|  |  |
|---|---|
| **REPORT DOCUMENTATION PAGE** | *Form Approved*<br>*OMB No. 0704-0188* |

|  |  |  |
|---|---|---|
| **1. REPORT DATE** *(DD-MM-YYYY)*<br>14-11-2006 | **2. REPORT TYPE**<br>THESIS | **3. DATES COVERED** *(From - To)* |

**4. TITLE AND SUBTITLE**
AIR FORCE RECRUITMENT: A GEOGRAPHIC PERSPECTIVE.

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
CAPT ROSS JASON J

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
ARIZONA STATE UNIVERSITY

**8. PERFORMING ORGANIZATION REPORT NUMBER**
CI07-0004

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
THE DEPARTMENT OF THE AIR FORCE
AFIT/ENEL, BLDG 16
2275 D STREET
WPAFB OH 45433

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

**15. SUBJECT TERMS**

| **16. SECURITY CLASSIFICATION OF:** |  |  | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** |  | 112 | **19b. TELEPHONE NUMBER** *(Include area code)* |

# AIR FORCE RECRUITMENT: A GEOGRAPHIC PERSPECTIVE

by

Jason J. Ross

A Thesis Presented in Partial fulfillment
of the Requirements for the Degree
Master of Arts

Arizona State University

November 2000

# ABSTRACT

Beginning in 1973 the Armed Forces implemented an All Volunteer Force recruitment policy. Since then, the Armed Forces have relied heavily upon propensity studies in order to make recruitment policy decisions. By aligning with the adage "the best predictor of the future is the past", this study used past recruit's home addresses in order to develop models designed to predict areas of recruitment instead of relying upon propensity studies. Descriptive and inferential statistics were used to create and evaluate both non-spatial and spatial auto correlated models to determine the best method for predicting recruitment. Ultimately, the research conclusively found that different areas of the country are inclined to recruitment; suggesting the use of statistical measures based on Home of Record information instead of propensity studies is a better method for predicting recruitment.

AIR FORCE RECRUITMENT: A GEOGRAPHIC PERSPECTIVE

by

Jason J. Ross

has been approved

November 2006

APPROVED:

______________________________, Chair

______________________________

______________________________

Supervisory Committee

ACCEPTED:

______________________________

Director of the School

______________________________

Dean, Division of Graduate Studies

# DEDICATION

For my God, my family, and my country.

The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

# ACKNOWLEDGEMENTS

I thank my family for the opportunity to begin and the support to complete this program. I thank my committee for the advice and the persistence. To the people within the School of Geographic Sciences, together we laughed, together we complained, and together we transgressed the graduate experience, I will forever remember you as, "friends".

## TABLE OF CONTENTS

# LIST OF TABLES

## LIST OF FIGURES

Figure Page

Figure Page

# CHAPTER 1 - INTRODUCTION

## Overview

As the United States and the Air Force struggle to pay for the Global War on Terrorism, new equipment, rising health care costs, and retirement pensions, efforts are being made to transform the Air Force into a more efficient organization. This transformation effort has led to force shaping initiatives (downsizing), Base Realignment and Closings, and also organizational embracement of Sigma Six techniques in order to reduce waste. With waste reduction in mind, today's Air Force leaders could use an accurate model to predict where people are joining its ranks, allowing them to better utilize their resources.

The end of the draft in 1973 signaled the beginning of the Department of Defense's (DoD) sole reliance on an All-Volunteer Force (AVF) to fill its ranks. Since then, military leaders and Congress have continually questioned the quality and quantity of new recruits, as the need for high-quality recruits is essential in order to maintain the military's high-tech weaponry and support equipment. For many, leaders have viewed the AVF as a success because every year since its inception the armed forces have met their recruitment goal, except for 1979 and 1999 (Fernandez, 1987). However, what is often overlooked is an overall steadily decline in enlistment propensity (a person's likelihood of enlistment) for the last thirty years.

Youth propensity studies have been the mainstay for recruitment prediction since the inception of the AVF. These studies focus on individual's socioeconomic status, gender, and race. In terms of geography, these studies generally group propensity trends into large regions of the United States but fail to look any closer than the state level. It is this arena that this study aims to improve by defining areas where strong recruitment aptitude exists instead of globalizing the geographic characteristics of recruitment propensity. By defining which geographic areas are promote recruitment, this study will attempt to accomplish what the propensity studies can not: predict areas where recruitment is most likely to occur.

To facilitate this endeavor, I plan to utilize the previous propensity studies' findings as a foundation and build upon it by adding "distance" to prediction methods. Using a wide variety of statistical techniques, this study seeks to enrich propensity studies by adding a geographic component to enlistment trends. Through the use of Geographical Information Systems technology these demographic patterns will be graphically displayed and analyzed to identify any social-spatial relationships that exist. By providing new methodologies to analyze the social-economic geography of past recruitment, the ultimate goal of this study is to create an accurate predictive model for future recruitment efforts.

## Research Question and Hypothesis

By looking beyond the propensity studies' socioeconomic recruitment factors, a thesis written by Jackson (1999) found that nearly half of all military recruits had a parent who was a veteran of the armed forces. Furthermore, I assumed that active-duty military and veterans have a tendency to live in close proximity to military installations, which serve as a place for employment, medical treatment, lower food costs, and a sense of home. Therefore, by using military installations as a static point of reference for high concentrations of active-duty and veteran personnel and assuming the installation's "military" influence decreases as distance increases, maybe it is possible to spatially predict enlistment. Hence, the overarching question for my thesis is summed up as, "Is distance to a military installation a better predictor of enlistment than traditional methods?"

My hypothesis is that using distance to military installations as a "grouping method" will prove to be a better predictor of enlistment than using the traditional political boundary grouping method. This study will set out trying to reject the null hypothesis, which is, "Distance to a military installation is not a significant predictor of recruitment."

## Geographic Properties, Study Area, Data Acquisition and Methods

This section outlines the study's statistical sources and defines the geographic properties, study area, data, and methods used to test the research question.

### Geographic Properties

In order to execute many of the statistical measures in this study, the use of geographic information systems was needed. The principal software used in this study was Environmental Systems Research Institute (ESRI) ArcGIS 9.1. A template was created using the North American 1983 Geographic Coordinate System datum along with the USA Contiguous Equidistance Conic projection. Standard parallels for the projection are located on the 33rd and the 45th parallels. Furthermore, the template's map units were set to "meters", thus, all distance measurements and geostatistical calculations were generated using these geographic properties.

### Study Area

Much concern was given to the study's spatial extent. Ultimately, the decision was made to focus on the contiguous United States (CONUS). Puerto Rico, Alaska, and Hawaii were considered but deemed inappropriate because of their geographic features. The islands were dismissed because anyone who joined the service from those locations would live near a military installation due to lack of these land mass. Alaska was dismissed for the opposite reason as it contains vast ranges of uninhabited areas. In short, because spatial proximity is a central component to this study, these territories needed to be removed from the study area else, their geographic properties were believed to skew the distance results.

In order to test the research question, CONUS was subdivided into two main datasets. The first dataset, labeled "counties", is comprised of the CONUS' county political boundaries (obtained from the Bureau of Transportation Statistics). This dataset was further subdivided into its urban and rural segments using a Bureau of Census Urbanized Areas shapefile. The resulting dataset contained 3,904 areal units representing the urban and rural segments of each county within CONUS (See Figure 1). The second dataset, labeled "buffers", divided CONUS into equidistance buffers surrounding military installations. This was created by first obtaining a shapefile from the Bureau of Transportation Statistics that contained all military installations currently being managed by the

Department of Interior. This shapefile was then compared to a list of major military installations found on the DoD's "Sites" website. The shapefile was edited to contain only those installations that were included on the DoD's list. Then, buffers were created at ten kilometer distances surrounding these military installations until the entire CONUS area was covered. Lastly, this buffer shapefile was subdivided once more using the same Bureau of Census Urbanized Areas shapefile used by the counties group. The resulting buffer shapefile

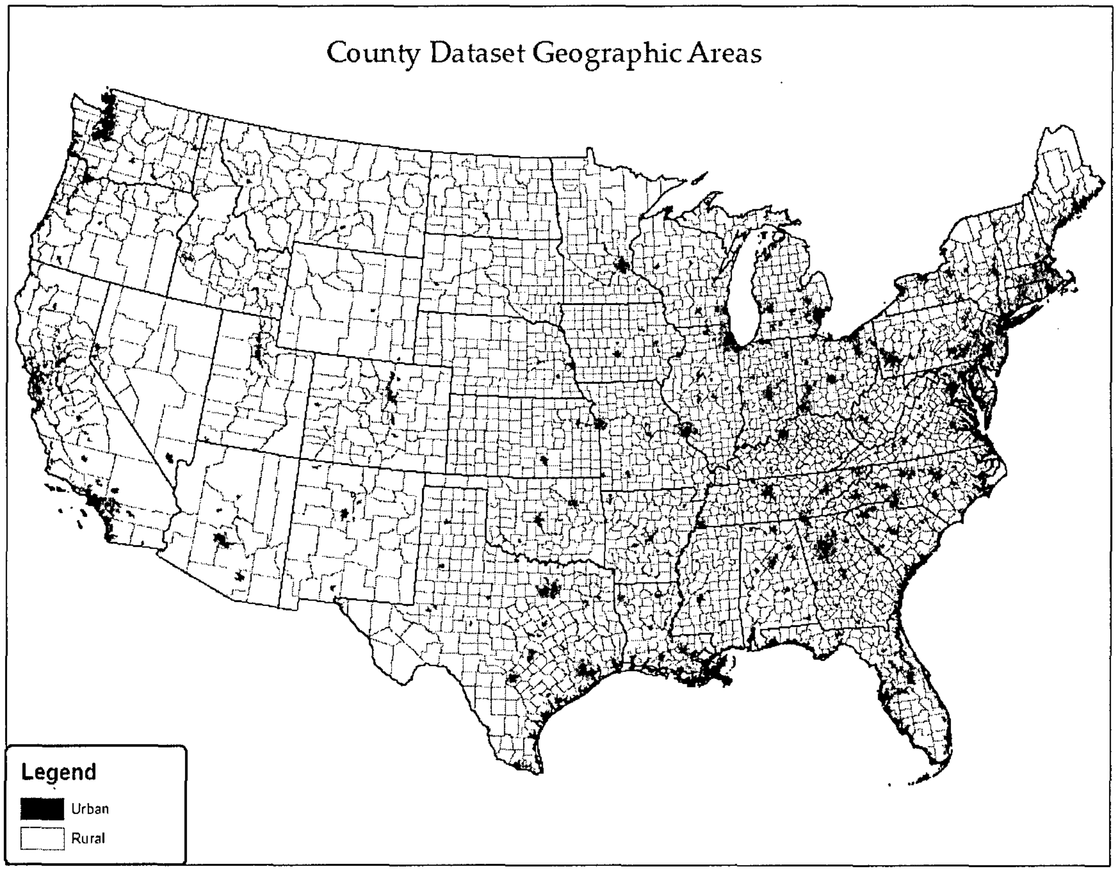


**Figure 1:** County Urban and Rural Extent

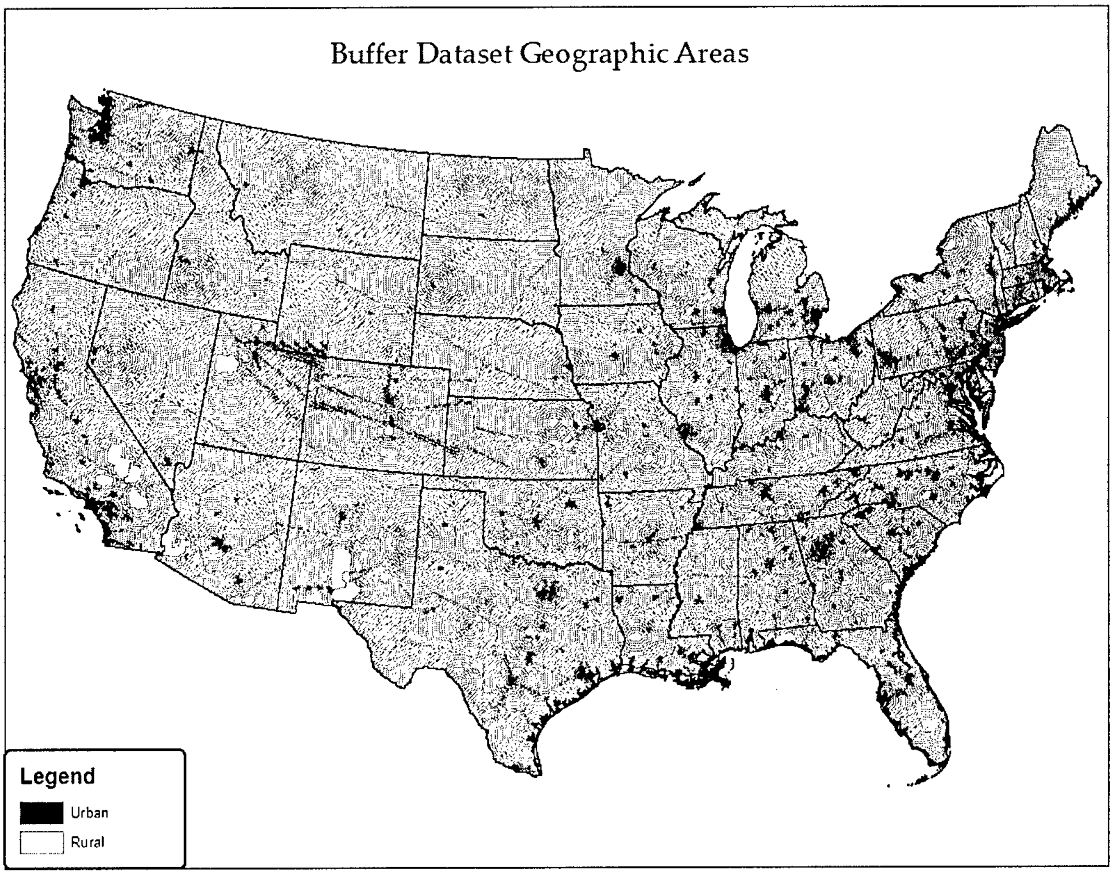


**Figure 2:** Buffer Urban and Rural Extent

contained 4,862 areal units, which divided CONUS into urban and rural ten-kilometer buffers (See Figure 2). A more detailed description of the creation of the buffer shapefile is included in the Data Manipulation chapter.

### Data Acquisition

The study's primary data source was obtained from the Air Force Recruitment Services. This contained "Home of Record" information (the address where a recruit lived at the time of enlistment) for every recruit that enlisted for Fiscal Years 2002 (HoR2002), 2003 (HoR2003), and 2004 (HoR2004). A fiscal year (FY) spans from October 1 through September 30. The Home of Record information is vital in that it illustrates an event that occurred in space and time, which is where someone was recruited.

Obtaining the demographic data from the Air Force Recruitment Services for each recruit would have been ideal for this study; however, due to privacy concerns this was not a plausible option. Therefore, the next-best source for demographic data was the Census 2000 SF3 dataset. The census demographic data were used to replicate the environment that promoted these events to occur. The crux of this study is to analyze these events and the environment surrounding them to determine if there is an underlying process that caused them to happen. If a process is identified then predicting future events is possible.

## Methods

Building upon the propensity studies research and using the framework within GIS, the methods employed will serve to construct spatial statistical models of socio-economic conditions in CONUS using the HoR2002 and HoR2003 data. These methods will be combined into two separate groups. The first group's methods assume no spatial autocorrelation exists within the data and the second group's methods assume spatial autocorrelation. In general, spatial autocorrelation is a numerical description that illustrates the inter-dependency among variables another across a spatial plane.

For the first section, Statistical Package for Social Sciences version 13.0 (SPSS) software will be used to evaluate the data using multiple linear regression, principal component analysis (PCA), and then another multiple linear regression using the results of the PCA. For the second group's methods will include point pattern analysis techniques using Nearest Neighbor, Moran's I and Getis' Ord GI, an ordinary kriging method, and Geographic Weighted Regression. The point pattern analysis and kriging techniques will utilize ESRI's ArcGIS geostatistical software whereas the Geographic Weighted Regression will be conducted through Geographic Weighted Regression 3.0 software. The rationale behind using these six techniques is that the first group's methods can easily be found within social science research, indicating a general acceptance for these techniques. Furthermore, the second group's methods have taken hold in geography research but are still quiet sparse in the social science arena. Therefore, as a geographer I want to ensure that cutting edge spatial statistical techniques are compared against traditional social science methods in order to answer the question at hand.

Each of these methods will be conducted on both the control group (Counties) and the experimental group (Buffers) and the results will be compared and contrasted to each other to determine if the experimental group has better predictability than the control group.

### Thesis Outline

In seeking to disprove the null hypothesis, this thesis is organized as follows. In Chapter 2, I review past literature in order to determine explore past works in recruitment research, distance decay theory, and geographic reference within propensity studies which will provide a roadmap for my research. In Chapter 3, I provide a detailed account of data acquisition and manipulation so that the study can be sufficiently validated. Chapter 4 contains the first portion of the data analysis which investigates the data using aspatial predictive methods which are: Multiple linear regression, a common statistical technique utilized for prediction which will also serve as a baseline by which all other methods will be compared; Principal Component Analysis (PCA), which minimizes the correlation between the independent variables; Multiple Linear Regression using the PCA results. Chapter 5 dissects the data with these spatial predictive methods: Point pattern analysis using Nearest Neighbor, Moran's I, and Getis' G, designed to locate geographic hotspots and cold spots; kriging; and Geographic Weighted Regression.

Geographic weighted regression, kriging and point pattern analysis could prove to be a better prediction device than the multiple linear regression models if spatial autocorrelation is found to be intrinsic within the data (See Figure 3).

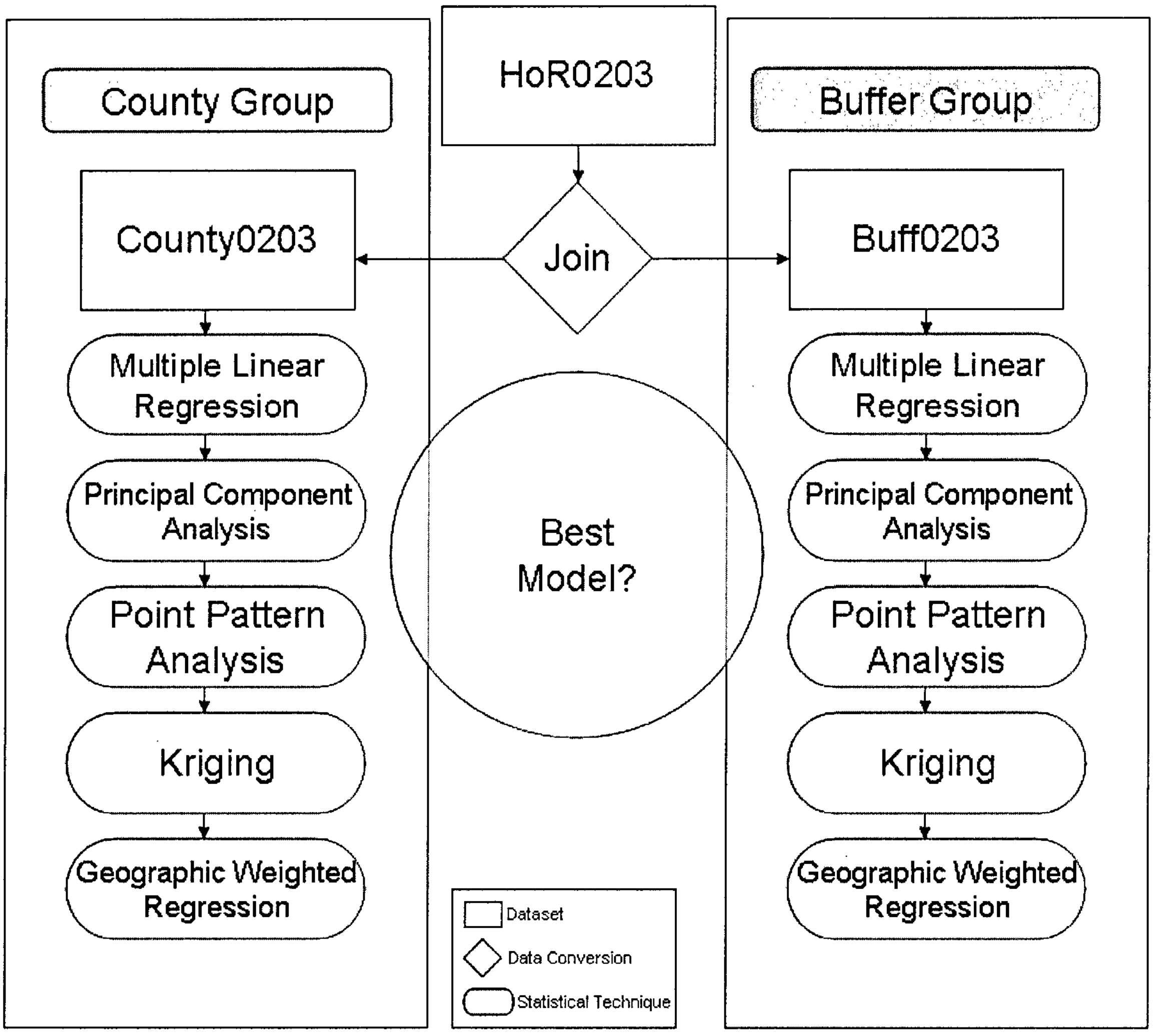

County Group
HoR0203
Buffer Group
County0203
Join
Buff0203
Multiple Linear Regression
Principal Component Analysis
Point Pattern Analysis
Kriging
Geographic Weighted Regression
Best Model?
Multiple Linear Regression
Principal Component Analysis
Point Pattern Analysis
Kriging
Geographic Weighted Regression
Dataset
Data Conversion
Statistical Technique

**Figure 3:** Methodology Flow Chart

Testing the chapter 4 and 5 models using the HoR2004 dataset will be the focal point for chapter 6 and ultimately decide which model best predicts recruitment. Lastly, Chapter 7 will serve as the conclusion chapter which will highlight the research's strengths, weaknesses, and suggestions for further research.

## Significance

Currently, the Air Force has implemented many different measures to entice people to join its ranks which range from enlistment bonuses to accelerated promotions, costing the Air Force millions of dollars annually. If this study's findings can be used to focus the limited Air Force recruitment monies into those geographic areas that are likely to produce recruits, then it is believed the Air Force would receive a greater return on their investment.

This study will benefit society at large in three main ways. First, it will give military recruiters a tool which will help them identify those areas which they should focus their attention. Next, it will benefit the military at large as it will help ensure that tomorrow's force will continue to be filled by high-quality recruits and not fall victim to gaps in recruitment numbers. Lastly, it will benefit the American taxpayer, as the monies used by the DoD will be spent more wisely.

# CHAPTER 2 – LITERATURE REVIEW

This chapter consists of four sections. The first section will briefly describe recruitment history and the second section will address the attributes and characteristics of distance decay theory. The third section will illustrate how previous propensity studies have depicted geographic variation and the last section will concentrate on the various demographic patterns that propensity studies have cited as recruitment predictors.

## Recruitment History

Malownowski (2005) expressed that, "failure to understand the [recruitment] past may cloud our judgment as recruiters view the future" (p. 349). The characteristics surrounding the United States' recruitment policies have varied through the course of American history with the two most significant strategies being conscription and volunteerism. Prior to World War II, the United States armed forces sustained a relatively small active force with the majority being in the Navy. During times of crisis, the United States called upon volunteers and also initiated the draft, or conscription services to fill its war-time needs. However, at the conclusion of World War II the small-active-force philosophy changed significantly as the United States maintained a large garrison force designed to deter aggression. During this period, conscription and volunteerism were both viable means of joining the service. This recruitment philosophy changed near the end of the Vietnam War era when the draft ceased and the AVF policy was instituted.

To ensure the armed forces continued to maintain its supply of new recruits, the DoD developed the Youth Attitude and Tracking Survey (YATS) which was given a survey to nearly 10,000 16-24 year olds between the years of 1975-1999 (Malinowski, 2005). This was the beginning of the propensity era, where the United States armed forces became concerned about a person's likelihood of serving on active duty. Due of its long history, Malinowski claims the YATS data was given considerable weight in the recruitment arena (Malinowski, 2005, p. 353), which explains why many recruitment research utilize these data.

In addition to the DoD's YATS data, the Monitoring the Future (MtF) project also serves as a resource for recruitment research (www.monitoringthefuture.org). Currently in its 32nd concurrent year of service, the MfF project is conducted by the University of Michigan annually and provides a source for social scientists studying the behavior and attitudes of today's youth. The MtF survey has been given to approximately 20,000 teens each year. In reference to military service, they are asked if they "*'definitely will', 'probably will', 'probably won't', 'definitely won't' serve in the military*" (Malinowski, 2005, p. 356).

Since the inception of the AVF, there have been two instances where the armed forces failed to meet their recruitment objectives. The first, and most severe, occurred in 1979 where the steadily deterioration of recruiting and high-quality recruits threatened AVF policies (Asch 2005). To curb the deterioration effects, Congress passed dramatic laws which significantly increased the military base pay in Fiscal Years 1980 and in 1981 and the Montgomery G.I. Bill was enacted, providing enhanced educational opportunities to service personnel. These drastic measures reversed the 1979 recruitment deterioration and allowed the armed forces met their recruitment quotas for another twenty years. In 1999, a less-severe recruitment crisis arose when the Air Force and Army failed to meet their goals. This crisis was cut short when Congress passed the Fiscal Year 2000 National Defense Authorization Act, which increased military pay, increased bonus ceilings, reformed military retirement options, and raised special incentive pay (Asch 2005). With the AVF being the sole source for military accession for over thirty years, it continues to be the mainstay of United States recruitment policy and with continual youth propensity monitoring, leaders are capable of adjusting incentives to meet current and future recruitment needs.

## Distance Decay

> *"Under the umbrella of spatial interaction and distance-decay, it has been possible to accommodate most model work in transportation, migration, commuting, and diffusion, as well as significant aspects of location theory."* (Olsson 1970, p. 223).

As the above quote illustrates, distance decay has played a substantial role in the fields of urban and economic geography in addition to other fields outside the discipline, like biology and economics; nevertheless, in the realm of social science, distance is typified as another variable along with more important indicators such as education, employment or age (Goodchild, 2004).

The Distance Decay theory originated from social physics in the 1960s and grew out of the gravity model developed by Stewart, Warntz, and Zipf (Eldridge and Jones, 1991). Stewart, when studying the geographic displacement of recent college undergraduates found that "the number of undergraduates or alumni of a given college who reside in a given area is directly proportional to the total population of that area and inversely proportional to the distance from the college" (Stewart, 1941, p. 89). He also states that his examination uncovered a remarkable "inverse distance 'law' or statistical regularity" (Stewart, 1941, p.89). In general, what Stewart discovered was that the relationship between an item and its origin decreases exponentially as distance increases. Furthermore, the affect that the item studied has on other items will be felt exponentially smaller the further one moves away from the origin. This statistical regularity became the foundation for Tobler's "First Law of Geography" which states, "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970, p. 236).

However, in the last twenty years, Tobler's law has come under some scrutiny. Critiques of the law state that distance is a relative term because it possesses a friction characteristic (Eldridge and Jones 1991). For example, the time to traverse ten miles over rough terrain will take longer than it would to traverse ten miles over smooth terrain. Therefore, Eldridge and Jones (1991) argue that equal distance intervals have spatially uneven effects on interaction and that distance, or the space between objects, is "warped". This effect is what they term as a "Warped Space" and conclude that distance decay is more of a contextual concept than a universal phenomenon.

Nevertheless, the effects of warped space can be counteracted by taking into consideration the spatial structure contained within the data. Fotheringham illustrates that the warped space happens because most models give a global view of the distance-decay parameter instead of focusing on the local (Fotheringham 1981). Hence, the development of new spatial modeling techniques occurred through the 1980s, which ultimately produced statistical software that focused on the local parameters instead of the global parameters (Fotheringham 2002). By focusing on the local parameter, the effects of warped space are minimized.

## Geographic Variation within propensity studies

To begin with, Malinowski addresses the overall attitude of geographic variation by attributing the limited geographic information contained within the YATS and MtF database (Malinowski, 2005). Both datasets break the country into four different sections: Northeast, North Central, South, and West. Since these are the main data sources for the majority of recruitment research, when research touches any aspect of geographic inquiry, it continually mirrors the above geographic regions.

Malinowski (2005) also addresses what appears to be a common theme in the literature, in that the South has maintained higher rates of military service when compared to other parts of the country. Segal and Segal (2004) validate this claim by arguing that nearly 40 percent of enlistees came from the South. However, Segal and Segal also note that the South has the highest recruit-eligible age population in the United States, which could account for the large percentage of enlistees. They also acknowledge that the western states, Nevada, Idaho, and Montana have the highest number of recruits when data are normalized by the recruit-eligible age.

Malinowski (2005) further points out that propensity attitudes within the different geographic areas typically move in unison, but at times move in opposite directions. He further makes the statement that these differences in propensity attitudes are crucial and that national recruitment policy should adjust to reflect the attitudes and values for the different geographic regions.

Perhaps the most in-depth look at geographic recruitment research is found in a Master of Arts thesis conducted by Paul Anthony Cannizzo in 1995. Cannizzo researched Air Force recruitment within the state of Ohio and utilized Geographic Information Systems technology to analyze these patterns (Cannizzo 1995). Cannizzo's methodology included geocoding enlistee's Home of Record and associated demographic information, which was obtained from the Air Force Recruiting Services, down to the county and census tract level. His research concluded that his study area displayed substantial spatial correlations for the locations of new recruits but his techniques were unable to identify a common predictive thread.

Cannizzo (1995) also had difficulty finding research that contained a significant geographic component to recruitment studies; however, he did cite Dandeker and Strachan (1993, p. 291), which called for more research in the area:

> ...spatial analysis research could produce detailed information on the social characteristics of the army's target populations for recruitment; the social/geographical settings in which they are located; and the ways in which these patterns cluster with other variables, especially those facilities/activities that are under the control of the army or subject to its influence. It would then be possible to indicate where recruitment resource inputs would be most likely to produce a reasonable return, and to test the reliability of these indicators by analyzing subsequent recruitment data.

A further review of Dandeker and Strachan's (1993, p.291) work reveals another critical component to their argument:

> The home addresses of those applying to join the army are important in understanding the recruitment process in three respects. First, they provide an objective measure of differences in absolute and relative application rates between localities. Second, the resultant patterns can be used to 'fine-tune' the targeting of recruitment budgets in order to maximize the flow of applicants. Finally, the incidence of applications can be related to the presence

> of service installations and military events, publicity, and school visits.

The bottom-line of Dandeker and Strachan's (1993) claim is that the location of military installations plays a pivotal role in advertising the military way of life to those that live nearby and that the use of home addresses gives a researcher the ability to isolate clusters and patterns within the data.

In defense of Dandeker and Strachan's (1993) theory of military installation proximity, Segal and Segal (2004) found that locations with a large military establishment affected how the community interacted with one another, which was notably different than areas without a military influence. They also found that in military installation areas, the population was among the least racially segregated areas in terms of population and employment. It is this localized social geography that is believed to promote military service to those who encounter its presence.

## Recruitment Demographic Patterns

In this section, I examine the recruitment demographic characteristics that exist throughout the recruitment literature. These demographic characteristics are derived from the MtF Project and the YATS data. The following subsections review the income, race, and military service of young adults.

### Income

In the late 1990's, the media claimed the strong economy was the source for the armed forces' recruitment woes (Malinowski, 2005). Nevertheless, Malinowski (2005) argues that these claims error because they only consider national statistics instead of local variations. He concludes although low unemployment rates may be related to recruitment rates but that overall, low unemployment rates are less important than other factors (Malinowski, 2005). Malinowski also indicated that per capita incomes had little predictive value for recruitment (Malinowski and Brockhaus 1999). On the other hand, in regards to macroeconomic opportunities (job availability or unemployment) a study conducted by the General Accounting Office in 2004 confirmed that when civilian unemployment rates are high, military enlistments increase (GAO 2005).

In addition to income attributes, education levels have a direct effect on recruitment. Segal and Segal (2004, p.10) state that, "enlistment is also predicted by parents' education" and cite that children of college education parents are less likely to serve in the military. Bachman et al. (2000) suggested that the general pattern is that the higher the parental education the lower level of military propensity.

Lastly, prior to the implementation of the AVF the draft was thought of as a social equalizer where every citizen had an equal opportunity to serve in the military. The concept of "equalization through a draft" argument is often questioned and many wonder if the poor are bearing the burden of military service. Kane (2005), in his demographic study of recruits before and after the September 11, 2001 terrorist attacks found that those in the middle class are more likely to join the military than those who are wealthy or those who are poor. Segal (2004) claims the poor either have health concerns or criminal pasts whereas the wealthy have more opportunities available to them; thus, making military service typically not an option. Kane's (2005) efforts did express that in the two years following the terrorist attacks on the World Trade Center, the greatest increase in recruitment levels came from those with higher incomes (Kane 2005).

### Race

To begin this category, research on recruitment has focused on the U.S. Census defined categories of race. These categories include: White alone, Black or African American alone, American Indian and Alaska Native alone, Asian alone, Native Hawaiian and Other Pacific Islander alone, Some other race alone, Two or More races, Hispanic or Latino White alone, Hispanic or Latino Black or African American alone, Hispanic or Latino American Indian and Alaska native alone, Hispanic or Latino Asian alone, Hispanic or Latino Native Hawaiian and Other Pacific Islander alone, Hispanic or Latino Some other race alone, Hispanic or Latino Two or more races. Malinowski (2005) illustrates that race is an important propensity determinant and shows that Blacks and Hispanics have a greater propensity than Whites (Malinowski, 2005). In a report detailing social-economic factors of today's military, Segal and Segal (2004) illustrate varying degrees of over- and under-representation of different races found in today's military ranks. They concur with Malinowski that the enlisted ethnic and racial population within the armed forces generally does not mirror society at large. Also, there exists a significant gender difference between military and civilian populations according to internal governmental reports ((DMDC (2005), GAO (2005)).

When comparing armed forces recruits before and after the September 11th terrorist attacks, Kane (2005) gives an overview of the general racial recruitment variation. In this report, he found that American Indian/Native Alaskans, Native

Hawaiians/Pacific Islanders, Blacks, and the Two or more races categories possessed an over-representation within the military. Those races that were under-represented were the Whites, Hispanics, which were slightly under-represented, and Asians, which were drastically below the national average (Kane 2005).

With the Hispanic population increasing its share of the total population each year, the armed forces are concerned that they are not being adequately represented within its ranks (Asch et al., 2005). However, they also report that Hispanic accessions have increased over the last decade (1994-2004) but their population still remains lower than the national average (Asch et al., 2005). Therefore, the armed forces have begun a stronger marketing campaign to attract Hispanic youth, such as the Army running commercials in Spanish (Asch et al., 2005).

### Military Service

In addition to regional differences, veteran status of one's parents is also a predictor of recruitment. Jackson (1999) found that 45 percent of Air Force enlistees had parents who had served in the military, and of these, 47 percent had served in the Army (Figures 3 and 4). She further illustrates that the 73 percent of respondents indicated that an enlistee's major influence in his/her decision to enlist was his/her parents (Jackson 1999).

In an effort to try to explain if prior military exposure due to a person's family serving in the military had a positive affect on their propensity to enlist, Brown (2005), through the use of Generalize Exchange Theory, found that people who had high military exposure were more likely to enlist than those with lower military exposure. Generalize Exchange Theory basically states that people seek to reciprocate those who benefit them; meaning, that people feel obligated to give back to an organization in which they have received something. In reference to military enlistees, Brown (2005) suggests that military exposure by an individual's parents serves as the vehicle in which an individual may feel obligated to join the service since he/she has received a benefit from the military as a result of his/her parents' service. Therefore as a predictor for recruitment, the location of veterans, active-duty military personnel and military installations serve as an exposure vehicle, which could be used to entice people to join its ranks.

# CHAPTER 3 – DATA ACQUISITION AND MANIPULATION

While a successful study is traditionally measured by how well the analysis supports (or does not support) the hypotheses, equally important is given to how well the study documents its methods. Accurate documentation is essential to scientific research because it validates the results and provides a clear roadmap for future research. Thus, a description of data acquisition and manipulation is required.

## Home of Record information and shapefile creation

The first data were obtained from the Air Force Recruiting Service which contained a list of Home of Record addresses for all Air Force recruits who enlisted through Fiscal Years (FY) 2002, 2003, and 2004, comprising of over one hundred thousand records. A Fiscal Year spans from October 1 through September 30 and a Home of Record address is where the recruit lived at the time of his/her enlistment. These records were received in Microsoft Excel format with three columns indicating the date entered service, street address, and city/state/zip variables. The Excel file was exported to a comma-separated text document which allowed for the city/state/zip variable to be manipulated and separated by commas. After all variables were separated by commas, the file was then exported back into a Microsoft Excel format which formed the foundation for data transformation to other software packages.

Every address outside the CONUS was excluded from the list, leaving 84,974 records which were then geocoded (assigned an X/Y coordinate) into Environmental Systems Research Institute's (ESRI) ArcGIS 9.1 software using ESRI's 2003 Streetmap USA for the reference layer, an enhanced TIGER 2000 streets dataset. The resulting confidence score for the geocoding processing yielded a 94.35 average. Only 6.2 percent of the addresses yielded a score between 60 and 80, with 80 being defined by ESRI as a "good match". None of the addresses fell below a score of 60, which was set as the minimum threshold for the geocoding processing. A distribution pattern analysis was conducted of those addresses with matches below 80 and revealed a random pattern. Therefore, no efforts were made to enhance the addresses' location in order to improve the resulting geocoding score nor were the addresses removed from the sample, as all original scores were at or above the minimum reliability level. These data established the foundation for a shapefile that included FY2002 and FY2003 Home of Record (HoR0203) addresses and another shapefile that included FY2004 (HoR04) addresses.

## Census information and shapefile creation

The 2000 Census SF3 data were used to represent the demographic characteristics of the areal units in this study because these data include veteran and active-duty military information. This vital information was reportable down to the block level; however, due to this study's extent, the block level was determined to be too refined. Therefore, urban and rural county information was determined to be most appropriate for this study instead of census block information for area descriptions.

In order to extract the applicable information from the 2000 Census, the entire SF3 data were downloaded from the census website. Using the index table, the appropriate fields were identified which reflected the county's race populations, age categories, income levels, education level, and military service. In all, 39 different variables were extracted from the data. Then, all population variables were normalized by dividing each variable's population number by its associated area's total population, leaving a number less than one.

In reference to the race variables, the census reflects 8 separate categories: White, Black, Indian/Eskimo, Asian, Hawaiian/Pacific Islander, Latin, Two or more races, and Other. However, the recruitment literature (Bachman et al. 2000) categorizes the racial classes into four: White, Black, Hispanic, and Other; and

Cannizzo (1995) and Melroy (1999) added an Asian category which included Hawaiian Islanders. This study will use five categories to represent the race population of the areal units: White, Black, Asian/Hawaiian Islanders (labeled "AHP"), Latin, and Other.

The Census broke down the ages of the area's recruit-eligible population into six categories: 18, 19, 20, 21, 22-24, and 25-29. I acknowledge that it is possible for someone to join the military at the age of 17; however, this is atypical and was not considered a significant age group to include. The recruitment literature showed the principle target age for military recruitment is 18-24 (Fricker 2003 and Segal and Segal 2004); hence, this study combined the 18, 19, 20, 21, and 22-24 variables into a recruitment age category labeled "Raw1824" where it was left in its raw count. Additionally, another variable was created that normalized the Raw1824 variable to reflect the percent of the age group from the county's total population, which was labeled "1824".

The remaining variables that were used in this study include active-duty military population (labeled "AD"), veteran population (labeled "Vets"), median household income (labeled "HHI"), unemployment rate (labeled "UnEmp"), PerCapita (labeled "PerCap"), and percent of population with no high school diploma (Labeled "NoHS"). The NoHS variable represents those individuals 25 years of age and older who do not have a high school diploma.

The construction of the shapefiles introduced in the overview section originated from two Bureau of Transportation Statistics files. The first shapefile consisted of the United States' counties and the other shapefile contained the 2000 Census Urbanized Areas. These two shapefiles were merged together to create a single shapefile that had areal units representing the United States urban and rural county elements. In the event a county contained more than one urban segment, the attribute table was analyzed and all urban elements inside a single county were merged together to create a single urban element, resulting in 3,904 areal units.

With the county urban and rural shapefile created, the census information was integrated into the shapefile's attribute table using the following methodology. The Census' "GeoID", Log Record Number (Logrecno), and county and state name, a conversion database was constructed to align the records with the shapefile's "FIPS ID" and county names. Once completed, census information was merged into the shapefile's attribute table.

Once the Census information was embedded into the county shapefile, the HoR0203 and HoR04 shapefiles were spatially joined to the county shapefile to transfer the Census information into the HoR0203 dataset and to create a "count" column in the county shapefile. The count column reflects the number of recruits that joined the service from each county section. Therefore, each address within the HoR datasets reflected the demographic characteristics of the areal unit in which they reside and each county within the county shapefile contained a number of recruits attribute.

### Buffer shapefile creation

The third dataset source was obtained from the Bureau of Transportation Statistics which contained a shapefile that included the nation's 406 military installations. These military installations were then compared against a list of major military installations found on the Department of Defense "Sites" website. The shapefile (milbase) was then edited to reflect only those 194 installations contained on the major military installation list, which the DoD categorizes as those who fall into a "Large" or "Medium" size installation. A large installation is defined as those sites that have a total plant replacement value greater than or equal to $1.5 billion dollars and a medium installation are those with a plant replacement value less than $1.5 billion dollars but greater than $800 million dollars (Department of Defense 2003).

Based off of the major military installation shapefile, a multi-ring buffer shapefile was created by making 102 un-dissolved rings at 10 kilometer intervals around each military installation, resulting in a shapefile that contained nearly twenty thousand overlapping buffers. In order to eliminate all overlapping areas, the buffers were clipped one by one until the overlapping areas were eradicated. This was accomplished using the attribute table and starting with the lowest-distant buffers (10 kilometers) and working up to the highest-distant buffers (1,020 kilometers), the buffers were selected individually and then clipped so no buffers overlapped. Using the United States shapefile from the Bureau of Transportation Statistics, the buffer shapefile was clipped to give it a CONUS shape.

As with the county shapefile, the buffer shapefile was merged with the Census' Urbanized Area shapefile to create a buffer shapefile that was broken into urban and rural segments. This shapefile was then edited to ensure that there existed only one rural and one urban area for each buffer unit using the same methodology in the creation of the urban/rural county shapefile. The resulting shapefile contained 4,862 urban and rural areal units that represented 10 kilometer distances away from military bases.

To transfer the census demographic information into the buffer shapefile, two different methods were considered. The first method considered required spatially joining the buffer shapefile to the county shapefile, which would give average demographic characteristics for each buffer. Each attribute value would be based on the average of each county that intersected the buffer. The second method considered required spatially joining the buffer shapefile to the HoR0203 shapefile to extract average demographic characteristics for each buffer. The attribute values would be the average demographic characteristics of the home of record addresses. However, this method would cause the buffer units with zero recruits to contain null demographic information; nevertheless, for those areas that crossed multiple county boundaries it would weight the average towards those areas that had multiple recruits.

It was decided that a combination of both methods was the most appropriate. First, the second method was used to reflect those areal buffer units that contained at least one recruit, as it is believed that this method most realistically represented reality. For those buffer units that contained zero recruits, the shapefile was merged against the county shapefile and the resultant average attributes were used for those records; then this shapefile was then renamed "Buff0203". In addition to gaining the demographic information from the home of record shapefiles, Buff0203 added a count column which counted the number of recruits that fell within each buffer.

**X/Y Coordinates assignment and Dependent variable creation**

In order to assign a single X/Y coordinate to each polygon for the data analysis that required points instead of polygons, I considered two options. The first option was to take the geometric centroid of each polygon. However, this particular technique posed a problem when dealing with the buffer shapefile, as most of the buffers are donut shaped. Hence, multiple buffers would each have the same coordinate because the centroid for these donut buffers would lie in the middle of the donut hole. The other technique considered was to take the mean X and the mean Y for the data points that lie within the polygons' borders and use those means to create the point for the polygon. In this study, a compromise was accomplished for both polygon datasets. The average X/Y method was used for those areal units that contained recruits and the centroid method was used for those areal units that contained zero recruits (see Figures 4 and 5).

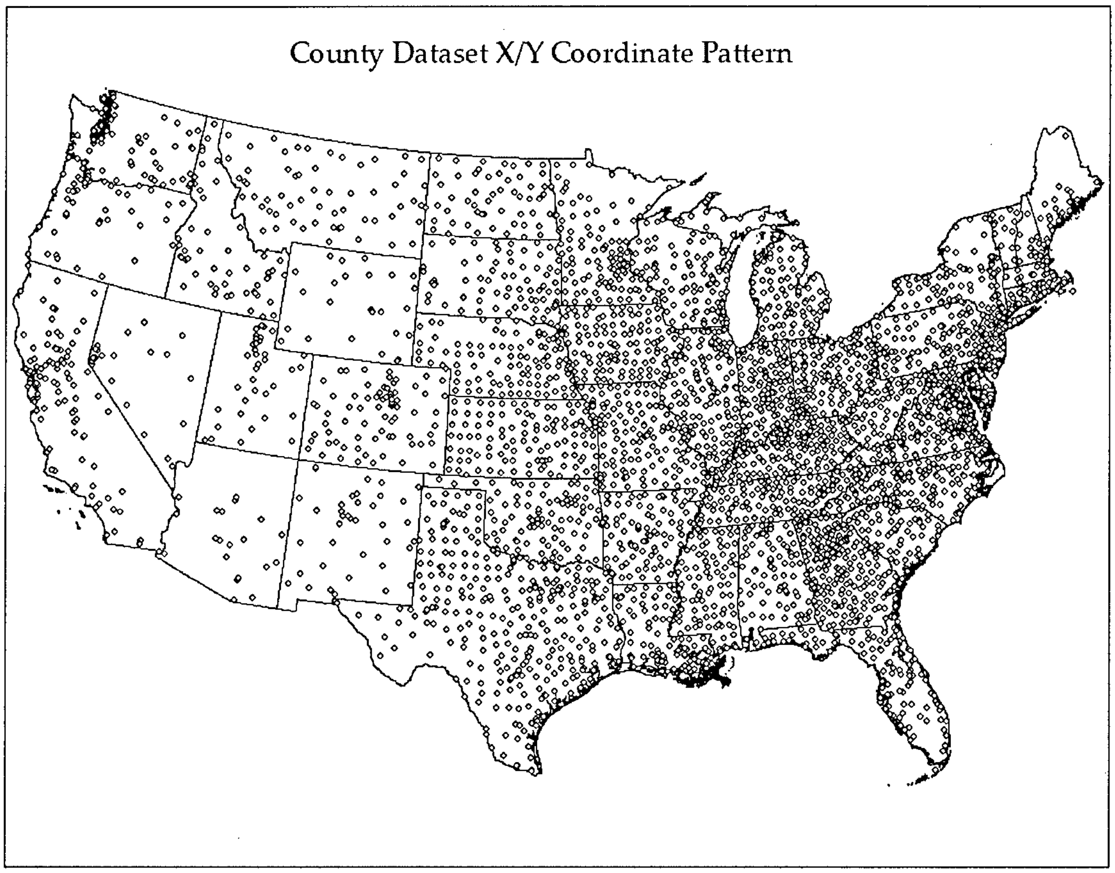


**Figure 4:** County Point Pattern

The dependent variable for this study was given considerable thought. The first argument originated with the thought that the Air Force would want to know the number of recruits per areal unit and is uninterested in the number being normalized by the population. The counter argument is that the Air Force does want to know this percentage to give the recruiters an idea of how many recruits they will garner for each age-eligible person that they meet. Therefore, it was decided to use both dependent variables. The first dependent variable, labeled "DV1", will be the raw count of enlistees where as the second dependent

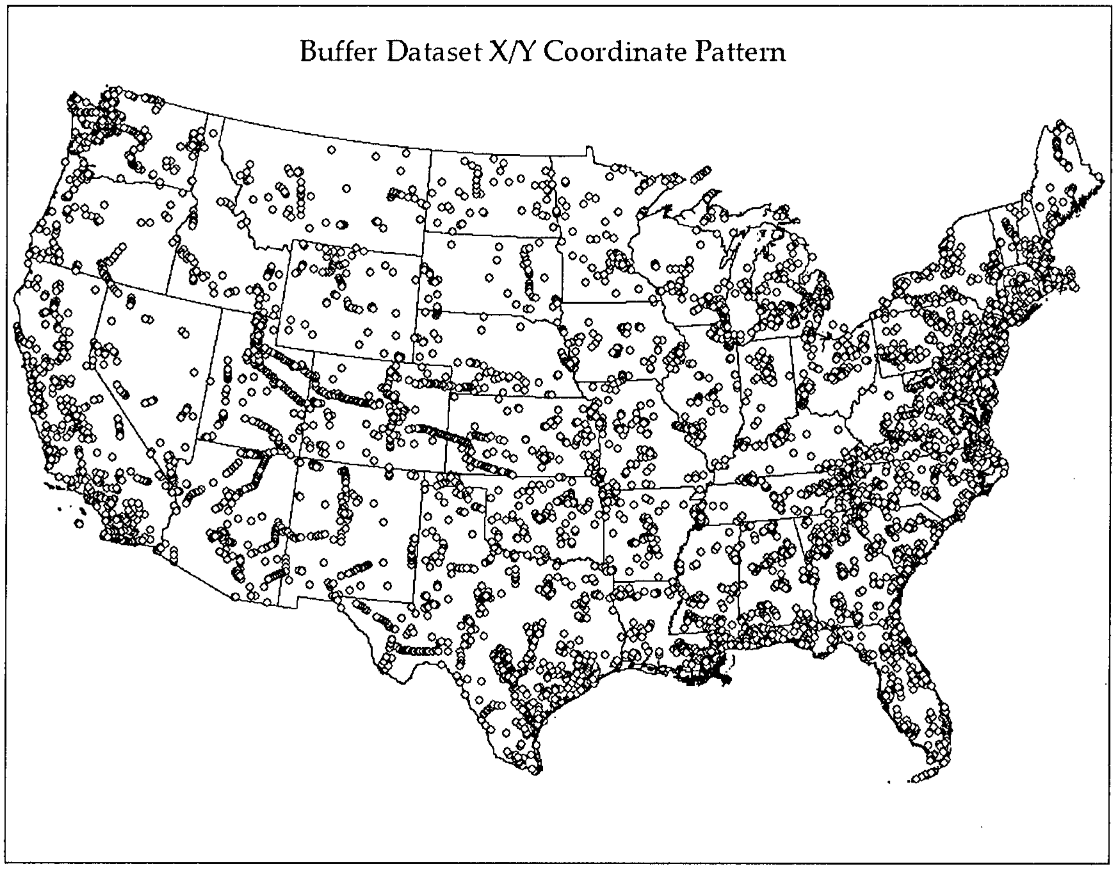


**Figure 5:** Buffer Point Pattern

variable, labeled "DV2" will be the raw count divided by the Raw1824 variable obtained from the census data.

The assignment of coordinates and dependent variables to the data completed the dataset's construction. To summarize, the County0203 and Buff0203 data matrices can be described as each record, or row, representing an urban or rural areal (county or buffer) unit and the variables, or columns, representing either 2000 Census demographic data, DV1 and DV2 variables, a corresponding X/Y coordinate or a distance to the nearest military installation variable (for the Buff0203 matrix). These two shapefiles were exported to a DBF IV file in preparation for statistical analysis.

# CHAPTER 4 – NON-SPATIAL AUTO-CORRELATED METHODS

Before analysis can be conducted, the datasets were evaluated for normality (a Gaussian distribution) by using the standardized coefficients of skewness and kurtosis. The calculated skewness and kurtosis statistics were compared against a t-value of 1.96, as the four datasets' degrees of freedom were above 120. Nevertheless, due to the large sample size, the standard error of skewness and standard error of kurtosis were exceedingly small for every variable. Therefore an effort was made to reduce the skewness and a kurtosis value to the lowest possible level through data transformation was needed.

A positive skewness existed in all of the variables with the exception of White, which was skewed negatively; therefore, it was determined that a natural log transformation was appropriate. However, due to log transformations' inability to transform any number equal or less than zero, necessary steps were taken to preserve the zero values within the data. Due to all the variables' original data being greater than or equal to one, each variable was assigned a new value by adding "1" to prevent the log transformation from converting the zero fields into "missing value" fields. Then, the natural log transformation was conducted. To complete the data manipulation, the transformed scores were then standardized to arrive at the values needed to conduct the analysis.

However, it should be noted that due to the high number of observations, which result in low standard error scores, all variables are "statistically" skewed in this study.

## Multiple Linear Regression

With the ultimate goal of this study being prediction, regression techniques will be a big discriminator in answering the question at hand; therefore, multiple linear regression was initially run on both the county and buffer datasets to establish a baseline. The primary criteria used in evaluating the regression outputs will be the coefficient of determination (adjusted R-Square), the Analysis of Variance F-Statistic and its related significance, the slope coefficient, the slope coefficient's t-value and related significance, and the distribution of the residuals.

In a regression analysis, the adjusted R-Square represents how well the independent variables contained within the model explain the variation of the dependent variable, which is by far the most quoted and representative means of explaining a regression's predictive capability. The Analysis of Variance's F-Statistic measures the degree of certainty that the model's ordinary least squares slope is unequal to zero, meaning it exhibits a relationship between the independent variables and the dependent variable. Likewise, the t-value identifies if the slope of the individual independent variables is unequal to zero, or in other words, if there exist a relationship between each independent variable and the dependent variable. Lastly, in order to adhere to the assumptions of the central limit theorem, the residuals must be normally distributed; therefore, an analysis of the residuals was conducted to ensure compliance.

### Adjusted R-Square, F-Statistic, Significance, and Residuals

These adjusted R-Square, F-Statistic, and significance measures reflect the global aspects of regression, as they give an indication of how well the model performed overall (See Table 1). This study's adjusted R-square values leave much to be desired. The county DV1 model performed the best with a .301 adjusted R-square, which was far superior to the rest of the models. The buffer DV1 model performed the next best with a .048 adjusted R-Square. Next, the buffer DV2 model obtained a .017 adjusted R-Square and the county DV2 model produced a .007 adjusted R-Square.

The F-Statistics followed the same pattern of performance with the county DV2 (3.389) performing worst, followed by the buffer DV2 (7.451), buffer DV1

(19.801), and finally the county DV2 (140.942). Despite the dismal values, all F-Statistics were found to be significant at the 95 percent level.

The buffer dataset's residuals varied greatly between the DV1 and DV2 models. DV1 produced a near-normal distribution with a slight spike in peakedness between the -2 and -1 standard deviations. The DV2 distribution pattern's kurtosis was significantly peaked. Furthermore, the P-P Plot diagrams reveal that the DV1 model was near normal with a slight deviation on the negative side of the mean, which represent the aforesaid peakedness, and the DV2 P-P Plot diagram clearly shows a strong negatively skewed distribution. See Appendix 1 for a detailed view of the results.

The county datasets' residuals closely mirrored those of the buffers with the DV1 possessing a normal distribution and DV2's being negatively skewed. Unlike the buffer DV1, the county DV1 was near-perfect with both the histogram and the P-P Plot indicating a normal distribution. The county DV2 mirrored the buffer DV2's dataset peakedness with a bimodal distribution pattern (See Appendix 1).

**Table 1:** Regression Global Results

| | Regression | | | |
|---|---|---|---|---|
| | Buffer DV1 | Buffer DV2 | County DV1 | County DV2 |
| Adjusted R-Square | 0.048 | 0.017 | 0.301 | 0.007 |
| F-Statistic | 19.801 | 7.451 | 140.942 | 3.389 |
| Significance | 0.000 | 0.000 | 0.000 | 0.000 |

## Slope Coefficients, t-statistic, and Significance

The slope coefficient is a numerical representation of the relationship between the dependent variable and the independent variable. A higher slope coefficient means the variables possess a greater relationship. The t-statistic is basically a confidence score where the higher the absolute value of the t-statistic, the more confident the relationship is a representation of reality.

A look at the regression's local statistics begins to reveal the disparity that exists between the county DV1 dataset and the others, in terms of its predictability (See Table 2). For the county DV1 dataset all but two variables were significant at the 95 percent level, which compared against the other three datasets which only had three to five significant variables. County DV1's two insignificant variables were AD and PerCap. However, due to county DV1's significantly larger adjusted R-Square its' slope coefficients deserve a closer look. Beginning with the two variables that had a negative correlation, White had a -

.838 coefficient and NoHS had a -.193 coefficient, which indicate that as the numbèr of recruits increase as the number of Whites and persons without a high school degree decrease. The variables with strong positive coefficents include Black, AHP, Other, UnEmp, and HHI with the three strongest being Black, Other, and HHI. With both Black and White variables having extremely strong coefficients suggests that these two variables had the strongest relationship; however, the AHP and HHI variables possessed the highest t-statistic which suggests otherwise. Regardless, AHP, HHI, White, Black, NoHS, UnEmp, and HHI were the strongest variables in this model, which appear to support the findings found in recruitment literature. In reference to the other three

**Table 2:** Regression Local Results. Grayed areas are insignificant

| | Buffer DV1 | | | Buffer DV2 | | | County DV1 | | | County DV2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Slope Coefficient | t-statistic | Significance | Slope Coefficient | t-statistic | Significance | Slope Coefficient | t-statistic | Significance | Slope Coefficient | t-statistic | Significance |
| Distance | -0.016 | -1.111 | 0.266 | 0.011 | 0.768 | 0.443 | | | | | | |
| White | 0.039 | 0.386 | 0.700 | 0.159 | 1.532 | 0.126 | -0.838 | -8.543 | 0.000 | 0.178 | 1.527 | 0.127 |
| Black | -0.063 | -0.735 | 0.463 | -0.159 | -1.840 | 0.066 | 0.856 | 9.855 | 0.000 | -0.161 | -1.559 | 0.119 |
| Latin | -0.125 | -6.584 | 0.000 | -0.048 | -2.481 | 0.013 | 0.085 | 4.490 | 0.000 | 0.025 | 1.125 | 0.261 |
| AHP | 0.012 | 0.598 | 0.550 | 0.049 | 2.470 | 0.014 | 0.238 | 13.145 | 0.000 | -0.042 | -1.962 | 0.050 |
| Other | 0.071 | 1.129 | 0.259 | -0.049 | -0.760 | 0.447 | 0.393 | 6.873 | 0.000 | -0.102 | -1.493 | 0.135 |
| Y1824 | -0.118 | -6.482 | 0.000 | -0.123 | -6.665 | 0.000 | 0.098 | 5.520 | 0.000 | 0.029 | 1.353 | 0.176 |
| NoHS | -0.149 | -6.597 | 0.000 | -0.040 | -1.731 | 0.083 | -0.193 | -9.286 | 0.000 | 0.040 | 1.607 | 0.108 |
| AD | -0.009 | -0.599 | 0.549 | -0.005 | -0.350 | 0.726 | 0.015 | 1.012 | 0.312 | 0.003 | 0.194 | 0.846 |
| Vet | -0.015 | -0.816 | 0.415 | -0.034 | -1.825 | 0.068 | 0.078 | 4.629 | 0.000 | 0.091 | 4.530 | 0.000 |
| UnEmp | 0.103 | 5.502 | 0.000 | -0.007 | -0.389 | 0.698 | 0.165 | 8.883 | 0.000 | -0.032 | -1.422 | 0.155 |
| HHI | -0.057 | -2.655 | 0.008 | -0.072 | -3.305 | 0.001 | 0.294 | 13.832 | 0.000 | 0.052 | 2.032 | 0.042 |
| PerCap | 0.001 | 0.056 | 0.955 | -0.026 | -1.546 | 0.122 | -0.006 | -0.370 | 0.711 | 0.022 | 1.123 | 0.262 |

regression models, their dismal adjusted R-square values deem their individual results inappropriate to discuss.

In reference to this study's question, the picture that the regression models paint is that distance thus far has failed to prove a significant method for predicting recruitment; whereas, the county DV1 regression model performed adequately in this respect. Furthermore, the results of the county DV1 mirror the results of Cannizzo's (1995) study, which tried using geographic information systems to predict recruitment in Ohio and ultimately arrived at an adjusted R-Square of .294.

**Principal Component Analysis**

In order to combat the effects that correlation between the independent variables exhibit on the coefficient of determination, Principal Component Analysis (PCA), a factor analysis technique, was conducted. Therefore, PCA was executed on the independent variables for both datasets using a varimax rotation which was set to extract components with Eigenvalues greater than one. The final solution resulted in 4 different components for the buffer dataset and 5 for the county dataset. The buffer dataset explained over 62 percent of the variance within the data and the county dataset explaining over 76 percent.

To begin with, it is important to note the PCA's performance on each dataset by using the communalities results, which give a score indicating how well the overall PCA represented each variable. These values range from zero (zero percent) to one (one hundred percent). For the buffer dataset, the extraction scores ranged from .599 to .819 with the exception of the distance variable which obtained a miserable score of .024. The county dataset's extraction values were exceedingly better which ranged from .582 to .962.

Beginning with the buffer dataset, the first extracted component explained 18.952 percent of the variance, which represented the income and educated-related variables (HHI, PerCap, and NoHS (negatively correlated)); this component was then renamed "Income_Ed". The second component explained 16.3 percent of the variation and represented the White, Black, and Y1824 groups, which was renamed "Wh_Bl_Age". Explaining 11.32 percent of the variance, the third component, renamed "Lat_Oth", represented Latin and other variables. The last component explained 9.565 percent of the variance and represented the AD and Vet populations, which was renamed "military".

The county dataset yielded similar results as the buffer dataset with the exception that it extracted one more component. The first extracted component was nearly identical to the buffer dataset which explained 26.907 percent of the variation and was renamed "Income_Ed". The second component explained 18.046 percent of the variation and represented the White and Black populations and was renamed "Wh_Bl". The third component, renamed "Lat_Oth", represented the Latin and other variables and explained 13.190 percent of the variation within the data. The fourth component, explaining 9.844 percent, represented the Y1824 and vet variables, which were negatively correlated renamed "Age". The last component was comprised of only the AD variable, which explained 8.607 percent of the variation.

Ultimately, the PCA accomplished the goal of minimizing the variables into four or five components that were completely uncorrelated, giving an opportunity to conduct multiple linear regression devoid of correlation influence. Furthermore, it is interesting that each dataset's resulting components were closely related to one another. One issue needing discussion is that in both grouping methods the income-related categories claimed the most variance explained which appears to support the recruitment-related literature that cites that income is a strong recruitment predictor.

### Multiple Linear Regression using PCA variables

As expected, an overall trend to these regression techniques produced adjusted R-Square values which were less than the original regression model, as the correlation effects of the first method enhanced the coefficients of determination. Nevertheless, this model did increase the F-Statistics and all continued to be statistically significant at the 95 percent level, suggesting this model is a more accurate representation of reality.

### Adjusted R-Square, F-Statistic, Significance and Residuals

As with the original regression method I will begin with the global statistics. The buffer DV2 model method was the worst performer of the four (See table 3) as it produced a .004 adjusted R-Square value and an F-Statistic of 5.928. The county DV2 model was the next worst performer with a coefficient of determination of .006 and a 5.355 F-Statistic. The buffer DV1 model produced an adjusted R-Square of .016 with a 20.86 F-Statistic. Lastly, the county DV1 model performed the best. It produced a coefficient of determination of .240 and a 247.201 F-Statistic. The residuals of the factor analysis regression model exhibited nearly-identical distributions to the original regression models with the exception that the kurtosis, which in every model was minimally reduced. Overall, the adjusted R-Square values were exceedingly low, although significant, suggesting there are many other reasons (variables) why individuals join the Air Force.

**Table 3:** Regression using Factor analysis components: Global Results

Factor Regression

| | Buffer DV1 | Buffer DV2 | County DV1 | County DV2 |
|---|---|---|---|---|
| Adjusted R-Square | 0.016 | 0.004 | 0.240 | 0.006 |
| F-Statistic | 20.860 | 5.928 | 247.201 | 5.355 |
| Significance | 0.000 | 0.000 | 0.000 | 0.000 |

### Slope Coefficients, t-statistic, and Significance

The slope coefficients in this technique mirrored those of the original regression model where most of the variables were insignificant (See Table 4). For both the buffer and county DV2 models, only one variable was significant in each, whereas in the buffer and county DV1 models, most if not all, were significant. For the buffer DV1 model, the Lat_Oth variable was insignificant, the Wh_Bl_Age category produced the highest slope and the other two produced negligible relationships. However, the county DV1 model performed well as with the original regression method and indicated that income and education remained the most prominent component in predicting recruitment with a slope coefficient of .401 and a 28.712 t-statistic. The next most significant variable was the AD component with a slope coefficient of .213 and a t-statistic of 15.242. The next three variables in order of significance were Age, and Wh_Bl, and Lat_Oth, which all produced t-statistics less than 10.

The regressions based on the factor analysis variables increased the F-Statistic dramatically, indication of a strong relationship within the data even if the adjusted R-Square value was smaller than the original regression techniques. Also, the factor analysis slope coefficients constantly were greater than their individual variable counterparts, further substantiating a strong relationship.

**Table 4:** Regression with Factor Analysis components Local results: Grayed areas are insignificant.

| | Buffer DV1 | | | Buffer DV2 | | |
|---|---|---|---|---|---|---|
| | Slope Coefficient | t-statistic | Significance | Slope Coefficient | t-statistic | Significance |
| Income_Ed | 0.031 | 2.191 | 0.028 | -0.006 | -0.394 | 0.694 |
| Wh_Bl_Age | -0.113 | -7.936 | 0.000 | -0.068 | -4.784 | 0.000 |
| Lat_Oth | 0.012 | 0.840 | 0.401 | 0.001 | 0.078 | 0.938 |
| Military | 0.055 | 3.867 | 0.000 | -0.012 | -0.870 | 0.384 |
| | | | | | | |
| | County DV1 | | | County DV2 | | |
| | Slope Coefficient | t-statistic | Significance | Slope Coefficient | t-statistic | Significance |
| Income_Ed | 0.401 | 28.712 | 0.000 | 0.036 | 2.260 | 0.024 |
| Wh_Bl | 0.095 | 6.811 | 0.000 | -0.022 | -1.393 | 0.164 |
| Lat_Oth | 0.094 | 6.766 | 0.000 | -0.012 | -0.755 | 0.450 |
| Age | 0.130 | 9.334 | 0.000 | -0.067 | -4.195 | 0.000 |
| AD | 0.213 | 15.242 | 0.000 | 0.020 | 1.249 | 0.212 |

## Chapter Summary

In reference to the question at hand, "Does distance from a military installation predict recruitment?" the results of the regression models do not look favorably to support the hypothesis. The distance variable was insignificant in the first two buffer regression models, was not represented well in the factor analysis, and was not included in any components for the final regression analysis. In regards to prediction, the county DV1 model performance appears to be in line with previous research by indicating income, education, and minorities as the major recruitment predictors. In terms of the hypothesis, the county DV1 factor regression model did extract that AD was the second-most important variable in predicting recruits, giving a ray of hope for supporting the hypothesis.

However, with adjusted R-Square scores below 30 percentage points, the models do not fit the data very well. A possible source for this problem is that neither Ordinary Least Squares regression nor PCA take into account the spatial attributes inherent in the data. In order to adjust for this assumption, three spatially auto-correlated analysis techniques will be executed in order to improve upon the goodness of fit measure: Spatial Point Pattern Analysis, kriging, and Geographic Weighted Regression.

# CHAPTER 5 – SPATIAL AUTO-CORRELATED TECHNIQUES

## Point Pattern Analysis

When comparing items across a geographic plane, Tobler's law suggest nearer items tend to be more similar than those farther away. When similar items are found to be located near one another, this concept is known as clustering. The clustering affect contained within a dataset is of primary concern for much of the geographic inquiry in this study, as clustering can oftentimes lend itself to prediction. If clustering can show which areas contain a large percentage of recruits, then a closer look at those areas is warranted to uncover the underlying trends that could potentially serve as a recruitment predictor. In order to test to see if phenomena are clustered instead of simply looking at a map, statistics are used to confirm or deny the existence of similarity within a group of points.

These techniques offer a global and local measure to determine if the events are clustered (similar) or dispersed (different). The global feature simply looks at the dataset as a whole to determine whether the dataset could have happened in a random environment. This global value is measured in normal standard deviations, identifying if the dataset is random, dispersed or clustered.

Likewise, for the local measure, each observation is assigned an individual value that describes its relationship to its neighbors.

Nearest neighbor is an exceptional initial technique that is used which simply analyzes the distribution of data points to give an overview of the dataset's spatial pattern. Nearest Neighbor calculates the Euclidean distance between each point and every other point to determine which point is its nearest neighbor. After all points have an associated nearest neighbor, the average is taken in addition to an expected average distance being calculated. This expected value is compared against the actual observed pattern in the data to produce a R Value. The R value ranges from zero, indicating complete clustering, to 2.149, indicating completely dispersed, with a value of 1 being completely random. The R value is then standardized to obtain its associated Z-Score which is compared to a t-distribution table. If the Z-Score is greater than 1.96 standard deviations or less than -1.96 standard deviations then it is considered statistically significant, given it has more than 120 degrees of freedom. However, the weaknesses with Nearest Neighbor technique are that it considers all points as equals and it does not take into account the study area's anomalies (i.e. rivers, ranges, deserts, etc). These effects can be minimized by using the following techniques.

Moran's I is a statistical measure that looks at local variation to see how each point (or polygon) differs from the dataset's mean. This calculation is then compared to the rest of the data points in the dataset in order to distinguish if the item is similar or different than the surrounding items. A large global Moran's I value indicates that items are surrounded by items with similar values (positive spatial autocorrelation). On the other hand, a small global Moran's I value shows that the item is dispersed because it is surrounded by others with dissimilar values (negative spatial autocorrelation). The question the global Moran's I answers is, "Is the data spatially auto correlated?" with the null hypothesis being that it is not auto correlated. In addition to the global value, each event can be given a local Moran's I value so the values can be mapped to determine the locations of positive and negative spatial autocorrelation.

Getis Ord GI is another technique for testing spatial autocorrelation. Getis Ord's "General G" statistic builds upon Moran's I by indicating whether the clusters contain high or low values; thus, creating what is referred to as "Hot Spots" and "Cold Spots". However, the global value is unable to determine if both hot and cold spots exist in the data, it simply tells which one is more prevalent. However, the local values can be used to identify the hotspot and coldspot locations.

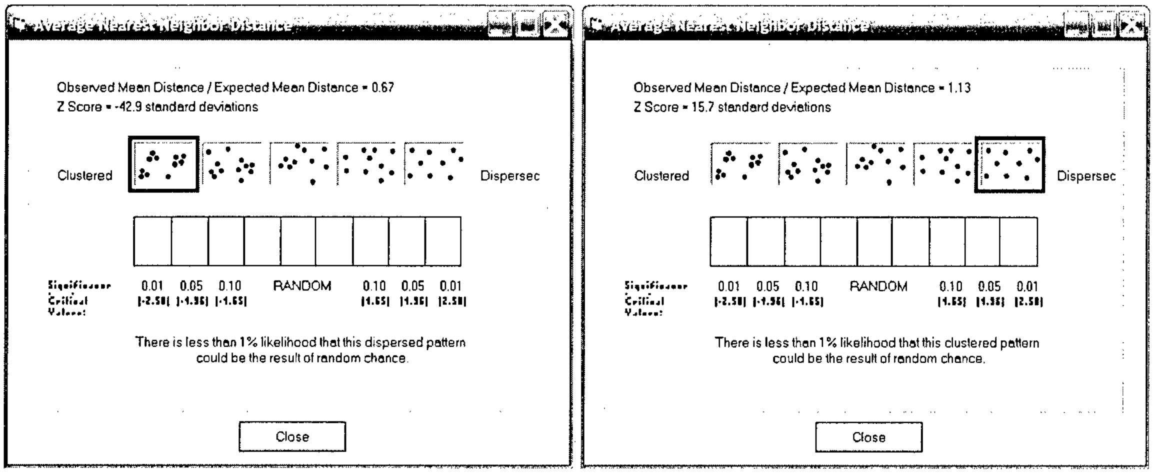
The results of the point pattern analysis will be outlined in the following manner. The Nearest Neighbor, global and local Moran's I, and global and local Getis' GI, will be conducted on both the buffer and county dataset in order to compare and contrast the results. A summary of the similarities and differences between the county and buffer datasets will be discussed at the end.

### Nearest Neighbor

The Nearest Neighbor analysis for both datasets yielded dissimilar results (see Figure 6). The buffer dataset possessed an extreme clustering characteristic with a Z Score of -42.9 standard deviations and a R Value of .67. The county dataset's characteristic was nearly opposite with a dispersed pattern and a Z Score of 15.7 standard deviations and an R Value of 1.13. These global results give a general view that illustrate the buffer dataset is significantly clustered and the county is dispersed, but these indications merely look at the data point and not the individual values associated with the locations.

### Moran's I

The global Moran's I was calculated for both dataset's dependent variables (See Figures 7), which reflected similar patterns as the Nearest


Average Nearest Neighbor Distance
Clustered
Dispersec
Significance 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01
Critical Values: |-2.58| |-1.96| |-1.65| |1.65| |1.96| |2.58|
Close
Average Nearest Neighbor Distance
Clustered
Dispersec
Significance 0.01 0.05 0.10 RANDOM 0.10 0.05 0.01
Critical Values: |-2.58| |-1.96| |-1.65| |1.65| |1.96| |2.58|
Close

Observed Mean Distance / Expected Mean Distance = 0.67
Z Score = -42.9 standard deviations

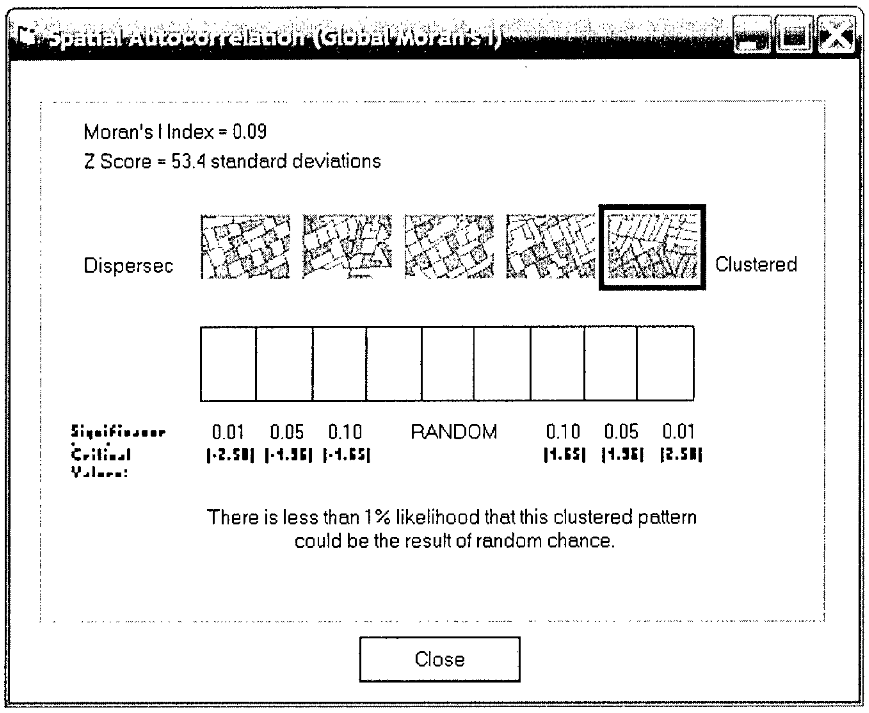
There is less than 1% likelihood that this dispersed pattern could be the result of random chance.

Observed Mean Distance / Expected Mean Distance = 1.13
Z Score = 15.7 standard deviations

There is less than 1% likelihood that this clustered pattern could be the result of random chance.

**Figure 6:** Buffer and County Nearest Neighbor Results

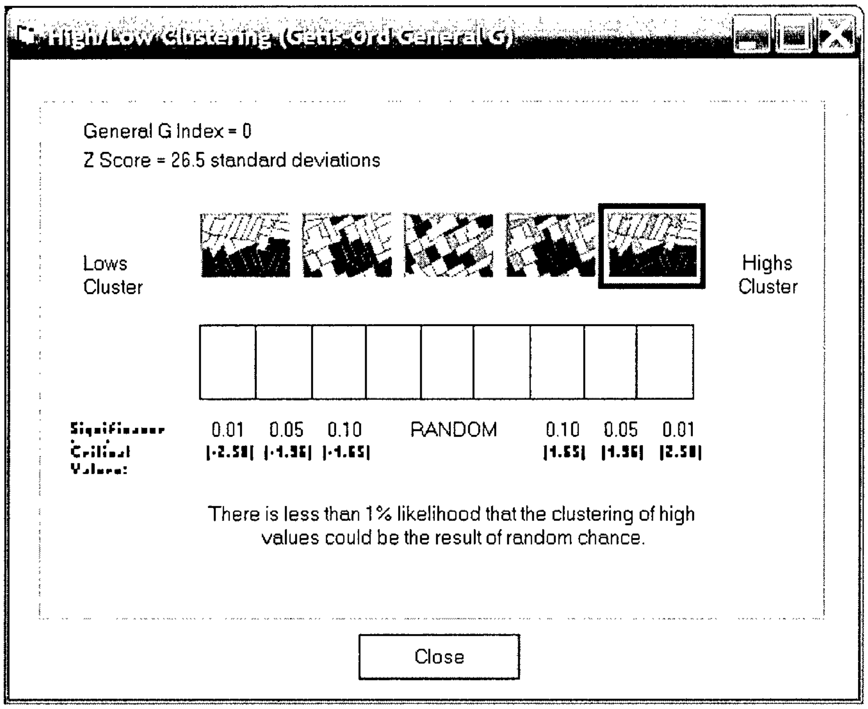
Neighbor for the buffer dataset but dissimilar results for the county dataset. For the buffer dataset, the strongest indicator belongs to DV1 which possessed 94.9 standard deviations and the DV2 produced 21.8 standard deviations. The county dataset's strongest indicator was the DV1 dataset with a 6 standard deviations and the DV2 producing a 1.5 standard deviations. These results show that the buffer dataset was extremely clustered and the county datasets exhibited a slight clustering for the DV1 variable and a random disbursement for the DV2 variable. Regardless of the individual results, this global point pattern technique reveals spatial auto-correlated within the data except for the county DV2 variable; however, what Moran's I fails to interpret is if the auto-correlation is based off of high or low values.



Moran's I Index = 0.09
Z Score = 53.4 standard deviations

There is less than 1% likelihood that this clustered pattern could be the result of random chance.

**Figure 7:** Moran's I Output Example



General G Index = 0
Z Score = 26.5 standard deviations

There is less than 1% likelihood that the clustering of high values could be the result of random chance.

**Figure 8:** Getis' Ord GI Output Example

### Getis' Ord GI

In reference to global Getis' GI, the results indicate a compelling story (See Figures 8). First, both buffer DVs exhibited a high-value clustering pattern with DV1 possessing the strongest indication with 22.7 standard deviations and DV2 produced a Z Score of 15.4 standard deviations with both datasets clustering high values. Both of these clusters were significant at the 99.9 percent level. For the county dataset, DV1 indicated a low-value cluster with a standard deviation of -4 significant at the 99.9 percent level. The county DV2 dataset produced a Z-Score of -0.2 standard deviations, which indicate that the point pattern exhibits a near-perfect random pattern. Furthermore, in reference to the difference between the two dependent variables, it should be noted that the DV1s produced stronger results than their DV2 counterparts, indicating that the DV1 methods exhibit a stronger spatial auto-correlation. Nevertheless, a closer look at the local values should reveal more about the spatial distribution of these indicators.

### Local Moran's I

As indicated in the previous sections, the global value gives a composite sketch of the dataset as a whole whereas the local assigns a value for each event. Therefore, the results of the local analyses can be viewed on a map to distinguish the variability over space. For display purposes, the county and buffer results will be shown using polygons instead of their corresponding points in order to ease the interpretation.

By recalling that Moran's I basically is an indicator of similarity/dissimilarity, by looking at a map one can interpret those areas that have similar values, values greater than 1.96 standard deviations, and those areas that are statistically dissimilar, which are areas that have a standard deviation less than -1.96. The buffer's global Moran's I indicated that the DV1 dataset was extremely clustered which was reemphasized by mapping the local Moran's I values. The geographic areas that were similar to its neighbors can be broadly defined as the Rocky Mountain West and the Midwest area ranging from the Northern Deep South up to the areas surrounding the Great Lakes and over to New England's metropolis (See Figure 9). When looking at the DV2 results, there were only a few areas that displayed any significance, which can be summed up as a few urban areas indicating similarity or dissimilarity (See Figure 10).

On the other hand, the county results were not so easily interpretable. For the County DV1 dataset, whose global statistic indicated a slight auto-correlation, (See Figure 11), the pattern appeared scattered with some similarity being displayed through portions of Florida and the South. The dissimilarity counties were scattered throughout the study area with no geographic area displaying a strong pattern. The county's DV2 results did not produce any pattern (see Figure 12), as there were only a few statistically significant counties sporadically scattered throughout the study area.

However, as discussed in the global Moran's I section, this technique simply displays the degree in which geographic areas behave; it does not indicate whether the behavior is high or low. Therefore, the local Getis' Ord GI method is utilized to determine which locations contain hotspots (those areas of high values) and cold spots (those areas of low values).

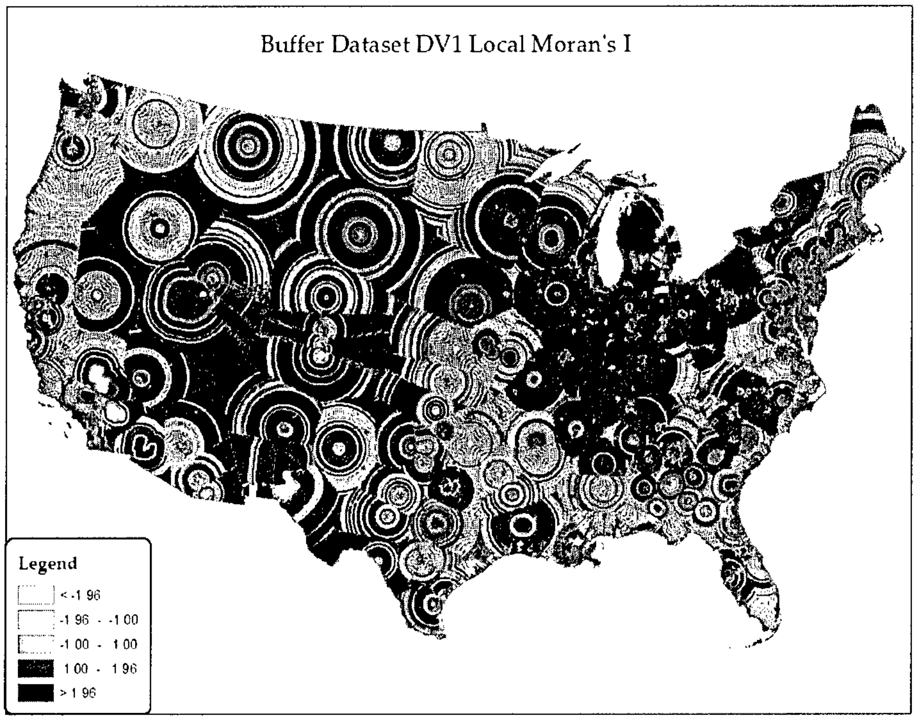


**Figure 9:** Buffer DV1 Local Moran's I

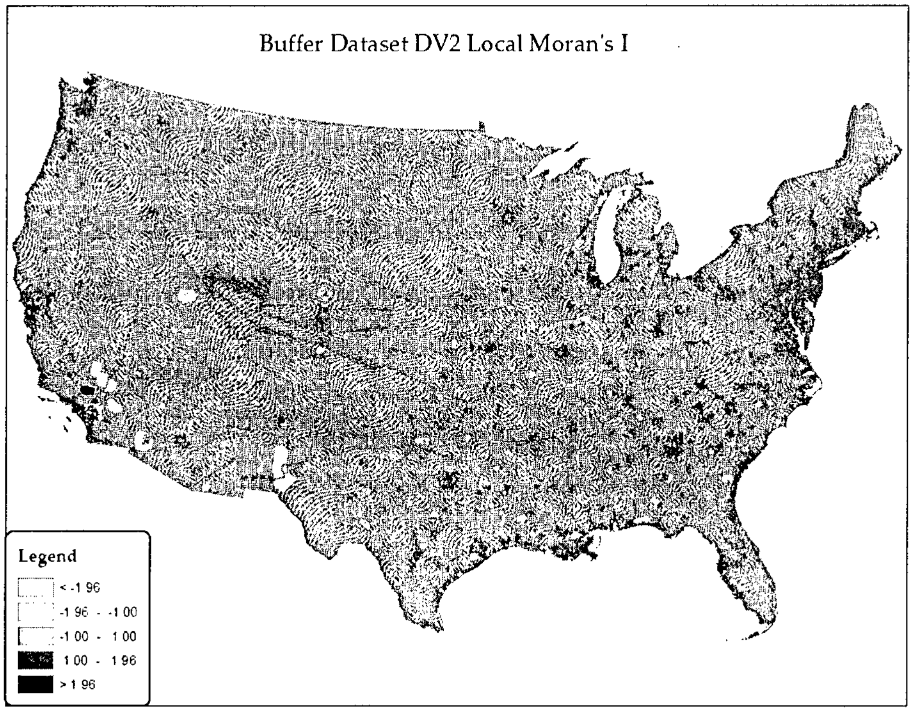


**Figure 10:** Buffer DV2 Moran's I

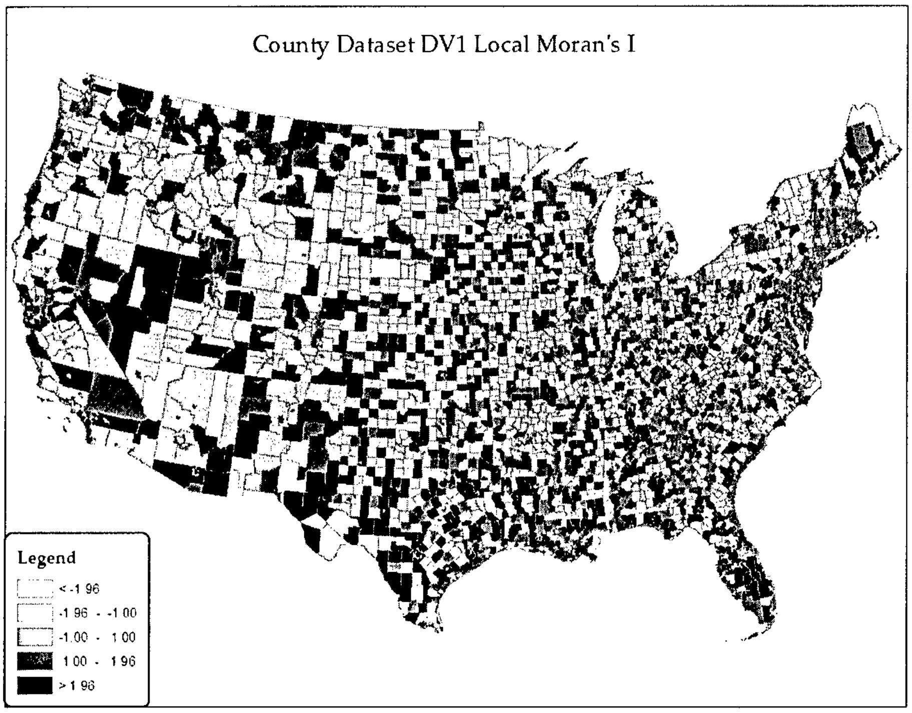


**Figure 11:** County DV1 Local Moran's I

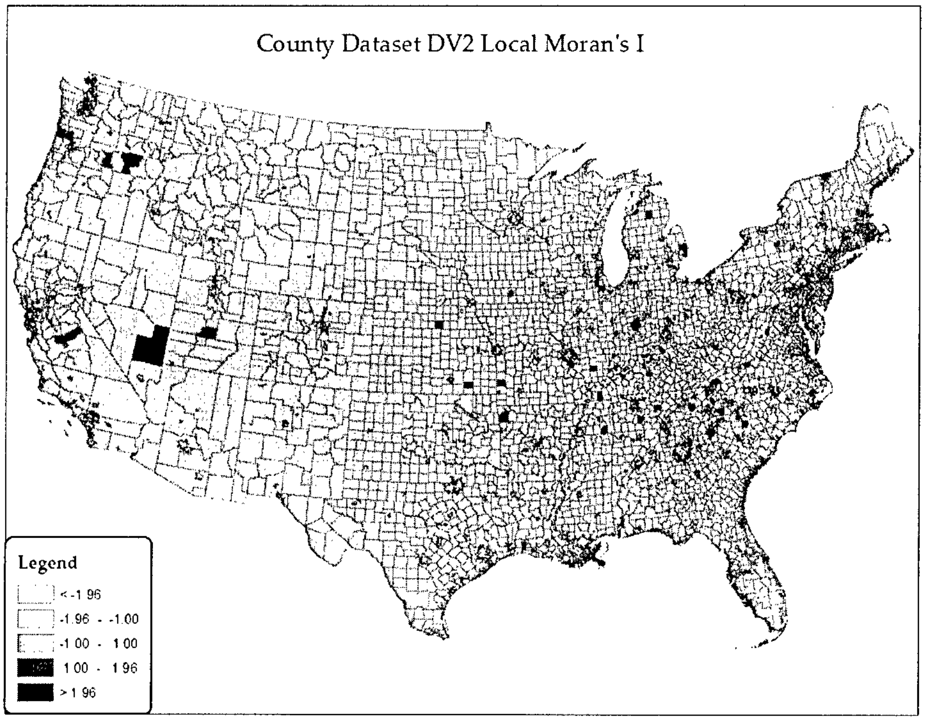


**Figure 12:** County DV2 Local Moran's I

### Local Getis' Ord GI

The Getis' Ord GI takes the Moran's I a step further by measuring an attribute's degree of spatial autocorrelation of and also determines where the high or low values are clustered. The buffer DV1 dataset revealed a strong sense of commonality between different regions of the study area with only a few not having significant hi or low values (See Figure 13). The areas of low value primarily were centered in the Rocky Mountain and Great Plain states whereas the high values were everywhere else with the exception of the Deep South and some Mid-Atlantic regions. The buffer DV2 results only showed a few hot spot areas with no cold spots. Generally speaking, the hot spot areas were centered on installations that are in close proximately to major urban areas (See Figure 14).

With the Getis' Ord GI technique, the county dataset performed nearly as well as the Moran's I with a fairly random distribution, which explains its near-random global statistic. However, a general trend can be identified for the DV1 dataset which suggests areas of high values can be found in the lower latitudes and upper latitudes of the study area whereas the low values can be found in the mid-latitudes (See Figure 15). However, this is speaking very generically as there were many highs and lows scattered throughout the entire study area. The DV2 variable was more random than the DV1 variable, which supports the global statistic's score as there were only a handful of statistically significant high values located throughout the study area's mid-latitude (See Figure 16).

Ultimately, what these point patterns indicate is that military-service attitudes vary throughout the United States, which should be taken into consideration when developing recruitment policy. Nevertheless, in regards to this study, by categorizing recruits via the buffer method more clearly illustrates the local variation than by grouping them by county.

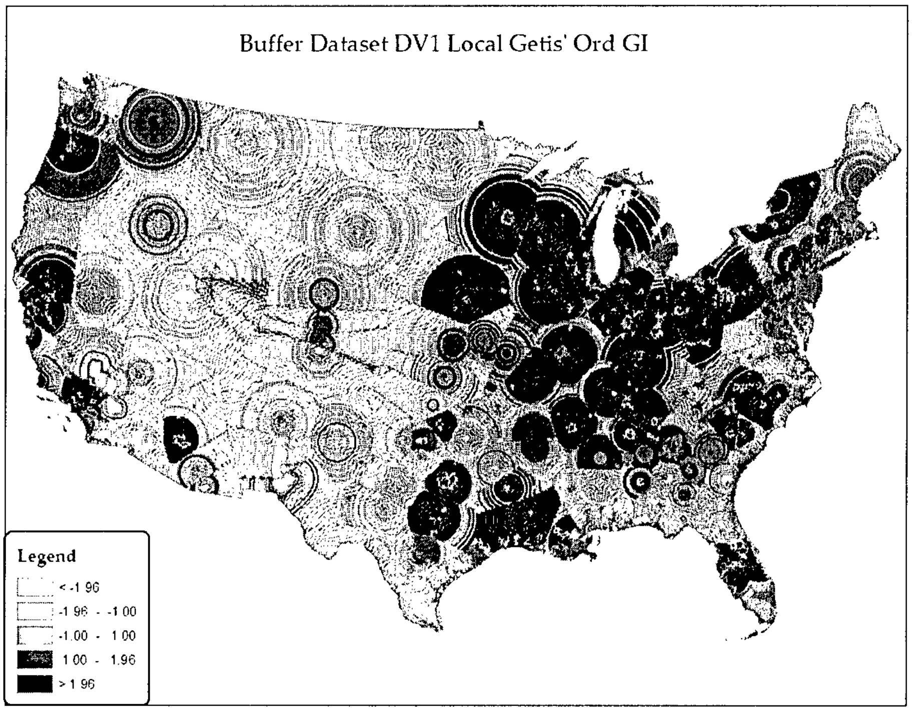


**Figure 13:** Buffer DV1 Local Getis' Ord GI

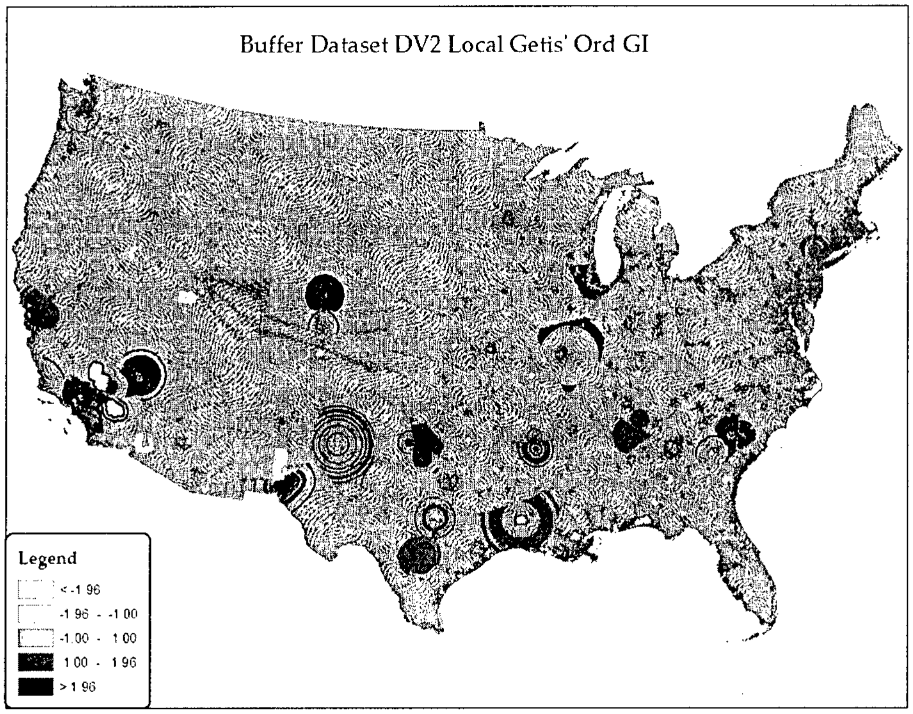


**Figure 14:** Buffer DV2 Local Getis' Ord GI

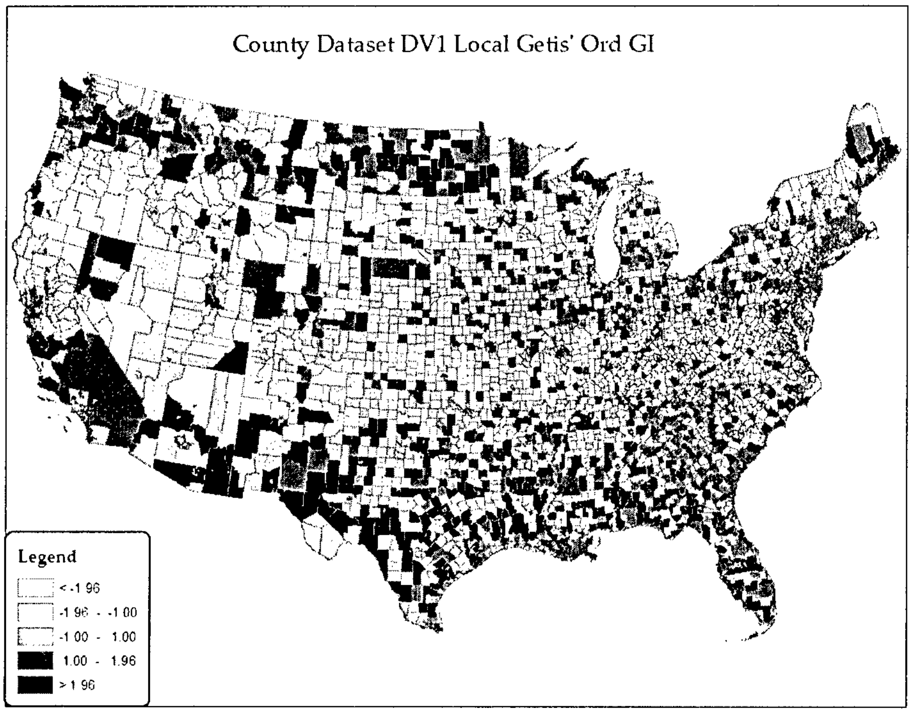


**Figure 15:** County DV1 Getis' Ord GI

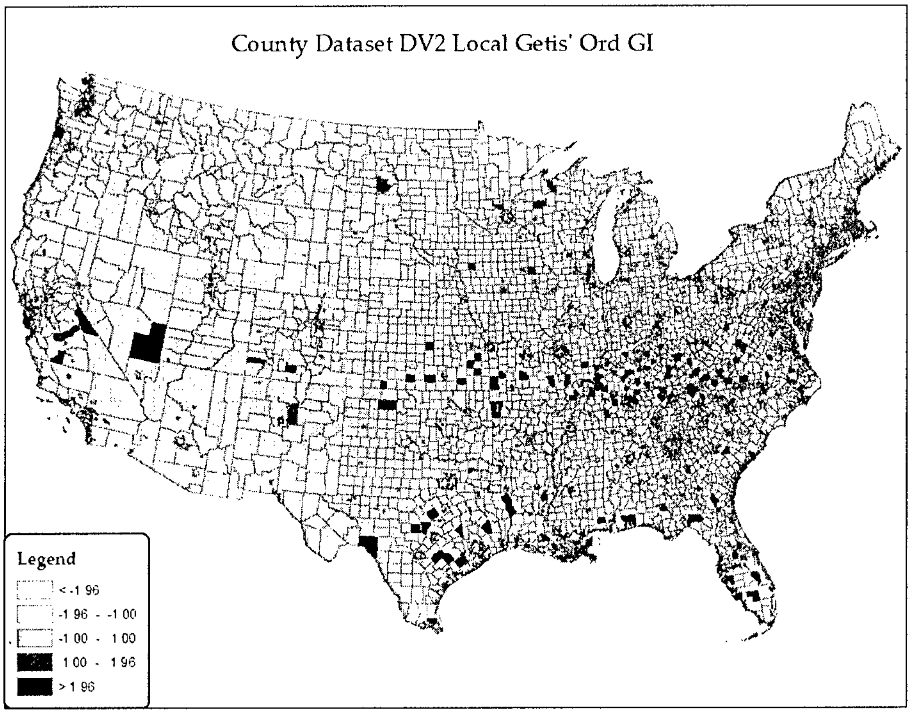
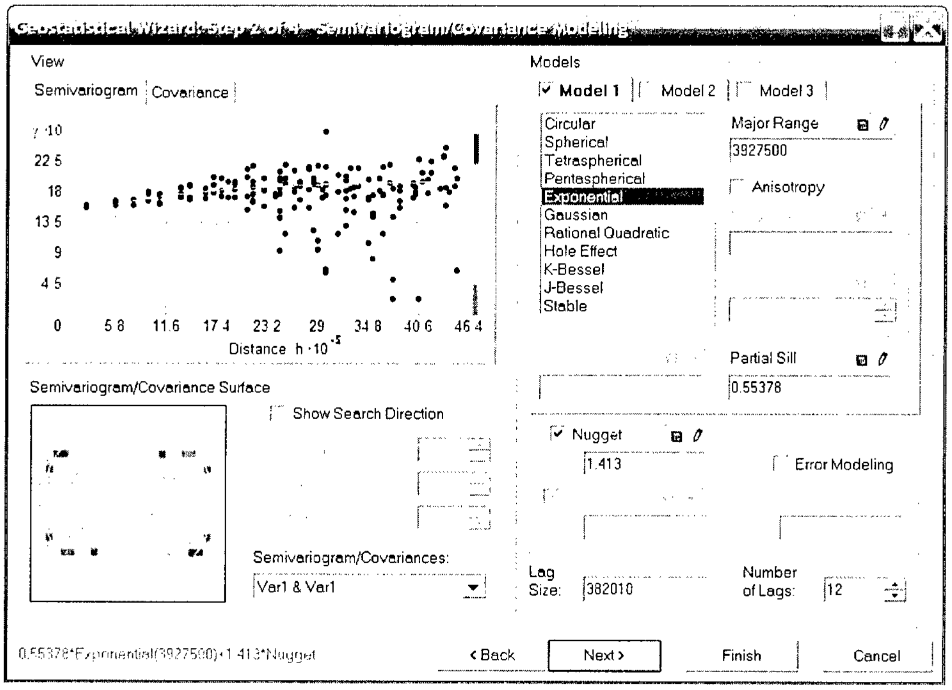


**Figure 16:** County DV2 Getis' Ord GI

## Kriging

Kriging is an interpolation technique that is able to predict values in areas where no values exist. As with the Moran's I and Getis' Ord GI, kriging assumes spatial autocorrelation when predicting values as it weights those items closer more than those farther away. Kriging produces a continuous surface based off of known value points. In this study, it is acknowledged that recruitment levels are not continuous surfaces like precipitation or air density, but this technique can draw areas of high and low recruitment without being bound by the county or buffer boundaries.

ESRI's ArcGIS geostatistical analyst allows users to choose from six different kriging techniques. As mentioned above, kriging is designed to use samples that are taken from a phenomenon that is continuous in space; therefore, since the number of recruits per areal unit is not a continuous phenomenon all the different kriging methods were analyzed. Fifty different techniques were used to explore the data to determine which method best fit the data. Ultimately, the tests found very little difference between the six different techniques; thus, for this study Ordinary Kriging was used to create prediction maps.



Geostatistical Wizard: Step 2 of 4 - Semivariogram/Covariance Modeling

☑ Model 1 | ☐ Model 2 | ☐ Model 3

Model: Exponential

Major Range: 3927500

☐ Anisotropy

Partial Sill: 0.55378

☑ Nugget: 1.413

☐ Error Modeling

☐ Show Search Direction

Semivariogram/Covariances: Var1 & Var1

Lag Size: 382010

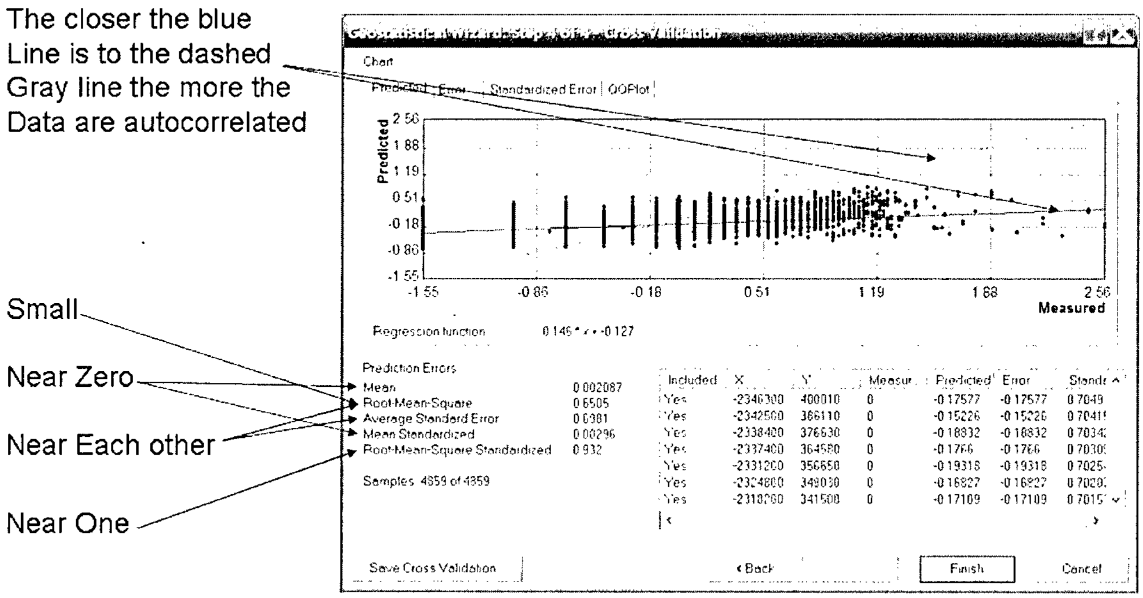
Number of Lags: 12

0.55378*Exponential(3927500)+1.413*Nugget

Figure 17: Semivariogram Example

Each kriging method was created by using the exponential semivariograms with no anisotropy or direction selected (see Figure 17). The search neighborhood included an eight-piece circle using 12 points for prediction with a minimum of 3 points per octant. In order to discuss which kriging method performed the best, the details will be discussed in detail in the following section.

The final step of ESRI's geostatistical analyst kriging method is a cross validation technique. The concept behind this technique is to remove each data point one at a time in order to predict its value, which is then compared to the measured value. Using the example in Figure 18, the cross validation output is



Regression function 0.146 * x + -0.127

Prediction Errors

| | |
|---|---|
| Mean | 0.002087 |
| Root-Mean-Square | 0.6505 |
| Average Standard Error | 0.6981 |
| Mean Standardized | 0.00296 |
| Root-Mean-Square Standardized | 0.932 |

Samples 4859 of 4859

| Included | X | Y | Measur | Predicted | Error | Stand |
|---|---|---|---|---|---|---|
| Yes | -2346300 | 400010 | 0 | -0.17577 | -0.17577 | 0.7049 |
| Yes | -2342500 | 386110 | 0 | -0.15226 | -0.15226 | 0.7041 |
| Yes | -2338400 | 376630 | 0 | -0.18832 | -0.18832 | 0.7034 |
| Yes | -2337400 | 364580 | 0 | -0.1766 | -0.1766 | 0.7030 |
| Yes | -2331200 | 356650 | 0 | -0.19318 | -0.19318 | 0.7025 |
| Yes | -2324800 | 349030 | 0 | -0.16827 | -0.16827 | 0.7020 |
| Yes | -2318700 | 341500 | 0 | -0.17109 | -0.17109 | 0.7015 |

http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?ID=2568&TopicName=Performing%20cross%2Dvalidation%20in%20[illegible]%20parameter%20selection[illegible]=4006&pid=2604

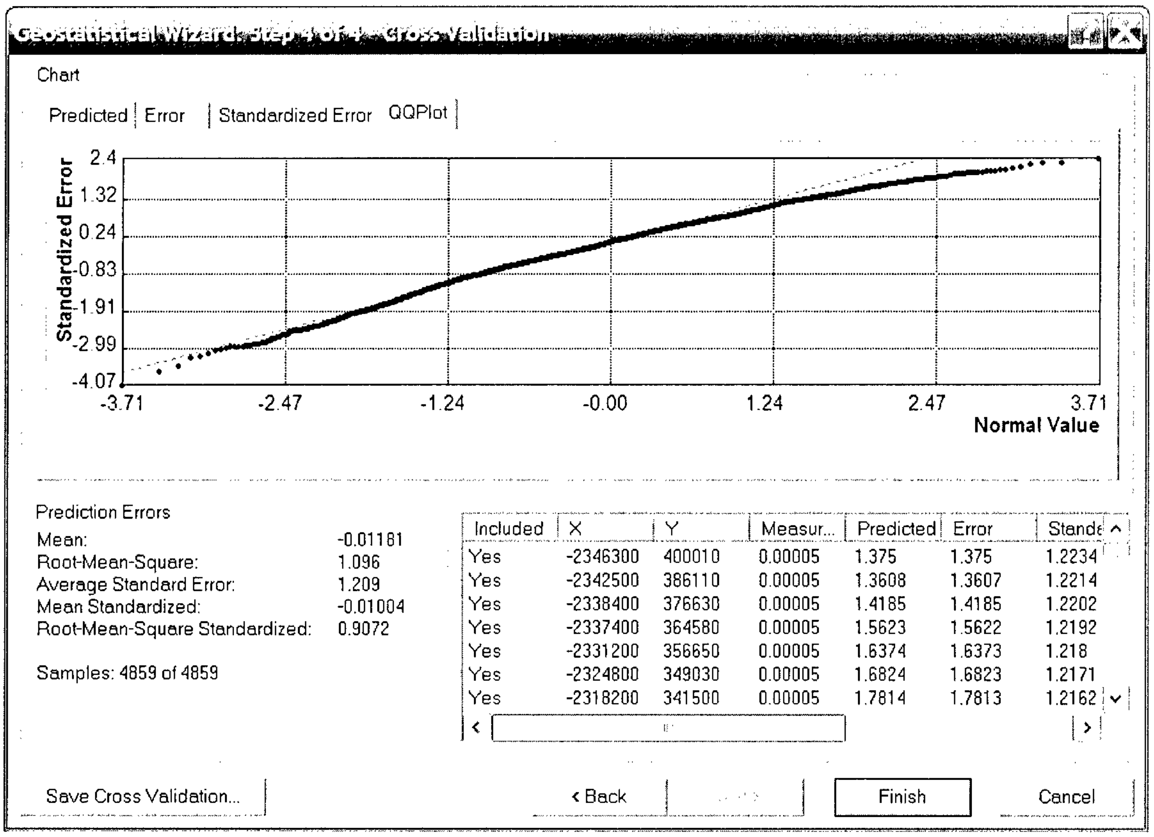
**Figure 18:** Cross Validation Output Example

interpreted by comparing multiple indicators. According to ESRI's technical documentation (Johnston, 2001) the blue line is the best fit line that runs through the data and it is optimal to have the points clustered close to the blue line. The closer the blue line is to the grey-dashed line the more spatial autocorrelated the data. A completely horizontal line would indicate that there was no spatial autocorrelation in the data. Next, both the mean prediction error and the mean standardized prediction error should be near zero, indicating that the errors are normally distributed. The root-mean-square prediction error and the average standard prediction error should be near one another, with the most optimal values approaching zero. Lastly, the root-mean-square standardized prediction error should be near one, indicating the kriging method performed well.

### Buffer Kriging Results

Beginning with the buffer DV1 dataset, the kriging method performed admirably as both the mean and mean standardized prediction errors were near zero and the root-mean-square and average standard error were near each other. Also, the root-mean-square standardized value was close to one. Lastly, the kriging's slope coefficient was .321, which is an admirable display of spatial autocorrelation. By reviewing the associated QQ plot (See Figure 18), it is evident that this kriging method had difficulty predicting the higher values within the data but performed well for those in the midrange (See Figure 18).

The weaknesses to the buffer DV1 kriging operation are that the root-mean-square and the average standard error, although close to one another, are not near zero. This indicates that when there is an error in predication, the predicted value is not close to the measured value. In regards to which areas this method predicted to have high number of recruits, it clearly highlighted the Midwest, Boston to Washington DC urban areas, the urban west coast areas, and



Prediction Errors

| | |
|---|---|
| Mean: | -0.01181 |
| Root-Mean-Square: | 1.096 |
| Average Standard Error: | 1.209 |
| Mean Standardized: | -0.01004 |
| Root-Mean-Square Standardized: | 0.9072 |

Samples: 4859 of 4859

| Included | X | Y | Measur... | Predicted | Error | Stande |
|---|---|---|---|---|---|---|
| Yes | -2346300 | 400010 | 0.00005 | 1.375 | 1.375 | 1.2234 |
| Yes | -2342500 | 386110 | 0.00005 | 1.3608 | 1.3607 | 1.2214 |
| Yes | -2338400 | 376630 | 0.00005 | 1.4185 | 1.4185 | 1.2202 |
| Yes | -2337400 | 364580 | 0.00005 | 1.5623 | 1.5622 | 1.2192 |
| Yes | -2331200 | 356650 | 0.00005 | 1.6374 | 1.6373 | 1.218 |
| Yes | -2324800 | 349030 | 0.00005 | 1.6824 | 1.6823 | 1.2171 |
| Yes | -2318200 | 341500 | 0.00005 | 1.7814 | 1.7813 | 1.2162 |

**Figure 19:** Buffer DV1 Kriging QQ Plot

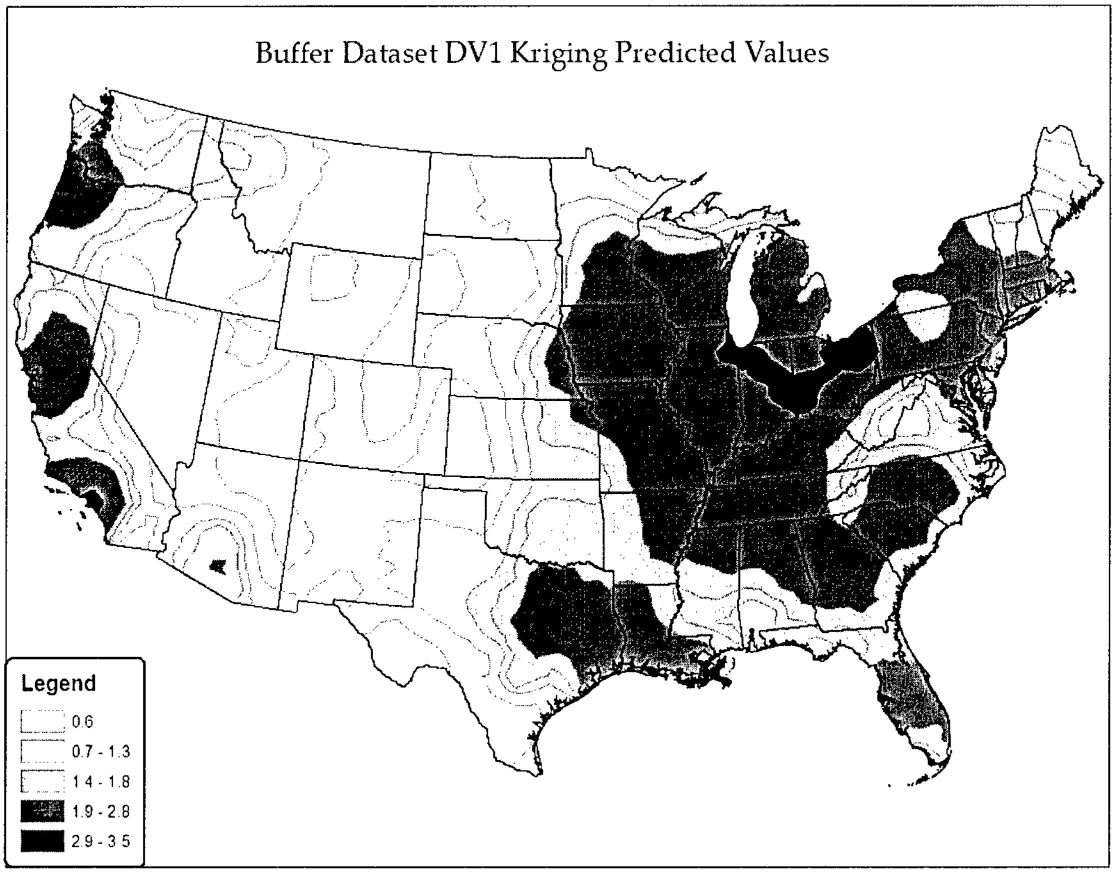
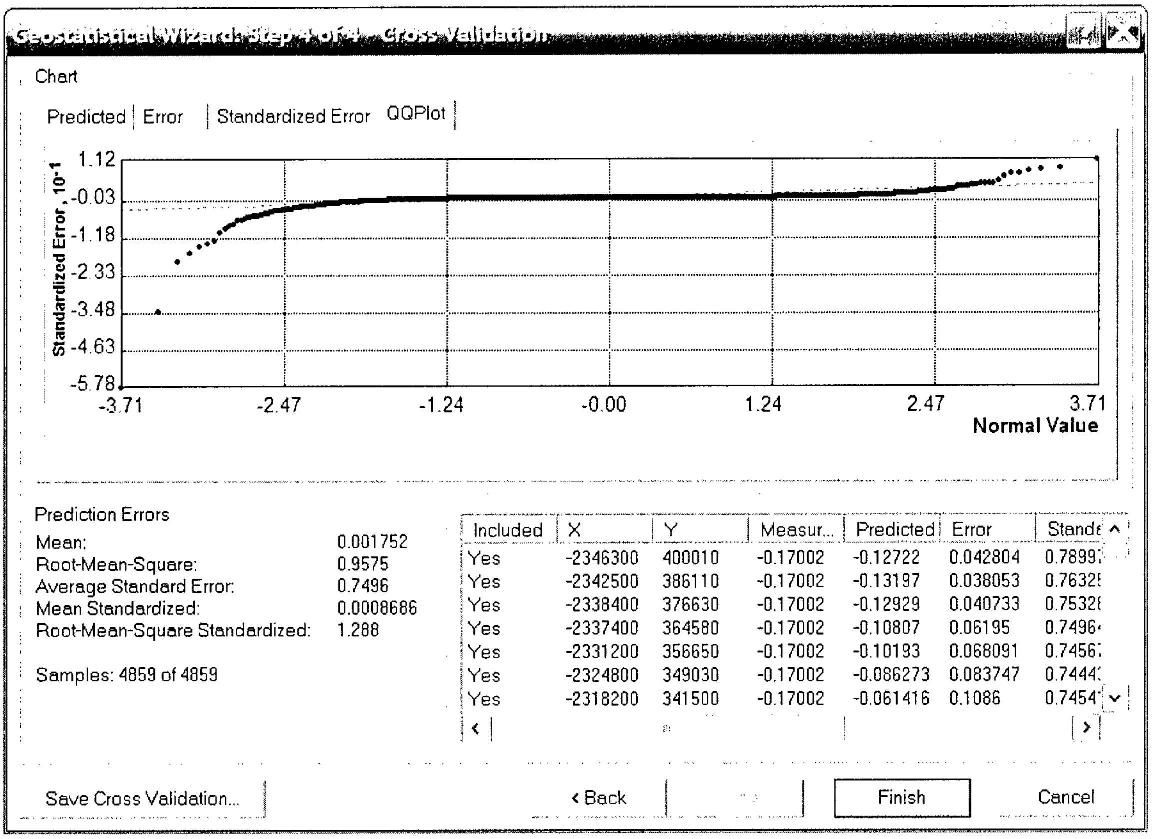


**Figure 20:** Buffer DV1 Kriging Prediction Map

portions of the South (See Figure 20). The low prediction areas were primarily the great plain states and the interior mountain regions.

In reference to the buffer DV2 dataset, the results were dissimilar to the DV1 data. Both the mean and mean standardized values were closer to zero than the DV1 dataset but the root-mean-square and average standard error scores were farther apart, which caused the root-mean-square standardized value to be larger than one. The slope coefficient was a paltry .151, less than half of DV1's coefficient, indicating a substantial departure of spatial autocorrelation. The QQ

Plot illustrates that the model had difficulty predicting the low areas and the high areas, but like DV1, did well in predicting those values in the middle (See Figure 21). Regardless of the weak results, this method did extrapolate four primary "hotspots" of recruit concentration which include the Los Angeles area, southeast Wyoming, and central Illinois. The low areas include the remainder of the study area (See Figure 22).



Prediction Errors

| | |
|---|---|
| Mean: | 0.001752 |
| Root-Mean-Square: | 0.9575 |
| Average Standard Error: | 0.7496 |
| Mean Standardized: | 0.0008686 |
| Root-Mean-Square Standardized: | 1.288 |

Samples: 4859 of 4859

| Included | X | Y | Measur... | Predicted | Error | Standa |
|---|---|---|---|---|---|---|
| Yes | -2346300 | 400010 | -0.17002 | -0.12722 | 0.042804 | 0.7899 |
| Yes | -2342500 | 386110 | -0.17002 | -0.13197 | 0.038053 | 0.7632 |
| Yes | -2338400 | 376630 | -0.17002 | -0.12929 | 0.040733 | 0.7532 |
| Yes | -2337400 | 364580 | -0.17002 | -0.10807 | 0.06195 | 0.7496 |
| Yes | -2331200 | 356650 | -0.17002 | -0.10193 | 0.068091 | 0.7456 |
| Yes | -2324800 | 349030 | -0.17002 | -0.086273 | 0.083747 | 0.7444 |
| Yes | -2318200 | 341500 | -0.17002 | -0.061416 | 0.1086 | 0.7454 |

**Figure 21:** Buffer DV2 Kriging QQ Plot

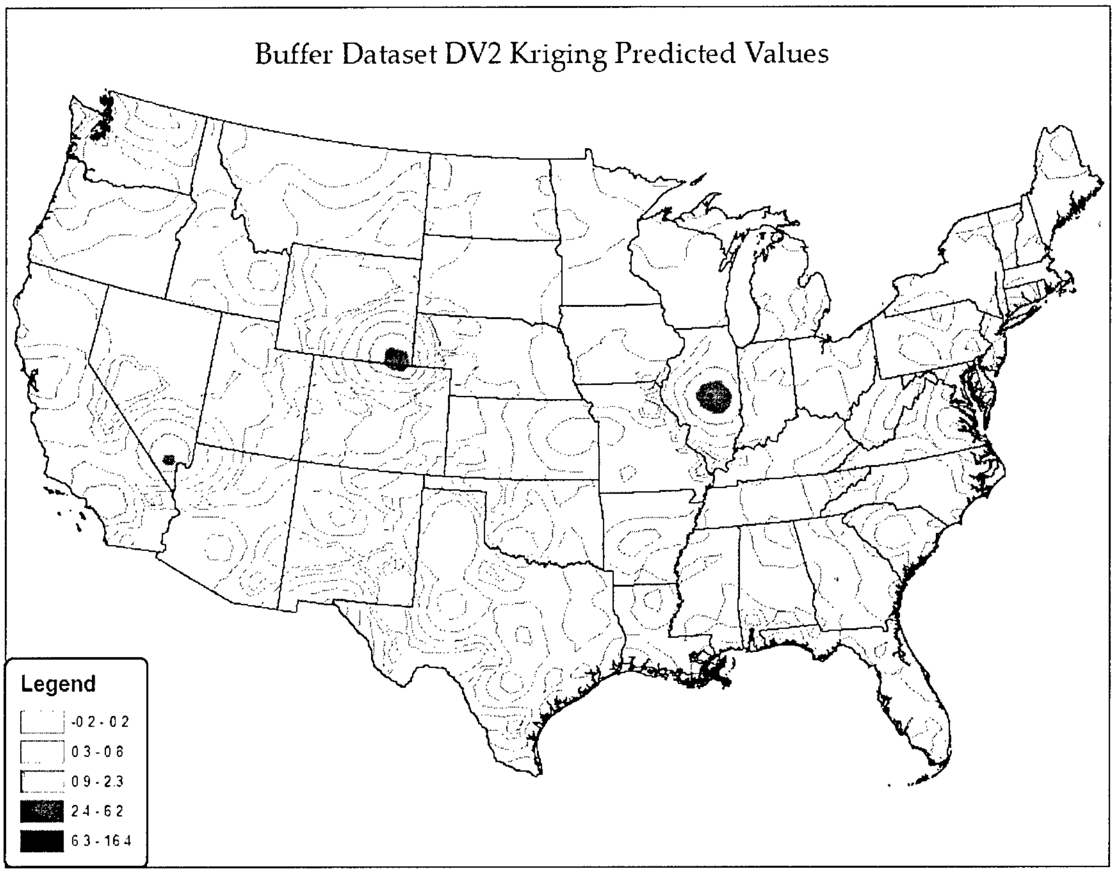
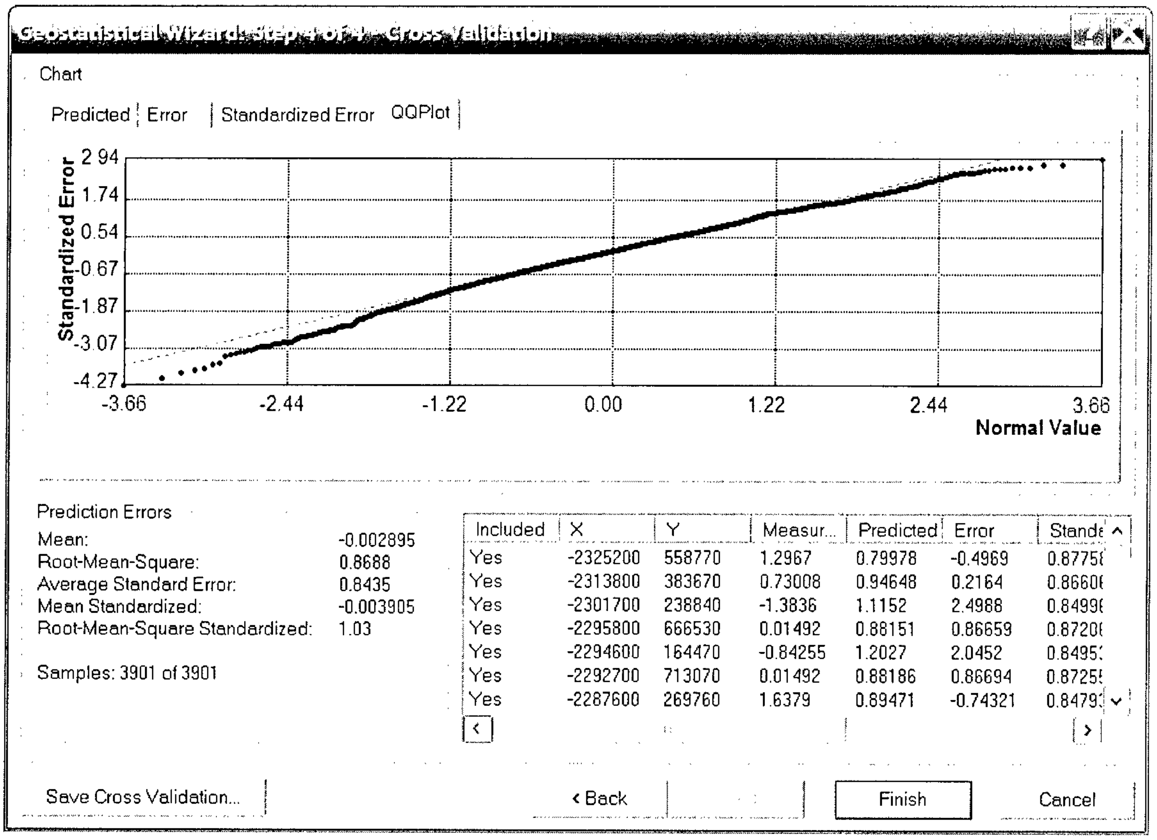


**Figure 22:** Buffer DV2 Kriging Prediction Map

## County Kriging Results

The county DV1 dataset's mean and mean standardized values were closer to zero and their root-mean square and average standard error scores were nearer each other than the buffer DV1 results; thus, causing the root-mean-square standardized value to be very close to one. These indications suggest the model performed exceedingly well in predicting the measured values; however, the slope coefficient, .238, did not measure up to buffer DV1's slope coefficient The

QQ Plot indicates that this model had difficulty predicting the low values and the extreme high values (See Figure 23). The county DV1 prediction results seem to mirror those of the buffer DV1 map, with the largest concentrations being on the California coast and the Midwest. However, the major difference is that the buffer highlighted the Midwest as the strongest region whereas the county dataset illustrates California as the highest recruitment contributor (See Figure 24).



Prediction Errors

| | |
|---|---|
| Mean: | -0.002895 |
| Root-Mean-Square: | 0.8688 |
| Average Standard Error: | 0.8435 |
| Mean Standardized: | -0.003905 |
| Root-Mean-Square Standardized: | 1.03 |

Samples: 3901 of 3901

| Included | X | Y | Measur... | Predicted | Error | Stand |
|---|---|---|---|---|---|---|
| Yes | -2325200 | 558770 | 1.2967 | 0.79978 | -0.4969 | 0.8775 |
| Yes | -2313800 | 383670 | 0.73008 | 0.94648 | 0.2164 | 0.8660 |
| Yes | -2301700 | 238840 | -1.3836 | 1.1152 | 2.4988 | 0.8499 |
| Yes | -2295800 | 666530 | 0.01492 | 0.88151 | 0.86659 | 0.8720 |
| Yes | -2294600 | 164470 | -0.84255 | 1.2027 | 2.0452 | 0.8495 |
| Yes | -2292700 | 713070 | 0.01492 | 0.88186 | 0.86694 | 0.8725 |
| Yes | -2287600 | 269760 | 1.6379 | 0.89471 | -0.74321 | 0.8479 |

**Figure 23:** County DV1 QQ Plot

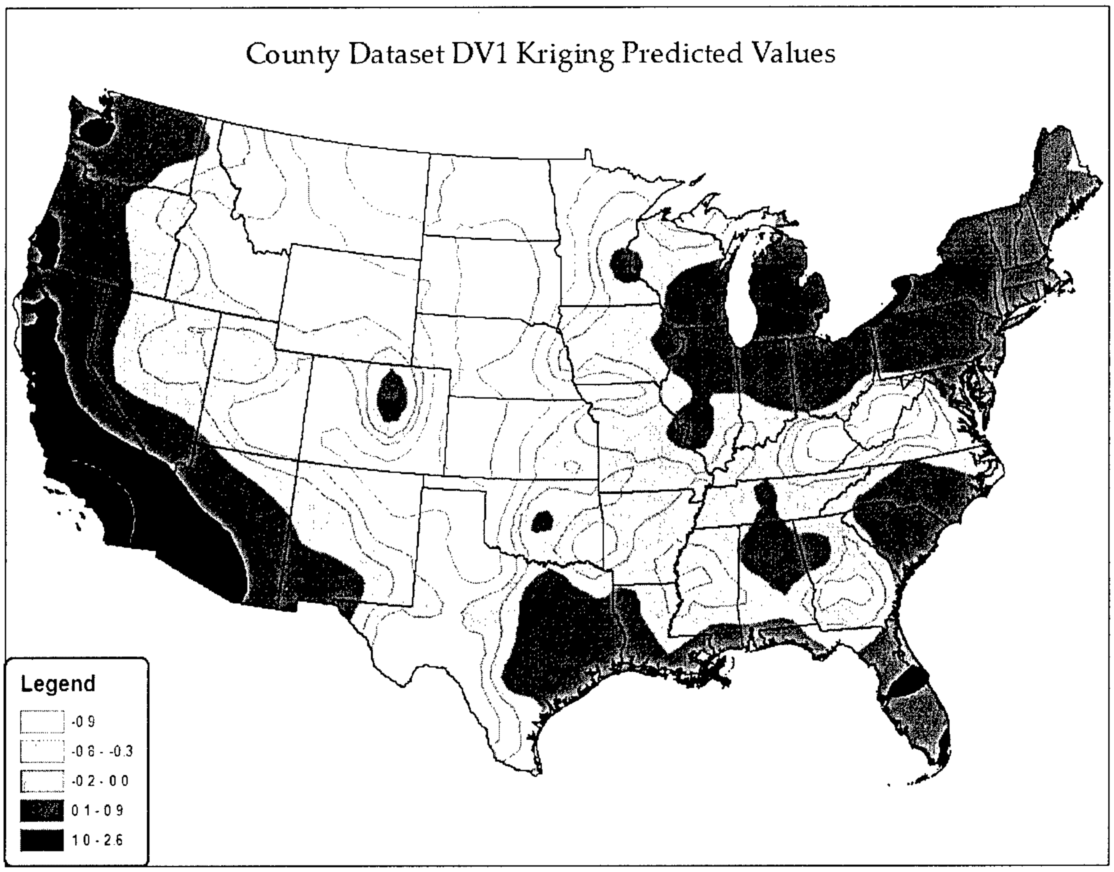
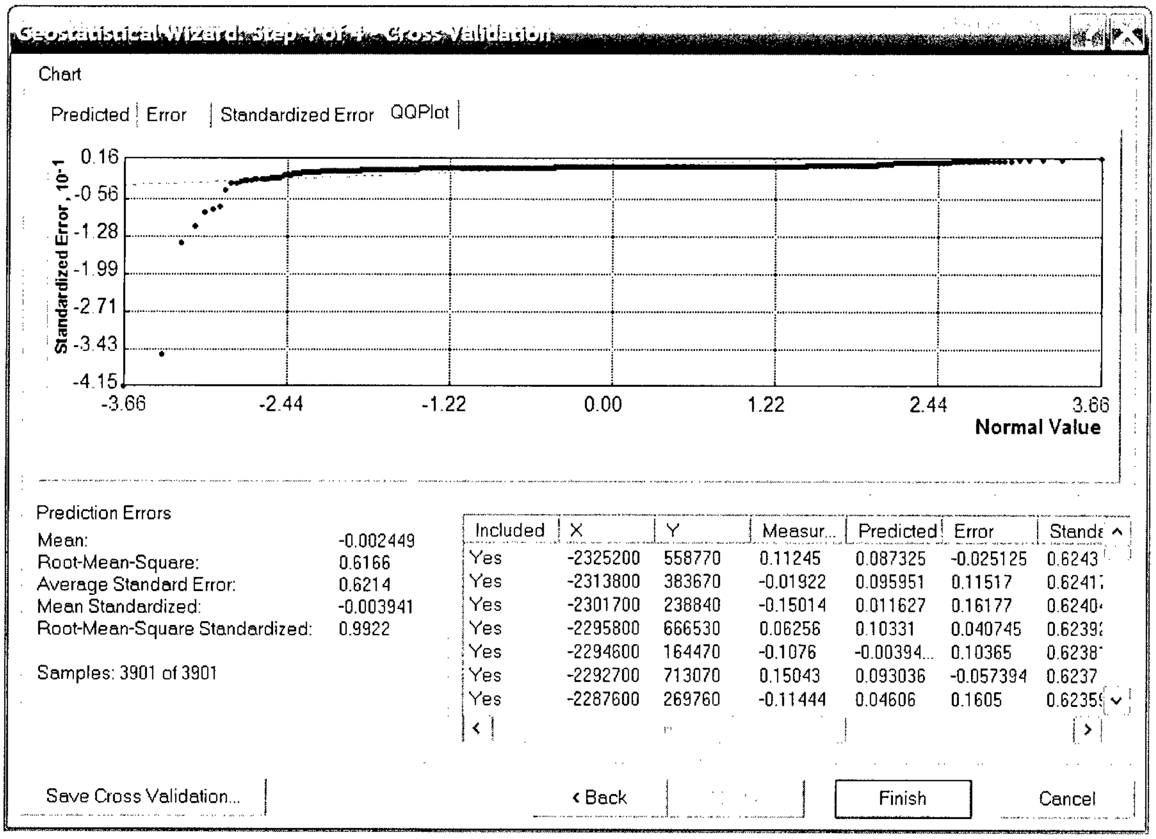


**Figure 24:** County DV1 Kriging Prediction Map

In terms of slope coefficient, the county DV2 model performed the worst of the four different models with a .045 slope coefficient, indicating nearly no spatial autocorrelation; however, in terms of the prediction errors, it performed the best. The mean and mean standardized scores were the closest to zero of any model, the root-mean-square and average standard error scores were closer to each other of the four, and it had the best root-mean-square standardized score (.9922) of them all. The QQ Plot indicates a rather shallow slope but most of the prediction plots are closely aligned with the prediction line with the exception of the extreme negative values, indicating this model had difficulty predicting the lowest values (See Figure 25). When looking at the prediction map (See Figure 26), the results generally claim the western United States stronger than the eastern, with the exception of a small pocket of high predictions around Washington D.C.



Prediction Errors

| | |
|---|---|
| Mean: | -0.002449 |
| Root-Mean-Square: | 0.6166 |
| Average Standard Error: | 0.6214 |
| Mean Standardized: | -0.003941 |
| Root-Mean-Square Standardized: | 0.9922 |

Samples: 3901 of 3901

| Included | X | Y | Measur... | Predicted | Error | Standa |
|---|---|---|---|---|---|---|
| Yes | -2325200 | 558770 | 0.11245 | 0.087325 | -0.025125 | 0.6243 |
| Yes | -2313800 | 383670 | -0.01922 | 0.095951 | 0.11517 | 0.6241 |
| Yes | -2301700 | 238840 | -0.15014 | 0.011627 | 0.16177 | 0.6240 |
| Yes | -2295800 | 666530 | 0.06256 | 0.10331 | 0.040745 | 0.6239 |
| Yes | -2294600 | 164470 | -0.1076 | -0.00394... | 0.10365 | 0.6238 |
| Yes | -2292700 | 713070 | 0.15043 | 0.093036 | -0.057394 | 0.6237 |
| Yes | -2287600 | 269760 | -0.11444 | 0.04606 | 0.1605 | 0.6235 |

**Figure 25:** County DV2 QQ Plot

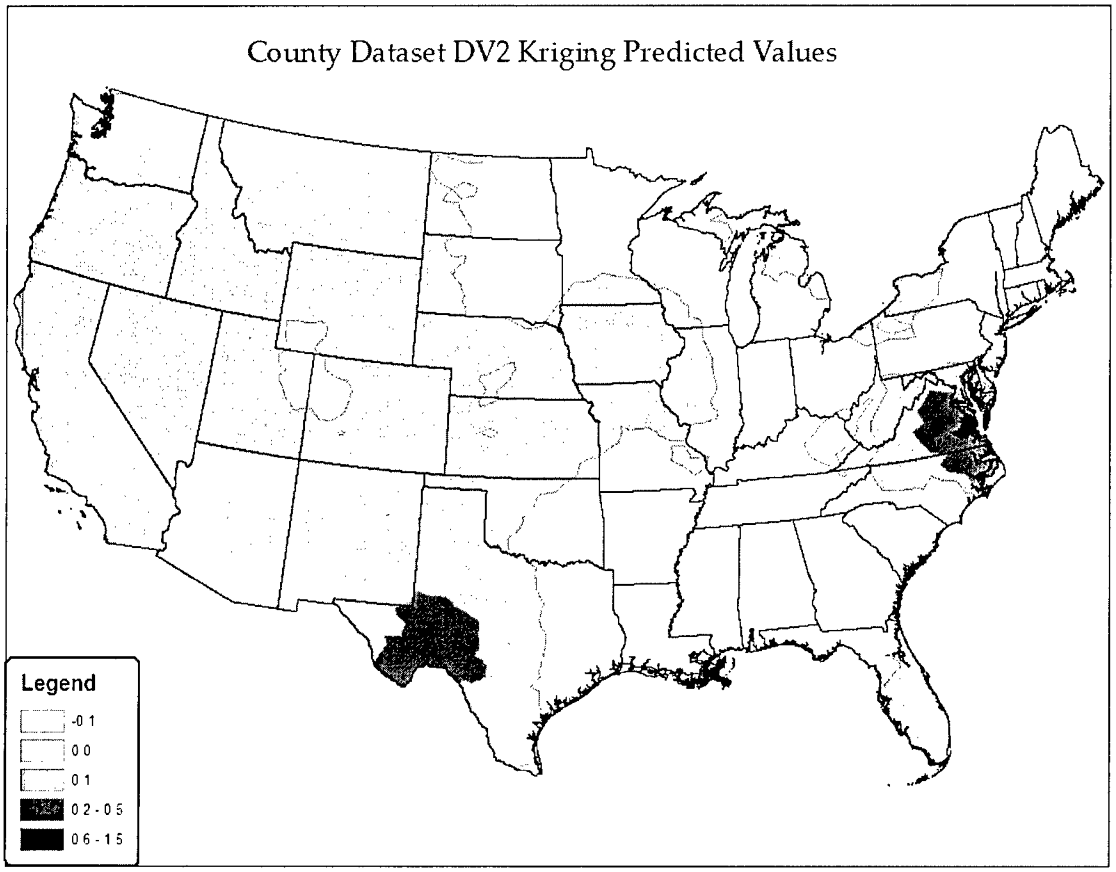


**Figure 26:** County DV2 Kriging Prediction Map

Ultimately, the results of the kriging operation closely resembled those of the Getis' Ord GI method, which produced areas of high and low values. However, because kriging assumes spatial autocorrelation where points are taken from a phenomenon that is continuous in space, the output for this study could be flawed. Considering these assumptions, the county DV2 model performed the best as root-mean-square and average standard error scores were closer to zero even though it had the worst slope coefficient.

## Geographic Weighted Regression

The final method used in this study was Geographic Weighted Regression, which is a regression technique developed by Stewart Fotheringham et al (2002) that takes into account local variation when predicting coefficients of determination. The concept that separates Geographic Weighted Regression from multiple linear regression is that multiple linear regression is a global model that treats all points are equal whereas Geographic Weighted Regression evaluates each point individually to determine the coefficient of determination for each location. Then, the overall adjusted R-Square average is assigned as the model's overarching coefficient of determination.

First, it is necessary to illustrate the framework used in the GWR model for duplication purposes. A Gaussian model with a fixed kernel and an AIC bandwidth minimization technique was used for each model. Once the model was ran, the model's adjusted R-Square, bandwidth, and independent variable significant scores were analyzed.

### Buffer GWR Results

The buffer DV1 dataset produced an adjusted R-Square value of .228, which was a significant improvement over the global .057 adjusted R-Square. All independent variables were found to be significant at the 99.9 percent level using a Monte Carlo significance test with the exception of Dist; hence, once again distance has proved to be an insignificant factor for predicting recruitment. When analyzing the residuals, the model largely over predicted the urban buffer units while predicting the rural areas better. The Buffer DV2 dataset did not fare as well. It produced a miniscule .027 adjusted R-Square which was only .006 points better than the global indicator. Furthermore, none of the independent variables were found to be significant at any level.

### County GWR Results

As expected, the county DV1 model performed the best overall with an adjusted R-Square .442 which was .142 points better than the global statistic. The Monte Carlo Significant test determined that the AD and PerCap variables were insignificant at any level and that the UnEmp variable was significant at the 95 percent level. All other variables were significant at the 99.9 percent level. Unlike the Buffer DV1 dataset, the residuals displayed a much more random pattern. The county DV2 only produced a .079 adjusted R-Square which was better than the global regression's .019 adjusted R-Square value. Like the buffer DV2 model, none of the variables were significant.

A look at the individual adjusted R-Square values for each dataset reveals which areas the models performed the best. For the buffer DV1 model, it performed the best in the desert Southwest, Floridian tip, the Great Plains, Rocky Mountain States, and portions of the Midwest (See Figure 27). It performed the weakest in the East Coast States. The buffer DV2 model illustrates that the weakest predictions were on the East Coast and gradually increase to the West with the strongest predictions being on the Pacific Coast (See Figure 28). The county adjusted R-Square exhibited very little pattern with the exception in the upper latitude areas that had high adjusted R-Squares whereas the rest of the study area displayed a random pattern (See Figure 29). The county DV2 pattern only re-emphasized the DV1 areas with the strongest areas being in the upper latitude states (See Figure 30).

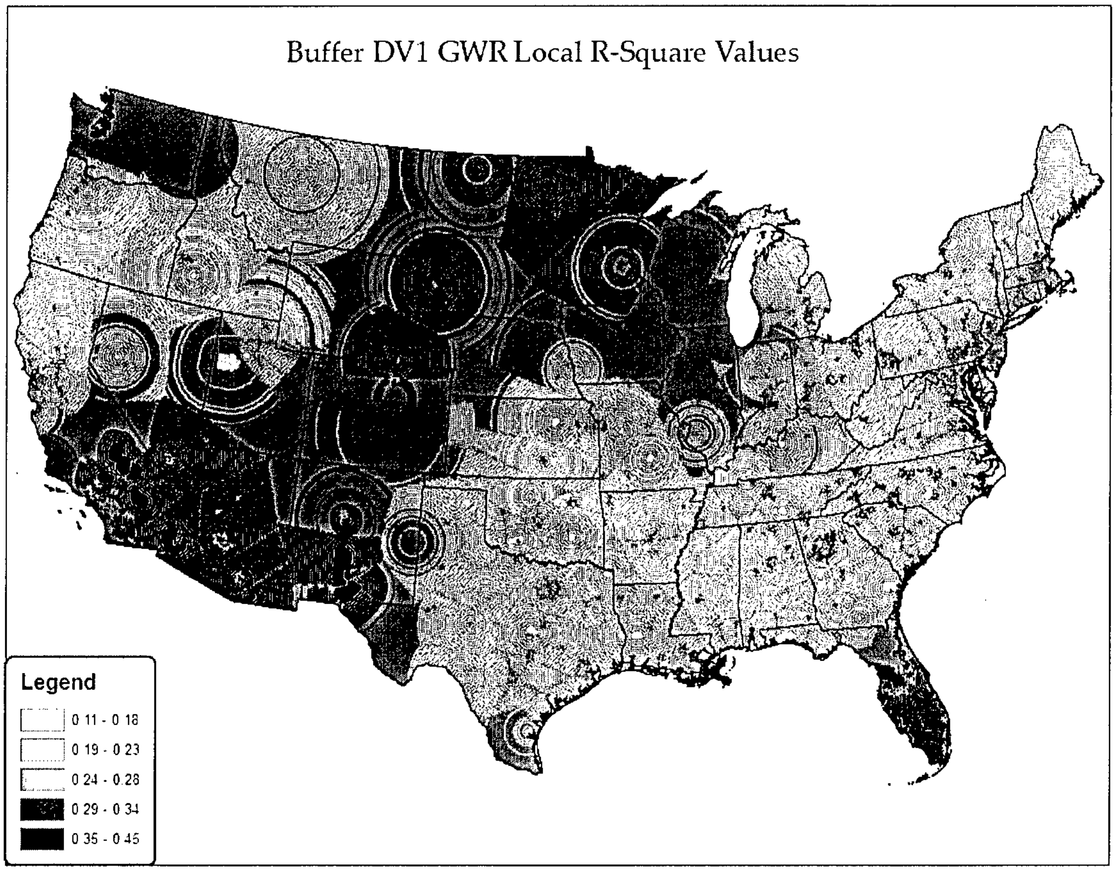


**Figure 27:** Buffer DV1 GWR Local R-Square Values

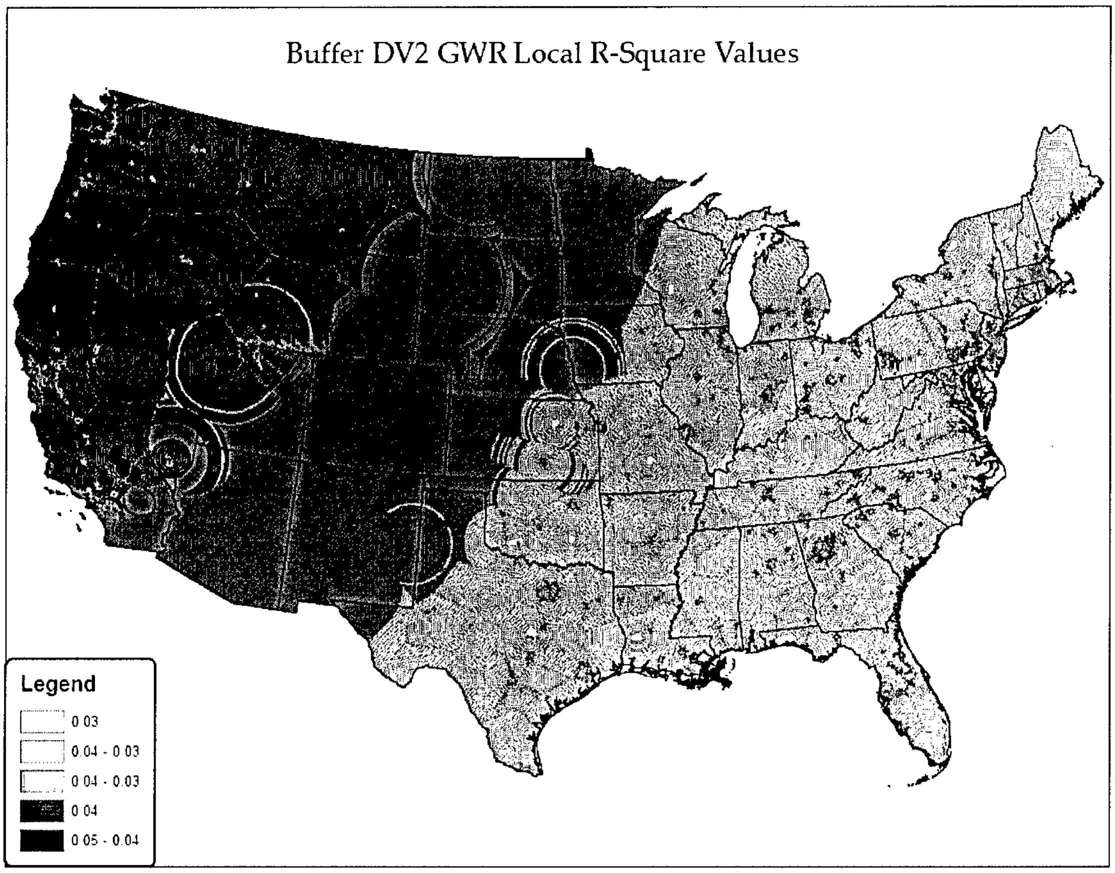


**Figure 28:** Buffer DV2 GWR Local R-Square Values

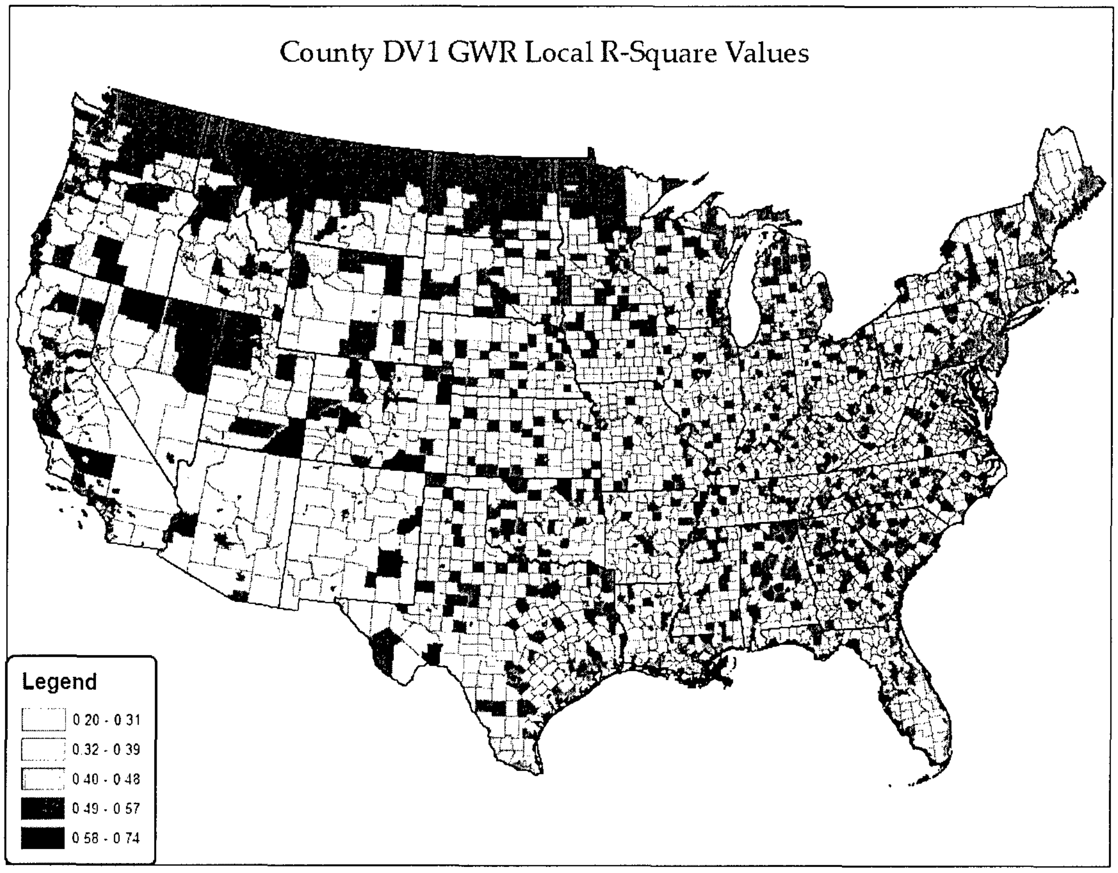


**Figure 29:** County DV1 GWR Local R-Square Values

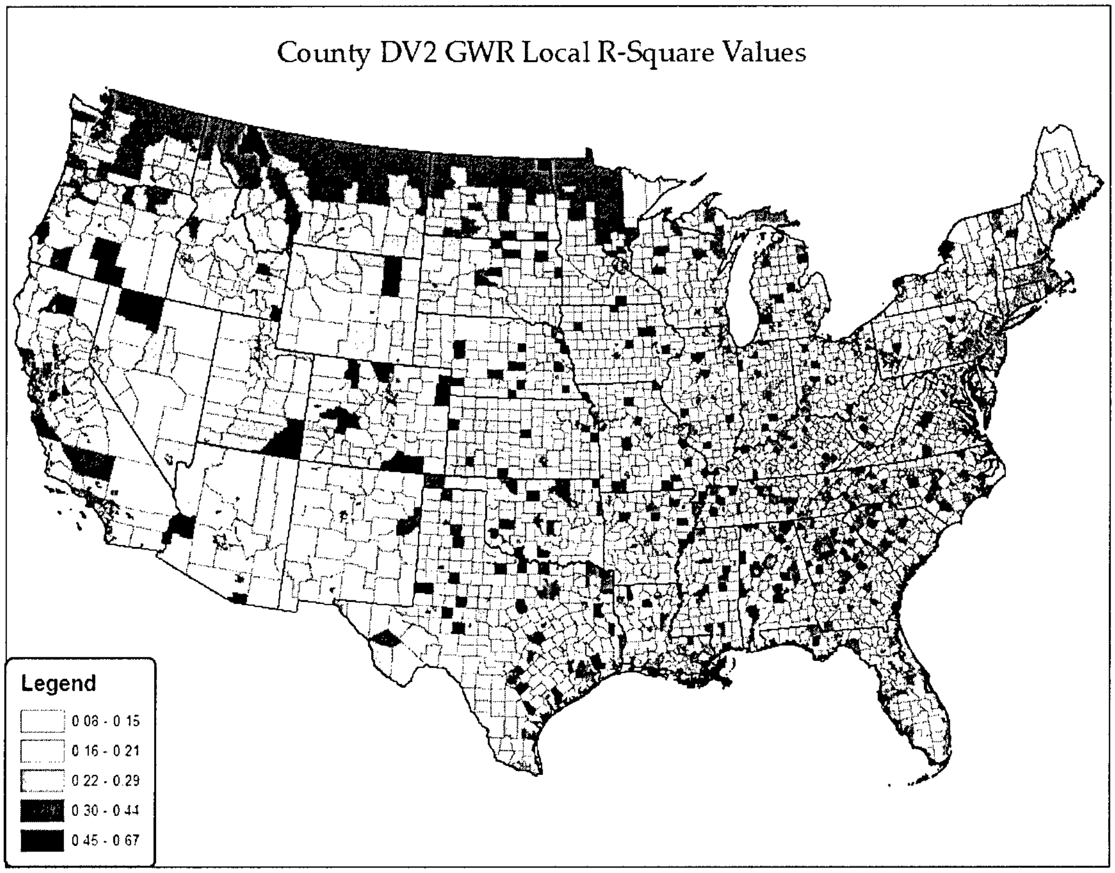


**Figure 30:** County DV2 GWR Local R-Square Values

Contrasting the prediction maps to the adjusted R-Square maps, the buffer datasets exhibited different spatial patterns between the two different maps but the county maps appear similar. The buffer DV1 extrapolated areas of high recruits being in the Midwest, upper Pacific coast and portions of the South while areas of low recruits pan through the great plain states. The DV2 model exhibited a general random pattern with no distinguishable pattern with highs and low areas being adjacent to one another throughout. Both of the county models produced predictive maps with random distribution pattern.

The GWR models appear to have performed best compared to the other models tested evidenced by the statistical indications with both buffer and county DV1 datasets performing better than their DV2 counterparts.

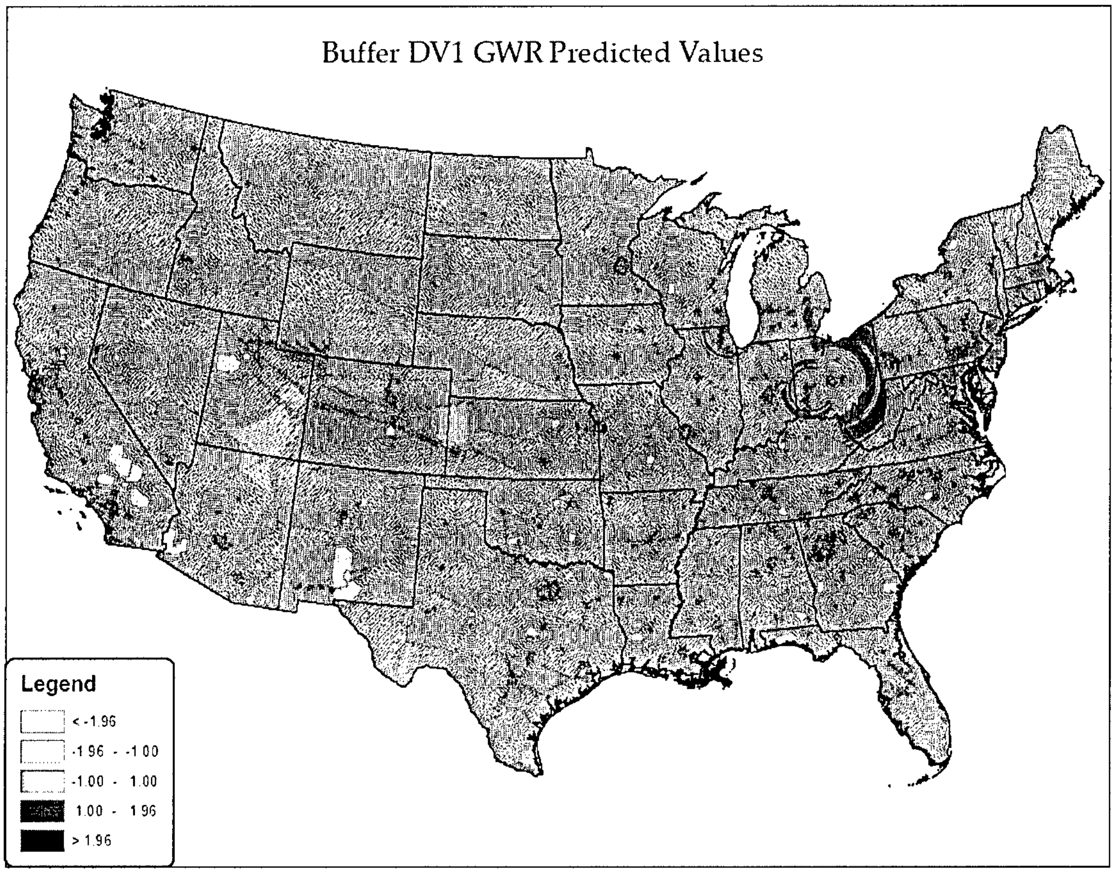


**Figure 31:** Buffer DV1 GWR Predicted Values

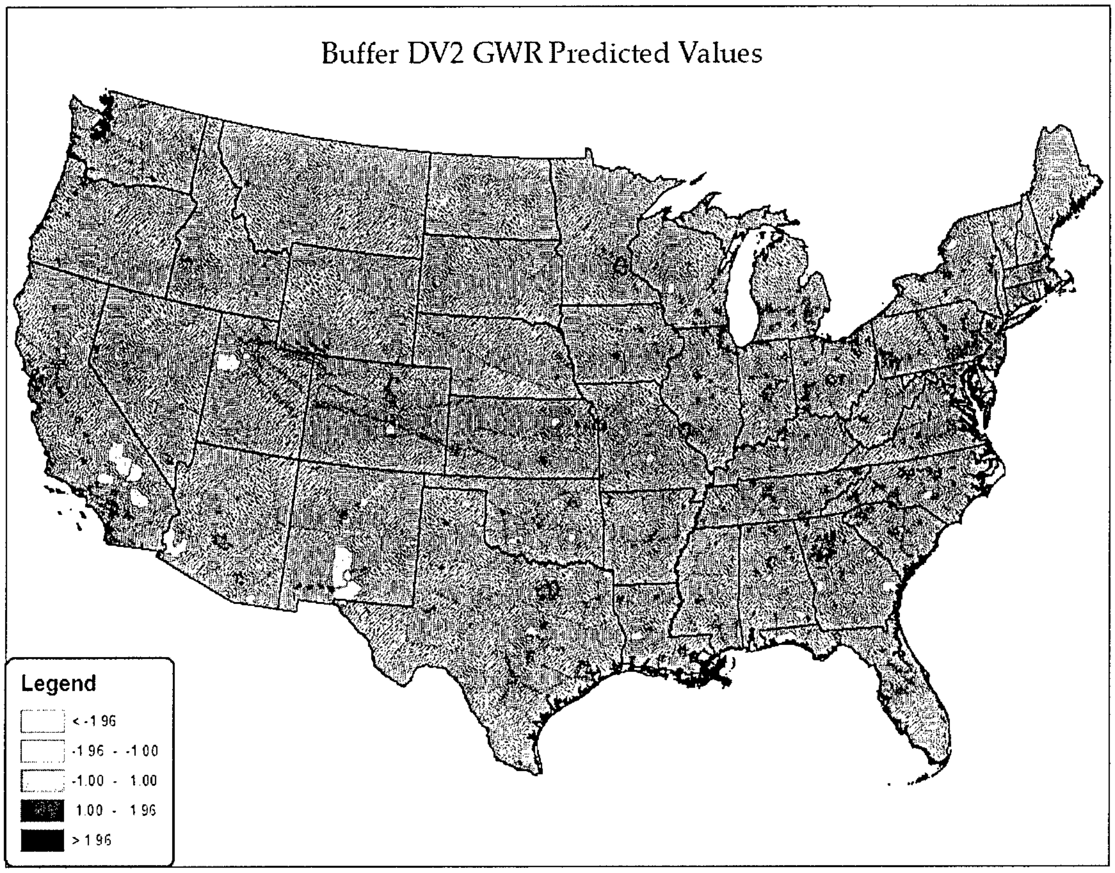


**Figure 32:** Buffer DV2 GWR Predicted Values

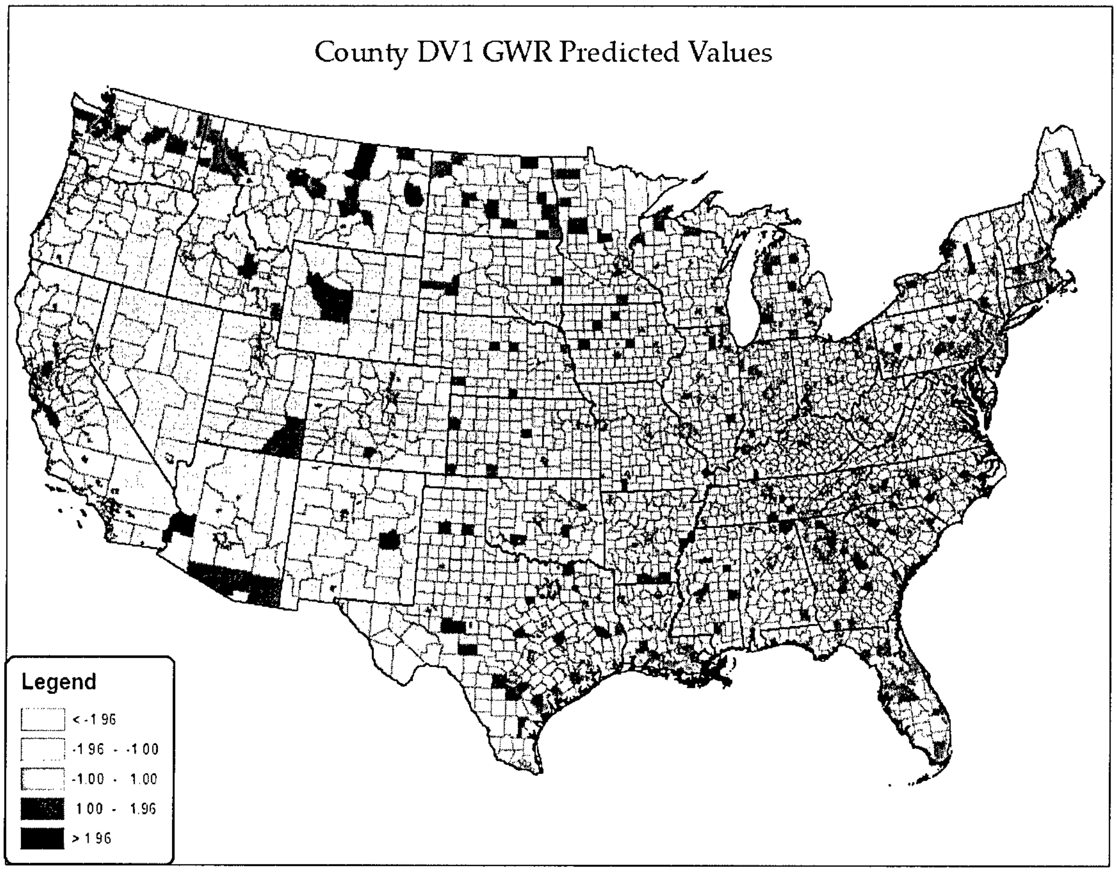


**Figure 33:** County DV1 GWR Predicted Values

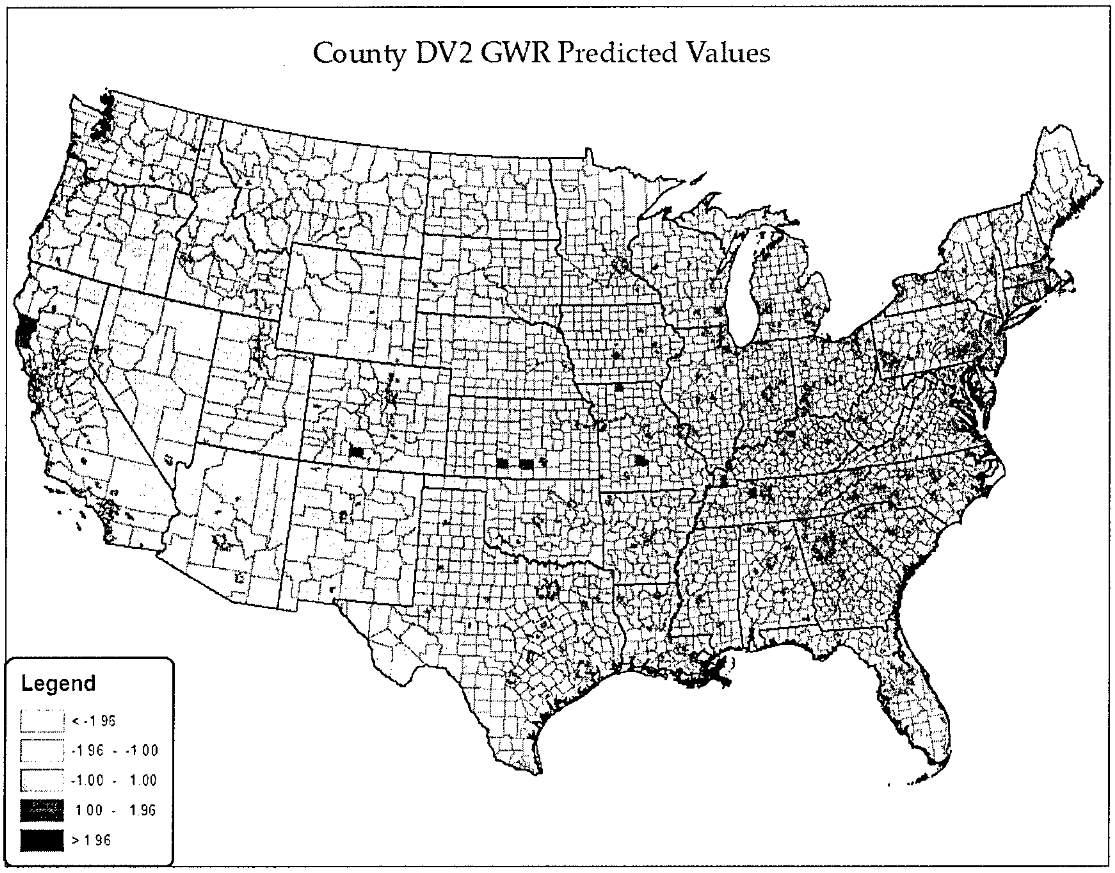


**Figure 34:** County DV2 GWR Predicted Values

## Chapter Summary

In retrospect, the spatially auto-correlated statistical techniques appear to have performed better than the multiple linear regression models by highlighting the local variation within the study area's extent that promote recruitment. This characteristic gives these techniques a competitive edge in creating a model to predict the locations where recruits originate. However, in terms of this study's goals, distance to a military installation has failed once again to prove its value as a recruitment predictor. Nevertheless, by relying solely upon the statistics to determine which model performed the best and not utilizing the models to predict recruitment would be an oversight.

# CHAPTER 6 – MODEL PERFORMANCE

Throughout this study the focus has been upon prediction. Therefore, instead of solely relying upon the statistical indicators to determine the best predictor, each model was used to predict the location of new recruits using the HoR2004 addresses. In order to accomplish this feat, the home address information was manipulated in the same fashion as the HoR0203 addresses. The predicted value results from the regression, factorial regression, Getis' Ord, kriging, and GWR models were then standardized in order to ensure commonality between all values. To determine how well the model predicted the FY2004 recruits, each model's predicted value was subtracted from the FY2004 value (measured value), leaving an error value. The model's variance statistic was then analyzed to determine how well the models performed, ultimately serving as the final discriminator as to which model performed the best (See Table 5).

Beginning with the buffer and county DV1 datasets, the county datasets performed better with the non-spatially correlated techniques and the GWR technique but the buffer dataset's local Getis' Ord and kriging techniques performed better. However, comparing the buffer and county DV2 models, the county dataset's methods outperformed the buffer dataset in every model.

**Table 5:** Performance Results

**County Descriptive Statistics**

| | N | Range | Minimum | Maximum | Sum | Mean | | Std | Variance |
|---|---|---|---|---|---|---|---|---|---|
| | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Statistic |
| Regression DV1 | 3,904 | 10.39 | -6.84 | 3.54 | 1.42 | 0.0004 | 0.01520 | 0.94981 | 0.902 |
| Factorial Reg DV1 | 3,904 | 13.75 | -10.90 | 2.84 | 1.42 | 0.0004 | 0.01614 | 1.00817 | 1.016 |
| Getis' Ord DV1 | 3,904 | 7.30 | -3.71 | 3.58 | 1.42 | 0.0004 | 0.01550 | 0.96871 | 0.938 |
| Kriging DV1 | 3,904 | 8.91 | -5.08 | 3.83 | 1.42 | 0.0004 | 0.01627 | 1.01650 | 1.033 |
| GWR DV1 | 3,904 | 5.71 | -3.35 | 2.36 | 1.42 | 0.0004 | 0.01298 | 0.81117 | 0.658 |
| Regression DV2 | 3,904 | 42.07 | -9.23 | 32.85 | -47.97 | -0.0123 | 0.01847 | 1.15424 | 1.332 |
| Factorial Reg DV2 | 3,904 | 40.52 | -7.56 | 32.96 | -47.97 | -0.0123 | 0.01868 | 1.16728 | 1.363 |
| Getis' Ord DV2 | 3,904 | 45.35 | -14.72 | 30.63 | -47.97 | -0.0123 | 0.01853 | 1.15749 | 1.340 |
| Kriging DV2 | 3,904 | 39.57 | -8.67 | 30.89 | -47.97 | -0.0123 | 0.01863 | 1.16414 | 1.355 |
| GWR DV2 | 3,904 | 52.79 | -21.78 | 31.01 | -47.97 | -0.0123 | 0.01698 | 1.06078 | 1.125 |
| Valid N (listwise) | 3,904 | | | | | | | | |

**Buffer Descriptive Statistics**

| | N | Range | Minimum | Maximum | Sum | Mean | | Std | Variance |
|---|---|---|---|---|---|---|---|---|---|
| | Statistic | Statistic | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic | Statistic |
| Regression DV1 | 4,862 | 14.38 | -8.36 | 6.03 | 0.00 | 0.0000 | 0.01805 | 1.25865 | 1.584 |
| Factorial Reg DV1 | 4,862 | 8.91 | -3.69 | 5.22 | 0.00 | 0.0000 | 0.01904 | 1.32777 | 1.763 |
| Getis' Ord DV1 | 4,862 | 7.85 | -3.78 | 4.07 | 0.00 | 0.0000 | 0.01258 | 0.87726 | 0.770 |
| Kriging DV1 | 4,862 | 7.79 | -3.25 | 4.54 | 0.00 | 0.0000 | 0.01344 | 0.93691 | 0.878 |
| GWR DV1 | 4,862 | 12.49 | -8.56 | 3.93 | 0.00 | 0.0000 | 0.01477 | 1.02976 | 1.060 |
| Regression DV2 | 4,862 | 53.34 | -7.39 | 45.96 | -0.01 | 0.0000 | 0.01930 | 1.34560 | 1.811 |
| Factorial Reg DV2 | 4,862 | 47.29 | -2.31 | 44.98 | -0.01 | 0.0000 | 0.01984 | 1.38368 | 1.915 |
| Getis' Ord DV2 | 4,862 | 52.81 | -12.61 | 40.20 | -0.01 | 0.0000 | 0.01726 | 1.20371 | 1.449 |
| Kriging DV2 | 4,862 | 68.06 | -21.55 | 46.51 | -0.01 | 0.0000 | 0.01781 | 1.24163 | 1.542 |
| GWR DV2 | 4,862 | 52.75 | -12.33 | 40.42 | -0.01 | 0.0000 | 0.01931 | 1.34671 | 1.814 |
| Valid N (listwise) | 4,862 | | | | | | | | |

When analyzing the performance differences between the aspatial methods and the spatial methods, the county DV1 methods produced mixed variance results. However, the buffer DV1 spatial methods performed much better than the aspatial methods. The same pattern that existed between the county and buffer spatial and aspatial DV1 models held true with the county and buffer spatial and aspatial DV1 models.

# CHAPTER 7 - CONCLUSION

In retrospect, the spatially auto-correlated statistical techniques performed better than the multiple linear regression models traditionally used in recruitment research when it comes to predicting where recruits originate. By focusing on the local instead of the global, recruitment prediction is given a competitive edge when it comes to finding the locations of new recruits. However, in terms of this study's question, distance to a military installation has failed to prove its value as a recruitment predictor.

Does distance to a military installation serve as a significant predictor of recruitment? This study failed to reject the null hypothesis and concludes that distance to a military installation does not serve well in predicting recruitment. However, even though the variable "distance" did not fare well statistically, the spatially auto-correlated techniques performed admirably compared to the non-spatially correlated counterparts, suggesting that distance does play a role indirectly.

## Discussion

The discussion session will be comprised of three sections. The first section addresses the buffer and county DV2 datasets and their relative poor performance in every statistical method. The second section will address the distance question by focusing on the buffer framework. Lastly, the final section will address the objective of predicting where recruits originate.

### Buffer and County DV2 Datasets

It is interesting that the DV2 variables did not perform well in this study, which is cause for some speculation for the reasons why. Perhaps one explanation lies with the technique of normalizing the independent variables. For the DV1 datasets, the population independent variables (race, veterans, and AD) were all normalized by the total population in the area, which was appropriate because the DV1 dependent variable included the raw number of recruits. On the other hand, for the DV2 variable, where the number of recruits was normalized by the 18-24 year old population, the population independent variables were still normalized by the entire population. Perhaps, if these independent variables were normalized by their associated 18-24 year old population, the performance of the model would increase. With the exception of this possibility, the reasons why the DV2 models performed so miserably elude me at this time.

### Why distance didn't make a difference?

As a result of the study's findings, a question needs to be asked. "Why did the buffer dataset perform so poorly?" The answer to this question may go unanswered for many years but there are many possible explanations. First and perhaps most apparent is the method which demographic data were placed into the buffer dataset. To best illustrate the problems with inherent with this technique, an example of the buffer units around Luke Air Force Base, Arizona will be utilized (see Figure 24). For the county dataset, Maricopa County's urban area contained the census information, which was different than the adjacent Maricopa County rural segment. However, due to the buffer dataset deriving its information from the county, adjacent buffer units within the county boundaries had similar information. In this example, the Maricopa County's urban 10 kilometer, 20 kilometer, 30 kilometer, 40 kilometer, 50 kilometer, 60 kilometer, and 70 kilometer information were identical. The same is true of the rural buffer segments within Maricopa County. Therefore, with no difference within the demographic information but different numbers of recruits, it would be extremely difficult for the statistical software to accurately infer predictions in the buffer model.

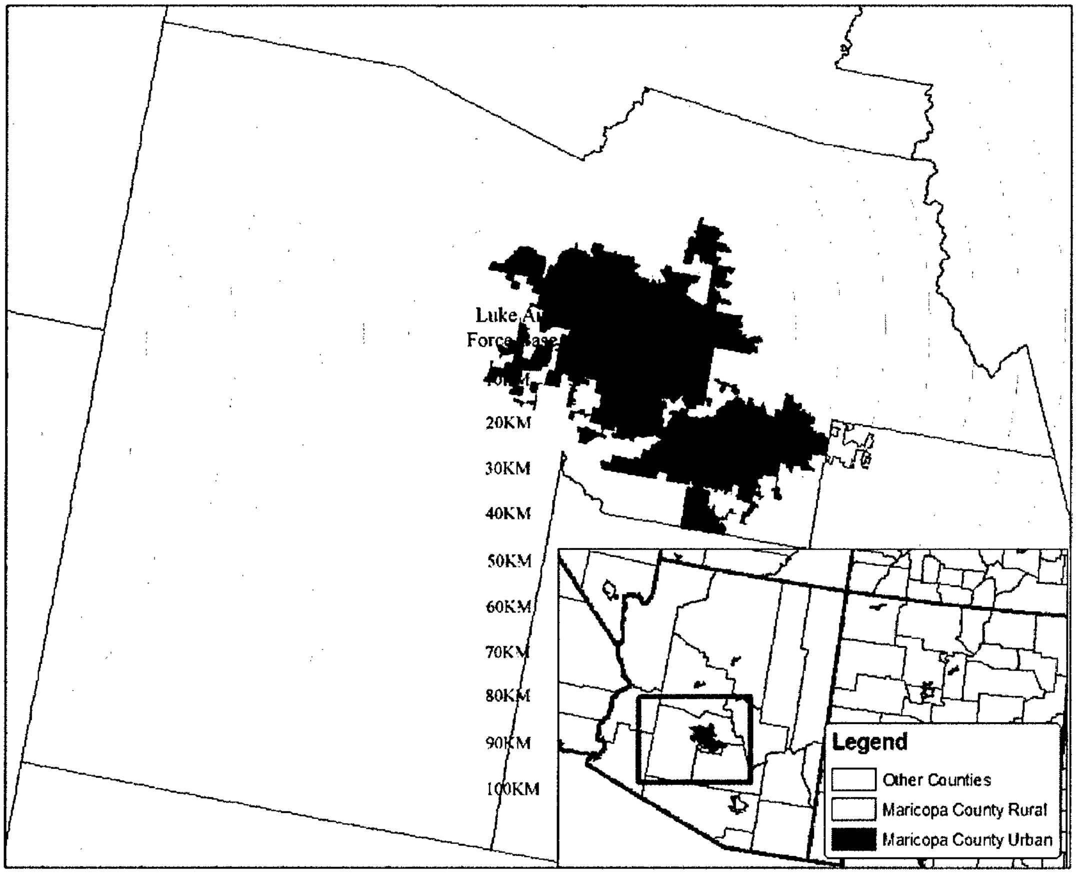


**Figure 24:** Maricopa County Buffer sections

Possibly another reason why the buffer dataset performed so poorly is the 10 kilometer framework. Although 10 kilometers was appropriate for this study given its hardware and software constraints, perhaps the influence a military installation exudes on the surrounding community is smaller than 10 kilometers. Therefore, a smaller buffer distance may illustrate the distance decay phenomena much clearer.

The final plausible reason for the buffer dataset's poor performance lies simply with the use of census data instead of the individual recruit's demographic information. Had the recruit's demographic information been available, it would have given two strengths to the study instead of causing a weakness. The first strength lies with the fact that the adjacent buffers would have been different instead of the same, as illustrated above; thus, giving the statistical software an advantage for making inferences about the dependent variable. The second strength with having the recruit's demographic information is that comparisons could be made against the census demographic information and the individual demographic information. These comparisons could show if the census information adequately represented those individuals joining the service.

### Recruit Location and Model Prediction

In reference to predicting where recruits originate, this section will discuss two methods: the buffer DV1 Getis Ord prediction model and the county GWR DV1 prediction model. By using only the recruit addresses, the buffer DV1 Getis Ord method displayed areas where recruits lived for the 2002/2003 dataset and also did an admirable job predicting the 2004 dataset's location. The prediction Texas, the Northeast, and the highest concentration in the Midwest (see Figure 25). Then, when testing this model's predictive capability, the variance of the

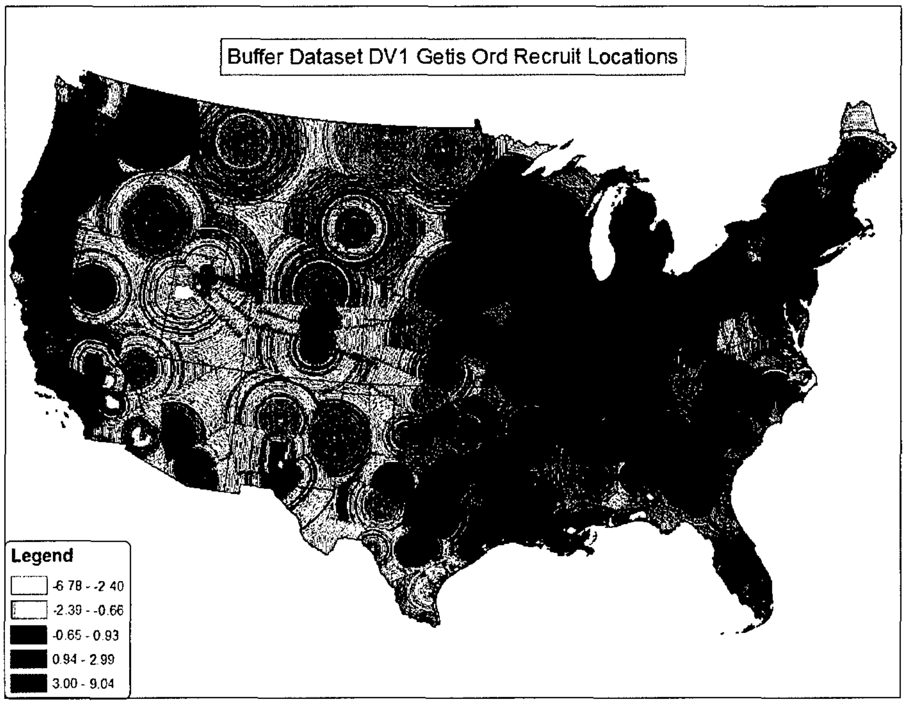


**Figure 25:** Getis Ord Prediction Model, showing areas of high and low recruits.

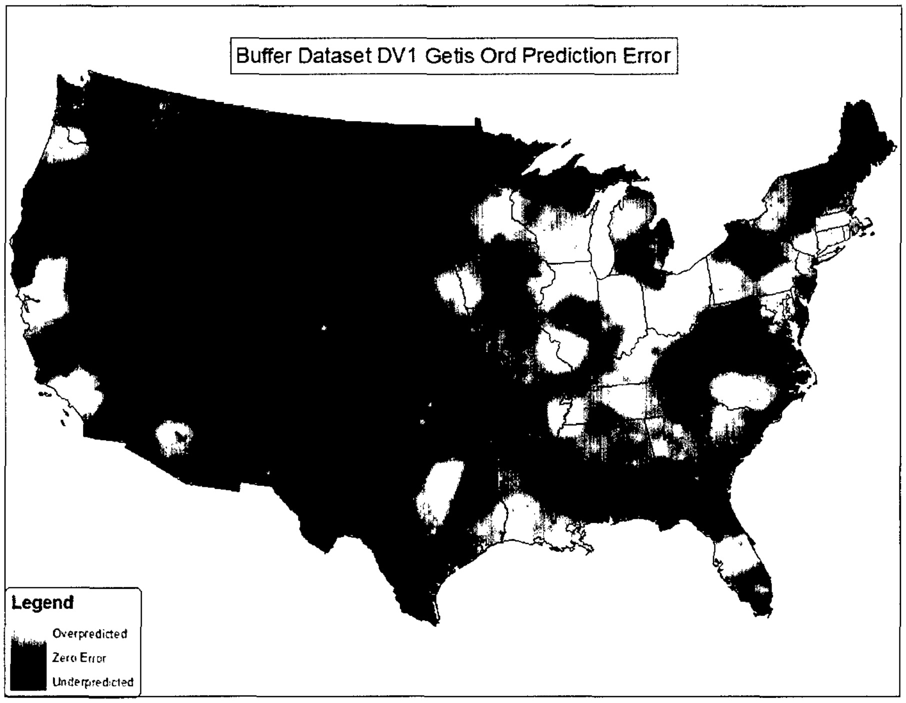


**Figure 26:** Getis Ord Performance Error, showing areas of over prediction error, no error, and under prediction error.

model's prediction error was .770. A look at the geographic variation of the error reveals which areas the model predicted well and which areas it did not. (see Figure 26). The majority of the error was in those areas that were included in the prediction's top and bottom twentieth percentile.

The county dataset's best predictor of recruitment produced different results than the buffer dataset. The GWR model predicted areas of high recruits and low recruits, but there exhibited no discernable pattern (see Figure 27). However, it must be noted that the GWR prediction's range was nearly one-third of the buffer dataset's. In terms of its performance error, the GWR model had difficulty predicting the same areas as the Getis Ord model; however, the GWR's prediction errors were reversed (see Figure 28).

Both the Getis Ord and the GWR error maps demonstrate one important message, "the location of some recruits can be predicted." Predicting locations of recruits instead of just the propensity of recruits gives the recruitment services a powerful tool, which can empower them to achieve their goals through an efficient and cost-effective means.

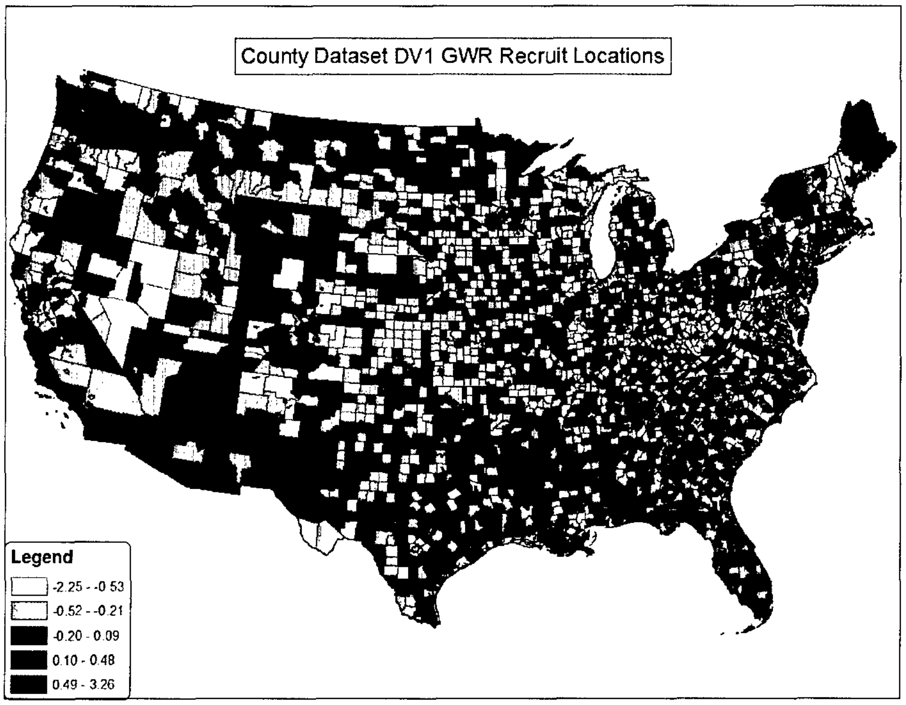


**Figure 27:** GWR Prediction Model, showing areas of high and low recruits.

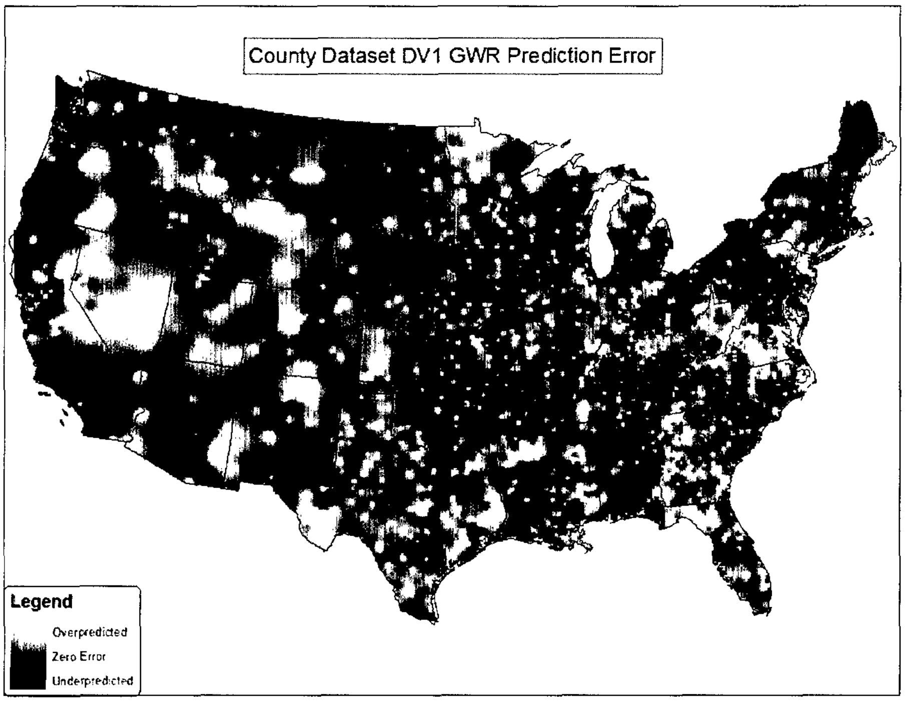


**Figure 28:** GWR Performance Error, showing areas of over prediction error, no error, and under prediction error.

### Propensity study comparison

When comparing this study's methods against the traditional propensity study methods, each method has its strengths and weaknesses. First, the propensity studies main strength is its length of service, for it can show trends over long periods of time. Another strength it possesses is that it also uses the demographic characteristics from the individuals. One of its weaknesses is that propensity is ineffective at predicting recruitment with 45 percent being its best predictability measure with another weakness is that it shows no geographic variation within its studies. By contrast, the GWR prediction model explained 44 percent of the variance within the data, essentially the same as the propensity studies, and showed geographic variation in the study area.

This study's methods have its strengths and weaknesses as well. Its strongest attribute is that it does show geographic variation in the number of recruits and the demographic attributes used to predict recruitment. Furthermore, an additional strength is that it improved upon previous research through its use of the GWR method in explaining the variation within the data. Its biggest weakness is that it used census demographic data to try to predict recruitment instead of using the individual recruit's information. The thought behind the census' use was trying to evaluate if different areas cultivated recruitment. In hind sight, focusing on the individual instead of the area would probably have yielded better results. The other main weakness to this study was the buffer framework as described in the above sections.

With the strengths and weaknesses of both methods being taken into consideration, what this study hopefully has shown is that geographic variation needs to be taken into consideration when predicting recruitment regardless of method. This is a necessity because there are differences in the population throughout the United States and it needs to be taken into consideration.

### Further Research

This study has broken away from the traditional means of recruitment research and attempted to look at recruitment from different perspective, the buffer framework. As with nearly any "new", it needs to be refined and improved upon in order to make it productive and accurate.

To begin with, it is important to note that the buffer framework shows its promise for an accurate prediction model for two reasons. The first reason is that in every method in this study, the buffer framework was able to clearly differentiate between different geographic regions, indicating a definite geographic component. The second reason is that by only using the coordinates of the recruits address and the Getis Ord method, it was able to predict recruits nearly as well as the county's GWR method, which shows its prediction ability.

The first improvement that needs to be made is that the individual recruit's demographic information must be utilized. Using this information will negate the weakness described above in the "why distance didn't matter" section. This change could effectively lead to showing that distance is a significant predictor of recruitment. However, as described above as well, perhaps the buffer distances need to be smaller, which is the second suggestion for improvement.

If the buffer distance was smaller, this could perhaps illustrate the military installations influence on its surrounding community better. However, in order to accomplish this technique, the researcher should ensure their hardware resources are capable of handling large statistical analyses that could potentially take weeks to finish its computations.

A method that would take into consideration the temporal aspects of the home of records, giving a stronger weight to more recent recruits and weaker weights to older recruit locations, would be an added benefit to prediction. By accounting for the temporal aspect in prediction, this would compensate for changes in demographic attitude patterns.

Another improvement to the demographic patterns lies with race categorization. Segal and Segal (2004) state that communities surrounding military installations are some of the least segregated in the county. With this in mind, instead of using individual race categories to predict recruitment, perhaps a measurement of segregation should be used. If a person feels at home in an un-segregated community, would that lead to feeling at home in the military?

Future research should implement veteran information that only include those veterans that served between the Korean War and the Persian Gulf War, as that roughly defines those individuals who would be in child-bearing age for an 18-24 year old recruit in 2006. Furthermore, if the recruits parent's veteran status can not be obtained, then the veteran information should be extrapolated at the census block group in order to refine the information, if hardware and software constraints can manage the large data sizes.

Lastly, if a method existed that could evaluate home of record locations without the need of a dependent variable, then perhaps this would be a feasible method for looking at this hypothesis in a new light. Near the end of this study, I explored a Nested Hierarchal Clustering technique with a software package named CrimeStat, which is designed to predict the location of crimes. This method did produce different areas that had significant clustering of recruits but due to time constraints the results of this method were not analyzed to the extent needed for publication.

Thirty years ago, propensity studies were recruited in order to assess youths' attitudes and their likelihood of enlisting into an all-volunteer force. These studies have served the country well in its long and distinguished history, but perhaps it is time to retire the propensity studies and recruit a new method which more accurately predicts who will enlist and where they will enlist. With transformation being at the forefront of the Air Force leaders agendas, now is the time to act. A time to implement a newer, leaner, more efficient way of evaluating recruitment, which could empowered the recruitment efforts to focus their limited recruitment resources in areas where they will receive a better return on their investment.

# References

Asch, B., J. A. Hosek, J. Arkes, C. C. Fair, J. Sharp, and M. Totten. 2002. *Military Recruiting and Retention after the Fiscal year 2000. Military Pay Legislation.* . RAND National Defense Research Institute.

Asch, B. J., C. Buck, J. A. Klerman, M. Kleykamp, and D. S. Loughran. 2005. *What Factors Affect the Military Enlistment of Hispanic Youth? A Look at Enlistment Qualifications.* RAND National Defense Research Instititute.

Bachman, J. G., D. R. Segal, P. Freedman-Doan, and P. M. O'Malley. 2000. Who Chooses Military Service? Correlates of Propensity and Enlistment in the U.S. Armed Forces. *Military Psychology,* Vol. 12, 1, pp. 1-30.

Cannizzo, P.A., 1995. *The Application of Geographic Information System Technology to United States Air Force Enlisted Recruiting: An Ohio Example.* Ohio State University.

Dandeker, C., A. Strachan. 1993. Soldier recruitment to the British Army: A spatial and social methodology for analysis and monitoring. *Armed Forces & Society.* Vol 19, 2, pp. 279-291.

Department of Defense. 2003. *Base Structure Report (A Summary of DoD's Real Property Inventory). Fiscal Year 2003 Baseline.* Office of the Deputy Under Secretary of Defense (Installations & Environment).

Eldridge, D. J. and J. P. Jones III. 1991. Warped Space: A Geography of Distance Decay. *Professional Geographer*, Vol. 43(4), pp. 500-511.

Fernandez, R. L. 1987. The Youth Population Decline and Prospects for Military Recruiting in the 1990s. Staff Working Paper. Congressional Budget Office, The Congress of the United States.

Fotheringham, S. A., C. Brunsdon, and M. Charlton. 2002. *Geographic Weighted Regression.* John Wiley & Sons Ltd. West Sussex, England, p. 24.

Fotheringham, S. A. 1981. Spatial Structure and Distance-Decay Parameters. *Annals of the Assoication of American Geographers*, Vol. 71., No. 3, pp. 425-436.

Fricker, R. D., and C. C. Fair. 2003. *Going to the Mines to Look for Diamonds. Experimenting with military Recruiting Stations in Malls.* RAND National Defense Research Institute.

GAO. United States Government Accountability Office. September 2005. *Military Personnel. Reporting Additional Servicemember Demographics Could Enhance Congressional Oversight.*

Goodchild, M. F. 2004. Social Sciences: Interest in GIS Grows. *ArcNews Online*, Spring. http://www.esri.com/news/arcnews/spring04articles/social-sciences.html

Last accessed: September 15, 2006

Jackson, L. C. 1999. *Recruiting Today's Youth: How Can We Increase youth Propensity to Join the Air Force during this Millennium.* Clark Atlanta University.

Johnston, K., J. M. Ver Hoef, and K. Krivoruchko. 2001. *Using ArcGIS Geostatistical Analyst.* ESRI Press, pp. 316.

Kane, T. 2005. *Who Bears the Burden? Demographic Characteristics of U.S. Military Recruits before and after 9/11. A Report of the Heritage Center for Data Analysis.* The Heritage Foundation. pp. 1-25.

Malinowski, J. C. and J. Brockhaus. 1999. Correlates of U.S. Army Recruiting Success in Texas. Paper presented at the Association of American Geographers' Annual Meeting, Honolulu, HI, March 1999.

Malinowski, J. C. 2005. Manning the Force: Geographic Perspectives on Recruiting. In *Military Geography: From Peace to War.*, eds. Palka, E. and F. A. Galgano. The McGraw-Hill Companies, Inc.

Olsson, G. 1970. Explanation, Prediction and Meaning Variance: An Assessment of Distance Interaction models. *Economic Geography,* Vol 46, p. 223

Segal, D. R. and M. W. Seagal. December 2004. America's Military Population. *Population Bulletin.* Vol. 59, No. 4. pp. 1-42.

Stewart, J. Q. 1941. An Inverse Distance Variation for Certain Social Influences. *Science,* Vol 93, No. 2404, pp. 89-90.

Tobler, W. 1970. A Computer Movie. *Economic Geography*, Vol. 46, pp. 234-240.