SPATIAL AND TEMPORAL ANALYSIS OF HEMORRHAGIC DISEASE

IN WHITE-TAILED DEER IN SOUTHEAST USA

by

BO XU

(Under the Direction of Marguerite Madden)

ABSTRACT

Hemorrhagic disease (HD) is a common disease transmitted by *Culicoides* midges in white-tailed deer that has caused significant mortality, mobidity and economic impact on recreational hunting throughout the United States. This dissertation provides a statistical analysis of the county-based spatial-temporal prediction and cluster detection of HD in white-tailed deer in southeast USA.

A spatial-temporal prediction model was constructed to predict HD occurrence from 1982 to 2000. Eleven principal factors which were reduced from 42 climatic and environmental factors derived from ground-based weather stations and remotely-sensed data were used as predictor variables and HD presence/absence data for each county in the study area as the dependent variable. A generalized linear mixed logistic model was used to consider the within-subject effect of the longitudinal data. A spatial dependency term was added to the model accounting for the influence of HD occurrence of adjacent counties on a particular county. The results show that wind speed, rainfall, land surface temperature and normalized difference vegetation index (NDVI) are significant factors in predicting HD occurrence. The total prediction accuracy is 65 percent when all four factors are considered for a five state area. The

prediction accuracy for individual years ranges from 27 percent to 96 percent. Remotely-sensed data prove to be informative and results in a higher prediction power than some climatic data.

Kulldorff's space-time scan statistic was applied to detect the spatial and space-time clusters in HD from 1980 to 2003. The results indicate that western and southern portions of Alabama, south of Alabama, central South Carolina, and the boundary between South Carolina and North Carolina are areas where high rate clusters of HD outbreaks occur. A maximum spatial window of 10 percent of the total population and a maximum temporal window of 25 percent of the study period are believed to be appropriate windows that include most of the clusters without leaving out subclusters. NDVI, wind speed and spatial dependency were found to be related to the HD clustering.

Future study with the integration of statistical, biological, geographical information system, and remote sensing information is expected to result in a more thorough understanding of this wide spread and economically influential wildlife disease.

INDEX WORDS: Hemorrhagic disease, Spatial-temporal, Prediction model, Clustering analysis

SPATIAL AND TEMPORAL ANALYSIS OF HEMORRHAGIC DISEASE IN WHITE-TAILED DEER IN SOUTHEAST USA

by

BO XU

B.S., Beijing Normal University, P.R.China, 1998

M.E.C., East China Normal University, P.R.China, 2001

M.S., The University of Georgia, 2007

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

SPATIAL AND TEMPORAL ANALYSIS OF HEMORRHAGIC DISEASE

IN WHITE-TAILED DEER IN SOUTHEAST USA

by

BO XU

Major Professor: Marguerite Madden

Committee: David E. Stallknecht
Kathleen C. Parker
Thomas W. Hodler

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2008

DEDICATION

This thesis is dedicated to my dear family: my parents, my brother, my sister-in-law, and my nephew for their deep love and unconditional support.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank all of those people who have been such an integral part of my life and helped me in this study. First and foremost, my deepest gratitude goes to my major professor, Dr. Marguerite Madden, for her excellent guidance. I cannot adequately express my appreciation for her dedication and warm support. It is her endless input and academic excellence that have seen this project to the end.

I would thank Dr. Stallknecht and the Southeastern Cooperative Wildlife Disease Study (SCWDS) at the University of Georgia for providing me the data used in the research, which made the project possible. His invaluable advice on the biological and epidemiological background helped me throughout my research and writing of this dissertation.

I am also grateful to Dr. Hodler and Dr. Parker for their insightful advice and suggestions, not only directly on this project but also in other academic areas which indirectly contributed to the completion of the thesis.

I want to thank my friends, Fuyuan Liang, Liang Zou, Bei Tu, Yanfen Le, and Nan Ma. Without their help and friendship, I would not achieve my goal smoothly.

Finally, I deeply thank my family: my parents, my younger brother, my sister-in-law, and my nephew. Although they are not with me during these years, I know they are always there for me, love me, support me and miss me. It is their unconditional love and endless patience that gives me the motivation to finish my dissertation.

.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 Background

In epidemiology, understanding the distribution of disease is paramount. It allows researchers to investigate hot spots of the disease outbreaks, identify possible causes and predict future trends of the disease. A subdiscpline in epidemiology, spatial epidemiology, has emerged to geographically describe and analyze indexed health data with respect to demographic, environmental, behavioral, socioeconomic, genetic, and infectious risk factors (Elliott and Wartenberg, 2004). Disease mapping is the prototype of spatial epidemiology, which dates back to the 1800s when maps of disease rates were drawn to visually demonstrate the distribution and possible causes of outbreaks of some infectious diseases such as yellow fever and cholera. Since then, spatial epidemiology has grown in sophistication, complexity, and utility (Walter, 2000).

Recently, with the advance of computer science, remote sensing technology, and geographic information science (GISci), a substantial amount of georeferenced data can be easily accessed and manipulated, which has greatly accelerated the development of spatial epidemiology. A number of spatial statistical models are proposed for the analysis of the spatial distribution of diseases such as heterogeneous Poisson process model (HEPP), Hybrid models, Bayesian models, hierarchical models, etc. (Lawson, 2001). A newer trend is to incorporate the time dimension into the spatial framework to explore the spatial-temporal distribution of disease.

There are numerous examples of the use of spatial modeling and disease mapping in human disease studies, however; in veterinary epidemiology, there has been less use of these techniques. Although some work has been done, since the 1980s, in a spatial frame, most of this work has centered on the distribution of disease vectors, and environmental factors, relate to

parasitic diseases (Durr, 2004). There have been few studies on the distribution of animal diseases, let alone the relationship between the distribution of animal disease and environmental factors.

Hemorrhagic disease (HD) is a common disease in white-tailed deer (*Odocoileus virginianus*). The first outbreak of HD in the USA dates back to the 1890s (Trainer, 1964; Hoff and Trainer, 1981). Since then, outbreaks of HD have been documented often, and can result in significant deaths among white-tailed deer. Veterinary researchers have done much work to identify vectors that transmit the disease. Previous research has shown that climate and environmental factors such as altitude, humidity, rainfall, and temperature affect the distribution of vectors and HD outbreaks (Nevill, 1971; Walker and Davies, 1971; Erasmus, 1975). Furthermore, there is seasonal and annual variation in HD occurrence (Couvillion et al., 1981). However, little research has been done to quantitatively investigate the distribution of HD in white-tailed deer in terms of spatial and temporal patterns and its relationship with various potential causal factors such as climate or environment. In the mean time, remotely-sensed data can be used for predicting vector-born disease in that remotely-sensed data can surrogate climatic variables which affect the survival and reproduction of vectors.

1.2 **Research objectives and study area**

This research is designed to investigate the spatial-temporal clustering or hot spots of the HD outbreaks in white-tailed deer ern USA, and construct a prediction model which relates the distribution of HD to the climatic factors and environmental factors derived from remotely-sensed data. The underlying hypothesis is a combination of climatic and environmental

factors such as temperature, precipitation, elevation, land use/land cover, etc. influences the spatial and temporal patterns of HD outbreaks in white-tailed deer through possible biological controls of the disease vector population dynamics, and/or herd health and immunity. This research will focus on the geographic and statistical correlation of the environment and HD distributions that may lead to a better understanding of the biological processes during the HD outbreaks. The hypotheses and expected results are:

1) There are spatial and temporal clusters in the distribution of HD in white-tailed deer.
2) There are spatial and temporal relationships between HD occurrence and climatic factors and environmental factors derived from remotely-sensed data.
3) A statistical prediction model can accurately predict the HD occurrence within the limits of available data and accuracy.

The objectives of this research are as follows:

1) Develop a 20-year spatial-temporal database of HD outbreaks/causal factors and a statistical model predicting HD distributions based on critical climate and environmental variables.
2) Use statistical methods to describe significant spatial and temporal clusters of HD outbreaks and areas of no outbreaks in white-tailed deer over a 20-year time period.
3) Assess possible causal factors affecting spatial and temporal patterns of HD distribution.

The study area in this research is the southeastern region of the USA including five states: Alabama, Georgia, South Carolina, North Carolina, and Tennessee where the HD in white-tailed deer is traditionally reported and considered endemic (Hoff and Trainer, 1981) (Figure 1.1). The

white-tailed deer can be found in wooded areas on the fringes of urban areas in southern Canada and most of the United States, except for the Southwest of USA, Alaska and Hawaii. Before the settlement of Europeans, there were between 23 and 40 million white-tailed deer in North America. After that, the population greatly decreased in USA due to habitat loss and unrestricted hunting. But by the 1950s, North America experienced an increase in white-tailed deer. At the end of the 20th century, the estimated population in USA alone was 14 to 20 million. (http://www.desertusa.com/mag99/june/papr/wtdeer.html). White-tailed deer have a number of negative economic impacts— they damage crops, vegetable gardens, fruit trees and personal property where their ranges overlap with human habitation. However, they also have invaluable positive values. For Native Americans and early European settlers, deer meat provided one of the most important sources of protein. White-tailed deer remain the most popular large game animals throughout most of the USA. Sport hunters harvest about 2 million white-tailed deer annually. The economic value of deer through license fees, meat, and hunter expenditures for equipment, food, and transportation can be measured in hundreds of millions of dollars (http://www.extension.org/pages/Deer). Thus deer population control and deer disease control are equally important.

The five states in southeast USA boasts a variety of physiographic regions including plateau, ridge and valley, hilly coastal plain, lower coastal plain, and piedmont (Miller and Robinson, 1995). These physiographic regions together with the mild winter, hot summer, plentiful rainfall, and large areas of forests provides various habitats to white-tailed deer as well as the vectors that transmit HD.

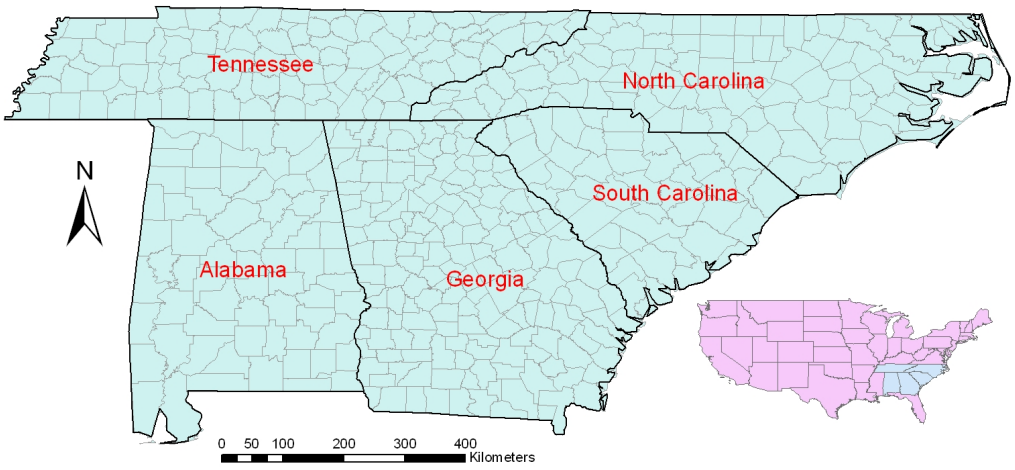


Figure 1.1: Study area: five states in southeast USA

The outbreaks of HD in white-tailed deer in southeastern states have a long history. In the 1940s, Ruff (1949) pointed out that extensive mortality of an unexplained fatal disease occurred at irregular intervals for many years in this region which was later believed to be similar to epizootic hemorrhagic disease (EHD). In 1954 and 1955, epizootics similar to HD appeared among the white-tailed deer population in southeast USA (Prestwood, et al., 1974). During the summer of 1971, the first documented HD outbreak in white-tailed deer in southeastern USA was reported, which caused significant die-offs (Thomas et al., 1974). The disease first occurred in South Carolina and then erupted almost simultaneously in Virginia, Tennessee, North Carolina, Kentucky, Georgia, and Florida. A widespread HD outbreak also occurred in 1980, whereby deer in 156 counties in 8 states had clinical evidence of exposure to HD (Couvillion et al., 1981). According to the Southeastern Cooperative Wildlife Disease Study (SCWDS) in the College of Veterinary Medicine, University of Georgia data from 1980 to 1989, contiguous instances of HD

were reported throughout the Southeast (Nettles et al., 1992). In 1998, Considerable die-offs appeared among white-tailed deer, marking a peak year of HD (Nettles and Stallknecht, 1992).

1.3 HD Data sources

The hemorrhagic disease data was collected on a county basis by SCWDS annually starting in 1980 and is the most comprehensive database for HD-related deer morbidity and mortality. For nearly 30 years, the researchers have mailed HD questionnaires to the state fish and wildlife agency in each state, and most of the state veterinary diagnostic laboratories. The surveillance also includes personnel in the U.S. Fish and Wildlife Service and the Animal and Plant Health Inspection Service, USDA. For 1980 and 1981, only the 16 southeastern states were polled. All states except Hawaii have been surveyed from 1982 to present. The researchers set four criteria for the surveillance reporting as follows:

1) Sudden, unexplained, high deer mortality reported during the late summer and early fall.
2) Necropsy diagnosis of HD as rendered by a trained wildlife biologist, a diagnostician at a State Diagnostic Laboratory or Veterinary College, or by SCWDS personnel.
3) Isolation of EHD or Bluetongue virus from a deer.
4) Observation of hunter-killed deer that showed sloughing hooves, ulcers in the mouth, or scars on the rumen lining as indirect evident of chronic HD occurrence.

Among them, criteria 1, 2, and 3 collectively reflect deer mortality, and criterion 4 indicates deer morbidity.

1.4 Dissertation structure

The dissertation structure is organized into five chapters. Chapter 1 is a brief introduction of the background, objectives, study area, mechanism, data sources and hypotheses. Chapter 2 gives a detailed literature review of the topics covered in this dissertation, including history, transmission, spatial and temporal distribution of HD, veterinary spatial epidemiology, geographic information systems and remote sensing in veterinary epidemiology, space-time clustering, and spatial-temporal modeling in veterinary epidemiology. The following two chapters are separate papers to be submitted to journals. In Chapter 3, a statistical spatial-temporal model is developed to predict the HD occurrence based on its relationship with the climatic factors measured by weather stations and environmental factors derived from remotely-sensed data. Chapter 4 describes the statistical detection of the spatial and space-time clusters of HD across the study area during a 20-year study period. Chapter 5 provides conclusions, limitations, and contributions of this dissertation.

CHAPTER 2

LITERATURE REVIEW

2.1 HD in white-tailed deer

2.1.1 History and clinical signs

Hemorrhagic disease is caused by both the bluetongue (BLU) and epizootic hemorrhagic disease (EHD) viruses, which are in the genus Orbivirus in the family *Reoviridae* (Hoff and Trainer, 1978). Hemorrhagic disease is transmitted by *Culicoides* midges (Nettles and Stallknecht, 1992) having life cycles composed of egg, four larval instars, pupa and adult stages. The immatures requires moisture and organic matter in damp or saturated soils, bogs, marshes, swamps, tree holes, and animal dung (Meiswinkel et al., 1994; Mellor, 1996). The life cycle ranges from 7 days in the tropics to 7 months in temperate areas. After *Culicoides* feed on a viremic white-tailed deer, the viruses are deposited into the posterior region of the vector's midgut. Then the virus attaches to the luminal surface of the gut cells, infects these cells and replicate in them. The salivary glands are infected when progeny virions are released through the basement membrane to the hemocoel. The viruses replicated in the salivary glands are transmitted to another deer by subsequent bite (Wittmann and Baylis, 2000). In the USA, two serotypes of EHD virus (serotypes 1 and 2) and five serotypes of BLU virus (serotypes 2, 10, 11, 13 and 17) have been isolated (Wieser-Schimpf et al., 1993; Stallknetht et al., 1996).

Because of their clinical and epidemiologic similarity, BLU and EHD field cases are generic diagnosed as HD (Hoff and Trainer, 1972; Hoff and Trainer, 1974; Nettles and Stallknecht, 1992). The HD is characterized by sudden onset, and the first sign is pyrexia which is a rise of temperature of the body. Clinical signs typically progress from hyperemia (an unusual high level of blood in some part of the body) to facial and cervical swelling, then lameness

followed by hemorrhage, sometimes sloughing of hooves, and finally ulceration (Hoff and Trainer, 1978) (Figure 2.1, 2.2 and 2.3). The infected deer often avoid and appear hypersensitive to sunlight (Hoff and Trainer, 1981). Figure 2.1, 2.2 and 2.3 shows the typical sings of HD in white-tailed deer.

2.1.2 Transmission

Culicoides were suspected to vector BLU virus in the early 1900s, and this was demonstrated in 1944 when BLU virus was isolated from wild-caught *Culicoides pallidepennis* in South Africa (DuToit, 1944; Hoff and Trainer, 1981). Subsequent research found BLU virus in *C.variipennis* in Texas (Price and Hardy, 1954), *C.milnei*, *C.tororoensis*, and *C.pallidepennis* in Kenya (Walker and Davies, 1971), *Culicoides* spp. in Australia (St.George et al, 1978), and *C.insignis*, *C.pusillus* and *C.filarifer/C.ocumarensis* in Central America and the Caribbean (Mo et al, 1994; Greiner et al., 1993). Experimental studies have shown that *C.variipennis* can serve as a biological vector of BLU virus (Foster et al., 1963; Jones and Foster, 1974; Hoff and Trainer, 1981).

The EHD virus was first isolated from *C.schultzei* in Nigeria in 1967 (Hoff and Trainer, 1981). In 1971, the virus was isolated from naturally infected *C.variipennis* during an epizootic of EHD involving white-tailed deer in Kentucky, USA (Jones et al., 1977; Hoff and Trainer, 1981). Foster et al (1977) demonstrated the biological transmission of this virus among deer by *C.variipennis* in their research.

Figure 2.1: Acutely ill fawn with characteristic signs of HD: lowered head, protruding tongue, and excessive salivation
(http://www.ag.auburn.edu/aaes/communications/highlightsonline/winter99/white-tail.html)

Figure 2.2: Deer exhibiting laid-back ears, depression, and reluctance to stand
(http://www.ag.auburn.edu/aaes/communications/highlightsonline/winter99/white-tail.html)

Figure 2.3: Deer in late-stage of HD, with frothy saliva due to congested lungs and labored breathing
(http://www.ag.auburn.edu/aaes/communications/highlightsonline/winter99/white-tail.html)

C.variipennis is considered the primary vector of BLU and EHD viruses in the USA (Osburn et al., 1981), and *C.insignis* is the other confirmed vector of BLU in the USA (Greiner et al., 1985). However, these two midges are not collected during outbreaks of HD in white-tailed deer (Smith et al., 1996). Based on the seasonality and relative abundance, *C.lahillei* and *C. stellifer* represent the two most possible vectors of EHD and BLU viruses for white-tailed deer ern USA (Mullen et al., 1985; Smith et al., 1996).

2.1.3 Spatial and temporal distribution

In the USA, outbreaks of HD among white-tailed deer have occurred since the 1890s (Trainer, 1964; Hoff and Trainer, 1981). However, significant die-offs were not documented until 1971 ern USA. Bluetongue was first reported in sheep in the USA in 1953, and then recognized in white-tailed deer in 1968. Epizootic hemorrhagic disease virus was first isolated from white-tailed deer in 1955 (Couvillion et al., 1981). Although BLU usually does not cause mortality in white-tailed deer, mortality due to EHD among white-tailed deer can be very severe. Experimental and field studies show that the mortality rate for EHD in white-tailed deer epizootics reaches 90 percent (Trainer, 1964; Hoff and Trainer, 1981).

Hemorrhagic disease has been documented throughout the southeast USA, extending to Texas in the west and New Jersey in the north (Stallknecht et al., 2002). The SCWDS data shows two contiguous geographic bands of HD occurrence in deer from 1980 to 1990 (Figure 2.4). The first one (1) is a transverse band from the southeast along the Missouri River northwestward until the Great Plains. The other band (2) is in coastal and northern California spreading to central Oregon and western Washington (Nettles et al., 1992).

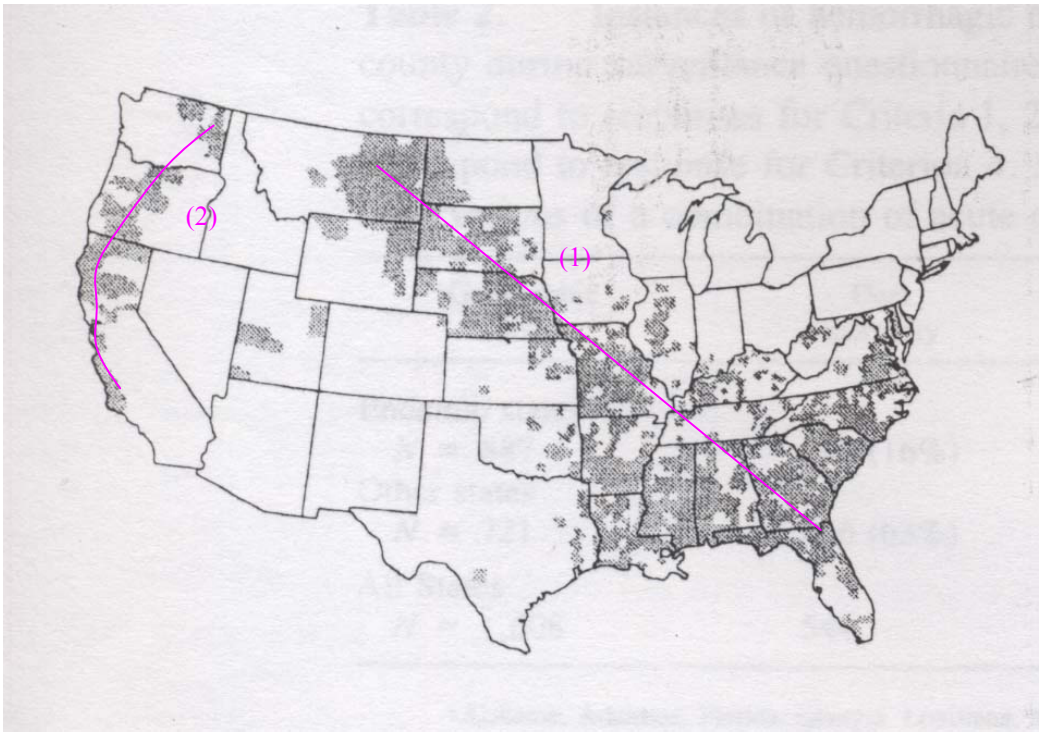


Figure 2.4: Reports of hemorrhagic disease in deer from 1980-1990 in the USA (Nettles et al., 1992, p140)

In their studies on HD among white-tailed deer from 1971 through 1980, Couvillion et al. (1981) found that 83 percent of all the HD cases examined were observed during August, September, and October. All peracute or acute cases occurred between June and November. The frequency of peracute or acute cases increased from July and to a peak in September, followed by a sharp decline in November. They concluded that there has been a biennial rise and fall as to the yearly pattern of reported HD cases, which also demonstrates the enzootic nature of this disease in Southeast USA. An eight- to ten-year cycle transmission is observed in epidemic regions. These temporal trends are probably related to the combined effects of herd immunity and climate and environment that influence vector populations (Stallknecht et al., 2002).

Erasmus (1975) pointed out that the factors that affect insect breeding and distribution such as altitude, humidity and temperature, will also influence the epizootiology of the disease. The geographic ranges of *Culicoides* are influenced by temperature, especially low temperatures which determine the distribution of the insects, whereas high temperatures adversely affect *Culicoides* adult survivorship and size (Wittmann and Baylis, 2000). Some previous studies have proved that temperature has been associated with bluetongue virus infection of ruminants. Rainfall strongly influences the increase of *Culicoides* numbers in spring and summer in South Africa +(Nevill, 1971). Walker and Davies (1971), in their survey of epidemiology of bluetongue in Kenya, postulated that there is a causal relationship between peak rainfall in April-May, peak numbers of *Culicoides* in May-June and peak bluetongue incidence in June-July. Because peak rainfall in April-May facilitates high larval survival and consequent rapid expansion of adult numbers that reaches a peak in May, June and July, thus a peak of BLU outbreaks occurs in the latter half of this peak. However, the relationship between rainfall and the abundance of *Culicoides* varies depending on the different sub-species of the vector. Some sub-species favor drier climate than humid climate. Wind speed and direction can also influence *Culicoides* distribution through their affect on the passive dispersal of the adults. For example, *Culicoides* can be carried as aerial plankton to a place up to 700 km away in winds at speeds of 10-40 km/h, at heights up to 1.5 km (Wittmann and Baylis, 2000).

2.2 Veterinary spatial epidemiology

Epidemiology is the study of disease in populations of humans or other animals. It attempts to discover the factors that are associated with or can protect humans or other animals

from disease. Recently, with the availability of geographically indexed health and population data, the increasing growth of computing and GIS, and the advances in statistical methodology (Elliott et al, 2000), a new subdiscipline in epidemiology – spatial epidemiology has emerged. The primary purpose of this subdiscipline is to describe and explain the spatial pattern of disease. It uses GIS, spatial statistical packages and remotely sensed images (Durr, 2004) to fulfill the following four tasks: disease mapping, geographical correlation studies, assessment of risk in relation to a point or line-source, and cluster detection and disease clustering (Elliott et al, 2000). There are some other terms regarding this aspect, such as environmental epidemiology, geographical epidemiology, and landscape epidemiology. Each of these focuses on a different research and methodology in spatial epidemiology. The concept of spatial epidemiology can be described as the following model (Figure 2.5).

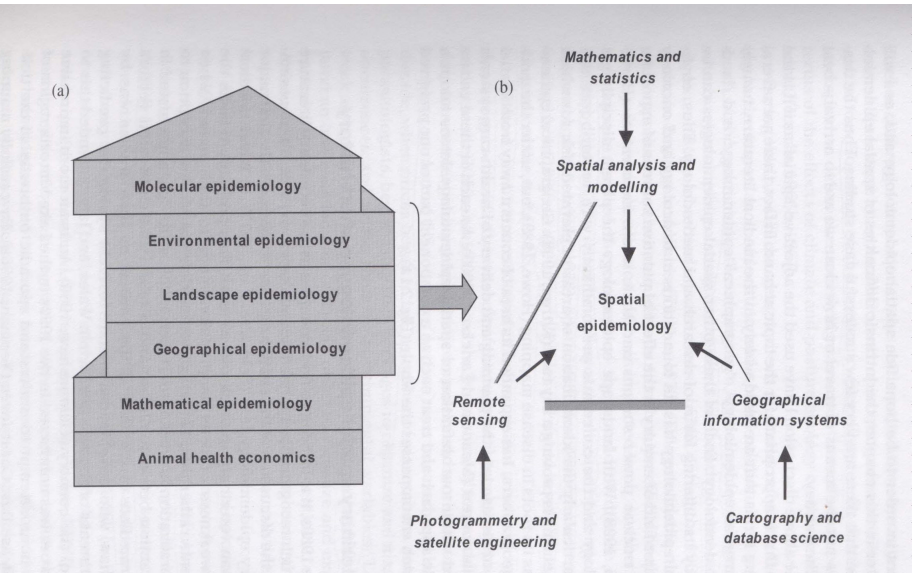


Figure 2.5**:** A conceptual model for spatial epidemiology: (a) its relationship to some other epidemiological disciplines that can be similarly defined as having a distinct viewpoint or approach; and (b) the source origins of its methodologies. Note that the list in (a) is not exhaustive (Durr and Gatrell, 2004).

Spatial epidemiology studies on human diseases and public health have been progressing rapidly due to the readily available health event data and socio-economic factors data. Numerous disease mapping and prediction methods have been advanced and applied. For examples, Xia and Carlin (1998) utilize Hierarchical Bayes model to map Ohio lung cancer mortality. Kleinschmidt et al. (2000) proposed a two-stage procedure to produce maps of predicted risk of malaria in Mali. Clements et al. (2006) develop a Bayesian geostatistical prediction model for the distribution of *Schistosoma baematobium* and *S. mansoni* in Tanzania in order to assist planning the implementation of mass distribution of the preventative medicine praziquantel. Waller et al (1997) propose a spatial-temporal model by extending existing Hierarchical Bayes methods to account for temporal effects and spatial-temporal interactions, and then illustrate the approach using the lung cancer rates data set in the state of Ohio. In addition to the above examples, a multitude of applications and methods have been conducted or proposed in human diseases in the past (see for examples, Clayton and Kaldor, 1987; Doll and Wakeford, 1997; Ghosh et al., 1998; Best and Wakefield, 1999; Pickle et al, 1999; Lawson, 2000; Sun et al, 2000; Lawson, 2001; Sheehan et al., 2001; Chaput et al., 2002; Mugglin et al., 2002; Ghebreyesus et al., 2003).

Veterinary spatial epidemiology, as the name indicates, is the spatial analysis of disease occurrence in veterinary science. Classic veterinary epidemiological analysis typically focuses on the animal dimension with experiments on the host and vectors, whereas temporal and spatial distributions are often explored with fairly basic methods. Veterinary spatial epidemiology is relatively much newer and few studies have been completed. Until the 1980s, there are few examples in the veterinary literature where much recognizable spatial analysis is evident. The

possible reason is that veterinary science focused on experiments rather than field observations during that period. One exception is the work by parasitologists whose interest is to explore the interaction between climate and disease via its effect on vectors and intermediate hosts (Durr, 2004). Until now, most studies are still emphasized in this area. Ollerenshaw (1966), for example, developed one of the earliest climate forecast systems for predicting acute outbreaks of *Fasciola hepatica* in Wales and England. Later, other researchers did even more impressive work in this scope. The real impetuses to the growth of spatial epidemiology, however, were the technical breakthroughs in computing and the availability of satellite imagery in the 1980s. Researchers, especially parasitologists, began using GIS packages and satellite images to map disease and relate them to environmental factors. Lessard et al. (1990) established landmark research by collating a large amount of data to visualize and investigate the spatial pattern of theileriosis across the whole continent of Africa.

The most popular use of satellite images in veterinary spatial epidemiology is to extract the normalized difference vegetation index (NDVI), which is closely correlated to the green vegetation biomass and thus indirectly to rainfall, and relate this index to vector habitat and vector abundance. Since the 1990s, with the emergence of user-friendly GIS and image processing packages, such as Arcview (ESRI), Mapinfo (Pitney Bowes MapInfo Corporation), Idrisi (Clark University), and ERDAS Imagine (Leica Geosystems), there is less need for the epidemiologists to get extensive GIS training and they can use the software easily to achieve certain goals. This trend has transferred spatial epidemiology into part of mainstream veterinary epidemiology and has been well illustrated by the number of papers presented in the successive

International Symposia on Veterinary Epidemiology and Economics (ISVEE) conferences during the 1990s (Durr, 2004). Pfeiffer (2000) also presented an overview of spatial analysis applications in veterinary epidemiology.

2.3 GIS and remote sensing in spatial epidemiology

A GIS is a system of hardware, software, and procedures designed to support the collection, manipulation, analysis, modeling, and display of geographically indexed data for solving complex problems (Burrough, 1986; Lo and Yeong, 2002). It is an essential and integral part of spatial epidemiology and provides a powerful tool for data visualization and exploration. For spatial epidemiological modeling, GIS has been used to provide input variables and display model output. A GIS also has the potential ability to be an integral component of a model, deriving information from other sources, feeding it into the model and storing the output.

The need to obtain information about the Earth's surface over space and time is the motivation for the use of remote sensing in epidemiological studies. Satellite images can provide regular, systematic and synoptic vegetation indices over the entire globe with image archives dating back to the early 1970s. They also can give high temporal data on a daily basis for time pattern recognition research.

Considering the climate factors, the collection and analysis of weather station data from a large number of sites involves considerable expense, time and effort. It has been demonstrated that remotely sensed images are correlated to a greater or lesser degree with certain climatic variables recorded on the Earth's surface (Hay et al., 1996; Baylis and Rawlings, 1998). Baylis et al.(1999) demonstrated that variables derived from satellite images performed better than the

climatic variables in their predictive modeling. They concluded that the reason may be that the remotely sensed data recorded more biologically relevant data than the weather stations, which also affected the distribution and abundance of vectors, in turn influencing the status of disease occurrence.

It is shown that easily available remotely sensed data set combined with GIS could be used to explain the observed patterns not only more efficiently, but also more effectively (Baylis et al., 1998; Baylis and Rawlings 1998; Baylis et al., 1999; Baylis et al., 2001). Recently there have been several studies on predicting vector distribution using remotely sensed data as surrogates for climatic data such as NDVI for vegetation, channels 4 and 5 of the Advanced Very High Resolution Radiometer (AVHRR) for temperature, cold-cloud duration (CCD) for rainfall, etc. (Linthicum et al., 1987; Rogers and Randolph, 1991; Wood et al., 1991; Rogers and Williams, 1993; Cross et al., 1996; Hay et al., 1996; Rogers et al., 1996; Estrada-Peña, 1997; Baylis and Rawlings, 1998; Baylis et al., 1999; Glass et al., 2000; Hay, 2000; Thomson and Conner, 2000; Baylis et al., 2001; Boone et al., 2000). The basis for the supposition that remotely sensed data can be used for vector-borne disease prediction is that pathogen transmission is facilitated by arthropods, whose survival and reproduction are affected by climate variables such as temperature and humidity, while the remotely sensed images are correlated with climate variables.

While there are a large number of Earth observation satellites, very few have found applications in epidemiology. The most important satellites used in epidemiology are the Landsat and the National Oceanic and Atmospheric Administration (NOAA), USA satellites (Durr and

Gatrell, 2004). Generally, remotely sensed images tend to have either high-spatial or high-temporal resolution, but not both. The Landsat satellite program has provided medium spatial resolution optical data of 80-m pixel sizes (1972-1983) and images of 30-m pixels (1982-present) This resolution may let researchers identify common land covers such as individual agricultural fields and make it ideal for comparing different types of vegetation, but it has a revisit period of 16 days, which means it gives few cloud-clear images of the Earth's surface each year (Hay et al., 1996). Advanced Very High Resolution Radiometer (AVHRR) sensors on-board the NOAA series, on the other hand, provide low-resolution 1000-m pixel images at a much greater temporal resolution (daily), which is most suitable for obtaining the cloud-free images. Other moderate resolution sensors include the USA Terra Sensors Moderate-resolution Imaging Spectroradiometer (MODIS) (up to 250-m resolution) and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (15-m resolution). Satellite programs in other countries such as the European SPOT1-5 (up to 10-m resolution), Indian IRS-1 (up to 5.8-m resolution), Russian RESURS (170-m resolution), Japanese JERS-1 (18-m resolution), and CBERS (20-m resolution) cooperated by China and Brazil all give Earth resource observations in different perspective (Lillesand and Kiefer, 1999).

Now with the advent of higher and higher spatial and temporal resolution satellite such as Ikonos (with revisit time of 11 days and spatial resolution of 1-m panchromatic and 4-m multispectral), QuickBird (with revisit time of 1 to 5 days and spatial resolution of 0.6-m panchromatic and 4-m multispectral), OrbView-3 (with revisit time of less than 3 days and spatial resolution of 1-m panchromatic and 4-m multispectral), and Worldview-1 (with revisit

time of 1.7 days and spatial resolution of 0.5-m panchromatic. There will be wider applications of remote sensing techniques in spatial epidemiology. Another option is the pointable sensors of constellations of small satellites that are increasingly being launched by multi-national coalitions. It requires a number of duplicate satellites in blocks constituting a constellation that provide adequate temporal, spatial and radiometric sampling (Huh, 2008, http://hurricane.lsu.edu/_unzipped/huh_paper1/huh_paper1.PDF)

2.4 Spatial-temporal modeling in veterinary spatial epidemiology

The increasing availability of high resolution (spatial, spectral and temporal) satellite image data provide information on Earth surface processes that can be linked to disease outbreaks via spatial-temporal models. In veterinary spatial epidemiology, the most often used models are logistic regression and discriminant analysis because the independent variables are usually binary data (used in logistic regression) or categorical data (used in discriminant analysis). During the last 10 years, a number of spatial modeling applications directly related to animal disease have been conducted. McKenzie et al. (2002) used a logistic regression model to predict the risk of *Mycobacteriumbovis* infection in wildlife using remotely sensed images, and revealed the presence of hotspots of tuberculosis infection on the basis of vegetation and slope information. Duchateau et al. (1997) produced a risk map of theileriosis outbreaks in Zimbabwe. They first applied principle components analysis (PCA) to climate factors to control for the multicollinearity between the variables, and the selected components were included in the logistic regression analysis. Baylis et al. (2001) applied discriminant models to the disease vector to detect areas of bluetongue infection risk in the Mediterranean. The model predicted three

abundance categories of *Culicoides imicola* on the basis of the temporal and spatial data of various remotely sensed climate variables. Estrada-Peña (1999) used 'Cokriging' to predict habitat suitability for *Boophilus microplus* ticks in South America by associating tick presence/absence data for selected locations with remotely sensed temperature and vegetation information. In his research, spatial dependence was taken into account. Pfeiffer et al. (1997) refined the work of Duchateau et al. (1997) with consideration of spatial dependence by utilizing generalized linear mixed logistic regression. The model included environmental and land-use risk variables, as well as a random effect to account for the local dependence between neighboring observations. Regarding HD, Ward (1994) tested 18 climatic variables for their association with the prevalence of bluetongue virus infection in the cattle herds in Queensland, Australia using logistic regression and concluded that the average daily maximum temperature and the average annual rainfall could explain most variability in the disease prevalence. All of the above models only consider spatial distribution of the disease, ignoring the time affect on the disease distribution and its temporal relations with explanatory factors. Thus a more realistic and comprehensive spatial-temporal model should be proposed for spatial epidemiology.

The initial methods proposed for spatial-temporal models were extensions of moving average methods for time series to include spatial patterns (Pickle, 2000). Later, log-linear regression methods were developed for mortality rates, assuming asymptotic normality and extra variation for the underlying Poisson or binomial data (Congdon, 1994). Afterwards, more methods were proposed, such as random fields (Handcock and Wallis, 1994), thin plate splines (Van der et al., 1995) and difference equations (Solna and Switzer, 1996). A number of the above

approaches assumed spatial stationarity (no variation across space) and isotropy (same in different directions), which generally do not hold for disease rate data. More recently, less restrictive Hierarchical Bayesian models and generalized linear mixed models have been proposed.

2.4.1 Bayesian models

Bernardinelli and Montomoli (1992) proposed a fully Baysian model in which both area-specific intercept and trend are modeled as random effects which allow for correlation among them, and perform an analysis of variation of risk for a given disease in space and time.

Hierarchical Bayes methods are increasingly popular tools for disease mapping because they permit smoothing of the fitted rates toward spatially local mean values, with more unreliable estimates receiving more smoothing. Zhu and Carlin (2000) developed a Hierarchical Bayesian model to analyze spatial-temporally misaligned data wherein the covariate is available on a grid that is a refinement of the regional grid for which the response variable is available, and where the regional boundaries may also evolve over time (Zhu and Carlin, 2000**)**. Waller et al. (1997) extended the Hierarchical Bayesian model to incorporate general temporal effects and space-time interactions, giving a hierarchical framework for modeling regional disease rates over space and time. This method requires careful implementation using Markov Chain Monte Carlo (MCMC) methods. Knorr-Held and Besag (1998) developed a dynamic model in hierarchical Bayesian framework to generate smoothness in time trends, so that estimates for any specific time can “borrow strength” from data at adjacent times. Parameter estimation from the above Bayes models has been by either a normal approximation or a MCMC resampling procedure

with an assumption of normal prior distributions (Pickle, 2000). The Bayesian models require continuous or count data as the dependent variable, which is not suitable for our research.

2.4.2 Generalized linear mixed models (GLMM)

Generalized linear mixed models (GLMM) are generalized linear models (GLM) with added random cluster and /or subject effects to account for the correlation of the data. In GLMM, the response distribution is defined conditionally depending on the random effects (Hartzel et al., 2001). It is a well-known tool in statistics for modeling non-normal data such as dichotomous, ordinal, nominal, and count data. For dichotomous data, a logistic or probit regression model is usually adopted with various methods for incorporating and estimating the influence of the random effects. The mixed-effects logistic regression model is arguably the most popular GLMM for analysis of multilevel dichotomous data. McCullagh (1980) proposed the proportional odds model based on the logistic regression formulation for the analysis of ordinal data in 1980. Since then, many of the GLMMs for ordinal data have been generalized from this model. As to nominal data, the mixed logistic model is also extended to fit nominal data (Hartzel et al., 2001). For count data, various types of Possion mixed models have been proposed. A review of some of these methods applied to longitudinal Possion data is given by Stukel (1993). The GLMM permits various spatial and temporal covariance structures, and does not require the assumption of stationarity or isotropy (Pickle, 2000).

2.5 Space-time clustering analysis

2.5.1 Space-time clustering techniques

Space-time clustering is a new branch in cluster detection. It is argued that disease cases occur close to each other in time as well as in space. Investigation of clustering of disease occurrence can provide valuable information on possible causes of the disease, and appropriate methods for disease control and prevention (Ward and Carpenter, 2000). Since the identification of space-time clustering based on visual inspection alone appears more difficult with the additional dimension of time, the statistical methods become urgent (Norström et al. 2000). The statistical methods can be divided into those based on the description of the space-time interaction and those based on the detection of cluster locations. The former one only gives a global statistical index indicating whether there are space-time clusters or not in the study area during the study period. This method is also called space-time interaction. The latter method can identify clusters with both a spatial and a temporal dimension, and is called a space-time cluster detection test (Kulldorff et al., 1998). Several techniques have been proposed and applied for space-time clustering analysis. The Knox test (Knox, 1964), the Barton's method (Barton and David, 1966), the Mantel regression (Mantel, 1967), and the k nearest neighbor test (Jacquez, 1996) are four popular methods to describe the space-time interaction. Kulldorff's space-time scan statistic is the one of the few methods for space-time cluster detection.

The Knox's test, proposed by Knox in 1964, quantifies space-time interaction based on defined space and time distances. The test statistic is a count of the number of pairs of cases that are separated by less than the critical space and time distances, and compared to the number of

events expected under a Possion model with a $\chi 2$ test (Knox, 1964; Mantel, 1967). Critics of the Knox's test mainly focus on the following two aspects: 1) subjective selection of critical space and time distance, and 2) invariant critical distance with changing population density, which is unreasonable because when population density increases, the distance from case to case will decrease (Jacquez, 1996).

Barton and David (1966) accept Knox's criterion of the number of close pairs but question the validity of the Possion test. The Barton's statistic is defined as the ratio of the squared distances among all temporal clusters to the mean squared distance between all cases. A statistic less than 1 indicates that there is a time-space interaction. The temporal clusters are those cases for which the time interval between clusters is smaller than the average interval range (Barton and David, 1966; Ward and Carpenter, 20000a). It partially solves the subjective selection of critical distances in Knox's test by using the actual location of events. However, in this method, small distances have less influence on the test statistic than do large distances do (Mantel, 1967).

Mantel's regression is the sum of the time distances multiplied by the spatial distances for all case pairs, and then standardized by the number of paired cases (Mantel, 1967). Mantel's regression is criticized for not being sensitive to non-linear dependence of time on space because it is based on a linear model. Although different data transformations are applied to investigate the non-linearities, the selection of an appropriate data transformation method is subjective (Jacquez, 1996; Ward and Carpenter, 2000a).

Jacquez (1996) extended the nearest-neighbor test for spatial clustering to a k nearest-neighbor test for space-time interaction. The test statistic is the count of the number of case pairs that are k nearest-neighbors both in time and space. It avoids the introduction of subjectivity because it does not need to specify critical time and space distances. Like Mantel's regression, it is also sensitive to both linear and non-linear dependencies between the space and time of occurrence of cases (Ward and Carpenter, 2000a).

Modifications of the above methods were also proposed and applied to practice, but they did not gain wide attention (Pinkel and Nefzger, 1959; Pinkel et al., 1963; Smith et al., 1976; Klauber, 1971). All of the above space-time interaction tests are designed for point locations of cases, and none of them take into account the dynamic change of the underlying population at risk. Kulldorff's space-time scan statistic not only allows the detection of actual geographical location and temporal period of clusters, but also incorporates data on the distribution of the population at risk (Norström et al. 2000). This is a great contribution because epidemiologists are usually interested in disease clusters only after the spatial variations of at-risk population density are adjusted. Otherwise, the results may be misleading. Another advantage of Kulldorff's space-time scan statistic is that it can be applied to point data as well as area data for the detection of spatial-temporal clusters (Kulldorff, 2006).

Kulldorff's space-time scan statistic is an extension of Kulldorff's spatial scan statistic and uses either a Possion-based model or a Bernoulli model (Norström et al. 2000). It imposes a cylindrical window in the three dimensions of space and time, with its base indicating space and height denoting time. The window is moved in space and time so that the base is centered on

each possible geographic position throughout the study area. For any given geographic position, the radius of the base varies continuously in size up to a specified upper limit, and the height of the cylinder also varies across all the possible time intervals up to a specified upper limit (Kulldorff et al., 1998; Song and Kulldorff, 2003). For each window, a likelihood value is computed based on the number of observed and expected cases within and outside the window. The window with the highest likelihood value is identified as the potential space-time cluster, and its likelihood value is compared with the likelihood under the null hypothesis to obtain the likelihood ratio. A p-value is also computed for this cluster (Kulldorff et al., 1998; Aamodt, 2006).

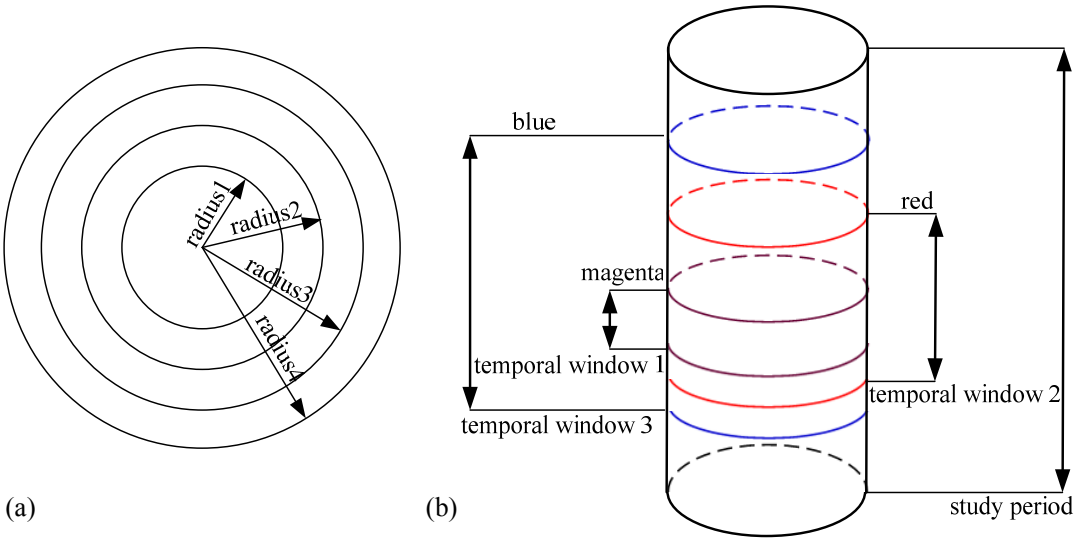


Figure 2.6: Schematic illustration of Kulldorff's space-time scan statistic method. (a): varying spatial windows for a target location; (b) varying temporal windows for a particular spatial window at a target location.

2.5.2 Space-time clustering applications

There are several applications of the investigation of space-time interaction in veterinary literature. White et al. (1989) applied the Knox's test and Mantel regression method to

investigate the space-time clustering of outbreaks of winter dysentery in cattle herds in Ithaca, New York. Paré et al. (1996) utilized Knox's test to analyze the space-time clustering of horses with *Salmonella krefeld* in their feces during hospitalization. In another research example conducted by Carpenter et al. (1996), Knox's test was adopted in the spatial-temporal analysis of fowl cholera outbreaks in turkeys between August 1985 and July 1986 in California. Singer et al. (1998) applied Barton's test to assess the space-time clustering of ampicillin- and tetracycline-resistant isolates of *Pasteurella multocida* and *P. haemolytica* from California cattle with pneumonia. Norström et al. (2000) used Kulldorff's space-time scan statistic, Knox's test, and Jacquez's k nearest-neighbor test to assess the presence of space-time clustering and interaction during an outbreak of acute respiratory disease in Norwegian cattle. Fuchs et al. (2000) investigated the space-time clustering of scabies in chamois in Austria using Knox's test and Mantel's regression. Ward and Carpenter (2000a) made use of blowfly capture data to demonstrate the use of various space-time clustering methods including Knox's test, Mantel's regression, Barton's test and Kulldorff's space-time scan statistic. They recommended the use of several methods together to increase the statistical power of the analysis. In Ward and Armstrong's (2000) study, the reported occurrence of blowfly strike in 57 sheep flocks in Queensland, Australia, was investigated for time-space clustering using Knox's method. Despite the above applications, space-time clustering investigations in veterinary science are not common possibly due to the following three reasons: 1) there is a general lack of information on the location of disease outbreaks and their timing; 2) it is not easy to explain and visualize the

concept of space-time interaction and cluster analysis; and 3) there are few readily available space-time clustering software packages.

In the spatial data mining field, several software packages that integrate visualization for exploring patterns hidden in the data set have been developed. These software packages are designed for interactive visual and/or statistical analysis of spatially and temporally referenced data, including clustering analysis. Some examples are CommonGIS, spatial mining for data of public interest (SPIN!), cubeview visualization system, and TerraSeer space-time information system (STIS). However, the clustering analysis in these software packages mainly focus on qualitative visualization aspects. Furthermore, the spatial and temporal distributions are analyzed separately with only simple statistical analysis on spatial clustering. Some other software packages that specialize in statistical space-time clustering analysis include CLUSTER, Stat!, and SaTScan™. CLUSTER and Stat! are designed for space-time interaction analysis, and SaTScan™ is used for space-time cluster detection.

CHAPTER 3

SPATIAL-TEMPORAL PREDICTION MODEL OF HEMORRHAGIC DISEASE IN WHITE-TAILED DEER IN SOUTHEAST USA: 1982 – 2000 [1]

[1] Xu, B., Stallknecht, D. E., Madden, M., Hodler, T. W. and Parker, K. C. To be submitted to *International Journal of Remote Sensing*.

Abstract: Spatial and temporal patterns of outbreaks of hemorrhagic disease (HD) in white-tailed deer in the USA may be related to spatial and temporal variance in environmental and climatic conditions. This paper proposes the development of a spatial-temporal database and statistic model to predict the geographic and temporal distribution of HD in white-tailed deer in southeast USA. The HD occurrence data available from the Southeastern Cooperative Wildlife Disease Study (SCWDS) dates back to the early 1980s. From this unique nation-wide and county-based HD survey data set, binary data on presence and absence of HD for individual counties in Alabama, Georgia, South Carolina, North Carolina, and Tennessee were extracted for the years 1982 to 2000. These dates coincide with available historical climatic data including temperature, rainfall, wind speed, and dew point for the five southeast states available from 1982 to 2000. In addition, archived remotely-sensed Advanced Very High Resolution Radiometer (AVHRR) satellite data were used to derive normalized difference vegetation index (NDVI) and land surface temperature data set. Other predictor variables included spatial dependency of HD in adjacent counties, time, and elevation. This study first applied principal component factor analysis to reduce the data volume and eliminate covariance between variables. Next a generalized linear mixed logistic model was used to develop a spatial-temporal statistical model, taking HD data as the dependent variable and the principal factors derived from the principal component factor analysis as predictor variables. The spatial dependency was considered by incorporating a spatial association term which evaluates the effect of HD occurrence in surrounding counties on a particular county. The results show that wind speed, rainfall, land surface temperature and NDVI were useful factors in predicting HD occurrence, with total

prediction accuracy of 65 percent when all four factors were considered for a five state area. The prediction accuracy for individual year ranges from 27 percent to 96 percent. Remotely-sensed data proved to be informative and resulted in a higher prediction power than some ground-based weather station climatic data.

Keywords: Hemorrhagic disease, Remote sensing, Spatial-temporal, Generalized linear mixed logistic model

3.1 Introduction

Hemorrhagic disease (HD) causes significant mortality in white-tailed deer and has been documented in the USA since 1890s (Trainer, 1964; Hoff and Trainer, 1981). The disease pathology, viruses causing the disease (bluetongue and epizootic hemorrhagic disease viruses), and the transmitting vectors (*Culicoides* midges) have been studied extensively by veterinary researchers (Hoff and Trainer, 1978; Nettles and Stallknecht, 1992). However, to date, few investigators to date have focused on the distribution of HD in white-tailed deer in relation to environmental factors that may be used to predict outbreaks of the disease. Such studies have been done with bluetongue virus in cattle by Ward (1994). Testing 18 climatic variables for their association with the prevalence of bluetongue virus infection in the cattle herds of Queensland, Australia, Ward concluded that ground-based measures of average daily maximum temperature and the average annual rainfall could explain the most variability in HD prevalence. Even in this study, Ward only applied simple ordinary linear regression without considering the spatial and temporal effects. Furthermore, it has been demonstrated that remotely-sensed images are correlated with certain climatic variables recorded on the Earth's surface (Hay et al., 1996; Baylis and Rawlings, 1998). Given the 30-year archive of global remote sensing data, it is now possible to correlate variance in climatic and environmental conditions with spatial and temporal distributions of HD outbreaks

The objective of this research is to develop a spatial-temporal statistical prediction model to investigate the relationship between occurrence of HD in white-tailed deer and various climatic/environmental factors derived from a combination of ground-based remotely-sensed

data. Section 3.2 provides a brief introduction to the use of remote sensing and GIS techniques in veterinary epidemiology and spatial-temporal modeling. Section 3.3 describes the study area and the mechanism for the project. Section 3.4 describes the HD, climatic, and remotely-sensed data that are utilized to construct the statistical model. In Section 3.5, the proposed spatial-temporal model is detailed, while Section 3.6 discusses the results and findings from the proposed model. Finally, Section 3.7 presents a summary and conclusion.

3.2 Background

3.2.1 Remote sensing and GIS in epidemiology

Remote sensing and geographic information system (GIS) techniques have been finding their way into epidemiological studies because of their value for obtaining and processing information about the Earth's surface over extensive areas and long time periods. Baylis et al.(1999), for example, demonstrated that variables derived from satellite images performed better than climatic variables measured by weather stations in their predictive modeling of the distribution of *Culicoides imicila* in southern Africa. They concluded the reason might be that remotely-sensed data recorded more biologically relevant data that affected the distribution and abundance of vectors influencing the status of disease occurrence. Further studies demonstrated easily available remotely-sensed data combined with GIS could be used to explain the observed patterns not only more efficiently, but also more effectively (Baylis et al., 1998; Baylis and Rawlings 1998; Baylis et al., 1999; Baylis et al., 2001).

Recently there have been several studies on predicting vector distributions using remotely-sensed data as surrogates for climatic data. Some examples are normalized difference vegetation index (NDVI) representing vegetation conditions, land surface temperature derived from channel 4 and channel 5 of the Advanced Very High Resolution Radiometer (AVHRR) substituting temperature, and cold-cloud duration (CCD) indicating rainfall (Linthicum et al., 1987; Wood et al., 1991; Rogers and Williams, 1993; Cross et al., 1996; Hay et al., 1996; Rogers et al., 1996; Estrada-Peña, 1997; Baylis et al., 1999; Boone et al., 2000; Glass et al., 2000; Hay, 2000; Thomson and Conner, 2000). However, almost all the previous studies concentrate on the relationship between the distribution of vectors of the disease and remotely-sensed (Durr, 2004). Few studies are conducted directly on the relationships between the distribution of the hosts of the disease, environmental factors and remotely-sensed data.

3.2.2 Spatial-temporal model

Modeling is a necessary technique for the explanation and prediction of the relationship between disease and affecting factors, usually environmental and social in origin. Some work has been done to predict disease occurrence with climatic or remotely-sensed data using modeling (Rogers and Randolph, 1993; Roberts et al., 1996; Rogers et al., 1996; Beck et al., 1997; Snow et al., 1998; Glass et al., 2000; Lindsay and Thomas, 2000; Baylis et al., 2001; Hales et al., 2002). In these studies, simple linear discriminant analysis, logistic regression, and ordinal linear regression were applied assuming that the occurrence of disease and transmission of viruses are spatially and temporally independent, which violates both the mechanism behind disease

distribution and the spatial and temporal dependency. Some studies did consider the spatial dependency using essentially two methods (Augustin et al., 1996; Thomson et al., 1999; Kleinschmidt et al., 2000). One assumes that the spatial covariance between points is a function of distance between them, and by analyzing the variogram of the residuals from the fitted model, the spatial correlation in the data can be estimated. The other method explicitly adds an additional spatial covariate term into the model. The additional term is a distance or proximity-weighted average of values for each point or area. These studies take spatial dependency into account, but none of them incorporate temporal aspects.

Traditional longitudinal statistical analyses such as marginal models and generalized linear mixed models (GLMM) have been applied in epidemiology to consider the covariance between repeated measurements for one subject at different times (Fienberg et al., 1985; Van Marter et al., 1990; Lauer et al., 1997; Rogan et al, 2001; Hedeker, 2003). These models, on the contrary, consider time effects, but not the spatial correlation. Many more complex Bayesian hierarchical models and GLMM have been proposed over the years and have gained wide acceptance for application in spatial-temporal disease mapping (Waller et al., 1997; Knorr-Held and Besag, 1998; Xia and Carlin, 1998; Bohning et al., 2000; Pickle, 2000; Zhu and Carlin, 2000). However, they are not extensively applied to prediction models in epidemiology.

3.3 Study area and mechanism

3.3.1 Study area

The southeast USA has a long history of HD occurrence in white-tailed deer. As early as 1949, Ruff (1949) pointed out that extensive mortality of an unexplained fatal disease in white-tailed deer occurred at irregular intervals for many years in this region which was later believed to be similar to HD. Since then, occurrences of HD were traditionally reported and caused significant die-offs, such as the outbreaks in 1971 (Thomas et al., 1974), 1980 (Couvillion et al., 1981), and 1998 (Nettles and Stallknecht, 1992). Furthermore, precipitating antibodies to bluetongue virus (BLU) and Epizootic Hemorrhagic disease (EHD) are consistently detected in white-tailed deer annually in this region (Stallknecht et al., 1991). Annual data collected by the Southeastern Cooperative Wildlife Disease Study (SCWDS) in the College of Veterinary Medicine, University of Georgia indicates that contiguous instances of HD have been reported throughout the Southeast since 1980 to present (Nettles et al., 1992). Thus, our study focuses on five states in the southeastern region of the USA: Alabama, Georgia, South Carolina, North Carolina, and Tennessee (Figure 3.1).

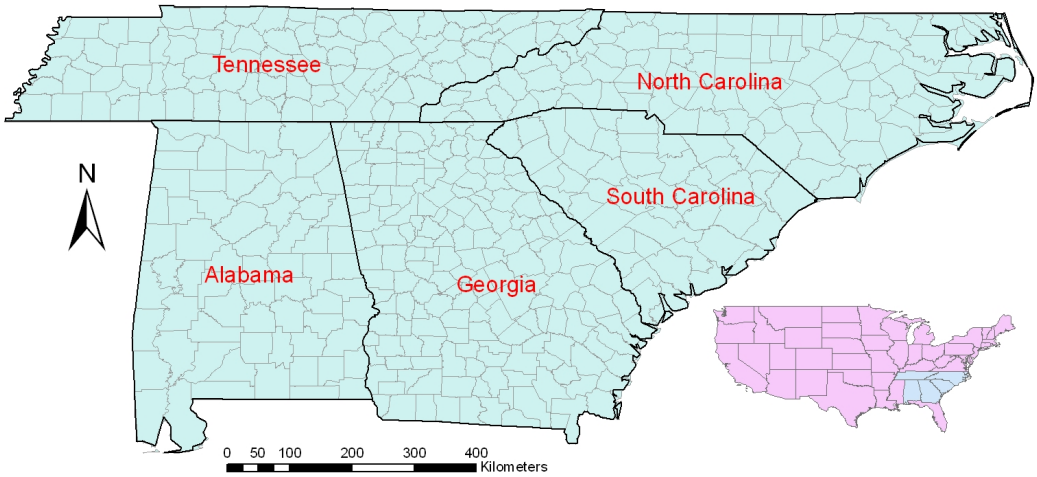


Figure 3.1: Study area: Alabama, Georgia, South Carolina, North Carolina, and Tennessee

3.3.2 Mechanism

Given the importance of many aspects of vector behavior and biology to the transmission of vector-borne diseases, it is apparent that the distribution and intensity of such diseases is usually dependent upon the abundance and distribution of the vectors, which in turn depend on climate and environment (Hay et al., 1996). Hemorrhagic disease is a vector-borne disease transmitted by species of the genus *Culicoides*. As a result, environmental factors affect the epidemiology of the disease through their effects on the distribution, size and abundance of the vector population, which, in turn, is thought to account for the seasonal occurrence of HD discovered by previous investigators (Walker and Davies, 1971; Ward, 1994; Hay et al., 1996; Wittmann and Baylis, 2000). Environmental factors also directly influence the epidemiology of the disease by affecting the distribution and growth of the herds. There is, therefore, a causal relationship between the occurrence of HD and environmental factors. Furthermore, studies also

show that climatic data can be surrogated by remotely-sensed data (Hay et al., 1996; Baylis and Rawlings, 1998; Baylis et al., 1999). These conditions constitute the theoretical basis that I use to construct a prediction model depicting the relationship between the presence/absence of HD and the climatic and remotely-sensed data (Figure 3.2).

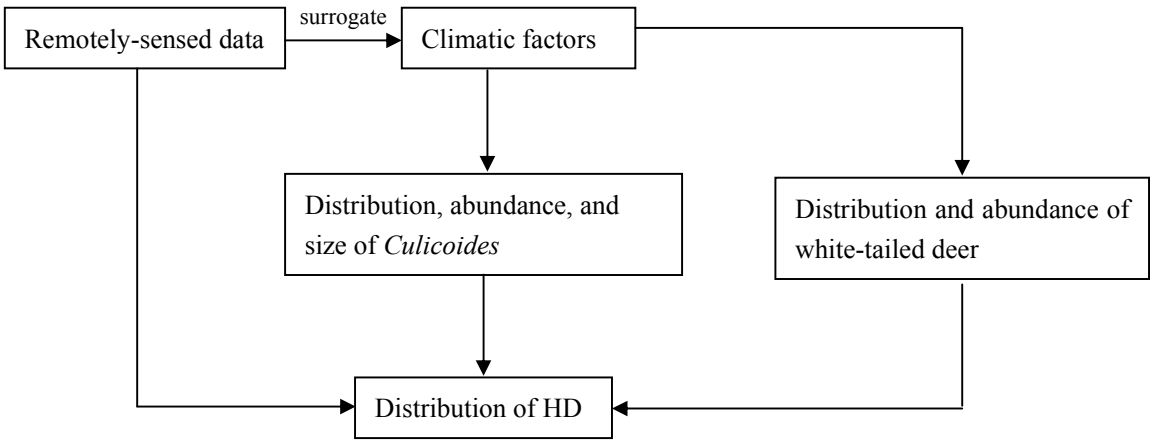


Figure 3.2: Mechanism of the model

3.4 Data sources

To construct a statistical prediction model for HD distributions, HD occurrence data for each county in the five states are needed as the response variable, and various climatic data and remotely-sensed data are used as predictor variables.

3.4.1 HD data

The HD occurrence data are collected on a county basis through the surveillance project conducted by SCWDS. This surveillance project has been performed annually beginning in 1980 and is the most comprehensive database for white-tailed deer morbidity and mortality anywhere.

The researchers mail questionnaires to the state fish and wildlife agency in each state, plus most of the state veterinary diagnostic laboratories. The surveillance also includes personnel in the U.S. Fish and Wildlife Service and the Animal and Plant Health Inspection Service, USDA. All states, except Hawaii, were surveyed from 1982 to present. The researchers set four criteria for the surveillance reporting as follows:

1) Sudden, unexplained, high deer mortality reported during the late summer and early fall.

2) Necropsy diagnosis of HD as rendered by a trained wildlife biologist, a diagnostician at a State Diagnostic Laboratory or Veterinary College, or by SCWDS personnel.

3) Isolation of EHD or bluetongue virus from a deer.

4) Observation of hunter-killed deer that showed sloughing hooves, ulcers in the mouth, or scars on the rumen lining as indirect evident of chronic HD occurrence.

Among them, criteria 1, 2, and 3 are collectively called deer mortality, and criterion 4 is deer morbidity. From the original report of each state, a file reporting the HD occurrence by years and counties from 1982 to 2000 was created for this study. The data are binary, if there is HD occurrence in the county in the specific year, it is represented as 1, otherwise, it is 0 (Table 3.1). Those counties that do not report any occurrence in any year are discarded from the research.

Table 3.1: HD data explanation

Value	Explanation
0	No HD exists in the county in the year
1	HD exists in the county in the year

3.4.2 Climatic data

The climatic data were obtained from the national climatic data center (NCDC). The original data are hourly measures of temperature, rainfall, dew point, and wind speed at 161 weather stations located throughout the study area. These hourly measures were processed to obtain the daily maximum, minimum, and mean data. After that, the daily data were aggregated to monthly average maximum, monthly average minimum, monthly average mean, monthly highest, and monthly lowest data. The monthly data were then aggregated to corresponding yearly data and period data from April to October when HD is mostly likely to occur. Since these data were attached to the 161 weather stations, an interpolation technique (kriging) with a cell size of 30 meters was applied in ArcGIS to create a continuous surface of values across the study area for each year and each period. Finally, the interpolated data were subject to zonal statistics using counties as zonal areas to produce the zonal average value of the corresponding climatic variables for each county in ArcGIS. Figure 3.3 gives the weather stations, interpolation results and zonal statistics results for mean temperature in 1980. All together, there are 30 climatic variables with 15 yearly-based variables and corresponding 15 period-based variables(Table 3.2).

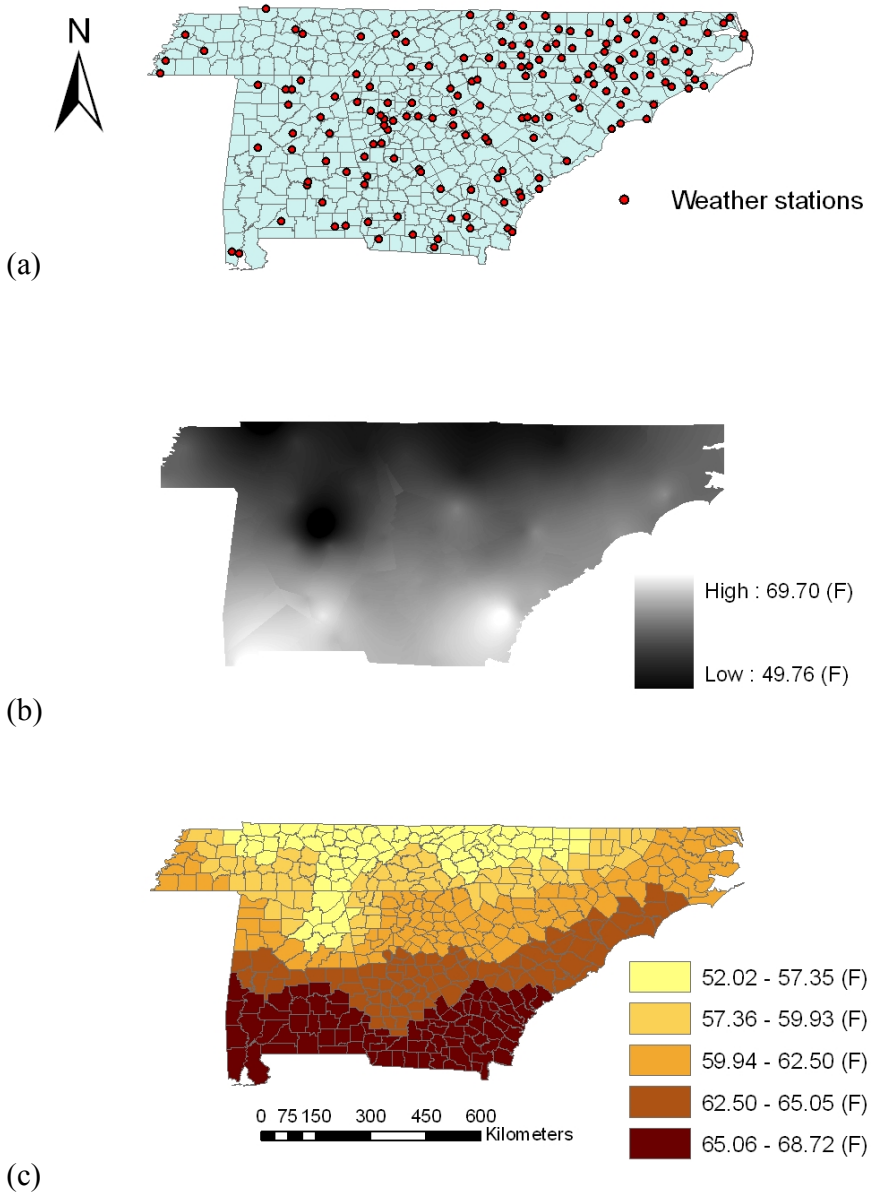


Figure 3.3: Data processing for mean temperature in 1980: (a) Weather stations in study area; (b) Temperature interpolation results for entire study area; and (c) Zonal statistics results for each county

Table 3.2: Climatic variables

Variable Description	Variables for the year	variables for the period from April to October
Mean dew point	Davg	D'avg
Mean average maximum dew point	Dmax	D'max
Mean average minimum dew point	Dmin	D'min
Mean rainfall	Ravg	R'avg
Mean average maximum rainfall	Rmax	R'max
Mean average minimum rainfall	Rmin	R'min
Mean temperature	Tavg	T'avg
Highest temperature	Thigh	T'high
Lowest temperature	Tlow	T'low
Mean average maximum temperatures	Tmax	T'max
Mean average minimum temperatures	Tmin	T'min
Mean wind speed	Wavg	W'avg
Highest wind speed	Whigh	W'high
Mean average maximum wind speed	Wmax	W'max
Mean average minimum wind speed	Wmin	W'min

3.4.3 Remotely-sensed data

As mentioned above, some remotely-sensed data can be surrogates of environmental indices. It is also argued that remotely-sensed data may produce better results because they not only capture measures over a larger area rather than a point measurement recorded by a weather station, but also the data may reflect the biophysical characteristics which contribute to the occurrence of disease and transmission of viruses. In this research, I use NDVI index and land surface temperature derived from channels 4 and 5 of the AVHRR as predictor variables to see whether remotely-sensed data can be used to predict HD occurrence.

NDVI index is calculated as follows:

$$NDVI = (Ch2-Ch1)/(Ch2+Ch1) \qquad \text{Equation 3.1}$$

Where Ch1 is the channel 1 reflectance (visible red, 0.58 – 0.68μm) in AVHRR

Ch2 is the channel 2 reflectance (near infra-red, 0.72 – 1.10μm) in AVHRR

This vegetation index is theoretically a specific measure of chlorophyll abundance and energy absorption (sometimes described as greenness), and now is extended to cover vegetation biomass (Tucker et al., 1983), coverage (Tucker et al., 1985), seasonal rainfall (Linthicum et al. 1987), and phenology (Justice et al., 1985) in a range of ecosystems.

The relationship between land surface temperature (Ts) and thermal infrared channel 4 (10.3 – 11.30μm) and channel 5 (11.5 – 12.50μm) brightness temperatures in AVHRR (Ch4 and Ch5, respectively) is provided in Equation 3.2.

$$Ts = Ch4 + 3.33*(Ch4-Ch5) \quad \text{Equation 3.2}$$

Ch4 gives a brightness temperature based on Planck's law which quantifies the spectral radiance of electromagnetic radiation from a black body as a function of wavelength. The second part of the equation modifies the estimate to allow for attenuation of the signal by the atmosphere (Price, 1984; Sugita and Brutasaert, 1993; Cooper and Asrar, 1989; Baylis et al., 1999).

The NDVI index and Channel 4 and Channel 5 were obtained from the NOAA/NASA Pathfinder AVHRR Land program (ftp://disc1.gsfc.nasa.gov/data/avhrr/). The Pathfinder Program produces data set derived from the observations made by AVHRR on the "afternoon" NOAA operational meteorological satellites (NOAA-7, -9, 11) from 1982 to 2000. In this research, the 10-day composite of NDVI, AVHRR channel 4 brightness temperature, AVHRR channel 5 brightness temperature with spatial resolution of 8km x 8km were used to obtain the final remotely-sensed variables for each year and period in the study area. The NDVI index is

already provided by the Pathfinder AVHRR Land program and can be directly downloaded from the NASA data center. Surface temperature is calculated from Channel 4 and Channel 5 of AVHRR using Equation 3.2. The original data are compressed binary files scaled to 8 bit (.NDVI) and 16 bit (Channel 4 and Channel 5). After downloading, these files were decompressed in the linux system, and imported to PCI Geomatics software as .pix file format. The imported files were then exported to ERDAS Imagine software and rescaled to original 32 bit data. Land surface temperatures were calculated according to Equation 3.2. Subsequently, maximum, minimum, mean values were extracted for each year and each period in ERDAS Imagine. In ArcGIS, the above images were subset to the study area and subject to zonal statistics to obtain the final data for each county.

The final remotely-sensed variables are listed in Table 3.3. Combined with the availability of HD data and explanatory data, all together, 366 of the 467 counties in the five states were included in the study, and 101 counties with no HD occurrence reported were excluded.

3.4.4 Elevation data

The elevation data were used as a predictor variable in the model because it is thought the elevation can influence the distribution of *Culicoides*. Since the analysis is based on the county level, the mean elevation of each county is calculated for the input. The original 30-meter resolution elevation data were downloaded from the National Map Seamless Server run by the U. S. Geological Survey (USGS) which offers seamless USGS National Elevation Dataset (NED) (http://ned.usgs.gov/). After downloading, the mean elevation of each county was calculated

using zonal statistic function of Spatial Analyst Extension in ArcGIS with county map as the zone base.

Table 3.3: Remotely-sensed variables

	Remotely-sensed Variable description	Variables for the year	Variables for the period from April to October
Lmax	Mean land surface temperature	Lmax	L'max
Lmean	Mean average maximum land surface temperature	Lmean	L'mean
Lmin	Mean average minimum land surface temperature	Lmin	L'min
Nmax	Mean NDVI	Nmax	N'max
Nmean	Mean average maximum NDVI	Nmean	N'mean
Nmin	Mean average minimum NDVI	Nmin	N'min

3.5 Spatial-temporal model

In this study, each of the 366 counties in the study area is measured repeatedly from 1982 to 2000 on a yearly basis (19 years). All together, there are 6954 observations for response variable and predictor variables. The response variable is 1 if HD is present and 0 if HD is not present. The hypothesis is that the probability that HD is present in a county is dependent on the climatic factors, remotely-sensed data, elevation, time, as well as spatial dependency (whether or not the disease occurs in its neighboring counties).

3.5.1 Principal component factor analysis

The original data contain 30 climatic variables, 12 remotely-sensed variables, one elevation variable, one time variable, and one spatial dependency variable, or 45 total variables. Furthermore, it is argued that correlation exists between and within climatic variables and remotely-sensed variables, which violates the assumption of independency between predictor variables for linear regression. Thus, before constructing the spatial-temporal model, the climatic and remotely-sensed variables were subject to principal component factor analysis for data reduction and correlation elimination. Principal components factor analysis combines traditional factor analysis and principal components analysis together. Although it applies the same methods used in traditional common factor analysis, it does not analyze the common variance as in factor analysis. Instead, it considers the total variance as in principle components analysis. On the other hand, different from principle components analysis, it yields a component loading matrix that can be rotated for ease of interpretation.

Table 3.4 shows the principal component factors after performing the principal component factor analysis. Two criteria are commonly used to help choose the optimal number of factors for later analysis. The first is the Kaiser criterion proposed by Kaiser (1960). According to this criterion, I can retain only factors with eigenvalues greater than 1. The second criterion is the graphical scree test proposed by Cattell (1966), which plots the eigenvalues in a line plot. It is suggested to find the place where the smooth decrease of eigenvalues appears to level off to the right of the plot (Cattell, 1966). Those factors with engenvalues greater than the eigenvalue corresponding to leveling off of the plot were retained. It is argued that the Kaiser

criterion sometimes retains too many factors, while the scree test retains too few (Linn, 1968; Tucker et al., 1969).

Figure 3.4 shows the scree plot with a line indicating eigenvalue = 1. From the scree plot, I can see that the eigenvalues level off at a value of 1.2, and the factors to the left of this point should be retained. That is, the first 8 factors should be used to express the total variance in the data set. Whereas based on Kaiser's criterion, 11 factors should be chosen. By examining the cumulative percentage in Table 3.4, the first 8 factors only explain 78.28 percent of the total variance, while the first 11

Table 3.4: Principal component factors

Factor	Eigenvalue	Proportion	Cumulative
1	10.37224	0.247	0.247
2	5.11598	0.1218	0.3688
3	4.69331	0.1117	0.4805
4	3.51211	0.0836	0.5641
5	3.01552	0.0718	0.6359
6	2.62788	0.0626	0.6985
7	1.88508	0.0449	0.7434
8	1.65531	0.0394	0.7828
9	1.18764	0.0283	0.8111
10	1.14059	0.0272	0.8382
11	1.10147	0.0262	0.8645
12	0.72896	0.0174	0.8818
13	0.64036	0.0152	0.8971
14	0.58059	0.0138	0.9109
15	0.47861	0.0114	0.9223
16	0.43482	0.0104	0.9326
17	0.40559	0.0097	0.9423
18	0.38336	0.0091	0.9514
19	0.34532	0.0082	0.9596
20	0.3159	0.0075	0.9672
21	0.22062	0.0053	0.9724
22	0.15841	0.0038	0.9762
23	0.12246	0.0029	0.9791
24	0.11002	0.0026	0.9817

25	0.09528	0.0023	0.984
26	0.0885	0.0021	0.9861
27	0.08295	0.002	0.9881
28	0.07209	0.0017	0.9898
29	0.06153	0.0015	0.9912
30	0.05561	0.0013	0.9926
31	0.05247	0.0012	0.9938
32	0.04381	0.001	0.9949
33	0.03741	0.0009	0.9958
34	0.03366	0.0008	0.9966
35	0.03187	0.0008	0.9973
36	0.02502	0.0006	0.9979
37	0.02308	0.0005	0.9985
38	0.01709	0.0004	0.9989
39	0.01466	0.0003	0.9992
40	0.0122	0.0003	0.9995
41	0.0112	0.0003	0.9998
42	0.00944	0.0002	1

factors can explain 86.45 percent. So in our analysis, I choose the first 11 factors for input as predictor variables in the spatial-temporal model.

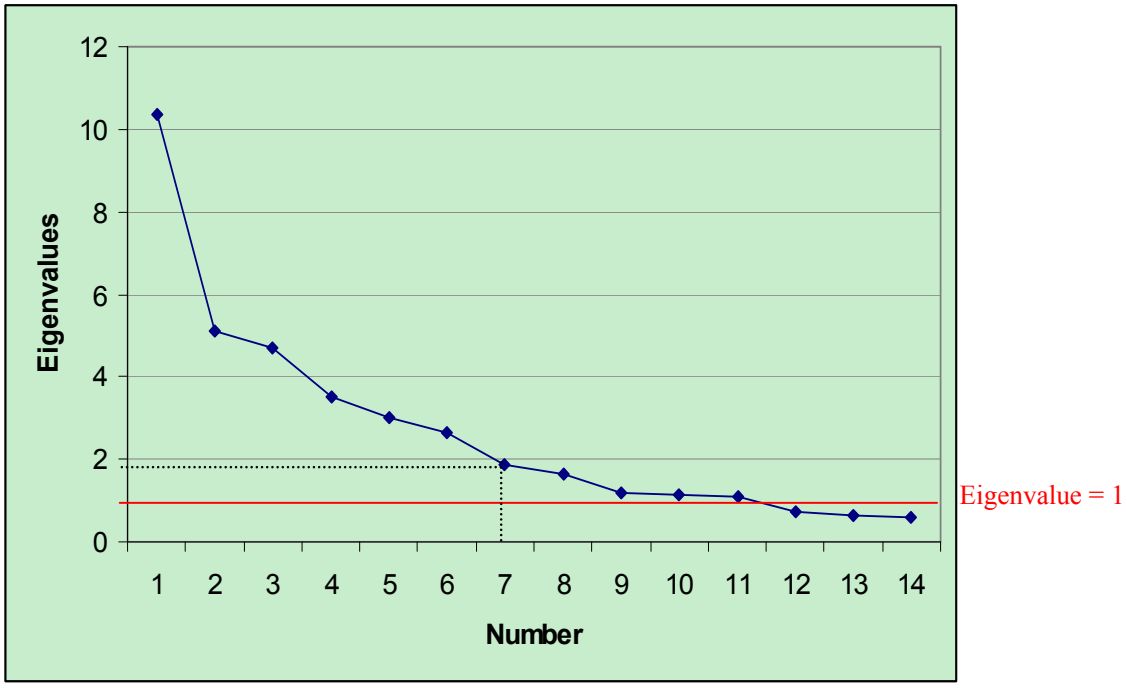


Figure 3.4: Scree-test for principle component factor analysis

Table 3.5 shows the rotated factor loadings for each of the 11 factors. As mentioned above, rotated factor loadings make it easier to interpret the correlation between factors and original variables. The underlying premise is that by rotating the axes of the scatterplot of factor loadings, a much clearer pattern of loadings can be obtained without changing the relative correlations between factors and original variables. Various rotation strategies are proposed such as varimax, quartimax, and equamax (Hill and Lewicki, 2007). In our research, the most popular varimax rotation strategy was used. The factors and its interpretation are listed in Table 3.6.

3.5.2 Spatial dependency

Statistical prediction modeling of disease is often complicated by spatial association, which typically gives positive correlations between observations spatially close to each other. This dependency violates the basic assumptions underlying standard linear regression (Thomson et al., 1999). If the modeling does not take spatial dependency into account, the resulting estimates may be inaccurate. In this research, an additional spatial dependency term (SA) is calculated from the original data and added into the model to account for spatial association. The computation of SA is shown in Equation 3.3.

Table 3.5: Factor loadings and corresponding variables

Factor 1		**Factor 2**		**Factor 3**		**Factor 4**	
Variable	Loading	Variable	Loading	Variable	Loading	Variable	Loading
dminp	0.906	nmeany	0.959	wavgy	0.943	lmeanp	0.923
davgp	0.905	nmeanp	0.929	wavgp	0.902	lmaxp	0.848
dmaxp	0.896	nmaxp	0.9	wminy	0.861	lminp	0.797
tavgp	0.889	nmaxy	0.885	wminp	0.828	lmeany	0.747
tmaxp	0.865	nminp	0.658	wmaxy	0.782	lmaxy	0.306
tlowp	0.832	nminy	0.649	wmaxp	0.746	tmaxy	0.266
tminp	0.62	lmaxy	0.157	lmaxy	0.257	nminp	0.21
davgy	0.358	lmeany	0.138	rminy	0.199	tavgy	0.189
Factor 5		**Factor 6**		**Factor 7**		**Factor 8**	
Variable	loading	Variable	loading	Variable	loading	Variable	loading
whighp	0.917	thighp	0.226	ravgp	0.923	rminp	0.959
whighy	0.906	tmaxp	0.156	thighp	0.895	nminp	0.147
wmaxp	0.468	tminp	0.139	rmaxp	0.714	rmaxp	0.103
tminp	0.317	tavgp	0.096	tlowy	0.206	davgp	0.086
wmaxy	0.315	lmaxp	0.073	whighy	0.095	dminp	0.083
tlowy	0.292	wavgp	0.067	whighp	0.083	dmaxp	0.079
rmaxy	0.28	tmaxy	0.066	lmeanp	0.072	lminy	0.077
nminp	0.185	rminp	0.06	lmaxp	0.061	tavgp	0.072
Factor 9		**Factor 10**		**Factor 11**			
Variable	loading	Variable	loading	Variable	loading		
dminy	0.336	nmaxp	0.302	lminy	0.911		
rminy	0.308	nmeanp	0.255	lmeany	0.386		
nminp	0.291	wmaxy	0.244	nmaxy	0.151		
lminp	0.169	nmaxy	0.211	dminy	0.082		
wminy	0.143	rmaxp	0.154	davgy	0.072		
wminp	0.109	wavgy	0.111	tminy	0.07		
ravgp	0.076	rminp	0.046	lminp	0.069		
lminy	0.058	wmaxp	0.039	tavgy	0.064		

$$SA_i = \frac{\sum_{k=1}^{n} w_{ik} y_k}{\sum_{k=1}^{n} w_{ik}}$$ Equation 3.3

Where SA_i is the spatial dependency term for county i

w_{ik} is the spatial weight of county k on county i

n is the number of counties that are adjacent to county i

y_k is the HD response value for county k

Table 3.6: Factor and interpretation

Factor	Climatic and remotely-sensed variable
1	period temperature, period dew point
2	period NDVI, yearly NDVI
3	yearly wind speed, period wind speed
4	yearly land surface temperature, period land surface temperature
5	period and yearly wind speed, period and yearly temperature
6	period temperature
7	period rainfall
8	period minimum rainfall, period minimum NDVI
9	Yearly minimum dew point, yearly minimum rainfall
10	period NDVI, yearly maximum wind speed
11	yearly land surface temperature

For polygons such as counties, w_{ik} can be calculated as distance between centroids of polygons, or as binary output with 1 denoting adjacency between two polygons and 0 denoting non-adjacency, or based on the length of shared borders between two polygons. In this research, the length of shared borders is used to calculate w_{ik} because it makes sense that if two counties have a long shared border, the deer and the vectors are more likely to move from one county to the other. Thus the surrounding county k that has a longer shared border with county i carries more weight in county i. w_{ik} is computed in The TerraSeer Space-Time Intelligence System (STIS) software. The output from STIS is a spatial weight matrix between each county and every other county. The weights are then multiplied by the corresponding HD occurrence and summed to determine the spatial dependency (SA) for each county.

Besides the 11 principal factors (f1 to f11) resulting from the principal component factor analysis, SA, elevation (E) and time (T) variables are also integrated into the statistical model. Table 3.7 is the correlation matrix between all the 14 predictor variables.

Table 3.7: Correlations between predictor variables

	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	SA	E	T
f1	1.00													
f2	0.00	1.00												
f3	0.00	0.00	1.00											
f4	0.00	0.00	0.00	1.00										
f5	0.00	0.00	0.00	0.00	1.00									
f6	0.00	0.00	0.00	0.00	0.00	1.00								
f7	0.00	0.00	0.00	0.00	0.00	0.00	1.00							
f8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00						
f9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00					
f10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00				
f11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00			
SA	0.09	0.01	-0.11	0.03	0.04	0.04	-0.14	0.26	0.04	-0.07	0.01	1.00		
E	-0.32	0.35	-0.25	-0.26	-0.06	-0.03	-0.01	-0.06	-0.03	0.40	-0.04	-0.06	1.00	
T	0.21	0.32	0.00	-0.07	0.53	-0.16	-0.01	-0.14	-0.33	0.06	-0.24	0.04	0.00	1.00

As is expected, there is no correlation between the 11 factors. However, correlation exists between the other three variables and the 11 factors as well as within the three variables. There are varying opinions in the literature concerning what level of correlation constitutes multicollinearity. Jensen (1967) advanced the conservative and liberal views. The conservative view is to assume multicollinearity if two variables have a correlation coefficient greater than 0.5, whereas the multicollinearity in the liberal view is with a correlation greater than 0.9. Even if I adopt Jensen's conservative view, only the correlation between f5 and time (highlighted in brown) can be regarded as causing multicollinearity. Thus when building the model, I should be careful

if these two variables are included at the same time. Other variables that have correlations greater than 0.2 (highlighted in yellow) also should also be paid attention to for the sake of interpretation.

3.5.3 Model construction

The data are typical longitudinal data which means measurements of the same subject or individual (in this case county) are taken repeatedly through time, thereby allowing the direct study of change over time (Rabe-Hesketh and Skrondal, 2005). The purpose of longitudinal analysis is to capture the within-subject changes in response over time because it is believed that the repeated measurements for the same subject or individual over time are correlated with each other. For binary response data, GLMM which has been increasingly popular in applied epidemiology, is an extension of the generalized linear model (GLM) by the inclusion of random effects in the predictor.

The underlying premise of GLMM is that individuals in the population are assumed to have their own subject-specific mean response trajectories over time and a subset of the regression predictors vary randomly from one individual to another. Thus, the mean response is modeled as a combination of population-averaged effects which is regarded to be shared by all individuals and subject-specific effects that are unique to a particular individual. The population-averaged effects are called fixed-effects, and the subject-specific effects are referred as random effects (Fitzmaurice et al., 2004). In this research, since the response variable is dichotomous, I will use generalized linear mixed logistic model, in which logit link is utilized in the model, namely,

$$g(P_{ij}) = \text{logit}(P_{ij}) = \log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \alpha + \beta * X'_{ij} + b_i + e_{ij} \qquad \text{Equation 3.4}$$

where P_{ij} is the probability of presence for subject i at occasion j

$1- P_{ij}$ is the probability of absence for subject i at occasion j

α is the population-averaged intercept

β is the vector of population-averaged coefficients for predictors

X'_{ij} is the vector of predictor variables for subject i at occasion j

b_i is the subject-specific random effect for subject i

e_{ij} is the measurement or sampling errors for subject i at occasion j

The outcome from the above linear regression is the log odds of presence. One should make the following transformation (Equation 3.5) (Twisk, 2003) to get the final probability of presence P_{ij}.

$$P_{ij} = \frac{1}{1+\exp\left[-\left(\alpha + \beta * X'_{ij} + b_i + e_{ij}\right)\right]} \qquad \text{Equation 3.5}$$

However, common GLMM used in epidemiology assumes that the subjects or individuals are independent of each other, which does not hold in this research because there is spatial dependency between counties. The common GLMM only considers temporal change, but not spatial covariates.

In this study, based on the generalized linear mixed logistic model, I develop an approach that allows the incorporation of spatial dependency by adding a spatial association term (SA). Thus the final model can predict the spatial-temporal distribution for HD in white-tailed deer. Let i denote the county, and j denote the repeated HD observations nested in each county. Therefore, i = 1, …..366, and j = 1, ……19. The proposed spatial-temporal model is constructed as follows:

$$\log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \alpha + \beta * X'_{ij} + \alpha_1 * SA_{ij} + \alpha_2 * T + \gamma * IR' + b_i + e_{ij} \qquad \text{Equation 3.6}$$

Where P_{ij} is the probability of HD occurrence in county i in year j

α is the population-averaged intercept

X'_{ij} is the vector of f1 to f11 for county i in year j

β is the vector of population-averaged coefficients for f1 to f11

SA_{ij} is the spatial dependency for county i in year j

α_1 is the population-averaged coefficients for spatial dependency

T is time in years

α_2 is the population-averaged coefficients for time

IT' is the vector of interactions between predictor variables

γ is the population-averaged coefficient for interactions

b_i is the subject-specific random effect for county i

e_{ij} is the measurement or sampling errors for county i in year j

3.6 Results and discussion

Twenty percent of the county-based data were randomly selected as testing data for validating the model. The remaining 80 percent were used to estimate the model. That is, from the total 6954 observations for 366 counties during the 19 years, 292 counties during the 19 years (5548 observations) were utilized for estimation of the model. Among them, there are 730 HD presence observations and 4818 HD absence observations. In the reserved testing data, the total 1406 observations are composed of 272 HD presence observations and 1134 HD absence

observations. The dependent variable is HD occurrence, and the predictor variables are f1-f9, SA, E, T, and possible interactions between them.

There are 14 predictor variables all together, which can result in numerous alternative models. The goal of this study was to find the optimal prediction model. The information criterion statistic: Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were adopted to select the best model. They choose the best model by compromising model fit and model complexity. Lower AIC and BIC values imply either fewer explanatory variables, better fit, or both. The models with lower AIC and BIC are, therefore, better models than those with higher AIC and BIC.

First, the model was performed using all the predictor variables without interaction (model1). The output is shown in Table 3.8.

The Wald test in Table 3.8 is a generalized Wald test for all predictor variables and follows a χ^2 distribution with 14 degrees of freedom. In other words, the Wald test is to evaluate the importance of all the regression coefficients. Its ρ-value (Prob > chi2 = 0.0000) indicates that Wald test in model1 is highly significant. By closely examining the output, I can conclude that variables f1, f5, f6, f11 and elevation are not significant at $\rho = 0.05$, and should be dropped from the model. The standard deviation of the random effect is denoted as sigma_u. I can calculate the square of sigma_u to obtain the variance of the random effect which is 0.7874 in model1. The underlying idea is that the overall unexplained variance is divided into two parts, one is the variance of the random effect, and the other is related to the remaining 'errors' such as measurement or sampling error. The rho value is an estimation of the intraclass correlation

coefficient (ICC) which is computed as the variance of the random effect divided by the total unexplained variance. So rho is an indication of within-subject dependency. In model1, the variance of the random effect accounts for 19 percent of the total remaining variance. The Likelihood-ratio test of rho provides information on the importance of allowing the random effect. The difference between this model and a similar model without a random effect is 171.46, which follows a χ^2 distribution with one degree of freedom, and which is highly significant (Prob >= chibar2 = 0.000). Thus, it is necessary to include a random effect in this particular model.

Table 3.8: Output of model1

			Wald chi2(14) =336.38	
Log likelihood = -1886.4728			Prob > chi2 = 0.0000	
hd	Coef.	Std. Err.	z	P>z
f1	0.0683802	0.061432	1.11	0.266
f2	-0.176547	0.0829118	-2.13	0.033
f3	-0.2267664	0.0554919	-4.09	0
f4	0.2426384	0.066439	3.65	0
f5	-0.0564875	0.0578344	-0.98	0.329
f6	0.0896043	0.0514959	1.74	0.082
f7	-0.244984	0.038012	-6.44	0
f8	0.4420369	0.0408429	10.82	0
f9	0.3944866	0.0639763	6.17	0
f10	-0.1815679	0.0655136	-2.77	0.006
f11	0.1081336	0.0631405	1.71	0.087
SA	1.403131	0.1779767	7.88	0
dem	0.00028	0.0005537	0.51	0.613
time	0.089485	0.0158888	5.63	0
Intercept	-3.542563	0.2328271	-15.22	0
/lnsig2u	-0.2389647	0.1544118		
sigma_u	0.8873797	0.0685109		
Likelihood-ratio test of rho=0: chibar2(01) =171.46 Prob >= chibar2 = 0.0000				

The t-test in model1 indicates that some insignificant predictor variables should be excluded. Theoretically, if the predictor variables are independent of each other, the drop of one or more variables will not affect the significance of another variable. However, in this research, although the value is relative low, correlations do exist between some predictor variables. The study proceeded by dropping the insignificant variables one at a time, each time dropping the most insignificant variable from the previous model, until all predictor variables are significant. The final model includes f3, f4, f7, f8, f9, f10, SA, and T, the same as the significant variables in model1. This proves that the low correlations in the data have little influence on the significance of predictor variables. Note that the two variables with their correlation greater than 0.5 (f5 and T) do not appear together in model2, which saves the efforts of identifying the true source of the two variables in interpreting the model. The output of model2 is shown in Table 3.9.

Table 3.9: Output of model2

Log likelihood = -1891.3665			Prob > chi2 = 0.0000	
hd	Coef.	Std. Err.	z	P>z
f3	-0.2550442	0.0536297	-4.76	0
f4	0.238649	0.0650699	3.67	0
f7	-0.2317267	0.0368216	-6.29	0
f8	0.4268687	0.0385337	11.08	0
f9	0.3706288	0.0609242	6.08	0
f10	-0.1376361	0.0553749	-2.49	0.013
SA	1.45695	0.1774919	8.21	0
time	0.0656585	0.0093733	7	0
intercept	-3.25705	0.1372925	-23.72	0
/lnsig2u	-0.2433885	0.1542304		
sigma_u	0.885419	0.0682793		
rho	0.1924395	0.0239684		
Likelihood-ratio test of rho=0: chibar2(01) =171.46 Prob >= chibar2 = 0.0000				

Next, various combinations of interactions between predictor variables were added to model2 to see whether the interactions can contribute to the dependent variable. The AIC and BIC for some of the models with significant interaction terms as well as the AIC and BIC for model1 and model2 are shown in Table 3.10.

Table 3.10: AIC and BIC for selected models

Model	**Interaction term**	**Degree of freedom**	**AIC**	**BIC**
model1	None	16	3804.95	3910.88
modelf11sw	None	12	3794.41	3873.87
modelf3f10	None	11	3800.91	3873.74
mode_f3f8	None	11	3800.82	3873.65
modelf7sw	None	11	3800.74	3873.57
modelf3ft	None	11	3800.06	3872.89
modelf9f10	None	11	3799.89	3872.72
modelf8sw	None	11	3797.20	3870.04
model2	None	10	3802.73	3868.94
modelf7t	None	11	3793.56	3866.39
model_f4T	f4*T	11	3792.83	3865.66
model_f9SA	f9*SA	11	3792.80	3865.63
model_f8T_f9SA_f3T	f8*T, f9*SA, f3*T	13	3778.22	3864.29
modelf8T_f7T_f4T	f8*T, f7*T, f4*T	13	3774.51	3860.58
model_f8T	f8*T	11	3787.27	3860.11
model_f8T_f8SA	f8*T, f8*SA	12	3780.41	3859.86
model_f4T_f9SA_f4T	f4*T, f9*SA, f4*T	13	3772.87	3858.94
model_f8T_f9SA	f8*T,f9*SA	12	3777.58	3857.03
modelf8T_f7T	f8*T, f7*T	12	3776.90	3856.35
model_f8T_f7T_f9SA	f8*T, f7*T,f9*SA	13	3766.60	3852.67

It turns out that AIC and BIC generally decrease with the addition of interaction terms. The full model with all predictor variables (model1) has the highest AIC and BIC. However, there are some minor inconsistencies between AIC and BIC. In this research, BIC was chosen as the primary criterion because BIC is preferred by many applications because it is a Bayesian procedure (Burnham and Anderson, 2004). Furthermore, it imposes a greater penalty for additional parameters than does AIC. The model with the smallest BIC is the model with three interaction terms: f8*T, f7*T, and f 9*SA (model_f8T_f7T_f9SA). See Table 3.11 for the output of model_f8T_f7T_f9SA.

Table 3.11: Output of model_f8T_f7T_f9SA

			Wald chi2(11) = 358.28	
Log likelihood = -1870.2998			Prob > chi2 = 0.0000	
hd	Coef	Std. Err	z	P>z
f3	-0.13	0.06	-2.31	0.02
f4	0.22	0.07	3.33	0.00
f7	-0.92	0.18	-5.07	0.00
f8	1.20	0.17	6.92	0.00
f9	0.29	0.07	3.89	0.00
f10	-0.23	0.06	-3.84	0.00
SA	1.22	0.19	6.54	0.00
T	0.05	0.01	4.62	0.00
f8t	-0.10	0.02	-4.33	0.00
f7t	0.05	0.01	3.63	0.00
f9SA	0.73	0.21	3.39	0.00
intercept	-3.15	0.14	-22.36	0.00
/lnsig2u	-0.23	0.15		
sigma_u	0.89	0.07		
rho	0.20	0.02		
Likelihood-ratio test of rho=0: chibar2(01) =174.2 Prob >= chibar2 = 0.0000				

The likelihood-ratio test of rho is significant for all the three models: model1, model2 and model_f8T_f7T_f9SA, which means the random effect should be considered in all three models. The Wald tests are also significant for each model, thus all the coefficients are significant as a whole to each model. By further examination, I see that the difference of AIC and BIC between model2 and model_f8T_f7T_f9SA is only 36.13 and 16.27, respectively, but the latter model is much more complicated with three interaction terms between continuous variables. Interactions between continuous variables are rarely analyzed in practical analysis, as Jaccard and Turrisi (2003), and Aiken and West (1991) argue that there is little work done to include interactions between continuous variables in the model because of the difficulty of interpretation. To determine which of the two models should be chosen as the optimal model, testing data were applied to validate the models. If one model predicts the presence and absence of HD more successfully than the other, it should be chosen as the optimal model.

Since the result of the model is the probability of HD presence for a particular county in a particular year, before validation, a cut-off probability must be determined where the probabilities above this value indicate presence and probabilities below the value indicate absence (Hinely, 2006). With the training data, a graph of cut-off probability vs. the percentage correctly predicted for both HD presence and absence is produced. See Figure 3.5 and Figure 3.6 for the graphs of model2 and model_f8T_f7T_f9SA, respectively.

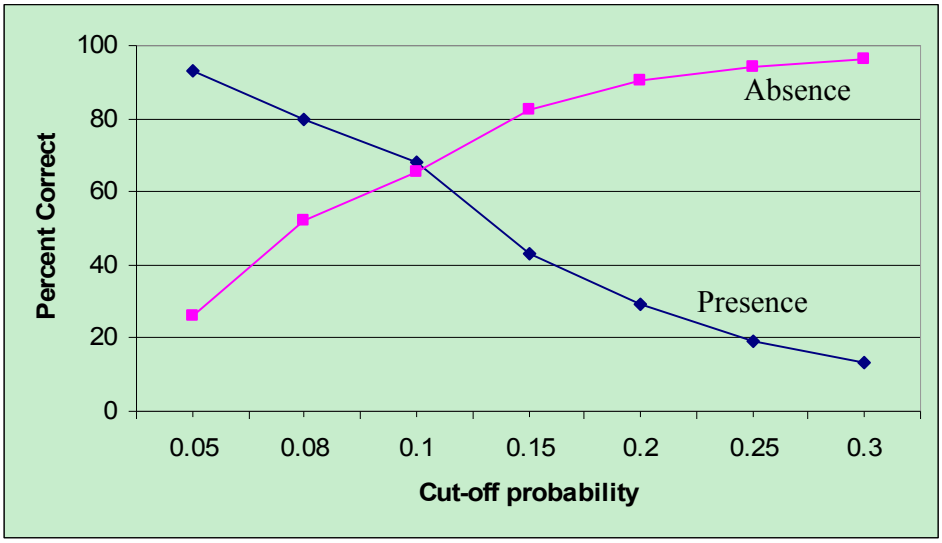


Figure 3.5: Cut-off analysis for model2

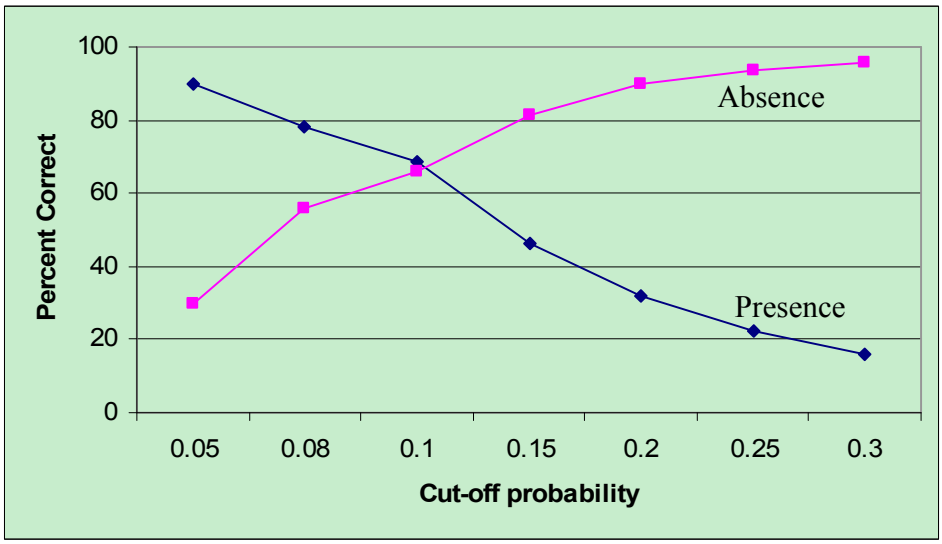


Figure 3.6: Cut-off analysis for model_f8T_f7T_f9SA

Both graphs show a decreasing tendency of the prediction accuracy for HD presence and an increasing trend of the prediction accuracy for HD absence. The place where two lines cross is the cut-off point that maximizes both the presence accuracy and absence accuracy. The cut-off

points for both models are approximately the same, which is at the probability of 0.1. This cut-off probability is then used for testing data in both models. The prediction results are shown in Table 3.12 and Table 3.13.

Table 3.12: Prediction results for model2

		Predicted			
Observed		Presence	Absence	Total	Accuracy
	Presence	174	98	272	0.64
	Absence	395	739	1134	0.65
	Total	569	837	1406	0.65

Table 3.13: Prediction results for model_f8T_f7T_f9SA

		Predicted			
Observed		Presence	Absence	Total	Accuracy
	Presence	165	107	272	0.61
	Absence	387	747	1134	0.66
	Total	552	854	1406	0.65

For the total 272 observed HD presence observations in the testing data, 172 observations are correctly predicted in model2, and 165 observations are correctly predicted in model_f8T_f7T_f9SA. The predicted presence accuracy is 0.64 and 0.61 for model2 and model_f8T_f7T_f9SA, respectively. For the total 1134 observed HD absence observations, 739 observations are correctly predicted by model2, and 854 observations are correctly predicted by model_f8T_f7T_f9SA, which results in predicted absence accuracy of 0.65 and 0.66,

respectively. The overall accuracy for both models is exactly the same. The predicted presence accuracy in model2 is 3 percent greater than that in model_f8T_f7T_f9SA, whereas the predicted absence accuracy in model_f8T_f7T_f9SA is only 1 percent higher than that in model2. It can be concluded that model2 is better than model_f8T_f7T_f9SA. Furthermore, as I have discussed, model2 is much simpler than model_f8T_f7T_f9SA, and is easier to interpret, model2 was chosen as the optimal model for this study. Figure 3.7 to Figure 3.25 give map comparisons of observed and predicted HD presence and absence across the study area from 1982 to 2000 using model2. The prediction accuracy for individual year ranges from 27 percent to 96 percent. The years with over-predictions of HD presence are 1982, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1994, 1996, 1997, 1998, 1999, and 2000. Among them, the years of 1989, 1994, 1997, 1998, 1999, and 2000 have considerable over prediction of HD presence (2 times higher than observed HD presence). The predicted HD presence shows a strong spatial contiguity over the study area which is the result of the strong relationship between spatial dependency and response variable in model2.

The output of model2 shows that f3, f7 and f10 are negatively related to the log odds of HD occurrence, while f4, f8, f9, time and SA are positively related to the log odds of HD occurrence. Since the logit form in model2 made it difficult to interpret the regression coefficients directly, the coefficients are first transformed into odds ratios (exp(β)). The odds ratios (exp(β)) always range from zero to infinity, with 1 being the center. For a negative coefficient, its odds ratio varies between 0 and 1. For a positive coefficient, its odds ratio is from 1 to infinity. This causes confusion when interpreting the negative and positive coefficients in the

usual way as ordinary regressions. Thus, for those coefficients with negative values (i.e., coefficients for f3, f7, f10, and the intercept), I further divide their odds ratios with 1 to obtain their reciprocals. Table 3.14 gives the transformation results.

Table 3.14: Model coefficient transformation

Variable	Coefficient (β)	Odds Ratio (exp (β)	1/Odds Ratio
f3	-0.255	0.7749	1.2905
f4	0.2386	1.2695	NA
f7	-0.2317	0.7932	1.2607
f8	0.4269	1.5325	NA
f9	0.3706	1.4486	NA
f10	-0.1376	0.8714	1.1476
SA	1.457	4.2928	NA
time	0.0657	1.0679	NA
intercept	-3.2571	0.0385	25.9740

The final interpretation of the model is as follows: with regards to those variables with positive coefficient (f4, f8, f9, SA, and time), the coefficient odds ratio for a particular variable is interpreted by assuming that if all other variables remain unchanged, a one-unit increase of this variable will cause the odds of the dependent variable being in the higher group by the corresponding odds ratio. As to those variables with negative coefficient (f3, f7, f10), it was interpreted by assuming that if all other variables remain constant, a one-unit decrease of a particular variable will cause the odds of dependent variable being in the higher group by the reciprocal of the corresponding odds ratio. From Table 3.14, it was found that the odds ratio of spatial dependency (SA) is the highest among all the variables with positive coefficients. With one-unit increase of spatial dependency, a county has a 4.3 times higher odds of being present of HD, which shows a strong spatial effect on HD occurrence and distribution. F3 has the highest

reciprocal of odds ratio among the variables negatively related to HD, with one-unit decrease of f3 where a county has a 1.3 times higher odds of being present of HD.

With reference to Table 3.6, it can be concluded that wind speed, rainfall, land surface temperature and NDVI were the most important factors for the prediction of HD occurrence. Here it should be noted that the climatic temperature is not included in the prediction model. Instead, land surface temperature calculated from AVHRR channel 4 and channel 5 is incorporated, which reinforces the arguments that remotely-sensed data have higher prediction power than climatic data in some occasions.

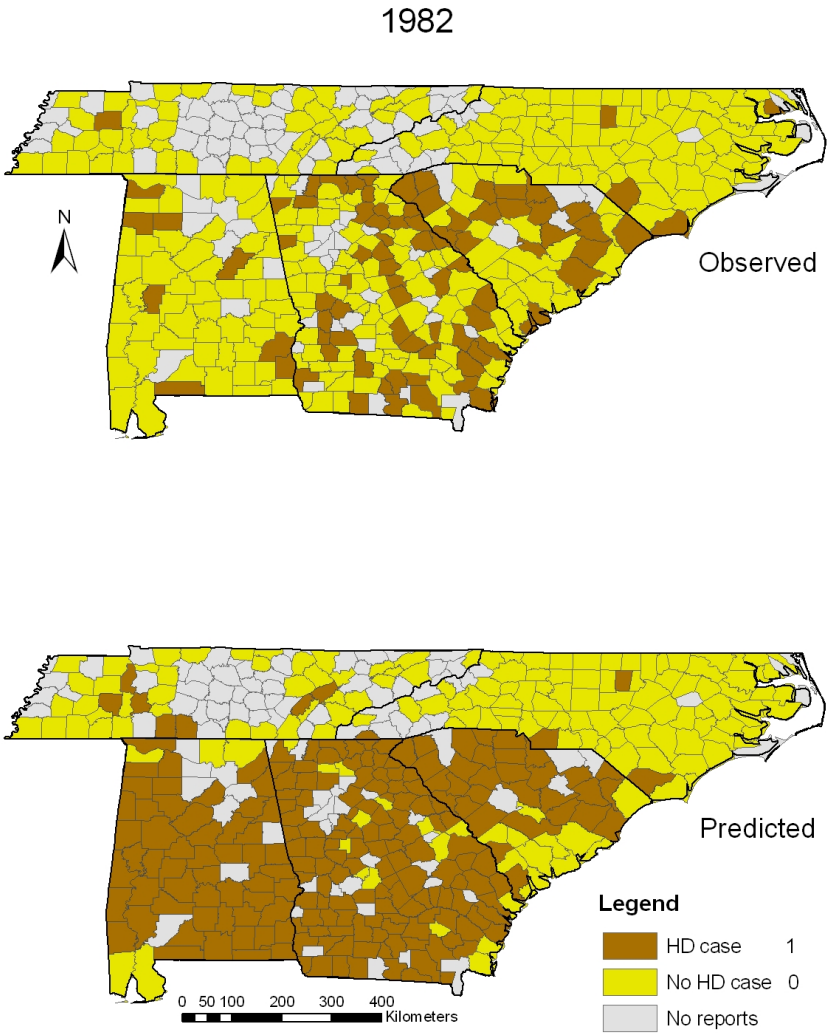
3.7 Conclusion

In this study, a spatial-temporal statistical model is proposed to predict the HD occurrence of white-tailed deer in southeast USA on the basis of individual counties using climatic data and remotely-sensed data from 1982 to 2000. The model is based on generalized linear mixed logistic model which can deal with longitudinal data taking the within-subject effect into account. The spatial dependency is considered by incorporating a spatial association term that indicates the influence of surrounding counties on a particular county. Principal component factor analysis is first applied to reduce data volume and remove correlations between predictor variables. The results show that wind speed, rainfall, land surface temperature, and NDVI are valuable explanatory factors to predict HD occurrence, with total prediction accuracy of 65 percent. Remotely-sensed data prove to be useful and give a higher prediction power than some ground-based climatic data.

This conclusion is consistent with previous studies that humidity and temperature affect

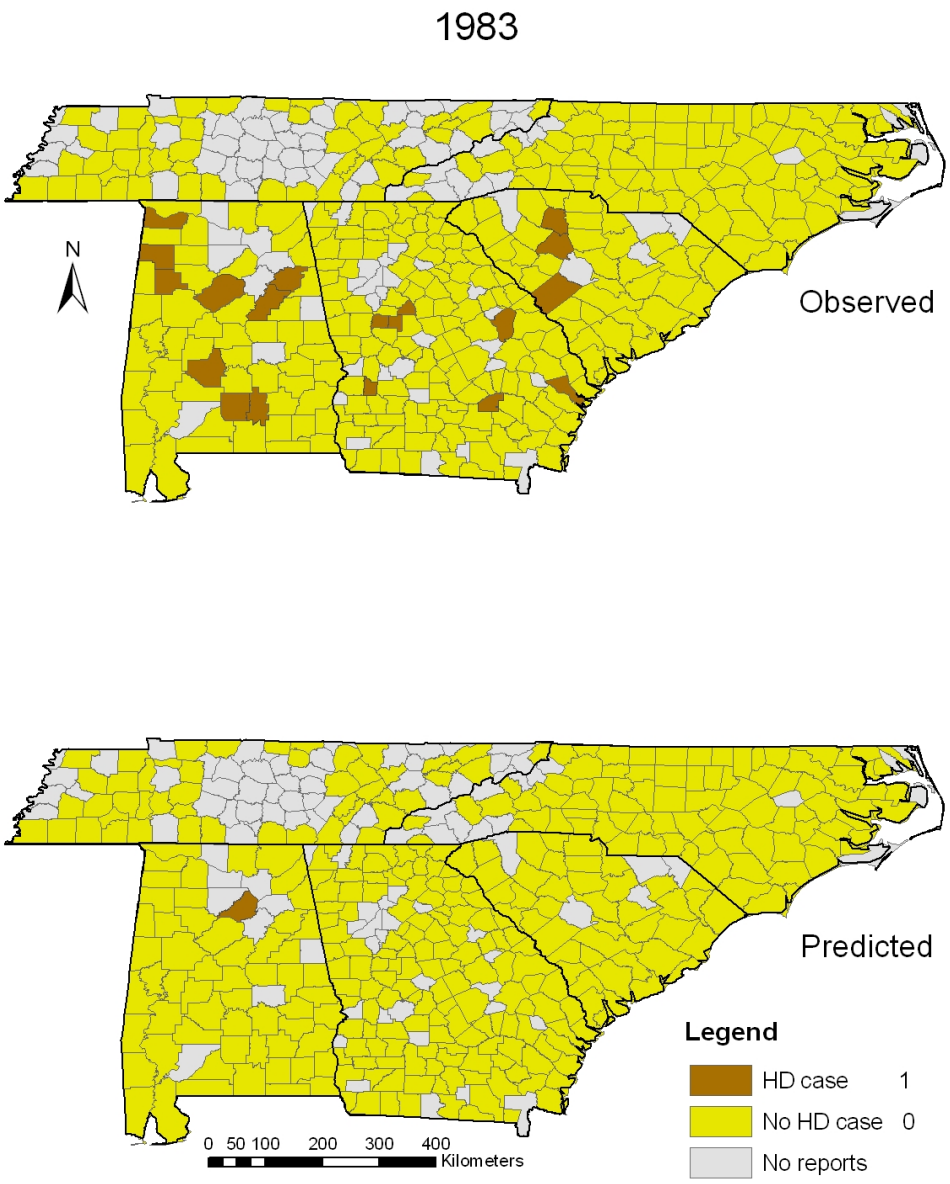
the epizootiology of HD (Erasmus, 1975). Temperature greatly influences the geographic ranges of *Culicoides*. Low temperature determines the distribution of the insets, while high temperature adversely affect *Culicoides* adult survivorship and size (Wittmann and Baylis, 2000). Rainfall strongly influences the increase of *Culicoides* numbers in spring and summer (Nevill, 1971). Walker and Davies (1971) postulated that there is a causal relationship between peak rainfall in April-May, peak numbers of *Culicoides* in May-June and peak bluetongue incidence in June-July. Wind speed can also affect *Culicoides* distribution through their influence on the passive dispersal of the adults. For example, *Culicoides* can be carried as aerial plankton to a place up to 700 km away in winds at speeds of 10-40 km/h, and at heights up to 1.5 km (Wittmann and Baylis, 2000). NDVI represents the vegetation biomass on the ground which is indicative of *Culicoides* habitat. All these factors are related to the distribution and occurrence of HD through their influence on the vectors.

Given the overall prediction accuracy of 65 percent, there should be other factors that contribute to the HD occurrence, but are not included in the model. Furthermore, in the statistics field, there are some debates on whether it is appropriate to first apply principal component analysis, and then use the resulted principal factors as later input. However, despite its difficulty in interpreting the principal factors, the principal component factor analysis is an ideal method to reduce data volume and remove correlations. It can substantially simplify the analysis without greatly decreasing the model's prediction power.



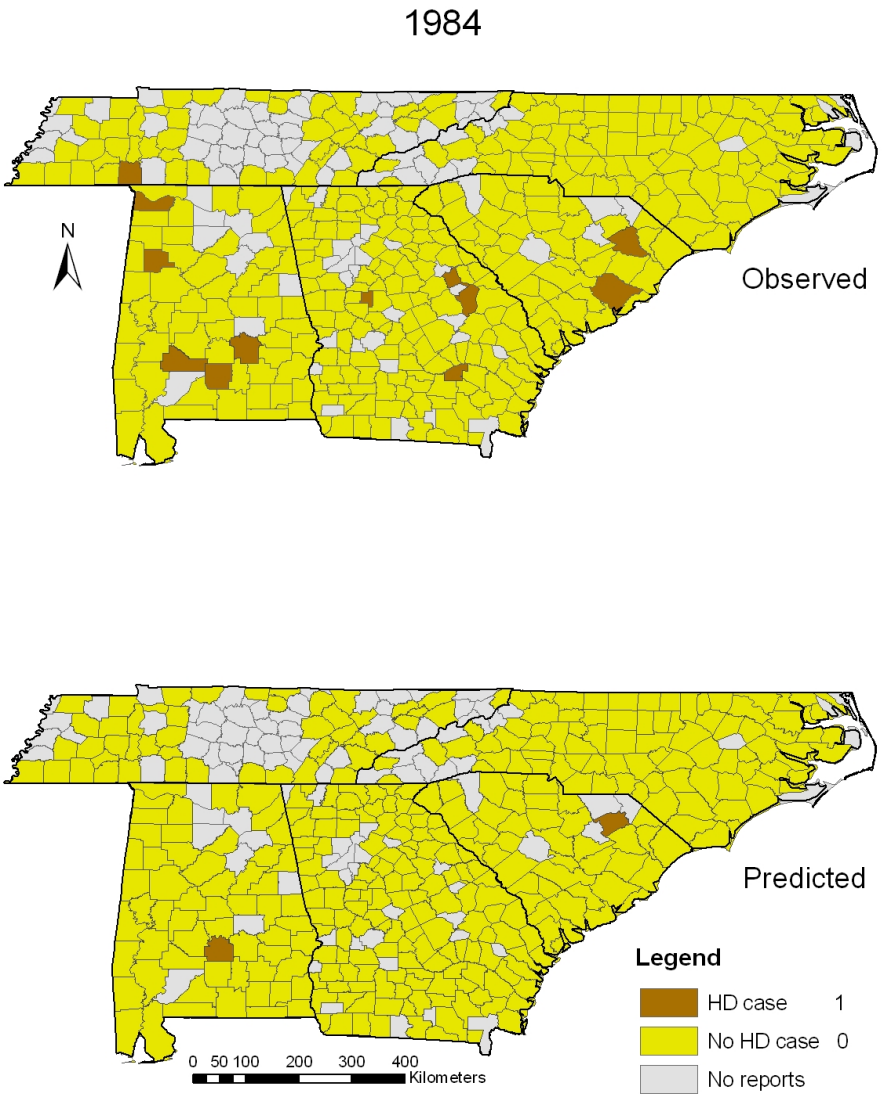
1982	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	68	16	84	0.81
	Absence	146	136	282	0.48
	Total	214	152	366	0.56

Figure 3.7: Observed and predicted HD cases (1982)



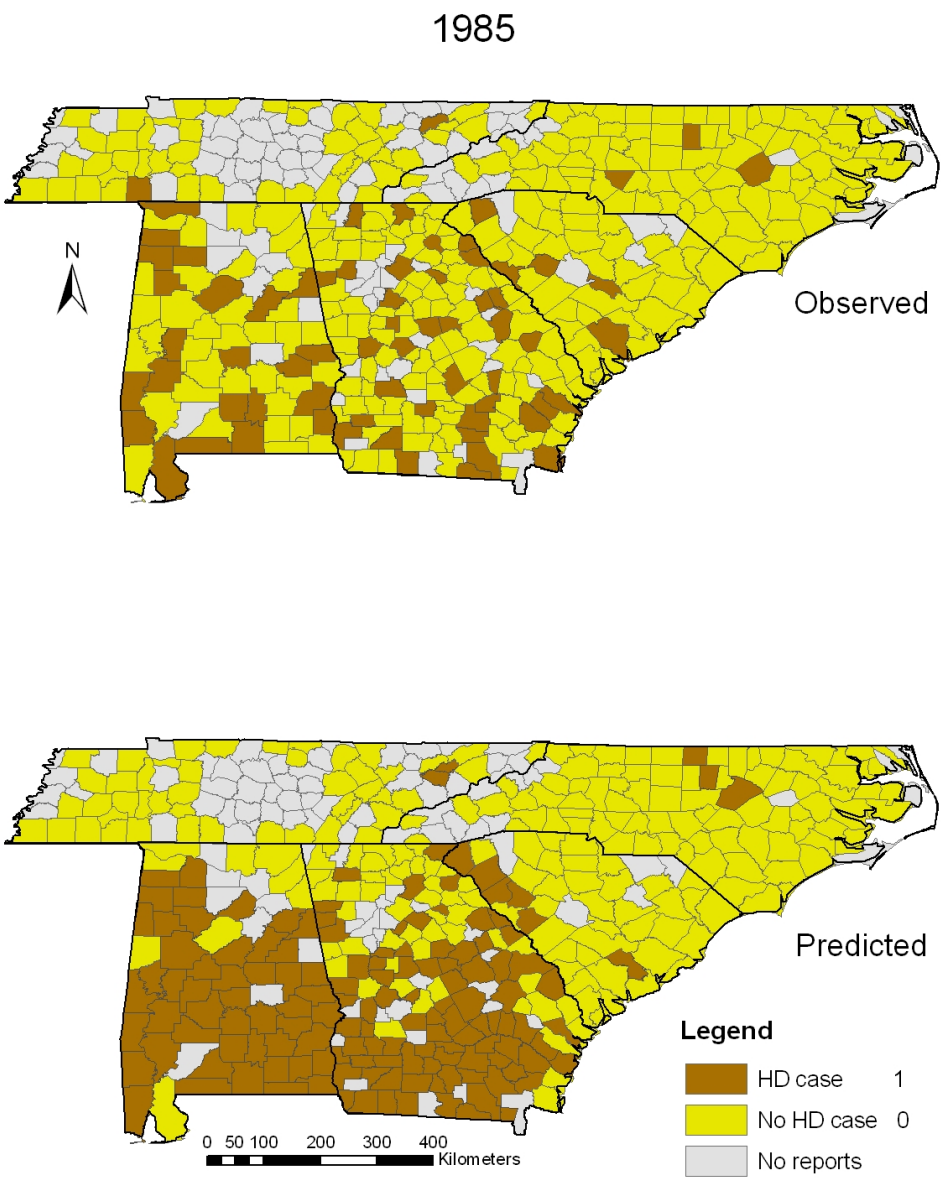
1983	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	0	19	19	0.00
	Absence	1	346	347	1.00
	Total	1	365	366	0.95

Figure 3.8: Observed and predicted HD cases (1983)



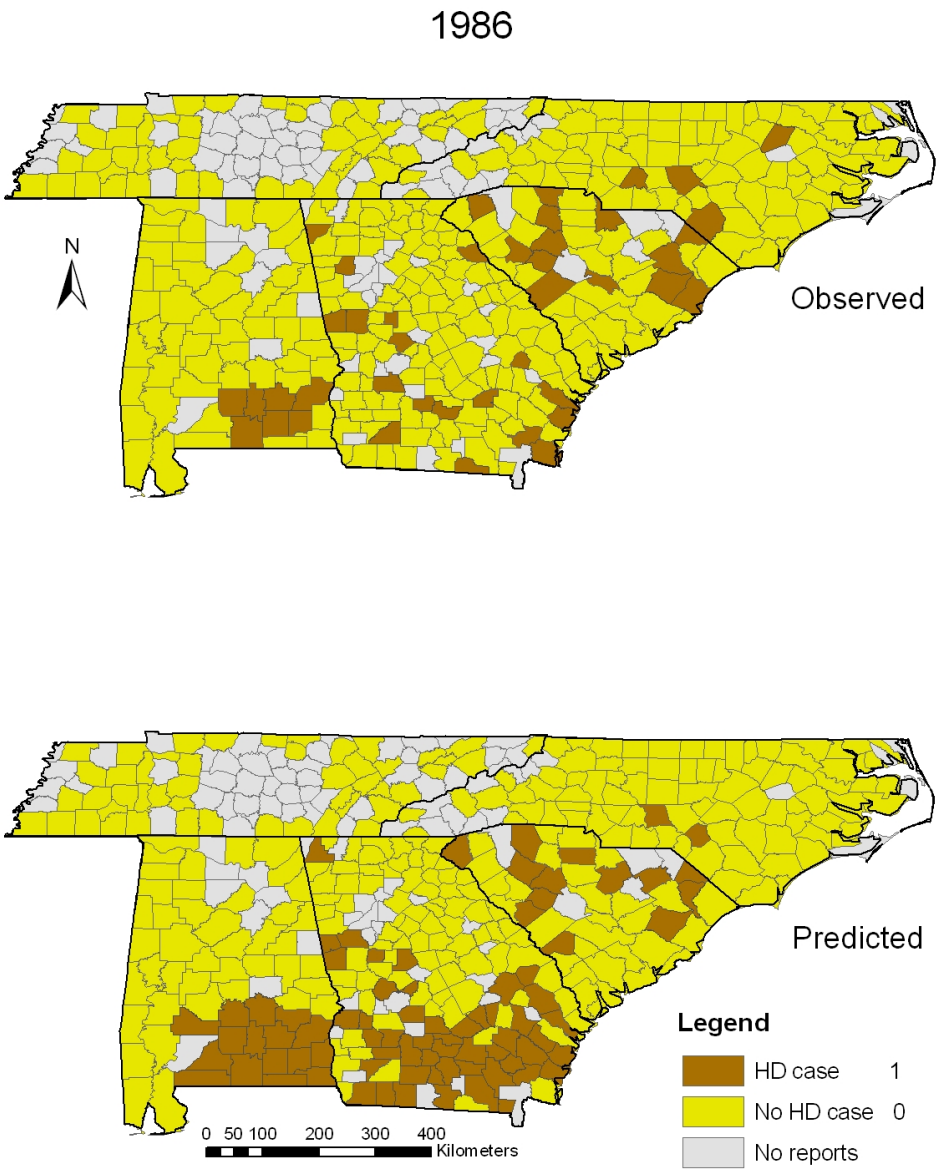
1984	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	0	12	12	0.00
	Absence	2	352	354	0.99
	Total	2	364	366	0.96

Figure 3.9: Observed and predicted HD cases (1984)



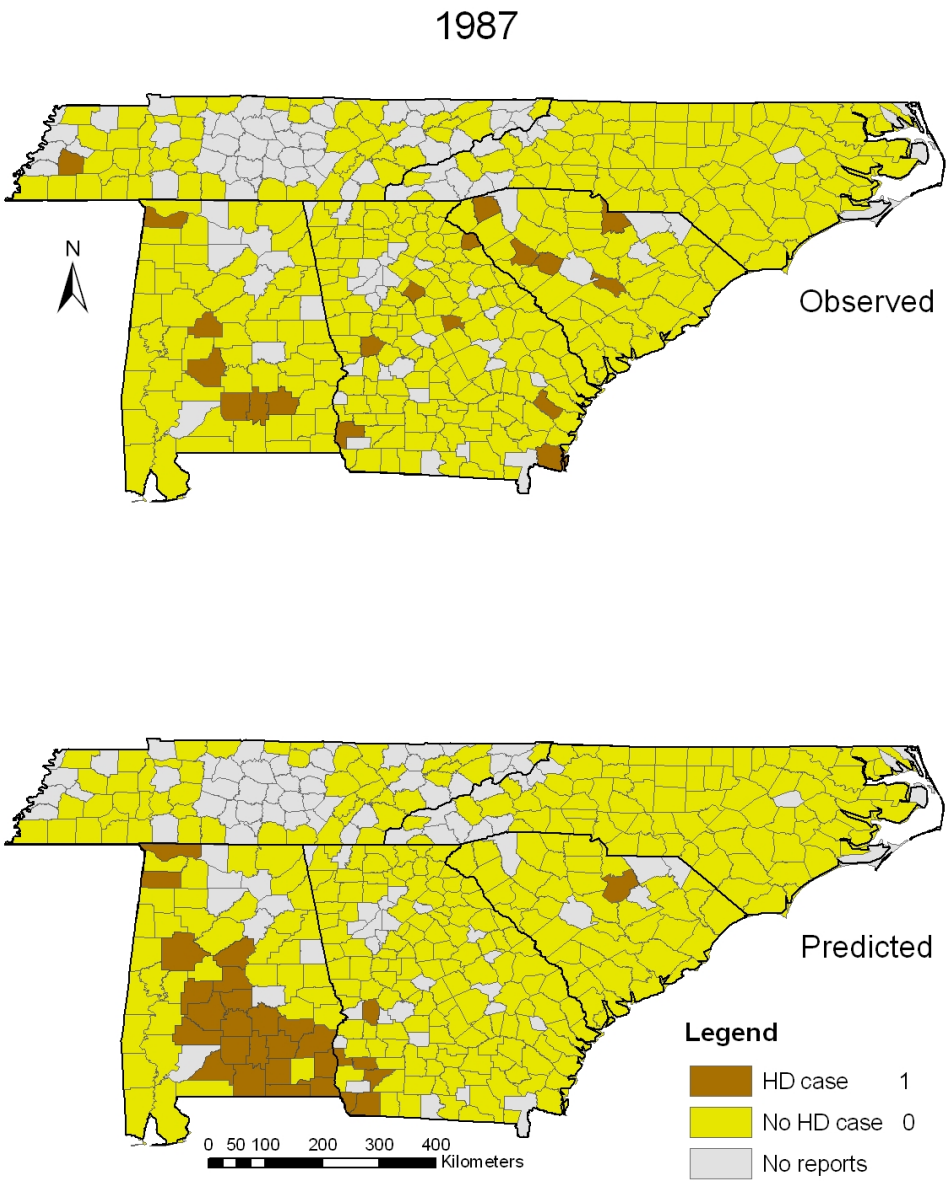
1985	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	38	30	68	0.56
	Absence	106	192	298	0.64
	Total	144	222	366	0.63

Figure 3.10: Observed and predicted HD cases (1985)



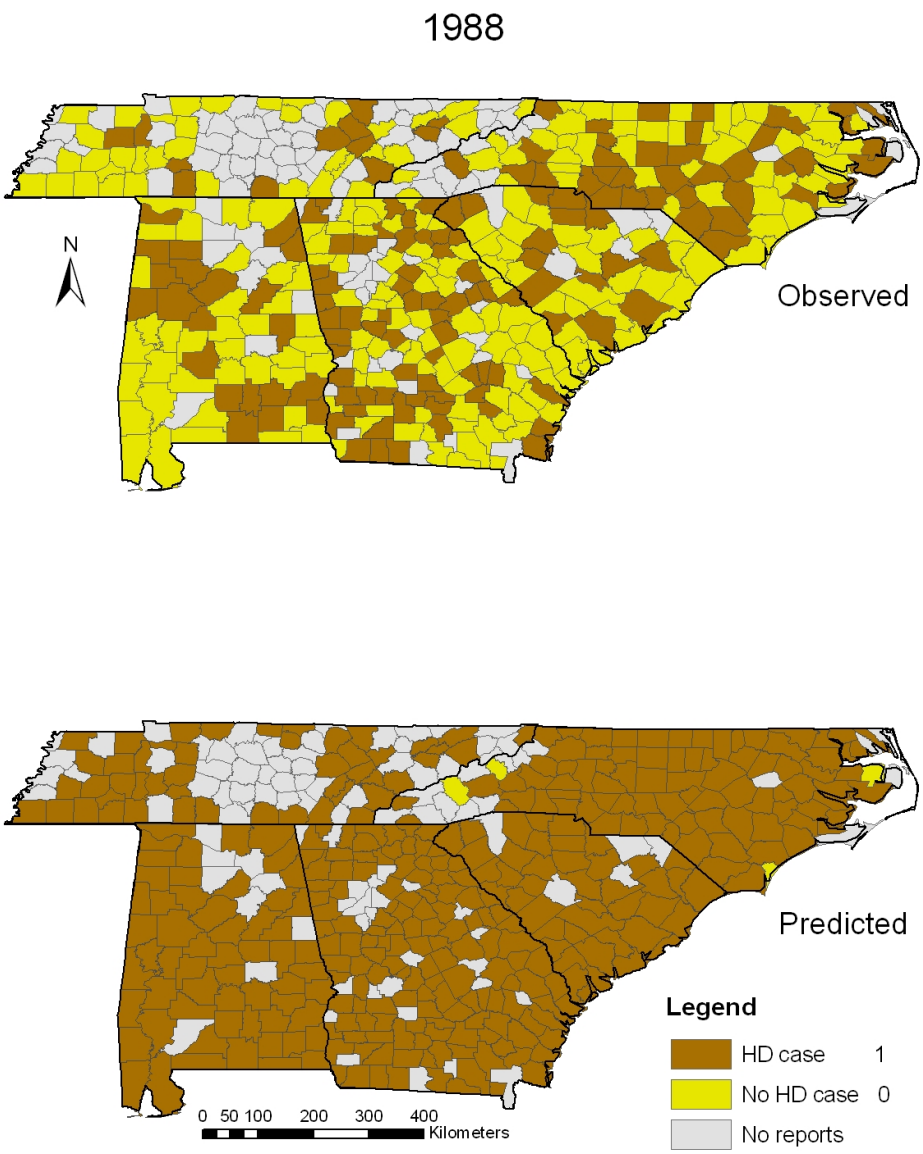
1986	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	16	26	42	0.38
	Absence	63	261	324	0.81
	Total	79	287	366	0.76

Figure 3.11: Observed and predicted HD cases (1986)



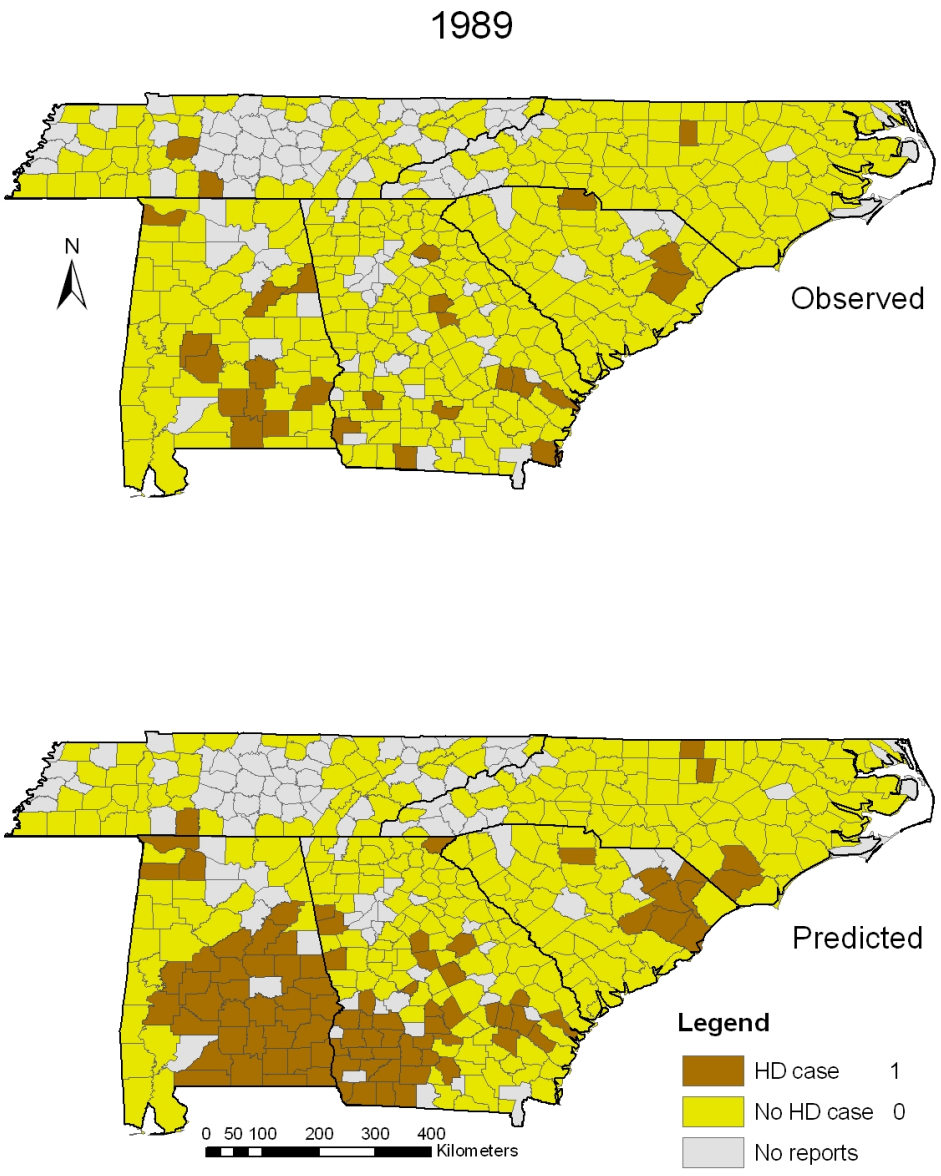
1987	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	4	15	19	0.21
	Absence	25	322	347	0.93
	Total	29	337	366	0.89

Figure 3.12: Observed and predicted HD cases (1987)



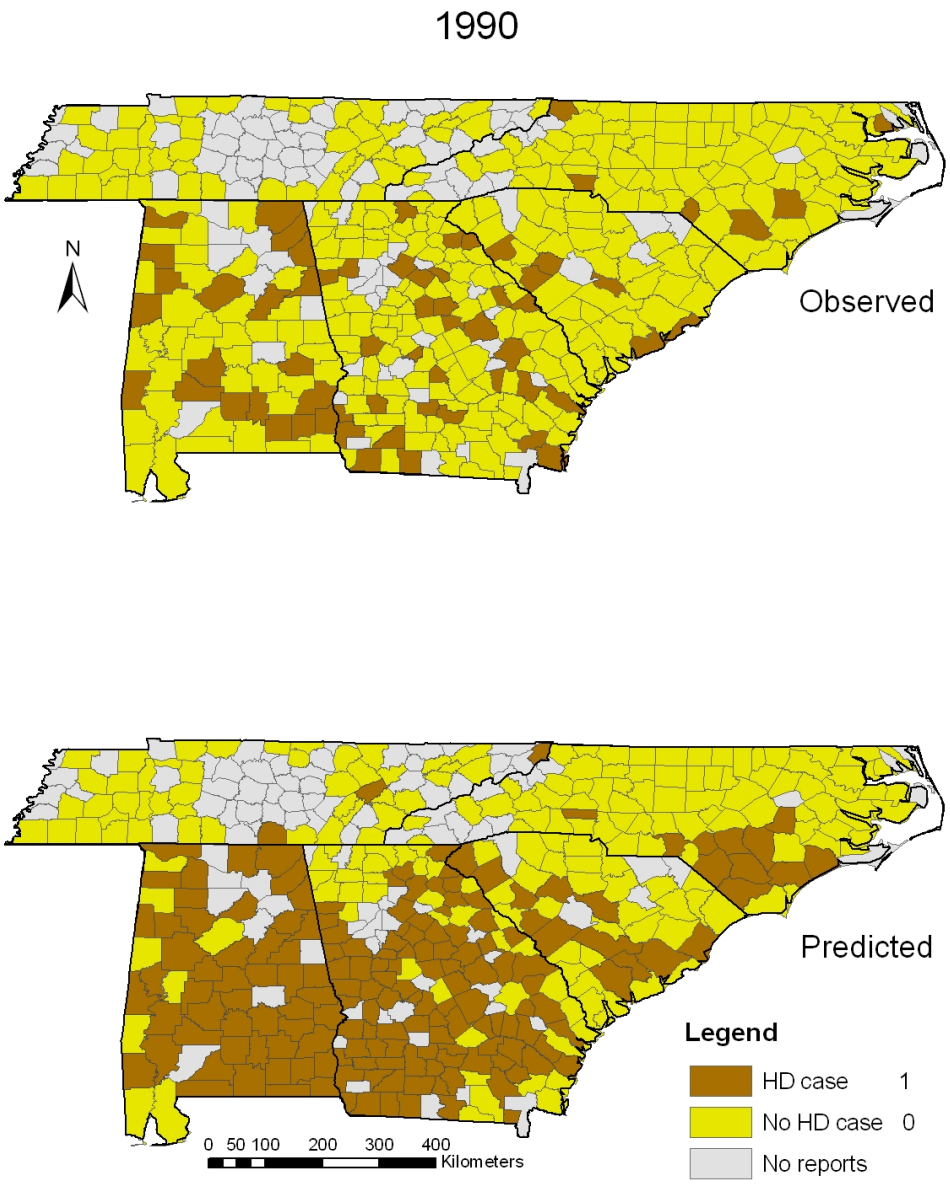
1988	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	144	2	146	0.99
	Absence	218	2	220	0.01
	Total	362	4	366	0.40

Figure 3.13: Observed and predicted HD cases (1988)



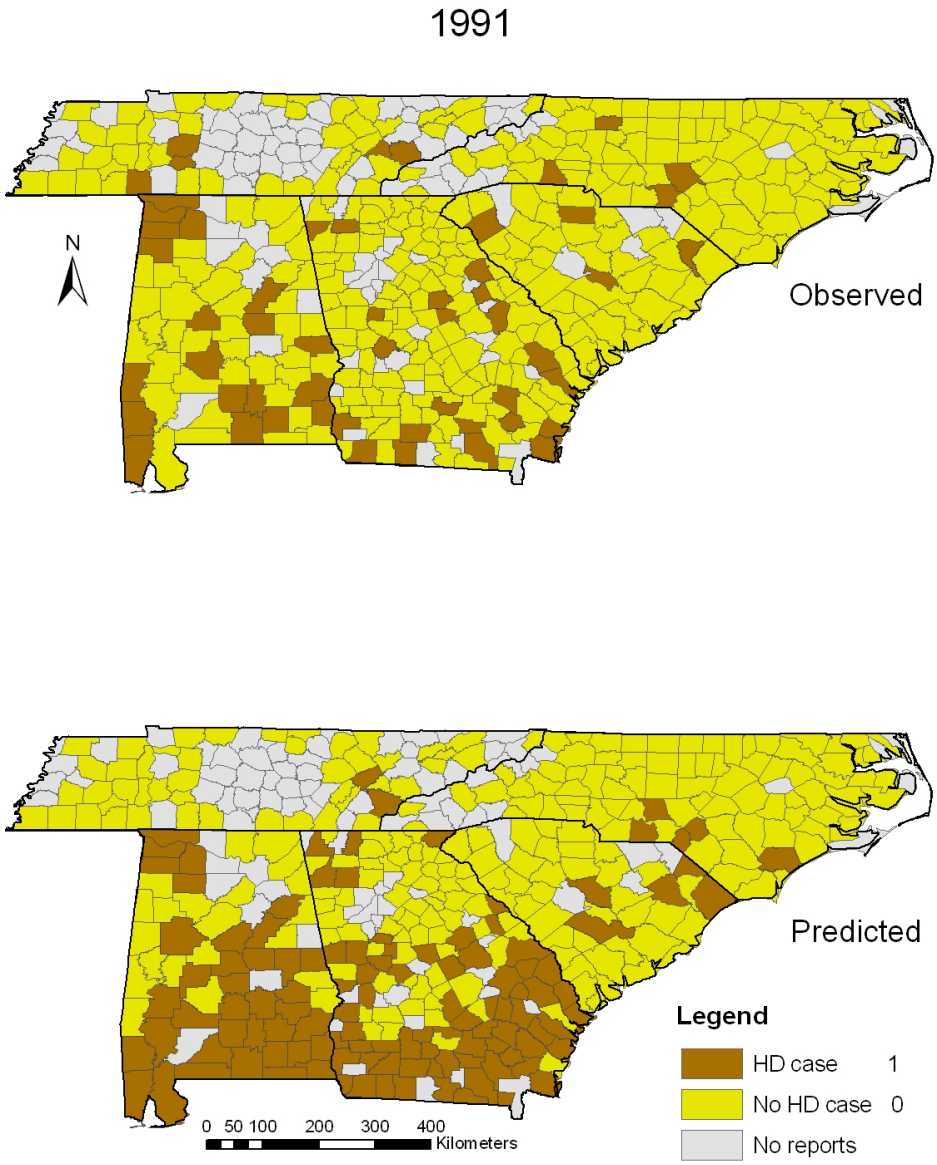
1989	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	16	12	28	0.57
	Absence	75	263	338	0.78
	Total	91	275	366	0.76

Figure 3.14: Observed and predicted HD cases (1989)



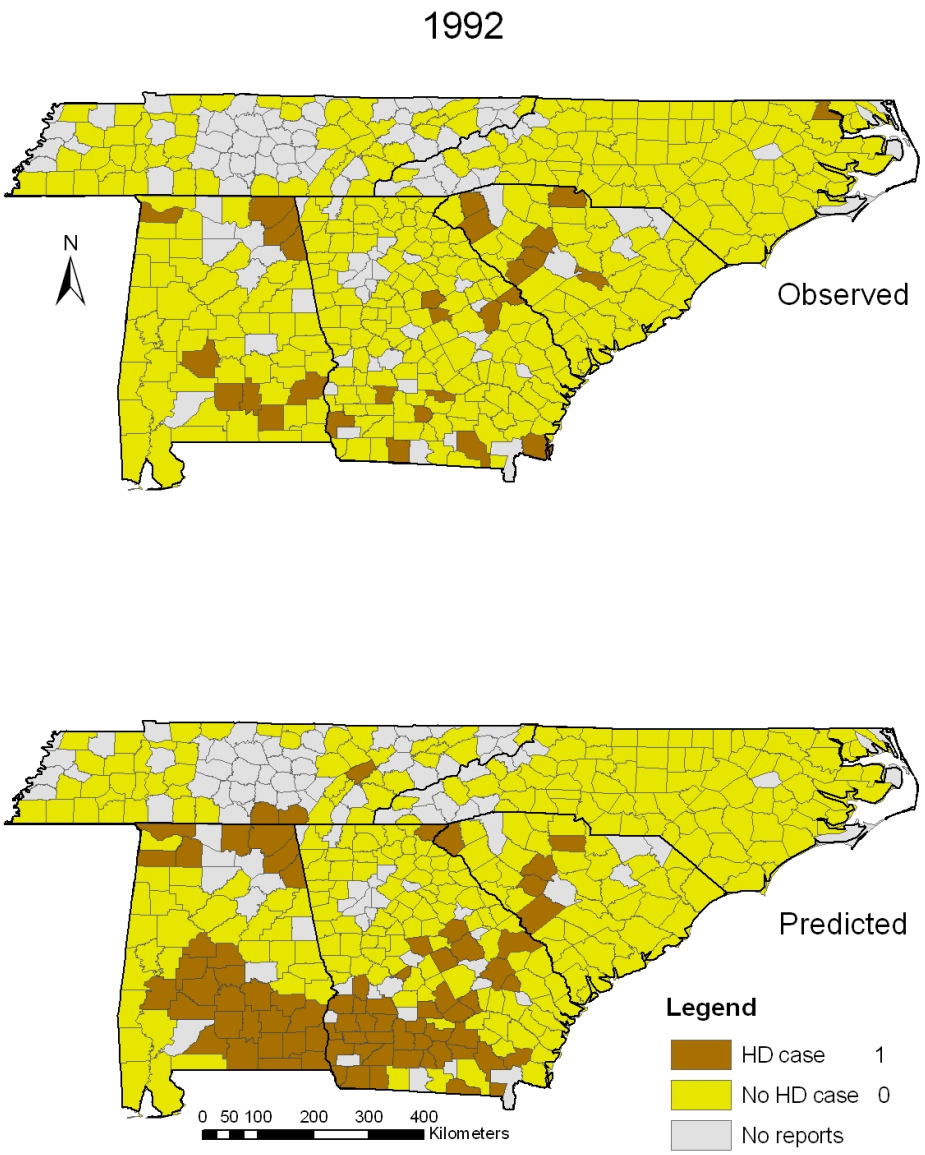
1990	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	38	24	62	0.61
	Absence	135	169	304	0.56
	Total	173	193	366	0.57

Figure 3.15: Observed and predicted HD cases (1990)



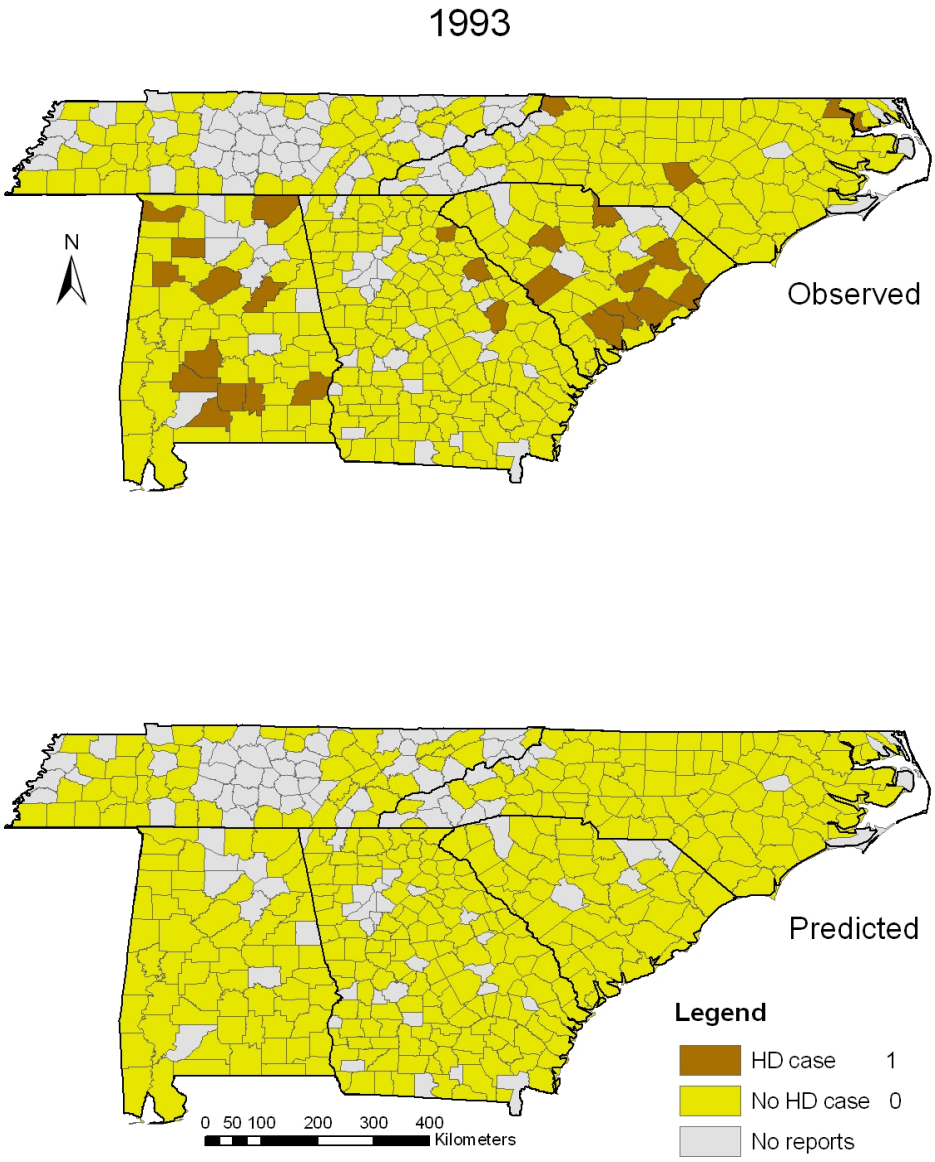
1991	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	24	30	54	0.44
	Absence	93	219	312	0.70
	Total	117	249	366	0.66

Figure 3.16: Observed and predicted HD cases (1991)



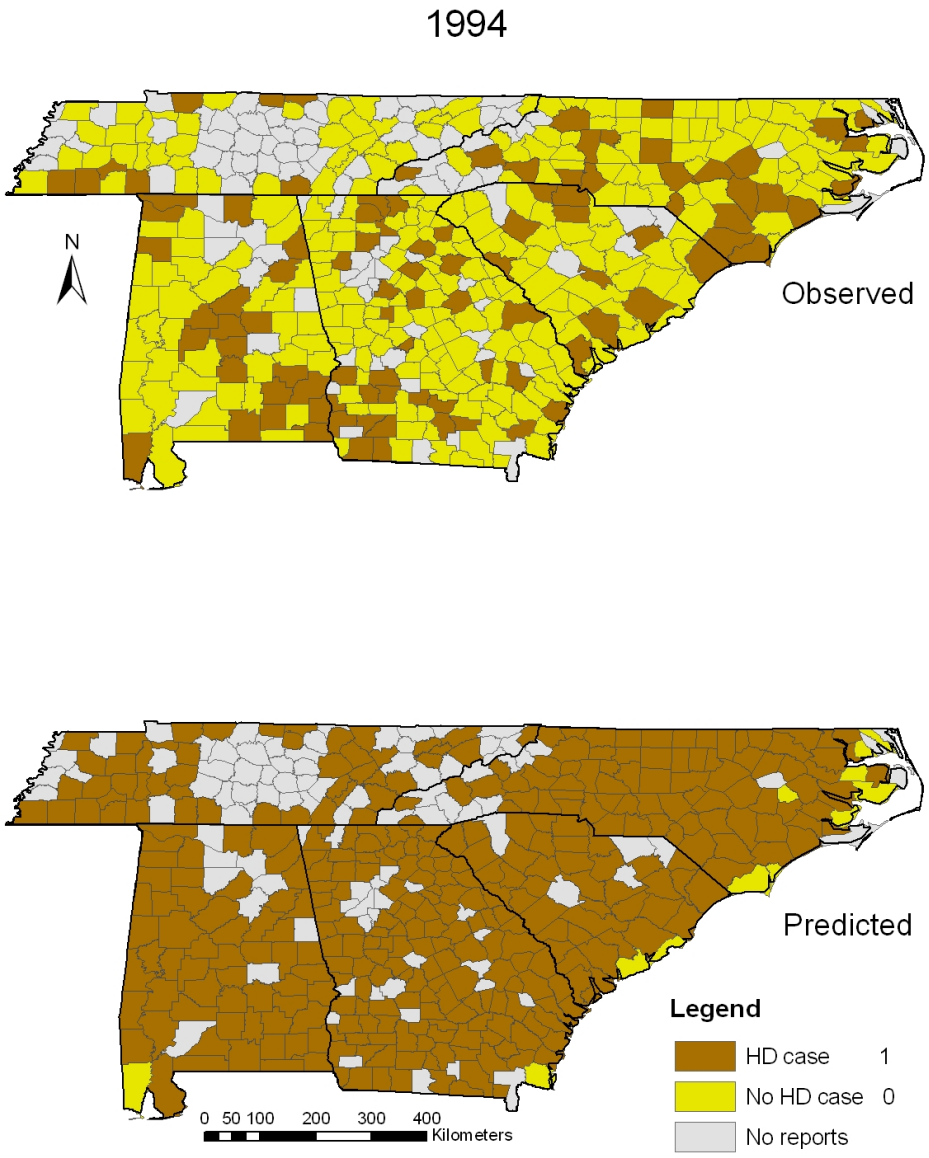
1992	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	13	15	28	0.46
	Absence	64	274	338	0.81
	Total	77	289	366	0.78

Figure 3.17: Observed and predicted HD cases (1992)



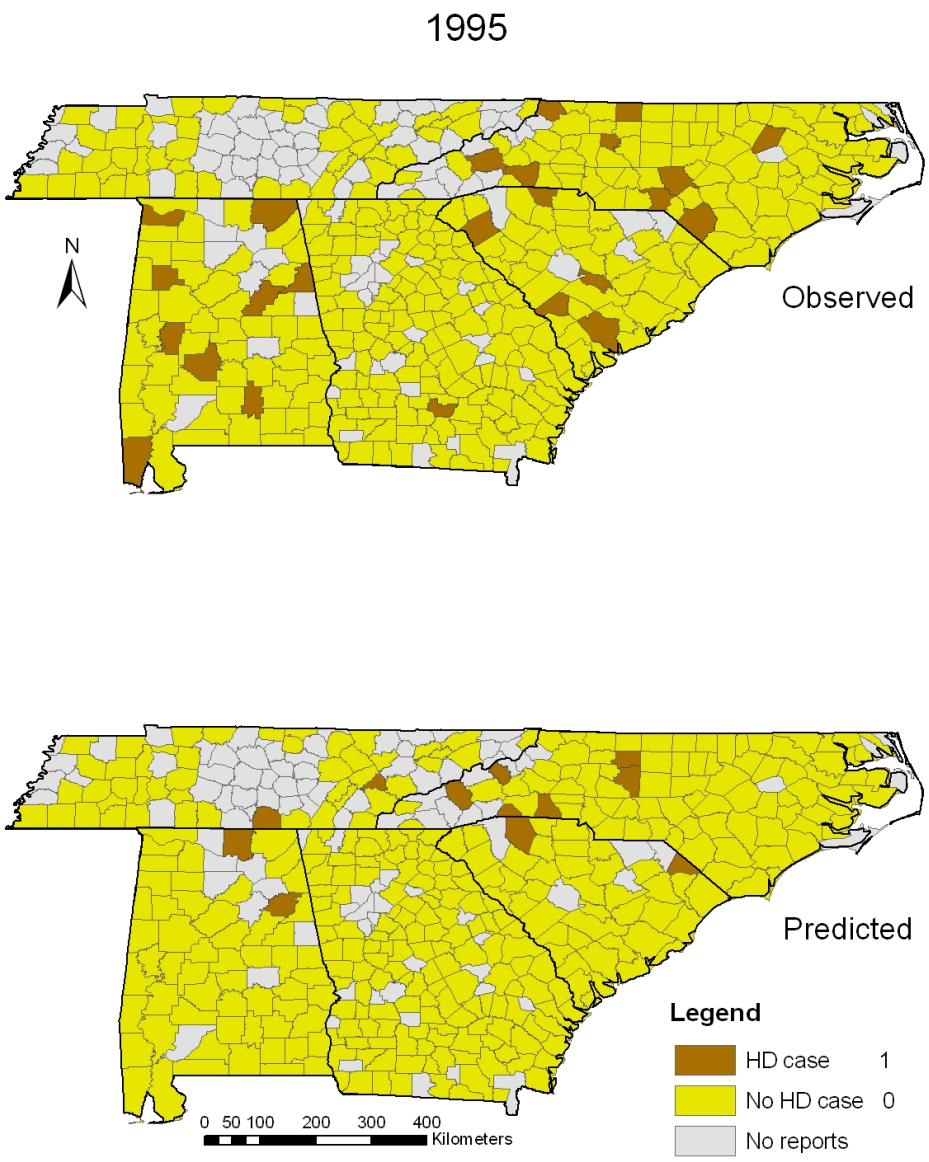
1993	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	0	28	28	0.00
	Absence	0	338	338	1.00
	Total	0	366	366	0.92

Figure 3.18: Observed and predicted HD cases (1993)



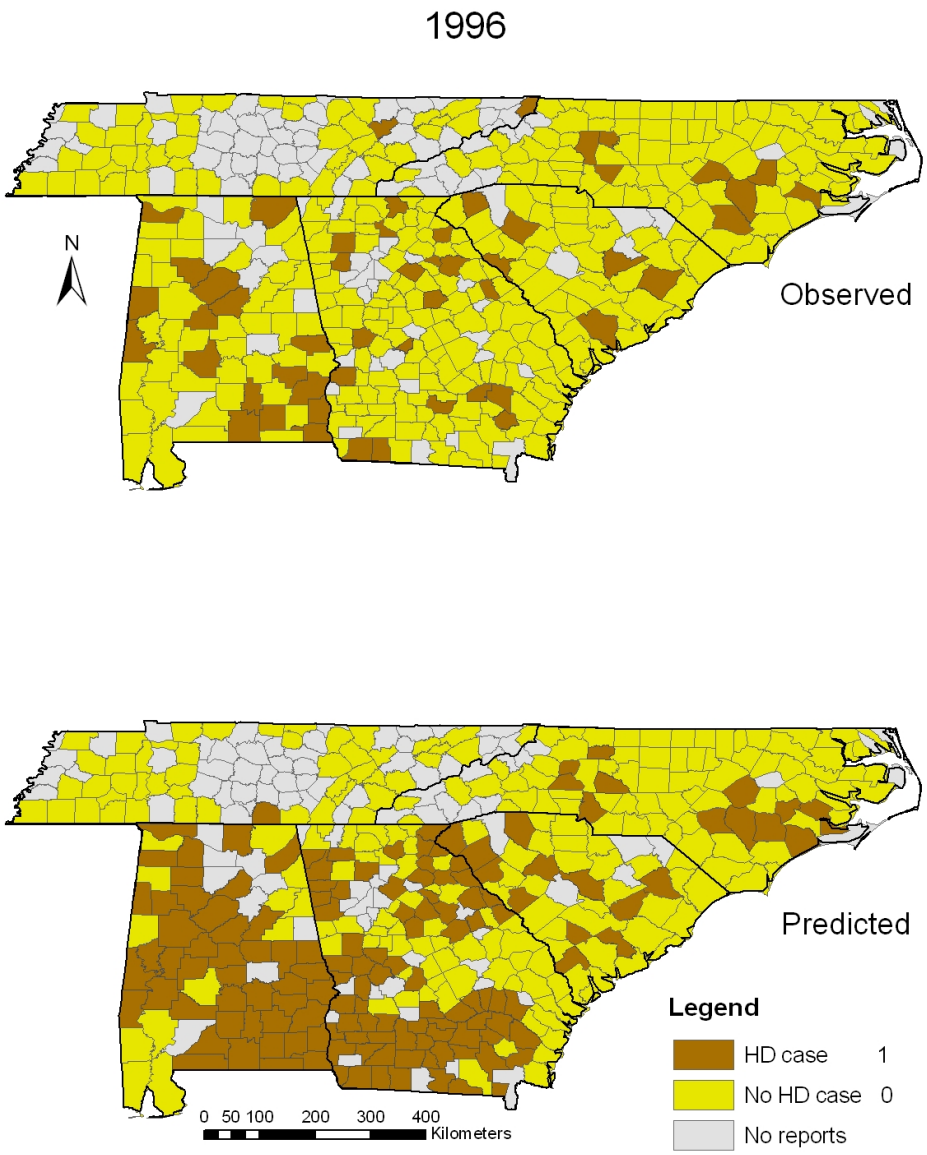
1994	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	93	5	98	0.95
	Absence	262	6	268	0.02
	Total	355	11	366	0.27

Figure 3.19: Observed and predicted HD cases (1994)



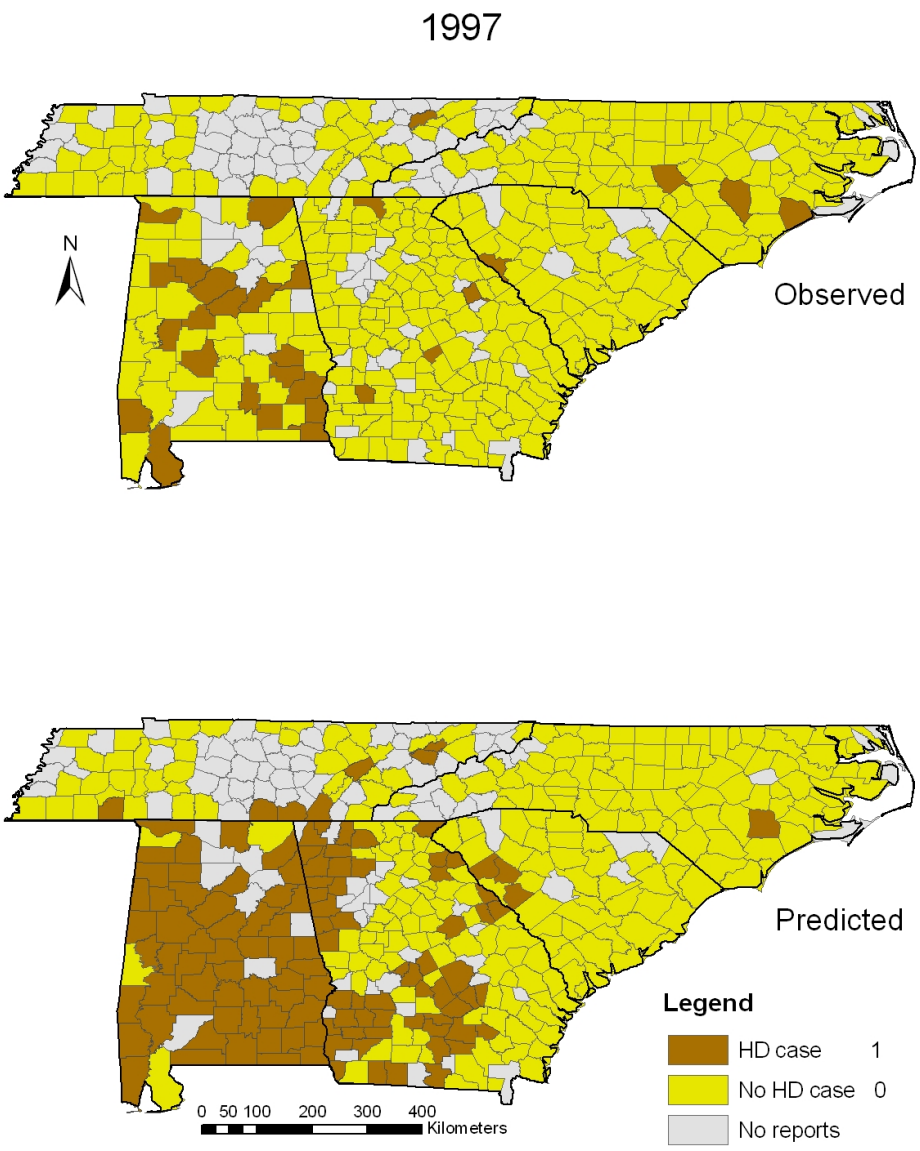
1995	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	0	24	24	0.00
	Absence	12	330	342	0.96
	Total	12	354	366	0.90

Figure 3.20: Observed and predicted HD cases (1995)



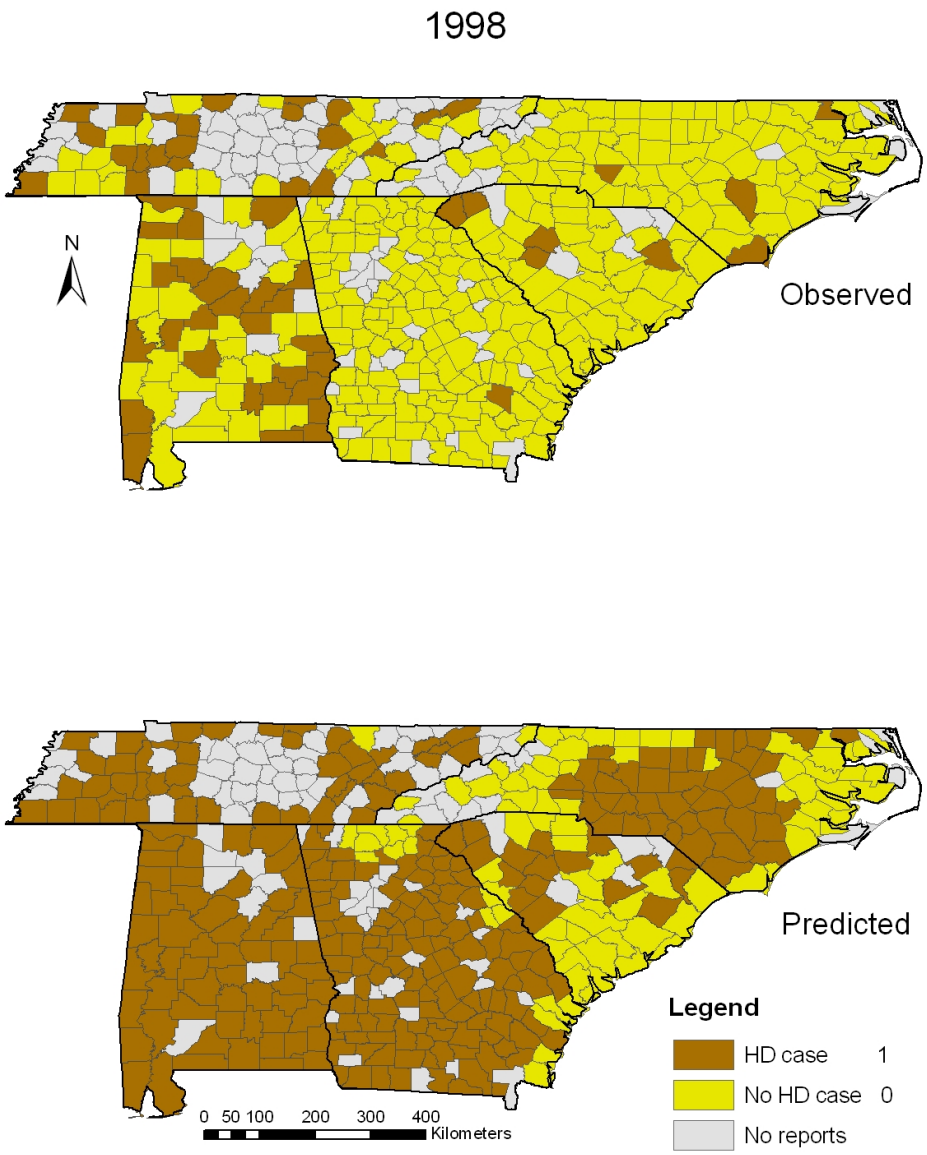
1996	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	22	30	52	0.42
	Absence	124	190	314	0.61
	Total	146	220	366	0.58

Figure 3.21: Observed and predicted HD cases (1996)



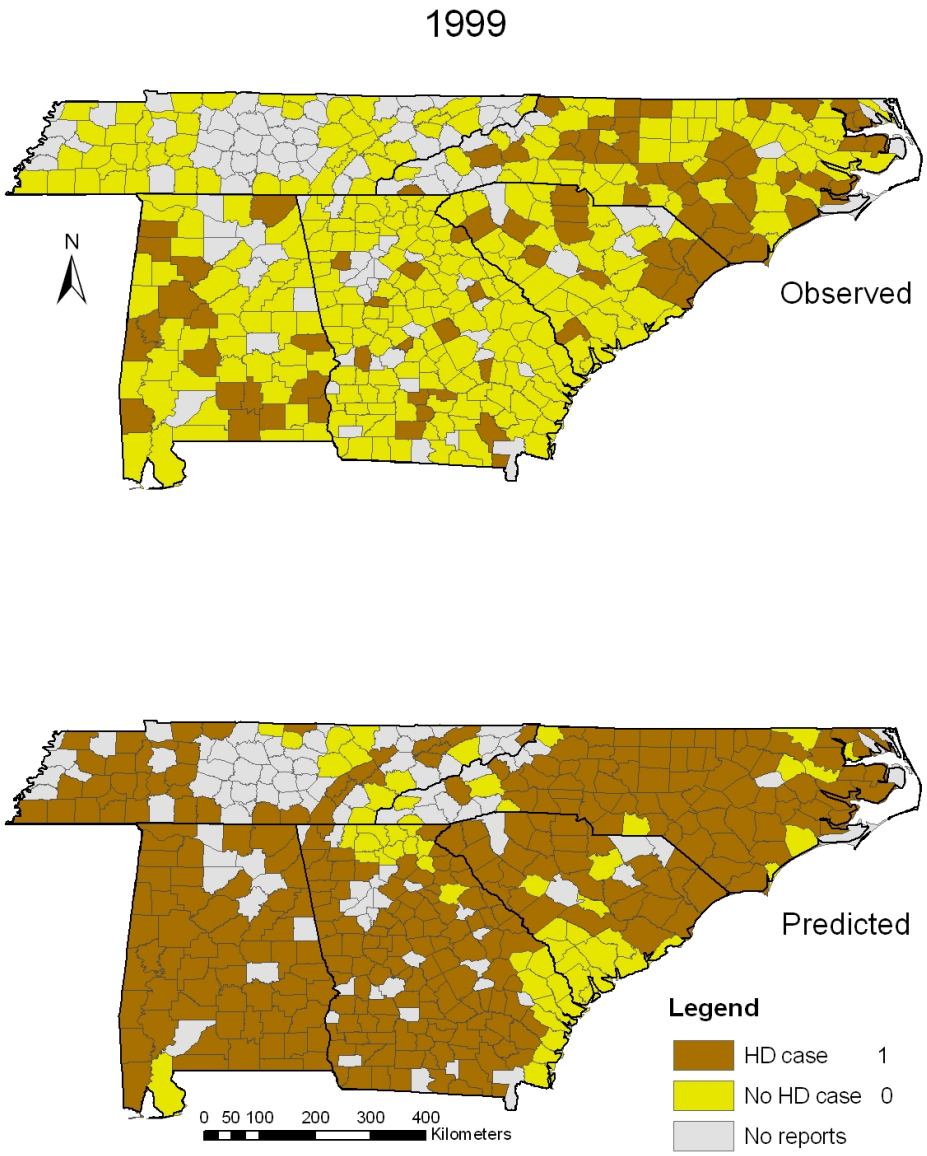
1997	Predicted			
Observed	Presence	Absence	Total	Accuracy
Presence	18	11	29	0.62
Absence	102	235	337	0.70
Total	120	246	366	0.69

Figure 3.22: Observed and predicted HD cases (1997)



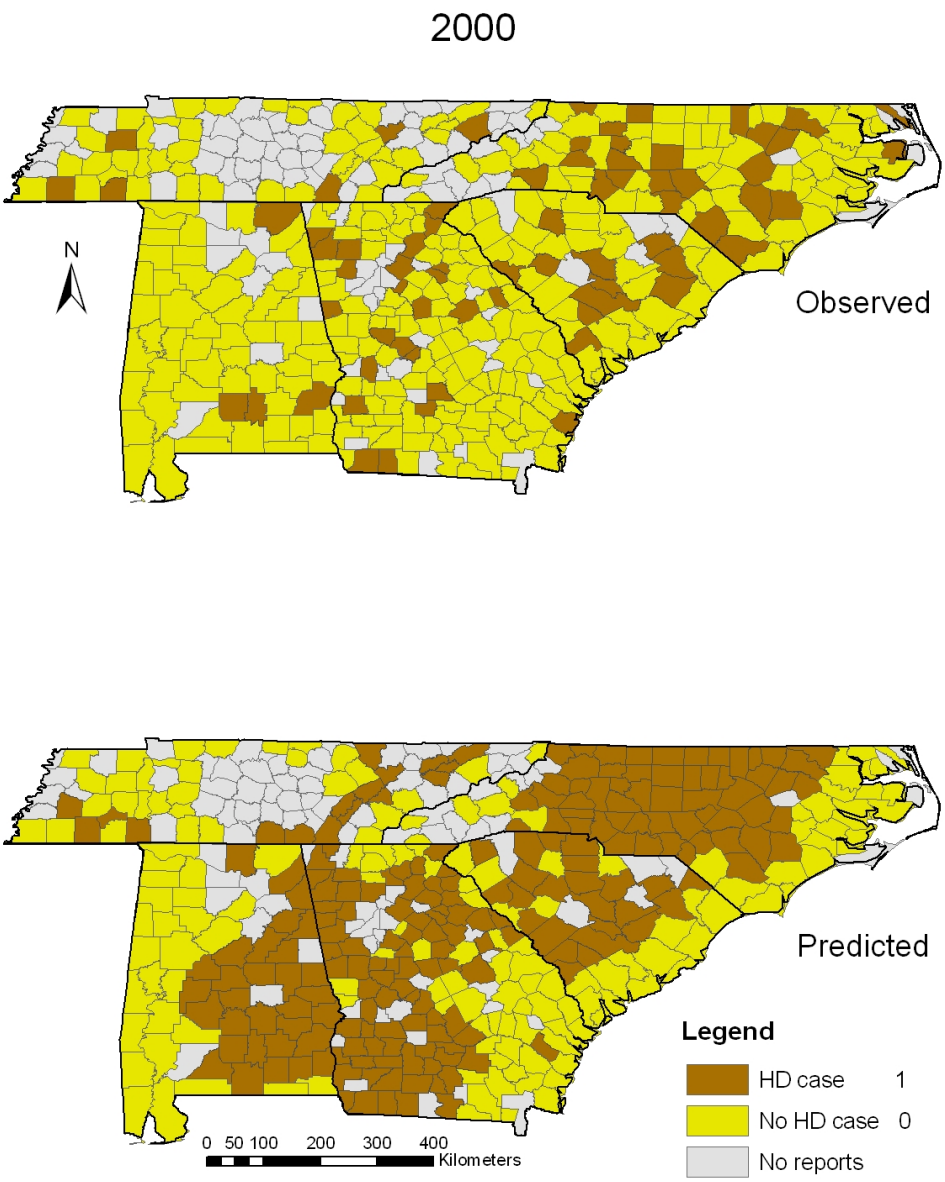
1998	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	56	3	59	0.95
	Absence	226	81	307	0.26
	Total	282	84	366	0.37

Figure 3.23: Observed and predicted HD cases (1998)



1999	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	70	10	80	0.88
	Absence	239	47	286	0.16
	Total	309	57	366	0.32

Figure 3.24: Observed and predicted HD cases (1999)



2000	Predicted				
Observed		Presence	Absence	Total	Accuracy
	Presence	50	21	71	0.70
	Absence	163	132	295	0.45
	Total	213	153	366	0.50

Figure 3.25: Observed and predicted HD cases (2000)

References

Augustin, N. H., Mugglestone, M. A., and Buckland, S. T., 1996. An autologistic model for the spatial distribution of wildlife, *The Journal of Applied Ecology*, 33 (2): 339-347.

Baylis, M., Bouayoune, H., Touti, J., and Hasnaoui, H., 1998. Use of climatic data and satellite imagery to model the abundance of Culicoides imicola, the vector of African horse sickness virus, in Morocco. *Medical and Veterinary Entomology,* 12: 255-266.

Baylis, M., Meiswinkel, R., and Venter, G. J., 1999. A preliminary attempt to use climate data and satellite imagery to model the abundance and distribution of *Culicoides imicola* (Diptera: Ceratopogonidae) in southern Africa. *Journal of the South African Veterinary Association,* 70: 80-89.

Baylis, M., Mellor, P. S., Wittmann, E. J., and Rogers, D. J., 2001. Prediction of areas around the Mediterranean at risk of bluetongue by modeling the distribution of its vector using satellite imaging, *Veterinary Record,* 149: 639-643.

Baylis, M. and Rawlings, P., 1998. Modeling the distribution and abundance of *Culicoides imicola* in Morocco and Iberia using climatic data and satellite imagery, *African Horse Sickness*, Arcives of Virology (Supplement) 14 (P. S. Mellor, M. Baylis, C. Hamblin, C. Calisher, and P. P. C. Mertens, editors), Vienna, Springer,137-153.

Beck, L. R., Rodriguez, M. H., Dister, S. W., Rodriguez, A. D., Washino, R. K., Roberts, D. R., and Spanner, M. A., 1997. Assessment of a remote sensing-based model for predictiong malaria transimission risk in villages of Chiapas, Mexico, *Am. J. Trop. Med. Hyg.*, 56(1): 99-106.

Bohning, D., Dietz, E., and Schlattmann, P., 2000. Space-time mixture modeling of public health data. *Statistics in Medicine*, 19: 2333-2344.

Boone, J. D., McGwire, K. C., Otteson, E. W., DeBaca, R. S., Kuhn, E. A., Villard, P., Brussard, P. F., and St. Jeor, S. C., 2000. Remote sensing and geographic information systems: Charting Sin Nombre Virus infections in deer mice, *Emerging Infectious Diseases*, 6(3): 248-258.

Burnham, K. P. and Anderson, D. R., 2004. Multimodel inference: understanding AIC and BIC in model selection, *Sociological Methods and Research*, 33: 261-304.

Cattell, R. B., 1966. The scree test for the number of factors, *Multivariate Behavioral Research*, 1: 245-276.

Cooper, D. R. and Asrar, G., 1989. Evaluating atmospheric correction models for retrieving surface temperatures from the AVHRR over a tallgrass prairie, *Remote Sensing of Environment*, 27: 93-102.

Couvillion, C. E., Davidson, W. R., Pearson, J. E., and Gustafson, G. A., 1981. Hemorrhagic disease among white-tailed deer from 1971 through 1980, *Proceedings of the United States Animal Health Association,* 85: 522-537.

Cross E. R., Newcomb, W. W., and Tucker C. J., 1996. Use of weather data and remote sensing to predict the geographic and seasonal distribution of Phlebotomus papatasi in southwest Asia, *American Journal of Tropical Medicine and Hygiene*, 54: 530-536.

Durr, P. A., 2004. Spatial epidemiology and animal disease: introduction and overview. *GIS and Spatial Analysis in Veterinary Science* (P. A. Durr and A. C. Gatrell, editors), Wallingford, UK ; Cambridge, MA, CABI Pub, 35-67.

Eisa, M., Karrar, A. E., and Elrahim, A. H., 1979. Incidence of bluetongue virus precipitating antibodies in sera of some domestic animals in the Sudan, *Journal of Hygiene*, 83: 539-545.

Erasmus, B. J., 1975. The epizootiology of bluetongue: the African situation, *Australian Veterinary Journal*, 51: 196-203.

Estrada-Peña, A., 1997. Epidemiological surveillance of tick populations: a model to predict the colonization success of Ixodes ricinus (Acari: Ixodidae), *European Journal of Epidemiology,* 13: 581-586.

Fienberg, S. E., Bromet, E. J., Follmann, D., Lambert, D., and May, S. M., 1985. Longitudinal analysis of categorical epidemiological data: a study of three mile island, *Environmental Health Perspectives*, 63: 241-248.

Glass, G. E., Cheek, J. E., Patz, J. A., Shields, T. M., Doyle, T. J., Thoroughman, D. A., Hunt, D. K., Enscore, R. E., Gage, K. L., Irland, C., Peters, C. J., and Bryan, R., 2000. Using remotely sensed data to identify areas at risk for hantavirus pulmonary syndrome, *Emerging Infectious Disease*, 6(3): 238-247.

Hales, S., Wet, N. E., Maindonald, J., and Woodward, A., 2002. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model, *The Lancet*, 360: 830 – 834.

Hay, S. I., 2000. An overview of remote sensing and geodesy for epidemiology and public health application, *Advances in Parasitology*, 47: 1-35.

Hay, S. I., Tucker, C. J. Rogers, D. J. and Packer, M. J., 1996. Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease, *Annals of Tropical Medicine and Parasitology,* 90(1): 1-19.

Hedeker, D., 2003. A mixed-effects multinomial logistic regression model, *Statistics in Medicine*, 22: 1433-1446.

Hill, T., and Lewicki, P., 2007. *Statistics: Methods and Applications*, StatSoft, Tulsa, OK

Hinely, A. J., 2006. *GIS-based Habitat Modeling Related to Bearded Capuchin Monkey Tool Use*, Master thesis, the University of Georgia, Athens, Georgia.

Hoff, G. L. and Trainer, D. O., 1978. Bluetongue and epizootic hemorrhagic disease viruses: their relationship to wildlife species, *Advances in Veterinary Science and Comparative Medicine,* 22: 111-132.

Hoff, G. L. and Trainer, D. O., 1981. Hemorrhagic disease in wild ruminants, *Infectious Diseases of Wild Mammals* (Davis, J. W., L. H. Karstad, and D. O. Trainer, editors) Ames: Iowa State University Press, 45-53.

Homan, E. J., Mo, C. L., Thompson, L. H., Barreto, C. H., Oviedo, M. T., Gibbs, E. P. J., and Greiner, E. C., 1990. Epidemiologic study of bluetongue viruses in Central America and the Caribbean: 1986-1988, *American Journal of Veterinary Research*, 51: 1089-1094.

Jensen, R. E., 1967. A multiple regression model for cost control: assumptions and limitations, *The Accounting Review*, 42: 265-273.

Justice, C. O., Townshend, J. R. G., Holben, B. N., and Tucker, C. J., 1985. Analysis of the phenology of global vegetation using meteorological satellite data, *International Journal of Remote Sensing*, 10: 1539-1561.

Kaiser, H. F., 1960. The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, 20: 141-151.

Kleinschmidt, I., Bagayoko, M., Clarke, G., Craig, M., and Sueur, D. L., 2000. A spatial statistical approach to malaria mapping. *International Journal of Epidemiological*, 29: 355-361.

Knorr-Held, L. and Besag, J., 1998. Modeling risk from a disease in time and space, *Statistics in Medicine*, 17: 2045-2060.

Lauer, R. M., Clarke, W. R., and Burns, T. L., 1997. Obesity in childhood: the Muscatine study, *Acta Padiatrica Scandinavica,* 38: 432-437.

Lindsay, S. W., and Thomas, C. J., 2000. Mapping and estimating the population at risk from lymphatic filariasis in Arica, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94: 37-45.

Linn, R. L., 1968. A Monte Carlo approach to the number of factors problem, *Psychometrika*, 33: 37-71.

Linthicum, K. J., Bailey, C. L., Davies, F. G. and Tucker, C. J., 1987. Detection of Rift Valley fever viral activity in Kenya by satellite remote sensing imagery, *Science,* 235: 1656-1659.

Nettles, V. F., Davidson, W. R., and Stallknecht, D. E., 1992. Surveillance for hemorrhagic disease in white-tailed deer and other wild ruminants, 1980-1989, *1992 Proceedings of Annual Conference of Southeastern Association of Fish and Wildlife Agencies*, 138-146.

Nettles, V. F. and Stallknecht, D. E., 1992. History and progress in the study of hemorrhagic disease of deer, *Transactions of the North American Wildlife and Natural Resources Conference,* 57: 499-516.

Nevill, E. M., 1971. Cattle and *Culicoides* biting midges as possible overwintering hosts of bluetongue virus, *Journal of Veterinary Research*, 38: 65-72.

Pickle, L.W., 2000. Exploring spatio-temporal patterns of mortality using mixed effects models, *Statistics in Medicine*, 19: 2251-2263.

Price, J. C., 1984. Land surface temperature measurement for the split window channels of the NOAA 7 advanced very high resolution radiometer, *Journal of Geographysical Research*, 89: 7231-7327.

Roberts, D. R., Paris, J. F., Manguin, S., Harbach, R. E., Woodruff, R., Rejmankova, E., Polanco, J., Wullschleger, B., and Legters, L. J., 1996. Predictions of malaria vector distribution in belize based on multispectral satellite data, *American Journal of Tropical Medicine and Hygiene*, 54(3): 304-308.

Rogan, W. J., Dietrich, K. N., Ware, J. H., Dockery, D. W., Salganik, M., Radcliffe, J., Jones, R. L., Ragan, N. B., Chisolm, J. J., and Rhoads, G. G., 2001. The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead, *New England Journal of Medicine*, 344: 1421-1426.

Rogers, D. J., Hay, S. I., and Packer, M. J., 1996. Predicting the distribution of tsetse flies in west Africa using temporal Fourier processed meteorological satellite data, *Annals of Tropical Medicine and Parasitology*, 90(3): 225-241.

Rogers, D. J. and Williams B. G., 1993. Monitoring trypanosomiasis in space and time. *Parasitology*, 106: 77-92.

Snow, R. W., Gouws, E., Omumbo, J., Rapuoda, B., Craig, M. H., Tanser, F. C., Sueur, D. L., and Ouma, J., 1998. Models to predict the intensity of Plasmodium falciparum transmission: applications to the burden of disease in Kenya, *Transactions of the royal society of tropical medicine and hygiene*, 92: 601-606.

Stallknecht, D. E., Kellogg, M. L., Blue, J. L., and Person, J. E., 1991. Antibodies to bluetongue and epizootic hemorrhagic disease viruses in a barrier island white-tailed deer population, *Journal of Wildlife Diseases*, 27(4), 668-674.

Sugita, M. and Brutasaert, W., 1993. Comparison of land surface temperatures derived from satellite observations with ground truth during FOFE, *International Journal of Remote Sensing,* 14: 1659-1676.

Thomas, F. C., Willis, N., and Ruckerbauer, G., 1974. Identification of viruses involved in the 1971 outbreak of hemorrhagic disease in southeastern United States white-tailed deer, *Journal of Wildlife Diseases*, 10: 187-189.

Thomson, M. C., and Conner, J., 2000. Environmental information systems for the control of arthropod vectors of disease, *Medical and Veterinary Entomology*, 14: 227-344.

Thomson, M. C., Connor, S. J., D'alessandro, U., Rowlingson, B., Diggle, p., Cresswell, M., and Greenwood, B., 1999. Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results, , *American Journal of Tropical Medicine and Hygiene*, 61(1): pp. 2-8.

Trainer, D. O., 1964. Epizootic hemorrhagic disease of deer, *Journal of Wildlife Management,* 28: 377-381.

Tucker, L. R., Koopman, R. F., and Linn, R. L., 1969. Evaluation of factor analytic research procedures by means of simulated correlation matrices, *Psychometrika*, 34: 421-459

Tucker, C. J., Vanpraet, C., Boerwinkel, E. and Gaston, A., 1983. Satellite remote sensing of total dry matter production in the Senegalese Sahel, *Remote Sensing of Environment*, 13: 461-474.

Twisk, Jos W. R., 2003. *Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide*. Cambridge, U.K.; New York: Cambridge University Press.

Van Marter, L, J., Leviton, A., Kuban, K. C. K., Pagano, M. and Allred, E. N., 1990. Maternal glucocorticoid therapy and reduced risk of bronchopulmonary dysplasia, *Pediatrics*, 86: 331-336.

Walker, A. R., and Davies, F. G., 1971. A preliminary survey of the epidemiology of bluetongue in Kenya, *Journal of Hygiene*, 69: 47-60.

Waller, L. A., Carlin, b. P., Xia, H., and Gelfand, A. E., 1997. Hierarchical spatio-temporal mapping of disease rates, *Journal of the American Statistical Association*, 92: 607-617.

Ward. M. P., 1994. Climatic factors associated with the prevalence of bluetongue virus infection of cattle herds in Queensland, Australia, *Veterinary Record,* 134: 407-410.

Wittmann E. J. and Baylis, M., 2000. Climate change: effects on *Culicoides*-transmitted viruses and implications for the UK, *The Veterinary Journal*, 160: 107-117.

Wood, B. L., Beck, L. R., Washino, R. K., Hibbrard, K. A., and Salute J. S., 1991. Estimating high mosquito-producing fields using spectral and spatial data, *International Journal of Remote Sensing*, 12: 621-626.

Xia, H., and Carlin, B. P., 1998. Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality, *Statistics in Medicine*,17: 2025-2043.

Zhu, L. and Carlin, B. P., 2000. Comparing hierarchical models for spatio-temporally misaligned data using the deviance information criterion, *Statistics in Medicine*, 19: 2265-2278.

CHAPTER 4

SPATIAL AND SPACE-TIME CLUSTERING ANALYSIS OF HEMORRHAGIC DISEASE IN WHITE-TAILED DEER IN SOUTHEAST USA: 1980 - 2003 [2]

[2] Xu, B., Stallknecht, D. E., Madden, M., Parker, K. C., and Hodler, T. W. To be submitted to *Journal of Wildlife Diseases*.

Abstract: Clustering analysis can provide valuable information for spatial epidemiology on possible causes of disease as well as its distribution. Although previous studies have revealed some spatial and temporal patterns of hemorrhagic disease (HD) in white-tailed deer, no statistical methods have been applied to detect its spatial and temporal clusters. This research uses the Kulldorff's space-time scan statistic to analyze whether there are any clusters in the distribution of HD in white-tailed deer in five southeastern states, and if yes, where and when do the clusters occur. The HD occurrence data are binary presence/absence data acquired annually on a county basis from 1980 to 2003. Purely spatial clustering analysis for the whole study area, space-time clustering analysis, and purely spatial clustering analysis by years were applied to the counties in Alabama, Georgia, South Carolina, North Carolina, and Tennessee. The results show that there are statistically significant spatial clusters and space-time clusters in the study area during the study period. Some clusters reoccur every several years. Almost all the clusters are restricted to 10 percent of the total population and 25 percent of the study period. The most evident high rate clusters are located along the west boundary of Alabama, southern Alabama, central South Carolina, and along the boundary between South Carolina and North Carolina. These detected clusters can be used to identify possible causes and future prevention of HD outbreaks.

Keywords: Hemorrhagic disease, Kulldorff's space-time scan statistic, Clustering analysis

4.1 Introduction

Hemorrhagic disease (HD) is a common disease in white-tailed deer (*Odocoileus virginianus*) which results in significant mortality and subsequent economic impacts on recreational hunting. The first documented outbreak of HD in USA dates back to the 1890s (Trainer, 1964; Hoff and Trainer, 1981). Since then, extensive work has been done in veterinary science to study the clinical signs and the mechanisms of the disease. Now it has been clear that HD is caused by the bluetongue (BLU) and epizootic hemorrhagic disease (EHD) viruses (Hoff and Trainer, 1978) and transmitted by arthropod *Culicoides* midges (Nettles and Stallknecht, 1992). Clinical signs and lesions of HD in white-tailed deer can include hyperemia, facial and cervical swelling, lameness, hemorrhage, sloughing of hooves, and ulceration (Hoff and Trainer, 1978). The infected deer often avoid and appear hypersensitive to sunlight (Hoff and Trainer, 1981).

In addition to pathologic studies, the spatial and temporal distributions of HD in white-tailed deer have been explored. According to a long-term and national surveillance conducted by the Southeastern Cooperative Wildlife Disease Study (SCWDS) in the College of Veterinary Medicine, University of Georgia, two contiguous bands of HD occurrence are found in the USA. The first one is a transverse band from the southeast along the Missouri River northwestward to the Great Plains. The other band is in coastal and northern California spreading to central Oregon and western Washington (Nettles et al., 1992). As to the temporal distribution, Couvillion et al. (1981), in their studies on HD among white-tailed deer from 1971 through 1980, found that all peracute or acute cases occurred between June and November. The frequency of

peracute or acute cases increased in July, peaked in September, and declined sharply in November. Approximately 83 percent of all the HD cases examined were observed during August, September, and October. They also concluded that there is a biennial rise and fall as to the yearly pattern of reported HD cases.

Although these previous analyses provided overall trends in the spatial and temporal distribution of HD outbreaks, the space and time dimensions were studied separately, and did not consider the interaction between space and time on the clusters of HD distribution. The objective of this study is thus to statistically test whether HD in white-tailed deer in the southeastern USA exhibit clusters in space and in space-time. If yes, then where and when do the clusters occur and how have the geographical clusters evolved with time?

In Section 2, some background information on spatial and space-time clustering techniques are introduced. Section 3 describes the HD data from 1980 to 2003 and the study area consisting of five states in southeast USA: Alabama, Georgia, South Carolina, North Carolina, and Tennessee. Section 4 shows the Kulldorff's space-time scan statistic that is used in the study, and based on that, Section 5 gives the spatial and space-time clustering results of HD in white-tailed deer in the study area during the 24-year study period. The paper ends with a short conclusion in Section 6.

4.2 Clustering analysis

Cluster detection is one of the four tasks in spatial epidemiology (i.e., disease mapping, geographical correlation studies, the assessment of risk in relation to a point or line-source, and cluster detection) (Elliott et al., 2000). Investigation of clusters of disease occurrence can provide

valuable information on possible causes of the disease, and appropriate methods for disease control and prevention (Ward and Carpenter, 2000). It is argued that disease cases occur close to each other in time as well as in space. Statistical methods for space-time clustering can be divided into two categories: space-time interaction and space-time cluster detection. The former gives only a global statistical index indicating whether there are space-time clusters or not in the study area during the study period. Space-time interaction such as the Knox's test (Knox, 1964), the Barton's method (Barton and David, 1966), Mantel's regression (Mantel, 1967), and the k nearest neighbor test (Jacquez, 1996), all are designed for point locations of cases. None of them, however, take into account the dynamic change of the underlying population at risk, which is important because when population density increases, the spatial distance from case to case will decrease (Jacquez, 1996).

Space-time cluster detection methods identify clusters together and test their significance. Kulldorff's space-time scan statistic is one of the few methods for space-time cluster detection. It not only allows the detection of actual geographical location and temporal period of clusters, but it also considers the distribution of the at-risk population density (Norström et al. 2000). Meanwhile, Kulldorff's space-time scan statistic can be applied to point data as well as area data for clustering analysis (Kulldorff, 2006).

Researchers in veterinary science have been adopting the above methods to explore space-time clusters in practical applications of disease investigations (White et al., 1989; Carpenter et al., 1996; Paré et al., 1996; Ekstrand and Carpenter, 1998; Singer et al., 1998; Fuchs et al., 2000; Ward and Carpenter, 2000a; Ward and Armstrong, 2000; Norström et al.,2000).

However, there is still much work to be done. The relative lack of space-time clustering in veterinary science is possibly due to the following three reasons: a) it is usually difficult to acquire data for the locations of the disease and the corresponding time; b) interactions among space-time phenomena are often complex and difficult to analyze and visualize the clusters; and c) there are few readily available and mature space-time clustering analysis software packages. This study will utilize a unique data set of HD outbreaks in conjunction with a recent software package to analyze the spatial and space-time clusters in southeast USA from 1980 to 2003.

4.3 Study area and data sources

4.3.1 Study area

The study area in this research is the southeastern region of the USA, where the HD in white-tailed deer has been long reported and considered endemic (Hoff and Trainer, 1981). As early as 1949, Ruff (1949) pointed out that extensive mortality of an unexplained fatal hemorrhagic disease occurred at irregular intervals for many years in this region which was later believed to be similar to HD. In 1954 and 1955, epizootics similar to HD appeared among white-tailed deer population in the southeastern USA (Prestwood, et al., 1974). During the summer of 1971, the first documented HD outbreak of white-tailed deer, which caused significant die-offs, was reported. The disease first occurred in South Carolina and then erupted almost simultaneously in Virginia, Tennessee, North Carolina, Kentucky, Georgia, and Florida (Thomas et al., 1974). A widespread HD outbreak occurred in 1980, whereby deer in 156 counties in 8 states had clinical evidence of exposure to HD (Couvillion et al., 1981). In 1998, substantial die-offs appeared among white-tailed deer, marking a peak year of HD (Nettles and

Stallknecht, 1992). Furthermore, precipitating antibodies to BLU and EHD are consistently detected annually in white-tailed deer in this region (Stallknecht et al., 1991). According to SCWDS from 1980 to 1989, contiguous instances of HD were reported throughout the Southeast (Nettles et al., 1992). It is, therefore, paramount to explore the distribution trends and detect clusters of HD in this area for future prediction and prevention.

Although the HD occurrence data set acquired by SCWDS is nation-wide, the focus of this research only includes five states: Alabama, Georgia, South Carolina, North Carolina, and Tennessee (Figure 4.1). These five states cover large areas of forest and open fields that constitute the habitat of white-tailed deer. In addition, the mild winter, hot summer, and plentiful rainfall, together with various physiographical regions such as ridges and valleys, piedmont and plains, all provide favorable living conditions for white-tailed deer and *Culicoides* midges.

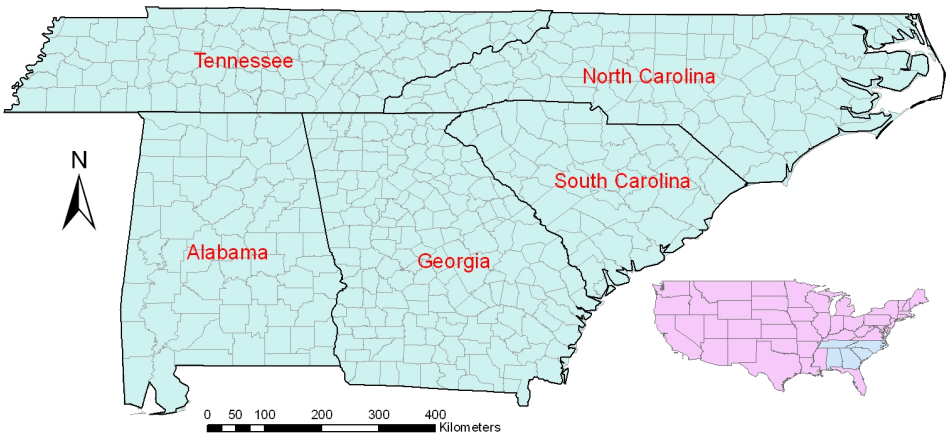


Figure 4.1: HD cluster analysis study area

4.3.2 HD data sources

The HD occurrence data are collected on a county basis through the USDA funded surveillance project conducted by SCWDS that has been performed annually since 1980. It is believed to be the most comprehensive database for white-tailed deer morbidity and mortality anywhere. The researchers mail questionnaires to the state fish and wildlife agency in each state, as well as most of the state veterinary diagnostic laboratories. The surveillance also includes personnel in the U.S. Fish and Wildlife Service and the Animal and Plant Health Inspection Service, USDA. For 1980 and 1981, only the 16 southeastern states were polled. In 1982, the surveillance was expanded to include all states except Hawaii. From the report of each state, a file is created to compile the reported HD occurrence by year and county. A subset of the data set was obtained for this study that included HD occurrence from 1980 to 2003 within the five state study area. The data are binomial, which means there are only 1s and 0s. If there is HD occurrence in the county in the specific year, it is represented as 1, otherwise, it is 0 (Table 4.1). Those counties that do not report any HD occurrence in any year were discarded from this research. All together, 371 counties in the five states are included in the study and 96 counties are excluded.

Table 4.1: HD data explanation

Value	Explanation
0	No HD exists in the county in the year
1	HD exists in the county in the year

Although the statistical space-time methods used in this paper allow point or area data, the HD data set is based on administrative counties. The reported presence or absence of HD for each county carries equal weight regardless of the varying sizes of the counties. Therefore, the county area data were converted to point data by utilizing the centroid of each county as the location of the HD presence or absence. The longitude and latitude of the centroid were calculated in a geographic information system (GIS) software package ArcGIS (ESRI, Redlands, California).

4.4 Methods

Kulldorff's space-time scan statistic is an extension of the Kulldorf's spatial scan statistic. In the spatial scan statistic, a theoretical circular window is placed on the map of the study area. This window is sequentially centered on a target location for which the surrounding area or neighborhood will be assessed. At each target location, the window radius varies continuously from zero to an upper limit specified by the user (at most 50 percent of the study area) (Figure 4.2a). For each location and size of the scanning window, a likelihood value is calculated using a likelihood function based on the number of observed and expected cases within and outside the window (Kulldorff et al., 1998; Aamodt, 2006). The null-hypothesis is there is no cluster of occurrence within the window, and the alternative hypothesis is that there is an elevated risk within the window as compared to outside. The window with the maximum likelihood value constitutes the most likely cluster. A likelihood ratio is calculated by comparing the maximum likelihood with the likelihood under the null hypothesis. The distribution under the null-hypothesis and the corresponding p value are obtained by repeating the same analytic

exercise on a large number of random replications of the data set generated under the null hypothesis in a Monte Carlo simulation (repeated random replications of the data set) (Smith et al., 2000).

Kulldorff's space-time scan statistic adds an additional time dimension to the window, which constitutes a cylindrical window in the three dimensions of geographic space (longitude and latitude) and time, with its base indicating space and height denoting time. The window is moved in space and time so that the circular base is centered on each possible geographic position throughout the study area (Figure 4.2b). For any given geographic position, the radius of the base varies continuously in size from zero to an upper limit (at most 50 percent of the total population), and the height of the cylinder also varies across all the possible time intervals from zero to an upper limit (at most 90 percent of the study period) (Kulldorff et al., 1998; Song and Kulldorff, 2003). Like the spatial scan statistic, finally I get the cylindrical window with the highest likelihood value is determined and its likelihood ratio and the corresponding p value are then calculated.

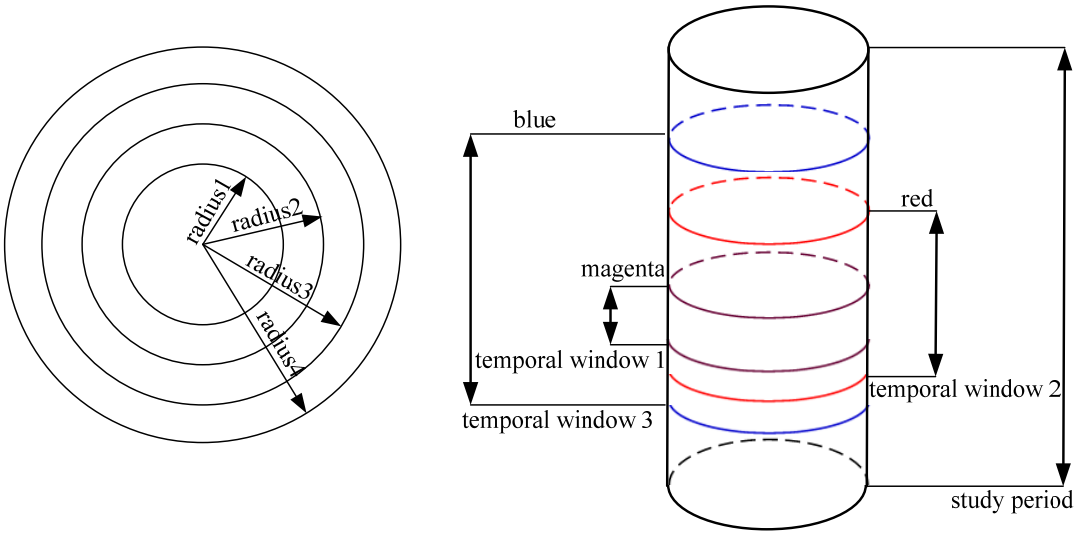


Figure 4.2: Schematic illustration of Kulldorff's space-time scan statistic method. (a): varying spatial windows for a target location; (b) varying temporal windows for a particular spatial window at a target location.

SaTScan™ version 7.0 is a free software package for spatial, temporal, and space-time clustering analysis (National Cancer Institute, Bethesda, Maryland, 2006). Kulldorf's space-time scan statistic is the theoretical core of SaTScan™. Different models also are incorporated in this software including the Bernoulli model, Poisson model, space-time permutation model, ordinal model, exponential model, and normal model. The ideal model for this study is the Bernoulli model which only requires information on the location of a set of cases and controls, while no underlying at-risk population distribution is needed. Each HD presence is represented by 1 as a case, and each HD absence is represented by 0 as a control. The location of each case or control is represented by the centroid of the county, thereby, converting broad-scale county level by viewing each county as a point in a large geographic area. The likelihood function for the Bernoulli model is as follows (Kulldorff, 2006):

$$L = \left(\frac{c}{n}\right)^{c}\left(\frac{n-c}{n}\right)^{n-c}\left(\frac{C-c}{N-n}\right)^{n-c}\left(\frac{(N-n)-(C-c)}{N-n}\right)^{(N-n)-(C-c)} I(\ ) \qquad \text{Equation 1}$$

Where L is likelihood value

C is total number of cases

c is observed number of cases within the window

$I(\ )$ is indicator function

n is total number of cases and controls within the window

N is total number of cases and controls in the data set

There are all together 8904 observations across the study area during the 24-year study period, with 1435 cases and 7469 controls. The SaTScan™ software can detect purely spatial, purely temporal and space-time clusters. In this analysis, the first step was to apply the purely spatial scan statistic to all the observations during the 24-year study period without considering the time aspect. The upper limit of the spatial window (radius) was set to 50 percent, 25 percent and 10 percent of the total population, respectively, to look for possible subclusters.

Second, the space-time scan statistic was performed to detect space-time clusters, allowing the spatial variations over the entire time period to be analyzed in a single model. The time aggregation was set to a one-year interval so that the space-time clusters could be detected for consecutive years as well as for one year. The maximum temporal scan window was set to 90 percent, 50 percent, and 25 percent of the study period, respectively.

The final step was to apply the purely spatial scan statistic to the cases and controls in the study area for each year to see whether there were spatial clusters in a particular year and how

these clusters vary across years. In this step, the spatial scan window was set to a maximum size of 10 percent of the total population.

For each of the three steps, both high and low rates of the clusters were tested, which means the output consists of clusters where the number of cases are higher than expected (high rates) as well as clusters where the number of cases are lower than expected (low rates). Thus I() in Equation 1 is equal to 1 for all windows since both high and low rates are tested. For statistical inference, 999 Monte Carlo replications were conducted to balance the accuracy and the processing time. In addition to the most likely cluster, SaTScan^{TM} also identifies secondary clusters in the data set, and orders them according to their likelihood ratio. In this study, I define primary clusters as those clusters with p-value less than or equal to 0.05, and secondary clusters as those clusters with p-value less than or equal to 0.1.

4.5 Results and discussion

4.5.1 Purely spatial clustering analysis during entire study period

The purely spatial clustering analyses with the maximum spatial window of 50 percent, 25 percent and 10 percent of the total population were performed for all the 24-year study period. The results are the same for the analyses with maximum spatial window of 50 percent and 25 percent, indicating that the detected clusters are robust when the radius exceeding 25 percent of the total population. The results for maximum spatial window of 50 percent and 25 percent (Situation 1) and 10 percent (Situation 2) of the total population are shown in Table 4.2 and Table 4.3 which list the cluster ID(Cl), latitude of the central location (Lat), longitude of the central location (Lon), radius (km), number of counties (Co), p-value (p), observed cases (Obs),

expected cases (Exp), and relative risk (RR) (the ratio of observed cases and expected cases) for high (HR) and low (LR) rate clusters. The primary (Pr) or secondary (Se) cluster is defined by its p value, <=0.05 and <=0.1, respectively. The results displayed in Figure 4.3 and 4.4 depict primary and secondary high rate clusters of HD presence as shades of red and orange, respectively, while primary and secondary low rate clusters indicating clusters of HD absence are depicted as bright and pale yellow, respectively. Individual counties identified as belonging to clusters listed in Table 2 and Table 3 are labeled as numbers for high rate (presence) and letters for low rate (absence) clusters.

For both Situations (i.e., spatial windows of 50 percent and 25 percent), the detected high rate clusters are exactly the same. That is, there are five high rate clusters, 4 primary and 1 secondary. The most statistically significant high rate cluster (Cluster 1) consists of 10 counties of Alabama bordering Georgia, with 103 observed HD cases over the 24-year study period. The relative risk is 2.79, which indicates 179 percent more cases than would have been expected under the null hypothesis. The second high rate cluster (Cluster 2) is found in the northwest of Alabama adjacent to Tennessee including 15 counties. It has 108 observed HD cases compared to the 58.02 expected cases, with a relative risk of 1.93. The third high rate cluster (Cluster 3) is located at the center of South Carolina, with only a single county, 17 HD cases and a relative risk of 4.44. The fourth high rate cluster (Cluster 4) is at the north boundary of Georgia and South Carolina, covering three counties in South Carolina, and two counties in Georgia with 40 HD cases occurring in Cluster 4 during the 24 years, compared to 19.34 expected cases. There is one secondary high rate cluster in South Carolina approximately midway between Cluster 3 and

Table 4.2: Results of purely spatial clustering analysis of HD in white-tailed deer during entire study period (spatial window <= 50 percent or 25 percent of total population): Situation 1

		Cl	°Lat	°Lon	Radius (km)	Co (No.)	p-value	Obs	Exp	RR
HR	Pr	1	31.40	-85.99	76.61	10	0.001	103	38.68	2.79
		2	33.80	-87.30	114.79	15	0.001	108	58.02	1.93
		3	33.67	-80.78	0.00	1	0.001	17	3.87	4.44
		4	34.89	-82.73	63.48	5	0.013	40	19.34	2.10
	Se	5	34.01	-81.73	33.62	2	0.077	20	7.74	2.61
LR	Pr	A	36.55	-85.54	260.30	57	0.001	120	220.47	0.50
		B	36.36	-78.41	151.07	29	0.005	70	112.17	0.60

Table 4.3: Results of purely spatial clustering analysis of HD in white-tailed deer during entire study period (spatial window <= 10 percent of total population): Situation 2

		Cl	°Lat	°Lon	Radius (km)	Co(No.)	p-value	Obs	Exp	RR
HR	Pr	1	31.40	-85.99	76.61	10	0.001	103	38.68	2.79
		2	33.80	-87.30	114.79	15	0.001	108	58.02	1.93
		3	33.67	-80.78	0.00	1	0.001	17	3.87	4.44
		4	34.89	-82.73	63.48	5	0.009	40	19.34	2.10
	Se	5	34.01	-81.73	33.62	2	0.059	20	7.74	2.61
LR	Pr	A	36.47	-86.46	189.79	31	0.001	48	119.91	0.38
		B	36.36	-78.41	151.07	29	0.003	70	112.17	0.60
		C	36.18	-82.85	119.31	14	0.009	26	54.15	0.47

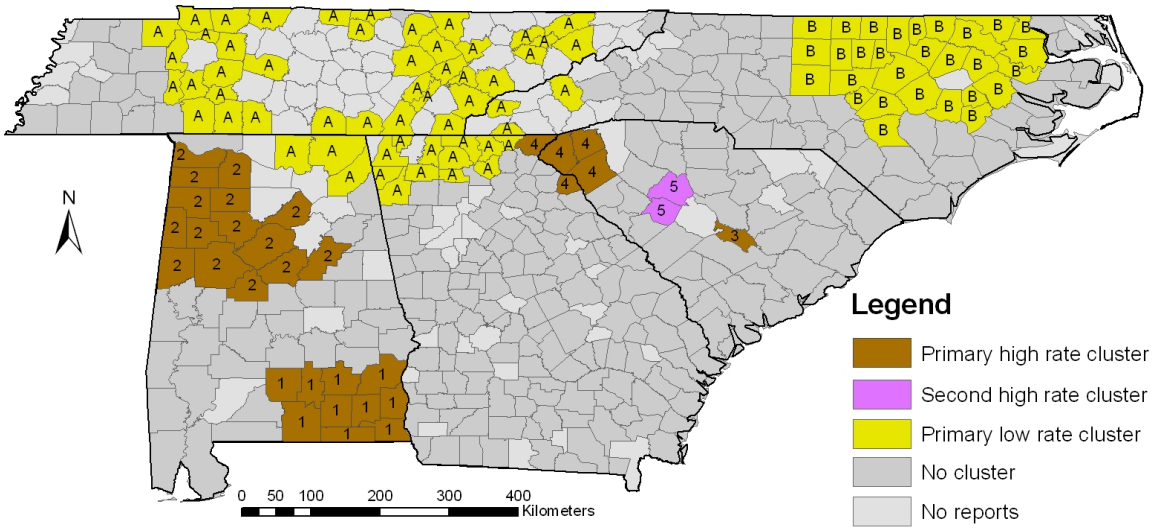


Figure 4.3: Purely spatial clustering analysis of HD in white-tailed deer during entire study period (spatial window <=50 percent or 25 percent of total population): Situation 1, numbers indicating high rate clusters, letters indicating low rate clusters.

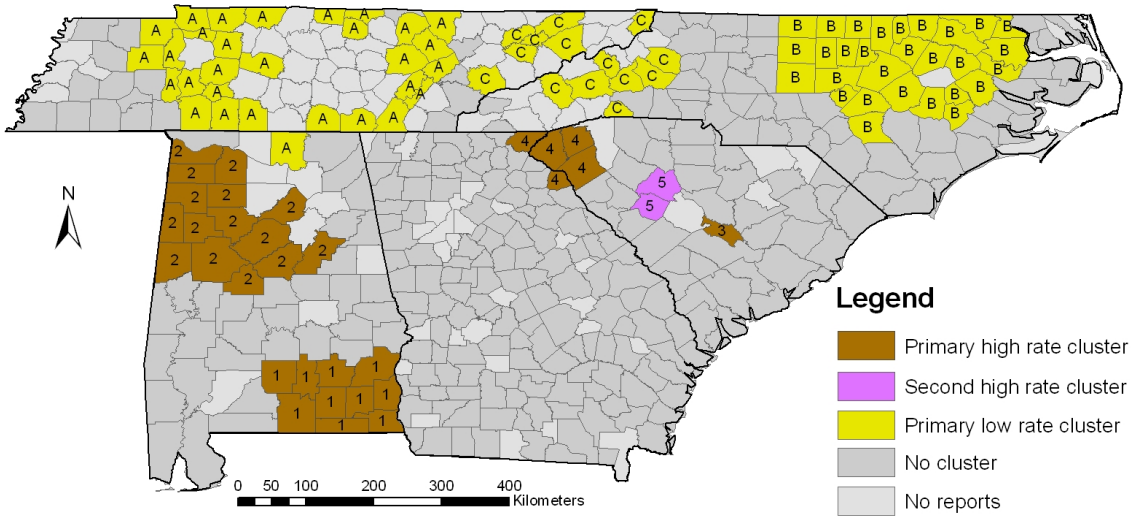


Figure 4.4: Purely spatial clustering analysis of HD in white-tailed deer during entire study period (spatial window <=10 percent of total population): Situation 2, numbers indicating high rate clusters, letters indicating low rate clusters.

Cluster 4 including 2 counties with 2.61 times more cases than are expected, and the significant level is 0.077.

With regard to the low rate clusters, there are some differences between the two Situations. For Situation 1 (spatial window <=50 percent or 25 percent of the total population), two primary clusters are found. Cluster A is situated at the boundary of Alabama, Tennessee, Georgia, and North Carolina, covering 57 counties. Most of the counties come from Tennessee (39 counties), 12 counties are from Georgia, and counties are in North Carolina, and 3 lie in Alabama. This Cluster extends over a large spatial area, with 120 cases observed, compared to the 220.47 expected cases. The second low rate cluster is located in the north of North Carolina, where 70 cases occurred during the whole study period in these 29 counties, whereas 112.17 cases should have been expected under the null hypothesis. For Situation 2 (spatial window <=10 percent of the total population), three low rate clusters are specified. Cluster A is essentially a subcluster of the Cluster A in Situation 1, consisting of 30 counties in Tennessee, and 1 county in Georgia. Within this Cluster, 48 HD cases actually occurred compared to the 119.91 expected cases under null hypothesis. Cluster B is exactly the same as the Cluster B in Situation 1. Cluster C is approximately another subcluster of Cluster A in Situation 1, concentrating on the boundary between Tennessee and North Carolina, with 7 counties in North Carolina included and 7 counties in Tennessee. It should be noted that the counties detected as part of Cluster A in north Georgia in Situation 1 are not identified as low rate cluster any more in Situation 2. However, more counties in North Carolina near Tennessee are detected as low rate cluster in Situation 2 than in Situation 1.

The purely spatial clustering analysis over the entire study period shows that there are high rate spatial clusters as well as low rate clusters in the study area during the years between 1980 and 2003. The clusters are relatively robust with varying maximum spatial windows (50 percent, 25 percent and 10 percent of the total population), and they are essentially located at the boundaries between two or three states. Alabama has the largest area of high rate clusters during the study period, whereas Tennessee and North Carolina both have large areas of low rate clusters. For both Situations, high rate clusters range from 1 to 15 counties. In Situation 1, low rate clusters range from 29 to 57 counties, while low rate clusters in Situation 2 range from 14 to 31 counties.

4.5.2 Space-time clustering analysis

Section 4.5.1 reveals that if the maximum spatial window is set to 10 percent of the total population, I can identify subclusters of those clusters resulting from 50 percent of the total population setting. However, there is no large difference in the results for the different window settings, which means the clusters are relatively robust. In the space-time clustering analysis, therefore, only 10 percent of the total population was used for the maximum spatial window setting. For the maximum temporal window setting, the 90 percent (Situation 3), 50 percent (Situation 4), and 25 percent (Situation 5) of the study period were used, respectively, to see the space-time clusters at different time scales. Table 4.4, Table 4.5, and Table 4.6 list the Cluster ID (Cl), latitude of the central location (Lat), longitude of the central location (Lon), radius (km), number of counties (Co), p-value (p), observed cases (Obs), expected cases (Exp), relative risk (RR), and cluster period (period) for high (HR) and low rate (LR) spatial-time clusters grouped

by primary (Pr) and secondary (Se) clusters of the three Situations. Figure 4.5, Figure 4.6, and Figure 4.7 provide a visual examination and comparison of these clusters. There are 3 low rate primary clusters in Situation 3. Cluster A consists of 35 counties that mostly lie in Tennessee, with a relative risk of 0.31 for the period from 1980 to 1997. Cluster B finds itself in the west of Georgia bordering Alabama with 12 observed cases compared to 43.51 expected cases in the 27 counties from 1989 to 1998. Cluster C includes 35 counties of Georgia and 2 counties in the southwest of South Carolina. It is significant from 1992 to 2001 with a relative risk of 0.38. Figure 4.8, Figure 4.9, and Figure 4.10 illustrate the relative risk of each space-time clusters with their corresponding temporal periods.

Table 4.4: Results of space-time clustering analysis of HD in white-tailed deer
(temporal window <= 90 percent of the study period;
spatial window <=10 percent of total population) : Situation 3

		Cl	^{o}Lat	^{o}Lon	Radius (km)	Co (No.)	p-value	Obs	Exp	RR	Period
HR	Pr	1	31.40	-85.99	76.61	10	0.001	93	30.62	3.18	1985-2003
		2	34.23	-77.88	193.20	36	0.001	33	5.80	5.80	2002
		3	36.43	-81.50	164.73	33	0.001	28	5.32	5.35	2002
		4	33.28	-88.09	178.89	26	0.001	124	67.04	1.93	1988-2003
		5	34.22	-82.46	113.22	32	0.001	23	5.16	4.52	1980
	Se	6	34.78	-86.00	0.00	1	0.100	9	1.61	5.61	1992-2001
LR	Pr	A	36.53	-86.87	204.50	35	0.001	33	101.53	0.31	1980-1997
		B	32.88	-84.30	95.70	27	0.002	12	43.51	0.27	1989-1998
		C	31.55	-81.92	150.29	37	0.003	23	59.63	0.38	1992-2001

Table 4.5: Results of space-time clustering analysis of HD in white-tailed deer (temporal window <= 50 percent of the study period; spatial window <=10 percent of total population): Situation 4

		Cl	^{o}Lat	^{o}Lon	Radius (km)	Co (No.)	p-value	Obs	Exp	RR	Period
HR	Pr	1	34.23	-77.88	193.20	36	0.001	33	5.80	5.80	2002
		2	31.40	-85.99	81.75	11	0.001	65	21.27	3.15	1988-1999
		3	36.43	-81.50	164.73	33	0.001	28	5.32	5.35	2002
		4	34.22	-82.46	113.22	32	0.001	23	5.16	4.52	1980
		5	33.72	-87.74	166.69	26	0.001	85	41.90	2.09	1994-2003
LR	Pr	A	36.36	-85.67	188.02	30	0.001	0	29.01	0.00	1982-1987
		B	36.00	-88.93	104.46	15	0.002	4	29.01	0.14	1982-1993
		C	32.88	-84.30	95.70	27	0.002	12	43.51	0.27	1989-1998
		D	31.55	-81.92	150.29	37	0.003	23	59.63	0.38	1992-2001

Table 4.6: Results of space-time clustering analysis of HD in white-tailed deer (temporal window <= 25 percent of the study period; spatial window <=10 percent of total population): Situation 5

		Cl	^{o}Lat	^{o}Lon	Radius	Co	p	Obs	Exp	RR	Period
HR	Pr	1	34.23	-77.88	193.20	36	0.001	33	5.80	5.80	2002
		2	36.43	-81.50	164.73	33	0.001	28	5.32	5.35	2002
		3	31.73	-86.31	55.21	5	0.001	24	4.83	5.03	1986-1991
		4	34.22	-82.46	113.22	32	0.001	23	5.16	4.52	1980
		5	32.46	-83.67	104.90	34	0.003	21	5.48	3.87	1980
		6	33.00	-87.13	98.89	15	0.003	38	14.50	2.66	1994-1999
		7	31.87	-85.39	41.98	3	0.003	11	1.93	5.72	1996-1999
LR	Pr	A	36.36	-85.67	188.02	30	0.001	0	29.01	0.00	1982-1987

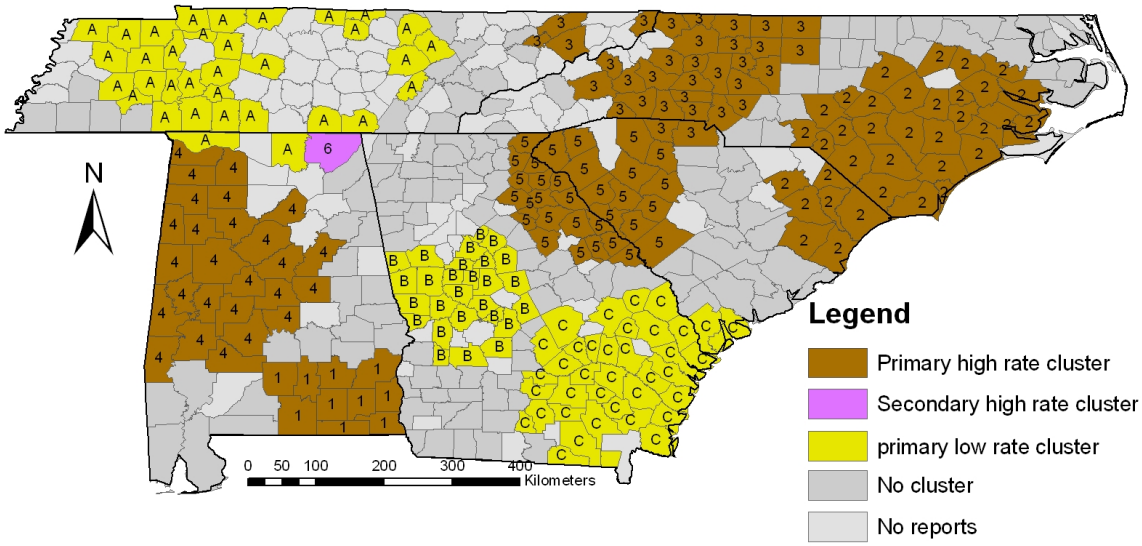


Figure 4.5: Space-time clustering analysis of HD in white-tailed deer (temporal window <= 90 percent of the study period; spatial window <=10 percent of total population): Situation 3, numbers indicating high rate clusters, letters indicating low rate clusters.

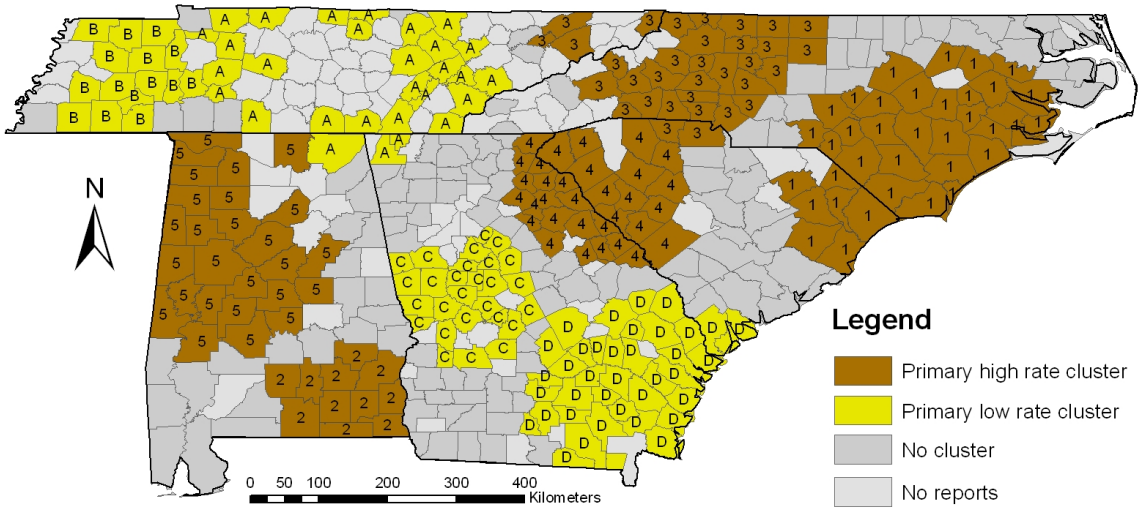


Figure 4.6: Space-time clustering analysis of HD in white-tailed deer (temporal window <= 50 percent of the study period; spatial window <=10 percent of total population): Situation 4, numbers indicating high rate clusters, letters indicating low rate clusters.

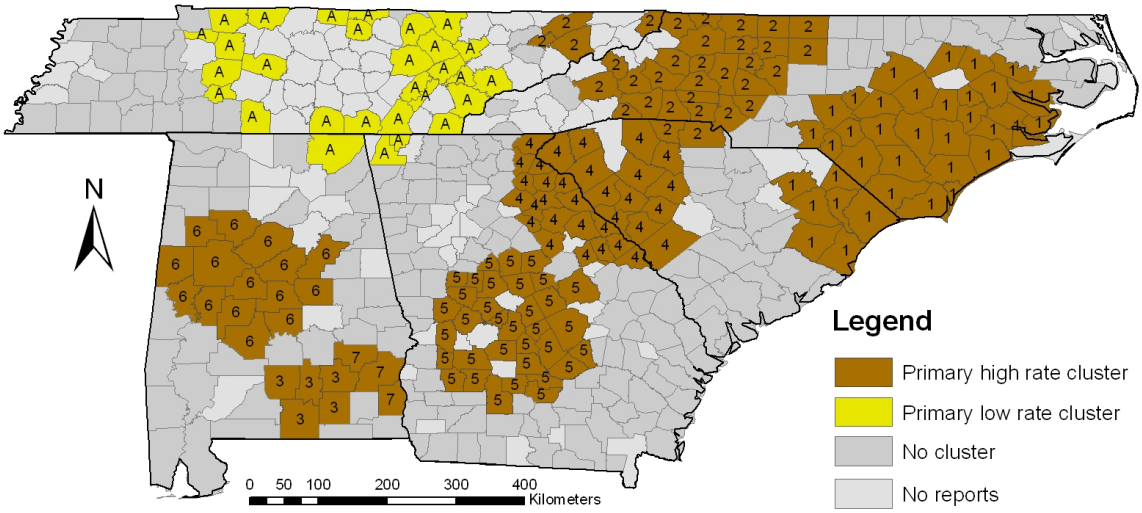


Figure 4.7: Space-time clustering analysis of HD in white-tailed deer (temporal window <= 25 percent of the study period; spatial window <=10 percent of total population): Situation 5, numbers indicating high rate clusters, letters indicating low rate clusters.

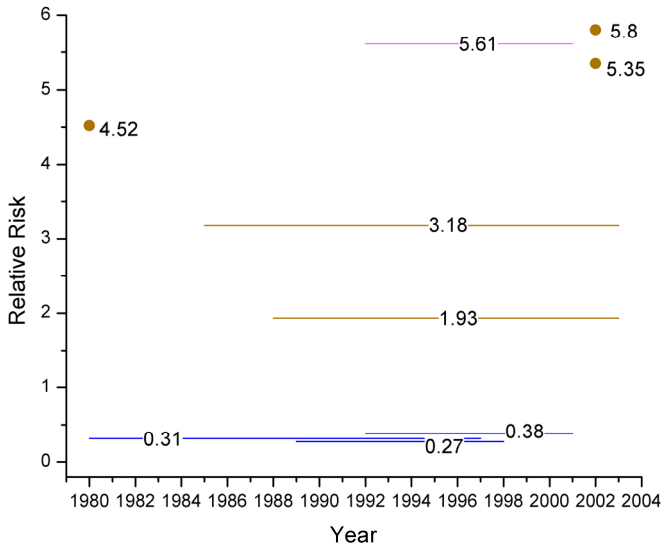


Figure 4.8: Relative risk and temporal periods for space-time clusters (temporal window <= 90 percent of the study period; spatial window <=10 percent of total population)

Note: brown color indicates primary high rate cluster, purple indicates secondary high rate cluster, and blue indicates primary low rate.

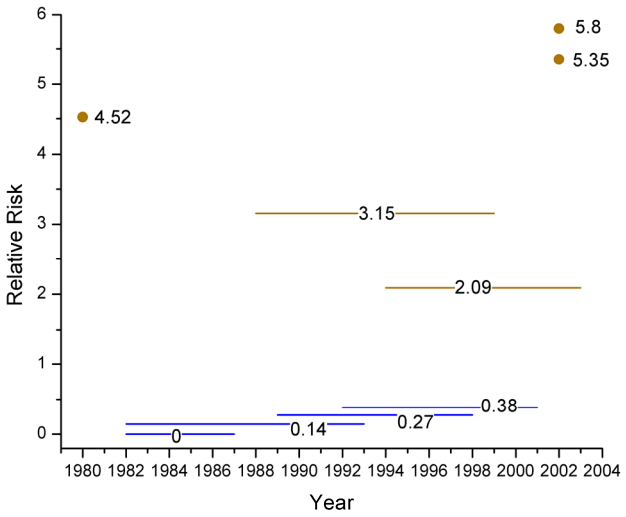


Figure 4.9: Relative risk and temporal periods for space-time clusters
(temporal window <= 50 percent of the study period;
spatial window <=10 percent of total population)

Note: brown color indicates primary high rate cluster, and yellow indicates primary low rate.

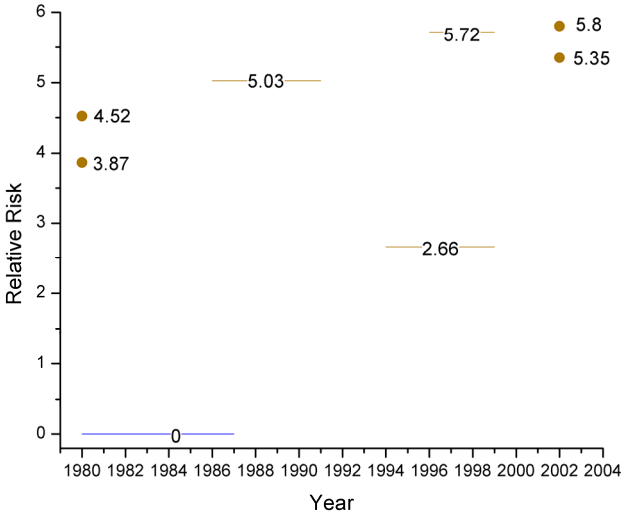


Figure 4.10: Relative risk and temporal period for space-time clusters
(temporal window <= 25 percent of the study period;
spatial window <=10 percent of total population)

Note: brown color indicates primary high rate cluster, and blue indicates primary low rate.

In Situation 3, 6 high rate clusters are identified, among which 5 are primary clusters and 1 is a secondary cluster. The most statistically significant primary cluster (Cluster 1) is located in southeast Alabama with 93 cases in the 10 counties contrasted to the 30.62 expected cases. The temporal period is from 1985 to 2003. Cluster 2 is found at the boundary of South Carolina and North Carolina along the coast. This cluster is significant in 2002 and is 480 percent higher than expected with 33 observed HD cases when 5.8 cases were expected. Cluster 3 is detected at the boundary of North Carolina, South Carolina, and Tennessee. Among the 33 counties, 2 are from South Carolina, 4 are from Tennessee, and the remaining 27 counties are in North Carolina. The significant period is year 2002 with a risk rate of 5.35. Cluster 4 situates at the northwest of Alabama including 26 counties, and is elevated 93 percent from 1988 to 2003. Cluster 5 is next to Cluster 3 around the north boundary of Georgia and South Carolina, covering 32 counties from both States with a relative risk of 4.52 in year 1980. In addition to the above 5 primary high rate clusters, there is another secondary high rate cluster (Cluster 6) comprised of a single county in Alabama bordering Tennessee. This cluster is significant from 1992 to 2001 with a p-value equal to 0.1 and 9 HD cases observed when 1.61 cases are expected during this period.

In Situation 4, Cluster 1 is the same as Cluster 2 in Situation 3. That means the 33 counties at the boundary between South Carolina and North Carolina along the coast in year 2002 became the statistically most significant high rate cluster in Situation 4. Cluster 2 in Situation 4 is approximately the same as Cluster 1 in Situation 3, with one more counties included and the significant period is from 1988 to 1999, which is a sub-period of Cluster 1 in Situation 3 (1985 – 2003). Cluster 3 in both Situations identifies exactly the same spatial area,

the same temporal period, and the same relative risk. Cluster 4 in Situation 4 is identical to Cluster 3 in Situation 3. Cluster 5 in Situation 4 reveals almost the same spatial area for Cluster 4 in Situation 3, but with shorter temporal period (1994 – 2003). There is no secondary high rate cluster discovered in Situation 4.

Regarding low rate clusters, there are 4 primary low rate clusters in Situation 4 compared with 3 in Situation 3. By closely examination, I can conclude that Cluster A and Cluster B in Situation 4 are approximately subclusters of Cluster A in Situation 3 both in spatial area and temporal period. Cluster C and Cluster D in Situation 4 are identical to Cluster B and Cluster C in Situation 3.

In Situation 5, the pattern is a little bit different. Because of the similarity of Situation 3 and Situation 4, the results of Situation 5 and Situation 4 are only compared. There are 7 primary high rate clusters and only one primary low rate cluster detected in Situation 5. Cluster 3 and Cluster 7 in Situation 5 are subclusters of Cluster 2 in Situation 4 both in space and in time. Spatially, there seems to be contradiction for Cluster 5 in Situation 5 compared with the same area in Situation 4. It is identified as a primary high rate cluster in Situation 5, but it covers some counties of Cluster C and Cluster D in Situation 4, which are low rate clusters. However, by examining the temporal period, I see there is no temporal overlap between Cluster 5 in Situation 5 and Cluster C and Cluster D in Situation 4, so no conflicts exist between these two analyses. The rest of the clusters are almost similar in both Situations.

In space-time clustering analysis, Situation 3 and Situation 4 discover almost the same pattern, with two low rate clusters in Situation 4 being the subclusters in Situation 3 both in

space and in time. The same clusters may have different significant level orders in the two Situations. The analysis reveals that the clusters are relatively robust for maximum temporal scan window set as 90 percent of the total study period and 50 percent of the total study period. For the temporal scan window less than and equal to 25 percent of the total study period, more high rate clusters are identified, whereas fewer low rate clusters are detected. Some clusters in Situation 5 are subclusters of those in Situation 3 and Situation 4. Despite the difference, most of the clusters have similar patterns for the three Situations. That is, the space-time clusters were mainly restricted to the temporal period of less than 25 percent of the study period.

Comparing the results of the space-time clustering analysis with the results of purely spatial clustering analysis in section 4.5.1, it can be concluded that, spatially, there are some similar clusters in both analyses. For convenience, here only Situation 1 in purely spatial clustering analysis and Situation 3 in space-time clustering analysis are examined. Cluster 1, Cluster 2, and Cluster A in Situation 1 almost cover the same area of Cluster 1, Cluster 4 and Cluster A in Situation 3. By examining the temporal periods of Cluster 1, Cluster 4 and Cluster A in Situation 3, three clusters were all found to extend long temporal periods of 1985-2003, 1988 – 2003, and 1980 – 1997, respectively. The convergence of the two analyses makes sense because the clusters extending along temporal periods in space-time clustering analysis should also be detected as clusters in purely spatial clustering analysis which takes the whole study period as a single time frame. Space-time clustering analysis detects more clusters than purely spatial analysis because some clusters only exist across short temporal periods, and cannot be identified as clusters when considering the study period as a whole. For example, Cluster 2

(temporal period: 2002) and Cluster 3 (temporal period: 2002) in Situation 3 are not detected as clusters in Situation 1. In a word, the purely spatial analysis essentially identifies the same clusters in space-time analysis that extends along temporal periods. However, there is an exception: Cluster 4 in Situation 1 is a subcluster of Cluster 5 in Situation 3 spatially, whereas temporally, the latter is a subcluster of the former one (temporal period of 1980 for Cluster 5 in Situation 3).

4.5.3 Purely spatial clustering analysis by individual year

The purely spatial clustering analysis was conducted across the study area by individual year to examine whether there are spatial clusters in particular years, whether spatial clusters exist in continuous years, and if yes, how the clusters evolve during those years. Table 4.7 lists the statistical results for maximum spatial window equal to 10 percent of the total population. Figure 4.11 and Figure 4.12 illustrate the number of clusters and the number of counties that were identified as clusters for each individual year, respectively. Figure 4.13 to Figure 4.27 show the visual results. Note, years where there are no statistically significant clusters are not depicted.

Spatial clusters exist for 15 years of the total study period of 24 years. There are more counties identified as clusters in 1980, 1982, 1993, 1998, 1999, and 2002, whereas fewer counties in 1986, 1989, and 1991. The largest number of counties that are detected as primary high rate clusters occurs in 1998 (63 counties), and 1980 is the year when primary low rate clusters cover maximum number of counties (65 counties). The number of clusters in each year ranges from 1 to 4, which is less than the number of clusters detected in previous analyses, possibly due to the reduced number of observed cases when analyzed by individual year. Most of

the clusters are high rate clusters. It seems some clusters do reoccur every several years. For examples: there is a high rate cluster in the west of Alabama in 1983, 1993, 1997, 1998 and 2001, the high rate cluster in the south of Alabama in 1986, 1987, 1994, and 1997, a high rate cluster along the boundary between South Carolina and North Carolina

Table 4.7: Results of purely spatial clustering analysis of HD in white-tailed deer by individual year (spatial window <= 10 percent of the total population)

Year			Cl	^{o}Lat	^{o}Lon	Radius (km)	Co (No.)	p-value	Obs	Exp	RR
1980	HR	Pr	1	33.67	-80.78	144.14	36	0.001	27	11.26	2.82
	LR	Pr	A	36.07	-88.07	215.45	37	0.001	0	11.57	0.00
		Pr	B	32.94	-86.25	130.40	28	0.009	0	8.75	0.00
1982	HR	Pr	1	34.52	-82.64	110.63	31	0.001	20	7.10	3.37
	LR	Pr	A	36.53	-86.01	215.16	37	0.021	0	8.48	0.00
		Se	B	36.49	-81.13	162.38	32	0.098	0	7.33	0.00
		Se	C	35.49	-77.68	135.40	32	0.098	0	7.33	0.00
1983	HR	Se	1	33.00	-87.13	160.15	31	0.067	8	1.67	7.31
1986	HR	Pr	1	31.73	-86.31	55.21	5	0.014	5	0.57	9.89
		Se	2	34.34	-80.59	130.66	29	0.094	11	3.28	4.18
1987	HR	Se	1	31.73	-86.31	36.06	3	0.055	3	0.16	21.65
1989	HR	Pr	1	31.73	-86.31	55.24	6	0.007	5	0.45	13.22
1991	HR	Pr	1	34.44	-87.84	53.94	5	0.029	5	0.75	7.18
1993	HR	Pr	1	33.92	-80.38	122.94	24	0.017	9	1.81	6.85
		Se	2	32.85	-87.95	205.73	35	0.067	10	2.64	5.33
1994	HR	Pr	1	31.15	-85.30	112.31	21	0.027	15	5.60	2.98
1997	HR	Pr	1	32.25	-87.79	197.50	35	0.001	14	2.83	8.40
1998	HR	Pr	1	36.53	-86.87	193.91	29	0.001	20	4.92	5.49
		Pr	2	32.85	-86.72	156.51	34	0.003	17	5.77	3.66
1999	HR	Pr	1	34.61	-78.56	153.91	34	0.001	20	7.33	3.30
	LR	Pr	A	36.50	-87.38	219.66	37	0.035	0	7.98	0.00
2001	HR	Se	1	32.59	-88.20	201.45	29	0.068	9	2.19	5.59
2002	HR	Pr	1	34.52	-77.91	176.54	33	0.001	31	14.23	2.46
		Pr	2	36.16	-80.67	90.28	18	0.001	18	7.76	2.49
	LR	Pr	A	36.53	-86.01	215.16	37	0.001	0	15.96	0.00
2003	HR	Pr	1	33.29	-87.53	188.25	35	0.001	21	7.08	3.73

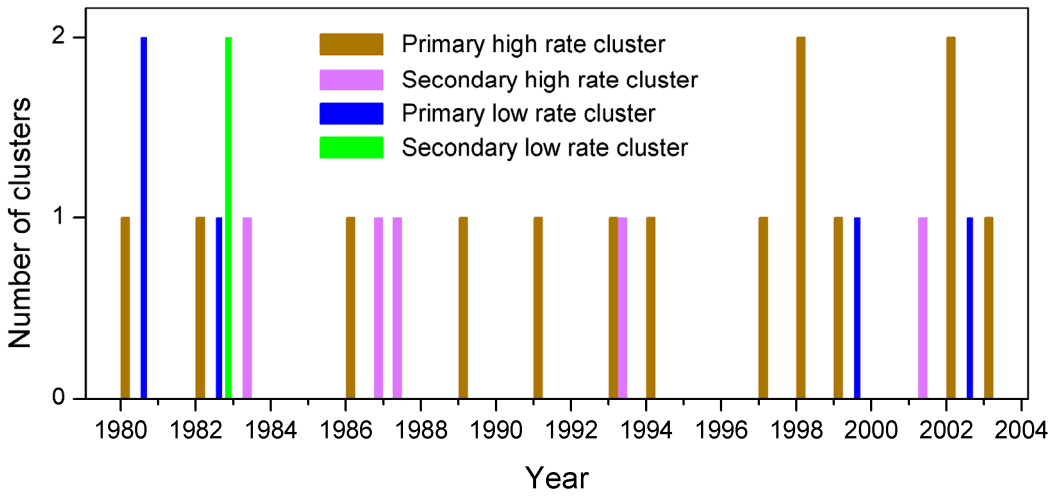


Figure 4.11: Number of clusters for individual year

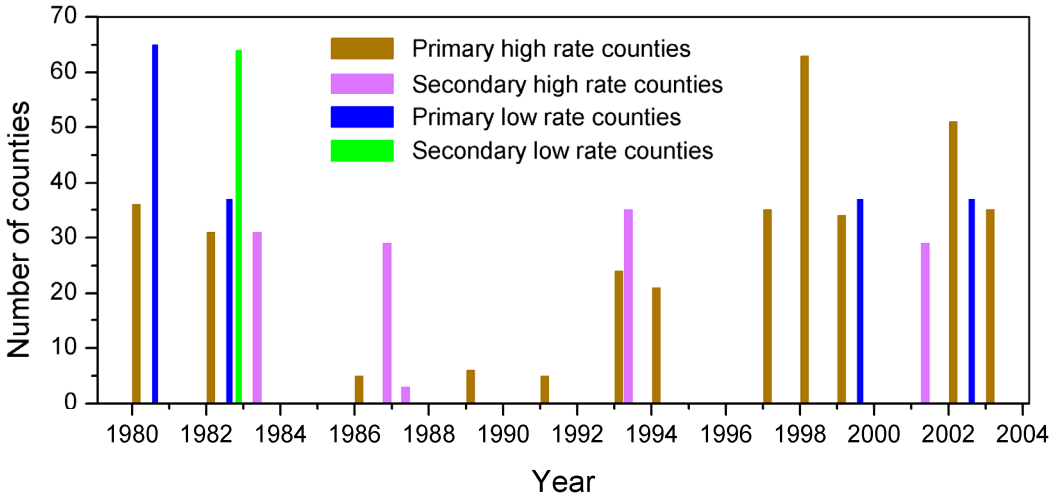


Figure 4.12: Number of counties identified as clusters for individual year

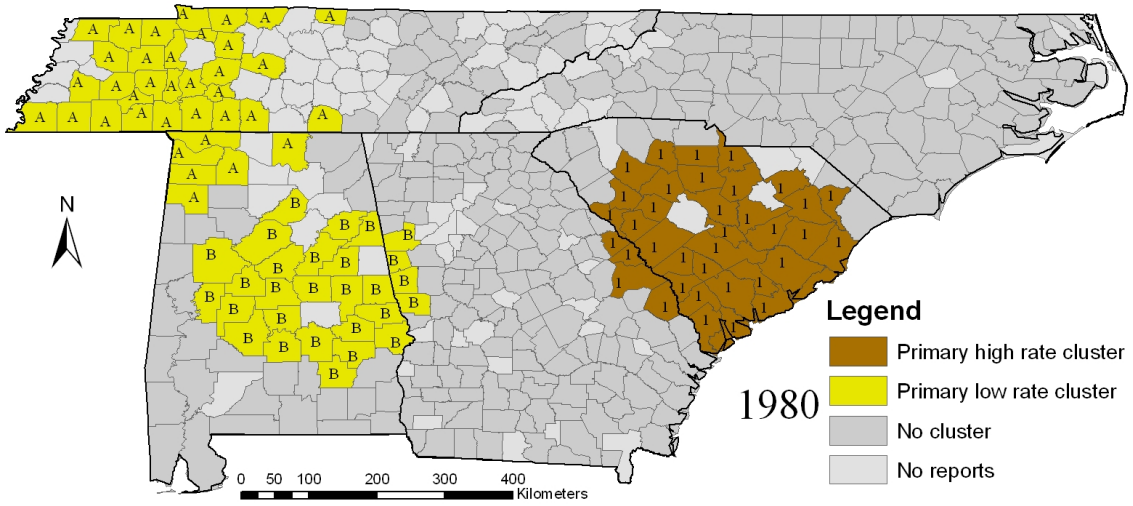


Figure 4.13: Purely spatial clustering analysis of HD in white-tailed deer (1980), numbers indicating high rate clusters, letters indicating low rate clusters.

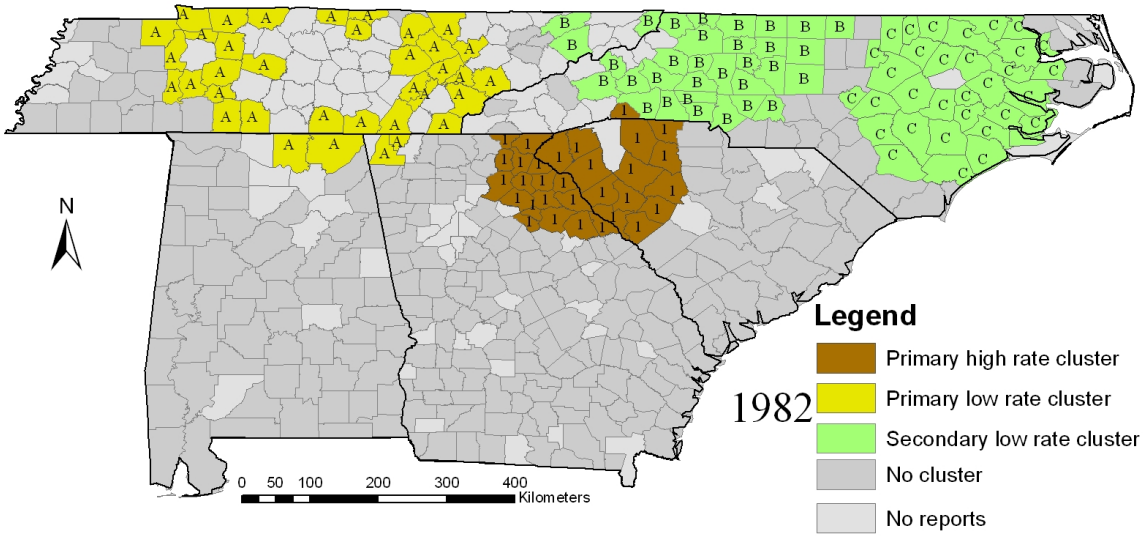


Figure 4.14: Purely spatial clustering analysis of HD in white-tailed deer (1982), numbers indicating high rate clusters, letters indicating low rate clusters.

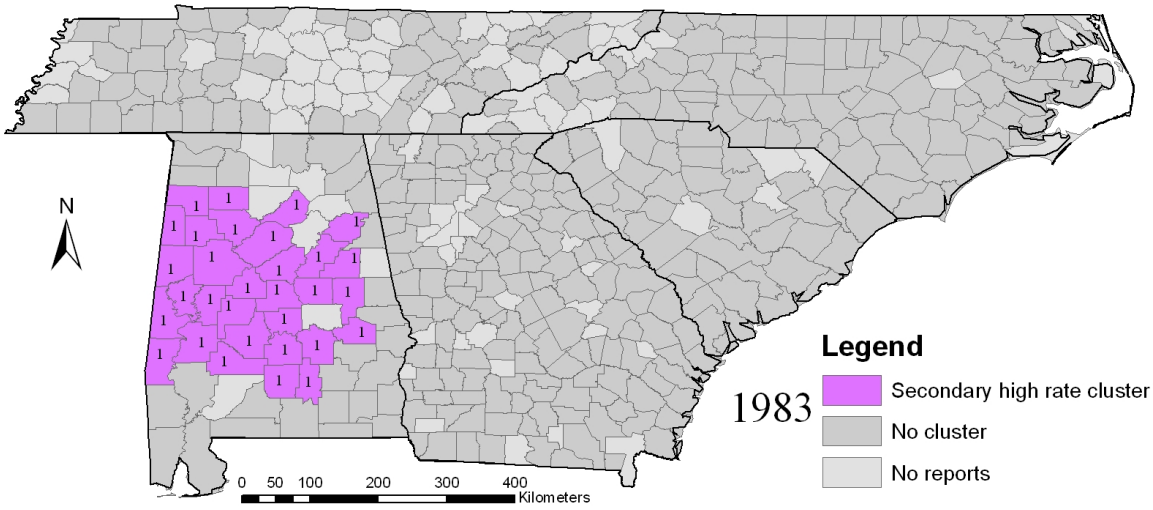


Figure 4.15: Purely spatial clustering analysis of HD in white-tailed deer (1983), numbers indicating high rate clusters, letters indicating low rate clusters.

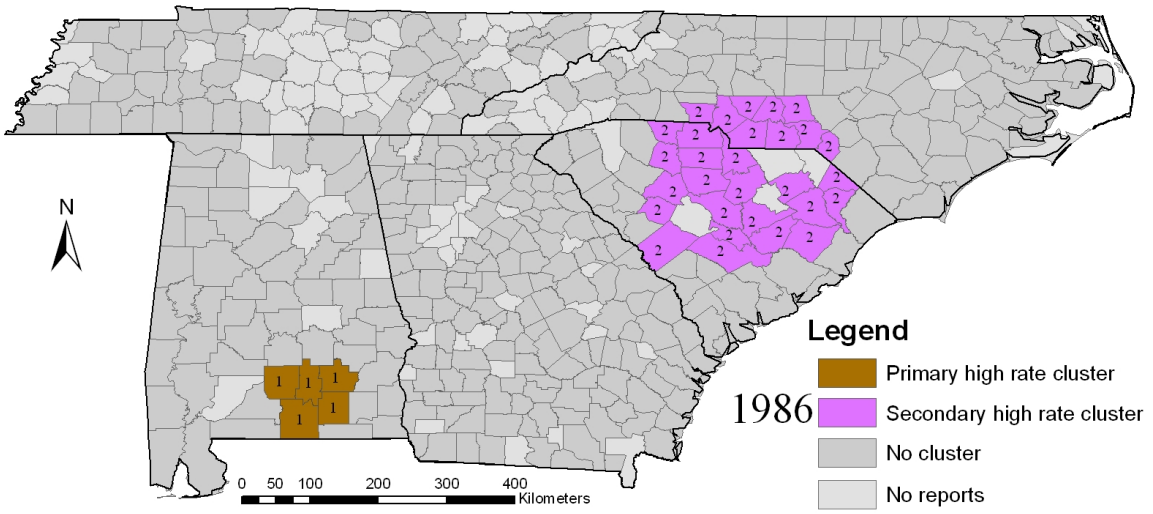


Figure 4.16: Purely spatial clustering analysis of HD in white-tailed deer (1986), numbers indicating high rate clusters, letters indicating low rate clusters.

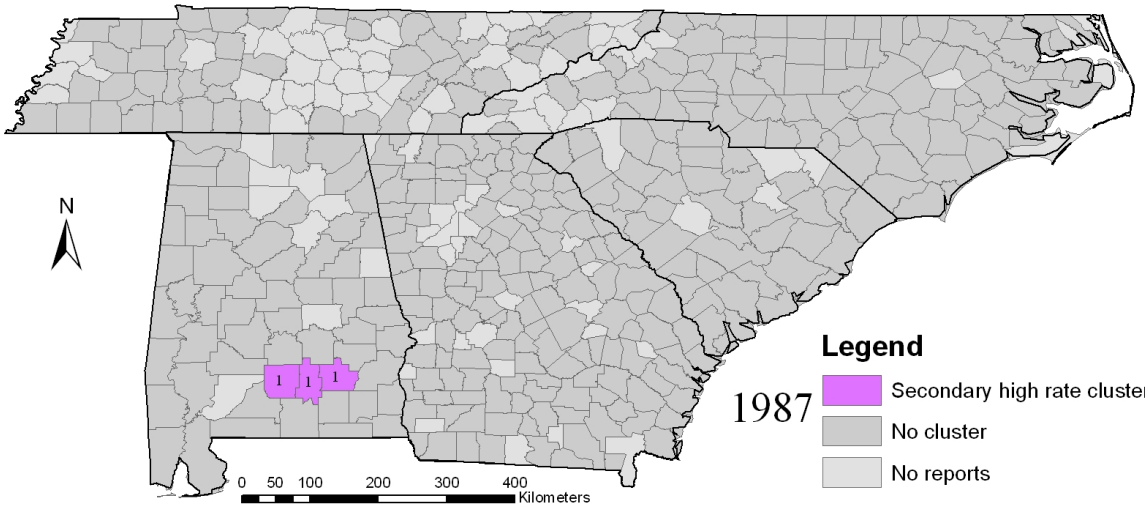


Figure 4.17: Purely spatial clustering analysis of HD in white-tailed deer (1987), numbers indicating high rate clusters, letters indicating low rate clusters.), numbers indicating high rate clusters, letters indicating low rate clusters.

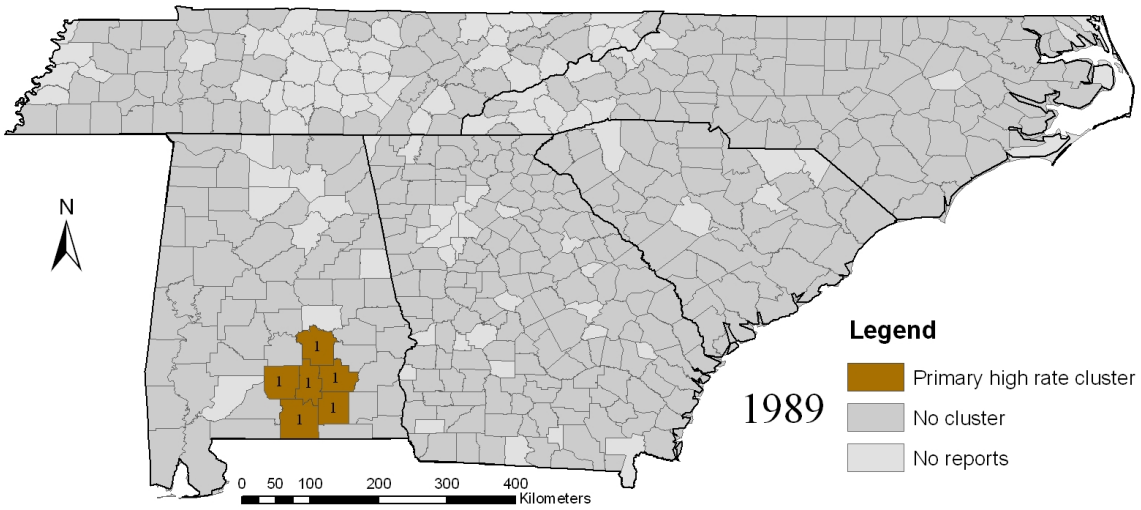


Figure 4.18: Purely spatial clustering analysis of HD in white-tailed deer (1989), numbers indicating high rate clusters, letters indicating low rate clusters.

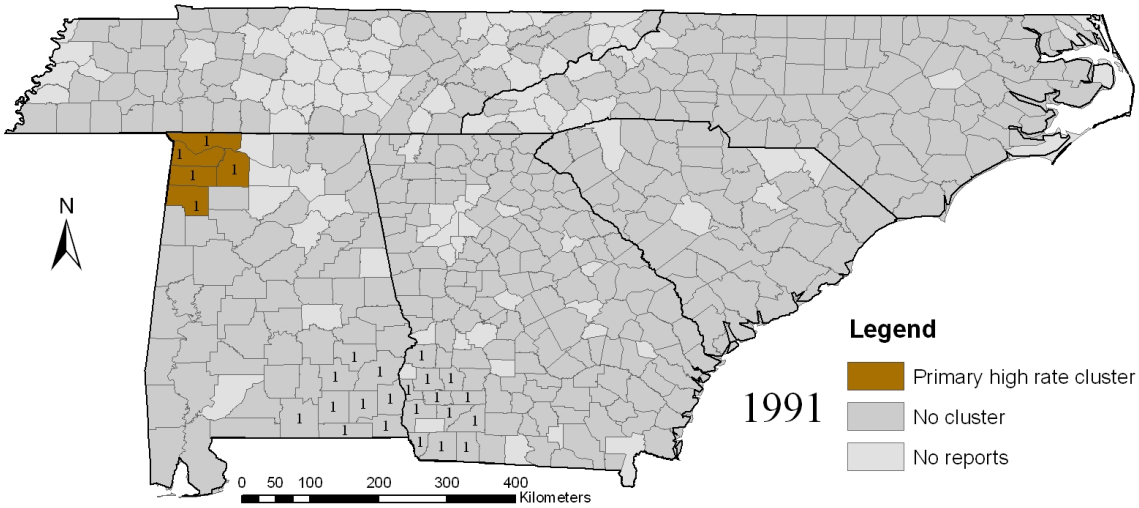


Figure 4.19: Purely spatial clustering analysis of HD in white-tailed deer (1991), numbers indicating high rate clusters, letters indicating low rate clusters.

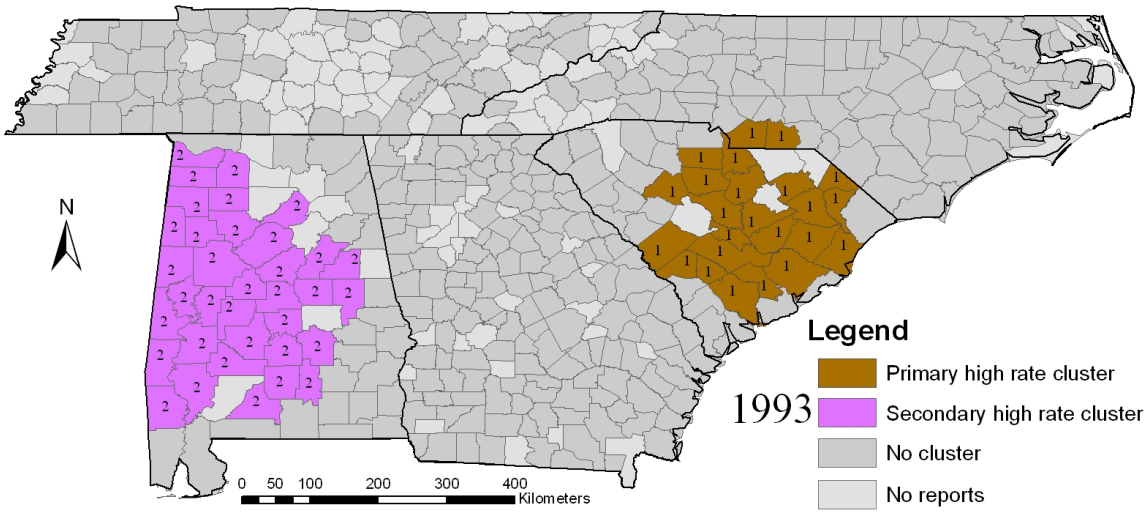


Figure 4.20: Purely spatial clustering analysis of HD in white-tailed deer (1993), numbers indicating high rate clusters, letters indicating low rate clusters.

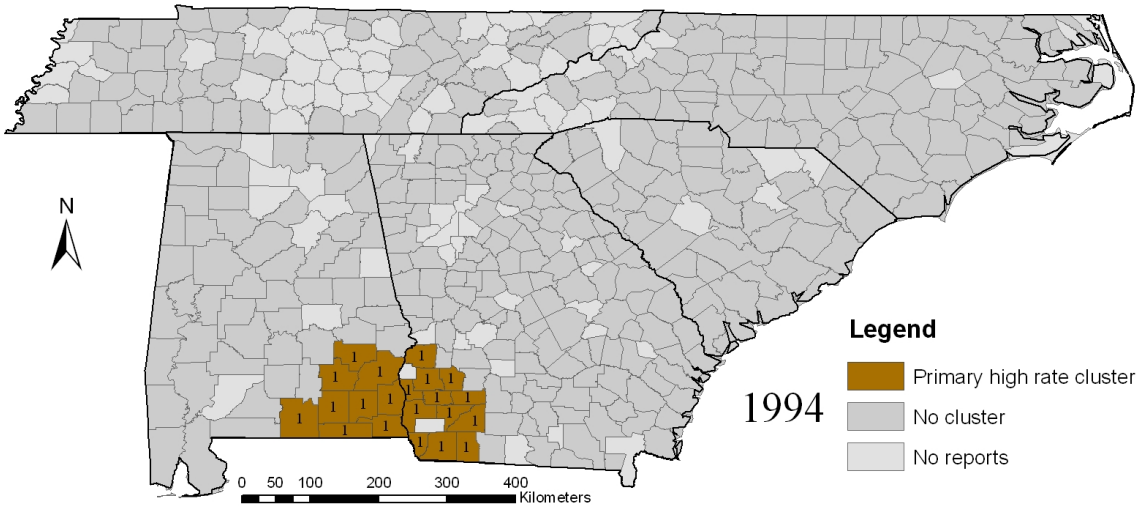


Figure 4.21: Purely spatial clustering analysis of HD in white-tailed deer (1994), numbers indicating high rate clusters, letters indicating low rate clusters.

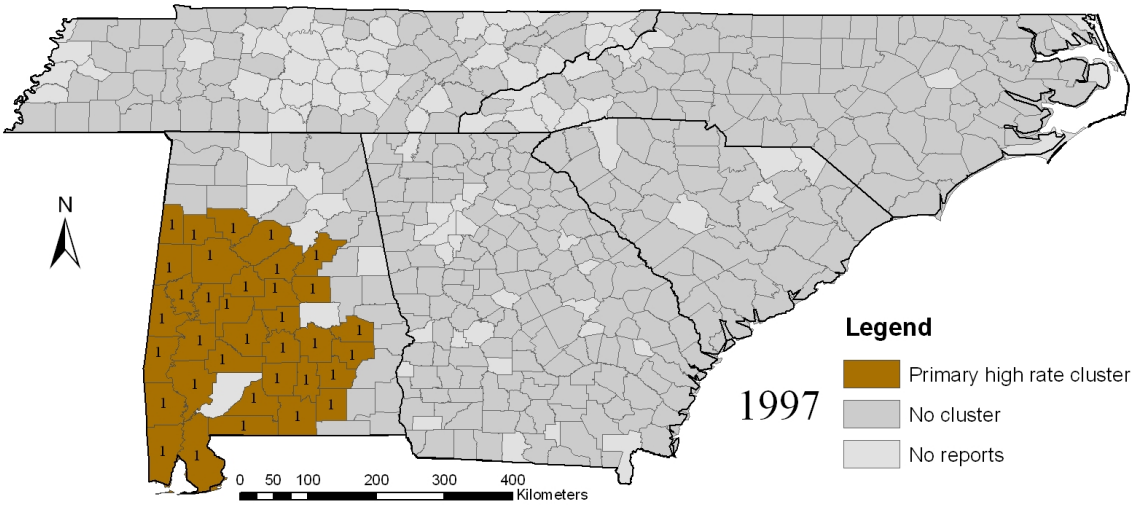


Figure 4.22: Purely spatial clustering analysis of HD in white-tailed deer (1997), numbers indicating high rate clusters, letters indicating low rate clusters.

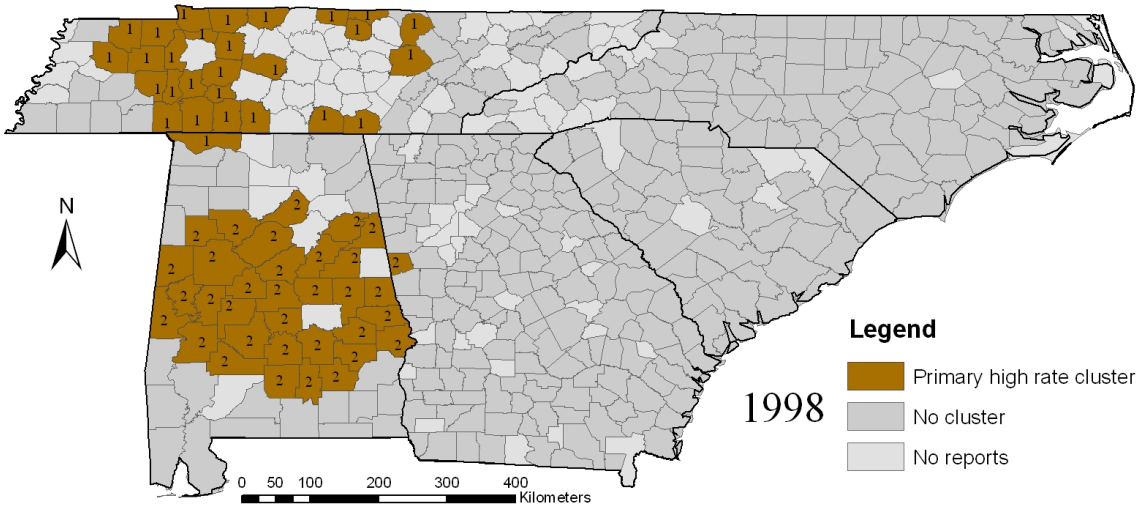


Figure 4.23: Purely spatial clustering analysis of HD in white-tailed deer (1998), numbers indicating high rate clusters, letters indicating low rate clusters.

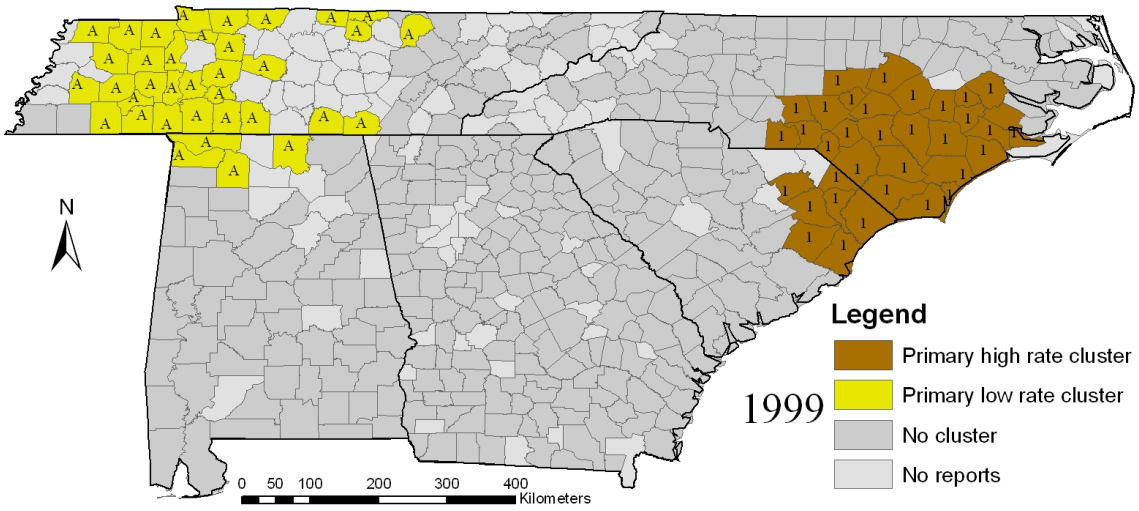


Figure 4.24: Purely spatial clustering analysis of HD in white-tailed deer (1999), numbers indicating high rate clusters, letters indicating low rate clusters.

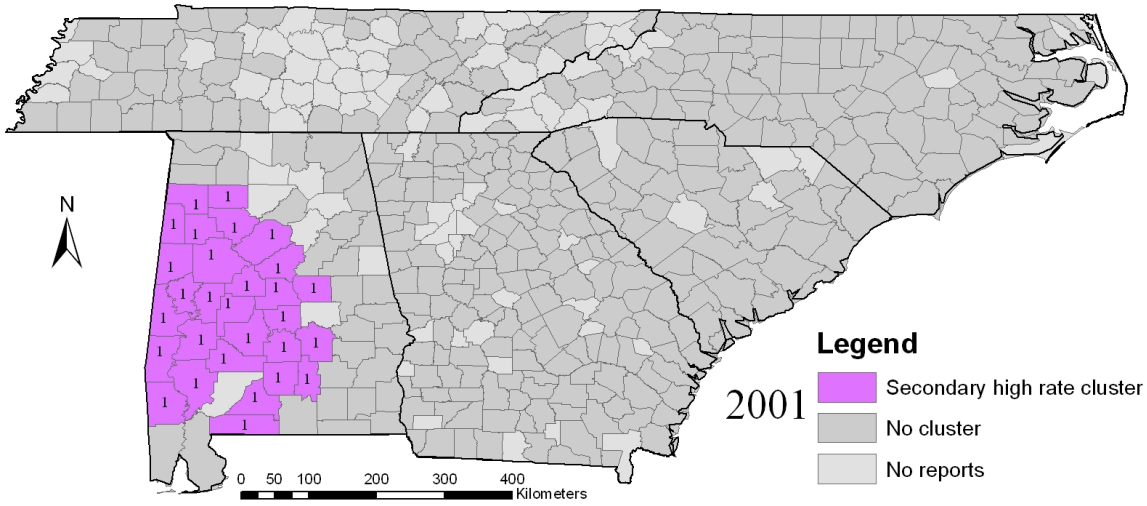


Figure 4.25: Purely spatial clustering analysis of HD in white-tailed deer (2001), numbers indicating high rate clusters, letters indicating low rate clusters.

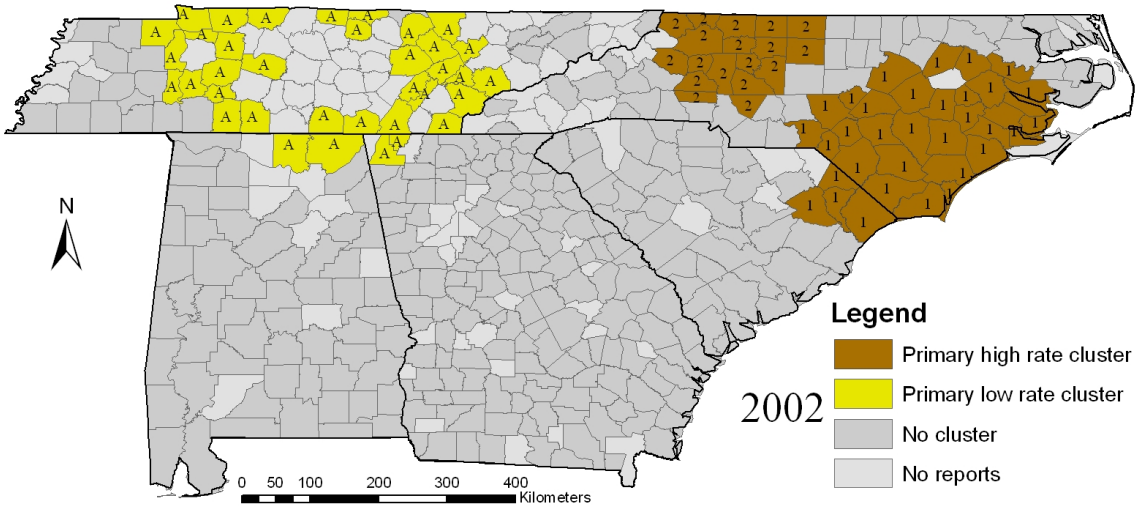


Figure 4.26: Purely spatial clustering analysis of HD in white-tailed deer (2002), numbers indicating high rate clusters, letters indicating low rate clusters.

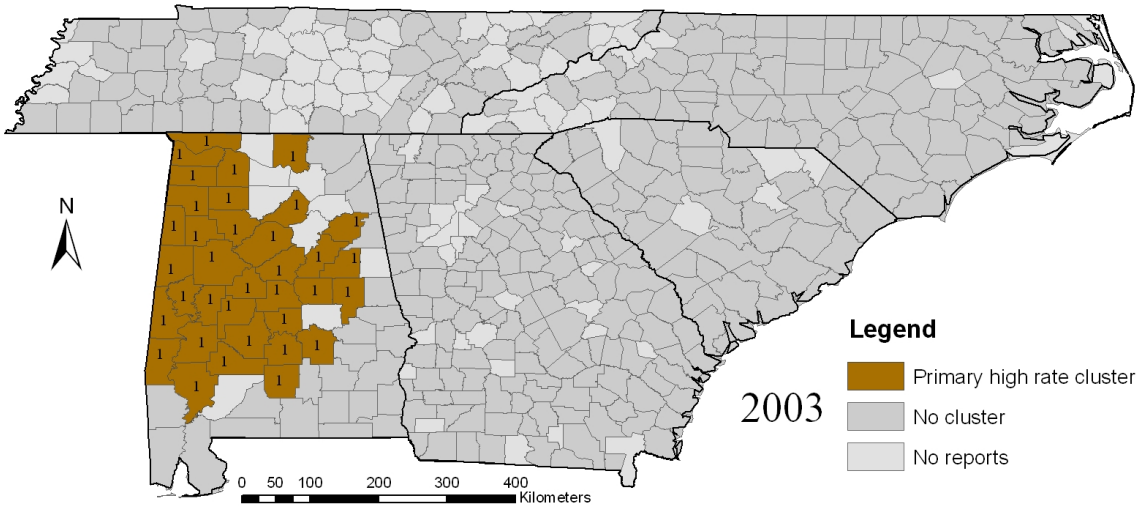


Figure 4.27: Purely spatial clustering analysis of HD in white-tailed deer (2003), numbers indicating high rate clusters, letters indicating low rate clusters.

along the coast in 1999, and 2002, and a high rate cluster in central South Carolina in 1980, 1986 and 1993. The low rate cluster in Tennessee also repeatedly appears in 1980, 1982, 1999, and 2002. I should be aware that for the first two high rate clusters which occur most often during the study period, their frequencies of reoccurring become higher in the last 10 years. These clusters are also identified in the previous purely spatial clustering analysis and space-time clustering analysis. By examining Situation 3 in space-time clustering analysis and the analysis in this subsection, I expect that Cluster 3 and Cluster 5 in Situation 3 should also be high rate clusters in 2002 and 1980, respectively, in the purely spatial clustering analysis by years. However, this is not the case, the reason is that in space-time clustering analysis, the detection of clusters is conditioned on all observations across the study area and during the study period. A cluster with a single year as its temporal period is the result of comparing the observed number of cases in that year with the number of the cases throughout the temporal period. In purely spatial cluster

analysis by individual year, the detection is only conditioned on the observations across the study area for each specific year. A space-time cluster with a single year as its temporal period, therefore, may not be reflected by the purely spatial cluster analysis in that year.

The three steps of cluster analysis: 1) purely spatial clustering analysis during the entire study area; 2) space-time clustering analysis; and 3) purely spatial clustering analysis by individual year, give a robust statistical analysis of cluster detection for HD in white-tailed deer in southeast USA from 1980-2003. The detection of clusters provides an initial understanding of the distribution of HD. After exploring when and where the clusters occur, further research is needed to investigate why the cluster exists in space and time. Although this issue is not within the scope of this study, some speculation will be made to discuss possible explanations for the spatial and temporal patterns of observed clusters.

In Chapter 3 of this dissertation, a statistical prediction model was constructed to predict HD occurrence, which identifies wind speed, rainfall, normalized difference vegetation index (NDVI), and land surface temperature calculated from channel 4 and channel 5 of the Advanced Very High Resolution Radiometer (AVHRR) to be useful to predict HD occurrence. The study applies principal component factor analysis to original explanatory data and used the resulting factor components as input to a later prediction model. The optimal prediction model includes Factor 3, Factor 4, Factor 7, Factor 8, Factor 9, Factor 10, spatial dependency and time as predictor variables (see Table 4.8 for the interpretation of the Factors). It may be hypothesized that the clusters detected in the current study may be related to those predictor variables. To roughly test this hypothesis, a purely spatial clustering analysis for the entire study period was

performed for each of theses variables to determine whether or not the resulting clusters coincide with HD clusters.

Due to the availability of predictive variables, the clustering analysis for predictive variables was conducted for 366 counties from 1982 to 2000. The spatial window was set to less than and equal to 10 percent of the total population. Figure 4.28 to Figure 4.34 show the results.

Table 4.8: Interpretation of factors

Factor	Climatic and remotely-sensed variable
3	yearly wind speed, period wind speed
4	yearly land surface temperature, period land surface temperature
7	period rainfall
8	period minimum rainfall, period minimum NDVI
9	Yearly minimum dew point, yearly minimum rainfall
10	period NDVI, yearly maximum wind speed

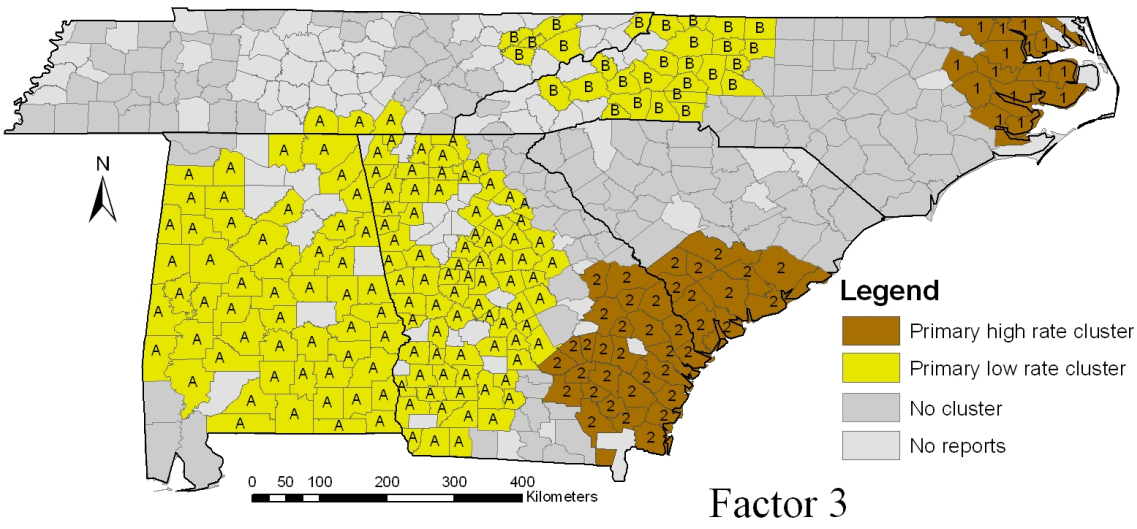


Figure 4.28: Purely spatial clustering analysis of Factor 3 (wind speed) during entire study period, numbers indicating high rate clusters, letters indicating low rate clusters.

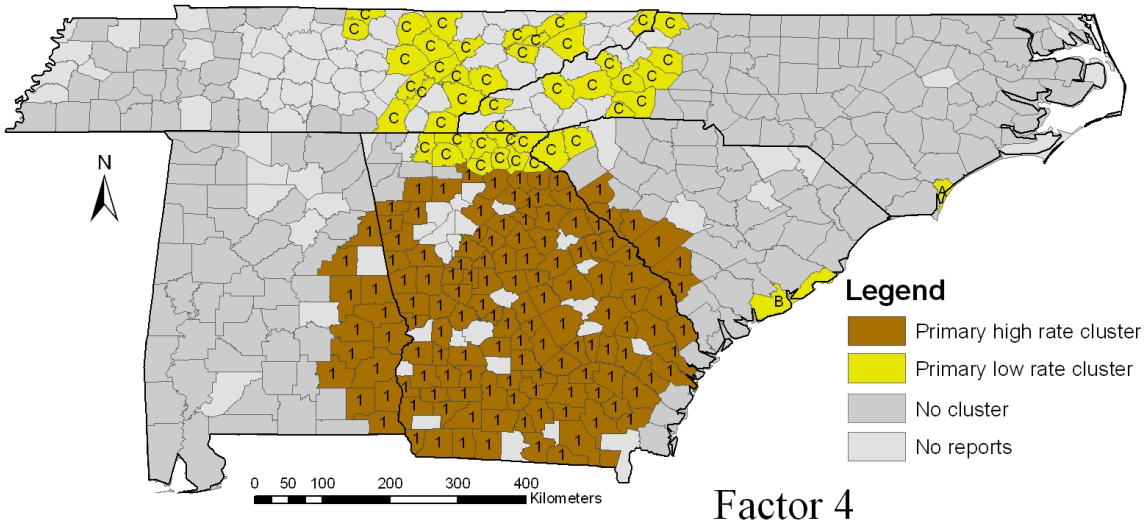


Figure 4.29: Purely spatial clustering analysis of Factor 4 (land surface temperature) during entire study period, numbers indicating high rate clusters, letters indicating low rate clusters.

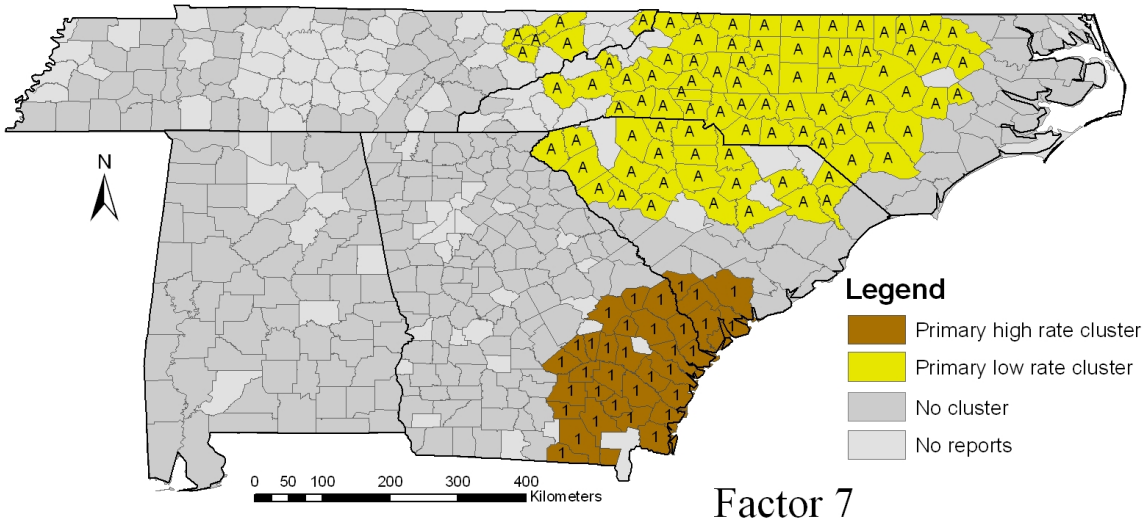


Figure 4.30: Purely spatial clustering analysis of Factor 7 (period rainfall) during entire study period, numbers indicating high rate clusters, letters indicating low rate clusters.

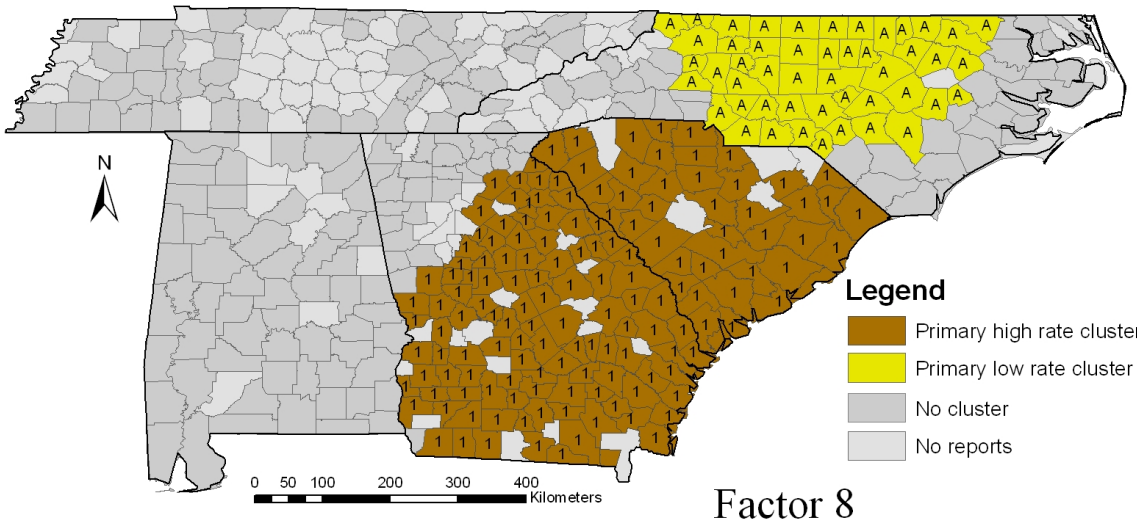


Figure 4.31: Purely spatial clustering analysis of Factor 8 (minimum rainfall) during entire study period, numbers indicating high rate clusters, letters indicating low rate clusters.

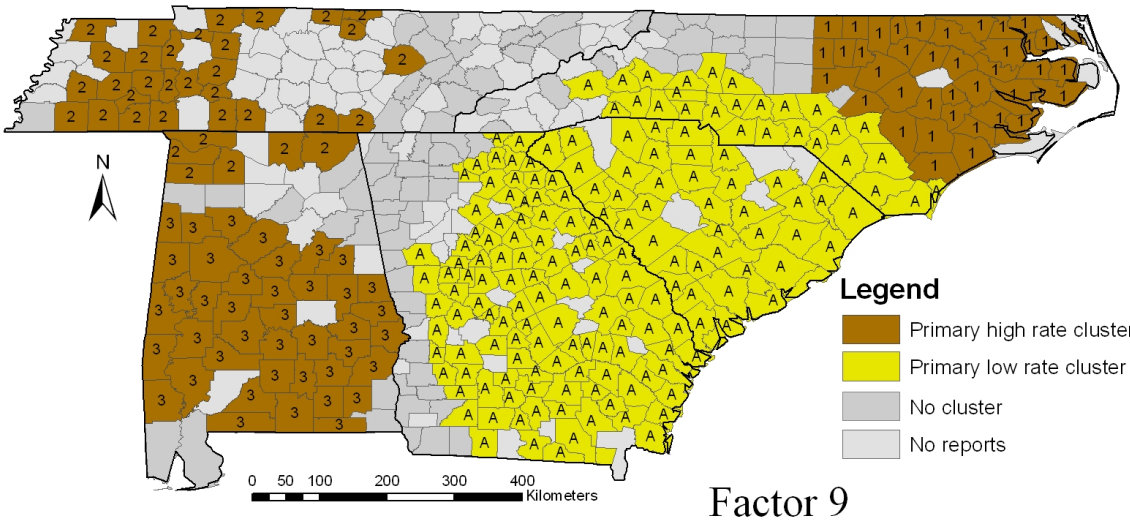


Figure 4.32: Purely spatial clustering analysis of Factor 9 (minimum dew point and minimum rainfall) during entire study period, numbers indicating high rate clusters, letters indicating low rate clusters.

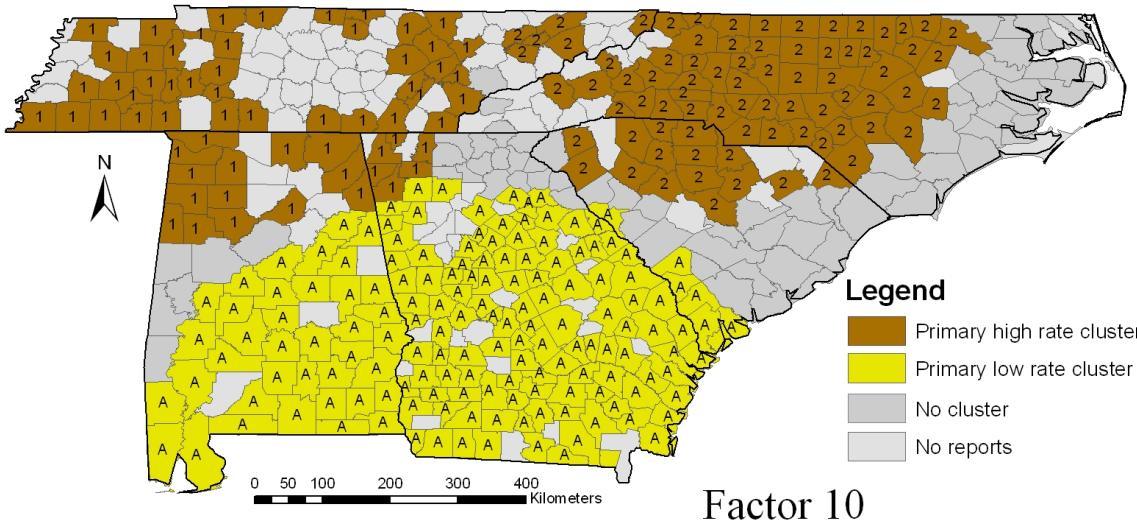


Figure 4.33: Purely spatial clustering analysis of Factor 10 (period NDVI and maximum wind speed) during entire study period, numbers indicating high rate clusters, letters indicating low rate clusters.

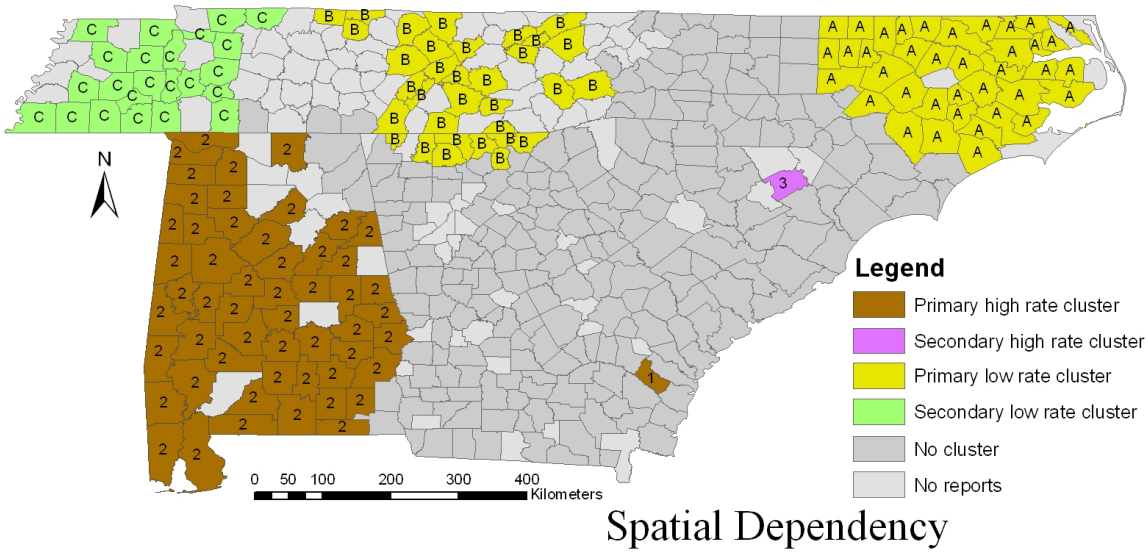


Figure 4.34: Purely spatial clustering analysis of spatial dependency during entire study period, numbers indicating high rate clusters, letters indicating low rate clusters.

By comparing the clustering results of these factors with HD clusters resulting also from the maximum spatial window of 10 percent of total population (Figure 4.4), I see, the high rate clusters of HD occurrence in west of Alabama and southeast of Alabama are noted to coincide with low rate clusters of Factor 3 (wind speed) (Figure 4.28), whereas the low rate cluster of HD occurrence in northeast of North Carolina approximately coincides with the high rate cluster of Factor 3. For Factor 10 (period NDVI and maximum wind speed) (Figure 4.33), the whole study area is roughly divided into two parts by the north boundary of Alabama, Georgia, and South Carolina. The north part is covered by high rate clusters of Factor 10, while the south part mainly contains low rate clusters. This is the opposite of the general pattern of HD occurrence observed in Figure 4.4. There is a strong association between the high and low rate clusters of spatial dependency (Figure 4.34) and the corresponding high and low rate clusters of HD clusters (Figure 4.4). The above conclusion is consistent with the previous study that Factor 3 and Factor 10 are negatively related to HD occurrence, while spatial dependency is positively related to HD occurrence. From visual examination, it seems the general trend of other predictive factors do not overlap with HD occurrence. Since Factor 3 mainly represents wind speed and Factor 10 is largely explained by NDVI, I can infer that wind speed, NDVI and spatial dependency may be related to the HD clustering. Although this is only a preliminary exploration, that should be followed by future detailed investigation on the reasons of HD clustering. These factors are expected to highly influence the distribution, abundance and availability of the Vector *Culicoides* midges and the spread of HD.

4.6 **Conclusion**

This study explores geographical and temporal clusters of HD in white-tailed deer in the five states from 1980 to 2003, using three cluster analysis methods that identify similar general clusters from different aspects. In the mean time, each cluster analysis approach reveals some unique information as well. Western and southern portion of Alabama, and along the north boundary between Georgia and South Carolina were detected as areas of high rate HD clusters by both purely spatial clustering analysis during the entire study period and space-time clustering analysis. Purely spatial clustering analysis for the entire study period further revealed several counties in central South Carolina as high rate clusters, and space-time clustering analysis identified the area at the boundary of South Carolina and North Carolina along the coast and the area in the west of North Carolina bordering Tennessee as high rate clusters during short temporal periods. Among those clusters, western and southern Alabama, along the boundary between South Carolina and North Carolina, and central South Carolina were also detected as repeatedly HD occurrence areas by the purely spatial clustering analysis by individual year. Although researchers tend to be most interested in high rate clusters, this study also detected some low rate clusters that are associated with areas of lower than expected HD occurrence. The one low rate cluster that was discovered by all three clustering approaches lies in Tennessee. However, it should be noted that this might indicate a problem of reporting in this area.

Space-time clustering analysis identifies more clusters than purely spatial clustering analysis in that the former can also detect clusters existing for a short period while the latter analysis only detect clusters for the entire study period. This research indicates that the spatial

and temporal clusters are restricted to 10 percent of the total population and 25 percent of the study period, respectively, which means clusters larger than these two thresholds rarely occur. This study also gives a rough investigation of possible reasons for HD clustering based on previous studies, and concludes that NDVI, wind speed and spatial dependency may be related to the HD clustering identified in this investigation.

References

Aamodt,G, Samuelsen S. O, and Skrondal, A. A., 2006. Simulation study of three methods for detecting disease clusters, *International Journal of Health Geographics*, 5:15.

Barton, D. E., and David, F. N., 1966. The random intersection of two graphs. *Research Papers in Statistics* (David, F. N., editor), New York: John Wiley $ Sons, Inc.

Carpenter, T. E., Hird, D. W., Snipes, K.P., 1996. A time-space investigation of the epidemiology of fowl cholera, *Preventive Veterinary Medicine*, 28: 159-163.

Couvillion, C. E., Davidson, W. R., Pearson, J. E., and Gustafson, G. A., 1981. Hemorrhagic disease among white-tailed deer from 1971 through 1980, *Proceedings of the United States Animal Health Association,* 85: 522-537.

Elliott, P., Wakefield, J. C., N. G, .Best, and Briggs, D. J., 2000. Spatial epidemiology: methods and applications, *Spatial Epidemiology: Methods and Applications* (P. Elliott, J. C.Wakefield, N. G. Best, and D. J. Briggs, editors), Oxford ; New York : Oxford University Press, 3-14.

Fuchs, K., Deuta, A. and Gressmann, G., 2000. Detection of space-time clusters and epidemiological examinations of scabies in chamois, *Veterinary Parasitology,* 92: 63-73.

Hoff, G. L. and Trainer, D. O., 1978. Bluetongue and epizootic hemorrhagic disease viruses: their relationship to wildlife species, *Advances in Veterinary Science and Comparative Medicine,* 22: 111-132.

Hoff, G. L. and Trainer, D. O., 1981. Hemorrhagic disease in wild ruminants, *Infectious diseases of wild mammals* (Davis, J. W., L. H. Karstad, and D. O. Trainer, editors) Ames: Iowa State University Press, 45-53.

Jacquez, G. M., 1996. A k nearest neighbor test for space-time interaction, *Statistics in Medicine,* 15: 1935-1949.

Knox, E. G., 1964. The detection of space-time interactions, *Applied Statistics,* 13: 25-29.

Kulldorff, M., 2006. SatScan User Guide for version 7.0, http://www.satscan.org/.

Kulldorff, M., Athas, W. F., Feuer, E. J., Miller, B. A. and Key, C. R., 1998. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American journal of Public Health,* 88: 1377-1380.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach, *Cancer Research,* 27: 209-220.

Nettles, V. F., Davidson, W. R., and Stallknecht, D. E., 1992. Surveillance for hemorrhagic disease in white-tailed deer and other wild ruminants, 1980-1989, *1992 Proceedings of Annual Conference of Southeastern Association of Fish and Wildlife Agencies*, 138-146.

Nettles, V. F. and Stallknecht, D. E., 1992. History and progress in the study of hemorrhagic disease of deer, *Transactions of the North American Wildlife and Natural Resources Conference,* 57: 499-516.

Norström, M., Pfeiffer, D. U. and Jarp, J., 2000. A space-time cluster investigation of an outbreak of acute respiratory disease in Norwegian cattle herds, *Preventive Veterinary Medicine*, 47: 107-119.

Paré, J., Carpenter, T. E., Thurmond, M. C., 1996. Analysis of spatial and temporal clustering of horses with Salmonella krefeld in an intensive care unit of a veterinary hospital*, Journal of the American Veterinary Medical Association*, 209: 626-628.

Prestwood, A. K., Kistner, T. P., Lellogg, F. E., and Hayes, F. A., 1974. The 1971 outbreak of hemorrhagic disease among white-tailed deer of the southeastern United States, *Journal of Wildlife Disease*, 10: 217-225.

Ruff, F. J., 1949. “Black Tongue” in deer, *Mississippi Game and Fish*, 13 (4): 10-14.

Singer, R. S., Case. J. T., Carpenter, T. E., Walker, R. L., and Hirsh, D. C., 1998. Assessment of spatial and temporal clustering of ampicillin- and tetracycline-resistant strains of Pasteruella multocida and P.haemolytica isolated from cattle in California, *Journal of the* American *Veterinary Medical Association,* 212: 1001-1005.

Smith, K. L., DeVos, V., Bryden, H., Price, L. B., Hugh-Jones, M. E., and Keim, P., 2000. *Bacillus anthracis* diversity in Kruger national park, *Journal of Clinical Microbiology*, 38 (10): 3780-3784

Song, C. and Kulldorff, M., 2003. Power evaluation of disease clustering tests, *International Journal of Health Geographics*, 2:9

Stallknecht, D. E., Kellogg, M. L., Blue, J. L., and Person, J. E., 1991. Antibodies to bluetongue and epizootic hemorrhagic disease viruses in a barrier island white-tailed deer population, *Journal of Wildlife Diseases*, 27(4), 668-674.

Thomas, F. C., Willis, N., and Ruckerbauer, G., 1974. Identification of viruses involved in the 1971 outbreak of hemorrhagic disease in southeastern United States white-tailed deer, *Journal of Wildlife Diseases*, 10: 187-189.

Trainer, D. O., 1964. Epizootic hemorrhagic disease of deer, *Journal of Wildlife Management,* 28: 377-381.

Ward, M. P., Armstrong, R. T. F., 2000. Time-space clustering of reported blowfly strike in Queensland sheep flocks, *Preventive Veterinary Medicine*, 43: 195-202.

Ward, M. P. and Carpenter, T. E., 2000a. Analysis of time-space clustering in veterinary epidemiology, *Preventive Veterinary Medicine,* 43: 225-237.

Ward, M. P. and Carpenter, T. E., 2000b, Techniques for analysis of disease clustering in space and in time in veterinary epidemiology, *Preventive Veterinary Medicine*, 45: 257-284.

White, M. E., Schukken, Y. H., Tanksley, B., 1989. Space-time clustering of, and risk factors for, farmer diagnosed winter dysentery in dairy cattle, *Canadian Veterinary Journal,* 30: 948-951.

CHAPTER 5

CONCLUSION

Hemorrhagic disease (HD) has resulted in significant death in white-tailed deer since it was first documented in the United States in the 1890s. This disease is caused by both bluetongue virus (BLU) and epizootic hemorrhagic disease virus (EHD), which are transmitted by *Culicoides* midges. Extensive work has been accomplished in traditional epidemiology of HD, i.e., the isolation of BLU and EHD and studies on the vectors. However, analysis of long term survey data of HD occurrences such as the development of prediction models and cluster detection were not explored, especially spatially and temporally. Prediction modeling and cluster detection are two main tasks in spatial epidemiology. They are crucial for possible cause discovery and future prevention of diseases. With the survey conducted by Southeastern Cooperative Wildlife Disease Study (SCWDS) in the College of Veterinary Medicine, University of Georgia since 1980, a county-based longitudinal data set representing HD presence/absence were available to perform the above two analyses. This research attempts to investigate spatial-temporal aspects of HD occurrence in the southeastern USA where HD in white-tailed deer are traditionally reported. Accordingly, this study is divided into two parts.

The first part of this dissertation (Chapter 3) constructs a spatial-temporal statistical model to predict HD occurrence based on the observed HD presence and absence for each county in Alabama, Georgia, South Carolina, North Carolina, and Tennessee from 1982 to 2000. The explanatory factors are climatic data including temperature, rainfall, dew point, wind speed and remotely-sensed data, including normalized difference vegetation index (NDVI) and land surface temperature. Principal component factor analysis was applied to 42 independent variables to reduce the data volume and remove correlations between variables. A subset of 11

principal factors were then used for later input as predictor variables. In this study, each county is repeatedly observed from 1982 to 2000, and these repeated measurements should not be considered independent of one another. Thus the generalized linear mixed logistic model forms the basis of the spatial-temporal model which considered the within-subject effects. The spatial dependency is incorporated by adding a spatial association term in the model which evaluates the effect of HD occurrence in surrounding counties on a particular county. The final results show that wind speed, rainfall, land surface temperature, spatial dependency, and NDVI can be used for the prediction of HD occurrence with an overall prediction accuracy of 65 percent. In this prediction model, remotely-sensed data proved to be useful and gives a higher prediction power than some continuous surface climatic data based on weather station point measurements interpolated by kriging.

The second part of this dissertation (Chapter 4) explores the spatial and space-time clusters of HD in the five states from 1980 to 2003. Kulldorff's space-time scan statistic was adopted to detect where and when the clusters occur, as well as the level of their statistical significance. Purely spatial clustering analysis for the entire study period, space-time clustering analysis, and purely spatial clustering analysis by individual year were applied. By varying the settings for maximum spatial windows and temporal windows for neighborhood analysis when detecting clusters of high (presence) and low (absence) rates of HD to look for a more realistic spatial window and temporal window, the results showed that the three analyses methods revealed generally similar clusters. The most evident high rate clusters are located in the western and southern portions of Alabama, central South Carolina, and around the boundary along South

Carolina and North Carolina. A maximum spatial window of 10 percent of the total population and a maximum temporal window of 25 percent of the study period are believed to be appropriate parameters that identify most of the clusters without leaving out subclusters. Spatial clusters exist for 15 years from 1980 to 2003. There are more counties identified as clusters in 1980, 1982, 1993, 1998, 1999, and 2002, whereas fewer counties in 1986, 1989, and 1991. Some clusters reoccur every several years, such as the high rate cluster in the west of Alabama which appears in 1983, 1993, 1997, 1998 and 2001, the high rate cluster in the south of Alabama in 1986, 1987, 1994, and 1997, a high rate cluster around the boundary between South Carolina and North Carolina along the coast in 1999, and 2002, and a high rate cluster in central South Carolina in 1980, 1986 and 1993. The low rate cluster in Tennessee also repeatedly appears in 1980, 1982, 1999, and 2002. It should be noted that the reoccurrence frequencies of the cluster in the west of Alabama and the cluster in the south of Alabama become higher in the last 10 years. By examining the clusters of those significant predictive factors in Chapter 3, NDVI, wind speed and spatial dependency are found to be related to the HD clustering.

There are some limitations in this study. Firstly, the original HD data were collected by questionnaires, and there may be variation in observer skill and reporting effort among the survey participants, resulting in either under or over-reporting. However, these data do provide a sound overview of HD distribution in white-tailed deer. Secondly, the HD data used in this research are binary data based on individual counties. The exact number of HD cases and their exact locations are not available. Thus some useful and mature techniques that can only be applied to continuous data (local Moran's I, local G-statistic, Moran scatterplot, some

visualization techniques such as time-plot, Parallel coordinate visualization, geographic weighted regression model, etc.) cannot be used to explore this unique 24-year national data set. Finally, statistical approaches were adopted to describe and predict the distribution of HD outbreaks as well as areas where HD occurrences in white-tailed deer are not common. Although total prediction accuracy was 65 percent when four environmental factors of wind speed, rainfall, land surface temperature, and NDVI were considered for a five-state area, prediction accuracies for individual years were quite variable ranging from 27 percent to 96 percent. It is hoped that future work building on the findings of this research will lead to a full, biologically and physically based understanding of the processes that underlie and cause the periodic outbreaks of HD and the morbidity/mortality of white-tailed deer. Prediction models and clustering analysis based on sound biological understanding of the epidemiological processes of HD combined with spatial-temporal statistical methods are expected to provide wildlife managers with critical information for management decisions.

As mentioned before, previous analysis of HD in white-tailed deer mainly focus on the isolation of viruses and the distribution of vectors. This dissertation provides new directions in the study of HD: prediction modeling and clustering analysis, emphasizing on the spatial-temporal distribution, which expands the study scope of HD in white-tailed deer. The prediction model and cluster detection for HD in southeast USA also offer directions for future research on HD distributions, such as why those clusters occur, what other information can be obtained in order to improve the predictive power of the model, and how to prevent future HD outbreaks. Partnerships between veterinary scientists, wildlife managers, epidemiologists and

geospatial researchers will result in a more thorough understanding of this wide spread and economically influential wildlife disease. The inclusion of remote sensing demonstrates that readily available and relatively inexpensive remotely-sensed data can be used in epidemiology and can result in better disease prediction. More generally, this study contributes to the spatial epidemiology area in human diseases which continuously grows in importance with increasing human travel of globe. Contributions of this research are especially important in the veterinary spatial epidemiology area, which currently lacks attention on clustering detection and prediction models, let alone the more complicated spatial-temporal studies on these issues. The integration of GIS, remote sensing, and statistical methods provides entirely new information to traditional veterinary epidemiology that is spatially explicit, historical as well as current and readily available to all users via the internet and from various GIS data clearinghouses. The statistical methods used in geographic applications are often simple models including ordinary linear models, logistic models, and discriminant models. Recently, more models are proposed taking spatial dependency into consideration such as geographic weighted models. However, few spatial-temporal models are applied in geographic applications, especially for binary data. This dissertation provides some thoughts on this area, and the prediction results show that the generalized linear mixed logistic model proves to be useful in predicting the spatial-temporal HD occurrence in white-tailed deer in southeast USA.

CONSOLIDATED REFERENCES

Aamodt, G., Samuelsen S. O., and Skrondal, A., 2006. A simulation study of three methods for detecting disease clusters, *International Journal of Health Geographics*, 5 (15).

Augustin, N. H., Mugglestone, M. A., and Buckland, S. T., 1996. An autologistic model for the spatial distribution of wildlife, *The Journal of Applied Ecology*, 33 (2): 339-347.

Barton, D. E., and David, F. N., 1966. The random intersection of two graphs. *Research Papers in Statistics* (David, F. N., editor), New York: John Wiley $ Sons, Inc.

Baylis, M., Bouayoune, H., Touti, J., and Hasnaoui, H., 1998. Use of climatic data and satellite imagery to model the abundance of *Culicoides imicola*, the vector of African horse sickness virus, in Morocco. *Medical and Veterinary Entomology,* 12: 255-266.

Baylis, M., Meiswinkel, R., and Venter, G. J., 1999. A preliminary attempt to use climate data and satellite imagery to model the abundance and distribution of *Culicoides imicola* (Diptera: Ceratopogonidae) in southern Africa. *Journal of the South African Veterinary Association,* 70: 80-89.

Baylis, M., Mellor, P. S., Wittmann, E. J., and Rogers, D. J., 2001. Prediction of areas around the Mediterranean at risk of bluetongue by modeling the distribution of its vector using satellite imaging, *Veterinary Record,* 149: 639-643.

Baylis, M. and Rawlings, P., 1998. Modeling the distribution and abundance of *Culicoides imicola* in Morocco and Iberia using climatic data and satellite imagery, *African Horse sickness*, Arcives of Virology (Supplement) 14 (P. S. Mellor, M. Baylis, C. Hamblin, C. Calisher, and P. P. C. Mertens, editors), Vienna, Springer,137-153.

Beck, L. R., Rodriguez, M. H., Dister, S. W., Rodriguez, A. D., Washino, R. K., Roberts, D. R., and Spanner, M. A., 1997. Assessment of a remote sensing-based model for predictiong malaria transimission risk in villages of Chiapas, Mexico, *American Journal of Tropical Medicine and Hygiene*, 56(1): 99-106.

Bernardinelli, L. and Montomoli, C., 1992. Empirical bayes versus fully Bayesian analysis of geographical variation in disease risk, *Statistics in Medicine,* 11: 983-1007.

Best, N. and Wakefield, J. R., 1999. Accounting for inaccuracies in population counts and case registration in cancer mapping studies, *Journal of the American Statistical Association*, 95: 1076-1088

Bohning, D., Dietz, E., and Schlattmann, P., 2000. Space-time mixture modeling of public health data. *Statistics in Medicine*, 19: 2333-2344.

Boone, J. D., McGwire, K. C., Otteson, E. W., DeBaca, R. S., Kuhn, E. A., Villard, P., Brussard, P. F., and St. Jeor, S. C., 2000. Remote sensing and geographic information systems: Charting Sin Nombre Virus infections in deer mice, *Emerging Infectious Diseases*, 6(3): 248-258.

Burnham, K. P. and Anderson, D. R., 2004. Multimodel inference: understanding AIC and BIC in model selection, *Sociological Methods and Research*, 33: 261-304.

Burrough, P. A., 1986. *Principles of Geographical Information Systems for Land Resources Assessment*, Oxford: Oxford University Press.

Carpenter, T. E., Hird, D. W., Snipes, K.P., 1996. A time-space investigation of the epidemiology of fowl cholera, *Preventive Veterinary Medicine*, 28: 159-163.

Cattell, R. B., 1966. The scree test for the number of factors, *Multivariate Behavioral Research*, 1: 245-276.

Chaput, E. K., Meek, J. I., Heimer, R., 2002. Spatial analysis of human granulocytic ehrlichiosis near Lyme, *Connecticut. Emerging Infectious Diseases*, 8: 943-948.

Clayton, D. and Kaldor, J., 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43: 671-682.

Clements A. C. A., Lwambo, N. J. S., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J. P., Fenwick, A., and Brooker, S., 2006. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a *schistosomiasis* control programme in Tanzania, *Tropical Medicine and International Health*, 11(4): 490-503.

Congdon, P., 1994. Spatiotemporal analysis of area mortality, *Statistician*, 43: 513-528.

Cooper, D. R. and Asrar, G., 1989. Evaluating atmospheric correction models for retrieving surface temperatures from the AVHRR over a tallgrass prairie, *Remote Sensing of Environment*, 27: 93-102.

Couvillion, C. E., Davidson, W. R., Pearson, J. E., and Gustafson, G. A., 1981. Hemorrhagic disease among white-tailed deer from 1971 through 1980, *Proceedings of the United States Animal Health Association,* 85: 522-537.

Cross E. R., Newcomb, W. W., and Tucker C. J., 1996. Use of weather data and remote sensing to predict the geographic and seasonal distribution of Phlebotomus papatasi in southwest Asia, *American Journal of Tropical Medicine and Hygiene*, 54: 530-536.

Doll, R. and Wakeford, R., 1997. Risk of childhood cancer from fetal irradiation, *British Journal of Radiology*, 70: 130-139.

Durr, P. A., 2004. Spatial epidemiology and animal disease: introduction and overview. *GIS and Spatial Analysis in Veterinary Science* (P. A. Durr and A. C. Gatrell, editors), Wallingford, UK ; Cambridge, MA, CABI Pub, 35-67.

Durr, P. A. and Gatrell, A. C., 2004. The tools of spatial epidemiology: GIS, spatial analysis and remote sensing. *GIS and Spatial Analysis in Veterinary Science* (P. A. Durr and A. C. Gatrell, editors), Wallingford, UK ; Cambridge, MA, CABI Pub, 1- 33.

Duchatear, L., Kruska, R. L., and Perry, B. D., 1997. Reducing a spatial database to its effective dimensionality for logistic-regression analysis of incidence of livestock disease, *Preventive Veterinary Medicine,* 56: 51-62.

DuToit, R. M., 1944. The transmission of bluetongue and African horse-sickness by *Culicoides*, *Onderstepoort Journal of Veterinary Science and Animal Industry,* 19: 7-26.

Eisa, M., Karrar, A. E., and Elrahim, A. H., 1979. Incidence of bluetongue virus precipitating antibodies in sera of some domestic animals in the Sudan, *Journal of Hygiene*, 83: 539-545.

Elliott, P., Wakefield, J. C., N. G, .Best, and Briggs, D. J., 2000. Spatial epidemiology: methods and applications, *Spatial Epidemiology: Methods and Applications* (P. Elliott, J. C.Wakefield, N. G. Best, and D. J. Briggs, editors), Oxford ; New York : Oxford University Press, 3-14.

Elliott, P., and Wartenberg, D., 2004. Spatial epidemiology: current approaches and future challenges, *Environmental Health Perspectives*, 112: 998-1006.

Erasmus, B. J., 1975. The epizootiology of bluetongue: the African situation, *Australian Veterinary Journal*, 51: 196-203.

Estrada-Peña, A., 1997. Epidemiological surveillance of tick populations: a model to predict the colonization success of Ixodes ricinus (Acari: Ixodidae), *European Journal of Epidemiology,* 13: 581-586.

Estrada-Peña, A., 1999. Geostatistics and remote sensing using NOAA-AVHRR satellite imagery as predictive tools in tick distribution and habitat suitability estimations for *Boophilus microplus* (Acari: Ixodidae) in South America, *Veterinary Parasitology,* 81: 73-82.

Fienberg, S. E., Bromet, E. J., Follmann, D., Lambert, D., and May, S. M., 1985. Longitudinal analysis of categorical epidemiological data: a study of three mile island, *Environmental Health Perspectives*, 63: 241-248.

Foster, N. M., Breckon, R. D., Luedke, A. J., and Jones, R. H., 1977. Transmission of two strains of epizootic hemorrhagic disease virus in deer by *Culicoides variipennis*, *Journal of Wildlife Disease*, 13: 9-16.

Foster, N. M., Jones, R. H., and McC rory, B. R., 1963. Preliminary investigations on insect transmission of bluetongue virus in sheep, *American Journal of Veterinary Research,* 24: 1195-1200.

Fuchs, K., Deuta, A. and Gressmann, G., 2000. Detection of space-time clusters and epidemiological examinations of scabies in chamois, *Veterinary Parasitology,* 92: 63-73.

Ghebreyesus, T. A., Byass, P., Witten, K. H., Getachew, A., Haile, M., Yohannes, M., and Lindsay, S. W., 2003. Appropriate Tools and Methods for Tropical Microepidemiology: a Case-study of Malaria Clustering in Ethiopia, *Ethiopian Journal of Health Development*, 17:1-8.

Ghosh, M., Natarajan, K., Stroud, T., and Carlin, B. P., 1998. Generalized linear models for small-area estimation, *Journal of the American Statistical Association*, 93: 273-282.

Glass, G. E., Cheek, J. E., Patz, J. A., Shields, T. M., Doyle, T. J., Thoroughman, D. A., Hunt, D. K., Enscore, R. E., Gage, K. L., Irland, C., Peters, C. J., and Bryan, R., 2000. Using remotely sensed data to identify areas at risk for hantavirus pulmonary syndrome, *Emerging Infectious Disease*, 6(3): 238-247.

Greiner, E. C., Barber, T. L., Pearson, J. E., Kramer, W. L., and Gibbs, E. P .J., 1985. Orbiviruses from *Culicoides* in Florida, *Bluetongue and Related Orbiviruses* (Barer, T. L., M. M. Jochim, and B. I. Osburn, editors), New York: Alan R. Liss, 195-200.

Greiner, E. C., Mo. C. L., Homan. E. J., Gonzalez, J., Oviedo, M., Thopson, L. H., and Gibbs, E. P. J., 1993. Epidemiology of bluetongue in Central America and the Caribbean: Initial entomological findings, *Medical and Veterinary Entomology,* 7: 309-315.

Hales, S., Wet, N. E., Maindonald, J., and Woodward, A., 2002. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model, *The Lancet*, 360: 830 – 834.

Handcock, M. S. and Wallis, J. R., 1994. An approach to statistical spatial-temporal modeling of meteorological fields, *Journal of the American Statistical Association*, 89: 368-390.

Hartzel, J., Agresti, A., and Caffo, B., 2001. Multinomial logit random effects models, *Statistical Modeling,* 1: 81-102.

Hay, S. I., 2000. An overview of remote sensing and geodesy for epidemiology and public health application, *Advances in Parasitology*, 47: 1-35.

Hay, S. I., Tucker, C. J. Rogers, D. J., and Packer, M. J., 1996. Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease, *Annals of Tropical Medicine and Parasitology,* 90(1): 1-19.

Hedeker, D., 2003. A mixed-effects multinomial logistic regression model, *Statistics in Medicine*, 22: 1433-1446.

Hill, T., and Lewicki, P., 2007. *Statistics: Methods and Applications*, StatSoft, Tulsa, OK

Hinely, A. J., 2006. GIS-based Habitat Modeling Related to Bearded Capuchin Monkey Tool Use, Master thesis, the University of Georgia, Athens, Georgia.

Hoff, G. L. and Trainer, D. O., 1972. Bluetongue virus in pronghorn antelope, *American Journal of Veterinary Research,* 33: 1013-1016.

Hoff, G. L. and Trainer, D. O., 1974. Observations on bluetongue and epizootic hemorrhagic disease in white-tailed deer: (1) distribution of virus in the blood (2) cross-challenge, *Journal of Wildlife Diseases,* 10: 25-31.

Hoff, G. L. and Trainer, D. O., 1978. Bluetongue and epizootic hemorrhagic disease viruses: their relationship to wildlife species, *Advances in Veterinary Science and Comparative Medicine,* 22: 111-132.

Hoff, G. L. and Trainer, D. O., 1981. Hemorrhagic disease in wild ruminants, I*nfectious Diseases of Wild Mammals* (Davis, J. W., L. H. Karstad, and D. O. Trainer, editors) Ames: Iowa State University Press, 45-53.

Homan, E. J., Mo, C. L., Thompson, L. H., Barreto, C. H., Oviedo, M. T., Gibbs, E. P. J., and Greiner, E. C., 1990. Epidemiologic study of bluetongue viruses in Central America and the Caribbean: 1986-1988, *American Journal of Veterinary Research*, 51: 1089-1094.

Jacquez, G. M., 1996. A k nearest neighbour test for space-time interaction, *Statistics in Medicine*, 15: 1935-1949.

Jensen, R. E., 1967. A multiple regression model for cost control: assumptions and limitations, *The Accounting Review*, 42: 265-273.

Jones, R. H. and Foster, N. M., 1974. Oral infections of *Culicoides variipennis* with bluetongue virus: Development of susceptible and resistant lines from a colony population. *Journal of Medical Entomology*, 11: 316-323.

Jones, R. H., Roughton, R. D., Foster, N. M., and Bando, B. M., 1977. *Culicoides*, the vector of epizootic hemorrhagic disease in white-tailed deer in Kentucky in 1971, *Journal of Wildlife Disease,* 13: 2-8.

Justice, C. O., Townshend, J. R. G., Holben, B. N., and Tucker, C. J., 1985. Analysis of the phenology of global vegetation using meteorological satellite data, *International Journal of Remote Sensing*, 10: 1539-1561.

Kaiser, H. F., 1960. The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, 20: 141-151.

Klauber, M. R., 1971. Two-sample randomization tests for space-time clustering, *Biometrics*, 27, 129-142.

Kleinschmidt, I., Bagayoko, M., Clarke, G., Craig, M., and Sueur, D. L., 2000. A spatial statistical approach to malaria mapping. *International Journal of Epidemiological*, 29: 355-361.

Knorr-Held, L. and Besag, J., 1998. Modeling risk from a disease in time and space, *Statistics in Medicine*, 17: 2045-2060.

Knox, E. G., 1964. The detection of space-time interactions, *Applied Statistics,* 13: 25-29.

Kulldorff, M., 1998. Statistical methods for spatial epidemiology: tests for randomness, *GIS and Health* (Gatrell, A. and M., Lötönen, editors), Taylor and Francis, London, 49-62.

Kulldorff, M., 2006. SatScan User Guide for version 7.0, http://www.satscan.org/.

Kulldorff, M., Athas, W. F., Feuer, E. J., Miller, B. A. and Key, C. R., 1998. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American journal of Public Health,* 88: 1377-1380.

Lauer, R. M., Clarke, W. R., and Burns, T. L., 1997. Obesity in childhood: the Muscatine study, *Acta Padiatrica Scandinavica,* 38: 432-437.

Lawson, A. B. 2000. Cluster modeling of disease incidence via RJMCMC methods: a comparative evaluation, *Statistics in Medicine*, 19: 2361-2376.

Lawson, A. B., 2001. *Statistical Methods in Spatial Epidemiology*, New York : John Wiley.

Lessard, P., L'Eplattenier, R., Norval, R. A. I., Kundert, K., Dolan, T. T., Croze, H., Walker, J. B., Irvin, A. D. and Perry, B. D., 1990. Geographical information systems for studying the epidemiology of cattle diseases caused by *Theileria parva*, *Veterinary Record,* 126: 255-262.

Lillesand, T. M. and Kiefer, R. W., 1999. *Remote Sensing and Image Interpretation*, Fourth edition, John Wiley & Sons, Inc.

Lindsay, S. W., and Thomas, C. J., 2000. Mapping and estimating the population at risk from lymphatic filariasis in Arica, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94: 37-45.

Linn, R. L., 1968. A Monte Carlo approach to the number of factors problem, *Psychometrika*, 33: 37-71.

Linthicum, K. J., Bailey, C. L., Davies, F. G. and Tucker, C. J., 1987. Detection of Rift Valley fever viral activity in Kenya by satellite remote sensing imagery, *Science,* 235: 1656-1659.

Lo, C.P. and Yeung, A. K.W., 2002. *Concepts and Techniques in Geographic Information Systems*, Upper Saddle River, N.J. : Prentice Hall.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach, *Cancer Research,* 27: 209-220.

McCullagh, P., 1980. Regression models for ordinal data (with discussion), *Journal of the Royal Statistical Society*, Series B, 42: 109-142.

McKenzie, J. S., Morris, R. S., Pfeiffer, D. U. and Dymond, J. R., 2002. Application of remote sensing to enhance the control of wildlife-associated *Mycobacterium bovis* infection, *Photogrammetric Engineering and Remote Sensing,* 68: 153-159.

Meiswinkel, R., Nevill, E. M., and Venter, G. J., 1994. Vectors: *Culicoides spp*., *Infectious Diseases of Livestock with Special Reference to Southern Africa*, Volume 1 (Coetzer, J. A. W., G. R. Thomson, and R. C. Tustin, editors), Cape Town: Oxford Un Iversity Press.

Mellor, P. S., 1996. *Culicoides*: vectors, climate change and disease risk, *Veterinary Bulletin* 66: 301-306.

Miller, J. H. and Robinson, K. S., 1995. A regional perspective of the physiographic provinces of the southeastern United States, *Proceedings of the eighth biennial southern silvicultural research conferece*, November 1-3, 1994, Auburn, AL. 581-591.

Mo, C. L., Thompson, L. H., Homan, E. J., Oviedo, M. T., Greiner, E. G., Gonzalez, J., and Saenz, M. R., 1994. Bluetongue virus isolations from vectors and ruminants in Central America and the Caribbean, *American Journal Veterinary Research*, 55: 211-215.
Mugglin, A., Cressie, N., and Gemmel, I., 2002. Hierarchical statistical modeling of influenza epidemic dynamics in space and time, *Statistics in Medicine*, 21: 2703-2721.

Mullen, G. R., Hayes, M. E., and Nusbaum, K. E., 1985. Potential vectors of bluetongue and epizootic hemorrhagic disease viruses in cattle and white-tailed deer in Alabama, *Bluetongue and Related Orbiviruses* (Barber, T. L., M. M. Jochim, and B. I. Osburn, editors), New York: Alan R. Liss, 201-206.

Nettles, V. F., Davidson, W. R., and Stallknecht, D. E., 1992. Surveillance for hemorrhagic disease in white-tailed deer and other wild ruminants, 1980-1989, *1992 Proceedings of Annual Conference of Southeastern Association of Fish and Wildlife Agencies*, 138-146.

Nettles, V. F. and Stallknecht, D. E., 1992. History and progress in the study of hemorrhagic disease of deer, *Transactions of the North American Wildlife and Natural Resources Conference,* 57: 499-516.

Nevill, E. M., 1971. Cattle and *Culicoides* biting midges as possible overwintering hosts of bluetongue virus, *Journal of Veterinary Research*, 38: 65-72.

Norström, M., Pfeiffer, D. U. and Jarp, J., 2000. A space-time cluster investigation of an outbreak of acute respiratory disease in Norwegian cattle herds, *Preventive Veterinary Medicine*, 47: 107-119.

Ollerenshal, C. B., 1966. The approach to forecasting the incidence of fascioliasis over England and Wales 1958-1962, *Agricultural Meteorology,* 3: 35-53.

Osburn, B. I., McGowan, B., Heron, B., Lommis, E., Bushnell, R., Stott, J., and Utterback, W., 1981. Epizootiologic study of bluetongue: Virologic and serologic results, *American Journal of Veterinary Research,* 42: 884-887.

Paré, J., Carpenter, T. E., Thurmond, M. C., 1996. Analysis of spatial and temporal clustering of horses with Salmonella krefeld in an intensive care unit of a veterinary hospital, *Journal of the* American *Veterinary Medical Association*, 209: 626-628.

Pfeiffer, D. U., 2000. Spatial analysis – a new challenge for veterinary epidemiololgists, *Proceedings of the Annual Meeting of Society for Veterinary Epidemiology & Preventive Medicine, Edinburgh 29th-31st March, 2000* (Thrusfield, M. V. and E. A. Goodall, editors), Society for Veterinary Epidemiology and Preventive Medicine, Edinburgh, 86-106.

Pfeiffer, D. U., Duchateau, L., Kruska, R. L., Ushewokunze-Obatolu, U. and Perry, B. D., 1997. A spatially predictive logistic regression model for the occurrence of theileriosis outbreaks in Zimbabwe, *Proceedings of the VIII International Symposium on Veterinary Epidemiology & Economics,* Paris, 8-11 July, 12.12.1-12.12.3.

Pickle, L.W., 2000. Exploring spatio-temporal patterns of mortality using mixed effects models, *Statistics in Medicine*, 19: 2251-2263.

Pinkel, D. and Nefzger, D., 1959. Some Epidemiological Feature of childhood leukemia in the Buffalo, N. Y. Area. *Cancer*, 12: 351-359.

Pinkel, D., Dowd, J. E., and Bross, I. D. J., 1963. Some Epidemiological features of Malignant solid tumors of children in the Buffalo, N. Y., *Cancer*, 16: 28-33.

Prestwood, A. K., Kistner, T. P., Lellogg, F. E., and Hayes, F. A., 1974. The 1971 outbreak of hemorrhagic disease among white-tailed deer of the southeastern United States, *Journal of Wildlife Disease*, 10: 217-225.

Price, D. A. and Hardy W. T., 1954. Isolation of the bluetongue virus from Texas sheep Culicoides shown to be a vector, *Journal of American Veterinary Medical Association,* 124(925): 255-258.

Roberts, D. R., Paris, J. F., Manguin, S., Harbach, R. E., Woodruff, R., Rejmankova, E., Polanco, J., Wullschleger, B., and Legters, L. J., 1996. Predictions of malaria vector distribution in belize based on multispectral satellite data, *American Journal of Tropical Medicine and Hygiene*, 54(3): 304-308.

Rogan, W. J., Dietrich, K. N., Ware, J. H., Dockery, D. W., Salganik, M., Radcliffe, J., Jones, R. L., Ragan, N. B., Chisolm, J. J., and Rhoads, G. G.., 2001. The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead, *New England Journal of Medicine*, 344: 1421-1426.

Rogers, D. J. and Randolph, S. E., 1991. Mortality rates and population density of tsetse flies correlated with satellite imagery, *Nature*, 351: 739-741.

Rogers, D. J., Hay, S. I., and Packer, M. J., 1996. Predicting the distribution of tsetse flies in west Africa using temporal Fourier processed meteorological satellite data, *Annals of Tropical Medicine and Parasitology*, 90(3): 225-241.

Rogers, D. J. and Williams B. G., 1993. Monitoring trypanosomiasis in space and time. *Parasitology*, 106: 77-92.

Ruff, F. J., 1949. "Black Tongue" in deer, *Mississippi Game and Fish*, 13 (4): 10-14.

Sheehan, T. J., Gershman, S. T., MacDougal, L., Danley, R., Mrosszczyk, M., Sorensen, A. M., and Kulldorff, M., 2001. Geographical surveillance of breast cancer screening by tracts, towns and zip codes, *Journal of Public Health Management and Practice*, 6: 48-57.

Singer, R. S., Case. J. T., Carpenter, T. E., Walker, R. L., and Hirsh, D. C., 1998. Assessment of spatial and temporal clustering of ampicillin- and tetracycline-resistant strains of Pasteruella multocida and P.haemolytica isolated from cattle in California, *J. Journal of the* American *Veterinary Medical Association*, 212: 1001-1005.

Smith, K. E., Stallknecht, D. E., Sewell, C. T., Rollor, E. A., Mullen, G. R., and Anderson, R .R., 1996. Monitoring of *Culicoides* spp. At a site enzootic for hemorrhagic disease in white-tailed deer in Georgia., USA., *Journal of Wildlife Diseases,* 32: 627-642.

Smith, K. L., DeVos, V., Bryden, H., Price, L. B., Hugh-Jones, M. E., and Keim, P., 2000. *Bacillus anthracis* diversity in Kruger national park, *Journal of Clinical Microbiology*, 38 (10): 3780-3784

Smith, P. G., Pike, M. C., Till, M. M. and Hardisty, R. M., 1976. Epidemiology of childhood leukemia in Greater London: A search for Evidence of Transmission assuming a possible long latent period, *British Journal of Cancer*, 33: 1-8.

Snow, R. W., Gouws, E., Omumbo, J., Rapuoda, B., Craig, M. H., Tanser, F. C., Sueur, D. L., and Ouma, J., 1998. Models to predict the intensity of Plasmodium falciparum transmission: applications to the burden of disease in Kenya, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 92: 601-606.

Solna, K. and Switzer, P., 1996. Time trend estimation for a geographic region, *Journal of the American Statistical Association*, 91: 577-589.

Song, C. and Kulldorff, M., 2003. Power evaluation of disease clustering tests, *International Journal of Health Geographics*, 2:9

St..George, T. D., Standfast, H. A., Cybinski, D. H., Dyce, A. L., Muller, M. J., Doherty, R. L., and Carley, J. G., 1978. The isolation o a bluetongue virus from *Culicoides* collected in the northern territory of Australia, *Australia Veterinary Journal,* 54: 153-154.

Stallknecht, D. E. and Howerth, E. W., 2004. Epidemiology of bluetongue and epizootic haemorrhagic disease in wildlife: surveillance methods, *Vet. Ital*., 40(3): 203-207.

Stallknecht, D. E., Howerth, E. W., and Gaydos, J. K., 2002. Hemorrhagic disease in white-tailed deer: Our current understanding of risk, *Transactions of the 67th North American Wildlife and Natural Resourcs Conference,* Dallas, Texas, 75-86.

Stallknecht, D. E., Kellogg, M. L., Blue, J. L., and Person, J. E., 1991. Antibodies to bluetongue and epizootic hemorrhagic disease viruses in a barrier island white-tailed deer population, *Journal of Wildlife Diseases*, 27(4), 668-674.

Stallknecht, D. E., Luttrell, M. P., Smith, K. E., and Nettles, V. F., 1996. Hemorrhagic disease in white-tailed deer in Texas: A case for enzootic stability, *Journal of Wildlife Diseases,* 32: 695-700.

Stukel, T. A. 1993. Comparison of methods for the analysis of longitudinal interval count data, *Statistics in Medicine,* 12: 1339-1351.
Sugita, M. and Brutasaert, W., 1993. Comparison of land surface temperatures derived from satellite observations with ground truth during FOFE, *International Journal of Remote Sensing,* 14: 1659-1676.

Sun, D., Tsutakawa, R., Kim, H., and He, Z., 2000. Spatio-temporal interaction with disease mapping, *Statistics in Medicine*, 19: 2015-2035.

Thomas, F. C., Willis, N., and Ruckerbauer, G., 1974. Identification of viruses involved in the 1971 outbreak of hemorrhagic disease in southeastern United States white-tailed deer, *Journal of Wildlife Diseases*, 10: 187-189.

Thomson, M. C., and Conner, J., 2000. Environmental information systems for the control of arthropod vectors of disease, *Medical and Veterinary Entomology*, 14: 227-344.

Thomson, M. C., Connor, S. J., D'alessandro, U., Rowlingson, B., Diggle, p., Cresswell, M., and Greenwood, B., 1999. Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results, , *American Journal of Tropical Medicine and Hygiene*, 61(1): pp. 2-8.

Trainer, D. O., 1964. Epizootic hemorrhagic disease of deer, *Journal of Wildlife Management,* 28: 377-381.

Tucker, L. R., Koopman, R. F., and Linn, R. L., 1969. Evaluation of factor analytic research procedures by means of simulated correlation matrices, *Psychometrika*, 34: 421-459

Tucker, C. J., Vanpraet, C., Boerwinkel, E. and Gaston, A., 1983. Satellite remote sensing of total dry matter production in the Senegalese Sahel, *Remote Sensing of Environment*, 13: 461-474.

Twisk, Jos W. R., 2003. *Applied Longitudinal Data Analysis for Epidemiology: A practical Guide*. Cambridge, U.K.; New York: Cambridge University Press.

Van der Linde, A., Witzko, K-H., and Jockel, K-H., 1995. Spatial-temporal analysis of mortality using splines, *Biometrics*, 51: 1352-1360.

Van Marter, L, J., Leviton, A., Kuban, K. C. K., Pagano, M. and Allred, E. N., 1990. Maternal glucocorticoid therapy and reduced risk of bronchopulmonary dysplasia, *Pediatrics*, 86: 331-336.

Walker, A. R., and Davies, F. G., 1971. A preliminary survey of the epidemiology of bluetongue in Kenya, *Journal of Hygiene*, 69: 47-60.
Walter, L. A., 2000. Disease mapping: a historical perspective, *Spatial Epidemiology: Methods and Applications* (Elliott, P. J, Wakefield, N, Best, and D. J. Briggs, editors), Oxford: Oxford University Press, 223-252.

Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E., 1997. Hierarchical spatio-temporal mapping of disease rates, *Journal of the American Statistical Association*, 92: 607-617.

Ward. M. P., 1994. Climatic factors associated with the prevalence of bluetongue virus infection of cattle herds in Queensland, Australia, *Veterinary Record,* 134: 407-410.

Ward, M. P., Armstrong, R. T. F., 2000. Time-space clustering of reported blowfly strike in Queensland sheep flocks, *Preventive Veterinary Medicine*, 43: 195-202.

Ward, M. P. and Carpenter, T. E., 2000a. Analysis of time-space clustering in veterinary epidemiology, *Preventive Veterinary Medicine,* 43: 225-237.

Ward, M. P. and Carpenter, T. E., 2000b, Techniques for analysis of disease clustering in space and in time in veterinary epidemiology, *Preventive Veterinary Medicine*, 45: 257-284.

White, M. E., Schukken, Y. H., Tanksley, B., 1989. Space-time clustering of, and risk factors for, farmer diagnosed winter dysentery in dairy cattle, *Canadian Veterinary Journal,* 30: 948-951.

Wieser-Schimpf, L., Wilson, W. C., French, D. D., Heidner, H. W., and Foil, L. D., 1993. Bluetongue virus in sheep and cattle and *Culicoides* variipennis and C. stellifer (Diptera: Ceratopogonidae) in Louisiana, *Journal of Medical Entomology*, 30: 719-724.

Wittmann E. J. and Baylis, M., 2000. Climate change: effects on *Culicoides*-transmitted viruses and implications for the UK, *The Veterinary Journal*, 160: 107-117.

Wood, B. L., Beck, L. R., Washino, R. K., Hibbrard, K. A., and Salute J. S., 1991. Estimating high mosquito-producing fields using spectral and spatial data, *International Journal of Remote Sensing*, 12: 621-626.

Xia, H., and Carlin, B. P., 1998. Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality, *Statistics in Medicine,* 17: 2025-2043.

Zhu, L. and Carlin, B. P., 2000. Comparing hierarchical models for spatio-temporally misaligned data using the deviance information criterion, *Statistics in Medicine*, 19; 2265-2278.