forest * 3 + Ind * 5 + MixOth * 5 + Resid * 7 + WaterWet * 2 + Ag * 5 + CommServ * 5 + OpSp * 5 + PrimSec * 9 + Tertiary * 10

A brief discussion of the variable values and weights is presented below (see also Table 2.3). This is necessary for understanding the model output, and for future repetition of the study.

a. Slope

The slope categories were assigned based on the summary statistics of known site slope values (see Results). The majority of sites are located at lower slopes (~70%) and the category 0 – 1.0 was thus given a value of seven. Fewer sites occur between 1.1 – 3.0, and 3.1 – 10.0, so the assigned values are lower. No sites occur above 10.0 degrees, and the value assigned to 10.1 – 90.0 degrees equaled seven to indicate the importance of site absence in this category. The assigned weight (see Table 2.3) indicates the significance of slope in locating body disposal sites. Given Naismith's rule stating that perpetrators moving a victim's body are subject to time, distance, speed, and effort constraints (Rossmo 2000), slope is likely an important determinant in how far an offender will transport a body. Low slopes are assigned a positive weight and steep slopes are assigned a negative weight. Steeper slopes require more effort to move a victim's body, and offenders are therefore less likely to deposit bodies as slope increases.

b. Population Density

The population categories were assigned based on the summary statistics of known site population density values (see Results). The majority of sites (~70%) occur in low population density areas, and therefore the category 0 – 1,000 was given a value of seven. Fewer sites occur between 1,000.1 – 5,000 and 5,000.1 – 10,0000, so the assigned values are lower. Almost no sites were located at a population density above 10,000.1, so the category 10,000.1 – 22,041.2 was assigned a value of seven to indicate the importance of site absence in this category. The assigned weight indicates the significance of population density in locating body disposal sites (see Table 2.3). Environmental criminology research (Brantingham and Brantingham 1995;

Table 2.3. Deductive model independent variables and categories.

Variable	Category	Value (1-10)	Weight	Weighted value
Slope	0 – 1.0	7	7	49
	1.1 – 3.0	3	5	15
	3.1 – 10.0	1	3	3
	10.1 – 90.0	7	-5	-35
Population Density	0 – 1000.0	7	8	56
	1,000.1 – 5,000.0	2	5	10
	5,000.1 – 10,000.0	1	3	3
	10,000.1 – 22,041.2	7	-5	-35
Urban	Present	9	9	81
Forest	Present	2	3	6
Industrial	Present	1	5	5
Mixed and other urban	Present	1	5	5
Residential	Present	4	7	28
Water and Wetland	Present	1	2	2
Agriculture	Present	1	5	5
Commercial and services	Present	1	5	5
Open Space	Present	1	5	5
Primary and secondary roads	Present	7	9	63
Tertiary roads	Present	8	10	80

Rossmo 2000) indicates that crime occurs in more remote areas; therefore, lower population densities are assigned a positive weight, while the highest population densities are assigned a negative weight.

c. Presence/Absence Variables

The presence/absence variable values were assigned based on the summary statistics of known site presence (see Results). For example, the majority of sites are located in urban areas (~90%), giving urban area a value of 9. Similarly, few sites were located on forest land (~20%), giving it a value of 2. The urban area values were weighted more heavily (see Table 2.3), as urban areas are associated with the remoteness of an area and with population density (Rossmo 2000), suggesting that urban areas play a role in where an offender disposes of a body. The forest land values were weighted less (see Table 2.3), since densely forested land is likely an obstacle for moving a body long distances (Rossmo 2000). Industrial land, mixed and other urban land, commercial and services land, agricultural land, and open space values were weighted with a mid-range value (see Table 2.3). These land categories have a factor of remoteness, but are not as likely to drive disposal patterns as other factors of remoteness (Rossmo 2000). Residential land was weighted more heavily (see Table 2.3), as residential land is associated with the remoteness of an area and with population density. Water and wetlands was given a lower weight (see Table 2.3) despite research that indicates bodies are disposed of in water environments (Rossmo 2000; James 2001; Hakkanen et al. 2007). This is because most water recoveries were excluded from this research, as there is no way to tell from the case files where the body was deposited before it floated to its recovery location. Primary, secondary, and

tertiary roads were given high weights (see Table 2.3), since previous research has indicated the importance of distance to road in locating body disposal (Orengo 2006; Congram 2010).

D. Model Validation

The most effective means of testing a predictive model for accuracy and predictive ability is through field validation (Kohler and Parker 1986; Ebert 2004; Congram 2010). While ideal, this type of validation is not plausible for this project at this time. Field validation requires consistent updates from law enforcement and/or OCME on the location of body recovery site locations in Connecticut. Such validation can occur after a period of time passes. Data collection for this project ended in 2015. Data from 2016 and 2017 can be used to test the predictive model, if made available to the researcher. At the present moment, quantitative assessment of the model is possible by counting the number of known sites correctly predicted by the model and dividing this number by the total number of sites to get the positive predictability of the model. Dividing the correctly predicted non-sites by the total number of non-sites provides the negative predictability of the model. Additionally, eight body disposal site locations are known from neighboring Connecticut counties (Litchfield, Windham, Middlesex, Tolland). These eight sites will be tested to see if the predictive model can predict sites outside of the study area.

a. Kvamme's Gain Statistic

Another method of model validation is Kvamme's Gain Statistic (Kvamme 1990). The Gain statistic is calculated by testing the site prediction values on a set of known locations not included in the original model. In this case, the model validation set is the 35 known sites

withheld from the training set. The formula for calculating Kvamme's Gain statistic is as follows:

Gain = 1 – (% of area with high probability / % of sites with high probability).

The Gain statistic results range from – 1 to + 1 and provide a measure of the predictive ability of the probability surface. Positive values close to 1 indicate a model with high predictive value, while negative values close to -1 indicate a model with low predictive value. A Gain statistic of zero indicates the model is no better at making predictions than chance (Kvamme 1990; Vaughn and Crawford 2009).

b. Mann-Whitney U Test

The deductive model result values must be tested for independence using the Mann-Whitney U test available in the baseline package of "R" (R Core Team 2015). The deductive model results are only valuable if the predicted site values are significantly different from the predicted non-site values. When the site and non-site values are significantly different, it indicates that the model can differentiate between site and non-site locations.

IV. Summary

The methods discussed in this chapter are used to established that body disposal sites in Connecticut are patterned, nonrandom, and predictable. Exploratory spatial analyses, including Moran's I, Kernel Density Analysis, and Ripley's *K*-function identify if there are clusters of sites within the data set, or if sites are randomly distributed across the landscape. Two different data sets are used to create the inductive predictive model: the complete spatial randomness data set

and the population density data set. The complete spatial randomness data set included random non-sites simulated to mimic complete spatial randomness, and assumes that sites are equally as likely to occur in one location as they are in another within the study area. The population density data set included random non-sites simulated to mimic the population density of the study area, and assumes that sites occur where people occur on the landscape. The complete spatial randomness data set is also used to create the deductive model. Both inductive and deductive models are created to test if the "data-driven" models or the "theory-driven" models are best suited for predicting body disposal sites in Connecticut. The following chapter will discuss the results of these methods in relation to the objectives set forth in the Introduction.

CHAPTER FOUR
RESULTS

This chapter presents the results from the different analyses performed as outlined in Chapter 3. The spatial statistical test results inform on the patterned, inhomogeneous, and clustered nature of the known sites. Three predictive models are presented: one inductive model using data of known body disposal locations and random non-sites simulated to mimic complete spatial randomness; one inductive model using data of known body disposal locations and random non-sites simulated to mimic the population density of the study area; and one deductive model, created by referencing the known sites and utilizing theoretical assumptions of offender behavior from forensic anthropology and criminology. For each of the three predictive models, a table summarizing the results is presented, followed by the predictive model and a table presenting a classification matrix of correctly and incorrectly predicted sites and non-sites. The model validation results are presented, including Kvamme's Gain statistic, positive predictability of the model, negative predictability of the model, and success of the model in predicting the eight sites outside of the study area. An interpretation and detailed discussion of the significance of the results are in the following chapter.

I. Exploratory Spatial Data Analysis

A. Moran's I

Moran's I results indicate positive spatial autocorrelation is present in all variables (Table 3.1). Positive autocorrelation results show that the sites are non-homogeneously distributed throughout the study area. For example, the positive Moran's I result for "Tertiary roads" indicates more sites were present near tertiary roads than other areas, resulting in an uneven

Table 3.1. Results for Univariate Moran's I Test of Spatial Autocorrelation.

Variable	Moran's I Value*	Peak Distance (ft.)	p-value**
Population density	0.346	6560	**<0.0001**
Slope	0.185	8200	**0.003**
Urban	0.090	9840	**0.005**
Open space	0.424	1640	**0.0004**
Primary and secondary roads	0.256	2560	**0.010**
Tertiary roads	0.310	2560	**0.002**
Water and wetlands	0.099	6560	**0.024**
Residential	0.201	4920	**0.001**
Mixed and other urban	0.298	3280	**0.0001**
Agricultural	0.074	25000	**0.0004**
Industrial	0.424	1640	**0.0004**
Forest	0.604	3280	**<0.0001**
Commercial and services	0.580	4920	**<0.0001**

*Values range from -1 to +1, with -1 indicating perfect dispersion, 0 indicating a random spatial pattern, and +1 showing perfect clustering/spatial autocorrelation.
**Significant at 0.05.

spatial distribution of sites near tertiary roads across the study area. The positive Moran's I result indicates the data set is more spatially clustered than would be expected if the underlying spatial processes were random. The presence of autocorrelation is a sign that spatial statistical methods should be used over conventional statistical methods when possible.

B. Kernel Density Analysis

Visual assessment of the kernel density analysis output indicates the sites are not homogeneously distributed across the study area. The darker areas in the kernel density plot show that more sites are occurring near other sites in these areas than in other areas. Higher density areas are apparent within the study area, with three of the areas noticeably denser than the other five (Figure 3.1). This suggests that body deposition site locations occur non-randomly across the counties of Hartford, New Haven, and New London.

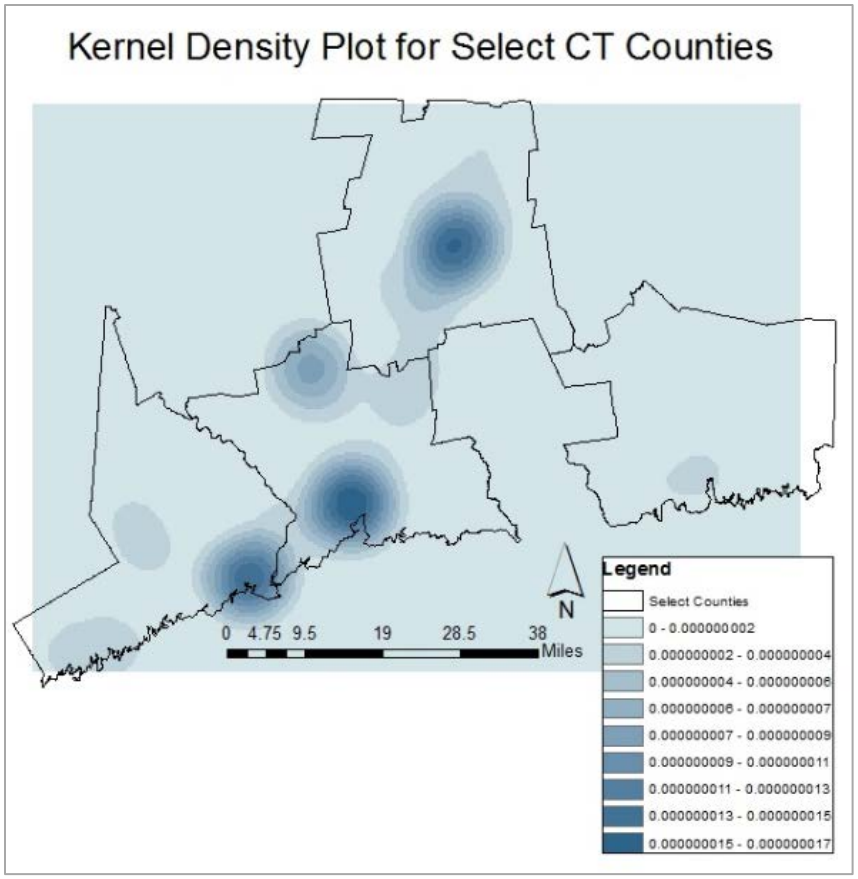


Figure 3.1. Kernel density analysis results for known sites.

C. Ripley's *K*-function

The results of Ripley's *K*-function analysis indicate significant site clustering from 0 – 70,000 ft., and significant site dispersion after 70,000 ft. (13.3 mi.). This indicates that sites are more likely to occur near other sites from 0 – 70,000 feet, and less likely to occur near other sites after 70,000 feet. Figure 3.2 displays the ObservedK and ExpectedK values with a 95% confidence interval. The ObservedK value is outside the bounds of the confidence interval, indicating it is significant. The ObservedK value rises sharply until about 20,000 ft. (3.8 mi.), where it continues to rise more slowly until 70,000 ft. when it drops below the ExpectedK value and becomes dispersed. While there is a lack of peaks in the graph to indicate areas of

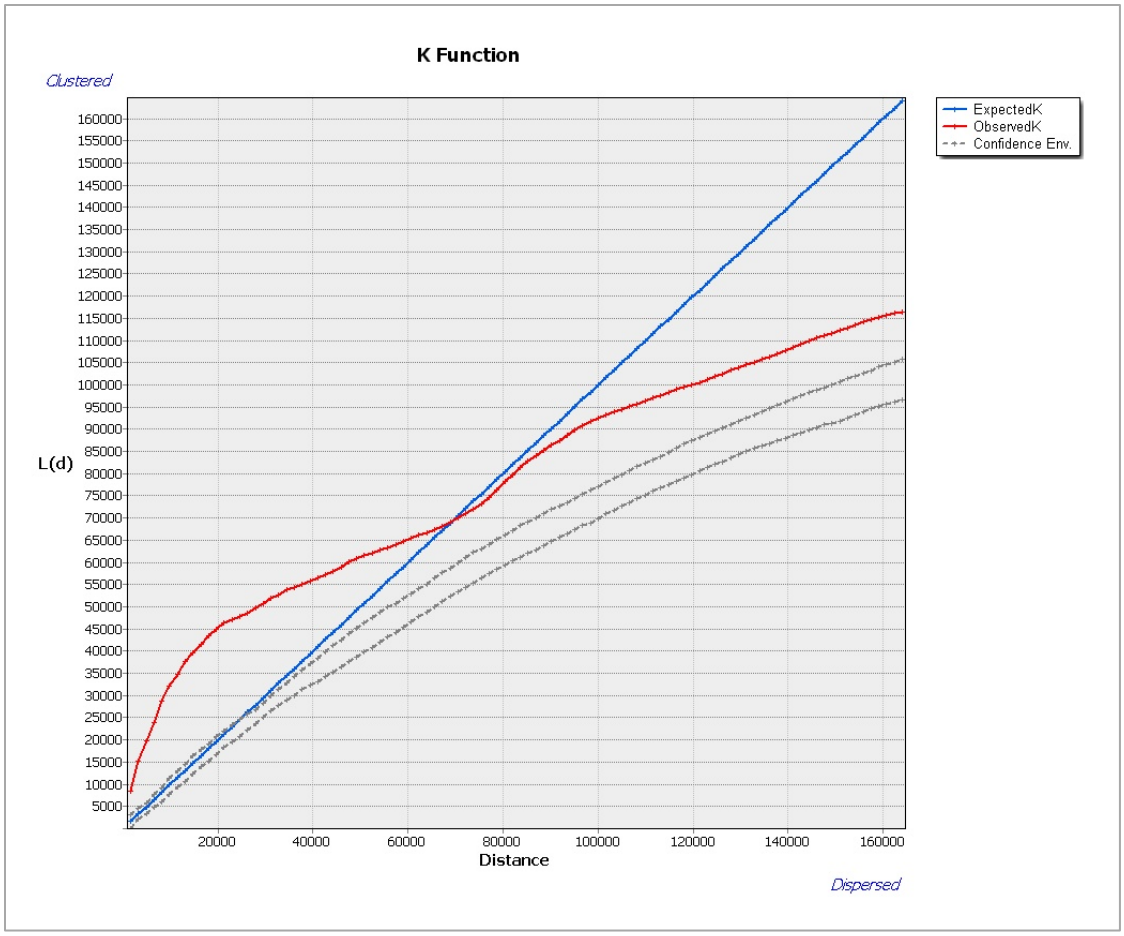


Figure 3.2. Ripley's *K*-function results for known site clustering and dispersion.

clustering important to investigate for body disposal sites, the distance bands that showed the highest amount of clustering occurred at 16,400 ft. (diff. = 25,125), 18,040 ft. (diff. = 25,324), and 19,680 ft. (diff. = 25,184). These distance bands indicate that within the study area, sites are more likely to occur near other sites (i.e., to cluster) within a range of 16,400 ft. (3.1 mi.) – 19,680 ft. (3.7 mi.).

II. Univariate and Bivariate Tests of Variables

A. Summary Statistics

Summary statistics for known body recovery site location (Table 3.2) indicate that the majority of sites are found in urban areas. Out of the seven land use categories, sites occurred most frequently on residential land, followed by forest land. Mixed and other urban land, as well as commercial and services land, had more sites than industrial, agricultural, and water and wetlands. Very few sites occurred in designated open spaces, such as parks and cemeteries. The majority of sites were within 1865 feet of primary and secondary roads, and within 267 feet of tertiary roads. Known sites occur most frequently on relatively flat ground and in lower population density areas (Table 3.3).

Most body disposal sites occur closer to tertiary roads (91%) then to primary and secondary roads (9%). The greatest distance from body disposal site location to primary and secondary roads was 1.63 miles, while the median distance was .19 miles (See Table 3.4). Distance from body disposal site to tertiary roads was even shorter, with a maximum of .19 miles (993.0 feet) and a median of .02 miles (89.3 feet) (See Table 3.4). The maximum distance from the victim's home to body disposal site for Connecticut residents was 48.23 miles, with a median

Table 3.2. Descriptive statistics for binary independent variables.

Variable	Present	Absent	% Present
Urban	157	15	92%
Forest	33	139	19%
Industrial	10	162	6%
Mixed and other urban	23	149	13%
Residential	70	102	41%
Water and wetland	7	165	4%
Agriculture	8	164	5%
Commercial and services	20	152	12%
Open space	11	161	6%
Primary and secondary roads	127	45	74%
Tertiary roads	135	37	78%

Table 3.3. Descriptive statistics for continuous independent variables.

Variable	Minimum	Maximum	Mean	Median
Slope	0.00 degrees	9.04 degrees	1.02 degrees	0.62 degrees
Population Density	0.00 ind/km^2	16296.20 ind/km^2	1665.19 ind/km^2	441.67 in/km^2

Table 3.4. Descriptive statistics for body site location (continuous variables).

Variable	Minimum	Maximum	Mean	Median
Distance from Victim's Home to Body Disposal Site	0.00 km	4236.00 km	46.03 km	3.17 km
	0.00 mi	2632.13 mi	28.60 mi	1.97 mi
Distance from Victim's Home to Body Disposal Site*	0.00 km	428.00 km	12.78 km	3.14 km
	0.00 mi	265.95 mi	7.94 mi	1.95 mi
Distance from Victim's Home to Body Disposal Site**	0.00 km	77.62 km	8.04 km	2.99 km
	0.00 mi	48.23 mi	5.00 mi	1.86 mi
Distance from Primary and Secondary Roads to Body Disposal Site	4.57 ft	8629.48 ft	1358.41 ft	977.01 ft
	0.001 km	2.63 km	0.41 km	.30 km
	.22 mi	1.63 mi	.25 mi	.19 mi
Distance from Tertiary Roads to Body Disposal Site	.35 ft	993.03 ft	169.40 ft	89.43 ft
	.0001 km	0.30 km	0.05 km	0.03 km
	6.63 e-05 mi	0.19 mi	0.03 mi	0.02 mi

*Without California case. **Only CT residents.

of 1.86 miles (Table 3.4). When the three non-Connecticut residents are included, the maximum distance was 2632.13 miles, with a median of 1.97 miles (Table 3.4).

B. Univariate Tests

a. Complete spatial randomness model

Chi-square tests were used to assess if the nominal presence/absence variables were significantly different between sites and non-sites. Significant variables include urban areas, forest land, mixed and other urban land, residential land, agricultural land, commercial and services land, distance to primary and secondary roads, and distance to tertiary roads (Table 3.5). Welch's two sample t-tests were used to assess if the continuous variables were significantly different between sites and non-sites. Both population density and slope were significant (Table 3.5).

Table 3.5. Univariate tests of significance for independent variables between site and non-site (Complete Spatial Randomness).

Variable	Statistical test	p-value*
Population Density	Welch's Two Sample T-test	**4.671 e-07**
Slope	Welch's Two Sample T-test	**0.0007832**
Urban	Chi Square	**4.688 e-12**
Forest	Chi Square	**4.742 e-08**
Industrial	Chi Square	0.1759
Mixed and Other Urban	Chi Square	**0.0005096**
Residential	Chi Square	**0.002715**
Water and Wetland	Chi Square	1
Agriculture	Chi Square	**0.03256**
Commercial and services	Chi Square	**0.01169**
Open Space	Chi Square	.975
Primary and Secondary Roads	Chi Square	**2.637 e-06**
Tertiary Roads	Chi Square	**9.830 e-11**

*Significant $p < 0.05$

b. Population density model

Chi-square tests were used to assess if the nominal presence/absence variables were significantly different between sites and non-sites. No variables were statistically significant (Table 3.6). Welch's two sample t-tests were used to assess if the continuous variables were significantly different between sites and non-sites. Neither population density nor slope were significant (Table 3.6).

Table 3.6. Univariate tests of significance for independent variables between site and non-site (Population Density).

Variable	Statistical test	p-value*
Population Density	Welch's Two Sample T-test	0.2685
Slope	Welch's Two Sample T-test	0.1463
Urban	Chi Square	1.0000
Forest	Chi Square	0.9110
Industrial	Chi Square	0.2254
Mixed and Other Urban	Chi Square	0.1300
Residential	Chi Square	0.4274
Water and Wetland	Chi Square	0.5022
Agriculture	Chi Square	0.2526
Commercial and services	Chi Square	0.2937
Open Space	Chi Square	0.5607
Primary and Secondary Roads	Chi Square	0.0544
Tertiary Roads	Chi Square	0.5977

*Significant $p < 0.05$

C. Bivariate Tests

Tetrachoric correlation tests were used to assess if the nominal presence/absence variables were significantly correlated with each other. Twelve variables were significantly correlated at $p < 0.0001$ (see Tables 3.7 – 3.10). Point-biserial correlation tests were used to assess if the nominal and continuous variables were significantly correlated with each other (Tables 3.7 – 3.12). Only two sets of variables were correlated: slope with forest, and population density with commercial and services (Tables 3.11 – 3.12). Pearson correlation tests were used

Table 3.7. Correlation of variables to urban areas.

Variable	Statistical Test	Correlation	p-value*
Slope	Point-Biserial	- 0.004	0.481
Population density	Point-Biserial	- 0.174	0.021
Forest	Tetrachoric	- 0.660	<**0.0001**
Industrial	Tetrachoric	- 0.064	0.228
Mixed and other urban	Tetrachoric	0.120	0.080
Residential	Tetrachoric	0.590	<**0.0001**
Water and wetland	Tetrachoric	- 0.380	<**0.0001**
Agriculture	Tetrachoric	- 0.380	<**0.0001**
Commercial and services	Tetrachoric	0.160	0.030
Open Space	Tetrachoric	- 0.029	0.368
Primary and secondary roads	Tetrachoric	- 0.041	0.317
Tertiary roads	Tetrachoric	0.470	<**0.0001**

*Significant $p < 0.001$

Table 3.8. Correlation of variables to open space.

Variable	Statistical Test	Correlation	p-value*
Slope	Point-Biserial	- 0.100	0.122
Population density	Point-Biserial	0.068	0.214
Urban	Tetrachoric	- 0.029	0.368
Forest	Tetrachoric	0.054	0.265
Industrial	Tetrachoric	0.190	0.128
Mixed and other urban	Tetrachoric	0.600	<**0.0001**
Residential	Tetrachoric	- 0.230	0.003
Water and wetland	Tetrachoric	0.390	<**0.0001**
Agriculture	Tetrachoric	0.190	0.013
Commercial and services	Tetrachoric	- 0.029	0.368
Primary and secondary roads	Tetrachoric	- 0.170	0.023
Tertiary roads	Tetrachoric	- 0.041	0.317

*Significant $p < 0.001$

Table 3.9. Correlation of variables to primary and secondary roads.

Variable	Statistical Test	Correlation	p-value*
Slope	Point-Biserial	0.189	0.013
Population density	Point-Biserial	- 0.012	0.444
Urban	Tetrachoric	- 0.041	0.317
Forest	Tetrachoric	- 0.100	0.122
Industrial	Tetrachoric	0.190	0.128
Mixed and other urban	Tetrachoric	0.130	0.064
Residential	Tetrachoric	- 0.130	0.064
Water and wetland	Tetrachoric	- 0.098	0.126
Agriculture	Tetrachoric	- 0.048	.288
Commercial and services	Tetrachoric	0.440	**<0.0001**
Open space	Tetrachoric	- 0.170	0.023
Tertiary roads	Tetrachoric	0.084	0.164

*Significant $p < 0.001$

Table 3.10. Correlation of variables to tertiary roads.

Variable	Statistical Test	Correlation	p-value*
Slope	Point-Biserial	0.110	0.100
Population density	Point-Biserial	- 0.212	0.006
Urban	Tetrachoric	0.470	**<0.0001**
Forest	Tetrachoric	- 0.450	**<0.0001**
Industrial	Tetrachoric	0.150	0.040
Mixed and other urban	Tetrachoric	0.220	0.005
Residential	Tetrachoric	0.380	**<0.0001**
Water and wetland	Tetrachoric	- 0.580	**<0.0001**
Agriculture	Tetrachoric	- 0.280	0.0004
Commercial and services	Tetrachoric	0.220	0.005
Open space	Tetrachoric	- 0.041	0.317
Primary and secondary roads	Tetrachoric	0.084	0.164

*Significant $p < 0.001$

Table 3.11. Correlation of variables to slope.

Variable	Statistical test	Correlation	p-value*
Population Density	Pearson	- 0.172	0.022
Urban	Point-Biserial	- 0.004	0.481
Forest	Point-Biserial	- 0.451	< **0.0001**
Industrial	Point-Biserial	0.145	0.045
Mixed and Other	Point-Biserial	0.032	0.355
Residential	Point-Biserial	0.113	0.093
Water and Wetland	Point-Biserial	0.060	0.242
Agriculture	Point-Biserial	0.100	0.122
Commercial and services	Point-Biserial	0.163	0.028
Open Space	Point-Biserial	- 0.100	0.122
Primary and Secondary Roads	Point-Biserial	0.189	0.013
Tertiary Roads	Point-Biserial	0.110	0.100

*Significant $p < 0.001$

Table 3.12. Correlation of variables to population density.

Variable	Statistical Test	Correlation	p-value*
Slope	Pearson	- 0.172	0.022
Urban	Point-Biserial	- 0.174	0.021
Forest	Point-Biserial	0.250	0.002
Industrial	Point-Biserial	0.122	0.077
Mixed and Other Urban	Point-Biserial	0.048	0.288
Residential	Point-Biserial	- 0.166	0.026
Water and Wetland	Point-Biserial	0.085	0.161
Agriculture	Point-Biserial	0.104	0.112
Commercial and services	Point-Biserial	- 0.325	< **0.0001**
Open Space	Point-Biserial	0.068	0.214
Primary and Secondary Roads	Point-Biserial	- 0.012	0.444
Tertiary Roads	Point-Biserial	- 0.212	0.006

*Significant $p < 0.001$

to assess if the continuous variables were significantly correlated with each other. Results indicated that they were not (Table 3.11 – 3.12).

III. Inductive Model

A. Complete Spatial Randomness (CSR) Model

a. Logistic regression

The final predictive model included mixed and other urban land, forest land, primary and secondary roads, tertiary roads, urban areas, slope, population density, and an interaction between forest land and slope (Table 3.13). The final predictive model is estimated as:

$$1 / (1 + (\exp (- (-1.608 + 2.358*MixOth - 0.868*Forest + 0.759*PrimSec + 0.795*Tertiary + 1.310*Urban - 0.573*Slope + 0.0003*PopDens + 0.613*Forest*Slope))))$$

The coefficient for mixed and other urban land use is positive, suggesting that body disposal sites occur more frequently on mixed and other urban property. Given the odds ratio, the odds of having a body disposal site on mixed and urban land while holding all other variables fixed are extremely high at 957% greater than the odds of having no body disposal site. The coefficients for primary and secondary roads, as well as for tertiary roads, are positive. The odds ratios suggest that, holding all other variables fixed, the odds of finding a body disposal site within 1865 feet of a primary or secondary road are 114% higher than not finding a site, while the odds of finding a body disposal site within 267 feet of a tertiary road are 121% higher than not finding a site. The urban coefficient is also positive, indicating body disposal sites are more likely to occur in urban areas. The odds ratio suggests the odds of finding a site on urban land are 271% greater than not finding a site. The coefficient of population density is positive, indicating sites

Table 3.13. Logistic regression results (Complete spatial randomness model).

Variable	Coefficient	Odds Ratio	p-value*
Intercept	- 1.608	0.200	**0.0009**
Mixed and other urban	2.358	10.570	**0.0305**
Forest	- 0.868	0.420	**0.0451**
Primary and secondary roads	0.759	2.136	**0.0165**
Tertiary roads	0.795	2.214	**0.0137**
Urban	1.310	3.706	**0.0011**
Slope	- 0.573	0.564	**0.0028**
Population density	0.0003	1.000	**0.0322**
Forest*Slope	0.613	1.846	**0.0047**

*Significant at < 0.05.

are more likely to be found in more densely populated areas. Holding all other variables fixed, the odds of finding a body disposal site are 0.03% higher than the odds of not finding a site with every unit increase (individual/km^2) in population density. Weighting the odds ratio by one standard deviation indicates that for every 2203.2 increase in population density, the odds of finding a site are 94% higher than not finding a site.

The coefficient for forest land is negative, indicating that sites occur less frequently on forest land. The coefficient of slope is also negative, indicating sites are more likely to be found at low slope degrees. There is a significant interaction between forest land and slope, which has a positive coefficient. Table 3.14 shows the effect of slope on the odds ratio of forest land within the range of slope values most common in the site data set. As slope increases and forest land is present, the odds of finding a body disposal site rise.

Table 3.14. Forest*Slope interaction odds ratios (Complete spatial randomness model).

Slope (degree)	Odds Ratio: Slope*Forest(Absent)	Odds Ratio: Slope*Forest(Present)
0	1.00	0.42
1	0.56	0.44
4	0.10	0.49
8	0.01	0.58

b. Predictive Model

The final CSR model has a predictive range of 1.71e-006 – 0.99. See Figure 3.3 for the whole range of predicted values. See Figure 3.4 for the model with a sectioning point of 0.5, with < 0.5 classified as low probability and ≥ 0.5 classified as high probability.

The predictive model reduces the study area of 1,713,339.1 acres to a high probability area of 574,630.2 acres (34% of original study area). When applying the model to the entire state of Connecticut, the predictive model reduces the area from 3,146,187.9 acres to 742,690.9 acres (24% of state area).

c. Model validation

Thirty four percent of the study area and 83% of the known testing sites fall within high probability areas, making the Gain statistic for this model +0.59. This positive value indicates that the model has positive predictability and is 59% more likely to predict body disposal site location than random chance.

A classification matrix of correctly predicted sites and incorrectly predicted sites indicates a positive predictive value of 78% (Table 3.15). Similarly, a classification matrix of correctly predicted non-sites and incorrectly predicted non-sites indicates a negative predictive value of 73% (Table 3.15).

Eight known site locations from Litchfield, Middlesex, Windham, and Tolland counties in Connecticut were used to test the predictive model outside of the original study area. Three of the eight sites were correctly predicted (38%). A very low percentage of the four counties was classified as high probability (12%). The Gain statistic for the CSR model applied to Litchfield, Middlesex, Windham, and Tolland was +0.68. The positive value indicates that the model has

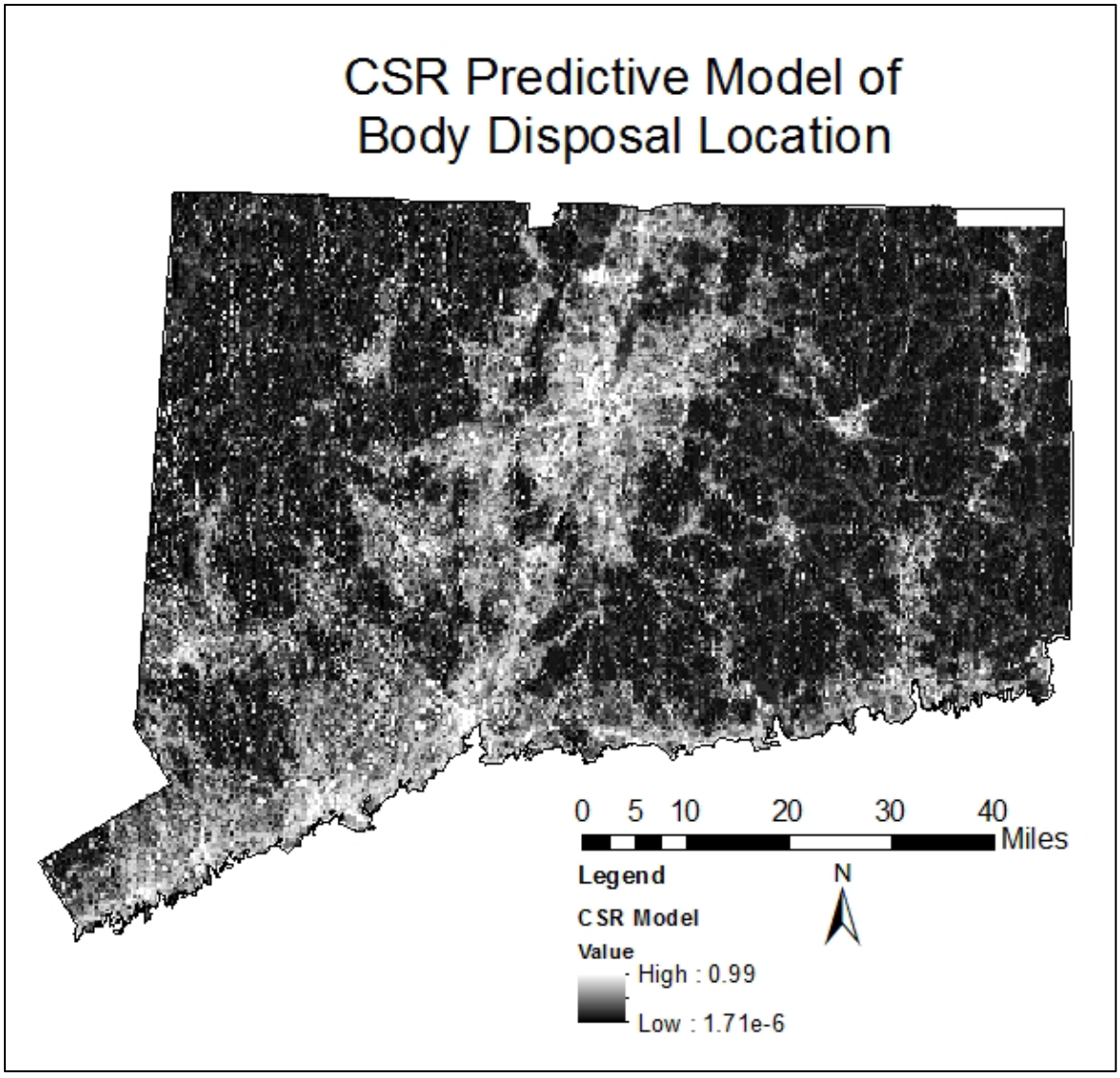


Figure 3.3. Complete spatial randomness predictive model, full range.

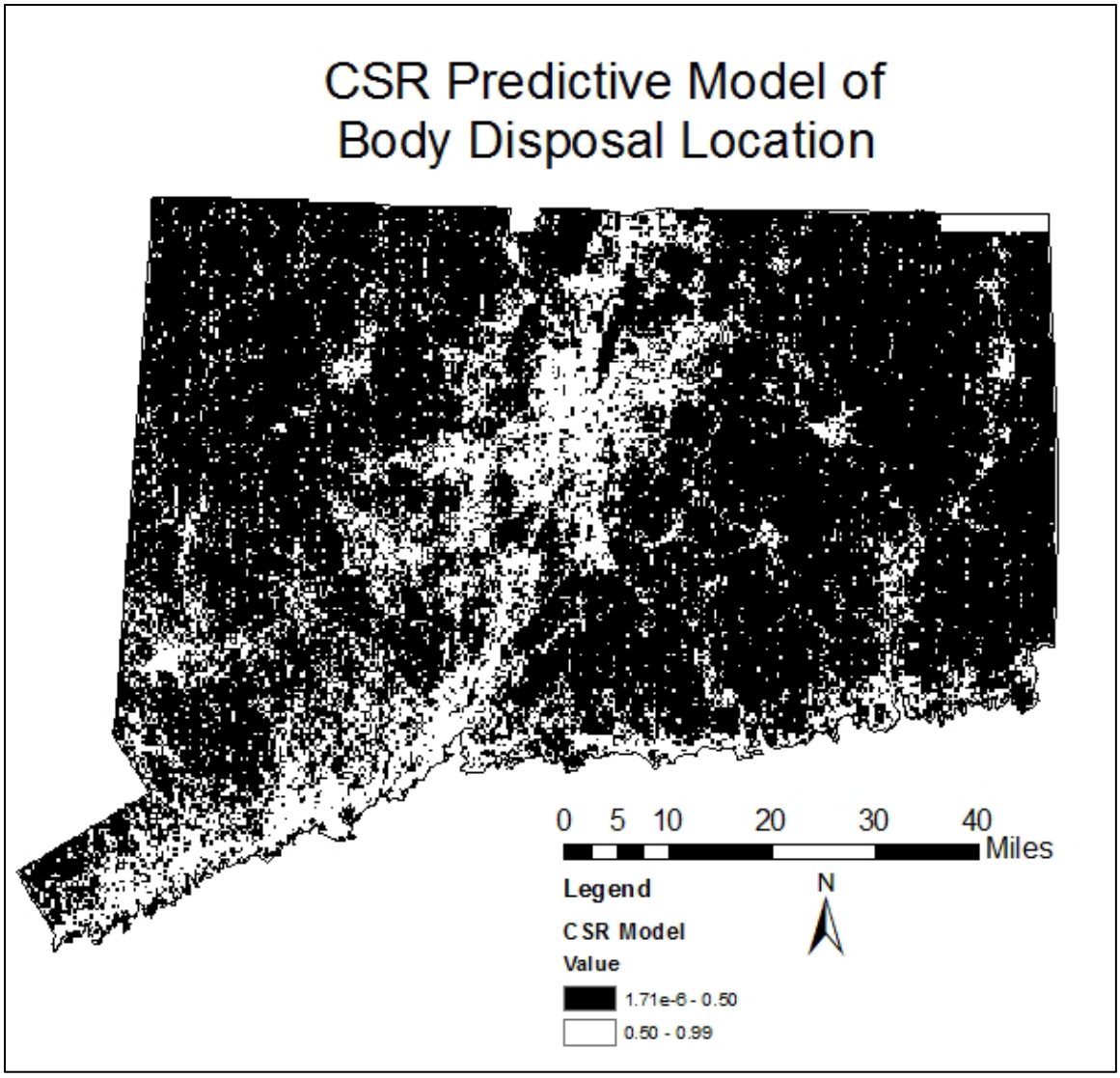


Figure 3.4. Complete spatial randomness predictive model, classified.

positive predictability and is 68% more likely to predict body disposal site location than random chance.

Table 3.15. Classification matrix of correctly and incorrectly predicted sites and non-sites (Complete spatial randomness model).

	Correctly Predicted	Incorrectly Predicted	Total*	Percentage Correct (%)
Sites	134	37	171	78
Non-Sites	124	46	170	73

*Some sites and non-sites did not fall within range of the model and were excluded.

B. Population Density (PD) Model

a. Logistic regression

The final predictive model included water and wetlands, residential, agriculture, forest, distance to primary and secondary roads, slope, population density, and an interaction between forest and slope (Table 3.16). The final predictive equation is estimated as:

1 / (1 + (exp (- (.945 – 1.900*WaterWet – 0.748*Resid – 1.461*Ag – 1.892*Forest + 0.531*PrimSec – 0.469*Slope – 0.000096*PopDens + 0.812*Forest*Slope))))

The coefficient for water and wetlands is negative, suggesting that body disposal sites occur less frequently in water and wetlands. The odds ratio for water and wetlands indicates that the odds of body disposal sites being present, holding all other variables fixed, are 85% less than the odds of no disposal site. Residential land also has a negative coefficient, suggesting that disposal sites are less likely on residential land. The odds ratio for residential land indicates the odds of having a body disposal site are 53% lower than the odds of not having a site if all other variables are held fixed. The coefficient for agriculture is negative as well. The odds ratio indicates the odds

Table 3.16. Logistic regression results (Population density model).

Variable	Coefficient	Odds Ratio	p-value*
Intercept	0.945	2.573	**0.0328**
Water and wetlands	- 1.900	0.150	**0.0156**
Residential	- 0.748	0.473	**0.0346**
Agriculture	- 1.461	0.232	**0.0120**
Forest	- 1.892	0.151	**0.0006**
Primary and secondary roads	0.531	1.701	0.0647
Slope	- 0.469	0.626	**0.0011**
Population density	- 0.0001	0.999	0.0546
Forest*Slope	0.812	2.252	**0.0010**

*Significant at < 0.05.

of having body disposal sites present, holding all other variables fixed, are 77% lower than the odds of having no site. Distance to primary and secondary roads had a positive coefficient, indicating sites are more likely to occur within 1865 feet of primary and secondary roads. The odds ratio for primary and secondary roads indicates that, holding all other variables fixed, the odds of having a site within 1865 feet of a primary or secondary road are 70% higher than the odds of not finding a site. The coefficient for population density is negative, indicating sites are less likely to occur in higher density areas. Holding all other variables fixed, the odds of finding a body disposal site are 1% lower than the odds of not finding a site with every unit increase (individual/km^2). Weighting the odds ratio by one standard deviation indicates that for every 2870.1 increase in population density, the odds of finding a site are 25% lower than the odds of not finding a site.

The coefficient for forest land is negative, indicating that sites occur less frequently on forest land. The coefficient of slope is also negative, indicating sites are more likely to be found at low slope degrees. There is a significant interaction between forest land and slope, which has a positive coefficient. Table 3.17 shows the effect of slope on the odds ratio of forest land within

Table 3.17. Forest*Slope interaction odds ratio (Population density model).

Slope (degree)	Odds Ratio: Slope*Forest(Absent)	Odds Ratio: Slope*Forest(Present)
0	1	0.15
1	0.63	0.21
4	0.15	0.60
8	0.02	2.34

the range of slope values most common in the site data set. As slope increases and forest land is present, the odds of finding a body disposal site rise.

b. Predictive Model

The final PD model has a predictive range of 7.23e-005 – 1. See Figure 3.5 for the whole range of predicted values. See Figure 3.6 for the model with a sectioning point of 0.5, with < 0.5 classified as low probability and ≥ 0.5 classified as high probability.

The predictive model reduces the study area of 1,713,518.7 acres to a high probability area of 622,852.8 acres (36% of original study area). When applying the model to the entire state of Connecticut, the predictive model reduces the area from 3,147,714.5 acres to 1,099,376.5 acres (35% of state area).

c. Model Validation

Thirty six percent of the study area and 62% of the known testing sites fall within high probability areas, making the Gain statistic for this model +0.42. This positive value indicates that the model has positive predictability and is 42% more likely to predict body disposal location than random chance.

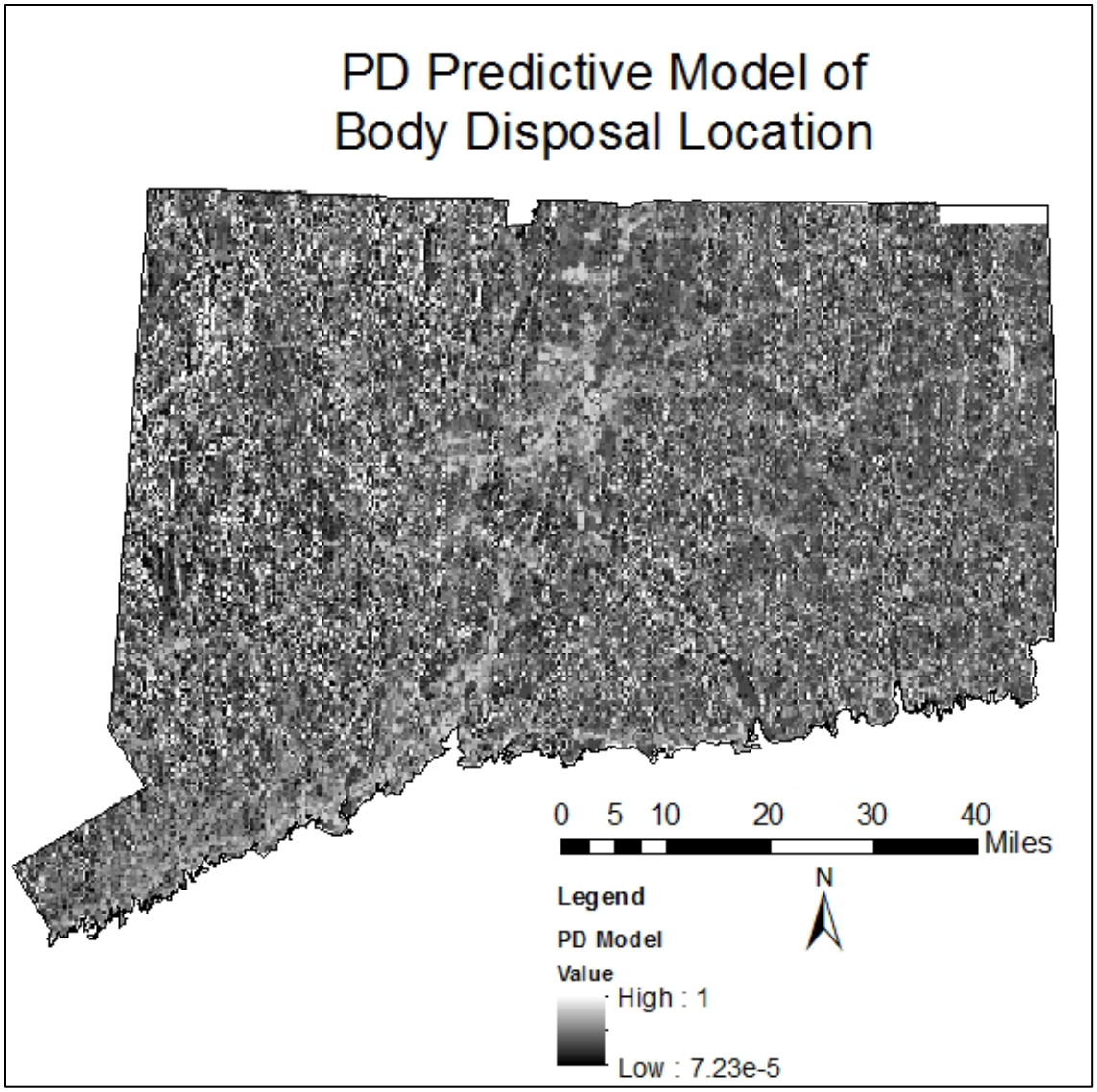


Figure 3.5. Population density predictive model, full range.

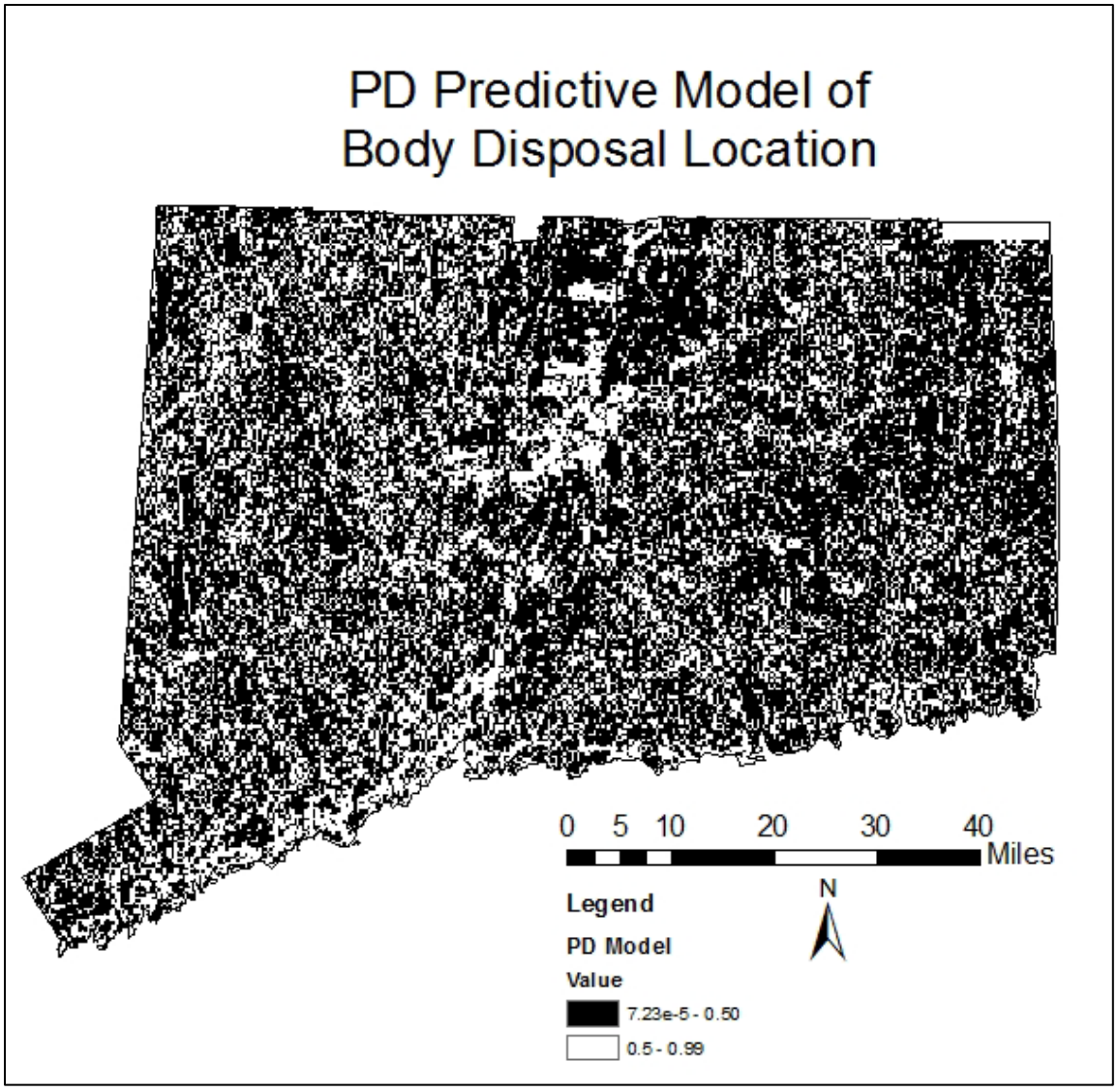


Figure 3.6. Population density predictive model, classified.

A classification matrix of correctly predicted sites and incorrectly predicted sites indicates a positive predictive value of 58% (Table 3.18). A similar classification matrix of correctly predicted non-sites and incorrectly predicted non-sites indicates a negative predictive value of 60% (Table 3.18).

Table 3.18. Classification matrix of correctly and incorrectly predicted sites and non-sites (Population density model).

	Correctly Predicted	Incorrectly Predicted	Total*	Percentage Correct (%)
Sites	99	72	171	58
Non-Sites	104	69	173	60

*Some sites and non-sites did not fall within range of the model and were excluded.

Eight known site locations from Litchfield, Middlesex, Windham, and Tolland counties in Connecticut were used to test the predictive model outside of the original study area. Three of the eight sites were correctly predicted (38%). Thirty three percent of the four counties were classified as high probability. The Gain statistic for the PD model applied to Litchfield, Middlesex, Windham, and Tolland was +0.13. The positive value indicates that the model has positive predictability and is 13% more likely to predict body disposal site location than random chance.

IV. Deductive Model

A. Results

The deductive model was created using the complete spatial randomness simulated points. A deductive model could not be created using the population density simulated points, as none of the variables were significantly different between site and non-site.

The final deductive model was split into a high probability area and a low probability area. The sectioning point was determined by looking at the table of cumulative percentage of area and cumulative percentage of sites (Table 3.19). Assessment of Table 3.19 indicates that the model predicts 50% of the data between values 23 and 277, and 50% of the data between values 278 and 362. As the optimal point in the inductive models was 0.5, or 50%, the same percentage was chosen for the deductive model. The full range of high and low values is depicted in Figure 3.7.

Fifty percent of the data is correctly predicted in 22% of the original study area with a sectioning point of 277. This narrows the study area from 1,712,710.5 acres to 353,363.0 acres. When 74% of sites are correctly predicted in 12% of the original study area, the study area is further narrowed to 183,012.4 acres. When applying the deductive model to the entire state of Connecticut, a sectioning point of 277 narrows the study area from 3,146,053.2 acres to 426,550.0 acres (13% of state area). When 74% of sites are correctly predicted in 7% of the study area with a sectioning point of 317, the search area is further narrowed to 222,299.9 acres.

Table 3.19. Deductive model map values, coverage, and proportion of sites (Deductive model).

Value	Cell Count	Cum. % of Area	All Sites	% of All Sites	Cum. % of All Sites
23 – 89	4893	13	1	1	1
90 – 124	3192	21	3	2	3
125 – 145	2825	28	0	0	3
146 – 169	3015	36	3	2	5
170 – 194	4689	48	11	6	11
195 – 217	1669	52	11	6	17
218 – 243	3906	62	16	9	26
244 – 277	6086	78	42	24	50
278 – 317	3794	88	42	24	74
318 – 362	4076	99	43	25	99

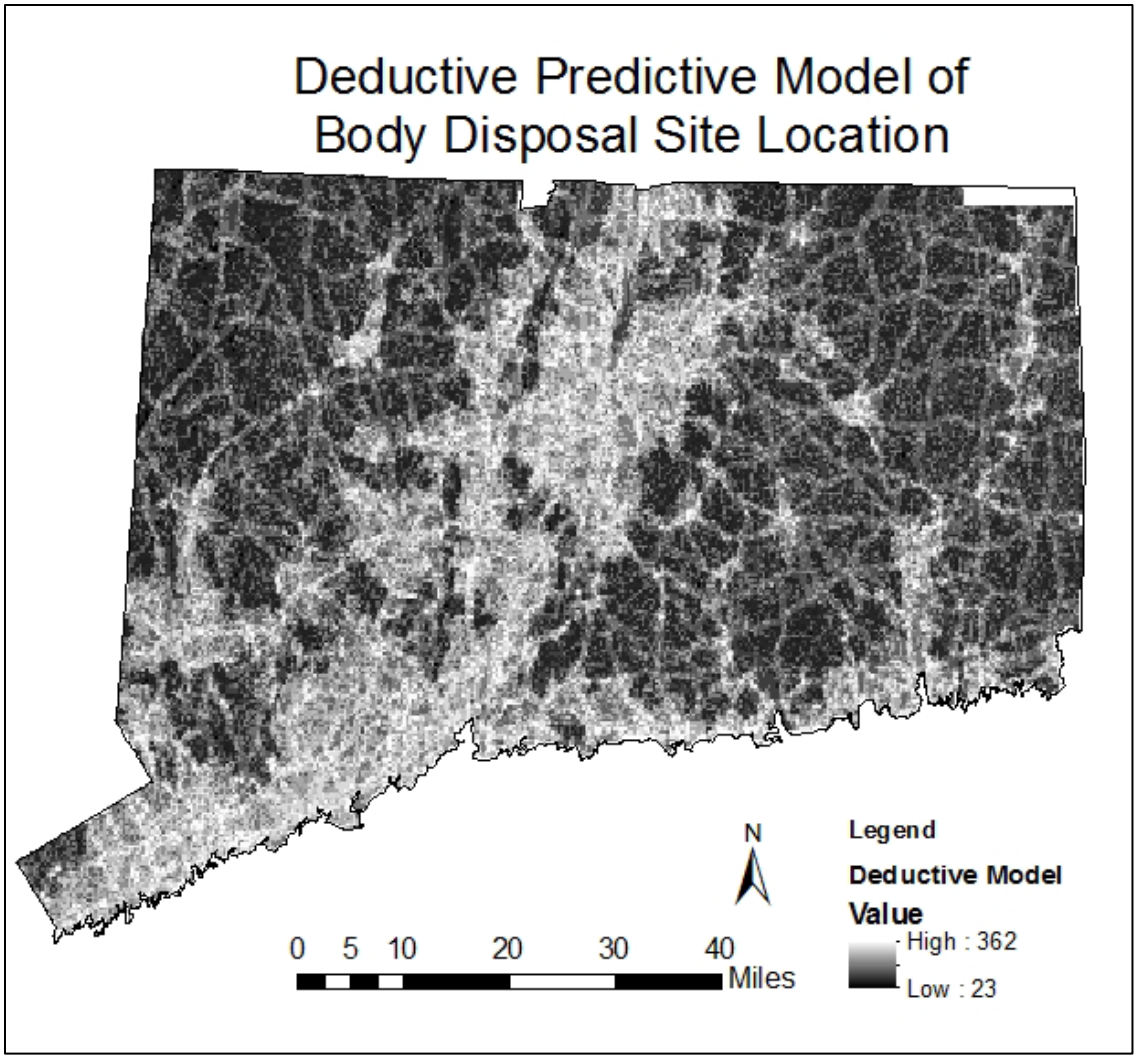


Figure 3.7. Deductive predictive model.

B. Model Validation

The Mann-Whitney U test was used to analyze the known testing sites for significance against the testing non-sites. The results indicate a significant difference, with a p-value of 1.05 e-14.

Using the optimal sectioning point of 277, the model was assessed using the Gain statistic. With 43% of known testing sites predicted, and 22% of the area categorized as high probability, the Gain statistic was +0.49. The positive value indicates positive predictability and is 49% more likely to predict body disposal site location than random chance.

Eight known site locations from Litchfield, Middlesex, Windham, and Tolland counties in Connecticut were used to test the predictive model outside of the original study area. One of the eight sites were correctly predicted (13%). A very low percentage of the four counties was classified as high probability (5%). The Gain statistic for the deductive model applied to Litchfield, Middlesex, Windham, and Tolland was +0.62. The positive value indicates positive predictability and is 62% more likely to predict body disposal site location than random chance.

V. Summary of Results

A sample of 173 known body disposal sites in Connecticut was used to develop and test three predictive models. The study area includes the counties of Hartford, New Haven, Fairfield, and New London.

Univariate Moran's I, Kernel Density analysis, and Ripley's *K*-function were used to test if the known sites were patterned and non-random within the study area. The results of the Moran's I test indicate that sites tend to occur near other known sites within the area. The Kernel Density analysis indicates sites are not homogeneously distributed across the landscape,

but are denser in some areas than others. The Ripley's *K*-function results indicate significant clustering of known sites up to 13.3 miles, with peak site clustering between 3.1 – 3.7 miles.

Descriptive statistics indicate that most sites occur in urban areas, at lower slopes, and in lower population densities. Most victims are found within a mile of a primary or secondary road, and less than a quarter mile from a tertiary road. On average, the bodies of Connecticut residents were found within 5 miles of their residence.

Univariate statistics conducted on the complete spatial randomness data set indicate that site and non-site variables are significantly different for most variables, except for industrial land use, water and wetlands, and designated open space. Not a single variable was significantly different when univariate tests were conducted on the population density data set. Bivariate statistics indicate that thirteen of the variables included in the analyses were significantly correlated. Many of these variables were land use categories correlated with urban area.

Three predictive models were created: two inductive and one deductive. The complete spatial randomness inductive model has the highest Gain statistic and successfully predicts sites and non-sites well. The population density inductive model is less useful, with a lower Gain statistic and poorer predictions of sites and non-sites. The deductive model predicts 50% of the sites in only 22% of the study area. Each model reduces the search area within the study area by over 50%.

The following chapter will discuss the results of the spatial statistical analyses and predictive models. The research objectives presented in the Introduction will be assessed in relation to these results, with an overall aim of identifying site patterning and choosing the predictive model best suited for modeling body disposal location in Connecticut.

CHAPTER FIVE
DISCUSSION

The purpose of this project was to investigate the feasibility of creating predictive models of body disposal location in the Connecticut medicolegal context. The project had three research goals: 1) to evaluate if body disposal sites in Connecticut are patterned and non-random; 2) to create predictive models of body disposal location in Connecticut; and 3) to determine which type of predictive model is best suited for creating models of body disposal location in Connecticut. This chapter will discuss these objectives in relation to the results presented in the previous chapter. Following this, two detailed examples of how law enforcement can utilize predictive models to minimize search areas are presented. The chapter ends with a discussion of recommendations for future predictive modeling efforts in the United States medicolegal context.

I. Spatial Statistics

The results of the exploratory spatial data analysis confirm the nonrandom pattern of body disposal locations in select Connecticut (CT) counties. The Moran's I results show strong positive autocorrelation, indicating that body disposal sites tend to occur near each other in the study area. The Kernel Density Analysis results also indicate that the sites are not distributed homogeneously throughout the study area. Visual assessment shows that the three densest areas correspond to the three of the main cities in Connecticut: Hartford, New Haven, and Bridgeport. The Ripley's *K*-function results indicate that not only are the bodies inhomogeneously distributed in space, they are significantly clustered up to 13.3 mi. Furthermore, these locations occur in similar areas, as the peak distance bands of the Ripley's *K*-function test indicate the peak clustering occurs when sites are 3.1 – 3.7 mi apart. Visual assessment of the known site

points superimposed on the CT town vicinity map (Figure 4.1), makes it clear that the sites occur within certain town vicinities (e.g., Waterbury, New Haven, Bridgeport, Hartford). Using the Measure tool in ArcGIS (ESRI 2014), the maximum east-west length of these town vicinities is calculated at average to be about 6 mi.; the maximum north-south length is similar. The significant clustering of known sites within town vicinities accounts for the peak cluster of sites 3.1 – 3.7 mi. apart.

II. Summary Statistics

Summary statistics indicate that the majority of sites occur in urban areas. The majority of sites also occur at low population densities (mean = 1,665.19 individuals/km^2; median = 441.67 individuals/km^2). This is similar to Kolpan (2015), who found the majority of her Florida cases in low population density (< 300 individual/km^2) urban areas. The combination of urban area with lower population density suggests that body disposal sites occur on the outskirts of higher population areas (Figure 4.2). This conclusion makes sense in the context of Rossmo's (2000) discussion of remote or "out of the way" locations, as offenders are avoiding the high population areas during body disposal. At the same time, the combination of urban with lower population density also confirms the underlying assumption that body disposal sites occur where people exist on the landscape (i.e., urban areas). These results justify the simulation of non-sites to mimic population density in the population density model.

The majority of sites were also found on residential land (42%). This is similar to the site characteristics presented by Rossmo (2000), who noted 45.2% of sites on residential land; yet, it is contradictory to Nethery (2004) who found the majority of sites in forested area (38%) and Hakkanen et al. (2007) who found the majority of sites in the woods (73.2%). The discrepancy

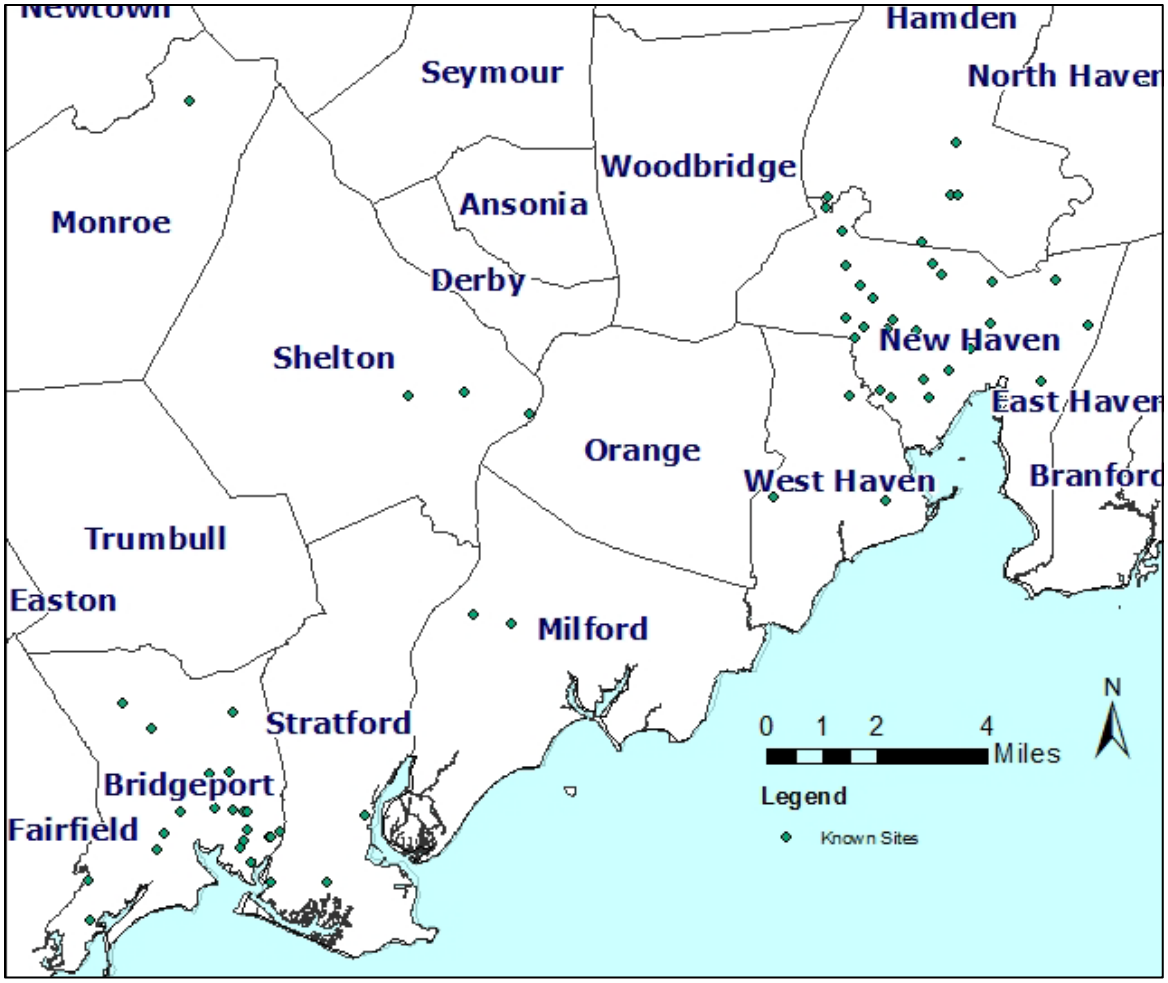


Figure 4.1. Example of known sites clustering in town vicinities.

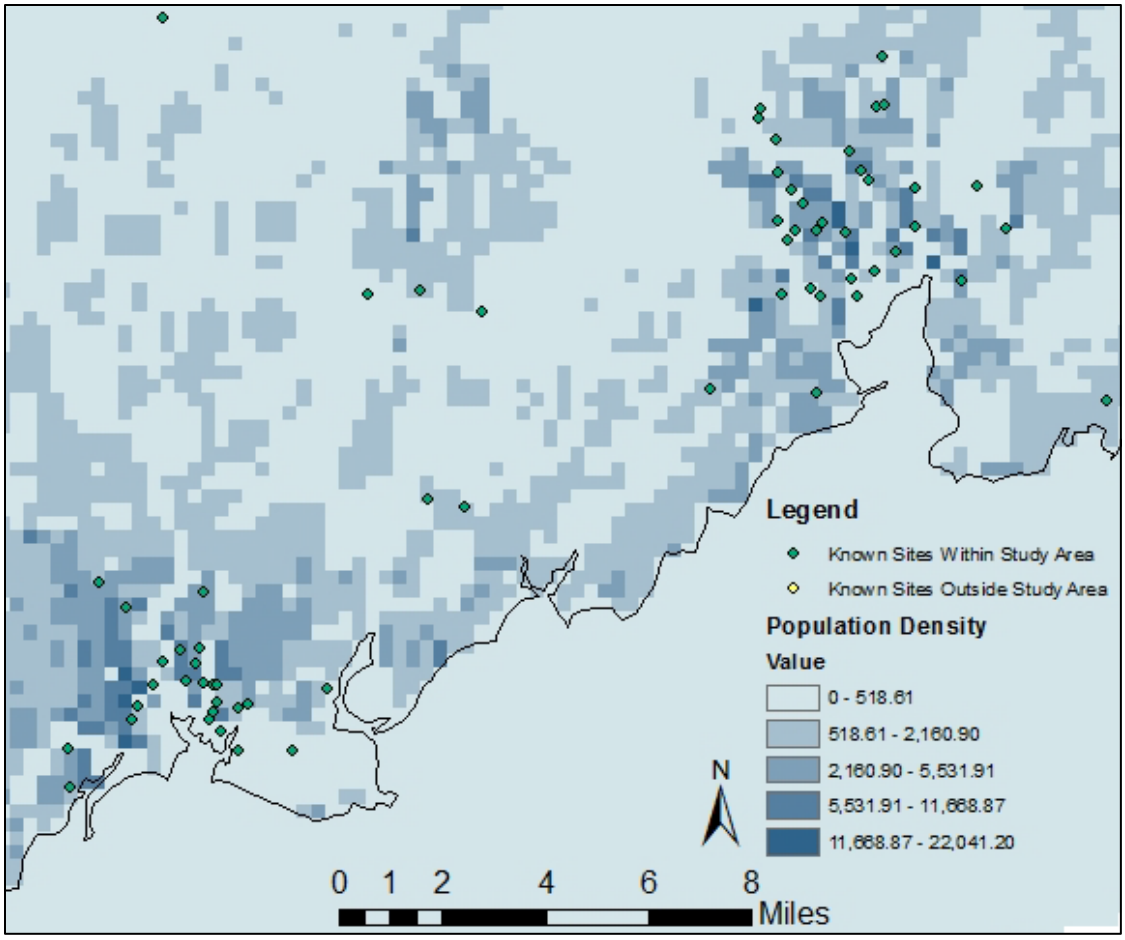


Figure 4.2. Known site relationship to population density.

between statistics may be due to different study locations. Rossmo (2000) uses a North American data set, Nethery (2004) uses a Canadian data set, and Hakkanen et al. (2007) use a Finnish data set. Having many body disposal sites on residential land in an urban area makes sense when taking distance from victim's home to body disposal into account. Much of the journey-to-crime literature indicates that there is a short distance from the victim's home to body disposal location (Fowler 1994; Hanfland et al. 1997; Nethery 2004; Hakkanen et al. 2007; Van Pattern & Delhauer 2007). The mean distance in these studies ranges from 6.2 mi. to 56.9 mi. The median ranges from 4.6 mi. to 12.3 mi. A similar pattern emerges in the distance from victim's home to body disposal sites in Connecticut (CT). When only taking CT residents into account, the mean distance is 5.0 mi. and the median distance is 1.9 mi. In urban areas, it is possible that land use matters less than low population density in choosing a body disposal location.

A large number of sites (31%) were also located on non-residential urban land (industrial, mixed and other urban, and commercial and services). This is a larger percentage than Rossmo (2000) reports (9.6%). These non-residential urban land use areas make sense as they are more remote locations; the areas share a low population density and depending on the site, can provide plenty of natural hiding places for a body (e.g. empty warehouse, machinery, large containers, etc.). Few bodies (6%) were in designated open spaces, such as cemeteries and parks. This was surprising, as open spaces fit the remoteness characteristic, especially at night.

Forest body disposal locations were less common than expected (20%). While similar to Rossmo's (2000) report of 21.2%, it is lower than Nethery (2004)'s report of 38% and very different from Hakkanen et al.'s (2007) report of 73.2%. The discrepancy may be due to the area included in each study. The current study area is selected by county in CT without

differentiating between urban and rural, while Hakkanen et al. (2007) look only at rural areas. The number of water and wetland body disposal sites in the current study (4%) is also lower than the reported numbers, which average around 20% (James 1991; Rossmo 2000; Nethery 2004; Hakkanen et al. 2007). This discrepancy is also likely due to the parameters of data collection. This study excluded most water recoveries, as the original body disposal site was unavailable.

The majority of body disposal sites occurred at low slope angles (mean = 1.02 degrees, median = 0.62 degrees). Having most of the sites at low slope angles aligns with criminological theory, which notes that higher slopes require more effort than lower slopes to move a body (Rossmo 2000). The distance to road values also aligned with previous journey-to-crime literature. Keppel and Birnes (1995) found most bodies within 147.6 ft. of the road. Burton (1998) located most bodies within 286.6 ft. of the road. Rossmo (2000) found all bodies within 450 ft. of the road, while Hakkanen et al. (2007) found the majority within 1640.4 ft. This study found the majority of disposal sites (66%) within 1640.0 ft. from primary and secondary roads, with a mean of 1358.4 ft. and a median of 977.0 ft. The distance to tertiary roads was even shorter, with the majority of disposal sites (79%) within 267 ft. of the road, with mean of 169.4 ft. and a median of 89.4 ft. Shorter distances to the road appear to be favorable for body disposal locations, which aligns with previous observations (Orengo 2006; Congram 2010; Rossmo 2000). Additionally, most cases were closest to tertiary roads (92%) as compared to primary and secondary roads (9%). This also aligns with criminological theory, since tertiary roads are typically in less populated areas and are more remote.

III. Univariate and Bivariate Tests of Significance

The univariate test results for significance between sites and CSR non-sites aligned with the summary statistics, with only industrial land, water and wetland, and open space not having a significant influence on body disposal location. It is unclear why industrial land and designated open space do not have a significant influence on body disposal location. It makes sense that water and wetlands do not have a significant influence on body disposal location, since water recoveries were excluded from this analyses.

The univariate test results for significance between sites and PD non-sites were very different than the CSR results. Not a single variable was significantly different between site and non-site. What this indicates is that population density drives a large portion of the observed patterning of disposal sites. When non-sites are randomly assigned to mimic population density, they become indistinguishable from known body disposal sites.

The bivariate test results for significant correlation between site variables were relatively unsurprising. Land use variables that overlap with urban areas were positively correlated with the urban variable (e.g. residential), and variables that are not found in urban areas were negatively correlated with the urban variable (e.g. agricultural, forest). While correlation does not imply causation, it is important to keep the correlated variables in mind for future modeling efforts; the correlated variables may have significant interactions.

IV. Inductive Model

A. CSR Model

The variables included in the CSR inductive model align with previous criminological literature. As previously discussed, theory supports the observation that more body disposal sites

will be found on mixed and other land use, within 1865 ft. of primary and secondary roads, within 267 ft. of tertiary roads, and occur at lower slope angles. The positive coefficient for urban areas was unexpected, but makes sense when paired with the summary statistics data on population density for known sites in urban areas. It was assumed that urban areas would have fewer body disposal sites because urban areas are high density areas with more potential witnesses to the crime. The summary statistics for sites indicate that urban areas can still have lower population densities, which proves the initial assumption that urban areas are always densely populated incorrect. The coefficient and odds ratio for population density is both logical and contradictory. The mean population density for known sites is 1,665.19 individuals/km^2, indicating sites can occur with increases in population density (i.e. not at 0 individuals/km^2). At the same time, there are no body disposal sites after 10,000 individuals/km^2, making the odds ratio contradictory to the data. It is possible that the equation is missing an interaction term that would address this conundrum. The interaction between slope and forest land indicates that alone, both forest land and high slope degrees decrease the odds of finding sites; yet, when the two are combined, they increase the odds. It may be that while most body disposal sites are not found on forest land (19%) and the majority of sites occur at low slope degrees (median = 0.62 degrees), the sites that are found on forest land occur at higher slopes.

The goal of a predictive model is to maximize the number of sites correctly predicted as high probability, while narrowing the study area to a small high probability area. Gibson (2005, in Verhagen 2007:135) argues that a good model has a Gain statistic of +0.86 or more. Verhagen (2007) notes that a Gain of +0.86 is difficult to attain and doubts if such high gains are truly attainable for most predictive models. Ducke and Munch (2005) note that Gain values around +0.5 are more likely, while Ebert (2000) observes that many predictive models will only

predict 60-70% of sites. With a Gain of +0.59, the CSR model does not meet Gibson's (2005) high standard of +0.86, but is above Ducke and Munch's (2005) +0.5 statistic. The CSR model narrowed the study area down to 34% of its original area, predicting 78% of all the known sites (83% of just the testing sites). This is 574,630.2 acres out of 1,713,339.1 acres. While the model bears improvement, the probability of locating body disposal sites is still 59% higher than random chance.

The CSR model correctly predicts the majority of sites close to the major cities in CT. The sites that are not correctly predicted are the sites outside the cities (Figure 4.3). When comparing the summary statistics of sites correctly and incorrectly predicted, the flaws of the model become more apparent (Table 4.1). The model failed to predict the majority of sites that occurred on forest land (64%). Sites occurring on higher slopes were also less likely to be predicted. The mean slope of correctly predicted sites is 0.93 degrees, while the mean slope of incorrectly predicted sites is 1.60 degrees. A similar issue occurred with population density. The mean population density for correctly predicted sites is 2,611.46 individuals/km^2, with a median of 685.17 individuals/km^2. The mean population density for incorrectly predicted sites is 334.66 individuals/km^2, with a median of 201.57 individuals/km^2. What the visual assessment of correctly v. incorrectly predicted sites and the comparison of summary statistics indicates is that the model works best in urban areas with higher population densities. While the urban variable was included in the model to try and account for the differences between urban and rural areas, it was not effective. Of the 15 sites that did not occur in urban areas, only 2 (13%) were correctly predicted. This suggests that future attempts at modeling need to incorporate variables or interactions that will better account for the difference between high and low population areas, or create two separate models for urban and rural areas.

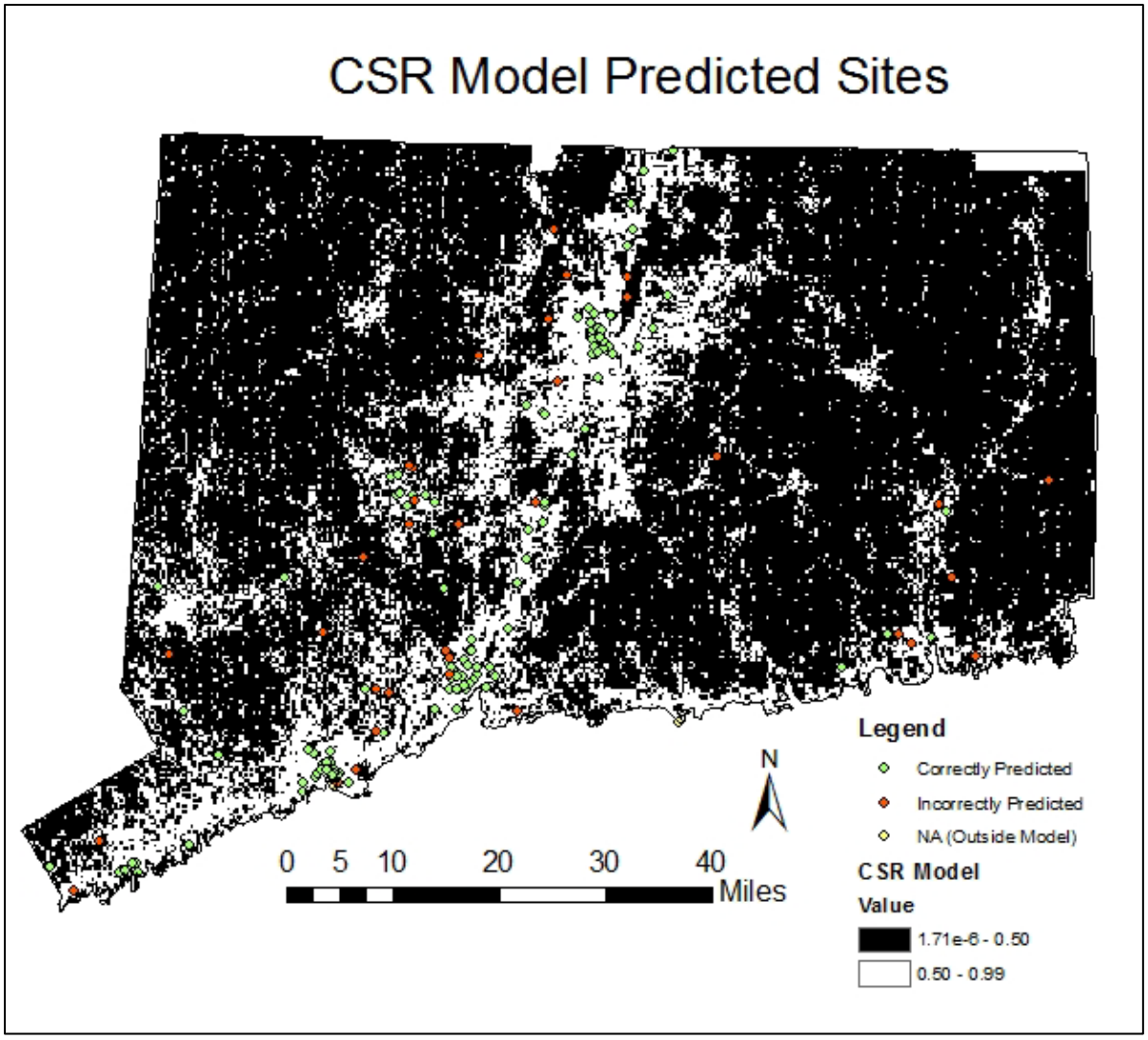


Figure 4.3. Complete spatial randomness model predicted sites.

Table 4.1. Summary statistics for predicted sites (Complete spatial randomness model).

Variable	Correctly Predicted (Present)	Correctly Predicted (Absent)	Incorrectly Predicted (Present)	Incorrectly Predicted (Absent)
WaterWet	5	128	2	36
Resid	61	72	9	29
MixOth	22	111	1	37
Ag	4	129	4	34
Ind	10	123	0	38
Forest	12	121	21	17
CommServ	19	114	1	37
PrimSec	107	26	20	18
Tertiary	110	23	5	33
Urban	131	2	25	13
OpSp	7	126	4	34
Slope	median = 0.43, mean = 0.93		median = 1.16, mean = 1.60	
PopDens	median = 685.17, mean = 2611.46		median = 201.57, mean = 334.66	

WaterWet = Water and wetlands, Resid = Residential, MixOth = Mixed and other urban land use, Ag = Agriculture, Ind = Industrial, CommServ = Commercial and services, PrimSec = Distance to primary and secondary roads, Tertiary = Distance to tertiary roads, OpSp = Designated open space, and PopDens = Population density.

The eight test sites in the four CT counties outside the study area support the need for interactions and/or two separate models for urban and rural areas. The CSR model correctly predicted three of the eight sites. The three correctly predicted sites were in urban areas; of the five incorrectly predicted sites, three occurred on forest land and four occurred in rural areas (Table 4.2).

B. PD Model

The population density model includes different variables than the CSR model. While the CSR included primarily positive coefficient variables, and therefore the model focused on site presence, the PD model includes primarily negative coefficient variables, and therefore focuses on site absence. The coefficients for water and wetland, agricultural land, and forest land align with summary statistics, in which these land use types had fewer sites. The negative coefficient for residential land also makes sense in regards to the summary statistics; although residential land had the majority of disposal sites, the total percentage (41%) was still below 50%. The PD model shares a negative coefficient for slope, a positive coefficient for distance to primary and secondary roads, and a positive coefficient for the forest and slope interaction with the CSR model. Unlike the CSR model, the PD model has a negative coefficient for population density. This is a more accurate representation of site relationship to population density than the positive coefficient in the CSR model. A histogram of site population density values indicates that sites are indeed less likely to occur at low population densities (Figure 4.4).

The PD model has a worse Gain statistic (+0.42) than the CSR model. The model narrows the study area down to 36% of the original area, and in that area, correctly predicts only 58% of known body disposal sites. The PD model does not meet the standards of Gibson (2005),

Table 4.2. Summary statistics for predicted sites outside study area (Complete spatial randomness model).

Variable	Correctly Predicted (Present)	Correctly Predicted (Absent)	Incorrectly Predicted (Present)	Incorrectly Predicted (Absent)
WaterWet	0	3	1	4
Resid	1	2	0	5
MixOth	0	3	0	5
Ag	2	1	1	4
Ind	0	3	0	5
Forest	0	3	3	2
CommServ	0	3	0	5
PrimSec	3	0	2	3
Tertiary	2	1	5	0
Urban	3	0	1	4
OpSp	0	3	1	4
Slope	median = 2.10, mean = 1.92		median = 0.98, mean = 2.19	
PopDens	median = 50.52, mean = 47.23		median = 34.19, mean = 56.65	

WaterWet = Water and wetlands, Resid = Residential, MixOth = Mixed and other urban land use, Ag = Agriculture, Ind = Industrial, CommServ = Commercial and services, PrimSec = Distance to primary and secondary roads, Tertiary = Distance to tertiary roads, OpSp = Designated open space, and PopDens = Population density.

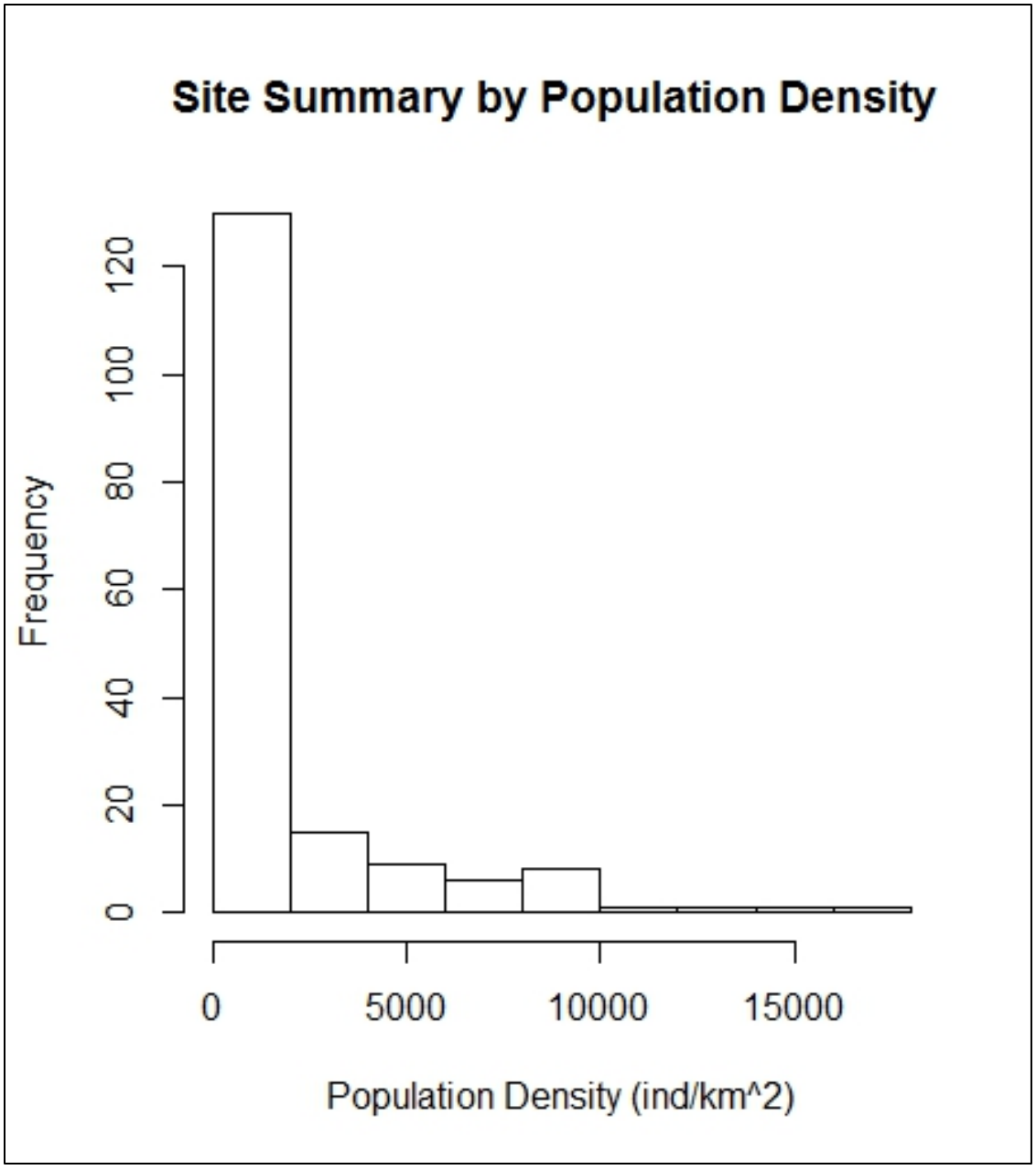


Figure 4.4. Histogram of site population density values.

Ducke and Munch (2005), or Ebert (2000); it does not have a Gain statistic of at least +0.86, it does not have a Gain statistic of at least +0.5, and it does not predict 60 – 70% of known sites.

The PD model is worse than the CSR model in predicting sites. While it still predicts many of the sites near the major cities (Figure 4.5), it does not predict as many as the CSR model. Furthermore, it is just as poor at predicting the sites outside of the densely populated urban areas as the CSR model is; the poor predictability in urban areas is not made up by high predictability in the rural areas (Table 4.3). The poor predictability of the PD model is evident when comparing population density summary statistics. The medians are similar between correctly predicted (447.8 individuals/km^2) and incorrectly predicted (306.9 individuals/km^2). The means are about 1,000 individuals/km^2 apart. What this indicates is that there is no obvious pattern in which sites were correctly predicted and which were not. While the CSR model had difficulty predicting all of the sites, it is clear from the summary statistics that the issue was that the model was missing sites with higher slopes, lower population densities, and in non-urban areas. The PD model had difficulty predicting sites in general. One reason the PD model may have poor predictability and a poor Gain statistic is that the model is over fitted. While the interaction between slope and forest was included because the two are significantly correlated, the interaction may be driving the model to over specificity, making it impractical for use. Future modeling attempts may want to leave the interaction out of the final model. Another reason for the difficulty in predicting site location may be that the sites and PD non-sites were not significantly different from each other, as determined by chi-square and Welch's two sample t-tests. If the sites are not significantly different, the model will have difficulty picking up strong enough differences to create successful predictions.

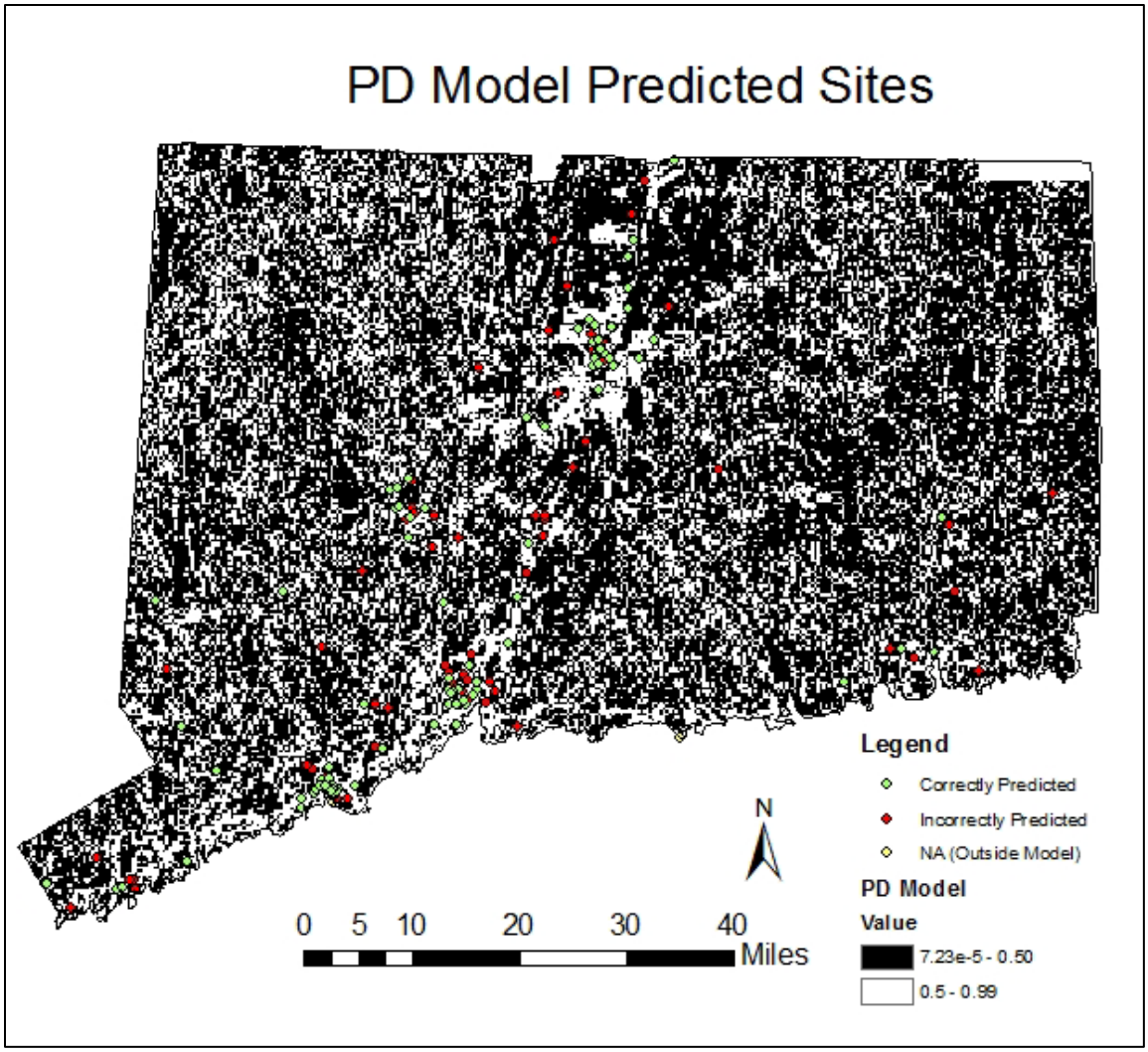


Figure 4.5. Population density model predicted sites.

Table 4.3. Summary statistics for predicted sites (Population density model).

Variable	Correctly Predicted (Present)	Correctly Predicted (Absent)	Incorrectly Predicted (Present)	Incorrectly Predicted (Absent)
WaterWet	3	96	4	68
Resid	42	57	28	44
MixOth	17	82	6	66
Ag	1	98	7	65
Ind	9	90	1	71
Forest	10	89	23	49
CommServ	17	82	3	69
PrimSec	85	14	42	30
Tertiary	84	15	50	22
Urban	95	4	61	11
OpSp	8	91	3	69
Slope	median = 0.51, mean = 0.98		median = 0.79, mean = 1.21	
PopDens	median = 447.8, mean = 1479.7		median = 306.9, mean = 2230.8	

WaterWet = Water and wetlands, Resid = Residential, MixOth = Mixed and other urban land use, Ag = Agriculture, Ind = Industrial, CommServ = Commercial and services, PrimSec = Distance to primary and secondary roads, Tertiary = Distance to tertiary roads, OpSp = Designated open space, and PopDens = Population density.

The eight test sites in the four CT counties outside the study area also indicate the poor predictability of the PD model. Only two of the eight sites were correctly predicted. Given the summary statistics of correctly versus incorrectly predicted sites, it appears that the model could not identify sites in urban areas, but successfully predicted sites on forest land (Table 4.4). The model had difficulty predicting the sites on higher slopes.

V. Deductive Model

The deductive model is more subjective than the inductive models. Any point could have been chosen as the optimal sectioning point, which would change the given Gain statistic for the model. Using the point at which 50% of the sites are predicted allows for the model to predict better than chance. In fact, the Gain statistic (+0.49) for the deductive model suggests that it is likely to predict sites 49% better than random chance. This Gain statistic is not as high as the CSR inductive model (Gain = +0.59), but is better than the PD inductive model (Gain = +0.42).

While the deductive model was split at 50% to provide a Gain statistic for model evaluation, this is not necessarily the best way to understand the model. Deductive models do not produce probabilities for each raster cell the way that logistic regression inductive models do. Rather, the high or low value is an indicator of site presence or absence. The quality of the model is evident in the map values, coverage, and proportion of sites (Table 3.19). For example, Table 3.19 indicates that the top values of 318 – 362 encompass 12% of the study area, and account for 26% of all sites. Table 3.19 also shows that 50% of the sites are predicted between values 277 – 362, which account for 22% of the study area. This summary of top values indicates this model predicts the majority of sites within a small high probability area, which is a sign of a good model.

Table 4.4. Summary statistics for predicted sites outside study area (Population density model).

Variable	Correctly Predicted (Present)	Correctly Predicted (Absent)	Incorrectly Predicted (Present)	Incorrectly Predicted (Absent)
WaterWet	0	2	1	5
Resid	0	2	1	5
MixOth	0	2	0	6
Ag	0	2	3	3
Ind	0	2	0	6
Forest	2	0	1	5
CommServ	0	2	0	6
PrimSec	1	1	4	2
Tertiary	2	0	5	1
Urban	0	2	4	2
OpSp	0	2	1	5
Slope	median = N/A, mean = 0.50		median = 2.20, mean = 2.62	
PopDens	median = N/A, mean = 35.89		median = 26.47, mean = 58.82	

WaterWet = Water and wetlands, Resid = Residential, MixOth = Mixed and other urban land use, Ag = Agriculture, Ind = Industrial, CommServ = Commercial and services, PrimSec = Distance to primary and secondary roads, Tertiary = Distance to tertiary roads, OpSp = Designated open space, and PopDens = Population density.

VI. Model Comparison

The difference in positive predictability and Gain statistic value for the two inductive models is almost certainly due to the different simulated points in each model. The CSR model assumed that non-sites could occur anywhere on the landscape, while the PD model assumed that the non-site density aligned with population density. The tests for independence are helpful to understand how these different assumptions reflect body disposal patterns. The fact that the PD simulated non-sites were not significantly different from the known sites indicates that population density is the main driver of body disposal site location. When non-sites were simulated to match the distribution of people on the landscape, they became indistinguishable from the known sites. Therefore, in this study most body disposal sites occur where people populate the landscape.

The PD model is valuable in identifying population density as an important driver of body disposal site location, but it is also a poor predictor of body disposal location. The PD model has difficulty differentiating site location when population density is controlled for. The CSR model is better at identifying the other factors of body disposal location. The CSR model picks up on the influence of distance to road more so than the PD model. Additionally, looking at the CSR tests for significance for independent variables (Table 3.5) it is evident that a model with CSR non-sites can parse out the influence of different variables in urban areas. For example, residential land and commercial and services land are both significant, but industrial land is not. The ability to differentiate between these different types of urban land is beneficial to understanding where body disposal sites may or may not be in urban areas.

Overall, the models created using the complete spatial randomness simulated non-sites performed better than the population density simulated non-sites (Table 4.5). Future modeling

Table 4.5. Comparison of model validation results.

Model	Gain	Positive Predictability	Negative Predictability	% Area in High Probability	High Probability (Acres)	Total Area (Acres)
Inductive CSR	0.59	78%	73%	34%	574,630.2	1,713,339.1
Inductive PD	0.42	58%	60%	36%	622,852.8	1,713,518.7
Deductive*	0.49	50%	80%	22%	353,363.0	1,712,710.5

*With a sectioning point of 277.

should continue to use CSR data sets, but should keep in mind that while these models perform better, it does not mean that they are unbiased. The CSR data sets do not account for the inhomogeneous distribution of body disposal sites across the landscape. The CSR inductive model performed the best in the given study area, with a Gain of 0.59. The model reduced the study area from 1,713,339.1 acres to 574,630.2 acres. While the deductive model only had a Gain of 0.49, the model did predict 50% of known sites within 22% of the study area. This narrowed the original study area of 1,712,710.5 acres to 353,363.0 acres. The CSR model had a higher Gain statistic, but the deductive model was still able to reduce the study area to a comparable area of high probability while continuing to correctly predict known sites. Future efforts at modeling should continue to explore the benefits of both inductive and deductive models.

When evaluating the models in terms of practical use, developing a deductive model is much easier and more user friendly than dealing with the constraints of logistic regression analyses and interpretations of odds ratios. By assigning both values and weights to the variables, the deductive model can simultaneously account for the driving influence of population density while also addressing the nuances of industrial land versus commercial and

services land. If future studies can continue to improve deductive models, these models may be the best approach for sharing predictive models with law enforcement.

VII. Examples of Use

Predictive models of body disposal location are not intended to replace the current methods of victim search and recovery; rather, they are intended to be yet another tool in the investigative tool kit. The models should be used in conjunction with witness testimony and other traditional investigative techniques. Like geographic profiling (Rossmo 2000), predictive models can be considered once the initial stages of the investigation have already taken place. The goal of introducing predictive modeling into investigations is to aid law enforcement in further narrowing search areas and maximizing resources after a general search area has already been established. The following section provides two examples of how predictive models can be implemented by law enforcement.

The first hypothetical scenario is as follows: law enforcement is searching for an individual that is missing and presumed dead. The offender's residence is unknown, but the victim lives at "Address A", Hartford, CT. Criminological literature suggests that a victim is often found within 20 miles of their residence. Summary statistics from this project indicate that victims who are Connecticut residents are found an average of five miles from their residence. Based on the literature and summary statistics from this project, law enforcement has decided to implement a search with a radius of eight miles from the victim's home.

Predictive models of body disposal location are to be used after an initial search area is established. In this scenario, the total search area is 128,503.8 acres, and is delineated on a map

with a circle to indicate the search radius (Figure 4.6). The predictive map information is added to the original map with the search radius. Areas of high probability are delineated by the white areas and areas of low probability are delineated by the black areas (Figure 4.6). In this scenario, the complete spatial randomness inductive model narrows the search area to 82,167.0 acres, which is a 36% decrease in the search area. By narrowing the search area, the predictive model allows law enforcement to prioritize certain areas within the established search radius while maximizing the available search resources.

The second hypothetical example is as follows: law enforcement is searching for an individual that is missing and presumed dead. The offender's address is unknown, but the victim lives at "Address B", Uncasville, CT 06382. Once again, a search radius of eight miles surrounding the victim's residence is established based on criminological literature and the summary statistics of this project.

In this scenario, the total search area is once again 128,503.8 acres. This search area is delineated on a map with a circle (Figure 4.7). The predictive map information is added, with high probability areas delineated by white and low probability areas delineated by black (Figure 4.7). In this scenario, the complete spatial randomness inductive model narrows the search area to 18,992.7 acres, which is an 85% decrease in the search area. By narrowing the search area, the predictive model once again allows law enforcement to prioritize certain areas within the established search radius while maximizing the available search resources.

VIII. Predictive Modeling of Body Disposal Sites: Lessons Learned

This study is the first to attempt to create predictive models of body disposal location in the United States. The resulting models indicate that in the select counties of Connecticut included in this study, body disposal sites are nonrandom across the landscape and can be

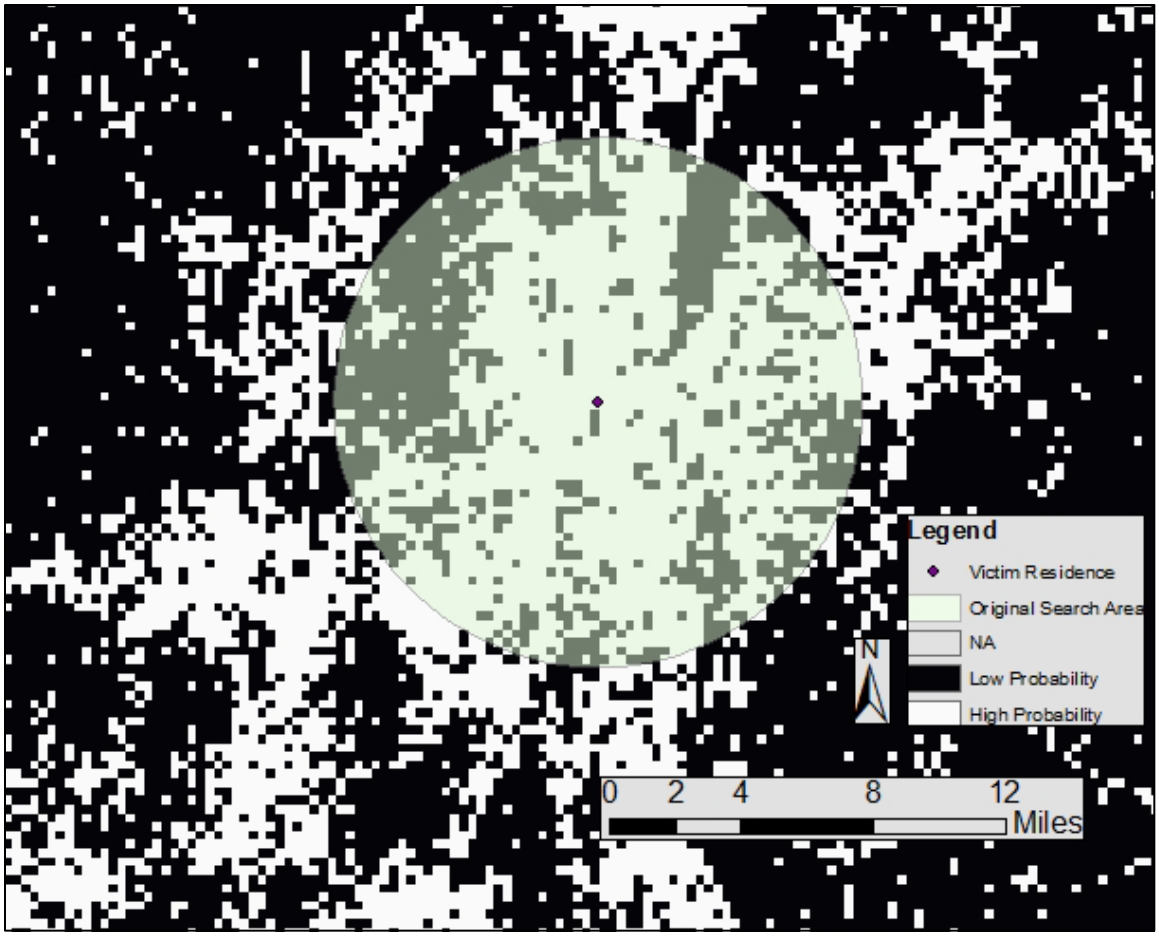


Figure 4.6. Example of complete spatial randomness model applied within an established search area (#1).

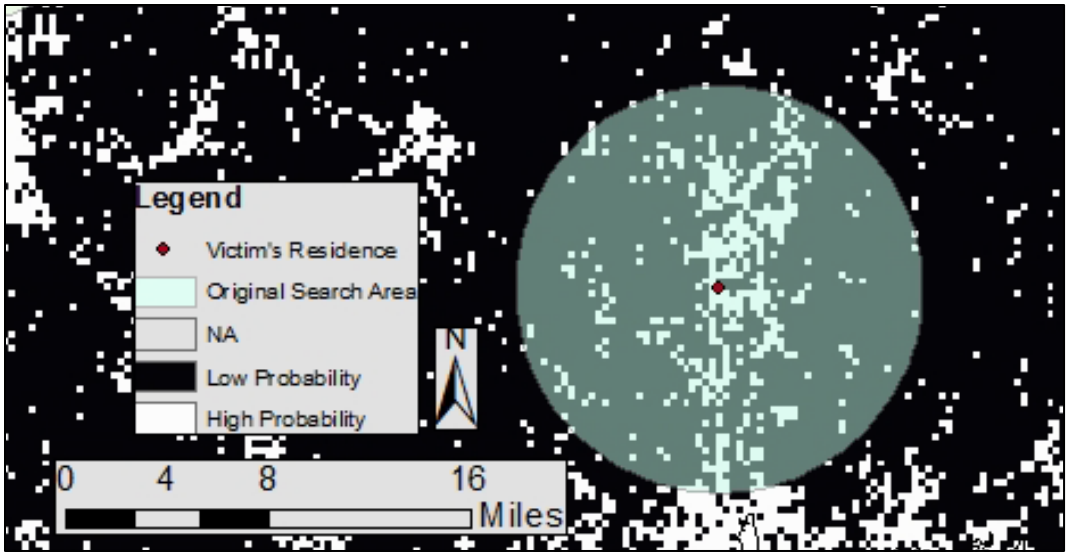


Figure 4.7. Example of complete spatial randomness model applied within an established search area (#2).

predicted based on associated variables, such as population density, slope, and land use. These models are a first step towards providing law enforcement with a tool to help narrow search areas and maximize resources, but as the first of their kind, they necessitate improvement before they can be implemented.

A. Context and Evaluation

The most pressing issue is that the model predicts well in urban areas, but fails to predict sites in rural areas. It is possible that tendency of the model to identify urban body disposal sites is due to ascertainment bias. The data set for this study is inherently biased; it only includes cases that have been located and reported to law enforcement. In this case, most of the known sites are in urban areas. The high frequency of sites on urban land may be due to the chosen study area; the select counties in CT have more area designated as urban than as rural. It could also be because there are more people to find bodies in urban areas than in rural areas. Having

most of the known sites in urban areas may bias the model to assign higher probability to urban areas than rural areas, when in reality, body disposal locations can occur with equal probability in rural areas.

While it is possible that ascertainment bias is influencing the observed results, the driving factor behind the failure to correctly predict both urban and rural sites is likely that urban and rural body disposal patterns are too different for one overarching explanatory model. A review of the criminological literature indicates that body disposal location summary statistics differ by urban/rural categorization and by country. For example, Hakkanen et al. (2007) report that 73.2% of bodies are found in forested area, but Rossmo (2000) reports 21.2% and Nethery (2004) reports 38%. Within these studies, Rossmo (2000) and Nethery (2004) do not distinguish between urban and rural, but Hakkanen et al. (2007) look specifically at rural cases. Additionally, Rossmo (2000) and Nethery (2004) are working in the United States and Canada, respectively, while Hakkanen et al. (2007) are studying body disposal location in Finland.

Future predictive modeling efforts should further examine the role of context in model success. At minimum, different predictive models should be created for urban areas and for rural areas. Researchers should also explore creating different predictive models for each state within the United States. For example, the models of body disposal location in Connecticut created in this thesis are unlikely to have the same predictive ability in the American Southwest, where steep forested land may be less common and areas of flat desert sand may be more common.

Context is also important in understanding the quality of the predictive models. As previously discussed in Chapter 2, researchers using spatial data face the modifiable areal unit problem, in which the results of a test will depend on the assigned study area. One example from this thesis highlights the importance of context in evaluating model predictive ability. The

reported Gain statistics for the models in this thesis were determined based on the given study area of Hartford, New Haven, Fairfield, and New London counties in Connecticut. When the Gain statistic is calculated using the entire state of Connecticut as the study area, the values are very different: the CSR inductive model Gain changes from 0.59 to 0.71; the PD model Gain changes from 0.42 to 0.44; and the CSR deductive model Gain changes from 0.49 to 0.74. All the models appear to improve in predictive ability based on the Gain statistic, although the number of sites correctly predicted do not change. The observed difference in Gain statistics is the result of including the other four counties in Connecticut, which are more rural than the counties included in the study area. As the three models predict well in urban areas, the addition of more rural areas decreased the percentage of high probability area and increased the percentage of low probability area, thereby increasing the Gain statistic for the models. This example highlights the importance of understanding the implication of the chosen study area when evaluating predictive models. Future evaluation of predictive models should continue to use the Gain statistic for validation, but this type of model validation must be taken at face value, with the understanding that it is shaped by the context in which it is applied.

The ability of each model to predict well in different contexts is also demonstrated in the provided examples of how the inductive CSR model can be used by law enforcement. In the first example, the model narrowed the search area by 36%, while in the second example the model narrowed the search area by 85%. The same model was used for each example, but the victim's residence was in two different counties. Once again, this example highlights the importance of understanding the implication of the chosen study area when using predictive models in the future.

B. Data Access

Splitting the data set into a training and a testing set allows the model to be tested with data that was not involved in the creation of the model. This is an efficient way to test the model when limited data is available. Additional validation of these models can occur with further field testing. For example, access to data from 2016 and 2017 would allow for further model validation on entirely new data in the study area.

Future models should also include updated, consistent geospatial data sets. Due to the limited amount of freely accessible geospatial data sets, not all the shapefiles in this study were the same map scale. When different map scales are used, each variable has a different level of detail within the analysis. Variables with a map scale of 1:24,000 will be more detailed than variables with a map scale of 1:100,000. When multiple scales are used in creating predictive models, the model can only be as detailed as the largest map scale. For example, the geospatial data sets included in this thesis range from 1:24,000 (Open Space) to 1:250,000 (Land Use); consequently, the predictive models can only have as much detail as is provided in the 1:250,000 scale. Obtaining maps at a scale of 1:24,000 is recommended for predictive modeling at the state level, as fewer feature boundaries are smoothed and simplified. Additionally, not all the geospatial data fall within the 1990 – 2015 timeframe defined by this project. For example, the land use data available for this project represents data from the years 1970 – 1985. Land use has certainly changed and evolved through time and it is unlikely that every land parcel from 1970 has the same land use designation in 2015. Future models should use geospatial data sets that correspond to the time frame in which the data was collected.

Another source of future model improvement is access to body disposal location coordinates in law enforcement records. The data obtained from the OCME did not contain

exact coordinates for each body disposal site. Coordinates were assigned based on information provided in the cases files, such as cross streets and addresses. This limits the reliability of the data for predictive variables such as distance to road. As a result, this project should be considered a low-resolution model. Finer grained models can be produced in the future with access to coordinates in law enforcement records.

CHAPTER SIX
CONCLUSION

Proving a homicide occurred without a body is nearly impossible and very few "no body" cases go to trial each year in the United States (Lundrigan 2016). Witness testimony can be unreliable, rendering locating a victim's body a difficult process (Fruzzetti et al. 1992). Investigations may be further complicated if the victim's body was moved to a secondary location, potentially delaying its discovery and subsequent recovery. Forensic site prediction models can augment relying solely on witness testimony or luck. The goal of this study was to test the feasibility of creating predictive models of body disposal sites in the United States using data collected from the Office of the Chief Medical Examiner, Connecticut.

While predictive modeling has successfully been used to identify archaeological sites since the 1980s, it remains a nascent method in forensic anthropology and medicolegal investigations. Orengo (2006) created a predictive model of body disposal site location for medicolegal cases in Leicestershire, United Kingdom. Congram (2010) utilized known site locations of mass graves to create a predictive model of mass grave location in Spain. The results of these studies indicate that the deposition of bodies, whether it be in a medicolegal context or in the context of armed civil conflict, is patterned and nonrandom. The current study contributes to this nascent field of research by applying predictive modeling in the United States medicolegal context.

Results from this study support Orengo (2006) and Congram's (2010) assertion that body disposal locations are patterned and nonrandom. Specifically, results show that sites tend to occur near each other in the study area and that there is significant clustering of sites up to 13.3 miles. Both inductive and deductive methods of predictive modeling indicate that predictive

models of body disposal location are indeed feasible. The CSR model narrowed the study area down to an area of high probability 66% smaller than the original size. Additionally, the CSR model predicted the sites of body disposal location 59% more likely than random chance. The deductive model was also successful. Fifty percent of the known sites were predicted in an area 78% smaller than the original study area. These results indicate that predictive models of body disposal location have a real possibility of narrowing search areas and maximizing resources for law enforcement when searching for missing victims.

The two hypothetical examples of how law enforcement can include predictive models in their search efforts emphasize the importance of predictive modeling. In the first example, the CSR inductive model narrowed the initial search area from 128,503.8 acres to 82,167.0 acres. In the second example, the CSR inductive model narrowed the initial search area from 128,503.8 acres to 18,992.7 acres. These examples show how predictive models can help prioritize locations when large areas of land are being searched. Including predictive models in the stages of planning allow law enforcement to maximize available search resources and minimize the search area.

While the predictive model results are promising, this study demonstrates the need for future modeling efforts. When non-site locations are simulated to mimic the distribution of people on the landscape, no variables significantly influence body disposal location. This indicates that population density is the main driver of body disposal sites. Using sites simulated to match complete spatial randomness, other significant variables emerge. Deductive models may be the best method to address the different levels of influence each variable has on the model, but further research is necessary to better understand how to model these trends.

The most pressing concern for future modeling efforts is how to model urban and rural environments. The models in this study indicate that urban body disposal site locations can be predicted with some level of confidence; the same cannot be said for rural body disposal locations. Further investigation is necessary to determine if the failure of the models to predict rural sites is due to ascertainment bias, or if it is because the patterns between urban and rural body disposal practices are too different to be captured in the same model. It is likely that the patterns between urban and rural body disposal practices cannot be explained by an overarching model. Context is important when applying predictive models, and future modeling efforts should create different predictive models for urban and rural areas. Different models should also be created for states across the United States that are representative of different population density distributions and different environmental landscapes.

Predictive models of body disposal location are not intended to replace the current methods of victim search and recovery; rather, they are intended to be yet another tool in the investigative tool kit. The models should be used in conjunction with witness testimony and other traditional investigative techniques. Like geographic profiling (Rossmo 2000), predictive models can be considered once the initial stages of the investigation have already taken place. The goal of introducing predictive modeling into investigations is to aid law enforcement in further narrowing search areas and maximizing resources.

REFERENCES CITED

Bevan, A. and J. Conolly
2006 Multiscalar Approaches to Settlement Pattern Analysis. In: *Confronting Scale in Archaeology: Issues of Theory and Practice*, edited by G. Lock and B. Molyneaux, pp. 217-34. New York: Springer.
Bivand, R.S., E.J. Pebesma, and V. Gomez-Rubio
2008 *Applied Spatial Data Analysis with R.* Springer, New York.
Brandt, R., B.J. Groenewouldt, and K.L. Kvamme
1992 An experiment in archaeological site location: modeling in the Netherlands using GIS techniques. *World Journal Archaeology* 24(2):268-282.
Brantingham, P. and P. Brantingham
1995 Criminality of place: Crime generators and crime attractors. *European Journal on Criminal Policy and Research* 3(3):1-26.
Brantingham, P.J. and P.L. Brantingham
1981 *Environmental Criminology*. Sage, London.
1984 *Patterns in Crime*. Macmillan, New York.
Brantingham, P.L. and P.J. Brantingham
1993 Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Environmental Psychology* 13:3-28.
1998 Environmental criminology: from theory to urban planning practice. *Studies on Crime and Crime Preventions* 7:1-30.
Burgess, E.W.
1925 The growth of the city: An introduction to a research project. In *The City*, edited by R.E. Park, E.W. Burgess and R.D. McKenzie, pp. 47-62. University of Chicago Press, Chicago.
Burton, C.
1998 The CATCHEM database: Child murder in the United Kingdom. Presentation at the International Homicide Investigators Association Symposium, Zutphen.
Carrer, F.
2013 An ethnoarchaeological inductive model for predicting archaeological site location: A case-study of pastoral settlement patterns in the Val di Fiemme and Val di Sole (Trentino, Italian Alps). *Journal of Anthropological Archaeology* 32:54-62.
Chainey, S. and J. Ratcliffe
2005 *GIS and Crime Mapping*. John Wiley & Sons, Chichester.
Congram, D.
2010 *Spatial analysis and predictive modeling of clandestine graves from rearguard repression of the Spanish Civil War*. PhD Dissertation. Department of Archaeology, Simon Fraser University, Burnaby.
2013 Deposition and Dispersal of Human Remains as a Result of Criminal Acts: *Homo sapiens sapiens* as a Taphonomic Agent. In *Manual of Forensic Taphonomy*, edited by J.T. Pokines and S.A. Symes, pp. 251-287. CRC Press, Boca Raton.
Congram, D., A. Green, and H. Tuller
2016 Mapping the Missing: A New Approach to Locating Missing Persons Burial Locations in Armed Conflict Contexts. In *Missing Persons: Multidisciplinary Perspectives on the Disappeared*, pp. 207 – 223, edited by D. Congram. Canadian Scholars Press, Toronto.

Connecticut Department of Energy and Environmental Protection
2017 GIS Data. Electronic document, http://www.ct.gov/deep/cwp/view.asp?a=2698&q=322898&deepNav_GID=1707%20#BaseMap, accessed August 25 2016.

Conolly, J. and M. Lake
2006 *Geographical Information Systems in Archaeology*. Cambridge University Press, Cambridge.

Cornish, D. and R. Clarke
1986 *The Reasoning Criminal: Rational Choice Perspectives on Offending*. Springer-Verlag, New York.

DiBiase, T.A.
2014 *No-body homicide cases: A practical guide to investigating, prosecuting, and winning cases when the victim is missing.* CRC Press, Boca Raton.

Dore, C.D. and L. Wandsnider
2006 Modeling for management in a compliance world. In *GIS and archaeological site location modeling*, edited by M.W. Mehrer and L.M. Wescott, pp. 75-95. CRC/Taylor & Francis, Boca Raton.

Ducke, B.
2015 Spatial Cluster Detection in Archaeology: Current Theory and Practice. In *Mathematics and Archaeology*, edited by J.A. Barcelo and I. Bogdanovic, pp. 352 – 368. CRC Press, Boca Raton.

Ebert, D.
2004 Applications of Archaeological GIS. *Canadian Journal of Archaeology* 28:319-341.

ESRI
2009 ArcGIS Desktop: Release 9.3. Redlands, CA: Environmental Systems Research Institute.
2014 ArcGIS Desktop: Release 10.3. Redlands, CA: Environmental Systems Research Institute.
2016a Resolution. Electronic document. http://support.esri.com/other-resources/gis-dictionary/term/resolution, accessed February 4, 2017.
2016b Incremental Spatial Autocorrelation. Electronic document, http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/incremental-spatial-autocorrelation.htm, accessed March 13 2017.

FBI
2016 Crime in the United States. Electronic document, https://ucr.fbi.gov/crime-in-the-u.s/2014/crime-in-the-u.s.-2014/tables/table-4, accessed April 3 2017.

Fischer, M. and A. Getis
2010 *Handbook of Applied Spatial Statistical Analysis.* Springer, New York.

Fowler, K.S.
1994 Analysis of dump sites of homicide victims in Jefferson Parish, Louisiana. Unpublished Master's thesis. Department of Geography and Anthropology, Louisiana State University, Baton Rouge.

Fruzzetti, A.E., K. Tolan, S.A. Teller, and E.F. Loftus

1992 Memory and eyewitness testimony. In *Aspects of memory, Vol. 1: The practical aspects*, 2nd ed., edited by M.M. Gruneberg and P.E. Morris, pp. 18 – 50. Taylor & Francis/Routledge, Florence.

Gaffney, V. and M. van Leusen

1995 Postscript – GIS, environmental determinism and archaeology: a parallel text. In *Archaeology and Geographical Information Systems: A European Perspective,* edited by G. Lock and Z. Stancic, pp. 367-382.

Haining, R.P.

2003 The Nature if Georeferenced Data. In *Handbook of Applied Spatial Analysis: Software Tools, Methods, and Applications*, edited by M.M. Fischer and A.G. Getis, pp.197-218. Springer, Berlin.

Hakkanen, H., K. Hurme, and M. Liukkonen

2007 Distance Patterns and Disposal Sites in Rural Area Homicides Committed in Finland. *Journal of Investigative Psychology and Offender Profiling* 4:181-197.

Hanfland, K.A., R.D. Keppel, and J.G. Weis

1997 *Case management for missing children homicide investigation*. Washington State Office of the Attorney General, Seattle.

Hodder, I.

1982 *Symbols in Action: Ethnoarchaeological Studies in Material Culture*. Cambridge University Press, London.

Hosmer, D.W. and S. Lemeshow

2000 *Applied logistic regression*. John Wiley, New York.

Hunter, J.R.

2002 Foreward: A pilgrim in forensic archaeology – a personal view. In *Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives*, edited by W.D. Haglund and M.H. Sorg, pp. xxv-xxxii. CRC Press, Boca Raton.

Illian, J., A. Paenttinen, H. Stoyan, and D. Stoyan

2008 *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons, Chichester.

James. E.

1991 *Catching serial killers*. International Forensic Services, Lansing.

Keppel, R.D. and J.G. Weis

1994 Time and distance as solvability factors in murder cases. *Journal of Forensic Sciences* 39:368-401.

Killam, E.W.

2004 *The Detection of Human Remains*. 2nd Ed. Charles C. Thomas, Springfield.

King, T.F.

1984 An overview of the archaeological law at the federal level. *American Archaeology* 4:115-118.

Kleinbaum, D.G.

1994 *Logistic Regression, a self-learning text*. Springer-Verlag, New York.

Kohler, T.A. and S.C. Parker

1986 Predictive models for archaeological resource location. In *Advances in Archaeological Method and Theory*, Vol. 9, edited by M.B. Schiffer, pp. 397-452. Academic Press, Orlando.

Kolpan, K.E.

2015 Utilizing geographic information systems (GIS) to analyze geographic and demographic patterns related to forensic case recovery locations in Florida. Paper presented at the 67th Annual Meeting of the American Academy of Forensic Sciences, Orlando.

Kvamme, K.L.

1983 Computer processing techniques for regional modeling of archaeological site locations. *Advances in Computer Archaeology* 1:26-52.

1984 Models of prehistoric site location near Pinyon Canyon, Colorado. In *Papers of the Philmont Conference on the archaeology of Northeastern New Mexico*, edited by C.J. Condie, pp.349-370.

1988 Using existing archaeological survey data for model building. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W.J. Judge and L. Sebastian, pp. 301-323. US Department of the Interior, Bureau of Land Management Service Center, Denver.

1989 Geographic information systems in regional archaeological research and data management. In *Archaeological Method and Theory*, Vol. 1, edited by M.B. Schiffer, pp. 139-203. University of Arizona Press, Tucson.

1990 The fundamental principles and practice of predictive archaeological modeling. In *Mathematics and information science in archaeology: a flexible framework*, edited by A. Voorrips, pp. 257 – 295. Holos-Verlag, Bonn.

2006 There and back again: Revisiting Archaeological Location Modeling. In *GIS and Archaeological Site Location Modeling*, edited by M.W. Mehrer and K.L. Wescott, pp. 2-35. Taylor & Francis, Boca Raton.

Levine, L.J., H.R. Campbell Jr., and J.S. Rhine

1984 Perpendicular Forensic Archaeology. In *Human Identification, Case Studies in Forensic Anthropology*, edited by T.A. Rathbun and J.E. Buikstra, pp. 87-95. Charles C. Thomas, Springfield.

Lock, G. and T. Harris

2006 Enhancing Predictive Archaeological Modeling: Integrating Location, Landscape, and Culture. In *GIS and Archaeological Site Location Modeling*, pp. 36-55. Taylor & Francis, Boca Raton.

Longley, P.A., M.F. Goodchild, D.J. Maguire, and D.W. Rhind

2011 *Geographic Information Systems & Science*. 4th ed. Wiley: Hoboken.

Lundrigan, S.

2016 A Review of Research into the Spatial Behaviour of Murderers and Implications for Investigations Involving Missing Murder Victims. In *Missing Persons: Multidisciplinary Perspectives on the Disappeared*, edited by D. Congram, pp. 184-206. Canadian Scholars' Press, Toronto.

Lundrigan, S. and D. Canter

2001 A multivariate analysis of serial murderers' disposal site location choice. *Journal of Environmental Psychology* 21:423-432.

Lynch, K.

1960 *The Image of the City*. MIT Press, Cambridge.

Manhein, M.H., G.A. Listi, and M. Leitner

2006 The application of geographic information systems and spatial analysis to assess dumped and subsequently scattered human remains. *Journal of Forensic Sciences* 51(3):469-474.

Maschner, H.D.G.
1996 *New methods, old problems: geographic information systems in modern archaeological research*. Southern Illinois University at Carbondale Press, Carbondale.

McLaughlin, J.E.
1974 *The Detection of Buried Bodies*. Andermac, Yuba City.

Mehrer, M.W. and K.L. Wescott
2006 *GIS and Archaeological Site Location Modeling.* CRC Press, Boca Raton.

Microsoft
2015 Microsoft Excel.

Morse, D., J. Duncan, and J. Stoutamire
1983 *Handbook of Forensic Archaeology and Anthropology*. Rose Printing.

Nethery, K.
2004 Non-familial abductions that end in homicide: An analysis of the distance patterns and disposal sites. Unpublished Master's thesis. School of Criminology, Simon Fraser University, Burnaby.

Oak Ridge National Laboratory
2016 LandScan Documentation. Electronic document, http://web.ornl.gov/sci/landscan/landscan_documentation.shtml, accessed October 2, 2016.

Orengo, H.A.
2006 Detection of Body Dump Sites and Clandestine Burials: A GIS-based Landscape Approach. In *Digital Discovery: Exploring New Frontier in Human Heritage, CAA2006. Computer Applications and Quantitative Methods in Archaeology,* edited by J.T. Clark and E.M. Hagemeister, pp. 185-190. Proceedings of 34th Conference, Fargo.

Peuquet, D.J.
1977 *Raster data handling in geographic information systems.* Geographic Information Systems Laboratory, Buffalo.

Pianka, E.R.
1974 *Evolutionary Ecology*. Harper and Row, New York.

R Core Team
2015 R: a language and environment for statistical computing. Vienna, Austria. R Foundation for Statistical Computing. http://www.R-project.org/.

Rathje, W.L.
1979 Modern Material Culture Studies. In *Advances in Archaeological Methods and Theory,* Vol. 2, edited by M.B. Schiffer, pp. 1-37. Academic Press, New York.

Rathje, W.L. and M.B. Schiffer
1982 *Archaeology*. Harcourt Brace Jovanovich, New York.

Reddick, Amanda J.
2005 *The Use of Spatial Techniques in the Location and Excavation of Contemporary Mass Graves.* Master's thesis. Institute of Archaeology and Antiquity, University of Birmingham, Birmingham.

Rengert, G.F., A.R. Piquero, and P.R. Jones
1999 Distance decay reexamined. *Criminology* 37:427-445.
Revelle, W.
2017 Package 'psych'. Electronic document, https://cran.r-project.org/web/packages/psych/psych.pdf, accessed Feb 20 2017.
Ridges, M.
2006 Understanding H-G behavioural variability using models of material culture: An example from Australia. In *GIS and Archaeological Site Location Modeling*, edited by M. Mehrer and K. Wescott, pp. 123-143. CRC Press, Boca Raton.
Ripley, B.D.
1976 The second order analysis of stationary point processes. *Journal of Applied Probability* 13:255-266.
1977 Modeling Spatial Patterns. *Journal of the Royal Statistical Society* 2:172-212.
1981 *Spatial Statistics*. Wiley, New York.
Rizopoulos, D.
2015 Package 'ltm'. Electronic document, https://cran.r-project.org/web/packages/ltm/ltm.pdf, accessed February 24 2017.
Rossmo, K.
2000 *Geographic Profiling*. CRC Press, Boca Raton.
Rupert, M.G., S.H. Cannon, J.E. Gartner, J.A. Michael and D.R. Helsel
2008 Using Logistic Regression to Predict the Probability of Debris Flows in Areas Burned by Wildfire, Southern California, 2003-2006. U.S. Geological Survey Open-File Report 2008-1370.
Safarik, M.E., J.P. Jarvis, and K.E. Nussbaum
2000 Elderly female serial sexual homicide: A limited empirical test of criminal investigative analysis. *Homicide Studies* 3(1):6-29.
SAS
2017 Usage Note 24455: Estimating an odds ratio for a variable involved in an interaction. Electronic document, http://support.sas.com/kb/24/455.html, accessed March 3 2017.
Sayer, D. and M. Wienhold
2013 A GIS-Investigation of Four Early Anglo-Saxon Cemeteries: Ripley's *K*-function Analysis of Spatial Groupings Amongst Graves. *Social Science Computer Review* 31(1):71-89.
Schwarz, K.R. and J. Mount
2006 Integrating spatial statistics into archaeological data modeling. In *GIS and Archaeological Site Location Modeling*, edited by M.W. Mehrer and K.L. Wescott, pp. 163-189. CRC Press, Boca Raton.
Schiffer, M.B.
1983 Toward the Identification of Formation Processes. *American Antiquity* 48(4):675-706.
Snook, B., R.M. Cullen, A. Mokros, and S. Harbot
2005 Serial murderers' spatial decisions: Factors that influence crime location choice. *Journal of Investigative Psychology and Offender Profiling* 2:147-164.
Steward, J.

1938 *Basin-plateau Aboriginal Sociopolitical Groups*. U.S. Government Printing Office, Washington D.C.

Streed, T.

1989 Dump-site Analysis and Behavioral Characteristics of Psychopathic Killers. Paper presented at the 41st annual meeting of the American Academy of Forensic Sciences, Las Vegas.

Szumilas, M.

2010 Explaining Odds Ratios. *Journal of Canadian Academy of Adolescent Psychiatry* 19(3):227-229.

Topologically Integrated Geographic Encoding and Referencing (TIGER)

2010 MAF/TIGER Feature Class Code (MTFCC) Definitions. Electronic document, http://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2010/TGRSHP10SF1AF.pdf, accessed March 13 2017.

Tobler, W.

1970 A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46:234-240.

Tomlinson, R.F.

1982 Panel discussion: technology alternatives and technology transfer. In *Computer assisted cartography and information processing: hope and realism*, edited by D.H. Douglas and A.R. Boyle, pp. 65 – 71. Canadian Cartographic Association, Ottawa.

Tuller, Hugh, U. Hofmeister and S. Daley

2008 Spatial analysis of mass grave mapping data to assist in the reassociation of disarticulated and commingled human remains. In *Recovery, Analysis, and Identification of Commingled Human Remains*, edited by B. Adams and J Byrd, pp.7-29. Humana Press, Totowa.

United States Census Bureau (USCB)

2017 State-based Metropolitan and Micropolitan Statistical Areas Maps. Electronic document, https://www.census.gov/geo/maps-data/maps/statecbsa.html, accessed March 13 2017.

United States Department of Agriculture (USDA)

2017 What is rural? Electronic document, https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/what-is-rural.aspx, accessed March 13 2017.

United States Geological Survey

2017 USGS DS 240: Enhanced Historical Land-Use and Land-Cover Data Sets of the U.S. Geological Survey. Electronic document, https://water.usgs.gov/GIS/dsdl/ds240/index.html, accessed August 30 2017.

University of Connecticut MAGIC

2017 Connecticut GIS Data. Electronic document, http://magic.lib.uconn.edu/connecticut_data.html, accessed August 37 2016.

van Leusen, P.M.

1996 GIS and locational modeling in Dutch archaeology: A review of current approaches. In *New Methods, Old Problems: Geographic Information Systems in Modern Archaeological Research*, edited by H.D.G. Maschner, pp. 177-197. Southern Illinois University at Carbondale Press, Carbondale.

2002 *Pattern to process: methodological investigations into the formation and interpretation of spatial patterns in archaeological landscapes.* Rijksuniversiteit Groningen.

Van Patten, I.T., and P.Q. Delhauer

2007 Sexual homicide: A spatial analysis of 25 years of deaths in Los Angeles. *Journal of Forensic Science* 52(5):1129-1141.

Vaughn, S. and T. Crawford

2009 A predictive model of archaeological potential: An example from northwestern Belize. *Applied Geography* 29:542-555.

Verhagen. P.

2007 *Case Studies in Archaeological Predictive Modelling*. Leiden University Press, the Netherlands.

Verhagen, P. and T.G. Whitley

2012 Integrating Archaeological Theory and Predictive Modeling: a Live Report from the Scene. *Journal of Archaeology Method and Theory* 19:49-10.

Verhagen, P., H. Kamermans, M. van Leusen, J. Deeben, D. Hallewas and P. Zoetbrood

2007 First thoughts on the incorporation of cultural variables into predictive modelling. In *Case Studies in Archaeological Predictive Modeling*, edited by P. Verhagen, pp. 203-208. Leiden University Press, the Netherlands.

Waller, L.A. and C.A. Gotway

2004 *Applied Spatial Statistics for Public Health Data.* John Wiley & Sons, Hoboken.

Warren, R.E. and D.L. Asch

2000 A predictive model of archaeological site location in the Eastern Prairie Peninsula. In *Practical Applications of GIS for Archaeologists: A Predictive Modeling Kit*, edited by K.L. Wescott and R.J. Brandon, pp. 5 – 32. Taylor & Francis, Philadelphia.

Wescott, K.L.

2000 Introduction. In *Practical Applications of GIS for Archaeologists*, edited by K.L. Wescott and R.J. Brandon, pp. 1-5. Taylor & Francis, Philadelphia.

Wescott, K.L. and R.J. Brandon (editors)

2000 *Practical Applications of GIS for Archaeologists*. Taylor & Francis, Philadelphia.

Wheatley, D. and M. Gillings

2002 *Spatial Technology and Archaeology*. Taylor & Francis, New York.

Willey, G.R.

1953 *Prehistoric Settlement Patterns in the Viru Valley, Peru*. Bulletin 155, Bureau of American Ethnology, Smithsonian Institution, Washington D.C.

Zuur, A.F., E.N. Ieno, and G.M. Smith

2007 *Analyzing Ecological Data*. Springer, New York.

APPENDIX

SUGGESTED PROTOCOL FOR CREATING BODY DIPOSAL SITE PREDICTION MODELS USING THE WEIGHTED MAP-LAYER APPROACH

1. **Create a geodatabase.** The company behind ArcGIS (ESRI) suggests that all data be organized and stored in a geodatabase. A geodatabase stores the files in an easy access folder, which can be shared with other ArcGIS users. The file storage in a geodatabase also optimizes performance and has few size limitations.
    a. Creating a geodatabase can be done in the ArcCatalog package, included in ArcGIS 10.3.
        i. Open ArcCatalog > browse for desired folder location in the "Catalog Tree" > Click on desired folder name > go to File > New > File Geodatabase. Change the filename to desired title.
2. **Compile all geospatial data.** The United States Geological Survey (USGS) hosts freely accessible geospatial data that includes elevation, orthoimagery, hydrography, transportation, land cover, and geology, which are available at http://www.nationalmap.gov. The United States Census Bureau also hosts freely accessible geospatial data that includes population counts, transportation, water features, and various state, county, and town level boundaries, which are available at www.census.gov/geo/maps-data/data/tiger.html. Many states also host their own geospatial datasets. For example, Connecticut GIS data is available through the University of Connecticut Library at www.magic.lib.uconn.edu/connecticut.data.html.
    a. Store all downloaded geospatial data in the geodatabase.
3. **Create a comma delimited file (.csv).** Using Microsoft Excel, create a simple spreadsheet of known body disposal site location. In the first column, assign the site an ID number (e.g., 1, 2, 3). The second column should be the longitude, and the third column should be the latitude. Additional columns can be added if further data is necessary to link with the body disposal site (e.g., victim age range, sex, ethnicity).
    a. Once the spreadsheet is complete, save the file as a comma delimited value (.csv). In Excel, go to File > Save As > Save as type: = CSV (Comma delimited).
4. **Import site location to ArcGIS 10.3.** In the ArcMap environment, go to File > Add Data > Add Data.

a. The .csv file will be added to the ArcMap work environment.
b. Right click on the .csv file > Display XY Data > X Field: = Longitude, Y Field: = Latitude.
c. Under Coordinate System of Input Coordinates, it will be unknown. Select > Edit > Geographic Coordinate Systems > World > WGS 1984 > Press OK > It will bring you back to the Display XY Data dialog box, click OK.
    i. A prompt will show up warning that there is no Object-ID field, click OK.
    ii. A new layer will appear named [filename].csv Events > Right click on the layer name > Data > Export > Export: All records, Output table: = click the folder to browse for previously created geodatabase, name the file (e.g., site location) > Click OK.
    iii. A prompt will appear asking if you want to add the exported data to the map as a layer > Click Yes.
    iv. At the top of the Table of Contents window, click on the "List by Drawing Order" tab. Here there will be two copies of the imported data, the one named [filename].csv Events, and the other name given to the exported data set. Right click on the [filename].csv Events > Remove.

5. **Create a set of random non-site points.** To test if the geospatial variables are significantly related to site location, each layer must be tested against a random distribution of simulated non-sites.
    a. ArcToolbox > Data Management Tools > Feature Class > Create Random Points.
    b. Name the Output Feature Class (e.g., RandomNonSites).
    c. Specify the number of random points to be generated (should be equal to the number of known sites).
6. **Import downloaded geospatial data.** To import the geospatial data go to File > Add Data > Add Data > Browse the geodatabase and select each of the geospatial data sets to be used.
    a. If the geospatial dataset is not already projected in WGS 1984, the Project tool will be required. Project is located in the ArcToolbox > Data Management Tools > Projections and Transformations > Project.

b. A Project prompt will appear. Click and drag the imported geospatial data set that needs projection into the Input Dataset or Feature Class.
c. The coordinate system used by the data set will appear in the Input Coordinate System.
d. In the Output Dataset or Feature Class, click the folder and browse for the geodatabase, name the file (e.g., ProjectedSlope).
e. Click the button to the right of the Output Coordinate System box. Select Geographic Coordinate Systems > World > WGS 1984 > Press OK > It will bring you back to the Project dialog box, click OK.
f. Repeat for all data sets.

7. **Prep the geospatial data.**
    a. Categorical geospatial data (e.g. land use) need to be separated into individual shapefiles for analysis.
        i. Right click on the geospatial data set > Click Open Attribute Table > Select one category (e.g., residential) and close attribute table.
        ii. Right click on the geospatial data set > Data > Export Data > Export: All records, Output table: = click the folder to browse for previously created geodatabase, name the file (e.g., residential) > Click OK.
        iii. A prompt will appear asking if you want to add the exported data to the map as a layer > Click Yes.
        iv. Repeat for each category in the geospatial data set.
    b. The geospatial data sets must be converted from polygon to raster format for analysis.
        i. ArcToolbox: Conversion Tools > To Raster > Polygon to Raster > select the file next to the Output Raster Dataset box and browse to find the geodatabase > Cell assignment type = Maximum Area > Click OK.
        ii. Repeat for each geospatial data set.
8. **Determine which geospatial variables correspond to known sites and simulated non-sites.** To determine which of the geospatial variables is associated with sites and non-sites, the Extract Multi Values to Points tool must be used. The Extract Multi Values to Points tool will add the value for each geospatial layer at the known body disposal sites to

the site attribute table. In ArcToolbox: Spatial Analyst Tools > Extraction > Extract Multi Values to Points.

a. Click and drag the known site data to the Input point features box.
b. Click and drag the desired geospatial data sets to the Input rasters box.
c. Click OK.
d. Repeat for non-site data.

9. Prep the data for analysis.

a. In ArcToolbox: Conversion Tools > Excel > Table to Excel.
    i. Click and drag the known site data to the Input Table box.
    ii. Click the folder and browse for the desired folder in which to export the table, name the file (e.g., SiteOutput).
    iii. Click OK.
    iv. Repeat for non-site data.
b. Model creation requires four spreadsheets: one will be labelled “TrainingData”, one will be labelled “SiteTrainingData”, one will be labelled “NonSiteTrainingData”, and one will be labelled “TestingData”.
    i. Open the SiteOutput Excel sheet. Add a column to the beginning. Label this first column “SiteType” and label each data entry “Site”.
    ii. Open the NonSiteOutput Excel sheet. Add a column to the beginning. Label this first column “SiteType” and label each data entry “NonSite”.
    iii. Open a new Excel spreadsheet, save it as “TrainingData”. Randomly select 80% of the site and 80% of the non-site data and transfer it to this spreadsheet. This data set will be used to create the model.
    iv. Open a new Excel spreadsheet, save it as “SiteTrainingData”. Take the 80% of sites selected for the training data set and transfer it to this spreadsheet. This data set will be used to determine the summary statistics and create the model.
    v. Open a new Excel spreadsheet, save it as “NonSiteTrainingData”. Take the 80% of non-sites selected for the training data set and transfer it to this spreadsheet. This data set will be used to create the model.

vi. Open a new Excel spreadsheet, save it as "TestingData". Take the remaining 20% of site and 20% of non-site data and transfer it to this spreadsheet. This data set will be used to test the model.

c. Enter the SiteTrainingData into the free statistical program "R" (R Core Team 2015). Enter the command summary(SiteTrainingData) to receive an output summarizing the data set. This will indicate the number of sites linked with each geospatial variable (e.g., number of sites on residential land use, number of sites on agricultural land use, average slope and population density for the data set).

10. Test the significance of site location against non-sites for each geospatial layer. The tables of site and non-site variables must be tested against one another for significance. The statistical test will depend on the nature of the geospatial layer.

a. Continuously distributed data sets. For continuously distributed data sets (e.g., slope, population density), a t-test will be used to test for significance.

i. Enter the SiteTrainingData and NonSiteTrainingData data into R. In R, a Welch's Two-Sample t-test can be performed with the command: t.test(SiteTrainingData$[geospatialdataset], NonSiteTrainingData$[geospatialdataset]). The function will return the t-score, degrees of freedom, and *p*-value. In this instance, the null hypothesis is that there is no difference between the sample means. If $p \leq 0.05$, the null hypothesis can be rejected and the two means are significantly different. This suggests that the geospatial layer has a significant impact on site location aside from random chance.

b. Categorical datasets. For geospatial layers organized by category (e.g. land use), a chi-square test will be used to test for significance.

i. Enter the TrainingData into R. Within R, a chi-square test can be used after the data is arranged in a contingency table. A contingency table is constructed using the table function from the MASS package (Vernables and Ripley 2002). For example, using residential land use the contingency table can be created with the command: tbl<-table(TrainingData$Residential, TrainingData$SiteType). The output is a

contingency table listing the frequencies of each of the residential types (e.g., present, absent) for the sites and non-sites.

ii. The chi-square test can be performed within the base R using the command: chisq.test(tbl). The output of the command will contain the χ^2 value, degrees of freedom, and associated *p*-value. Again, a *p*-value of 0.05 will be used to determine significance.

11. **Remove the geospatial layers that do not significantly contribute to known site location.** To remove the non-significant layers, within ArcMap, right click on the non-significant later and select Remove.

12. **Reclassify geospatial layer values based on significance results and summary statistics.** The weighted map-layering approach requires that geospatial map layers be weighted based on theory and knowledge of site relationship to the variables. Each map layer is assigned a value given the summary statistics of the data set and that value is weighted based on theory. Categorical variables are already prepared to be reclassified, but continuous variables need to be grouped into categories. Right click on the continuous geospatial data set > Properties > Symbology > Under Show: click Classified > Select the number of classes > Click Classify… and enter desired values in Break Values > Click OK > Click OK.

    a. **Reclassify values based on results.** The reclassified values will be determined by the summary statistics obtained in Step 9c. For example, if 70% of sites occur within 0.0 – 1.0°, this slope category will receive a value of 7. If 20% of sites occur within 1.1 – 3.0°, this slope category will receive a value of 2. If the remaining 10% of sites occur between 3.1 – 10.0°, this slope category will receive a value of 1. If no slopes occur above 10.1°, a larger value, such as 7, will be assigned to reflect that no sites occur here.

    b. **Reclassify tool.** To reclassify the raster values into assigned values, go to: ArcToolbox > Spatial Analyst Tools > Reclass > Reclassify > drag geospatial layer to Input Raster > set Reclass field to “Value” > enter the desired values > Set Output raster field to geodatabase, name accordingly (e.g. SlopeReclass) > Click OK.

    c. **Repeat for each geospatial layer.**

13. **Add reclassified raster together to create weighted map-layer predictive model.** Once each geospatial layer has been reclassified with the appropriate values, the weight for each variable is determined based on theory in criminology and forensic anthropology. For example, research indicates that distance to road significantly influences body disposal location (Congram 2010; Rossmo 2000). For this reason, distance to primary and secondary roads is assigned a weight of 9. Once the weights are decided, the composite prediction map will be created by multiplying each variable value by its weight and adding the resulting values together.
    a. Go to: ArcToolbox > Spatial Analyst Tool > Overlay > Weighted Sum > drag geospatial layers to Input rasters > set weight for each raster > click on the folder to the right of the Output raster box and browse for geodatabase, name file accordingly (e.g. DeductiveModel).
14. **Use Extract Value to Point to link model value to sites and non-sites.** Determine the model value for each site location and non-site location. Use the Table to Excel tool for each to export table to an Excel spreadsheet.
15. **Test the significance of the model to the randomly generated non-site data.** The model must be tested for independence from randomness.
    a. Within R, a Mann-Whitney U test can be performed through the base R package with the command: wilcox.test(Site$ModelValue, NonSite$ModelValue). Again, the *p*-value of 0.05 will be used to determine significance.
16. **Determine optimal sectioning point.** The optimal sectioning point is arbitrarily chosen with the weighted map-layer method. A recommended method of choosing the optimal sectioning point is to split the model range values into categories (e.g., create ten different categories), and create a table linking these categories with the percentage of area designated high probability, the cumulative % of area designated high probability, the number of sites correctly predicted, the percentage of sites correctly predicted, and the cumulative percentage of sites correctly predicted (see Table 3.9 for an example). The goal of the sectioning point is to pick a division with a small region of high probability that has a high percentage of correctly predicted sites.
17. **Determine Gain statistic.** Kvamme's Gain statistic can be calculated for the weighted map-layer model. Create another table like the one in Step 16. Include the same model

ranges, but only use the testing site data to calculate the percentage of sites correctly predicted. Using the sectioning point determined in Step 16, identify the percentage of the study area that is high probability and the percentage of sites correctly predicted in that area. Kvamme's Gain statistic is calculated as 1 – (% high probability area / % sites correctly predicted as high probability).

VITA

Annemarie Gundel attended Fordham University, College at Lincoln Center, in New York, New York for her undergraduate degree. She obtained her Bachelor of Arts degree in Art History and Anthropology in 2013. Annemarie attended field school in Alaska through Adelphi University. After graduating, she interned at the Klondike Gold Rush National Historic Park and conducted an archaeological survey of the area surrounding the Chilkoot Trail. Annemarie moved to Knoxville, Tennessee to pursue her Master of Arts degree in Biological Anthropology at the University of Tennessee. She accepted a two-year graduate research assistantship with Dr. Amy Mundorff in the Department of Anthropology, working with plant isotope analyses. Following this, she accepted a year-long graduate research assistantship with Dr. Michael Kenyhercz in the Department of Anthropology, scoring human decomposition rates. Annemarie graduated with a Master of Arts in Anthropology in May 2017.