5-7-2005

A GIS Model for Predicting Potential "High Risk" Areas of West Nile Virus by Identifying Ideal Mosquito Breeding Habitats

Robert Charles Wallis

A GIS MODEL FOR PREDICTING POTENTIAL "HIGH RISK" AREAS OF WEST NILE VIRUS BY IDENTIFYING IDEAL MOSQUITO BREEDING HABITATS

By

Robert Charles Wallis

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Geosciences
in the Department of Geosciences

Mississippi State, Mississippi

May 2005

A GIS MODEL FOR PREDICTING POTENTIAL "HIGH RISK" AREAS OF WEST NILE VIRUS BY IDENTIFYING IDEAL MOSQUITO BREEDING HABITATS

By

Robert Charles Wallis

____
William H. Cooke
Assistant Professor of Geosciences
(Director of Thesis)

____
John C. Rodgers, III
Assistant Professor of Geosciences
(Committee Member)

____
John E. Mylroie
Professor of Geosciences
Graduate Coordinator of the
Department of Geosciences
(Committee Member)

____
Philip B. Oldham
Professor and Dean of the College of
Arts and Sciences

Name: Robert Charles Wallis

Date of Degree: May 7, 2005

Institution: Mississippi State University

Major Field: Geosciences

Major Professor: Dr. William H. Cooke

Title of Study: A GIS MODEL FOR PREDICTING POTENTIAL "HIGH RISK" AREAS OF WEST NILE VIRUS BY IDENTIFYING IDEAL MOSQUITO BREEDING HABITATS

Pages in Study: 107

Candidate for Degree of Master of Science

West Nile virus has become a major risk to humans since its first appearance in New York City in 1999. Physicians and state health officials are interested in new and more efficient methods for monitoring disease spread and predicting future outbreaks. This study modeled habitat suitability for mosquitoes that carry West Nile virus. Habitat characteristics were used to derive risk maps for the entire state of Mississippi. Statistical significance tests yielded objective evidence for choosing among many habitat variables. Variables that were significantly correlated with diagnosed human cases for 2002 were combined in weighted linear algebraic models using a geographic information system (GIS). Road density, slope, and summer precipitation minus evaporation (P-E) were the most significant variables. GIS-based model results were compared with results from logistic regression models. The algebraic model was preferred when validated by 2003 human cases. If adopted, GIS-based risk models can help guide mosquito control efforts.

ACKNOWLEDGEMENTS

I would like to thank my committee members, Dr. John Mylroie and Dr. John Rodgers, for their help in completing this study. I would also like to thank Dr. Pat Gerard for his statistical analysis and Mr. John Cartwright for all of his help in formatting the figures and graphs. Finally, I want to especially thank my major professor, Dr. William Cooke, for all of his time, effort, collaboration, and guitar tips and tricks that eased the tension and helped in the completion of this project.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

TABLE		Page

LIST OF FIGURES

FIGURE		Page

CHAPTER I

INTRODUCTION

Computer technology has drastically improved over the last decade. This fact is best illustrated by the introduction of the modern, graphical user interfaced, geographic information system (GIS). A GIS is defined by ESRI (2005):

> an arrangement of computer hardware, software, and geographic data that people interact with to integrate, analyze, and visualize the data; identify relationships, patterns, and trends; and find solutions to problems. The system is designed to capture, store, update, manipulate, analyze, and display the geographic information. A GIS is typically used to represent maps as data layers that can be studied and used to perform analyses.

Geographic information systems are quickly becoming an important tool across multiple disciplines. Epidemiological research provides an excellent framework for the implementation of geo-spatial technologies. Physicians and state health officials are interested in new and more efficient ways to monitor current diseases and predict future outbreaks. This is where GIS can help.

This study attempts to predict mosquito habitat suitability and/or potential risk of West Nile virus for the entire state of Mississippi (Figure 1) by testing the usefulness of environmental variables in a predictive modeling scenario. The project relates mosquito habitat to general public risk in Mississippi from West Nile virus and specifically to natural resource managers and users of recreational facilities. Human case data for 2002 are used as the basis for modeling risk and human cases recorded in 2003 are used to validate the model results.

Previous studies that were designed to assess vectored disease risk, Malaria and Lyme disease for example, applied environmental variables in heuristically-based models (Glass et al., 1994; Beck et al., 1994; Nicholson and Mather, 1996). This heuristically-based, "seat of the pants," approach to modeling can be improved upon by thoroughly investigating each variable of interest in order to determine variable importance.

For this study, determination of variable significance and variable weights were investigated by two approaches: a process of argument and consensus building among 'experts' of diverse backgrounds and education, and a deterministic algorithmic approach with variable weights assigned through probability-based statistics (t-tests) followed by logistic regression.

Pertinent information about West Nile virus, mosquito biology, and previous modeling efforts are included as background information below. Methods used to develop the deterministic algorithmic models are discussed in the following chapter. Visual analysis of the spatial distribution of West Nile virus occurrences along with model output and predicted risk are also presented.

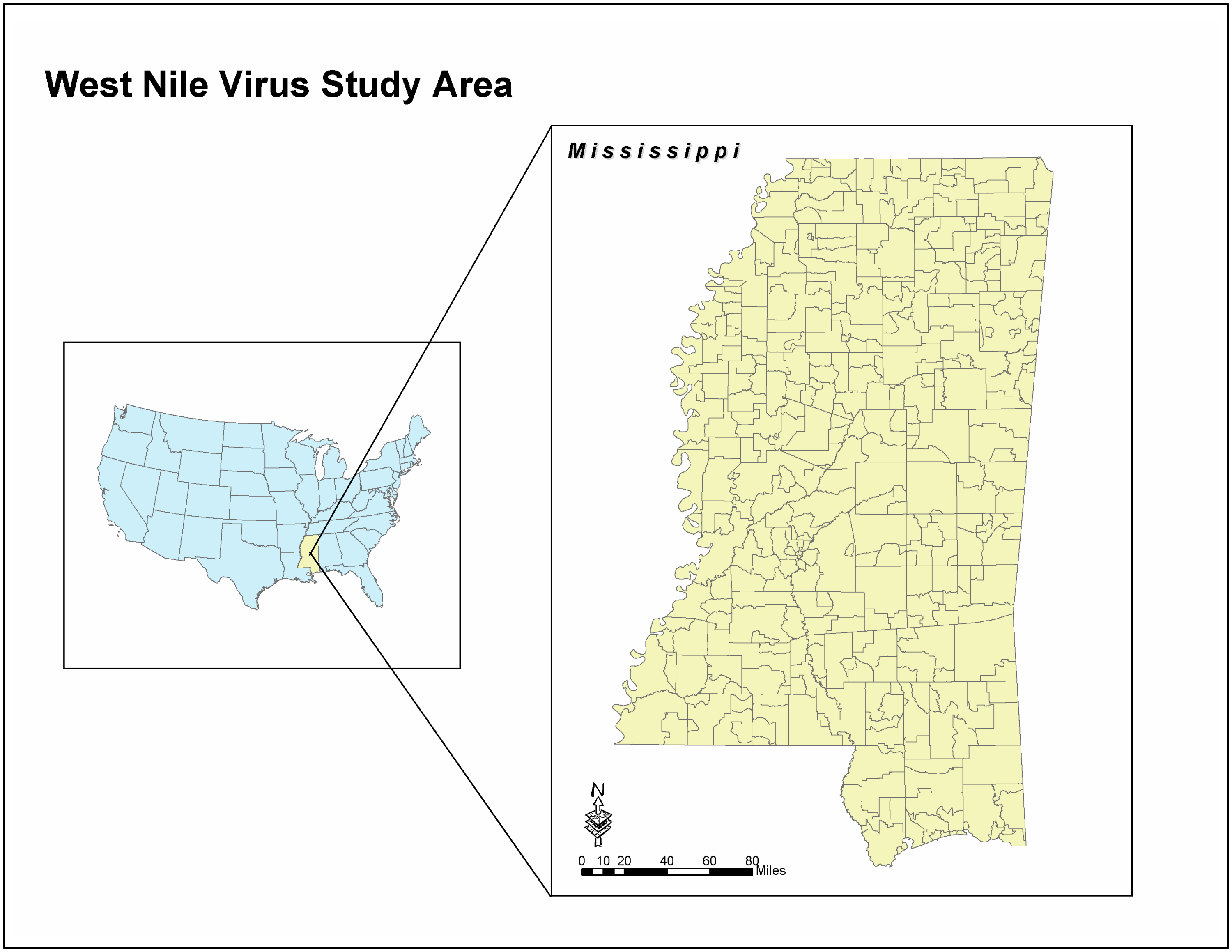

Figure 1: Map of Study Area

CHAPTER II

BACKGROUND INFORMATION AND LITERATURE REVIEW

This chapter examines background information and published or existing studies that use geographic information systems (GIS) to model biological systems for disease risk prediction. Focus is directed towards understanding how GIS has been used in epidemiological studies as well as the biology of mosquitoes and the West Nile virus. Disease and mosquito biology, GIS and biological systems, and the use of GIS in past and present modeling efforts to combat the virus are reviewed.

Disease Biology

West Nile virus, a member of the family Flaviviridae (genus Flavivirus), was first isolated in the West Nile district of Uganda in 1937 (Petersen and Roehrig, 2001; Guharoy et al., 2004; Chowers et al., 2001). It was identified from the blood of a febrile woman whose only known symptom was fever.

Sixty-two years later in 1999, the first U.S. case of West Nile virus was reported in New York City (Gea-Banaclocche et al., 2004; Peterson and Marfin, 2002; Guharoy et al., 2004; Petersen and Roehrig, 2001). "Within the past five years, West Nile virus has emerged as an important human, avian, and equine disease in the United States" (Guharoy et al., 2004, p.1235). The virus has spread rapidly, resulting in numerous human cases and several deaths. Every state, excluding Alaska and Hawaii, has reported an occurrence of West Nile virus. "In 2002, there were 4156 human cases with 284 deaths. In 2003, there were more than 9000 human cases and 220 deaths" (Gea-Banaclocche et al., 2004). In 2004, there were 2470 human cases and 88 deaths (CDC, 2005). These numbers indicate trends that lead to speculation. For example, it seems that in 2002, when the virus was first introduced, accurate diagnosis was difficult. Further, by 2003 the threat of the disease was known; as a result, everyone that showed symptoms resembling those of West Nile virus was probably diagnosed as having West Nile virus. Therefore, deaths decreased as a result of this inclusive diagnosis. Also, due to media attention, the general public began to take action by avoiding exposure to the most opportune times and places for mosquito contact. Finally, by 2004 it seems that doctors had become more efficient and accurate at diagnosing West Nile virus which helped decrease human deaths (Centers for Disease Control and Prevention, 2003).

Virology

The West Nile virus is a small, single-stranded RNA virus of the family Flaviviridae and genus Flavivirus and a member of the Japanese encephalitis virus antigenic complex (Guharoy et al., 2004; Petersen and Marfin, 2002; Gea-Banaclocche et al., 2004; Marra et al., 2004). The virus can be divided genetically into two lineages. Although two genetic lineages of West Nile virus have been identified, only members of lineage 1 have been associated with clinical human encephalitis in the United States (Petersen and Roehrig, 2001; Petersen and Marfin, 2002; Guharoy et al., 2004). "The West Nile virus responsible for the 1999 outbreak in New York City was a lineage 1 virus that circulated in Israel from 1997-2000, suggesting viral importation into North America from the Middle East" (Petersen and Marfin, 2002, p. 174). However, the means of its introduction will likely remain unknown (Petersen and Roehrig, 2001).

Ecology and Transmission

West Nile virus is maintained in an enzootic cycle involving several species of mosquitoes and birds before infecting humans (Guharoy et al., 2004). However; humans are considered dead-end hosts, insufficient to support the life cycle of the virus because of low-grade, transient viremia. (Gea-Banaclocche et al., 2004). Humans might not be hosts, but can become infected with the virus when bitten by an infected mosquito. West Nile virus infection is transmitted from birds to humans through the bite of mosquitoes (Guharoy et al., 2003). Mosquitoes become infected with West Nile virus when they feed on an infected host, usually a bird. Within about two weeks of becoming infected, a mosquito can transmit the virus in its saliva (Guharoy et al., 2004). There is some evidence that suggests warmer temperatures may shorten the 14 day cycle (Epstein, 2000, 2001; Dye, 2000; Monath and Tsai, 1987). During subsequent feedings, the mosquito injects this virus-laden saliva with each bite (Gea-Banaclocche et al., 2004). "Although *Culex pipiens*, *Culex restuans*, and *Culex quinquefasciatus* are probably the most important maintenance vectors in the eastern United States, it is unknown which species are most responsible for transmission to humans" (Petersen and Marfin, 2002, p. 174).

Regardless of which species are most responsible, the sick and elderly are at the highest risk of getting West Nile virus (Chowers et al., 2000; Petersen and Marfin, 2002; Gea-Banaclocche, 2004).

Mosquito Biology

Mosquito species such as the *Aedes aegypti* and *C. quinquefasciatus* are among those responsible for the transmission of most vector-borne diseases (Githeko et al., 2000). In addition, *Culex salinarius*, *C. restuans*, and *C. pipiens* have also been involved in the spread of vector-borne diseases (Epstein, 2001). There are numerous species of mosquitoes in Mississippi; however, only a few of them have been proven in the literature to be important arbovirus vectors (Table 1). According to Goddard (2002), some of the most important are *A. aegypti*, *Aedes albopictus*, *Ochlerotatus sollicitans*, *Ochlerotatus triseriatus*, *C. quinquefasciatus*, and *Psorophora columbiae*. The Yellow Fever Mosquito (*A. aegypti*) is found in shaded artificial containers (Gubler, 1989). Goddard (2002) adds that they have a flight range of 100-300 feet and usually bite during the morning or late afternoon. The Asian Tiger Mosquito (*A. albopictus*) has a life cycle similar to that of *A. aegypti.* They are most often found in tire piles. Their flight range is less than a ¼ mile. The Salt Marsh Mosquito (*O. sollicitans*) is a fierce biter, similar to *A. albopictus*. They rest on vegetation and have a flight range between 5 and 10 miles. The Tree Hole Mosquito (*O. triseriatus*) is another fierce biter. It has a short flight range and has the potential to carry forms of encephalitis. The Southern House Mosquito (*C. quinquefasciatus*) feeds on birds and humans and has an extremely short flight range. It is the major vector of St. Louis Encephalitis (Goddard, 2002). It is also involved with the West Nile virus in urban environments (Epstein, 2001). The Dark Rice Field Mosquito (*P. columbiae*) is a fierce biter that has a flight range of at least 10 miles. It is the major vector of several equine encephalitis cases (Goddard, 2002). What is concerning is that these mosquitoes may remain active throughout the year in southern states (Marfin et al., 2001). On the basis of these studies, the following conclusions may be drawn: a) competent mosquito vector species are found in urban and rural environments, b) flight ranges vary greatly from feet to miles and, c) competent mosquito vector species may be active year round.

Table 1
Important Mosquitoes of Mississippi

Mosquito	**"YellowFever" Aedes Aegypti**	**"AsianTiger" Aedes albopicus**	**"SaltMarsh" Ochlerotatus sollicitans**	**"TreeHole" Ochlerotatus triseriatus**	**"SouthernHouse" Culex quinquefasciatus**	**Culex pipiens**	**"DarkRice" Psorophora columbiae**
Habitat	Shaded artificial container and tree holes	Tire piles	Salt marshes, Flooded or not	Artificial Containers and tree holes	Sewers, septic tank Overflow, ditches and cesspools	**OPW	Freshwater pools and ditches
Flight Habits	100 – 300 ft.	< ¼ mile	5-10 miles up to 40 miles	Short	Extremely Short	Short	At least 10 miles
*Trans. Temps.	64ºF - 95ºF 17ºC - 35ºC	64ºF - 95ºF 17ºC - 35ºC	64ºF - 95ºF 17ºC - 35ºC	64ºF - 95ºF 17ºC - 35ºC	64ºF - 95ºF 17ºC - 35ºC	64ºF - 95ºF 17ºC - 35ºC	64ºF - 95ºF 17ºC - 35ºC
Limiting Temps.	32ºF - 104ºF 0ºC - 40ºC	32ºF - 104ºF 0ºC - 40ºC	32ºF - 104ºF 0ºC - 40ºC	32ºF - 104ºF 0ºC - 40ºC	32ºF - 104ºF 0ºC - 40ºC	32ºF - 104ºF 0ºC - 40ºC	32ºF - 104ºF 0ºC - 40ºC
‡Precip.	(HpLr/Hr)			(HpLr/Hr)	(HpLr/Hr)	(HpLr/Hr)	
Life Cycle to Trans.	10 - 20 days	10 - 20 days	7 - 10 days	28 days	10 - 14 days	10 - 14 days	7 - 10 days

**** Open, Polluted Water (High in organics)**
*** Peak disease transmission temperatures**
‡ High precipitation = low risk, two to three weeks after precipitation event = high risk

Breeding and Climate

According to Martens et al. (1997), breeding and egg laying, as well as mosquito longevity, are greatly influenced by temperature and precipitation. These influences will be discussed in the following sections. Reproduction rates are fairly inconsistent between the different species; they can be as short as a few days (*A. aegypti*) or as long as a few months (*A. albopictus* and *O. triseriatus*). Climate plays a major role in the time it takes for completion. The ability of vectors to breed and reproduce depends on whether they encounter motionless or rapidly moving water (Martens et al., 1997).

Gubler (1989) states that *A. aegypti* lay single eggs on the inside of containers at or above the water line. There has been a huge increase in the amount of these artificial containers that make ideal larval habitats for this mosquito. Under good conditions, larval development is completed in 6 to 10 days. The pupal stage lasts about two days (Goddard, 2002). "The life cycle can be completed within 10 days under good conditions or extend to three or more weeks under poor conditions" (Goddard, 2002, p. 35). *A. albopictus* has a similar life cycle as *A. aegypti.* Tire piles are the best place for *A. albopictus*, which like to breed in water filled containers (Hawley, 1991). *O. sollicitans* breeds in flooded salt marshes. However, breeding may occur in marsh areas not covered by water. Eggs that have remained dry for two weeks will hatch within minutes when flooded. Their life cycle can be completed in about 7 to 10 days during warm weather (Goddard, 2002). *Ochlerotatus taeniorhynchus* breeds in salt marshes or freshwater pools near those marshes. Breeding lasts from late spring until October. *C. quinquefasciatus*, like the majority of the *Culex* species, breed in organic waters. Eggs are laid on floating rafts of 50 to 400, which hatch within one or two days in warm temperatures. During cooler weather, several weeks may be required for complete development (Goddard, 2002). *P. columbiae* breeds in temporary freshwater pools and ditches and is very abundant in rice fields. Many broods are produced from April to October. Eggs are laid on flood-prone areas of low vegetation. At an average temperature of 26º C, larval stages can be completed in 5 days. The pupal stage lasts 1 to 2 days. “Areas that dry up and are reflooded every few days can produce a hatch with each flooding” (Goddard, 2002, p. 51). On the basis of these studies, the following conclusions may be drawn in regards to the modeling effort: a) breeding and egg laying are greatly influenced by temperature and precipitation and b) drought followed by precipitation increases the risk of mosquitoes.

Feeding and Climate

“Mosquitoes fall into four groups based on their feeding patterns. These are species that feed (i) primarily on mammals, (ii) primarily on birds, (iii) primarily on cold blooded vertebrates, and (iv) on a wide variety of hosts” (Edman and Taylor, 1968, p. 67). Edman and Taylor (1968) go on to say that mammal host feeding occurs in early summer, reaches a maximum between July and October, and is followed by a shift to avian host feeding, which dominates winter and spring. Day and Curtis (1989) agree that there is a seasonal feeding shift to mammals during the summer and autumn months.

“A combination of many factors results in successful host location and engorgement by mosquitoes. Host abundance is a key factor. Once found, non-defensive or incapacitated hosts are more easily fed on than defensive species" (Day and Curtis, 1989, p. 32). Host abundance may be a factor but the importance of vector abundance is an ongoing question. Conflicting reports of vector abundance and virus transmission appear in the literature (Day and Curtis, 1989). It can be concluded from these studies that host location and abundance are important to the modeling process.

Temperature Thresholds

Temperature plays an important role in the life cycle of mosquitoes and in the replication and transmission of diseases. Mosquitoes are critically dependent on climate for their survival and development. Climate circumscribes the distributions of mosquito borne diseases, while weather affects the timing and intensity of outbreaks (Githeko et al., 2000; Epstein et al., 1998). According to Patz et al. (1998) and Karl et al. (1995), minimum temperatures are now increasing at a disproportionate rate compared to average and maximum temperatures. This allows climate-sensitive vector-borne diseases to move into regions previously free of disease (Patz et al., 1998).

"The greatest effect of climate change on transmission is observed at the extremes of the range of temperatures at which transmission occurs; 14-18º C at the lower end and about 35-40º C at the upper end" (Githeko et al., 2000, p. 1136). Warmer temperatures speed the development of the parasites in mosquitoes, raising the odds of disease transmission (Epstein, 2000, 2001; Dye, 2000; Monath and Tsai, 1987). Cooler temperatures slow reproduction rates and disease replication; extreme cold weather kills adult mosquitoes, over-wintering eggs, and larvae (Githeko, 2000; Epstein 2000; Patz et al., 1998). There is a threshold temperature above which death is inevitable and a minimum temperature below which the mosquito cannot become active. Threshold temperatures for *Psorophora vivax* and *Psorophora falciparum* range between 14.5-15º C and 16-19º C. The optimal temperature for *Anopheles* survival lies between 20-25º C. *Aedes* are less responsive to ambient temperatures than *Anopheles* because they live mainly indoors (Martens et al., 1997). On the basis of these studies, the following conclusion may be drawn. Temperature influences mosquito abundance. This is important to the modeling process because according to Purvis (1993), temperature is one of the most important criteria that influence potential evaporation. Precipitation minus evaporation (P-E) is a variable used in the predictive models.

Precipitation Thresholds

"In addition to the direct influence of temperature on the biology of vectors and parasites, changing precipitation patterns can also have short and long term effects on vector habitats" (Githeko et al., 2000, p. 1137). High amounts of precipitation result in a greater potential to increase the number of breeding sites. A lack of precipitation is also important. Multi-month drought in spring and early summer was found to be associated with recent severe urban outbreaks of West Nile virus in the United States (Epstein, 2001). Monath and Tsai (1987) agree that outbreaks have been associated with drought. The combination of drought and rainfall is probably the key to outbreaks. Rains followed by drought seem to be the correct combination for these outbreaks. Excessive rainfall in January and February, in combination with drought in July, most often precedes outbreaks (Githeko et al., 2000). Day and Curtis (1989) found similar results. A wet July results in high mosquito abundance in August.

Humidity is an often-overlooked factor in the life cycle of mosquitoes and in the replication and transmission of diseases. "Rainfall raises the relative humidity particularly following dry periods, and relative humidity strongly influences mosquito flight and subsequent host-seeking behavior" (Day and Curtis, 1989, p. 36). The most adverse extremes of humidity can completely prevent mosquito host-searching flights. More in-depth research on the effects of humidity needs to be completed before a full understanding can be acquired (Day and Curtis, 1989). It can be concluded from these studies that the combination of drought and precipitation are important to mosquito habitat suitability and therefore are important to the modeling process.

GIS and Vector-Borne Diseases

Modeling the biology and transmission characteristics of vector-borne diseases is complex (Skidmore, 2002). Parsimonious models should maximize predictions without model over-parameterization. Existing GIS-based models are reviewed below for Lyme disease and Malaria, both of which are vector-borne diseases.

Lyme Disease

Lyme disease is a tick-transmitted bacterial infection that affects humans and domestic animals. Several studies on Lyme disease have demonstrated the ability to generate risk models using GIS. Glass et al. (1995) used a geographic information system to identify and locate residential environmental risk factors for Lyme disease.

They found that eleven of their fifty-three variables were associated with an increased risk of getting Lyme disease. After these significant variables were discovered, they generated a risk model that combined the geographic information system with logistic regression analysis (Glass et. al., 1995). It was concluded that "combining a geographic information system with epidemiologic methods could be used to rapidly identify risk factors of zoonotic disease over large areas" (Glass et. al., 1995, p. 944).

Similar to Glass et al., Nicholson and Mather (1996) also used GIS to identify factors that may regulate tick distributions and, thus, Lyme disease risk. Their findings were combined "to create a model that predicts Lyme disease transmission risk, thereby demonstrating the utility of incorporating geospatial modeling techniques in studying the epidemiology of Lyme disease" (Nicholson and Mather, 1996, p. 711).

Malaria

Malaria is a serious and sometimes fatal disease that is caused by a protozoan parasite which is transmitted by mosquitoes. Several studies on Malaria have demonstrated the ability to generate risk models using GIS. Beck et al. (1994) integrated remotely sensed data and GIS capabilities to identify villages with high vector-human contact risk. Their results indicated that villages with high Malaria vector-human contact risk can be identified using remote sensing and GIS technologies.

Srivastava et al. (2001) also developed a model that predicts Malaria risk. A predictive habitat model was developed for forest Malaria vector species using GIS and a Boolean operator to map areas where the species is likely to be found. Their results indicate that "GIS-based distribution can pinpoint areas of occurrence of *Anopheles dirus* at the micro-level, where species-specific environmental-friendly control measures can be strengthened" (Srivastava et al., 2001, p. 1133).

These studies suggest that GIS is a useful tool for modeling vector-borne diseases. In particular, Srivastava et al. (2001) points out that accurate delineation of favorable mosquito habitat is closely linked with disease risk.

GIS and West Nile Virus

Previous research on other vector-borne diseases, Lyme disease and Malaria, has demonstrated the ability to model risk of disease from these biological systems within a GIS. Review of current literature suggests that geographic information systems have primarily been used for monitoring and surveillance in combating West Nile virus. Very few GIS modeling efforts for West Nile virus have been published. This lack of predictive risk modeling presents a unique opportunity for using GIS to combat West Nile virus. This research moves beyond descriptive modeling and combines intuitive and deductive modeling philosophies for the development of a dynamic risk model.

West Nile Virus Surveillance

The Centers for Disease Control and Prevention (CDC) has one of the most sophisticated West Nile virus surveillance systems in the country. Known as ArboNet, the system helps states track West Nile and other mosquito-borne viruses (Centers for Disease Control and Prevention, 2003). Local and state public health departments share their data with the CDC, which provides real-time data on West Nile virus activity across the nation. The CDC also works in conjunction with the United States Geological Survey (USGS) to produce county maps of the entire United States that show bird, human, mosquito, sentinel, and veterinary cases of West Nile virus (USGS, 2004).

Pennsylvania also has a sophisticated surveillance program. The National Aeronautics and Space Administration (NASA) along with multiple state agencies have worked together to develop this West Nile virus surveillance system (Top Story GSFC, 2002). "The PA West Nile Virus Surveillance System (PAWNVSS) provides up-to-date information on where infected mosquitoes, birds, and humans have been reported throughout the state" (Top Story GSFC, 2002, p. 1). The data collected are combined in a GIS and used to create a county map of Pennsylvania that indicates in which counties West Nile virus has been reported. Pennsylvania agencies are currently using the PAWNVSS system to make daily decisions on the best places and times to spray for mosquitoes (Steitz and Ramanujan, 2002).

West Nile Virus Modeling

A unique modeling approach found in the literature is the Dynamic Continuous-Area Space-Time (DYCAST) model developed by a group at New York's Hunter College. The DYCAST model was developed to identify and monitor high-risk areas for West Nile virus in New York City (Theophilides et al., 2003). "It successfully identified areas of high risk for human West Nile virus infection in areas where five of seven human cases resided, at least 13 days prior to the onset of illness" (Theophilides et al., 2003, p. 843). The basis for this model is dead crow reports and a Knox Test for space-time interactions. Studies suggest that bird reports and the Knox test are biased. Kulldorff and Hjalmars (1999) state that the Knox test for space-time interaction is biased when there are geographical population shifts. Bird migration is definitely a geographical population shift. Also, Petersen and Roehrig (2001) state that although crows are by far the most identified species, this may reflect the lethality of infection in this species, rather than its importance as a reservoir host.

The Chicago Department of Public Health also uses a GIS model to predict West Nile virus risk. The LinksPoint VectorWatch geographic risk modeling system aids in the prevention of West Nile virus by identifying areas within the city where disease activity is present (LinksPoint, 2003). This model is based on the DYCAST model.

The previous models relied on dead bird reports with little emphasis on environmental risk factors. According to the Ames Research Center (2003), a group of students working for NASA created a West Nile virus risk model based on mosquito habitat suitability for Monterey County, CA. The group correlated ground observations with satellite imagery to identify countywide mosquito habitat. This resulted in a model that shows the location of at-risk humans who are 55 and older and their proximity to West Nile virus-carrying mosquito habitat. The group was also able to recommend additional mosquito surveillance in places where the county was not doing surveillance.

Bird data as an indicator species may have drawbacks. In Mississippi, some county health departments only test dead birds for West Nile virus until a positive WNV case is found, they do limited or no testing after that (Personal Communication, Sally Slavinski, 2004). The Environmental Risk Analysis Program (2002), from Cornell

University's Department of Communication, adds that cumulative counts of WNV-positive birds have ceased to be a useful indicator of WNV prevalence because reports of dead birds are handled differently in different places. Another obvious drawback to using bird cases for modeling is the necessity of a human being finding a dead bird and bringing it in for testing. Biases due to population density result in higher probability of bird detection in high population centers.

CHAPTER III

MATERIALS AND METHODS

This research was supported by a grant from the National Institutes of Health. The grant was administered through East Carolina State University and the Southern Coastal Agromedicine Center. The study was designed to assess risk for West Nile virus infection for the entire state of Mississippi.

Study Area

Mosquito habitat suitability was treated as a surrogate for potential human risk for West Nile virus infection. Data were acquired from a variety of sources. Some data were derived from other data sources through interpolative processes. When data were interpolated, the calculations were extended beyond the borders of Mississippi into Alabama, Tennessee, Arkansas, and Louisiana and then subset to the study area before analysis.

Raster and Vector Variables (GIS Layers)

GIS data are generally divided into two primary data structures, raster and vector. Vector data are stored as points, lines, and polygons while raster data are stored as a regular grid of cells. Continuous surface layers like elevation and its derivatives (slope, aspect) are usually stored as raster data and discrete data like soil type are usually stored as vector data. For GIS predictive modeling purposes, data are usually converted to the raster data structure. Often, discrete data like soil type have attributes (permeability, soil pH, etc.) that are stored as continuous data that are easily converted to the raster data structure.

Data Description

Numerous data layers were considered as potential modeling variables. Data layers were reviewed for usefulness to the modeling process based on several criteria including:

- Resolution of the data
- Quality of the data (collection and accuracy)
- Extent of the data coverage
- Importance to the modeling process based on literature review

After reviewing each potential GIS data layer based on these criteria, static and dynamic modeling variables were chosen. Static variables are those landscape variables like elevation that do not change through time or change very slowly. Dynamic variables (e.g. precipitation) are those whose temporality influences model results. For modeling purposes in this study, dynamic variables were summarized on a seasonal basis. Examples of these static and dynamic variables include:

Static Layers

- Digital Elevation Models and their derivatives
    - Aspect
    - Slope
- Soil Permeability
- Roads
    - Road density
- Perennial Streams
    - Stream density
- Intermittent Streams

- Stream density

Dynamic Layers

- Normalized Difference Vegetation Index (NDVI)
- Precipitation
- Evaporation

Digital Elevation Models and Derivatives

Ten-meter Digital Elevation Models (DEMs) by county were obtained from the Mississippi Automated Resource Information System (MARIS). The 10m DEM was a major improvement over the 30m DEM provided by the US Geological Survey (USGS). Slope, measured in degrees, was generated from the elevation data as a continuous variable ranging from 0 to 90 degrees.

Numerous layers were also obtained from MARIS. These included Mississippi roads, streams, and zip codes, as well as national parks, state parks, national forests, and wildlife refuge areas.

Soil Permeability

Soil Permeability, or the rate at which water flows through a soil in cubic centimeters per hour, was used to calculate drainage. Combining soil permeability with slope indicates the potential capability that a site has for water to pond and create mosquito habitat. A permeability grid based on States Soils Geographic Database (STATSGO) data was obtained from Pennsylvania State University.

Roads

The roads layer was originally obtained from MARIS; however, the layer was not as up-to-date as desired. The 2002 Census data roads layer was used in place of the data from MARIS. This vector layer was used as input to a GIS procedure for calculating road density.

Streams

Separate streams layers included Perennial and Intermittent streams. These vector layers were merged and used as input to a GIS procedure for calculating stream density.

Population

Census 2000 population data were summarized by zip code. These summarizations formed the basis for creating a continuous surface for population density, which helped normalize the West Nile occurrence data.

Normalized Difference Vegetation Index (NDVI)

NDVI is a ratio of the red and near infrared wavelengths and is commonly used in vegetation analyses to estimate vegetative cover (Lillesand et al., 2004). The National Oceanic and Atmospheric Administration's (NOAA) Moderate Resolution Imaging Spectroradiometer (MODIS) is a multi-spectral scanner that records several wavelengths including red and NIR. MODIS 14-day temporal composite data were used to calculate NDVI for use in this study.

Climatic Variables

Studies indicate that precipitation and evaporation are important variables for modeling disease risk when mosquitoes are vectors. The majority of the mosquitoes that carry the West Nile virus breed in open, stagnant water bodies. As a result, water input into the system would highlight potential breeding areas. However, precipitation alone does not give an accurate measurement of water input. Evaporation must be considered, since rainfall and evaporation yield estimates of the available water or "water balance." Precipitation and pan evaporation data for Mississippi were obtained from weather stations throughout the state for the 2002-year. Data were also obtained from the stations that border Mississippi in Alabama, Tennessee, Arkansas, and Louisiana. There are more stations that record precipitation than evaporation. However, because evaporation is more uniformly distributed across the landscape than precipitation, the lack of stations is less of a problem than if only a few stations recorded precipitation (Personal Communication, Christopher Bell, 2005).

Validation Data

West Nile virus positive human and bird cases by zip code were obtained from the Mississippi Department of Health (MDOH) for 2002 and 2003. Zip codes are higher resolution than county boundaries, 404 polygons as opposed to 82 polygons. It also should be noted that the human cases are a laboratorial diagnosis not a clinical diagnosis. Since clinical cases can be mis-diagnosed, the laboratorial data are suitable for training and validating the models. These data included the date of occurrence, the zip code, and the city name.

Data Preparation

The overall modeling approach required that all data have the same cell-size and that all variable "states" or levels be standardized for risk suitability. The 10m-County Digital Elevation Models were downloaded in a compressed format. All 82 counties were uncompressed and imported into the GIS software file format. The DEMs were reprojected from Mississippi State Transverse Mercator to USA Contiguous Albers Equal Area. Once projected, a mosaic was created from the individual county DEMs. The 82 counties were mosaiced into five groups due to GIS software processing and storage limitations. Each of the five mosaics were resampled to 30m and then combined (mosaiced) to form a statewide 30m DEM. This grid contained data gaps at some of the common county boundaries. The procedure used to remove these gaps employed a 3x3 focal mean filter. This filter looks at nine pixels within the roving window, averages them, and inserts that averaged value into the center pixel. The filter acts as a smoothing device to eliminate noise or in this case fill data gaps. The filtered DEM was merged with the original unfiltered DEM to create a seamless 30m DEM. The 'merge' routine fills the data gaps with the filtered grid values without changing all the values in the original grid. After the creation of the new 30m DEM, slope was derived, which was reclassified and divided into ten classes using the "Quantile" classification method. With the "Quantile" method, the range of possible values is divided into unequal-sized intervals so that the number of values is the same in each class. Classes at the extremes and middle have the same number of values. Because the intervals are generally wider at the extremes, this option is useful to highlight changes in the middle values of the distribution (ESRI, 2002). The lowest slope was given a rank of ten and the highest slope received a rank of one.

Unlike the excellent condition of the new, 30m DEM, the permeability grid obtained from Pennsylvania State University at 1-km cell resolution depicted sharp boundaries at cell transitions. Generally, resampling would improve the poor resolution; however, resampling the permeability grid to 120m from 1km was just not feasible. Each 1km grid cell would be broken down into eight, 120m cells. As a result, the permeability grid was converted to a point file. A spline interpolation was performed on the new permeability point file. This interpolation method estimates cell values using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points (ESRI, 2002). This improved the overall quality of the permeability layer, which resulted in a smooth transition between permeability classes. The output layer was multiplied by a "mask grid" of the state boundary shapefile. This “mask” confines the output to the extent of the state boundary. The permeability layer was divided into ten classes using the “Quantile” classification method. Finally, the lowest permeability was given a rank of ten, the highest permeability a rank of one.

Normalized Difference Vegetation Index (NDVI) was derived from MODIS imagery at 250m-resolution. Unlike the permeability layer, NDVI could feasibly be resampled to 120m. Each 250m grid cell would only be broken down into two, 120m cells. As a result, NDVI was resampled to 120m. This layer was also multiplied by the “mask grid.” The highest NDVI received a rank of ten and the lowest received a rank of one.

Perennial and intermittent stream shapefiles from MARIS were merged using a GIS “merge” function. A stream density grid was then created using the “Kernel” density type with a 2500m-search radius. With the kernel density calculation, the points or lines lying near the center of a raster cell's search area are weighted more heavily than those lying near the edge. The result is a smoother distribution of values (ESRI, 2002). The “mask grid” was applied to the output layer. The layer was then divided into ten classes using the “Quantile” classification method. The highest density received a rank of ten and the lowest stream density a rank of one.

Recent road data were available as 2002 TIGER files from the Census Bureau. The primary and secondary road layers were merged using a GIS “merge” function. This merged layer was handled the same way as the streams layer with the creation of a density grid using the “Kernel” density type with a 2500m-search radius. The highest road density received a rank of ten and the lowest density a rank of one.

Precipitation and evaporation data were provided by Dr. Charles Wax, the Mississippi State Climatologist. Pan evaporation is not truly representative of actual evaporation due to the differences in heating and exposure to wind from the pan environment to that of a pond or large body of water. Also, pan evaporation does not account for water loss to transpiration through plants. As a result, evaporation data was corrected by multiplying every entry by 0.8 (Bell, 2004). Missing data were filled with the monthly average for the station using the State Division number to find the value in the National Climatic Data Center (NCDC) database. Both precipitation and evaporation were provided in spreadsheet format, which included the daily averages for all twelve months and the station ID with its corresponding latitude and longitude. Point files were created from the precipitation and evaporation data. The created point files were used for interpolation. Both precipitation and evaporation for January – December 2002 were interpolated using the spline method. Spline interpolation techniques were chosen because this technique creates smooth transitions across the interpolation area. All twelve months of precipitation and evaporation data were multiplied by the “mask grid” to subset the layers to the Mississippi State boundary. Finally, each month of evaporation was subtracted from the corresponding month of precipitation to derive P-E. June, July, and August P-E were added to get the summer P-E. September, October, and November P-E were added to get the fall P-E. No other cases existed beyond these dates.

West Nile virus positive human and bird cases by zip code were obtained from the MDOH in spreadsheet format. Input errors such as a mis-keystroke during data entry, where the numbers in the zip codes for the same city were reversed, were corrected. Latitude and longitude for every zip code's polygon centroid were acquired from the CD Light, LLC website: www.zipinfo.com/search/zipcode.htm and added to the spreadsheet. If the looked-up zip code did not match the city name in the MDOH spreadsheet, the zip code was maintained and the city corrected. For these "problem" records, the zip codes were checked with the United States Postal Service records. After all errors were corrected and each zip code had its associated latitude, longitude, date, and number of occurrences attached to the spreadsheet, point files for 2002 and 2003 human and bird cases were created. In order to remain consistent with P-E, occurrences were separated by summer and fall. Summer included the months of June, July, and August while the fall included the months of September, October, and November. In order to eliminate population bias, the data were normalized by population. Population for each zip code was obtained from the website, www.joshskidmore.com/?_page=projects&_subpage=zipcode_database and then added to the spreadsheet. The total number of human occurrences of West Nile virus was divided by the total population, which resulted in a normalized set of occurrence data.

Methods

The spatial information product (SIP) for this project was a statewide West Nile virus risk map correlated to ideal mosquito breeding habitats for Mississippi. Natural resource areas and state parks were overlaid on this SIP and risk for each area calculated. Review of the literature on West Nile virus assumes that slope, NDVI, stream density, and other environmental variables are critical to the modeling process. As a result of the literature review and a "round table" discussion with a climatologist, a forester, a geoscientist, and a meteorologist, we proceeded with the first modeling effort. Modeling was carried out in the raster environment using static and dynamic variables. Even though NDVI is a dynamic variable, it was used as a static variable, a snapshot in time. The static variables, those that do not change (slope, aspect, road density, stream density, NDVI) and the dynamic variables, those that do change (precipitation and evaporation), were conditioned, ranked, and weighted in order to use map algebra in a linear additive modeling scenario. Weights were heuristically assigned based on the "round table" discussions.

There were three major parts to this study: data preparation/variable manipulation, statistical tests, and model construction. The majority of the effort for data preparation involved several steps to get the original occurrence data consistent and in a form that could be used in analysis. Once this was completed, the other variables were prepared for analysis. Each variable was converted to raster and conditioned in preparation for model generation. Variable "states" or levels for slope, road density, stream density, NDVI, summer P-E, and fall P-E were ordinated from 1-10 with 10 representing highest risk and 1 representing lowest risk.

The second portion of this study involved performing statistical tests to see how the variables correlated with WNV case occurrences and which variables were the most statistically important on a t-test basis. T-tests were made to test for differences between zip codes of West Nile virus occurrence and zip codes of non-occurrence at the 95% confidence interval using weighted and non-weighted case occurrences. Linear regressions were then applied for variables where significant differences existed for variables in zip codes of WNV occurrence vs. non-occurrence. Regressions helped to determine the strength of relationships between rate of infection and the variables of interest.

The last major portion of this study involved the creation of weighted linear additive models and a logistic regression model. For the additive models, each of the six variables was ranked in importance to the modeling effort based on their t-test probability level. Weights were calculated by dividing each individual rank by the total sum of the ranks. After the variables were ranked in order of significance and weights were assigned, linear additive models were constructed using map algebra techniques. Four linear additive models were created: 2002 Summer, 2002 Fall, 2003 Summer, and 2003 Fall. Due to low occurrence numbers by zip code and poor results relating rate of infection to any variable, logistic regression was investigated for modeling risk. Logistic regression, as used in epidemiology, is defined as a statistical method for calculating odds ratios for individual risk factors where a variety of risk factors may be contributing to the occurrence of disease (Wartenberg et al., 1996).

For the Logistic regression model, probability of occurrence in each zip code was calculated and linearized by taking the natural log of the probability of occurrence of West Nile virus in each zip code. Resulting probabilities were constrained between 0 – 1. Each zip code in the state was assigned a probability for occurrence of West Nile virus and the resulting probabilities were used in an Inverse Distance Weighted interpolation technique to calculate risk for the entire state. The Inverse Distance Weighted interpolation technique was chosen due to the fact that it assumes that the variable being mapped decreases in influence with distance from its sampled location. The logistic regression probabilities should decrease the farther you move away from the sampled location.

CHAPTER IV

RESULTS AND DISCUSSION

This chapter will focus on the results of the visual analysis, the statistical tests applied to the data, and the results from both the linear additive and logistic regression models. Variable ranks and weights along with issues dealing with the original data will also be discussed. Tables containing the original 2002 and 2003 human occurrence data are included in Appendix A.

Visual Analysis of the Spatial Distribution of Case Occurrences

Pattern of Case Occurrences vs. Population Density

Figure 2 displays the pattern of 2002, West Nile virus case occurrences against population density for the entire state of Mississippi. The diameter of the points is indicative of the number of occurrences. Higher numbers of occurrences result in larger diameters. Also, darker tones of blue indicate a lower population density while lighter tones indicate a higher population density. The pattern of case counts in relation to population centers indicates an urban problem. Clusters of large-diameter points are in close proximity to major metropolitan areas. However, when case occurrences are normalized by population, total number of occurrences divided by total population, a different picture is presented. Figure 3 illustrates this statement. Now the largest diameter points are located in places other than the major metropolitan areas.

Normalizing case count by population suggests that there are other variables that affect the pattern of West Nile virus occurrences.

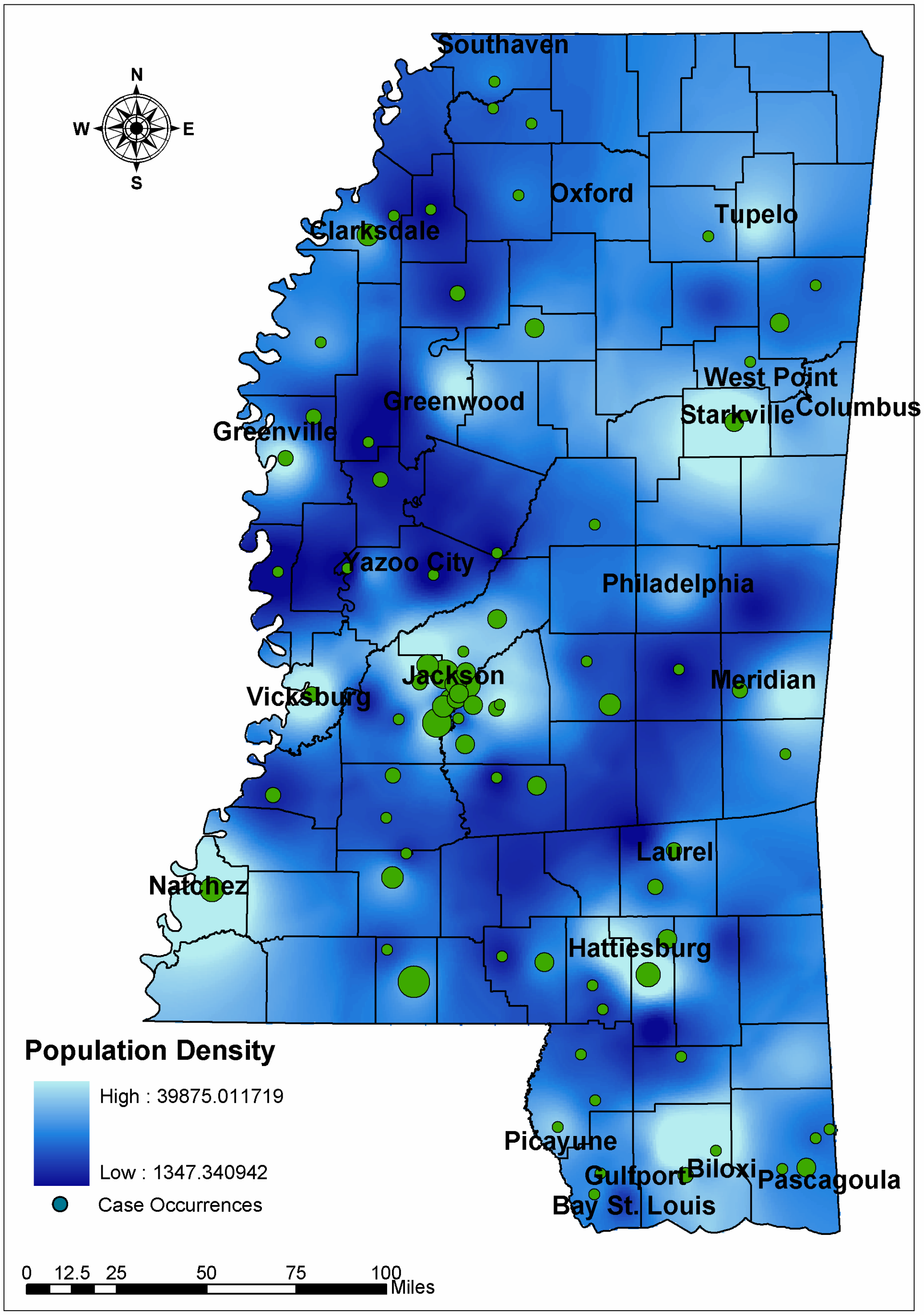

Figure2: Pattern of Case Occurrences vs. Population Density

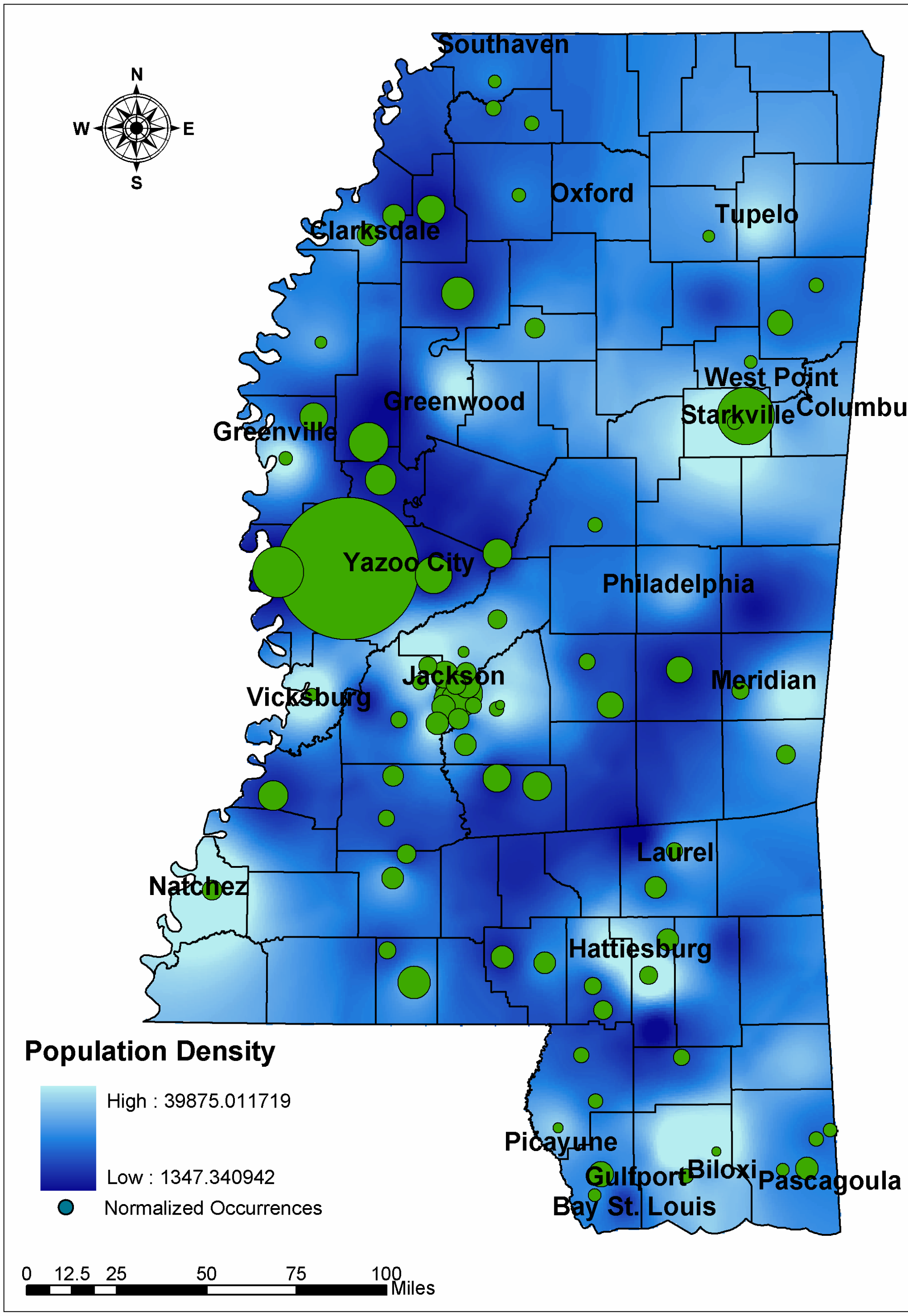

Figure 3: Pattern of Normalized Case Occurrences vs. Population Density

Pattern of Normalized Case Occurrences vs. Slope

Case occurrences were overlaid on each variable of interest in order to determine if there were relationships between the pattern of occurrences and the associated variables. It should be noted that Holly Bluff, the largest diameter point from Figure 3, was removed from each of the remaining figures for display purposes. Because of its size, the point was obscuring information below it.

Figure 4 shows the 2002, normalized case occurrences compared with slope. Steep slopes are represented by lighter tones. Darker tones indicate a more flat slope. Occurrences seem to be clustering around areas of gentle slope. Intuitively, this would make sense due to the fact that water is much more likely to pool in flat areas, resulting in higher mosquito habitat suitability. On this basis, it is surprising that there is only a small clustering of occurrences in the Mississippi Delta. The visual analysis suggests that, like population density, variables other than just slope are important.

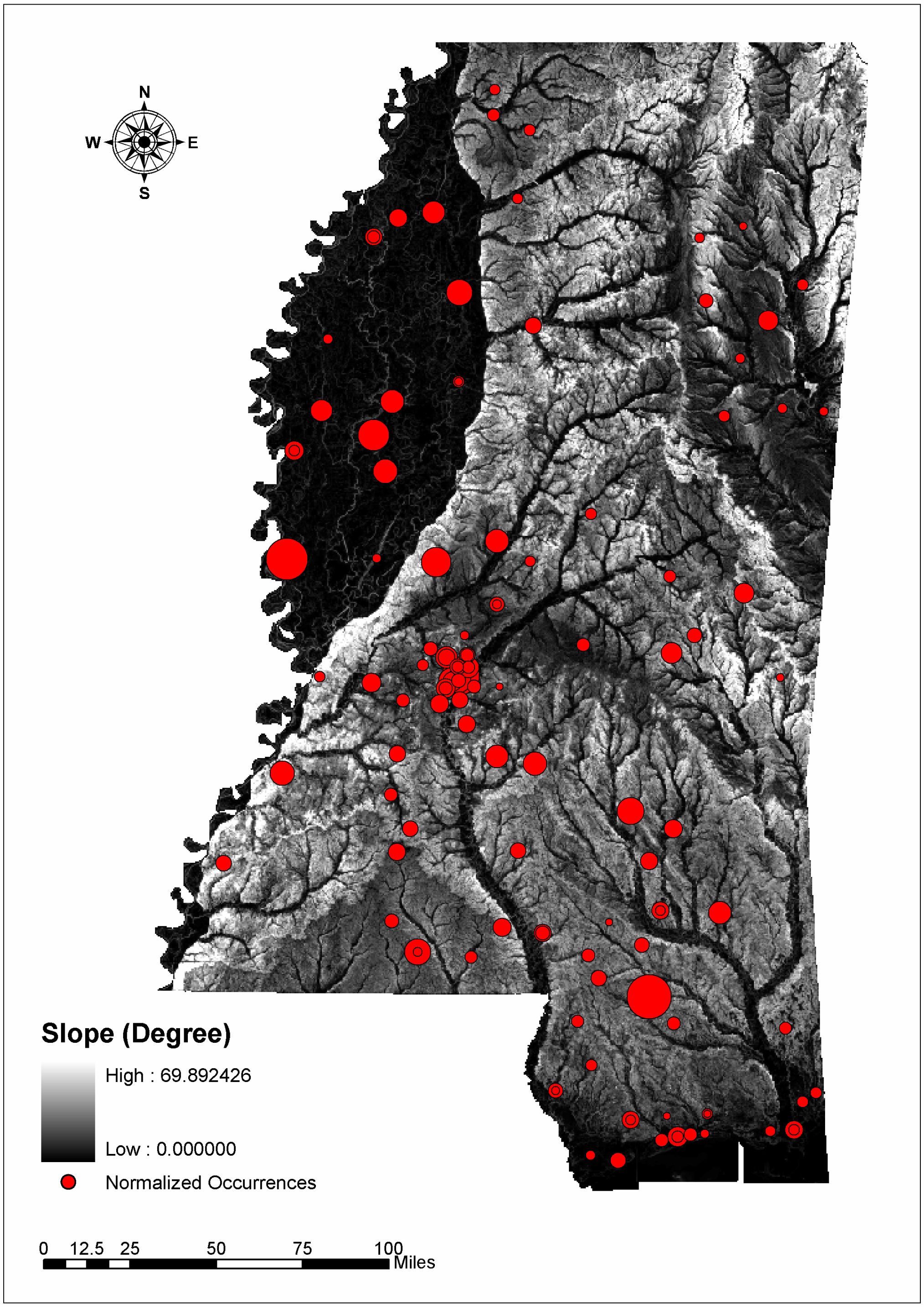

Figure 4: Pattern of Normalized Case Occurrences vs. Slope

Pattern of Normalized Case Occurrences vs. Soil Permeability

Figure 5 shows the normalized case occurrences compared with soil permeability. Higher values of soil permeability are displayed as lighter tones; lower values are displayed as darker tones. There seems to be a high to low gradient for permeability values from southeast to northwest across the state. The majority of the clusters are located in areas of lower permeability. This is intuitively appealing due to the fact that water is more likely to pond in areas of lower permeability.

Pattern of Normalized Case Occurrences vs. Road Density

Figure 6 shows the normalized case occurrences compared with the road density grid. Higher road densities are displayed as lighter tones, lower road densities are darker tones. Here, occurrences are clustered around areas of high road density, suggesting that there is a relationship between road density and West Nile virus occurrences.

Pattern of Normalized Case Occurrences vs. Stream Density

Normalized case occurrences were overlaid on the stream density grid (Figure 7). Similar to the previous figures, lighter tones represent higher stream density while darker tones represent a lower density. Unlike the clustering of occurrences in the areas of high road density, clusters of occurrences do not predominate in areas of high stream density. This was an unexpected result. Expectations that higher stream densities would result in a more suitable mosquito habitat were not substantiated by visual analyses of these data. Further investigation was needed.

Pattern of Normalized Case Occurrences vs. Normalized Difference Vegetation Index (NDVI)

Figure 8 shows the normalized case occurrences compared to NDVI. The normalized difference vegetation index is a standardized method of comparing vegetation greenness between satellite images. NDVI is preferred to more simple indices because it helps compensate for changing illumination conditions, surface slope, aspect, and other extraneous factors (Lillesand et al., 2004). Higher values of NDVI are represented by lighter tones and lower values are represented by darker tones. The values of NDVI decrease from south to north across the state. Occurrences seem to be clustered in areas of higher values of NDVI. A unique normalization approach presented by O'Sullivan and Unwin (2003) illustrates a possible connection between green-up and the pattern of case occurrences in 2002.

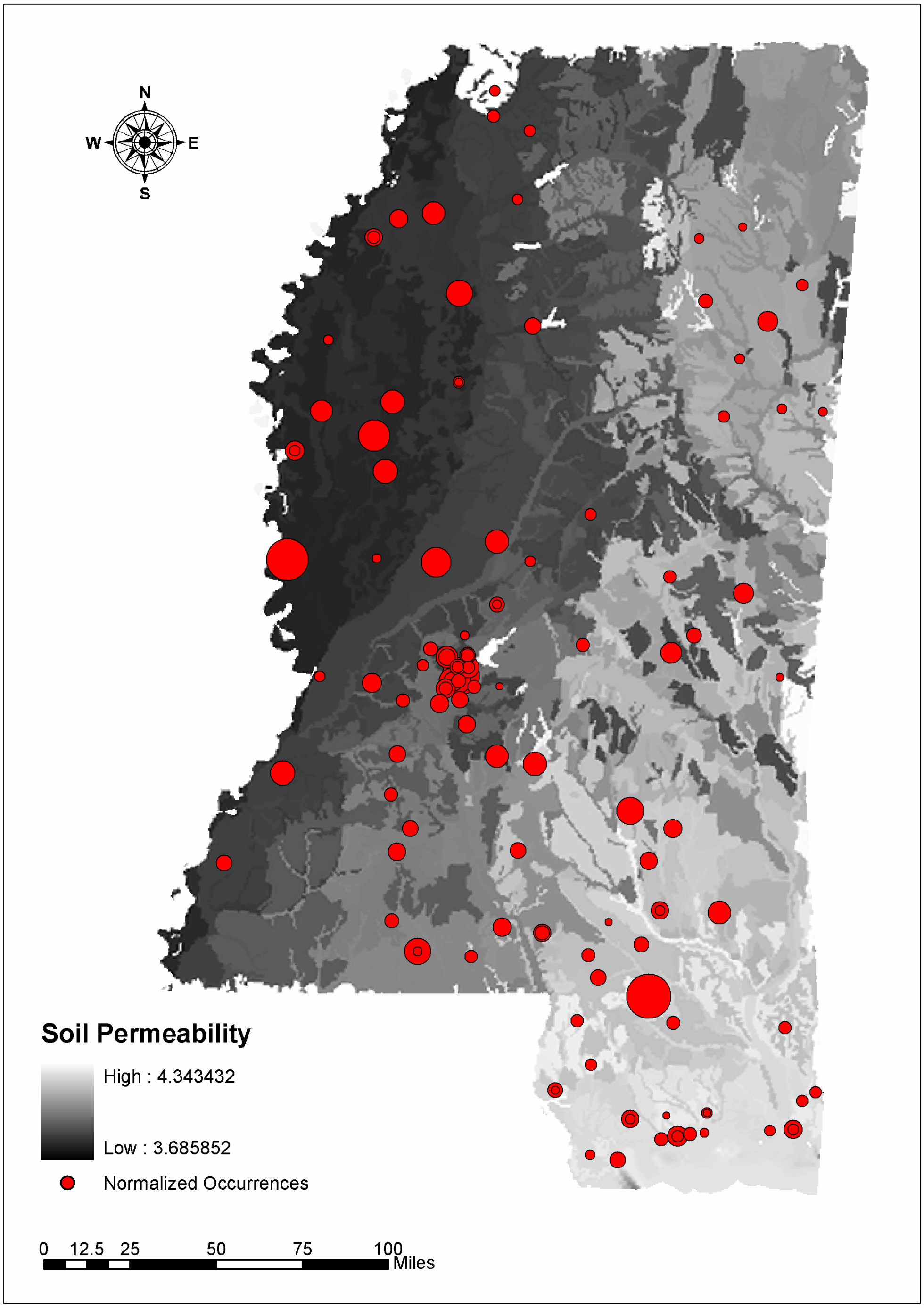

Figure 5: Pattern of Normalized Case Occurrences vs. Soil Permeability

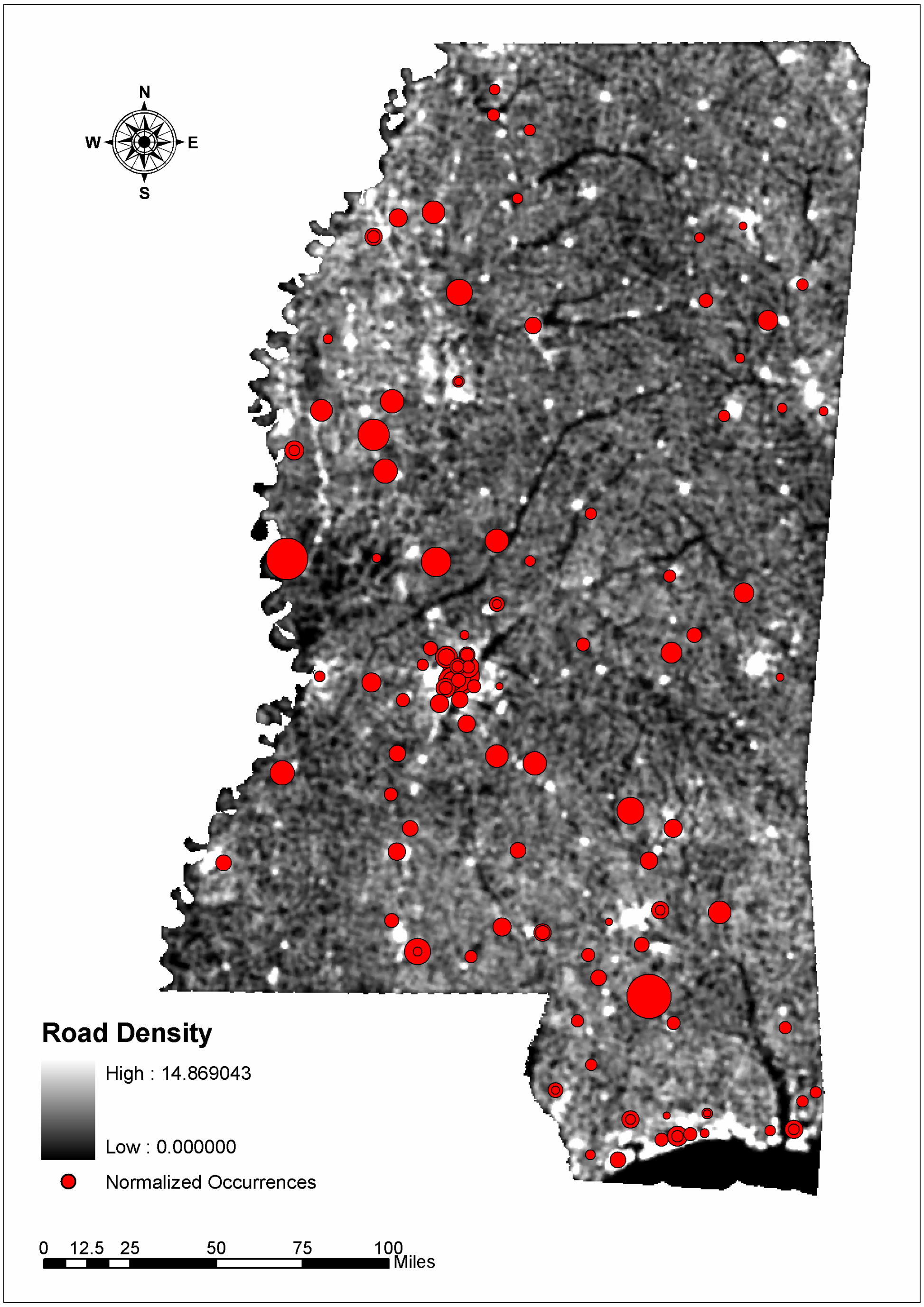

Figure 6: Pattern of Normalized Case Occurrences vs. Road Density

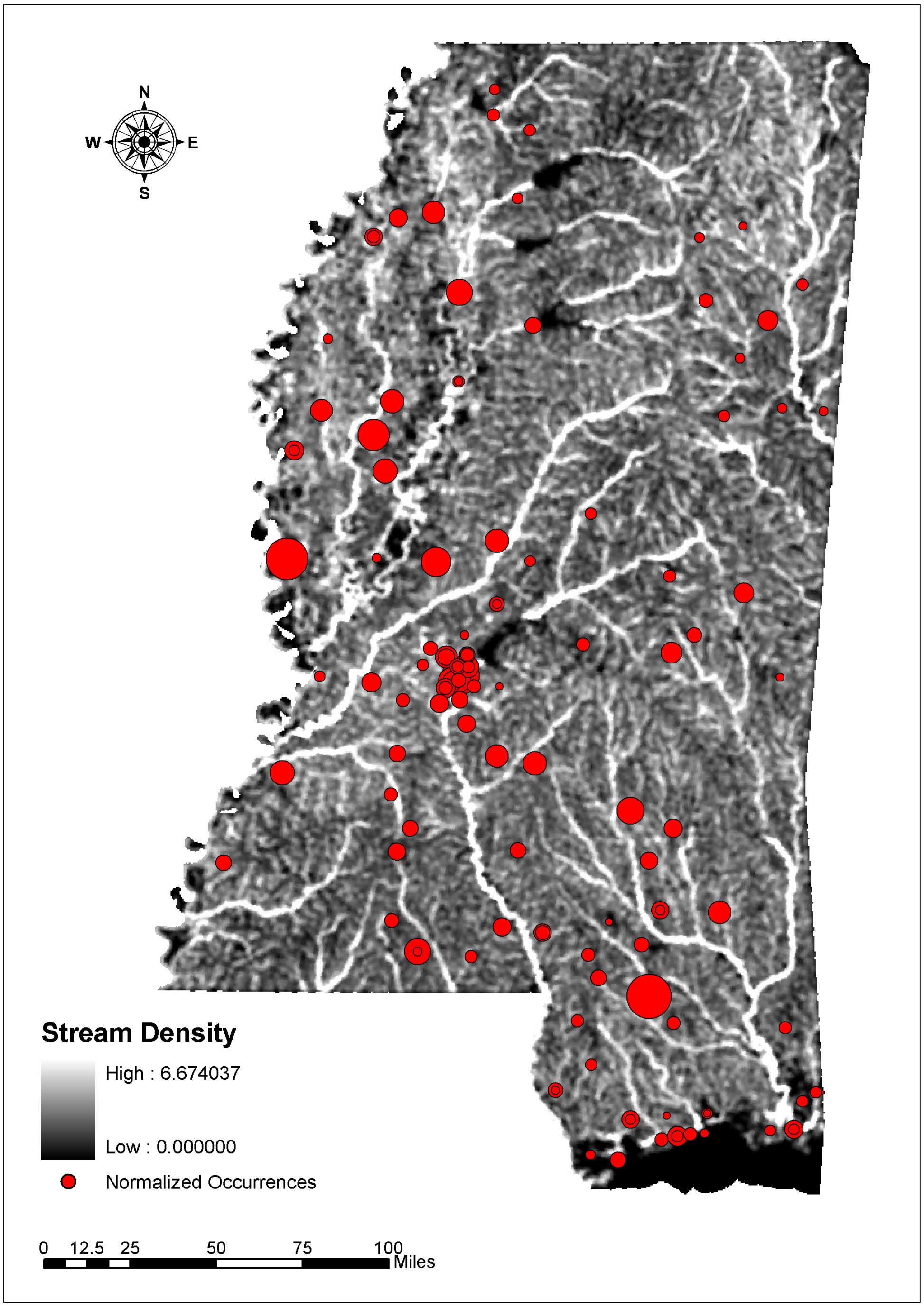

Figure 7: Pattern of Normalized Case Occurrences vs. Stream Density

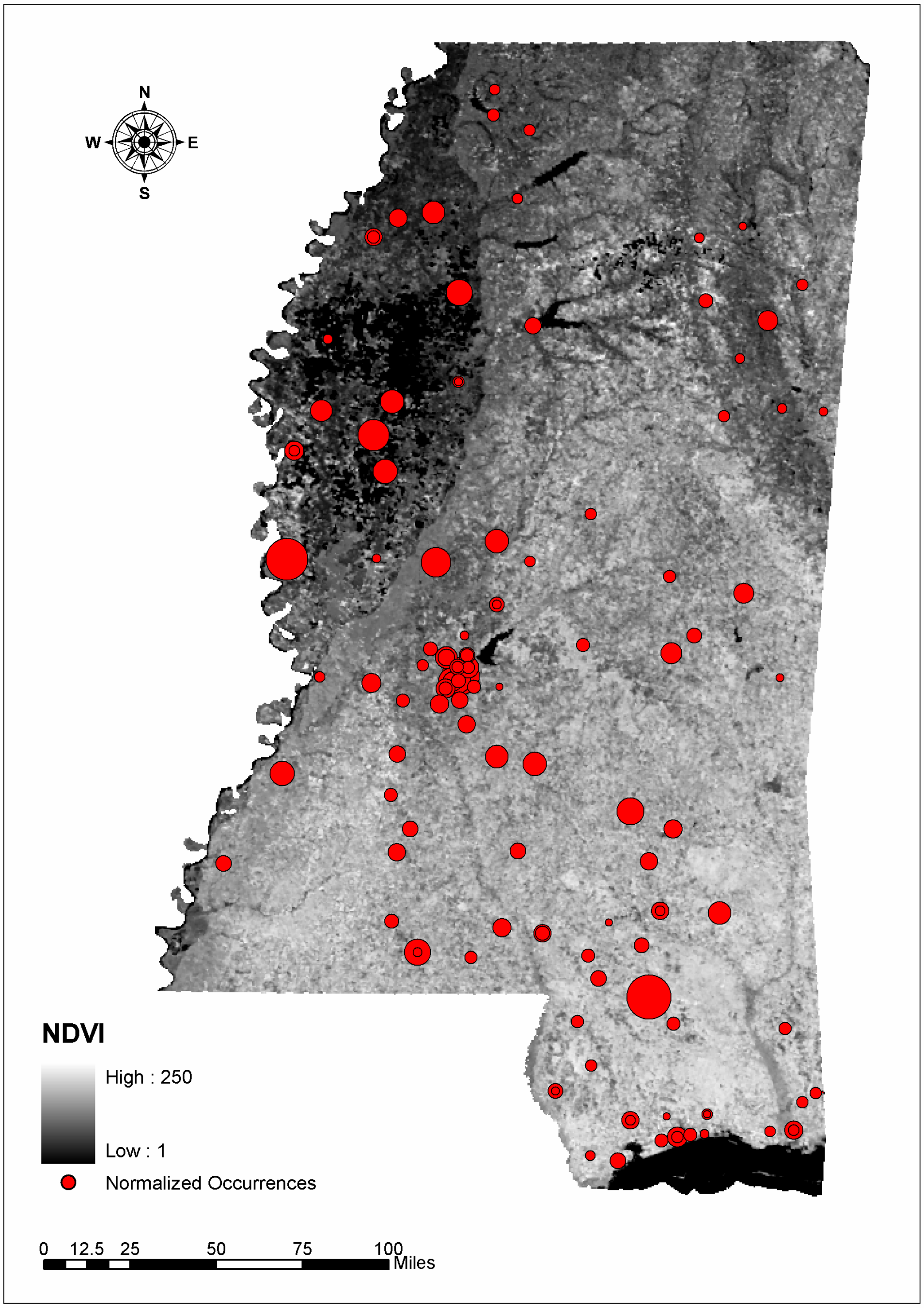

Figure 8: Pattern of Normalized Case Occurrences vs. Normalized Difference Vegetation Index (NDVI)

Pattern of Normalized 2002 Summer Case Occurrences vs. 2002 Summer Precipitation Minus Evaporation (P-E)

Figure 9 shows the normalized summer case occurrences compared with summer precipitation minus evaporation (P-E). Dark green tones indicate high values of P-E while lighter tones indicate lower values of P-E. It is difficult to determine if patterns exist. There are, however, clusters of occurrences within higher areas of P-E, suggesting that P-E may be an important variable in predicting mosquito habitat suitability and ultimately West Nile virus risk. This is intuitively appealing if one accepts the premise that as the amount of water increases the chances of mosquito habitat also increases.

Pattern of Normalized 2002 Fall Case Occurrences vs. 2002 Fall Precipitation Minus Evaporation (P-E)

Figure 10 shows the normalized fall case occurrences compared with fall P-E. As with Figure 9, dark green tones indicate higher values of P-E and light tones indicate lower values of P-E. The moisture regime here is more uniform than in the summer. Also, there are fewer occurrences in the fall. The occurrences that are present are located in areas of relatively high P-E, hinting to the fact that P-E may be an important variable in predicting mosquito habitat suitability. These results substantiate conclusions drawn by Githeko et al. (2000).

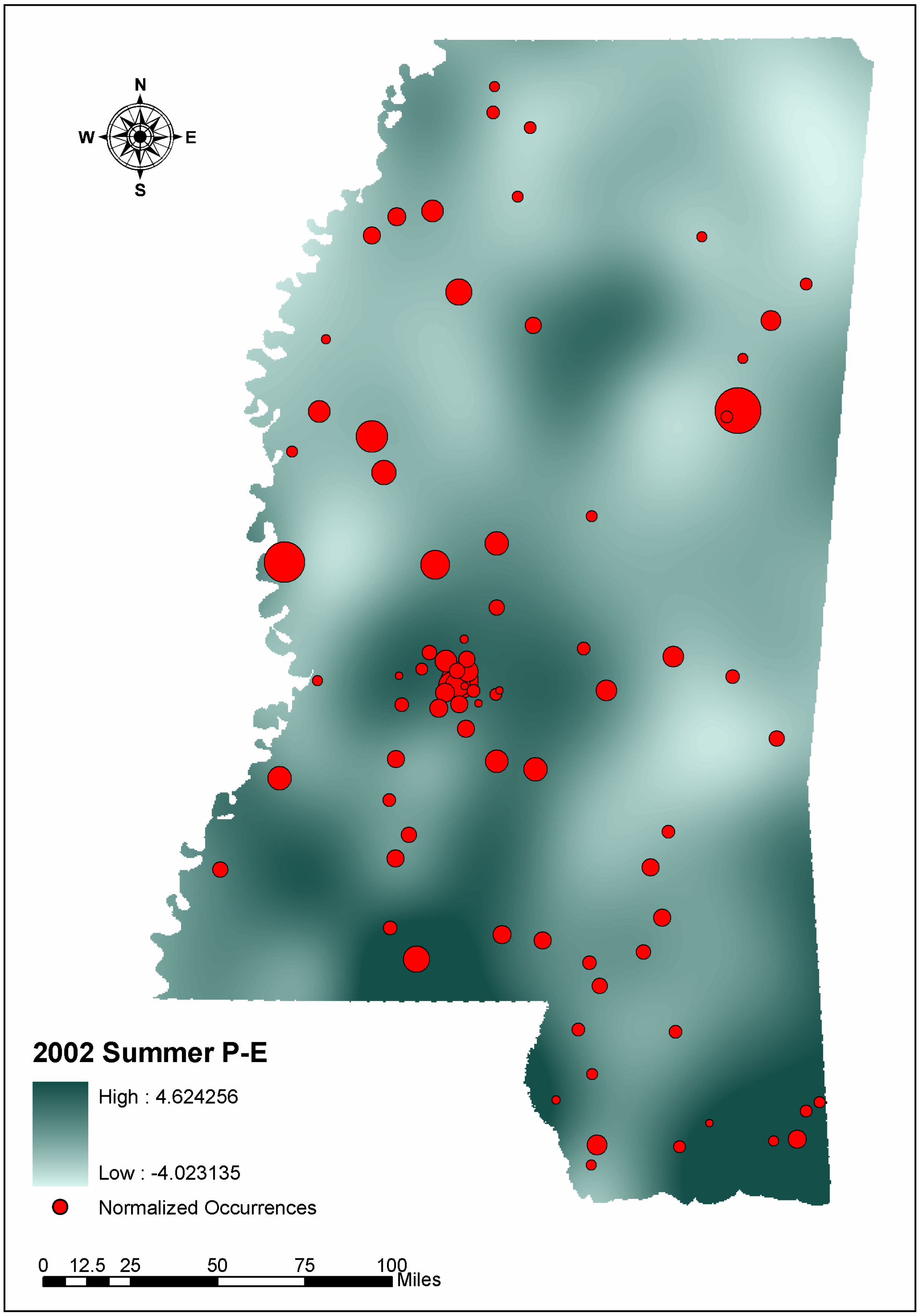

Figure 9: Pattern of Normalized 2002 Summer Case Occurrences vs. 2002 Summer Precipitation Minus Evaporation (P-E)

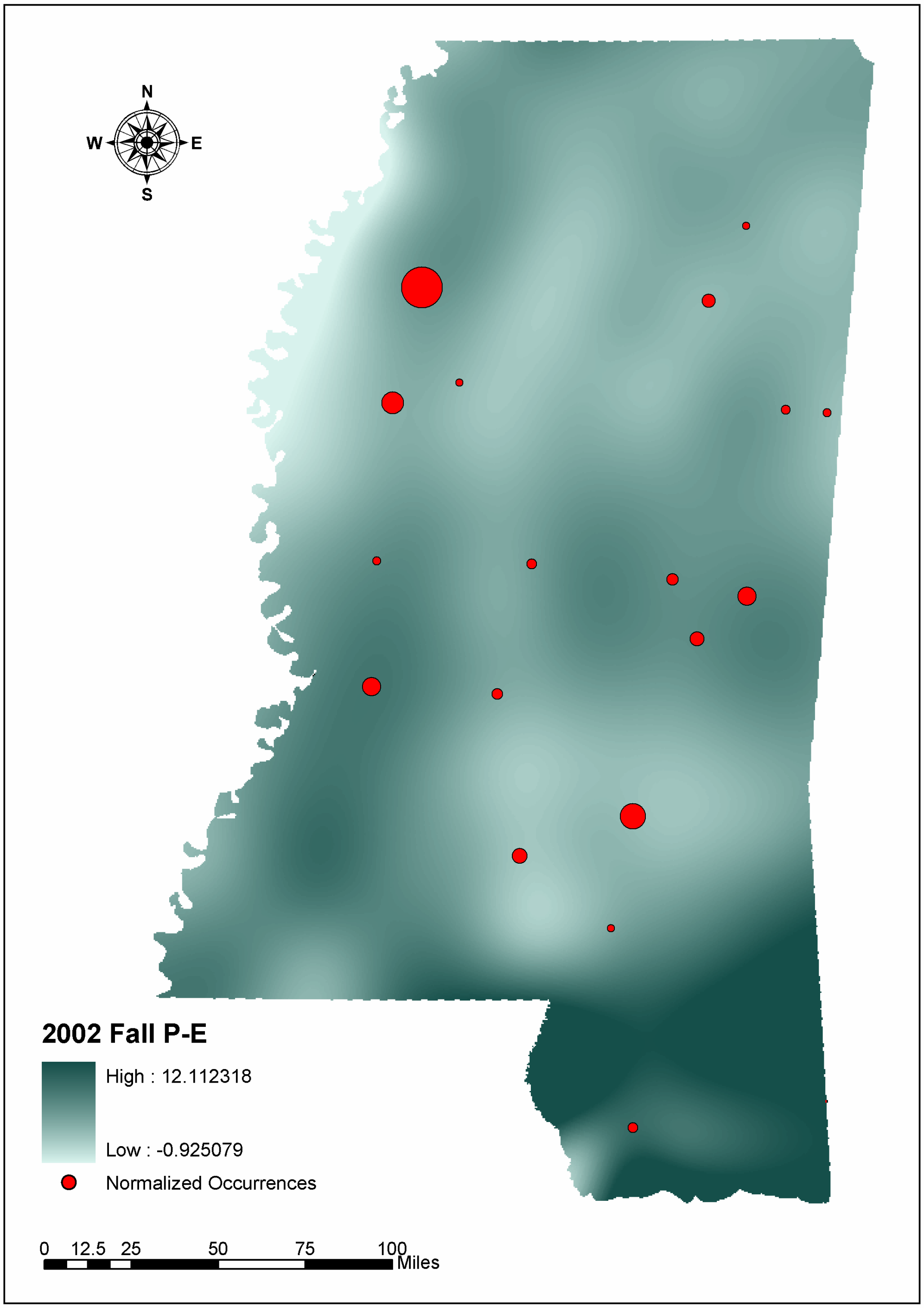

Figure 10: Pattern of Normalized 2002 Fall Case Occurrences vs. 2002 Fall Precipitation Minus Evaporation (P-E)

Pattern of Normalized 2003 Summer Case Occurrences vs. 2003 Summer Precipitation Minus Evaporation (P-E)

Figure 11 shows the normalized summer case occurrences compared with summer P-E. As with all of the other P-E figures, dark green tones indicate high values of P-E while lighter tones indicate lower values of P-E. Again, it is difficult to accurately determine if patterns exist; however, there seems to be clustering of larger diameter points in areas of higher values of P-E. There are more occurrences within areas of relatively higher P-E than in areas of lower P-E.

Pattern of Normalized 2003 Fall Case Occurrences vs. 2003 Fall Precipitation Minus Evaporation (P-E)

Figure 12 shows the normalized fall case occurrences compared with fall P-E. For this figure, P-E values seem to be more evenly distributed across the state, less concentrations of high and low values in a single location. Visually this figure, as opposed to the other P-E figures, displays the least correlation between high values of P-E and West Nile virus occurrence. Points are located in both areas of high and low values of P-E.

Although Figure 12 was less revealing than the other P-E figures, visualization of P-E variables suggests that a predominance of cases seem to fall into areas of higher relative moisture regimes. Visualization of the environmental variables suggests that patterns do exist but also raises more questions.

Visual analyses are a time-honored way of viewing patterns and speculating on the underlying processes that control the patterns (O'Sullivan and Unwin, 2003).

Today's GIS modeling capabilities can be combined with statistical analyses to help quantify these relationships and validate model outputs. The following sections present the results of the statistical analyses performed.

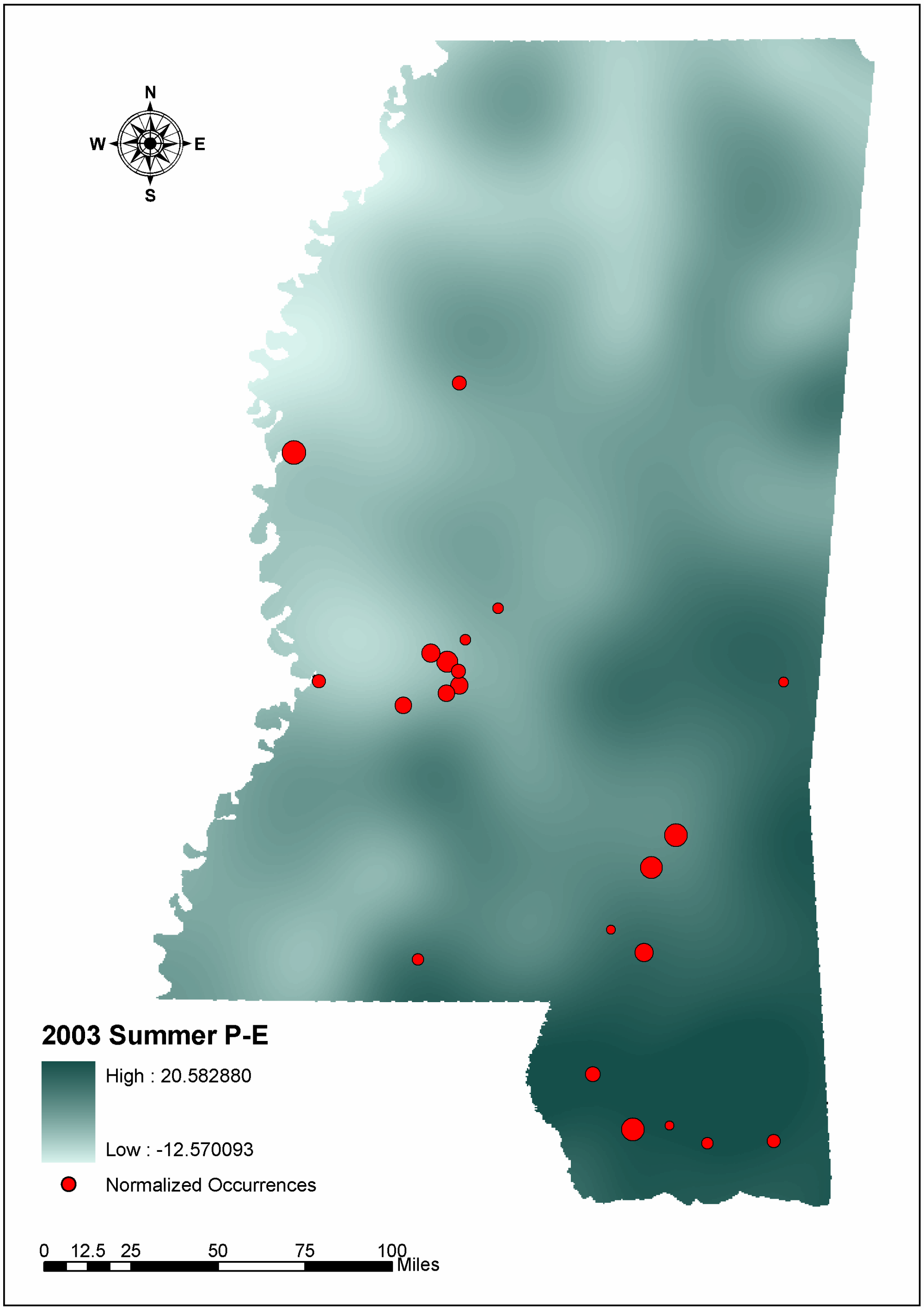

Figure 11: Pattern of Normalized 2003 Summer Case Occurrences vs. 2003 Summer Precipitation Minus Evaporation (P-E)

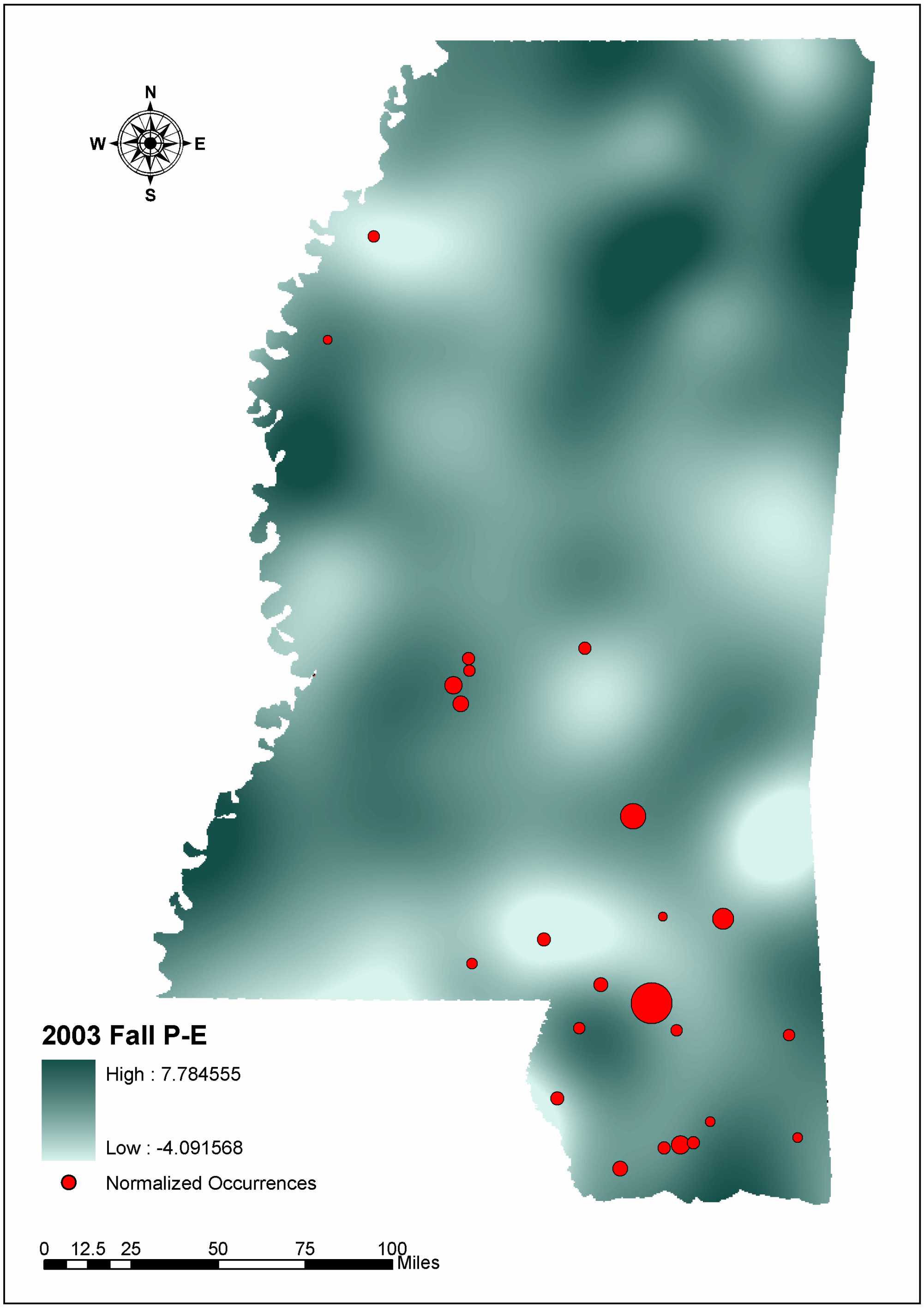

Figure 12: Pattern of Normalized 2003 Fall Case Occurrences vs. 2003 Fall Precipitation Minus Evaporation (P-E)

Statistical Tests for Each Variable of Interest

Data on West Nile virus infections are case occurrences summarized by zip code. The mean response for variables of West Nile virus occurrence versus variables of non-occurrence is compared using a t-test at the 95% confidence level. The two-tailed significance values were used for ranking variables. This is discussed in detail later in this chapter. Linear regressions were developed for the variables that showed significant differences between zip codes of West Nile virus occurrence versus zip codes of non-occurrence to determine the strength and direction of relationships between the significant variable and rate of occurrence.

T-Test for Slope

The t-test was performed to determine if there were significant differences between zip codes of occurrence and zip codes of non-occurrences for each variable. Table 2 shows the results of the t-test for slope weighted by case occurrence. It should be noted that within the "Group Statistics" table, one (1) represents zip codes with occurrences while two (2) represents zip codes without occurrences; this will hold true for the remainder of the t-test results. This test was performed with weighted occurrences. This means that if a zip code recorded more than one occurrence, the record was duplicated to match the number of occurrences within the t-test design. As a result of the high significance value (P-value = 0.001), equal variances were not assumed and its associated two-tailed significance value was recorded for variable ranking. It should be noted that the remaining t-tests were constructed with weighted occurrences.

Table 2
Results of the T-Test for Slope Weighted by Case Occurrences

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
SLOPE	1	164	6.3778	2.87779	.22472
	2	279	7.6212	3.77110	.22577

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SLOPE	Equal variances assumed	12.252	.001	-3.644	441	.000	-1.2435	.34122	-1.91408	-.57284
	Equal variances not assumed			-3.904	412.006	.000	-1.2435	.31854	-1.86963	-.61729

As previously mentioned in this chapter, linear regressions were developed for the variables that showed significant differences between zip codes of West Nile virus occurrence versus zip codes of non-occurrence. The goal of the regression procedure was to determine the strength and direction of relationships between the significant variable and rate of West Nile virus occurrence. Table 3 shows the results of the regression of case count on slope. An extremely weak linear relationship exists between case counts and slope ($R^2 = 0.011$). However, this relationship will become important during the ranking and weighting of the variables.

Table 3
Regression of Case Counts on Slope

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.104[a]	.011	-.004	1.66618

T-Test for Soil Permeability

Table 4 shows the results of the t-test for soil permeability. As with the t-test for slope, one (1) represents zip codes with occurrences while two (2) represents zip codes without occurrences. Equal variances are assumed for zip codes of West Nile virus occurrence versus non-occurrence based on a non-significant P-value (0.136). The test for equality of means resulted in a non-significant P-value (0.872). A review of the means 3.8485 (occurrences) versus 3.8479 (non-occurrences) suggests that soil permeability is not related to WNV occurrence. The associated two-tailed significance value is recorded for use later in variable rankings.

Table 4
Results of the T-Test for Soil Permeability

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
PERM	1	164	3.8485	.03785	.00296
	2	279	3.8479	.03994	.00239

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
PERM	Equal variances assumed	2.233	.136	.162	441	.872	.0006	.00386	-.00695	.00820
	Equal variances not assumed			.164	356.594	.870	.0006	.00380	-.00685	.00810

T-Test for Stream Density

The results of the t-test for stream density are shown in Table 5. Note the values for mean stream density, 1.1977 (occurrences) versus 1.1571 (non-occurrences). As with the means of permeability, there is little difference between mean stream densities within zip codes of occurrences versus zip codes of non-occurrences. A non-significant P-value (0.946) verifies this statement. Equal variances are assumed and the associated two-tailed significance value is recorded for use later in variable rankings. For these data, there is no evidence that a relationship exists between human occurrences and stream density.

T-Test for Road Density

Table 6 shows the results of the t-test for road density. Equal variances for zip codes of West Nile virus occurrence versus non-occurrence is not assumed based on a significant P-value (0.000) for the test of equal variances. The test for equality of means resulted in a significant P-value (0.000) leading to the assumption that road density is significantly different for zip codes of West Nile virus occurrence versus zip codes of non-occurrence. A review of the means 2.4841 (occurrences) and 1.2198 (non-occurrence) indicates that higher values of road density are related to WNV occurrence and increased risk. The associated two-tailed significance value is recorded for use later in variable rankings.

Table 5
Results of the T-Test for Stream Density

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
STREAM	1	164	1.1977	.24115	.01883
	2	279	1.1571	.24526	.01468

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
STREAM	Equal variances assumed	.004	.948	1.690	441	.092	.0405	.02398	-.00661	.08767
	Equal variances not assumed			1.697	346.391	.091	.0405	.02388	-.00643	.08750

Table 6
Results of the T-Test for Road Density

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
ROAD_DEN	1	164	2.4841	2.38049	.18588
	2	279	1.2198	.80401	.04814

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ROAD_DEN	Equal variances assumed	130.616	.000	8.123	441	.000	1.2643	.15564	.95836	1.57014
	Equal variances not assumed			6.584	185.105	.000	1.2643	.19202	.88543	1.64307

Table 7 shows the results of the regression of case count on road density. As with slope, a weak linear relationship exists between case counts and road density (R^2 = 0.219). This relationship will also become more important during the ranking and weighting of the variables.

Table 7
Regression of Case Counts on Road Density

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.468[a]	.219	.208	1.48011

T-Test for NDVI

Table 8 shows the results of the t-test for NDVI. The means for zip codes of occurrences and non-occurrences appear to be significantly different, 166.2011 for zip codes with occurrences and 161.5807 for zip codes without occurrences. Equal variances for zip codes of West Nile virus occurrence versus non-occurrence are not assumed based on a significant P-value (0.002) for the test of equal variances. The test for equality of means resulted in a non-significant P-value (0.105). As with the other test results, the associated two-tailed significance value is recorded for use later in variable rankings.

T-Test for 2002 Summer P-E

The results of the t-test for 2002 Summer P-E are shown in Table 9. Note the negative mean values for zip codes of non-occurrence. This is indicative of a drought condition. Negative values of P-E suggest that there is a water deficit. Equal variances are not assumed for zip codes of West Nile virus occurrence versus non-occurrence based on a significant P-value (0.001). The test for equality of means resulted in a significant P-value (0.000) leading to the assumption that 2002 Summer P-E is significantly different for zip codes of WNV occurrence versus zip codes of non-occurrence. A review of the means 0.2522 (occurrences) and -1.2213 (non-occurrences) indicates that higher values of 2002 Summer P-E are related to WNV occurrence and increased risk.

Table 8
Results of the T-Test for Normalized Difference Vegetation Index (NDVI)

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
NDVI	1	164	166.2011	27.10306	2.11639
	2	279	161.5807	31.65999	1.89543

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NDVI	Equal variances assumed	9.559	.002	1.562	441	.119	4.6204	2.95742	-1.19200	10.43278
	Equal variances not assumed			1.626	384.361	.105	4.6204	2.84109	-.96563	10.20641

Table 9
Results of the T-Test for 2002 Summer Precipitation Minus Evaporation (P-E)

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
SUM_PE	1	148	.2522	1.95908	.16104
	2	292	-1.2213	1.53177	.08964

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SUM_PE	Equal variances assumed	12.046	.001	8.655	438	.000	1.4735	.17025	1.13886	1.80809
	Equal variances not assumed			7.995	240.545	.000	1.4735	.18430	1.11042	1.83653

Table 10 shows the results of the regression of case count on 2002 Summer P-E. A weak linear relationship ($R^2 = 0.108$) exists between case counts and higher values of P-E. This will become more important during the ranking and weighting of the variables.

Table 10
Regression of Case Counts on 2002 Summer Precipitation Minus Evaporation (P-E)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.329[a]	.108	.095	1.59085

T-Test for 2002 Fall P-E

Table 11 shows the results of the t-test for 2002 Fall P-E. Equal variances are assumed for zip codes of West Nile virus occurrence versus non-occurrence based on a non-significant P-value (0.869). The test for equality of means resulted in a non-significant P-value (0.573). A review of the means 3.7574 (occurrences) and 3.5916 (non-occurrences) suggests that 2002 Fall P-E is not related to WNV occurrence. The associated two-tailed significance value is recorded for use later in variable rankings.

Table 11
Results of the T-Test for 2002 Fall Precipitation Minus Evaporation (P-E)

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
FALL_PE	1	17	3.7574	1.11890	.27137
	2	343	3.5916	1.18684	.06408

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
FALL_PE	Equal variances assumed	.027	.869	.564	358	.573	.1658	.29416	-.41274	.74428
	Equal variances not assumed			.595	17.832	.560	.1658	.27884	-.42044	.75198

T-Test for 2003 Summer P-E

Table 12 shows the results of the t-test for 2003 Summer P-E. Equal variances for zip codes of West Nile virus occurrence versus non-occurrence are not assumed based on a significant P-value (0.000). The test for equality of means resulted in a significant P-value (0.000) leading to the assumption that 2003 Summer P-E is significantly different for zip codes of WNV occurrence versus zip codes of non-occurrence. A review of the means 3.6259 (occurrences) versus -1.2213 (non-occurrences) indicates that higher values of 2003 Summer P-E are related to WNV occurrence and increased risk. The associated two-tailed significance value is recorded for use later in variable rankings. Although 2003 Summer P-E is significant, a linear regression was not developed as a result of the low ‘N’ of 45.

T-Test for 2003 Fall P-E

The results of the t-test for 2003 Fall P-E are displayed in Table 13. Equal variances are assumed for zip codes of West Nile virus occurrence versus non-occurrence based on a non-significant P-value (0.920). The test for equality of means resulted in a non-significant P-value (0.086). A review of the means 1.9773 (occurrences) and 2.5375 (non-occurrences) suggests that 2003 Fall P-E is not related to WNV occurrence. The associated two-tailed significance value is recorded for use later in variable rankings.

Table 12
Results of the T-Test for 2003 Summer Precipitation Minus Evaporation (P-E)

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
SUM_PE	1	45	3.6259	8.23223	1.22719
	2	292	-1.2213	1.53177	.08964

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SUM_PE	Equal variances assumed	579.550	.000	9.151	335	.000	4.8472	.52968	3.80530	5.88912
	Equal variances not assumed			3.939	44.471	.000	4.8472	1.23046	2.36813	7.32629

Table 13
Results of the T-Test for 2003 Fall Precipitation Minus Evaporation (P-E)

Group Statistics

	VALUE	N	Mean	Std. Deviation	Std. Error Mean
FALL_PE	1	41	1.9773	1.84901	.28877
	2	338	2.5375	1.98214	.10781

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
FALL_PE	Equal variances assumed	.010	.920	-1.721	377	.086	-.5602	.32553	-1.20025	.07991
	Equal variances not assumed			-1.817	51.810	.075	-.5602	.30824	-1.17875	.05841

Variable Manipulation

Variable Ranks

Each variable of interest was ranked in terms of its t-test probability level. It should be noted that t-tests can lead to acceptance of variables as significant 5% of the time (Type I error). For example, the rate at which you declare results to be significant when there are no relationships in the population. It is the rate of false alarms or false positives. Nevertheless, being aware of this error led to efforts to perform more advanced statistical procedures, i.e. Logistic Regression.

There were two sets of rankings for both 2002 and 2003, one for the summer model and one for the fall model. Summer P-E was removed for the fall model and Fall P-E was removed for the summer model. This resulted in different variable weights for each model. For the 2002 and 2003 summer models, the ranks from most important to least important were as follows: Road Density, Summer P-E, Slope, Permeability, NDVI, and Stream Density. For the 2002 and 2003 fall model, the ranks from most important to least important were: Road Density, Slope, Permeability, NDVI, Stream Density, and P-E Fall.

Variable Weights

Variable weights were determined by summing the ranks and then dividing each rank by that sum. As mentioned in Chapter III, each variable was conditioned from 10 - 1, with ten representing highest potential risk and one representing lowest potential risk. As a result of the conditioning, the final rankings needed to follow the same pattern from high to low. Because there are six variables for each model, the most important variable received a rank of one but a value of six.

Table 14 (2002 variable rankings) and Table 15 (2003 variable rankings) show each of the variables of interest, their t-test significance, R^2 values where appropriate, their rankings, and their assigned weights. The sum of the ranks equals twenty-one. Because road density is the most important variable, for both years, receiving a rank of one, you actually divide six by twenty-one to get a weight of 0.29. A rank of two results in a value of five and a weight of 0.24, and so forth, until each variable has its corresponding weight. Table 16 will help clarify this methodology.

Once the variables were ranked in order of t-test significance and weights were assigned, linear additive models were constructed using the conditioned variables for summer and fall. Linear additive models were constructed by multiplying each variable by its associated rank and then adding those products:

2002/2003 Summer Model

([road_density] * 0.29) + ([p-e_summer] * 0.24) + ([slope] * 0.19) + ([permeability] * 0.14) + ([ndvi] * 0.10) + ([stream_density] * 0.05)

2002/2003 Fall Model

([road_density] * 0.29) + ([slope] * 0.24) + ([permeability] * 0.19) + ([ndvi] * 0.14) + ([stream_density] * 0.10) + ([p-e_fall] * 0.05)

Table 14
Variable Manipulation for the Final 2002 Models

Variable	T-test Significance	R2 *	Rank Summer/Fall	Weight Summer/Fall
Road Density	.000	.219	1/1	.29/.29
Stream Density	.092	-	4/3	.14/.19
Slope	.000	.011	3/2	.19/.24
NDVI	.105	-	5/4	.10/.14
P-E Summer	.000	.108	2/0	.24/.00
P-E Fall	.573	-	0/5	.00/.10
Permeability	.872	-	6/6	.05/.05

* Calculated for significant variables

Table 15
Variable Manipulation for the Final 2003 Models

Variable	T-test Significance	R2 *	Rank Summer/Fall	Weight Summer/Fall
Road Density	.000	.219	1/1	.29/.29
Stream Density	.092	-	4/4	.14/.14
Slope	.000	.011	3/2	.19/.24
NDVI	.105	-	5/5	.10/.10
P-E Summer	.000	**	2/0	.24/.00
P-E Fall	.086	-	0/3	.00/.19
Permeability	.872	-	6/6	.05/.05

*Calculated for significant variables
** Not enough samples to calculate

Table 16
Explanation of Variable Ranks and Weights

Rank	Value	Variable	Weight
1	6	Road Density	6/21 = .29
2	5	Summer P-E	5/21 = .24
3	4	Slope	4/21 = .19
4	3	Stream Density	3/21 = .14
5	2	NDVI	2/21 = .10
6	1	Permeability	1/21 = .05
21			

Additive Model Results

Final 2002 Summer Additive Model

Figure 13 shows the output of the final 2002 summer model. As a method of visually validating the model, the normalized, 2003 summer occurrences were overlaid on the 2002 summer model. It is difficult to find an occurrence that did not appear in an area of "high" risk as determined by the model. Even though encouraging, the results were surprising. It was intuitively expected that the Mississippi Delta would be a high risk location; however, for this model run, that did not result. In fact, the Delta was relatively low risk as opposed to the Jackson metropolitan and Mississippi Gulf Coast areas. This is due in part to the precipitation regime. Referring back to Figure 9, Pattern of Summer Case Occurrences vs. 2002 Summer Precipitation Minus Evaporation (P-E), low values can be seen in the Delta. Also, 2002 Summer P-E displayed a high significance value, as determined by the t-test, resulting in a higher variable weight which exhibited greater influence on the model.

Final 2002 Fall Additive Model

Figure 14 shows the output of the final 2002 fall model. Similar to Figure 13, the normalized, 2003 fall occurrences were overlaid on the 2002 fall model as a way to visually validate the model results. As mentioned, it is difficult to find an occurrence that did not appear in an area of “high” risk as determined by the model. It should also be noted that, among the points representing the occurrences, the largest diameter points are in areas of relatively higher risk, for example, the southeastern portion of the state. This is the area with the highest risk and with the largest diameter points. There also seems to be a decrease in risk as you move from southeast to northwest across the state.

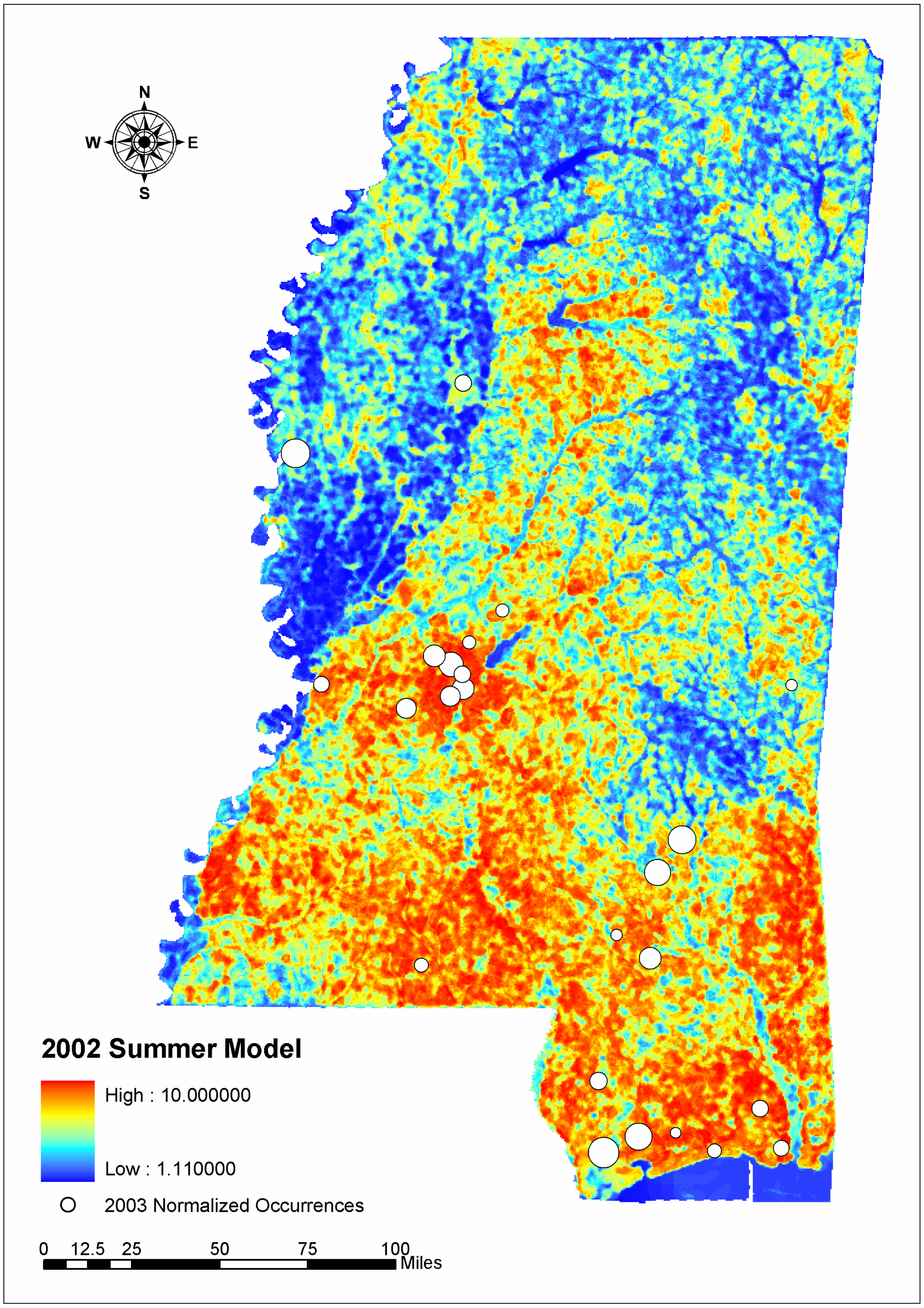

Figure 13: Final 2002 Summer Additive Model

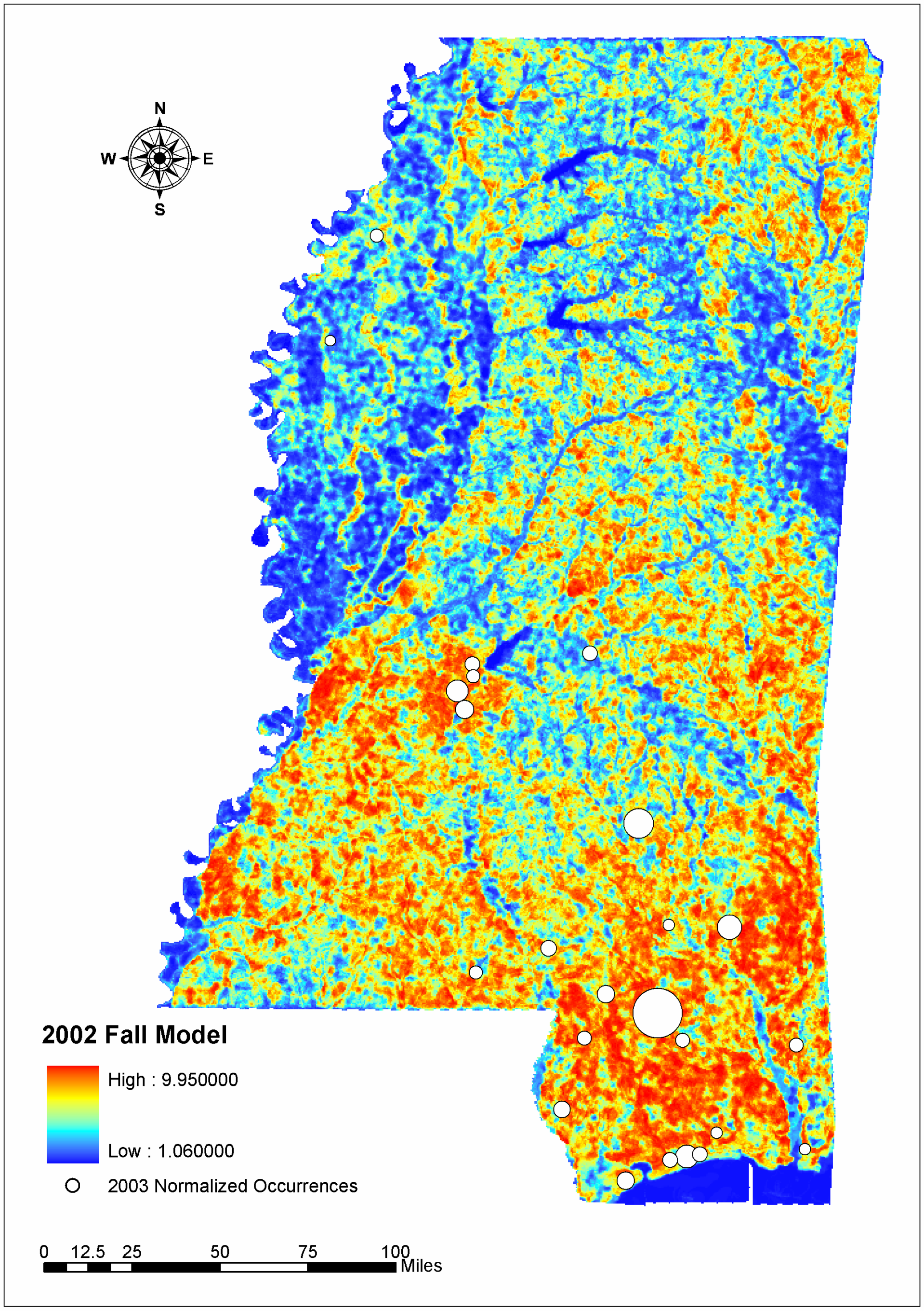

Figure 14: Final 2002 Fall Additive Model

Final 2003 Summer Additive Model

The results of the final 2003 summer model are shown in Figure 15. As with the previous models, the normalized, 2003 case occurrences were overlaid on the model as a way to visually validate the results. The Delta is relatively low risk, similar to the results from the final 2002 summer model. The southeast to northwest trend noticed in the previous model is even more pronounced for this model. Again, it is hard to find any case occurrences in areas of low risk. The largest diameter points are clustered in areas of highest risk.

Final 2003 Fall Additive Model

Figure 16 shows the results of the final 2003 fall model. As with all of the final additive models, the normalized, 2003 case occurrences were overlaid on the model as a way to visually validate the results. The general trend from southeast to northwest shown in the previous models was not depicted for the 2003 Fall Model. The risk can be explained in part by the precipitation regime for fall 2003, refer back to Figure 12. High values of P-E are scattered throughout the state.

Final Logistic Regression Model

For the Logistic Regression Model, as previously mentioned in Chapter III, the probability of occurrence in each zip code was calculated and linearized by taking the natural log of the probability of occurrence of West Nile virus in each zip code. Resulting probabilities were constrained between 0 – 1. Each zip code in the state was assigned a probability for occurrence of West Nile virus and the resulting probabilities were brought into the GIS system and interpolated across the state using an Inverse Distance Weighted (IDW) interpolation technique.

Figure 17 shows the results of the Logistic Regression Model. The resulting model shows the same trend as the previous additive models, a decreasing risk from southeast to northwest across the state. This consistency between the models is extremely encouraging. These results strengthen the additive model results which were based on less advanced statistics.

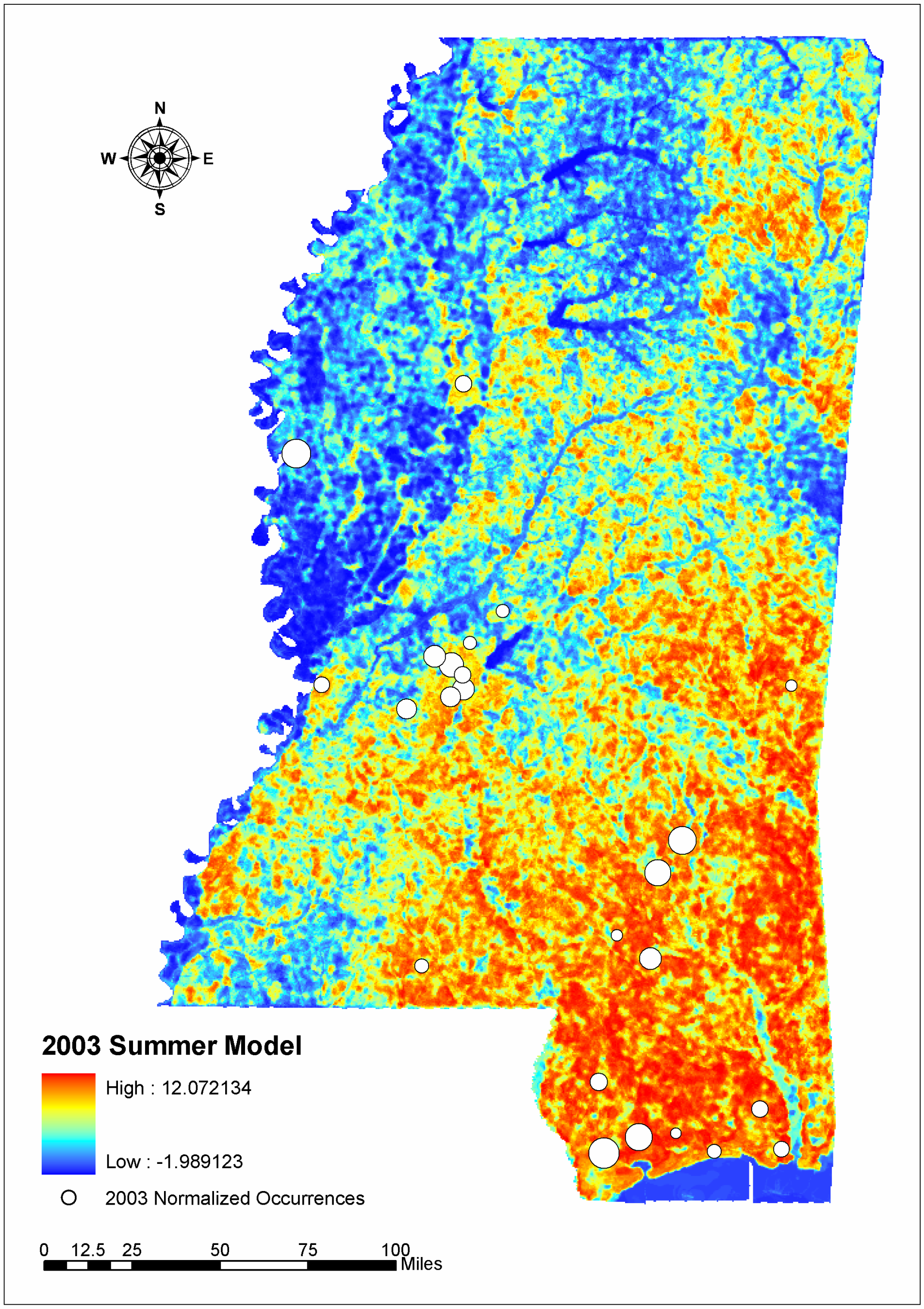

Figure 15: Final 2003 Summer Additive Model

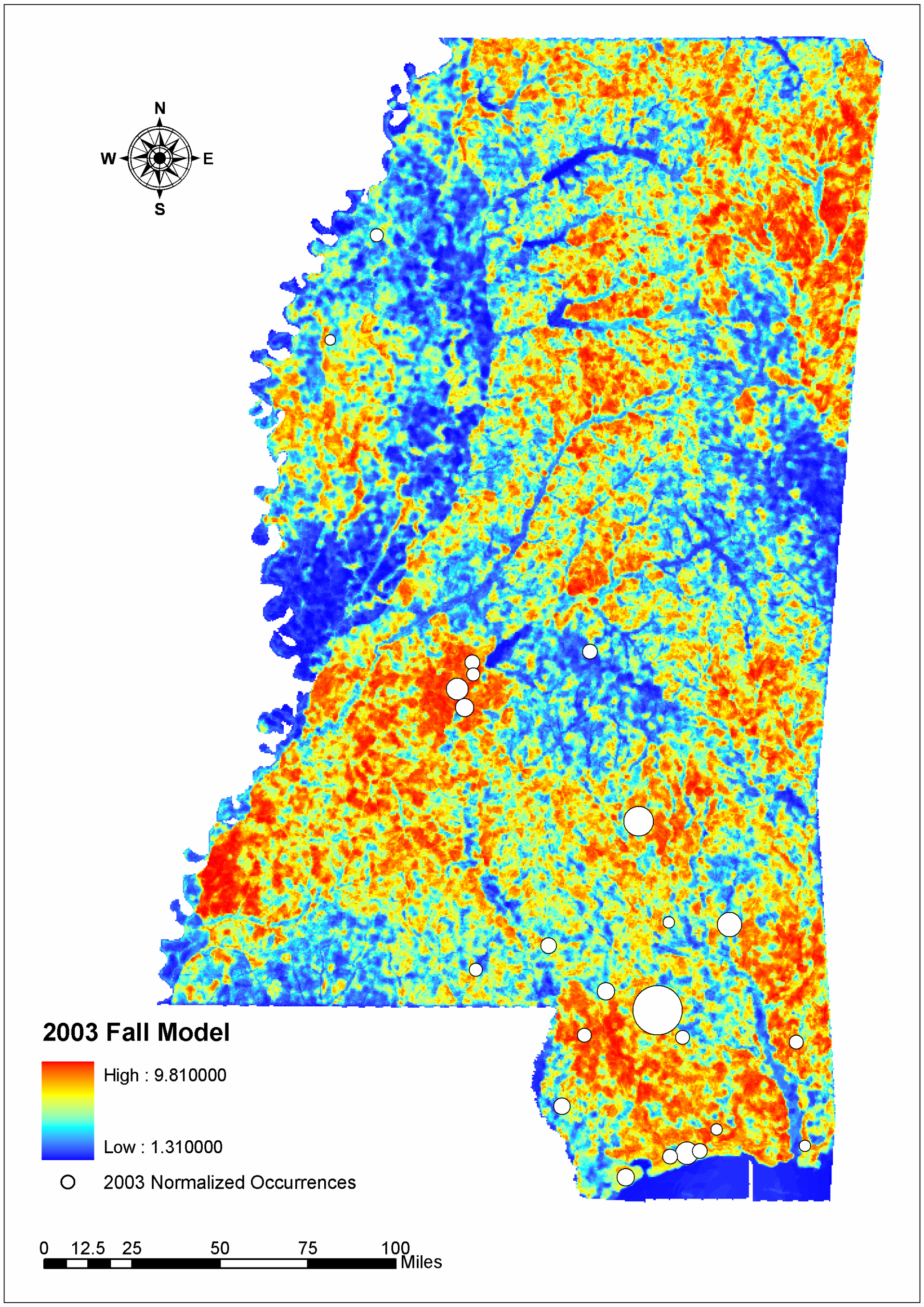

Figure 16: Final 2003 Fall Additive Model

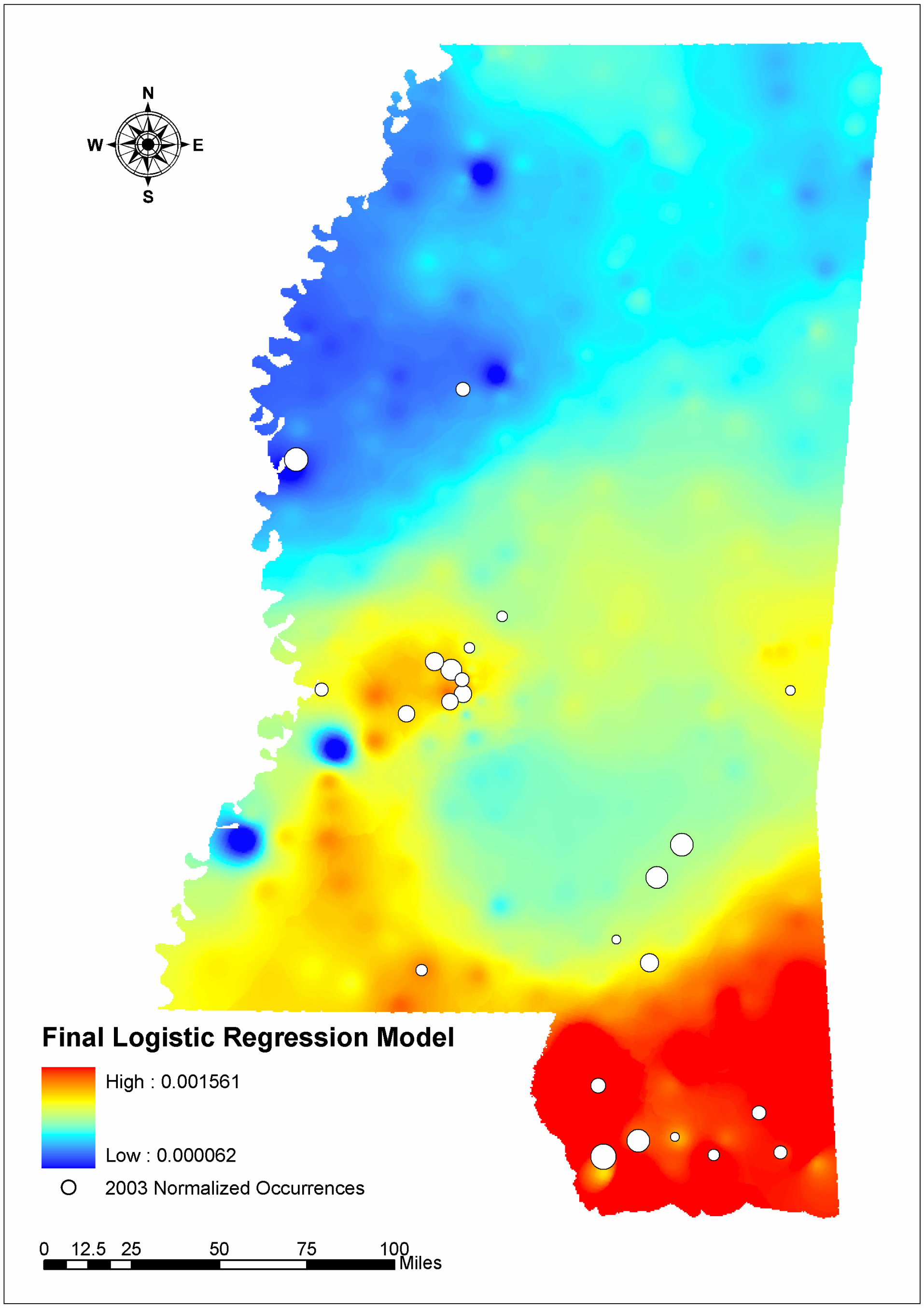

Figure 17: Final Logistic Regression Model

One of the main goals of this project was to relate mosquito habitat to general public risk in Mississippi from West Nile virus and specifically to natural resource managers and users of recreational facilities. In order to achieve this goal, predicted risk was determined for each state park and natural resource area in Mississippi as predicted by the Final 2002 Summer Additive Model as well as the Final Logistic Regression Model. Risk within each area of interest was determined by calculating the mean predicted risk using zonal statistics. The graphs of predicted risk for all state parks and all natural resource areas are provided in Appendix B. For the following figures, only the top-ten highest risked areas were graphed for interpretation.

Figure 18 shows the ten highest-risked state parks as predicted by the Final 2002 Summer Additive Model while Figure 19 shows the ten highest-risked state parks as predicted by the Final Logistic Regression Model. The Logistic Regression model and the Summer Additive Model agreed on seven out of the top-ten. The statistically-based Logistic Regression model approach agrees closely with the additive model results. Both models agree on seven out of ten state parks, (Lefleur's Bluff, Shepard, Roosevelt, Percy Quinn, Lake Lincoln, Paul B. Johnson, and Golden Memorial) with Lake Lincoln resulting in the same rank for both models.

Figure 20 shows the ten highest-risked natural resource areas as predicted by the Final 2002 Summer Additive Model while Figure 21 shows the ten highest-risked natural resource areas as predicted by the Final Logistic Regression Model. Similar to the previous results, there is general agreement for seven out of ten natural resource areas:

Gulf Island, Sandhill Crane, Homochitto, Desoto, Bogue Chitto, Bienville, and St. Catherine Creek.

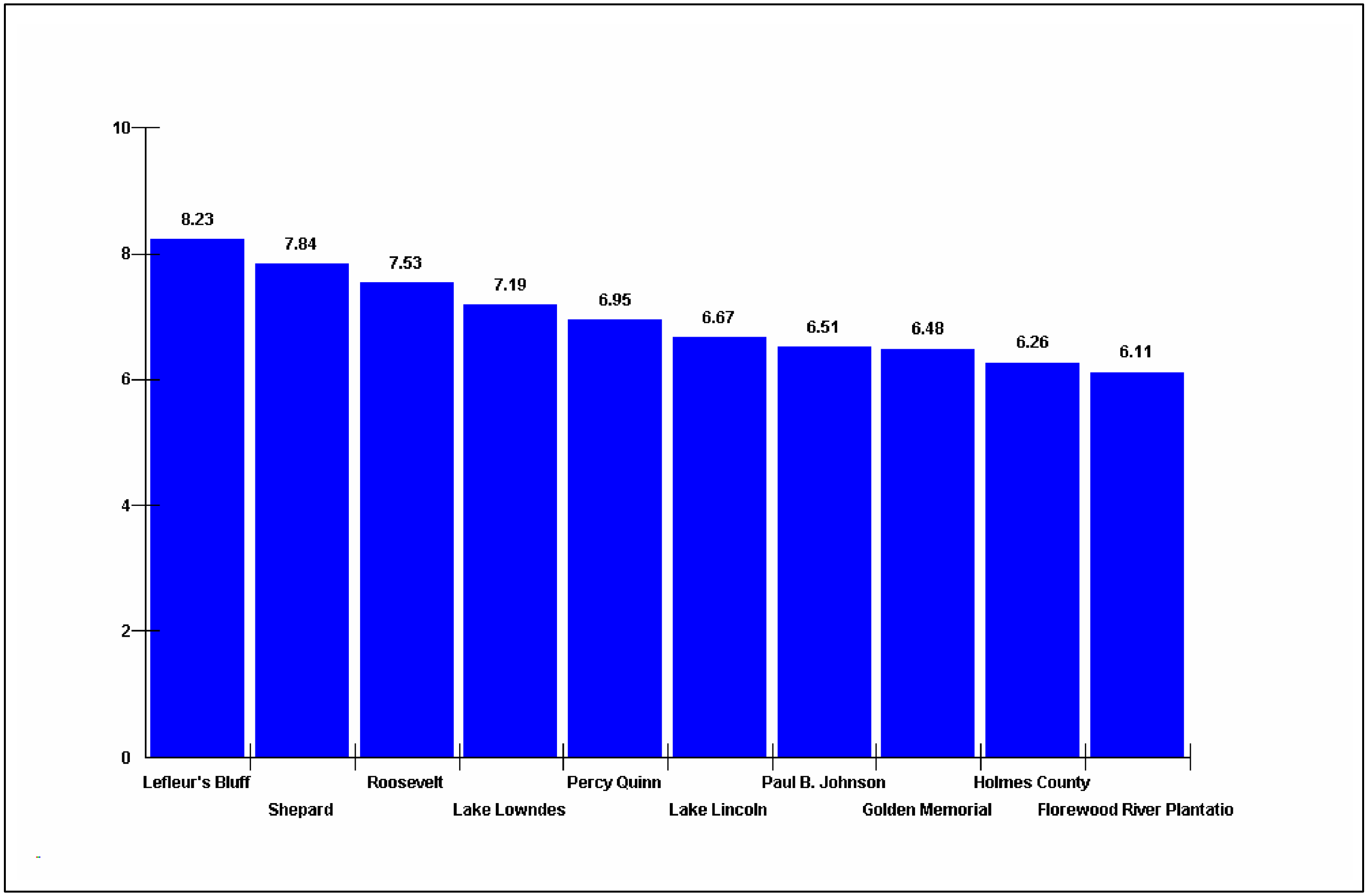

Figure 18: Top-Ten Highest Risked State Parks as Predicted by the Final 2002 Summer Additive Model

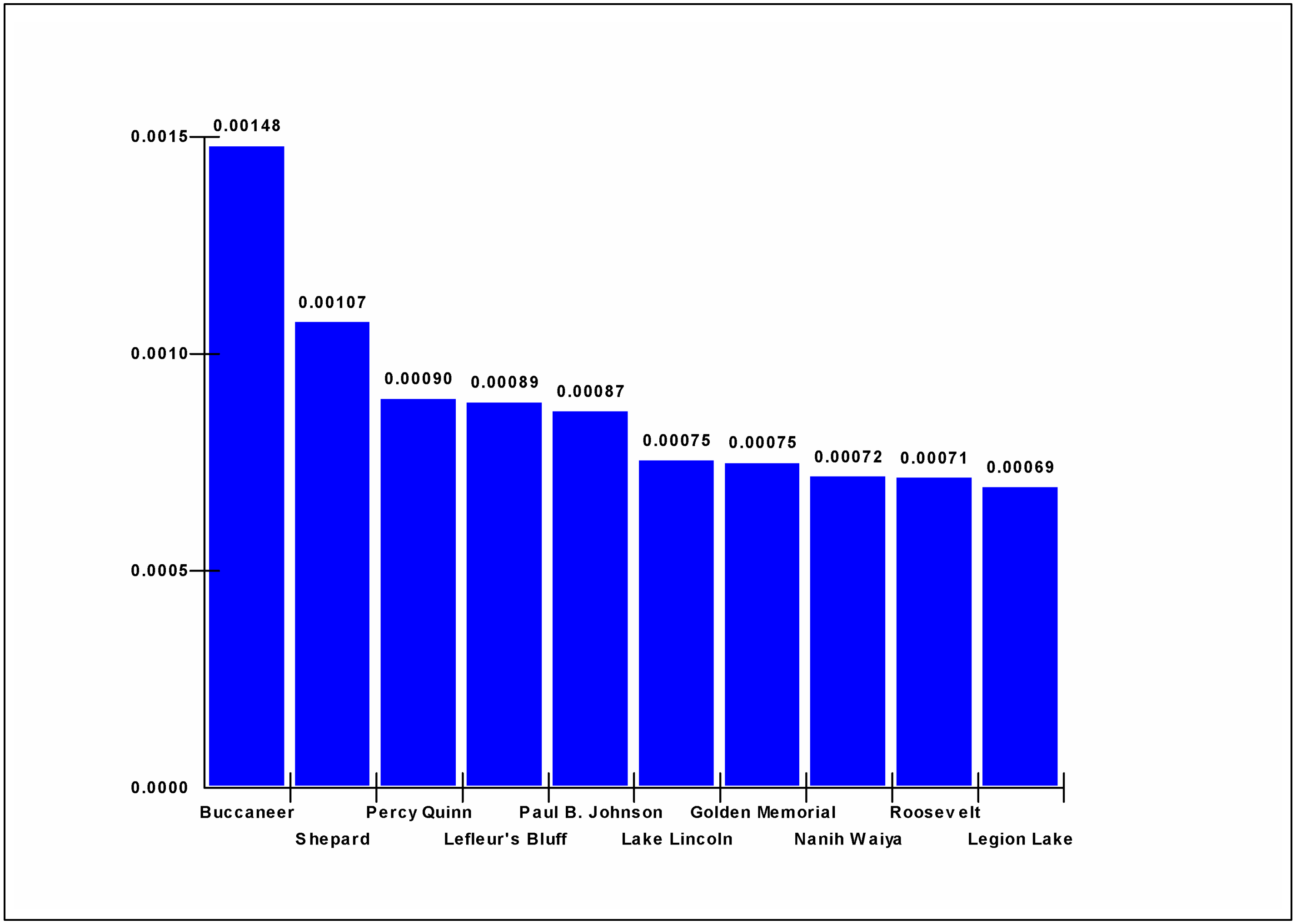

Figure 19: Top-Ten Highest Risked State Parks as Predicted by the Final Logistic Regression Model

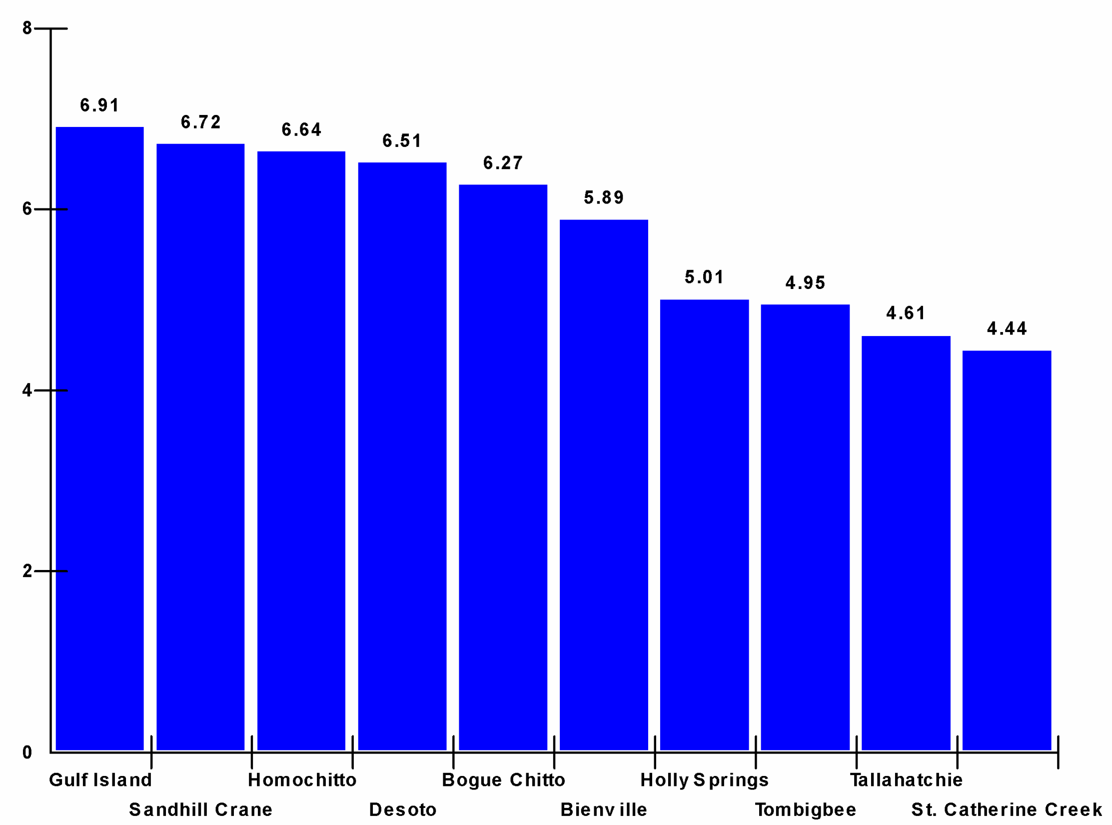

Figure 20: Top-Ten Highest Risked Natural Resource Areas as Predicted by the Final 2002 Summer Additive Model

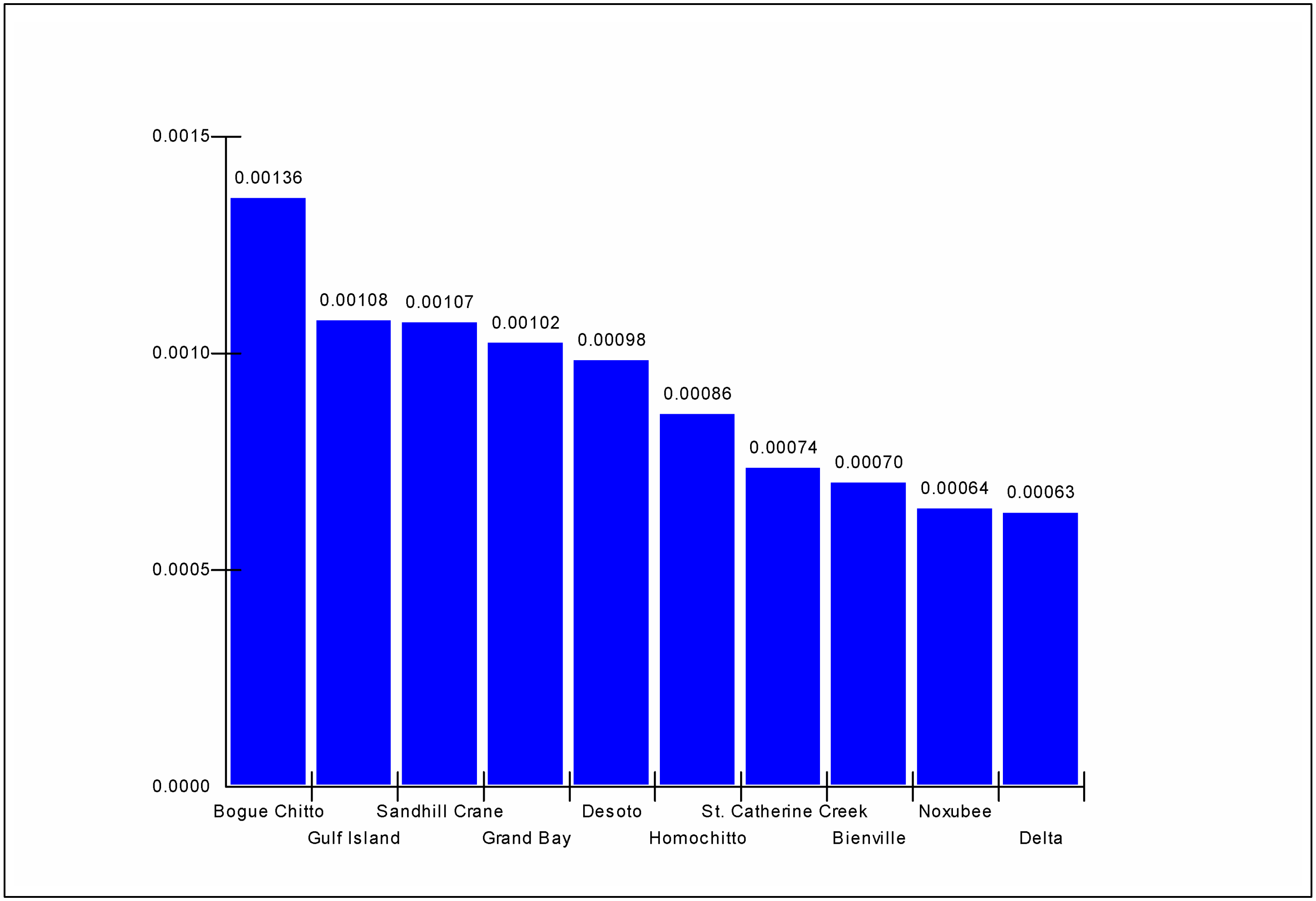

Figure 21: Top-Ten Highest Risked Natural Resource Areas as Predicted by the Final Logistic Regression Model

CHAPTER V

SUMMARY AND CONCLUSIONS

The purpose of this study was to develop a Mississippi state-wide raster model that predicts mosquito habitat suitability and/or potential risk of West Nile virus by testing the usefulness of environmental variables in a predictive modeling scenario. Two linear algebraic models were constructed, one for summer and one for fall, for each year beginning with 2002 and ending with 2003. An alternative statistically-based modeling approach using logistic regression was compared to the algebraic approach. The results of each model run were then used to calculate “risk” to the general public and specifically to Natural Resource Managers and users of recreational facilities. There were three major parts to this study: data preparation/variable manipulation, statistical tests, and model construction, each of which will be summarized below.

The majority of the effort in this first portion involved several steps to get the original occurrence data corrected and in a form that could be used in analysis. Once this was completed, the other variables were prepared for analysis. Each variable was converted to raster and conditioned in preparation for model generation. Before the models were created, statistical tests were performed which aided in variable ranking and weighting.

The second part of this study involved statistically testing each of the variables. T-tests were performed on each variable in order to determine if there were significant differences between the means of occurrence versus the means of non-occurrence. Linear regressions were developed for the variables that displayed significant differences to determine the strength and direction of relationships between the significant variable and the rate of West Nile virus occurrence. The results of the variable significance tests guided the variable weighting process for the algebraic modeling approach.

The last major portion of this study involved the creation of weighted linear additive models and a logistic regression model. For the additive models, the variables were ranked in terms of their t-test based significance and weights were assigned according to variable rankings determined on the basis of t-test probability levels. Four linear additive models were created: 2002 Summer, 2002 Fall, 2003 Summer, and 2003 Fall. Lastly, a logistic regression model was constructed. For this model, the probability of occurrence for each zip code was calculated and linearized by taking the natural log of the probability of West Nile virus occurrence in each zip code. Each zip code in the state was assigned a probability for occurrence of West Nile virus and the resulting probabilities were brought into the geographic information system and interpolated across the state which resulted in the final West Nile virus risk model.

Hard work and statistically-backed variables have resulted in a model that predicts mosquito habitat suitability. Models that predict mosquito habitat suitability are a surrogate for West Nile virus risk. Results of this study indicate that risk modeling for West Nile virus infections is feasible and inclusion of climatic variables results in a dynamic product with many unique applications. Monitoring weather conditions for dynamic stratification of the landscape offers unique mosquito control options.

Landscape stratification can also help optimize locations for mosquito pool sampling for West Nile virus. Natural resource managers and the general public can better prepare for their outdoor activities by knowing what the relative risk is for a given park, wildlife refuge, campground, or forest.

There are several strengths of this study. First, this study included two different modeling techniques that resulted in similar risk predictions. Second, correlations between landscape variables and West Nile virus risk were successfully determined. Third, the ease of modeling effort for the additive approach was demonstrated. Finally, additive modeling gives a landscape-based risk assessment at every cell location.

One weakness of this study that should be addressed concerns the original case occurrence data provided by the Mississippi Department of Health (MDOH). First, the data on West Nile virus infections are case occurrences by zip code. This presented a spatial problem that could have been avoided by using address-specific occurrence data; however, due to recent legislation and patient confidentiality issues, these data were unavailable. Secondly, the data that were available had inconsistencies between the number representing the zip code and the city associated with that zip code. Perhaps other methodologies could be developed for the correction of these data.

Several conclusions were reached from the completion of this project. They are as follows:

1. Birds are a poor indicator species for predicting West Nile virus risk.
2. Road density was the most important variable in predicting West Nile virus risk as determined by t-test and logistic regression results.

3. The general trend for risk decreases from southeast to northwest across the state.
4. Precipitation minus evaporation (P-E) is significantly different for areas of West Nile virus occurrence compared with areas of non-occurrence.
5. Reporting West Nile virus occurrences by zip code presents a spatial problem that should be corrected before this methodology can be applied to smaller scale studies.
6. Address specific occurrences would result in a more accurate model.

LITERATURE CITED

Ames Research Center, 2003, NASA Students Make West Nile Virus Risk Map for Monterey County, Available: www.spaceref.ca/news/viewpr.html?pid=12450.

Beck, L., Rodriquez, M., Dister, S., Rodriguez, A., Rejmankova, E., Ulloa, A., Meza, R., Roberts, D., Paris, J., Spanner, M., Washino, R., Hacker, C., and Legters, L., 1994, Remote Sensing as a Landscape Epidemiologic Tool to Identify Villages at High Risk for Malaria Transmission: American Journal of Tropical Medicine and Hygiene. v. 51, p. 271-280.

Bell, Christopher, 2004. Synthesis of a Serially Complete and Homogeneous Evaporation Data Set for the Southeastern Region of the United States, Mississippi State University Master's Thesis.

Centers for Disease Control and Prevention, 2003, Epidemic/Epizootic West Nile Virus in the United States: Guidelines for Surveillance, Prevention, and Control, Available: http://www.cdc.gov/ncidod/dvbid/westnile/resources/wnv-guidelines-aug-2003.pdf.

Centers for Disease Control and Prevention, 2005, 2004 West Nile Virus Activity in the United States (Reported as of January 11, 2005), Available: http://www.cdc.gov/ncidod/dvbid/westnile/surv&controlCaseCount04_detailed.htm.

Chowers, M., Lang, R., Nassar, R., Ben-David, D., Giladi, M., Rubinshtein, E., Itzhaki, A., Mishal, J., Siegman-Igra, Y., Kitzed, R., Pick, N., Landau, Z., Wolf, D., Bin, H., Mendelson, E., Pitlik, S., and Weinberger, M., 2001, Clinical Characteristics of the West Nile Fever Outbreak, Israel, 2000: Emerging Infectious Diseases. v. 7, p. 675-678.

Day, J., and Curtis, A., 1989, Influence of Rainfall on Culex nigripalpus (Ditera: Culicidae) Blood-Feeding Behavior in Indian River County, Florida: Annals of the Entomological Society of America, v. 82, p. 32-37.

Dye, C., 2000, Temperatures without Fevers: Science, v. 289, p. 1697-1698.

Edman, J., and Taylor, D., 1968, Culex nigripalpus: Seasonal Shift in the Bird-Mammal Feeding Ratio in a Mosquito Vector of Human Encephalitis: Science, v. 161, p. 67-68.

Environmental Risk Analysis Program, 2002, What's Going on with the West Nile Virus: Historical Summary by State and Country, Available: http://environmentalrisk.cornell.edu/WNV/Update/Update(A-E).php#

Epstein, P., 2000, Is Global Warming Harmful to Health?: Scientific American, v.282, p. 50-57.

Epstein, P., Diaz, H., Elias, S., Grabherr, G., Graham, N., Martens, W., Thompson, E., and Susskind, J., 1998, Biological and Physical Signs of Climate Change: Focus on Mosquito-borne Diseases: Bulletin of the Meteorological Society, v. 79, p. 409-417.

Epstein, P., Defilippo, C., 2001, West Nile and Drought: Global Change and Human Health, v. 2, p. 2-4.

ESRI, 2002, ArcGIS, Version 8.3. Help Index, keyword, Inverse Distance Weighting.

ESRI, 2002, ArcGIS, Version 8.3. Help Index, keyword, Kernel.

ESRI, 2002, ArcGIS, Version 8.3. Help Index, keyword, Quantile.

ESRI, 2002, ArcGIS, Version 8.3. Help Index, keyword, Spline.

ESRI, 2005, ESRI Support Center: GIS Dictionary, Available: http://support.esri.com/index.cfm?fa=knowledgebase.gisDictionary.search&search=true&searchTerm=GIS.

Gea-Banaclocche, J., Johnson, R., Bagic, A., Butman, J., Murray, P., Agrawal, A., 2004, West Nile Virus: Pathogenesis and Therapeutic Options: Annals of Internal Medicine, v. 140, p. 545-553.

Githeko, A., Lindsay, S., Confalonieri, U., and Patz, J., 2000, Climate Change and Vector-borne Diseases: a Regional Analysis: Bulletin of the World Health Organization, v. 78, p. 1136-1145.

Glass, G., Schwartz, B., Morgan, J., Johnson, D., Noy, Peter, and Israel, E., 1995, Environmental Risk Factors for Lyme Disease Identified with Geographic Information Systems: American Journal of Public Health, v. 85, p. 944-948.

Goddard, J. 2002, Setting up a mosquito control program: Bureau of Environmental Health, Mississippi State Department of Health. Updated July 2002. 52 pp.

Goddard Space Flight Center, 2002, Top Story: Pennsylvania's West Nile Virus Surveillance System Gets an Assist from NASA Data, Available: www.gsfc.nasa.gov/topstory/2002/20020830healthalliance.html.

Gubler, D., 1989, Aedes Aegypti and Aedes Aegypti-borne Disease Control in the 1990s: Top Down or Bottom Up: American Journal of Tropical Medicine and Hygiene, v. 40, p. 571-578.

Guharoy, R., Gilroy, S., Noviasky, J., and Ference, J., 2004, West Nile virus infection: American Journal of Health-System Pharmacy, v. 61, p. 1235-1241.

Hawley, W., 1991, Adaptable Immigrant: Natural History, v. 100, p. 55-59.

Karl, T., Knight, R., and Plummer, N., 1995, Trends in High-Frequency Climate Variability in the Twentieth Century: Nature, v. 377, p. 217-220.

Kulldorff, M. and Hjalmars, U., 1999, The Knox Method and Other Tests for Space-Time Interaction: Biometrics, v. 55, p. 544-552.

Lillesand, T., Kiefer, R., and Chipman, J., Remote Sensing and Image Interpretation, Fifth Edition. John Wiley & Sons, Inc. Hoboken, New Jersey, 2004.

LinksPoint, 2003, New Weapon in the War on West Nile Virus: Chicago Department of Public Health to Use Geographic Risk Prediction Analysis Solution from LinksPoint, Available: www.linkspoint.com/Chicago_WNV_pr.asp.

Marfin, A., Petersen, L., Millicent, E., Miller, J., Hadler, J., Farello, C., Werner, B., Campbell, G., Layton, M., Smith, P., Bresnitz, E., Cartter, M., Scaletta, J., Obiri, G., Bunning, M., Craven, R., Roehrig, J., Julian, K., Hinten, S., Gubler, D., and the ArboNet Cooperative Surveillance Group, 2001, Widespread West Nile Virus Activity, Eastern United States, 2000: Emerging Infectious Diseases, v. 7, p. 730-735.

Marra, P., Griffing, S., Caffrey, C., Kilpatrick, A., McLean, R., Brand, C., Saito, E., Dupuis, A., Kramer, L., and Novak, R., 2004, West Nile Virus and Wildlife: Bioscience. v. 54, p. 393-402.

Martens, W., Jetten, T., and Focks, D., 1997, Sensitivity of Malaria, Schistosomiasis and Dengue to Global Warming: Climatic Change, v. 35, p. 145-156.

Monath, T. and Tsai, T., 1987, St. Louis Encepahalitis: Lessons from the Last Decade: American Journal of Tropical Medicine and Hygiene, v. 37, p. 40S-59S.

Nicholson, M. and Mather, T., 1996, Methods for Evaluating Lyme Disease Risks Using Geographic Information Systems and Geospatial Analysis: Journal of Medical Entomology, v. 33, p. 711-720.

O'Sullivan, D. and Unwin, D., Geographic Information Analysis. John Wiley & Sons, Inc. Hoboken, New Jersey, 2003.

Patz, J., Martens, W., Focks, D., and Jetten, T., 1998, Dengue Fever Epidemic Potential as Projected by General Circulation Models of Global Climate Change: Environmental Health Perspectives, v. 106, p. 147-153.

Petersen, L. and Marfin, A., 2002, West Nile Virus: A Primer for the Clinician: Annals of Internal Medicine, v. 137, p. 173-179.

Petersen, L. and Roehrig, J., 2001, West Nile Virus: A Reemerging Global Pathogen: Emerging Infectious Diseases, v. 7, p. 611-614.

Purvis, John C, 1993. Pan Evaporation Records for the South Carolina Area. Southeastern Regional Climate Center, Columbia, South Carolina. http://www.dnr.state.sc.us/climate/sco/pan_evap.html.

Skidmore, A., Environmental Modeling with GIS and Remote Sensing. Taylor & Francis. London and New York, 2002.

Srivastava, A., Nagpal, B., Saxena, R., and Subbarao, S., 2001, Predictive Habitat Modeling for Forest Malaria Vector Species An. dirus in India – A GIS-Based Approach: Current Science, v. 80, p. 1129-1134.

Steitz D. and Ramanujan, K., 2002, NASA Researchers Developing Tools to Help Track and Predict West Nile Virus, Available: www.findarticles.com/p/articles/mi_pasa/is_200210/ai_423714074.

Theophilides, C., Ahearn, S., Grady, S., and Merlino, M., 2003, Identifying West Nile Virus Risk Areas: The Dynamic Continuous-Area Space-Time System: American Journal of Epidemiology, v. 157, p. 843-854.

USGS, 2004, National Wildlife Health Center: NWHC West Nile Virus Project, Available: http://www.nwhc.usgs.gov/research/west_nile/west_nile.html.

Wartenberg, D., Ramsey, D., Warner, J., Ober, D., and Murray, B., 1996, FACSNET Reporting Tools, Glossary of Terms, Available: www.facsnet.org/tools/ref_tutor/epidem/gloss.php3.

APPENDIX A

ORIGINAL WEST NILE VIRUS OCCURRENCE DATA

Table A1: West Nile Virus Positive Humans 2002

Date	City	Zip
8/22/2002	Natchez	39120
8/16/2002	Natchez	39120
9/29/2002	Natchez	39120
9/10/2002	Natchez	39120
8/19/2002	Natchez	39120
8/31/2002	Kosciusko	39090
7/27/2002	Cleveland	38732
10/5/2002	Houston	38851
8/19/2002	Port Gibson	39150
9/5/2002	Port Gibson	39150
8/6/2002	Quitman	39355
8/19/2002	West Point	39113
7/28/2002	West Point	39773
7/19/2002	Clarksdale	38614
8/16/2002	Clarksdale	38614
7/20/2002	Clarksdale	38614
7/29/2002	Clarksdale	38614
8/20/2002	Lyon	39645
8/26/2002	Crystal Springs	39059
9/2/2002	Crystal Springs	39059
8/12/2002	Crystal Springs	39083
8/19/2002	Wesson	39191
8/29/2002	Hernando	38632
6/27/2002	Hattiesburg	39401
9/20/2002	Hattiesburg	39401
7/30/2002	Hattiesburg	39401
7/1/2002	Hattiesburg	39401
7/30/2002	Hattiesburg	39401
9/10/2002	Hattiesburg	39402
7/25/2002	Petal	39465
8/29/2002	Petal	39465
7/30/2002	Petal	39465
8/5/2002	Grenada	38901
8/20/2002	Grenada	38901
8/25/2002	Grenada	38901
7/15/2002	Bay St. Louis	39520
8/7/2002	Kiln	39556
8/13/2002	D'Iberville	39532
8/31/2002	Gulfport	39501
8/10/2002	Gulfport	39501
9/21/2002	Pass Christian	39571
8/8/2002	Byram	39272
6/24/2002	Clinton	39056
8/18/2002	Clinton	39056
9/1/2002	Edwards	39066
7/1/2002	Jackson	39202
7/17/2002	Jackson	39202

Table A1: (Continued)

7/31/2002	Jackson	39202
9/27/2002	Jackson	39203
8/8/2002	Jackson	39203
7/31/2002	Jackson	39203
7/25/2002	Jackson	39203
8/1/2002	Jackson	39204
7/15/2002	Jackson	39204
7/15/2002	Jackson	39204
7/12/2002	Jackson	39204
8/2/2002	Jackson	39206
7/22/2002	Jackson	39206
7/17/2002	Jackson	39206
7/30/2002	Jackson	39206
8/16/2002	Jackson	39209
8/15/2002	Jackson	39209
8/23/2002	Jackson	39209
7/25/2002	Jackson	39209
7/15/2002	Jackson	39211
8/17/2002	Jackson	39211
7/25/2002	Jackson	39211
9/11/2002	Jackson	39211
7/12/2002	Jackson	39211
8/15/2002	Jackson	39211
8/21/2002	Jackson	39212
8/20/2002	Jackson	39212
8/7/2002	Jackson	39212
8/15/2002	Jackson	39212
7/22/2002	Jackson	39212
8/6/2002	Jackson	39212
8/2/2002	Jackson	39212
7/18/2002	Jackson	39213
7/25/2002	Jackson	39213
8/16/2002	Jackson	39213
8/31/2002	Jackson	39213
8/19/2002	Jackson	39213
7/16/2002	Jackson	39213
8/20/2002	Jackson	39213
7/31/2002	Jackson	39216
8/3/2002	Jackson	39216
8/5/2002	Jackson	39216
8/17/2002	Raymond	39154
8/22/2002	Pickens	39146
8/27/2002	Belzoni	39038
7/1/2002	Belzoni	39038
7/15/2002	Gautier	39553
8/15/2002	Moss Point	39563
8/8/2002	Moss Point	39563
9/4/2002	Moss Point	39563
8/16/2002	Pascagoula	39567

Table A1: (Continued)

8/22/2002	Pascagoula	39581
7/25/2002	Laurel	39443
10/3/2002	Prentiss	39474
7/26/2002	Ellisville	39437
8/21/2002	Ellisville	39437
9/12/2002	Laurel	39443
9/10/2002	Soso	39480
9/7/2002	Dekalb	39328
8/27/2002	Lumberton	39455
8/1/2002	Purvis	39475
8/16/2002	Meridian	39307
8/21/2002	Meridian	39307
12/14/2002	Carthage	39051
9/10/2002	Tupelo	38801
10/21/2002	Greenwood	38930
8/25/2002	Brookhaven	39601
7/29/2002	Brookhaven	39601
8/1/2002	Brookhaven	39601
7/23/2002	Brookhaven	39601
9/8/2002	Columbus	39701
9/1/2002	Columbus	39702
8/7/2002	Canton	39046
7/27/2002	Canton	39046
10/4/2002	Canton	39046
8/5/2002	Madison	39110
9/28/2002	Ridgeland	39157
8/23/2002	Ridgeland	39157
8/6/2002	Ridgeland	39157
8/2/2002	Columbia	39429
8/16/2002	Columbia	39429
8/29/2002	Columbia	39429
8/15/2002	Foxworth	39483
8/5/2002	Aberdeen	39730
8/2/2002	Aberdeen	39730
7/26/2002	Aberdeen	39730
8/11/2002	Amory	38821
9/7/2002	Philadelphia	39350
9/12/2002	Philadelphia	39350
10/19/2002	Union	39365
8/12/2002	Decatur	39327
8/18/2002	MS State	39762
8/12/2002	Starkville	39759
9/17/2002	Starkville	39759
8/23/2002	Starkville	39759
8/12/2002	Batesville	38606
8/2/2002	Carriere	39426
7/7/2002	Picayune	39466
7/29/2002	Poplarville	39470
7/30/2002	Magnolia	39648

Table A1: (Continued)

8/30/2002	McComb	39648
8/12/2002	McComb	39648
8/3/2002	McComb	39648
8/3/2002	McComb	39648
7/16/2002	McComb	39648
8/4/2002	McComb	39648
8/13/2002	McComb	39648
7/22/2002	Summit	39666
8/12/2002	Pontotoc	38863
7/28/2002	Marks	38646
7/11/2002	Brandon	39042
8/4/2002	Brandon	39047
12/12/2002	Brandon	39047
8/25/2002	Florence	39073
9/28/2002	Florence	39073
10/19/2002	Florence	39073
7/14/2002	Flowood	39232
8/19/2002	Pearl	39208
8/1/2002	Pearl	39208
7/30/2002	Pearl	39208
8/6/2002	Richland	39218
7/29/2002	Whitfield	39193
7/26/2002	Forest	39074
8/6/2002	Forest	39074
7/10/2002	Forest	39074
7/24/2002	Forest	39074
8/4/2002	Morton	39117
8/2/2002	Braxton	39044
7/30/2002	Mendenhall	39114
7/30/2002	Mendenhall	39114
9/2/2002	Mendenhall	39114
8/6/2002	Wiggins	39577
7/25/2002	Inverness	38753
9/12/2002	Moorhead	38761
7/27/2002	Charleston	38921
9/18/2002	Charleston	38921
9/5/2002	Sumner	38957
8/26/2002	Coldwater	38618
8/11/2002	Senatobia	38668
8/29/2002	Vicksburg	39180
9/22/2002	Vicksburg	39180
9/26/2002	Greenville	38701
8/31/2002	Greenville	38701
7/23/2002	Leland	38756
9/15/2002	Leland	38756
8/14/2002	Benton	39039
7/14/2002	Holly Bluff	39088
9/11/2002	Yazoo City	39194

Table A2: West Nile Virus Positive Humans 2003

Date	City	Zip
9/22/2003	Clarksdale	38614
9/7/2003	Clarksdale	38614
7/25/2003	Greenville	38701
8/10/2003	Greenville	38701
8/21/2003	Greenville	38701
8/15/2003	Greenville	38701
9/7/2003	Greenville	38701
9/3/2003	Greenville	38701
10/9/2003	Cleveland	38732
8/7/2003	Greenwood	38930
6/26/2003	Greenwood	38930
8/24/2003	Canton	39046
7/16/2003	Madison	39110
9/7/2003	Morton	39117
7/15/2003	Raymond	39154
9/7/2003	Ridgeland	39157
8/5/2003	Ridgeland	39157
8/16/2003	Vicksburg	39180
8/30/2003	Vicksburg	39180
8/7/2003	Jackson	39202
9/15/2003	Jackson	39203
8/13/2003	Jackson	39203
8/6/2003	Jackson	39204
8/13/2003	Jackson	39204
8/16/2003	Jackson	39206
7/30/2003	Jackson	39206
8/15/2003	Jackson	39209
5/14/2003	Jackson	39209
8/2/2003	Jackson	39209
8/1/2003	Jackson	39209
9/17/2003	Jackson	39211
9/24/2003	Jackson	39211
8/17/2003	Jackson	39213
8/25/2003	Jackson	39213
8/8/2003	Jackson	39213
8/6/2003	Jackson	39213
9/7/2003	Richland	39218
8/14/2003	Jackson	39236
8/21/2003	Jackson	39236
7/15/2003	Meridian	39301
8/25/2003	Hattiesburg	39401
9/7/2003	Hattiesburg	39401
6/2/2003	Hattiesburg	39401
1/17/2003	Hattiesburg	39401
7/25/2003	Hattiesburg	39401
8/24/2003	Hattiesburg	39402
10/11/2003	Brooklyn	39425

Table A2: (Continued)

8/23/2003	Carriere	39426
9/27/2003	Columbia	39429
9/8/2003	Columbia	39429
8/14/2003	Ellisville	39437
8/22/2003	Ellisville	39437
8/8/2003	Laurel	39440
8/18/2003	Laurel	39440
7/10/2003	Laurel	39440
8/18/2003	Laurel	39440
9/9/2003	Lucedale	39452
10/24/2003	Lucedale	39452
9/18/2003	Lumberton	39455
9/7/2003	Petal	39465
10/13/2003	Picayune	39466
7/30/2003	Picayune	39466
11/4/2003	Picayune	39466
9/12/2003	Poplarville	39470
10/20/2003	Richton	39476
9/15/2003	Richton	39476
10/2/2003	Soso	39480
10/5/2003	Gulfport	39501
7/21/2003	Gulfport	39501
8/19/2003	Gulfport	39501
7/20/2003	Gulfport	39501
8/24/2003	Gulfport	39501
7/26/2003	Gulfport	39501
8/13/2003	Gulfport	39503
9/1/2003	Gulfport	39507
10/8/2003	Gulfport	39507
8/1/2003	Biloxi	39531
10/10/2003	Biloxi	39532
8/1/2003	Biloxi	39532
8/15/2003	Gautier	39553
8/12/2003	Kiln	39556
10/22/2003	Long Beach	39560
9/1/2003	Long Beach	39560
10/12/2003	Moss Point	39563
7/15/2003	Ocean Springs	39565
7/16/2003	Pass Christian	39571
8/2/2003	Pass Christian	39571
9/27/2003	Pass Christian	39571
10/23/2003	Waveland	39576
9/11/2003	Wiggins	39577
7/27/2003	McComb	39648
10/11/2003	Tylertown	39667

APPENDIX B

PREDICTED RISK

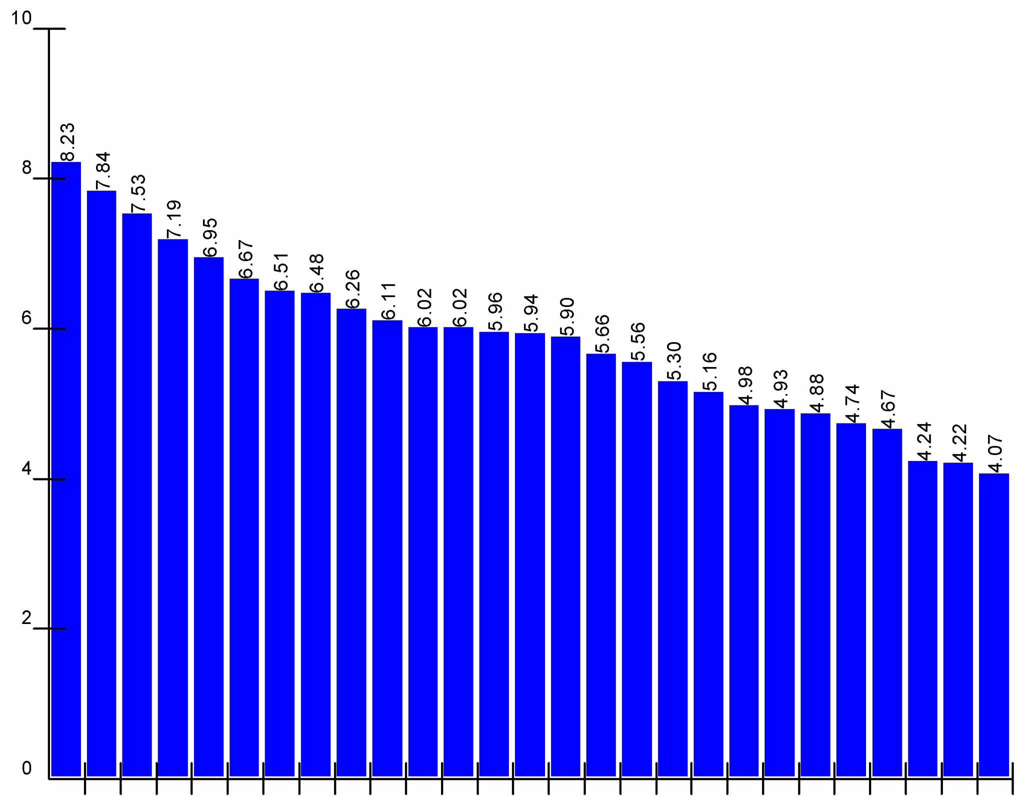

Figure B1: Risk for All State Parks as Predicted by the Final 2002 Summer Additive Model

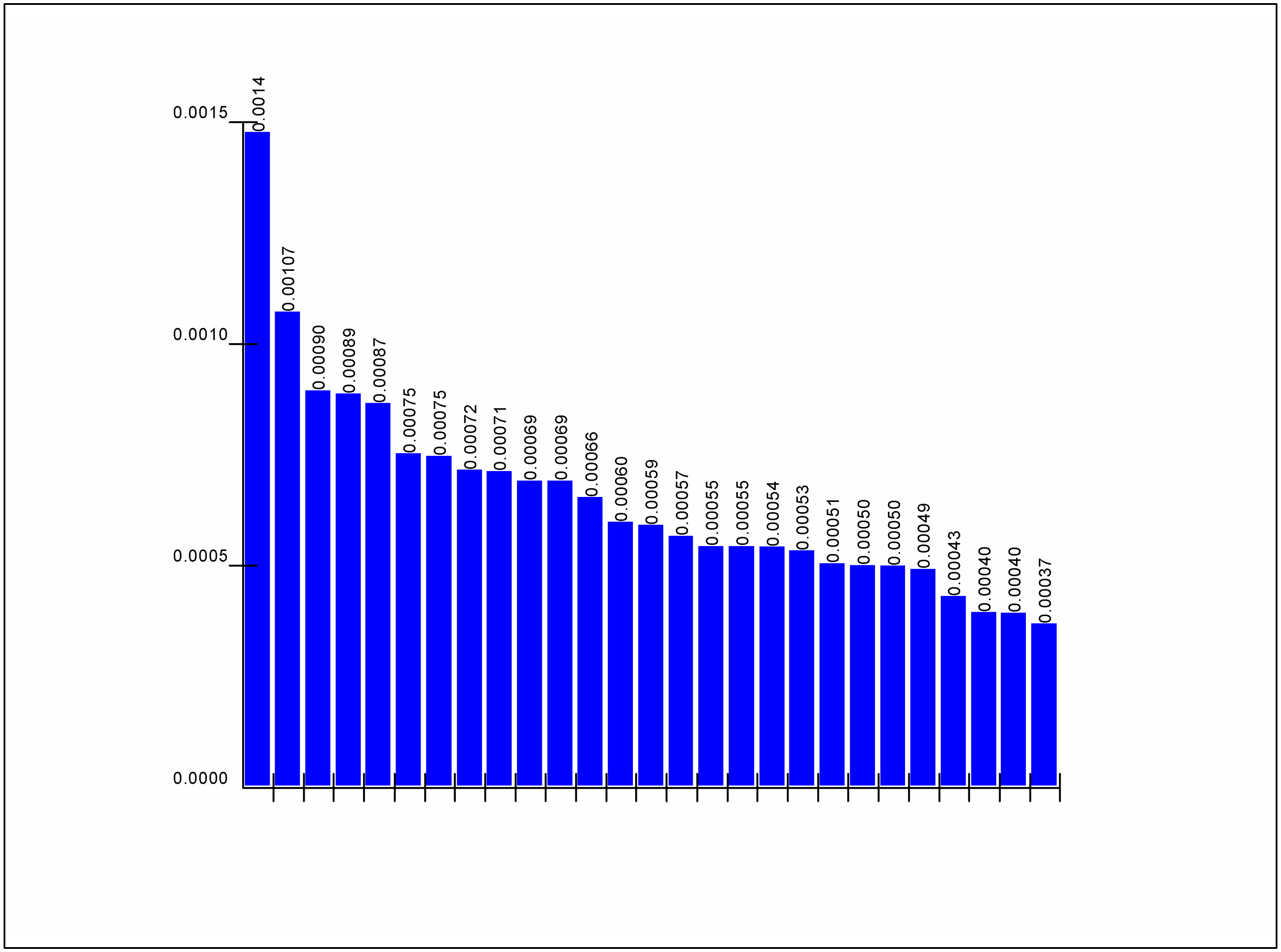

Figure B2: Risk for All State Parks as Predicted by the Final Logistic Regression Model

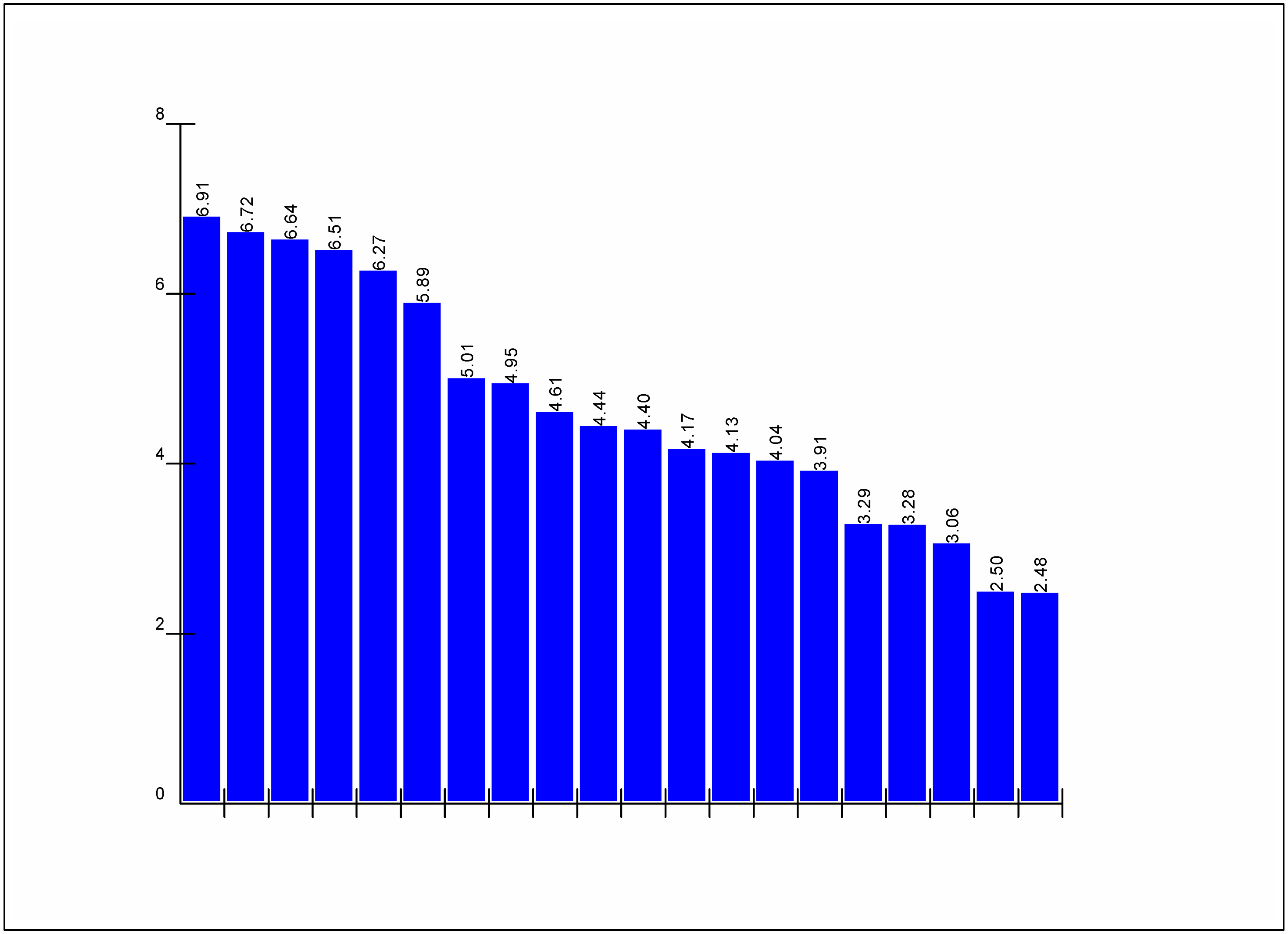

Figure B3: Risk for All Natural Resource Areas as Predicted by the Final 2002 Summer Additive Model

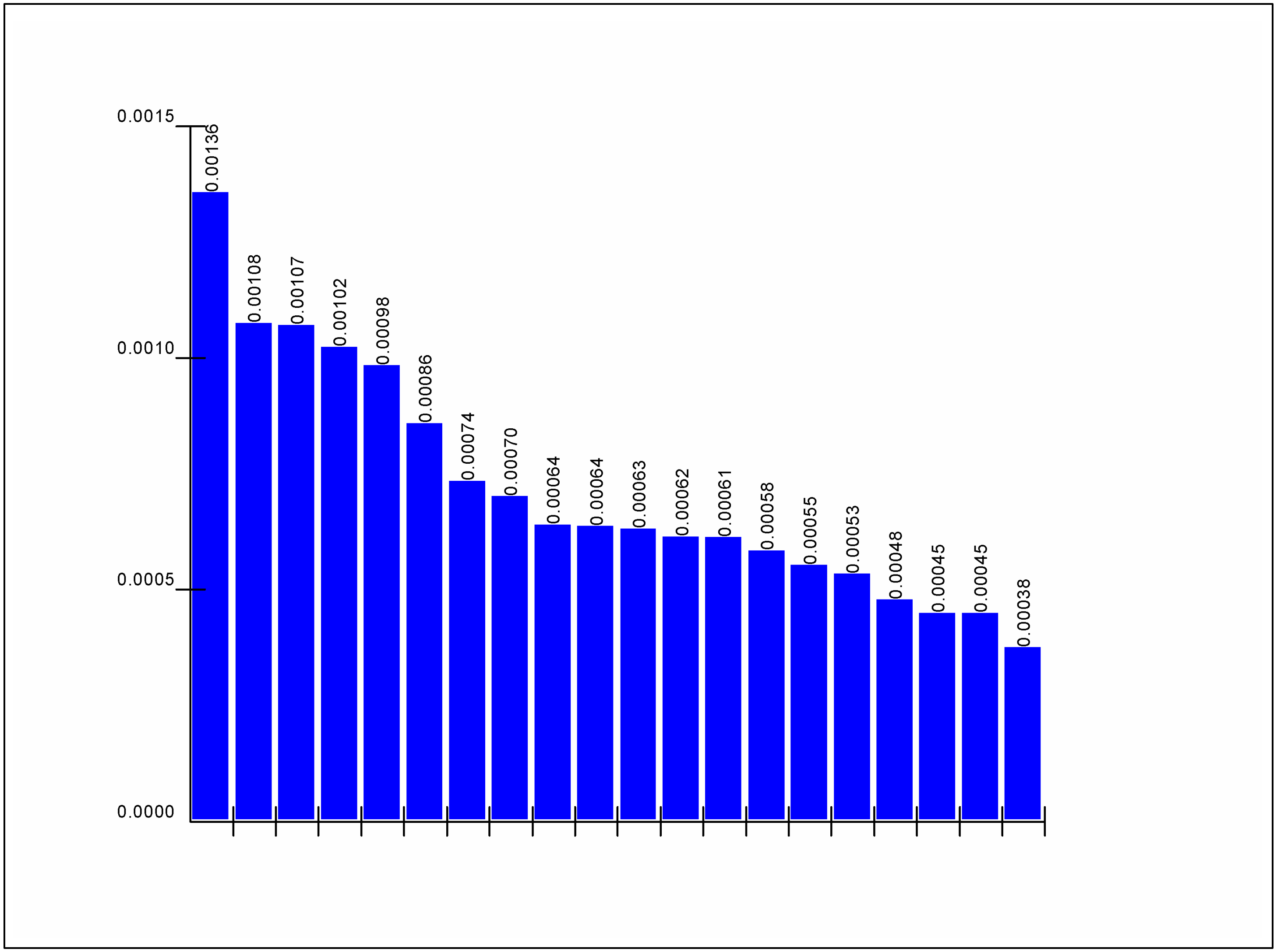

Figure B4: Risk for All Natural Resource Areas as Predicted by the Final Logistic Regression Model