2006

Evaluating the spatial ecology of anthrax in North America: examining epidemiological components across multiple geographic scales using a GIS-based approach

Jason Kenna Blackburn

Louisiana State University and Agricultural and Mechanical College, jblack6@lsu.edu

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Social and Behavioral Sciences Commons

Recommended Citation

https://digitalcommons.lsu.edu/gradschool_dissertations/2425

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
EVALUATING THE SPATIAL ECOLOGY OF ANTHRAX IN NORTH AMERICA:
EXAMINING EPIDEMIOLOGICAL COMPONENTS ACROSS MULTIPLE GEOGRAPHIC SCALES USING A GIS-BASED APPROACH

A Dissertation

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the Requirements for the degree of Doctor of Philosophy

In

The Department of Geography and Anthropology

by

Jason Kenna Blackburn
B.S., Louisiana State University, 2001
M.S., Louisiana State University, 2003
May 2006
“Man, wow, there’s so many things to do, so many things to write! How to even begin to get it all down and without modified restraints and all hung-up on like literary inhibitions and grammatical fears…” Dean Moriarity

From *On the Road* by Jack Kerouac
To Mom, Dad, and both my families-

The Blackburn’s and the Holley’s

Special thanks to Todd and Ashley Hymel for introducing me to the great state of Louisiana, it will forever be my adopted home

In loving memory of

Ken Moore

(1945 – 1995)

*He gave us more in a short time than many give in a life time... And gave us all one hell of an adventure to get him home...*
ACKNOWLEDGEMENTS

First, I want to personally thank Dr. Andrew Curtis, my major professor and lab director. Dr. Curtis first persuaded me to stay at LSU to pursue graduate studies after meeting in his undergraduate Digital Cartography and Introductory GIS classes in 2000. Not only did Dr. Curtis find a place for me in the department, he provided me with the ultimate freedom; freedom to pursue other disciplines, other interests, new techniques, and a wide range of methodologies. Dr. Curtis allowed me to develop a research program that bridged my interests in wildlife, ecology, disease, and spatial analysis. This has been a real gift and I am a better researcher because of it. Dr. Curtis has always been open-minded to my various interests, patient with deadlines, flexible with methodologies, and supportive of my fieldwork. He has become a mentor, a collaborator, a co-author, and a friend. Thanks, Chief.

Dr. Martin Hugh-Jones and I first began working together after I literally grabbed him in the vet school parking lot late one afternoon in the spring of 2003 to review my Master’s thesis and join my graduate committee as that degree approached the two minute warning. He graciously accepted, provided useful insight into antibiotic resistance in sharks, and we have been running hard on new methodologies for fieldwork and GIS analyses to further our understanding of anthrax ever since. In the three years we have worked together I have had the great pleasure of traveling, discussing, reviewing, revising, rethinking, brainstorming, and studying one of the great diseases with one of the disease’s great pioneers.

Dr. Mark Mitchell has been another influential collaborator and professor during my tenure at LSU. From hand capturing free swimming nurse sharks and searching high and low for whale sharks in Belize, dogfish tracking in New England, bull shark netting in the Gulf of Mexico, to tortoise mapping in Louisiana, Mark and I have had some good old fashion
adventures, and have managed to come home with data nearly every time. Dr. Mitchell has supported many of my projects both financially and intellectually. He has offered great insight into all of my research projects and has always promoted trying new techniques and approaches; Mark truly appreciates thinking outside of the box. Additionally, I am forever indebted to Mark for introducing me to my wife. It is impossible to think of life any differently.

Dr. Michael Leitner has also provided great insights, new ideas, and fresh perspectives on this and other research. He is always willing to open his door at any hour he is in and entertain discussion on GIS topics and spatial analyses.

Dr. Patrick Hesp was willing to take on a biogeography/GIS/Medical Geography hybrid project with little hesitation. He has a wealth of knowledge on the effects of spatial phenomena on environmental processes and is not afraid to talk ecology in a geographic context. Dr. Hesp also has a wide range of field experience that proved useful in developing the environmental data collection for Chapter 3.

Dr. Kenneth McMillin served as the Dean’s representative for this committee and showed patience with me and interest in the work. He provided useful insight into the development and completion of the project and dissertation.

I also want to thank Dr. Bruce A. Thompson of the LSU Coastal Fisheries Institute and Dr. Kam Biu-Liu from Geography. Bruce was my Master’s committee co-advisor and continues to be a great friend and collaborator. While I now focus much of my attention on the geography of wildlife and zoonotic diseases, Bruce and I continue to collaborate on fisheries projects and scheme on the uses of GIS to further our knowledge of fish ecology in Louisiana. Bruce has always opened his home to me and truly is a great friend. Dr. Liu first introduced me to the
department and though we have differing philosophies on the discipline of biogeography, I owe him for opening the door to Geography at LSU.

This work would not have been possible without the help and support of several individuals. I would like to thank Mr. Buba Culp and his family for the long hours, extra effort and tireless help to complete the fieldwork in Texas. Likewise, Roy Hurley, Shane Gonzales, Mark Gonzales, and Gerardo Boquin all provided the help necessary to successfully capture and release white-tailed deer. Without them, there would have been no telemetry project. Dr. Hesp provided the four-wheeler used to complete the fly sampling work and without it, we may still be out in Texas trying to complete the sampling. Projects that require extensive fieldwork can only be completed with a network of people. I am lucky to have had a group so dedicated.

Dr. Kristina McNyset, of the US EPA (WED/ORD) in Corvalis, Oregon provided a tremendous amount of help and guidance on the implementation, interpretation and fundamental principles of the GARP modeling approach. The work with Dr. McNyset has lead to not only better work here, but a fruitful collaboration and plans for future work together both in disease modeling and my other scientific passion, fish ecology. Dr. David Rogers, of the TALA Research Group at Oxford University, provided great insights into modeling paradigms, accuracy metrics, a wealth of literature, and also provided the NDVI data set used in Chapters 2 and 3. Mr. Ricardo Scachetti-Pereira developed the DesktopGARP implementation of the GARP algorithm. Without his efforts, this study would not have been possible. The open source programming community has much to offer the ecological modeling and GIS communities. Their work is largely unfunded, yet critically important. They deserve much recognition.

Data sets were provided from a number of sources for the development of the continental data set used in Chapter 2 and 3. Dr. Neil Dyer provided North Dakota outbreak locations, Drs.
Russ Daly and Mike Wimberly provided localities for South Dakota. Mr. David Wiklund provided data from Minnesota. The Texas data was a combination of data from WHOCC field investigations and the Centers for Disease Control outbreak response in 2001. CDC data were provided to Dr. Martin Hugh-Jones. Dr. Lane Foil and the LSU Entomology department provided insights into Nzi trap design, fly identification, and several discussions on the role of flies in the anthrax cycle. Dr. Ted Hadfield collaborated on anthrax testing in deer and flies for these projects.

I would like to add a special thanks to my families. My parents Scott and Joanne Blackburn have always supported my academic endeavors. It has been a long haul from Grossmont College to the completion of this degree and they have always been there. This would not be possible without them. They have also been supportive of the long-distance haul from my native Boulevard, California to my adopted home in Louisiana. My move to Louisiana was not only academic, it was permanent. My parents have welcomed this move and exerted a great deal of effort to make it part of their home as well. I also want to thank the Holley family. Since meeting their daughter in Mark Mitchell’s office in the spring of 2002, it has been non-stop. They have always been supportive and have welcomed me into the family with a full embrace. Eric, Cathie, Trevor, Grandma, and Grandpa have also ridden out a long road with the brother/uncle/grandson off chasing sharks one day, tunas the next, deer the next and country hopping along the way. Thanks to all of you for all the miles.

Finally, none of this would have happened without the love of my life, Mary Claire Holley Blackburn, my wife. Without the love, support, and tolerance for thousands of dead flies, long days, late nights, lonely weeks-on-end apart, and her skill as a veterinarian, the transition
from sharks to zoonotic diseases would not have seemed so normal. It is true what Van Morrison says, “She gives me love, love, love, love, crazy love…” Thanks, Baby.

Funding was provided by the LSU WHOCC Laboratory, the Civilian Research and Development Foundation, the RV Ranch, the International Aquatic and Terrestrial Conservation Medicine and Biotelemetrics Laboratory, and private support from the Blackburn Family, the Holley Family, and Dr. Martin Hugh-Jones.
TABLE OF CONTENTS

DEDICATION ........................................................................................................... iii

ACKNOWLEDGEMENTS........................................................................................... iv

LIST OF TABLES ....................................................................................................... xi

LIST OF FIGURES ..................................................................................................... xii

ABSTRACT ................................................................................................................ xiv

CHAPTER 1. INTRODUCTION.................................................................................... 1
  1.1 Introduction ....................................................................................................... 1
  1.2 Expected Significance ....................................................................................... 6
  1.3 References ....................................................................................................... 8

CHAPTER 2. PREDICTING THE DISTRIBUTION OF ANTHRAX, *Bacillus anthracis*, IN THE CONTIGUOUS UNITED STATES ........................................10
  2.1 Introduction ....................................................................................................... 10
  2.2 Materials and Methods .................................................................................. 26
  2.3 Results ............................................................................................................ 34
  2.4 Discussion ....................................................................................................... 45
  2.5 References ....................................................................................................... 59

CHAPTER 3 PREDICTING THE DISTRIBUTION OF ANTHRAX IN MEXICO .................. 64
  3.1 Introduction ....................................................................................................... 64
  3.2 Materials and Methods .................................................................................. 65
  3.3 Results and Discussion .................................................................................. 69
  3.4 References ....................................................................................................... 74

CHAPTER 4 SPATIO-TEMPORAL PATTERNS OF HEMATOPHAGOUS FLIES: POTENTIAL VECTORS FOR ANTHRAX INFECTION ........................................... 77
  4.1 Introduction ....................................................................................................... 77
  4.2 Materials and Methods .................................................................................. 80
  4.3 Results ............................................................................................................ 85
  4.4 Discussion ....................................................................................................... 91
  4.5 References ....................................................................................................... 94

CHAPTER 5 THE CASE-MULTIPLIER HYPOTHESIS: POSITIVE CONFIRMATION OF ANTHRAX SPORES IN NECROPHILIC FLIES IN WEST TEXAS ............. 97
  5.1 Introduction ....................................................................................................... 97
  5.2 Methods and Results ...................................................................................... 98
  5.3 References ....................................................................................................... 99
LIST OF TABLES

Table 2.1  The confusion matrix used to derive error metrics in ecological modeling……17
Table 2.2  Specific locality data used to derive GARP models……………………………..27
Table 2.3  Environmental coverages used to derive GARP models……………………………..29
Table 2.4  Model sample sizes and accuracy metrics for GARP model development and validation………………………………………………………………………36
Table 2.5  Dominant rules from the 10 models in the final model best subset (Model 1)…………………………………………………………………………39
Table 2.6  Character loadings and percentage of variance explained by principal components (PC) I and II………………………………………………………53
Table 2.7  Character loadings and percentage of variance explained by principal components (PC) I and II………………………………………………………54
Table 2.8  Descriptive statistics for environmental variables at 156 known anthrax sites………………………………………………………………………………54
Table 2.9  Descriptive statistics grouped by individual state for 156 unique anthrax locations………………………………………………………………………………55
Table 4.1  Descriptive statistics for the three time periods analyzed in this study………….85
Table 6.1  Summary of individual deer tracked during this study……………………………110
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Conceptual flow chart of the ecological components that represent anthrax</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>Environmental coverages used to develop GARP models for anthrax in the</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>continental United States.</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>Distribution of case localities (red) and 5,000 random locations (gray)</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>from the lower 48 states used to construct an environmental envelope for</td>
<td></td>
</tr>
<tr>
<td></td>
<td>anthrax.</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>GARP outputs for anthrax in the U.S.</td>
<td>37</td>
</tr>
<tr>
<td>2.4</td>
<td>GARP prediction for potential distribution of anthrax, <em>Bacillus anthracis</em></td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>in the continental United States.</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>Maps of dominant presence (red color ramp – P) and absence (blue color</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>ramp – A) rules for each of the top 10 best models in the subset.</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Map displaying a selection of the dominant rule set assignment for a single</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>task.</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>Decadal persistence of anthrax by county.</td>
<td>49</td>
</tr>
<tr>
<td>2.8</td>
<td>Counties positive for anthrax from 1915 – 1955.</td>
<td>50</td>
</tr>
<tr>
<td>2.9</td>
<td>Counties positive for anthrax from 1956 – 2005.</td>
<td>51</td>
</tr>
<tr>
<td>2.10</td>
<td>Overlay of positive anthrax counties 1996 – 2005.</td>
<td>52</td>
</tr>
<tr>
<td>2.11</td>
<td>Principal components (PC) I and II based on six environmental variables</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>for 4,622 locations randomly derived from across the lower 48 states.</td>
<td></td>
</tr>
<tr>
<td>2.12</td>
<td>Principal components (PC) I and II based on six environmental variables</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>for 156 anthrax locations.</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Predicted distribution of anthrax in the U.S. and Mexico.</td>
<td>71</td>
</tr>
<tr>
<td>3.2</td>
<td>Close-up of the GARP predictions for anthrax in Mexico.</td>
<td>72</td>
</tr>
<tr>
<td>3.3</td>
<td>Rule set map of dominant presence rules (red) and absence rules for</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Mexico.</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Nzi fly trap setup.</td>
<td>84</td>
</tr>
</tbody>
</table>
Figure 4.2  Kernel density derived hotspots for biting flies during the anthrax season..................................................................................................................87
Figure 4.3  Density surfaces over base maps.................................................................88
Figure 4.4  Interpolated wind speeds from recordings at trap sites.................................89
Figure 4.5  $G_i^*$ cluster locations at 500m and 2000m..................................................90
Figure 6.1  Deer population estimates between 1995 and 2005....................................108
Figure 6.2  Anthrax case numbers between 2001 and 2005 on a study ranch in west Texas.................................................................109
Figure 6.3  MCP’s (red lines) for seven deer tracked between 30 June and 22 August 2005.............................................................................................................111
Figure 6.4  MCP’s (red lines) for seven deer tracked between 30 June and 22 August 2005.............................................................................................................112
Figure 6.5  Location of home ranges relative to protein feeders.................................113
Figure 6.6  A collared deer capture by a motion-sensitive camera.................................114
Figure 7.1  An expanded version of Figure 1.1...............................................................122
ABSTRACT

This dissertation explores the spatial ecology and potential pathways of infection of anthrax, Bacillus anthracis, in North America. A multi-scale approach was used to evaluate the components required for disease agent survival in the environment, interactions with wildlife, and the potential role that vectors play in anthrax transmission. First, ecological niche modeling with the Genetic Algorithm for Rule-set Production (GARP) was used to predict the geographic distribution of anthrax in the continental U.S. using case data from outbreaks between 1957 and 2005. These results were then used to produce the first quantitative, continental scale predictions of anthrax in Mexico. At the meso-scale, the route of transmission in white-tailed deer is unknown, despite a large number of outbreaks in wild deer in Texas in recent years (2001 – 2005). To determine the interactions between deer and potential anthrax sources, two pilot studies were conducted on 1) the distribution of biting flies in relation to anthrax cases to evaluate the potential role of hematophagous flies as vectors, and 2) the summer home ranges of deer in relation to fly densities and carcass locations. The results of the GARP studies support the use of the technique for modeling the niche of this disease and suggest a central corridor of anthrax habitat from southwest Texas to the Canadian border, with disjunct areas in the Pacific Northwest and California. Mexico’s predicted areas were extensions of the Texas and California ranges. The deer study suggests that deer interactions with spores occur within a limited home range in Texas and long-distance movement of spores is unlikely by individual deer. Biting fly densities were highest in areas of known anthrax infection and lowest in areas where case-positive deer have not been identified, suggesting that flies may play a role in disease transmission, either through mechanical transmission or through increased nuisance that leads to immuno-suppression in deer. This dissertation presents the first continental-scale predictions for
the geographic distribution of anthrax in the U.S. and Mexico. Additionally, this is the first known study to evaluate spatial patterns between known cases, fly densities, and animal movements.
CHAPTER 1 INTRODUCTION

1.1 Introduction

This dissertation is a multi-scaled exploration of the spatial ecology and geographic distribution of anthrax in North America. Anthrax remains an important zoonotic disease in livestock and wildlife nearly worldwide (Hugh-Jones and de Vos 2002). Despite being a disease of antiquity, the spatial ecology of Bacillus anthracis, the etiological agent of anthrax, remains poorly understood (Gates et al. 1995, Smith et al. 2000, Hugh-Jones and De Vos 2002). B. anthracis is a Gram-positive, rod-shaped, soil-borne bacterium, which primarily affects herbivorous ungulates and secondarily affects humans (van Ness 1971, Hugh-Jones and de Vos 2002). The classical infection pathway of anthrax is through direct interaction between the bacterium and the affected host species, most likely through soil ingestion or inhalation (Gates et al. 1995). However, there is evidence that secondary infection pathways such, as arthropod transmission, through both hematophagous and necrophilic insects, and less so, direct contact between individual hosts can play a role in outbreak promotion (Gates et al. 1995, Hugh-Jones and De Vos 2002). In any proposed transmission mechanism, there must be interaction between multiple ecological players. First, B. anthracis must be present in the environment. This indicates that the environment must be able to sustain the etiological agent within the physiological constraints of the species’ ecological niche (Grinell 1917). Second, the affected ungulate species has to interact with the environment, also indicating that the host and disease agent must share ecological and geographic space. Third, if insects can secondarily transmit the organism between hosts, there must be an overlap between these species. Consequently, there is a complex series of interactions between species and environments in the epidemiology of anthrax infections in wildlife or livestock in natural environments. This study aims to
compartmentalize these potential interactions and identify the specific geographical space in which these interactions occur.

In mammalian hosts, anthrax exists and replicates as vegetative cells (Gates et al. 1995). Vegetative cells can release proteins that, when combined become toxic to the host, and maybe fatal in herbivorous ungulates. Bacilli are then released back into the environment post-death when scavengers open the infective carcass. When vegetative cells are exposed to oxygen or low nutrient environments, they form a hard exosporium that is resistant to desiccation. These spores can survive long periods of time in soil, until the appropriate environmental conditions provide exposure to an appropriate host species (sporulation; Dragon and Rennie 1995, Hugh-Jones and De Vos 2002). While it has been observed experimentally and anecdotally in the field that anthrax spores have affinities for alkaline soils and high calcium levels (van Ness 1971, Dragon and Rennie 1995, Gates et al. 1995, Smith et al. 2000) these physiological affinities are still poorly understood. In general, it has been hypothesized that low pH soils, and soils that lack calcium, lead to quick desiccation of spores. Gates et al. (1995) suggested that the loss of calcium from the soil can trigger spores to prematurely break down the exosporium and die from a lack of nutrients necessary for survival as vegetative cells. Additionally, most studies conducted on anthrax soil preferences have been either simple geographic associations based on case distributions (van Ness and Stein 1956) or limited to small geographic areas (Smith et al. 2000).

Anthrax is considered a non-contagious, soil-borne disease that is primarily contracted through soil/host interactions, and direct inhalation and/or ingestion are considered the primary pathways of infection (Gates et al. 1995). However, most discussions on disease transmission in ungulate herds are speculative and based on post-mortem investigations of affected animals.
While a number of papers address regional scale (Dragon et al. 1994, Hugh-Jones and de Vos 2002) and local scale (De Vos 1990, Lindeque and Turnbull 1994, Smith et al. 1999, Smith et al. 2000) and ecological factors that promote anthrax spore survival, no quantitative spatial analyses are available at the continental scale to define the geographic limits of the species. Chapter 2 of this dissertation provides the first such continental-scale estimates of the geographic distribution of anthrax for the United States (U.S.).

In Chapter 2, ecological niche modeling using the Genetic Algorithm for Rule-set Production (GARP), satellite-derived and field-based environmental measurements (e.g. temperature, precipitation, soil pH, etc), and multi-variate analyses are employed to quantitatively define the geographic and ecological space of spore-promoting environments for the lower 48 states at the continental scale. This study was designed to identify the ecological parameters that best describe anthrax case distributions from recent and historical outbreaks and construct both a predictive spatial model of *B. anthracis* distribution, and a multi-dimensional environmental envelope that defines the physiological tolerances of anthrax spores. This is the first such application of the GARP modeling system to anthrax and one of the first ecological modeling studies on the disease in the New World at the continental scale. This study evaluates anthrax at the smallest geographic scale in this dissertation and addresses the macro-scale distribution of *B. anthracis*.

Chapter 3 is an expansion of Chapter 2 and presents the first predicted geographic distribution of anthrax in Mexico. The GARP modeling approaches allows for the results of one geographic area to be projected to another using environmental coverages. Chapter 3 provides a description of the important ecological parameters for anthrax in Mexico based on models developed for the U.S.
While Chapters 2 and 3 aim to define the broad-scale geographic distribution of the disease agent, such continental-level modeling cannot be used to understand the specific interactions between affected animals and anthrax spores. These interactions take place at a local level (meso-scale), within a given individual animal’s daily activity space or at the intersection of animals and potential transmission vectors. Chapter 4 is a meso-scale pilot study designed to address the spatial distribution and density of hematophagous flies during an anthrax outbreak season on an affected ranch in west Texas. While a number of papers anecdotally identify biting flies as a potential mechanical vector for anthrax (Mohiyuddeen and Krishna Rao 1958, De Vos 1990, Gates et al. 1995, Hugh-Jones and De Vos 2002), there is limited data available on the specifics of fly population dynamics or spatial distribution during the anthrax season. A pilot study was initiated to address the hypothesis of uniform versus clustered fly distributions. Non-baited net traps were deployed to systematically collect flies at a series of fixed sampling locations across a study ranch with a well documented and mapped anthrax history. Sampling was conducted throughout the summer period (June – August 2005) and two hotspot analyses were used to quantify the density and environmental relationships of fly abundance with respect to anthrax. This study reflects the growing trend in evaluating spatially explicit patterns of disease vectors in epidemiological studies (e.g. Jeffery et al. 2002, Getis et al. 2003, Cecere et al. 2004), and the importance of discerning spatial patterns in disease systems (Pfeiffer and Hugh-Jones 2002).

Necrophilic flies have also been implicated in anthrax transmission during outbreaks (Braack and De Vos 1990, Gates et al. 1995). Braack and De Vos (1990) examined the spatial dispersion of recaptured marked flies in relation to an anthrax positive animal carcass and determined that individual flies, despite having the potential for long distance movements, were
most likely to defecate or regurgitate on vegetation in the immediate vicinity (few meters) of the carcass, limiting transmission through fecal/regurgitation contamination to a relatively small spatial area. In this scenario, browsing ungulate species ingest contaminated vegetation and the number of cases from the index case – the carcass nearby, is increased. Hugh-Jones and De Vos (2002) hypothesized that this same phenomenon likely occurs in North American wildlife, such as white-tailed deer, *Odocoileus virginianus*, that exhibit browsing behaviors. However, no confirmatory data are available to substantiate this hypothesis in North America. To address this hypothesis, individual necrophilic flies were collected from dead deer found on a ranch in Texas during the 2005 anthrax season. Chapter 5 presents the rationale, collection methodology, and diagnostics employed to address this question in North America and defines the *case-multiplier hypothesis* for necrophilic insects in disease transmission.

Chapter 6 further addresses the mechanisms of disease spread at the meso-scale by considering deer movements within the same study area of the fly investigations. This follows the Gates et al. (1995) suggestion that the role of behavior in wildlife species is under appreciated in the scientific literature on anthrax and should be investigated to understand infection and transmission, especially in wildlife. Chapter 6 considers species-specific behavior during the anthrax season (summer months) using radio-telemetry, movement-sensitive cameras, and GIS-based analyses to determine the home ranges of white-tailed deer, on the study ranch from Chapters 4 and 5. Individual deer were randomly selected for radio collaring from all habitats on the study ranch (approximately 7406 hectares). Animals were selected from areas of the ranch known to have high infection rates in past years and from areas without any known cases. Animals were re-located throughout the 2005 summer periods (June – August and spot checked in October) using standard telemetry triangulation techniques and GPS locations. Home
range estimates were derived using a series of GIS-based analyses. Habitat classifications derived from LandSat 7 TM+ satellite data were used to describe the habitat types in each home range and to describe the habitats where anthrax cases have been documented on the study ranch.

This chapter was designed to introduce a methodology for estimating the potential movement space utilized by deer during a normal anthrax season. This is the first known study to specifically address the movement behaviors of white-tailed deer with respect to anthrax interaction and infection potential.

1.2 Expected Significance

This dissertation addresses multiple components that together make up the spatial ecology of anthrax disease. While many of the results presented in this dissertation represent first attempts at describing these ecological components, a great deal of important geographic and biological data are presented herein. Figure 1.1 illustrates these various components and identifies the scientific workflow necessary to study and relate them. This figure is expanded in Chapter 7 to illustrate the results of this dissertation effort and highlight areas of future research in anthrax ecology.

While vaccination remains the primary means of managing the disease in livestock, there is no realistic vaccination protocol for wildlife. In addition, livestock vaccination is primarily reactionary (used only once an outbreak has begun; Coker 2002) so increased surveillance and a full understanding of disease ecology are critical to improving management strategies for the disease. However, disease surveillance is expensive, and disease cases can be difficult to find, and more difficult to confirm (especially in wildlife). The use of GARP modeling can be employed to improve the current understanding of anthrax habitats and potentially aid in directing future surveillance. Likewise, this study will show how GARP modeling and other
Figure 1.1. Conceptual flowchart of the ecological components that represent anthrax disease.
quantitative mapping approaches can be employed to identify potential areas of risk from
existing cases and used to predict where disease spread through natural or anthropogenic causes
could lead to new areas of spore survival. There is limited knowledge on the role that animal
behavior and insects play in disease transmission and this dissertation introduces a first look at
the spatial relationships between wildlife and flies on an affected ranch in the endemic zone of
Texas. These studies can lead to an expanded methodology for determining specific pathways of
infection in wildlife and advance our general understanding of natural anthrax transmission.

1.3 References

Braack, L.E.O. and De Vos, V. 1990. Feeding habits and flight ranges of blow-flies 
(Chrysoma spp.) in relation to anthrax transmission in the Kruger National Park, South

Coker, P.R. 2002. Bacillus anthracis Spore Concentrations at Various Carcass Sites.
Doctoral dissertation, Louisiana State University. 77pp.

of reinfection by Triatoma infestans (hemiptera: reduviidae) following insecticide
spraying in a rural community in northwestern Argentina. American Journal of Tropical
Medicine and Hygiene 71(6): 803-810.

De Vos. V. 1990. The ecology of anthrax in the Kruger National Park, South Africa.
Salisbury Medical Review (Supplement) 68: 19-23.

Canadian Veterinary Journal 36: 295-301.

Gates, C.C., Elkin, B.T., Dragon, D.C. 1995. Investigation, control, and epizootiology of
anthrax in a geographically isolated, free-roaming bison population in northern Canada.
Canadian Journal of Veterinary Research 59: 256-264.


Getis, A., Morrison, A.C., Gray, K., Scott, T.W. 2003. Characteristics of the spatial pattern of
the dengue vector, Aedes aegypti, in Iquitos, Peru. American Journal of Tropical
Medicine and Hygiene 69 (5): 494-505.


CHAPTER 2  PREDICTING THE DISTRIBUTION OF ANTHRAX, *BACILLUS ANTHRACIS*, IN THE CONTIGUOUS UNITED STATES

2.1. Introduction

Anthrax is a disease that remains a problem in many countries worldwide for herbivorous livestock and wildlife species (Smith et al. 2000). Despite being a disease of antiquity, little is known about the spatial ecology of anthrax or the specific geography of environmental conditions that promote long-term survival of *Bacillus anthracis*, the causative agent of anthrax (Gainer and Saunders 1989, Kaufmann 1990, Smith et al. 1999). Much of the literature addresses the ubiquitous nature of *B. anthracis* as a soil-borne bacteria and many of these studies recognize the environmental constraints on long-term survivability of *B. anthracis* spores in soils (e.g. soil pH, calcium levels; van Ness and Stein 1956, van Ness 1959a,b, 1971, Dragon and Rennie 1995 Smith et al 2000, Coker 2002). While early literature argued that *B. anthracis* could replicate in soil (e.g. van Ness 1971, Kaufmann 1990), the current literature supports that *B. anthracis* only replicates in the animal host and can then survive long periods of dormancy in soil (Smith et al. 2000). Despite a body of literature defining the specifics of anthrax outbreaks regionally, and a body of work on the specific soil parameters required to maintain *B. anthracis* in the environment, few studies have evaluated the geographic distribution of these environmental characteristics using satellite-derived environmental data and multi-variate analyses. Smith et al. (2000) used spatio-temporal analyses and environmental data to relate soil conditions to anthrax outbreaks and spore persistence in the Kruger National Park in Africa, but no such detailed spatial analyses exist for North America.

Anthrax was most likely introduced into the United States (U.S.) by earlier European colonists (van Ness 1971). Anthrax was a large problem in domestic American livestock and wildlife through the 1950s (Stein 1945, Stein and van Ness 1955). Stein (1945) showed an
annual increase in the number of counties affected by anthrax between 1915 and 1944. Stein and van Ness (1955) reported continued increase in the spatial distribution of the disease up until the mid-1950s. With the introduction of a mass produced vaccine (Stein and van Ness 1955), disease management improved and the distribution of counties impacted reduced from the 1950s to the present day. Anthrax however, still remains a problem in both livestock and wildlife in certain parts of the United States (U.S.) (Hugh-Jones and de Vos 2002). While vaccination is inexpensive and readily available, it is often used in reaction to an outbreak more so than a disease preventative (Coker 2002), which consequently increases the likelihood of continued outbreaks in environments that promote long-term spore survival. Likewise, the current vaccine is administered through injection and only useful for livestock or farmed wild species that can be handled. Therefore, in areas like west Texas where the disease remains a problem in white-tailed deer (Hugh-Jones and de Vos 2002), vaccination is not a realistic form of disease control until an effective oral vaccine can be developed (and then work will be needed to determine likelihood of consumption). As anthrax remains a problem in both livestock and wildlife, each of which can incur high economic costs, a clear understanding of the spatial ecology of the disease is essential. Likewise, given that the disease remains enzootic, surveillance efforts must be targeted to capture areas of greatest risk for infection, and monitor all potential components of the disease (hosts, reservoirs, potential vectors). However, disease surveillance is expensive, requires multi-agency networks, and a multi-disciplinary approach. These networks must be clearly identified and the goal of the surveillance explicit. This study aims to identify continental scale patterns of anthrax that may further the current understanding of the disease and target areas of risk that require further research or surveillance.
The purpose of this study is to evaluate the specific geography of anthrax in North America using GIS-based analyses and ecological niche modeling. The goals of this study are to 1) define an environmental envelope for *B. anthracis* using multi-variate analyses, and 2) define the spatial distribution of the fundamental niche for the species using a machine-learning algorithm and satellite derived environmental data. This study is an experimental test of the usefulness of ecological niche modeling to predict a soil-borne, non-contagious bacterial disease. Additionally, this study evaluates the long-term persistence of anthrax in the lower 48 U.S. states and the relationship between persistence and model predictions. Defining the spatial extent of anthrax can be useful for generating new research hypotheses about disease persistence and for targeting surveillance efforts to areas of greatest risk of potential disease presence.

### 2.1.1 Ecological Niche Modeling

Ecological niche modeling (ENM) is the science of predicting the potential geographic and ecological distributions of species’ through the analysis of relationships between a combination of environmental variables (e.g. – temperature, precipitation, elevation, etc) and species’ locality data. A rich literature is available on a number of modeling approaches and accuracy metrics for defining and validating distributions. Many of these techniques bridge the use of multi-variate analyses, such as principal components analysis (Robertson et al. 2001), discriminant function analysis (Rogers 2000, 2006), or hybrid mulit-technique approaches (Stockwell and Peters 1999), with geographic information system (GIS) technology.

The fundamental principles of ecological niche modeling are based on the evolutionary relationships between species and environmental variables first defined as a species’ niche by Grinnell (1917). Grinnell defined the niche of a species as the ecological space where a population could maintain without immigration. Hutchinson (1957) expanded this to describe a
Holt and Gaines (1992) further defined the ecological niche within the constructs of evolution as the mean phenotype of the species. In this sense, the niche drifts as a population drifts and therefore is conserved across evolutionary time. Peterson et al. (1999) later confirmed this and determined that conservation of niches can be modeled with ecological niche approaches.

The fundamental niche is further sub-divided into a realized niche by biological processes such as competition (MacArthur 1972), species’ dispersal mechanisms (Peterson 2003), and historical factors such as extirpation (Peterson 2003). The realized niche is a limited sub-region of the fundamental niche, where the latter represents all potential ecological space for a species to exist, and the first represents that realized portion of geographic space where the species actually occurs (Hutchinson 1957, MacArthur 1972). In order to properly interpret ecological modeling results, it is important to understand this basic difference and ensure that biological relationships between sister taxa and evolutionary mechanisms for dispersal are understood, or at least recognized in poorly understood ecological systems, such as anthrax.

2.1.2 A Description of the GARP Model – a Machine-Learning Algorithm

This study was completed using a Genetic Algorithm for Rule-Set Production (GARP) to develop an ecological niche model for anthrax in North America. GARP is a presence-only modeling technique that relates point distributions (here anthrax outbreak locations) to environmental parameters (the environmental coverages) using a combination of methods (rules) that best describe the factors associated with species presence (Stockwell and Peters 1999, Stockwell and Peterson 2002). GARP rasterizes point locations (latitude/longitude pairs) of species presence to rasterized ecological variables (such as satellite-derived data and interpolated field measurements). GARP generates presence/absence predictions based on a set of
heterogeneous rules (rule-set) derived from a series of rule types in an iterative process. GARP employs four specific IF/THEN rule types in model development: 1) atomic rules – where predicted locations are defined by a specific environmental variable (e.g. IF temperature = [22°C] AND precipitation = [380 mm] THEN species = present/absent); 2) range rules – where predicted locations are defined by a range of variables (e.g. temperature = [18 - 22 °C] AND precipitation = [350 – 380 mm]); 3) negated range rules – where prediction locations are defined as values outside of a defined range (e.g. If range not temperature = [18 - 22 °C] AND precipitation = [350 – 380 mm]); 4) logit rules – where predicted locations are fit to a logistic regression model with the environmental variables (Stockwell and Peters 1999). These rules are used to determine non-random associations between environmental parameters and point distributions. GARP can be considered a super-set of individual modeling approaches, as range rules are essentially bioclimatic rules (e.g. Box et al. 1993) and several models use only logistic regression (e.g. Manel et al. 2001), and should have higher predictive accuracy than any single modeling approach (Stockwell and Peters 1999).

GARP is considered a genetic algorithm because the rules used to define species distribution can “evolve” from rule to rule. This modeling approach was inspired by early computational models based on principles of evolution (Holland 1975). In other words, the components of “IF” statements can be modified from one rule to the next to create new rules (Stockwell and Peters 1999). One example of a rule modification is crossover recombination. Following (Stockwell and Peters 1999), two range rules may be defined as:

Rule 1: IF temperature = [22, 37] AND elevation = [125, 500], THEN species = present

Rule 2: IF temperature = [19, 22] AND soil pH [6.2, 7.0], THEN species = present
Crossover can occur between rules 1 and 2 to define rules 3 and 4, where:

Rule 3: IF temperature = [22, 37] AND soil pH \([6.2, 7.0]\), THEN species = present

Rule 4: IF temperature = [19, 22] AND elevation = \([125, 500]\), THEN species = present

Crossover then is the recombination of partial rule definitions between rules (notice the change in position of the bold characters between rules). GARP can also produce point mutations, where the values of a specific variable are changed within a rule. For example:

Rule 5: IF temperature = [22, 37] AND elevation = [125, 500], THEN species = present

A point mutation is specific to the numerical values within a parameter, so under such mutation Rule 5 would become:

Rule 6: IF temperature = [19, 22] elevation = [125, 500], THEN species = present

In addition to crossover and point mutations, GARP can employ other rule modifiers, including deletions (where particular parameters are removed from a rule) and insertions (where particular parameters are added to a rule; Kluza and McNyset 2005).

Each GARP model process begins with the random selection of a rule type, and the rule-set is developed iteratively until a maximum of 50 rules have been defined to generate a rule-set. GARP utilizes this rule-set procedure to associate the environmental conditions at point locations...
to all areas inside of the defined study boundary. In this way GARP is searching both geographic space and ecological space to find solutions that define species’ presence. In GARP this is defined by the environmental coverages provided.

GARP is an iterative process whereby rules are selected, modified, tested and incorporated or rejected to develop a final rule-set (series of logical rules) that define the niche of the target species (McNyset 2005). Given the random variability generated in genetic rule development (from rule modifications), these model approaches are stochastic and can vary in the predicted distributions from model to model (McNyset 2005). Predictive accuracy is measured at each rule step using error components derived from predictions to determine the inclusion or exclusion of a rule to the rule-set. GARP generates a confusion matrix at each rule step to calculate predictive accuracy and statistical significance of predictability using randomly generated training and testing data (Anderson et al. 2003).

The confusion matrix is a 2x2 table that incorporates four elements for deriving accuracy metrics (Fielding and Bell 1997; Table 2.1). Element $a$ represents known species localities that are correctly predicted as present by the model. Element $b$ represents known absence localities predicted incorrectly as present (pseudo-absences selected within the model – see below). Element $c$ represents known species localities that are incorrectly identified as absent by the model. Element $d$ represents known absence localities correctly identified as absent by the model. Elements $a$ and $d$ represent correctly predicted areas of presence and absence (correct classifications), while elements $b$ and $c$ represent areas incorrectly over- and under- predicted, respectively. Element $c$ is defined as omission, where false negatives are defining areas of true presence as absence. Element $b$ is defined as commission, where false positives are defining
Table 2.1. The confusion matrix used to derive error metrics in ecological modeling

<table>
<thead>
<tr>
<th>CONFUSION MATRIX</th>
<th>Actual Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
</tr>
<tr>
<td>Present</td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td></td>
</tr>
</tbody>
</table>

To generate rule-sets, GARP intrinsically re-samples map pixels with replacement using the point data of species localities for training and testing (Anderson et al. 2003). GARP randomly partitions input data points into training data (for model building) and testing data, for intrinsic model fitting and evaluation. First, 1,250 map pixels are randomly chosen with replacement from those pixels with known localities (defined as the training set of data points). The quantity of $a$ in the confusion matrix is the number of pixels from those 1,250 that are predicted presence. The quantity of $c$ is the number outside of the predicted areas. Therefore, $a + c = 1,250$ for models where all pixels are predicted as present or absent. This may differ in models where $x$ (some number) of those 1,250 are not predicted by the rule-set developed and those values are coded as ‘no data’. GARP then repeats the re-sampling with replacement process for pixels outside of the training set. 1,250 pixels are re-sampled from the remaining pixels (pixels with no confirmed data points from the training data set). This process leads to the selection of pseudo-absence points, defined as background points in GARP (Stockwell and Peters 1999, Anderson et al. 2003). Background pixels categorized as predicted presence represent $b$. Background pixels categorized as predicted absence represent $d$. Therefore, $b + d = 1,250$ (unless there are cells not predicted, see $a + c$ above). The lack of true absence data
creates asymmetry in the confusion matrix (Anderson et al. 2003). Presence only data sets provide little error in presence, but high potential for high rates of pseudo-absence. In reality, \( e \) is representative of true omission; \( b \) includes both true and apparent commission, a value that cannot be immediately evaluated, without true absence data (Anderson et al 2003). Following Anderson et al. (2003) apparent commission errors are defined as pixels that could promote species presence that are correctly predicted, but with no verified records of species presence from those areas. This error can arise from more than one source. First, it could be as simple as geographically incomplete survey data. Likewise, historical reasons, such as extirpation – regionalized extinction of a species from only part of its range, or anthropogenic land use modifications, might limit the current distribution of a species to patches within a large ecological region that could otherwise support the species (Soberron and Peterson 2005). Biotic interactions- especially competition - can also isolate the distribution of a species. Competition can broadly be defined as any biological mechanism that increases the population of one species’ at the detriment of another species (MacArthur 1972). Apparent commission error is manifested in GARP-like applications (Anderson et al. 2003). This is an important concept because the accuracy metrics for internal rule-set selection, model development, and final user evaluation of the models are based on the components of the confusion matrix. The goal of GARP, and other modeling approaches (e.g. Rogers 2000, Gibson et al. 2004), is to determine which ecological and geographic space can and cannot support species presence, therefore commission is a critical error term to evaluate/interpret (Anderson et al. 2003).

A number of accuracy metrics can be derived from the confusion matrix (see Fielding and Bell 1997 for a full review of all methods). Two measures, the Kappa statistic and the \( \chi^2 \)-statistic are frequently reported in modeling papers. Both methods are preferred because they
take advantage of all four elements of the confusion matrix in their respective calculations (Fielding and Bell 1997). However, because pseudo-absence is generated using a re-sampling technique, pseudo-absence points vary between model runs. Therefore, element $b$ has the potential to vary more so than other presence/absence modeling approaches. Multiple authors have reported that Kappa and $\chi^2$ are sensitive to sample size (Fielding & Bell 1997, Anderson et al. 2003) and $\chi^2$ can have high significance in models with unacceptable omission rates (Anderson et al. 2003). In the specific case of GARP, the use of re-sampling to derive pseudo-absence from a large number of background pixels can lead to over-estimation of apparent commission. Other modeling approaches such as logistic regression require absence values be defined by the modeler. Therefore, measures of $b$ will be relative to the number of absence values available for the model (or derived and provided by the modeler).

Biologically it is difficult, and perhaps misleading, to define absence from a simple lack of presence data (Anderson 2003). The majority of modeling papers available are based on data collected either from biological field surveys, museum records, or published localities. These methods are often biased toward presence data and may not represent true absence of a species (Anderson 2003, Anderson et al. 2003). This is especially true over large geographic scales, where it is nearly impossible to derive a model training set from a single data source. While systematic surveys of similar taxa (e.g. Manel et al. 2001) over relatively small geographic areas may allow for a measure of both presence and absence, this is unlikely to be the case in studies based on multiple data sets, where rigorous sampling design would not be equal across data sets (Peterson 2001). Likewise, it is important to differentiate between true absence of species and sampling bias that may exclude the collection of individual organisms (Remsen and Good 1996, Carlson and Cortés 2003). Gear bias – physical limitations in sampling equipment or diagnostic
tools that may exclude the collection of a species even when it is present – has to also be coupled with sampling seasonality and potential local conditions (e.g. local weather during collection) that can affect the presence or absence of a species. In other words, not collecting a specimen in one sampling site might mean that individuals are present, but excluded by collecting gear used. Or the species may be present in an area at night, but not during the day (diel migration), therefore daylight sampling would define the species as absent. It is important to qualify if absence data is representative of species’ non-presence or really represents the lack of a species’ niche at the site of collection. Because of this, many authors have introduced a number of methods for deriving pseudo-absence data. Rogers (2006 in press) generates pseudo-absence points at some minimum and maximum spatial distance from any given presence point (for example an absent point may be no closer that 0.5 degree and no farther than 1.0 degree from a present point). Similar to GARP, the Rogers method employs a large number of absence points to capture large areas of ecological space and, if through re-sampling, pseudo-absences are predicted as present, they are re-classified based on probabilities. Engler et al. (2004) generated pseudo-absences randomly, and through the use of a multi-variate factor analysis of presence points in ecological space (termed an ENFA) and found that ENFA-derived pseudo-absences increased the performance of a generalized linear model at predicting target species’ distributions. In many ways these methods are similar to GARP, with the exception of sub-sampling protocols. For example, Rogers (2006 in press) randomly sub-samples an equal number of absence points to match the number of presence points. This approach not only allows for pseudo-absence assignment, but reduces disparity between element b and other elements within the confusion matrix, addressing sensitivity in Kappa.
Several authors have employed the area under the curve (AUC) in a receiver operating characteristic (ROC) analysis to determine predictive power in spatially explicit modeling techniques (Manel et al. 2001, Wiley et al. 2003, Gibson et al. 2004, McNyset 2005, Adjemian et al. 2006). The ROC analysis is a threshold independent assessment of model quality derived from a plot of sensitivity (true positive rate; y-axis) versus 1 – specificity (error or true negative rate; x-axis) constructed from each model set to determine if models are predicting better than random (Centor 1991, Zweig and Campbell 1993, McNyset 2005, Rogers 2006). Likewise, ROC, as employed by McNyset (2005), does not sub-sample pseudo-absence data to match the sample size of presence localities, and AUC’s are based on all pixels of presence and all pixels of absence. Historically, ROC approaches come from the medical diagnostic literature (e.g. Hanley and McNeil 1982). ROC provides a measure of diagnostic power (predictive power in ecological terms) of some technique (group of predictor variables from a series of tests – e.g. radiographs) to a gold standard for the ailment being diagnosed. The goal of the diagnostic technique is to determine if disease is present or absent, while in ecological terms the prediction tool aims to predict the presence of a species (using environmental data – e.g. bioclimatic variables; see Manel et al. 2001 Figure 1). In modeling terms, the predictions of presence and absence from a modeling technique are being compared to random predictions. The AUC of a given model set is compared to the AUC of a random prediction using a z-test. Successful models have AUC scores approaching 1.0 (a perfect model or a measure of reality), the higher the AUC the better the model is predicting presence/absence. Models predicting no better than random will have an AUC approaching 0.5 (Hanley and McNeil 1982). Wiley et al. (2003) and McNyset (2005) rigorously tested GARP at predicting species distributions of various taxa with
varying sample sizes and environmental coverages. Both studies utilized AUC scores to evaluate model quality and aid in model selection.

Anderson et al. (2003) reviewed the GARP algorithm and interpretations of Kappa, $\chi^2$ and a number of techniques for optimizing GARP based on evaluations of omission and commission. Current implementations of GARP, DesktopGARP (DG; available for free download at [www.lifemapper.org/desktopgarp/](http://www.lifemapper.org/desktopgarp/)) allow the user to set a best subsets procedure that evaluates omission and commission at each model output and sets criteria for final model selection. Because of the stochastic nature of GARP, it is critical to develop multiple models to evaluate inter-model variation (Anderson et al 2003). This is not exclusive to GARP, but rather a situation that exists in all ecological models that search broad geographic and ecological space for solutions based on random subsets of presence and absence. Rogers (2006 *in press*) employs a similar technique for deriving multiple models and evaluating distribution predictions and an average probability of presence across all models (usually 100). Optimal models in GARP are those that compromise between omission and commission. A high rate of omission suggests that models are over-fitting environmental variables to species’ locations and under-predicting the full fundamental niche of the species. Likewise, a high rate of commission would suggest that the model is over-predicting the distribution of the fundamental niche and expanding species’ distributions (Anderson et al. 2003). DG employs a best subset procedure to maximize model outputs by selecting models with user defined omission thresholds (e.g. 5% or less, 10% or less, etc). The user can select the total number of models to subset out of the total model run (e.g. select 20 models under 10% omission from a total of 100 models run). Secondarily, the best subsets procedure then clips the upper and lower user defined percentage of commission rates from those models to produce a best subset. Anderson et al. (2003) determined that eliminating
the lower 25% and upper 25% of commission error produced the most conservative model subset with the highest predictive power. By selecting the 50% commission error the omissions subset (20 in this example) is reduced to ten models that best optimize the relationship between omission and commission.

GARP models are based on spatially explicit locations of species’ presence and environmental coverages. Because of this GARP outputs can be directly viewed and manipulated within a GIS. GARP outputs are rasterized coverages of the study area representing presence pixels as 1’s and absence pixels as 0’s. To evaluate multiple model outputs, these individual model coverages can be summed using standard map algebra (e.g. model\textsubscript{A} pixels + model\textsubscript{B} pixels = model\textsubscript{AB}; where A pixels = 1 and B pixels = 1, AB pixels = 2; likewise, where A pixels = 0 and B pixels = 1, AB pixels = 1; and so on). The summation of models allows the user to identify geographic areas where none, some, or all of the models predict presence or absence (Kluza and McNyset 2005). The greater the number of models that agree, the more certainty there is in the probability of correct prediction classification (Ron 2005). Likewise, similarity across models indicates stability in the modeling system. This coupled with relatively high AUC scores for independent test data indicate greater certainty that the environmental coverages selected are explaining the variation in the species’ distribution data set used for training. Anderson et al. (2003) showed that summation of models selected during a best subset procedure have better overall predictability than summarizing a large number of individual models. In other words, 10 models evaluated for low omission and commission produce more conservative and better fit models than the summation of 100 or more individual models.

There must also be trade-off between training and testing data for model building and evaluation. In general, most statistical models have an increase in power associated with an
increase in sample size. Several studies have evaluated the required sample size needed for different modeling systems. Pearce and Ferrier (2000) reported a necessary 250 training locations for high accuracy models using generalized linear models and generalized additive models. Stockwell & Peterson (2002) reviewed accuracy in GARP relative to sample size and found a sharp increase as the sample size of presence localities approached 20 with accuracy leveling off until highest accuracies were reached with 50 sample points. McNyset (2005) found that even when limiting overall sample sizes to test the GARP algorithm for predicting species distributions (sample sizes for training data ranged from 15 to 74 points), GARP accuracies were high across a number of taxa. Adjemian et al. (2006) also performed a robustness analysis to evaluate GARP by repeatedly reducing sample size within a species to evaluate predictive power. Adjemian et al. (2006) showed that a sample size of six or greater was adequate for producing GARP models with high prediction accuracy for a flea species that can transmit plague. While model quality does increase with sample size, these studies support that sample sizes greater than 10 are adequate, 20 are better, and sample sizes of 50 points will maximize prediction power in GARP. This allows for adequate independent test data (a “hold out sample” in Fielding and Bell 1997) to be set aside for independent testing of GARP outputs for the generation of external measures of omission, commission, and ROC (McNyset 2005).

2.1.3 Applying GARP to anthrax: does GARP work with disease data?

A vast number of studies have confirmed the usefulness, success, and applicability of GARP to a wide range of species across terrestrial and aquatic taxa (e.g. terrestrial – Peterson 2001, Peterson and Vieglais 2001, Anderson et al. 2002, Raxworthy et al. 2003; aquatic – Wiley et al. 2003, Kluza and McNyset 2005, McNyset 2005). Additionally, a number of studies have applied the GARP modeling system to disease vectors. Peterson et al. (2002) applied GARP to
the distribution of potential mammalian reservoirs and triatomine insects in Mexico to predict areas of high risk for Chagas disease at the continental scale for Mexico. Costa et al. (2002) employed a similar approach to GARP model the distribution of Triatoma brasiliensis, an important Chagas vector in Brazil. This latter study, when combined with multi-variate analyses was used to differentiate individual populations of vectors across their geographic range and to define the ecological space utilized by the populations. A third study was conducted by Beard et al. (2003) to predict Chagas risk in previously under-sampled areas of Texas based on museum specimens of triatomine insects and a limited number of canine disease cases. All three studies produced both statistically significant and biologically useful maps of Chagas distribution. This in turn allowed for the development of new hypotheses about reservoir/vector relationships, potentially under sampled areas of disease presence, and the identification of areas in Texas where disease surveillance should be increased. Adjemian et al. (2006) also found GARP to be highly effective at identifying the distribution of plague, Yersinia pestis, in California, based on museum specimens and statewide environmental coverages. This latter study confirms GARP’s ability to identify the fundamental niche for species with limited dispersal patterns.

GARP has also been employed to predict distributions of diseases with no vectors (Ron 2005). Ron (2005) modeled the New World distribution of an exotic frog fungus that may be partially responsible for the rapid decrease in global amphibian populations. This study modeled the potential environments where the fungus could survive using published records of disease presence. Additionally, Ron (2005) was able to project the results from New World models onto Old World geography and verify GARP models with Old World locality data. Peterson et al. (2004) also successfully employed GARP and multi-variate analyses to model the distribution and environmental envelope of the poorly understood filoviruses. This last study identifies new
regions of interest for surveillance and fieldwork in the search for previously unknown reservoirs and vectors for the disease.

2.2 Materials and Methods

2.2.1 Anthrax Data Set Development

A GIS database of anthrax outbreak regions and specific localities within the 48 contiguous states was developed from a variety of data sources. GIS manipulations were performed in ArcGIS 9.0 and ArcView 3.2a (ESRI, Redlands, CA). Two data sets were used for this study. First, a data set of specific outbreak events (point data – latitude/longitude) was developed for inclusion in an ecological niche model analysis. These data were assimilated from the time period 2000 – 2005, with the exception of a 1957 outbreak report that could be mapped at the point level for Oklahoma. Point level maps were derived from latitude and longitude coordinate pairs collected by field personnel, address matched to farm front gates from diagnostic laboratory records, or heads-up digitized over high resolution satellite imagery using field reports and paper maps as guides to case locations. Point location accuracy estimates ranged from +/- 15m from a given carcass location (field investigation data – GPS coordinates) to +/- 2km for address matched farm locations and heads-up digitized outbreak locations. Point data (wildlife and livestock) for model building were available from six states representing three regions of anthrax outbreaks for the contiguous 48 states: 1) the Dakotas Region (North Dakota, South Dakota, Minnesota), 2) the Southern Region (Oklahoma, Texas), and 3) the Western Region (Nevada). Table 2.2 summarizes the sample sizes and methods of data collection for each of the states used in this analysis. Point level data from all states were combined into a single data set containing the latitude and longitude of each outbreak location in decimal degrees.
Table 2.2 Specific locality data used to develop GARP models.

<table>
<thead>
<tr>
<th>Outbreak location</th>
<th>Outbreak dates</th>
<th>Resolution</th>
<th>N</th>
<th>Geocoding technique</th>
<th>Accuracy estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dakota Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota*</td>
<td>2000 - 2005</td>
<td>Farm location</td>
<td>60</td>
<td>GPS coordinates</td>
<td>+/- 2 km</td>
</tr>
<tr>
<td>North Dakota**</td>
<td>2005</td>
<td>Farm location</td>
<td>88</td>
<td>address matching</td>
<td>+/- 2 km</td>
</tr>
<tr>
<td>South Dakota‡</td>
<td>2005</td>
<td>Farm location</td>
<td>49</td>
<td>address matching</td>
<td>+/- 2 km</td>
</tr>
<tr>
<td>Southern Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oklahoma†</td>
<td>1957</td>
<td>pasture locations</td>
<td>21</td>
<td>heads-up digitizing</td>
<td>+/- 2 km</td>
</tr>
<tr>
<td>Texas‡</td>
<td>2001 - 2005</td>
<td>carcass locations</td>
<td>122</td>
<td>GPS coordinates</td>
<td>+/- 15 - 100m</td>
</tr>
<tr>
<td>Western Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevada§</td>
<td>2002</td>
<td>carcass locations</td>
<td>32</td>
<td>GPS coordinates</td>
<td>+/- 15 - 100m</td>
</tr>
</tbody>
</table>

Data sources:
*Minnesota Board of Animal Health
**North Dakota State University Veterinary Diagnostic Laboratory
‡Sout Dakota State University Agriculture Extension Office & GIS Center for Excellence
†Oklahoma Department of Agriculture
TU S Centers for Disease Control and LSU WHOCC field investigations
§U.S. Department of Agriculture APHIS

Second, a decadal, county-level database was constructed from published and unpublished sources to evaluate persistence of the disease. Stein (1945) and Stein and van Ness (1955) published a series of decadal county-level maps as part of a continual review of the status of anthrax in North America. These maps pertained to the period between 1915 and 1955. Hugh-Jones (unpublished data) developed a digital database of the original Stein maps and updated them through 2004 using various sources including personal communications with local, state, and regional veterinary officers, Promed reports (www.promedmail.org), veterinary extension news letters, and ranch owner communications. The county-level data set was updated through 2005 based on World Health Organization Collaborating Center for Remote Sensing and GIS for Public Health, Louisiana State University (LSU WHOCC) field investigations and communication with veterinary personnel responsible for disease reporting in affected states. The county level data indicate the presence or absence of disease (1 present, 0 absent) for a given decade and are representative of the entire 48 state study area.
2.2.2 Environmental Data Sets

A set of environmental coverages was constructed from publicly available satellite-derived climatic and biological parameters. A total of 19 variables were downloaded from the WorldClim data set representative of various temperature and precipitation measurements (www.worldclim.org; Hijmans et al. 2005). An additional 13 environmental variables including temperature and vegetation measures (mean NDVI) were provided by the TALA research group at Oxford, University (available in Hay et al. 2006 in press). Two continuous soil parameters from the STATSGO data set (soil moisture, soil pH; www.ncgc.nrcs.usda.gov/products/datasets/statsgo) were used to incorporate measures of known ecological factors that promote _B. anthracis_ survival in the environment. Both variables were converted to raster format using the ArcGIS 9.0 Spatial Analyst extension. All environmental coverages were re-sampled to 0.10 degree² (~8 x 8 km²) using the GARP data sets extension for ArcView 3.2 provided with the software application and clipped to the boundary of the 48 contiguous U.S. states. All data sets were prepared using ERDAS Imagine v 8.7 (Leica GeoSystems, St. Gallen, Switzerland), ArcGIS 9.0 and ArcView 3.2a. Final coverages for GARP modeling were output for GARP using the GarpDataSets extension for ArcView 3.2a. Coverages were converted into a single GARP data set using the DG Dataset Manager application.

All sets of variables were tested in a series of combinations in GARP, first eliminating variables that represented similar parameters (e.g. mean annual temperature from WorldClim and land surface temperature from TALA) prior to model testing. Combinations of coverages were evaluated using a jackknife procedure (N-1 variables are used to build models iteratively until all possible combinations of variables have been used; Peterson et al. 2003). The jackknife
procedure is useful for eliminating variables that lead to over-fitting (Peterson and Cohoon 1999). To evaluate jackknife results a correlation matrix was derived from a set of models using the N-1 procedure and the measure of omission error for a 20 model set (McNyset 2005). Environmental variables were excluded from the final variable set if they increased omission between model outputs. A combination of jackknife evaluations and systematic model development and omission evaluation lead to the selection of the environmental coverage set. A final coverage set of six environmental variables was used that captured temperature, precipitation, elevation, soil moisture and pH, and vegetation (mean NDVI; Table 2.3). Figure 2.1 illustrates each of the six environmental variables used in this study.

2.2.3 GARP Model Development:

Two GARP modeling experiments are presented in this study. Both models were constructed using the rule-set writing and mapping application of DG v1.1.3. Both models were created using the same environmental coverages. Models were varied by manipulating the sample size used in each modeling approach. Both models were built using 100 initial models

<table>
<thead>
<tr>
<th>Environmental Variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean annual temperature (°C)</td>
<td>WorldClim (Higmans et al. 2005)</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td>WorldClim (Higmans et al. 2005)</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>(Hay et al. 2006 in press)</td>
</tr>
<tr>
<td>Soil Moisture (lowest liquid limit as % of weight)</td>
<td>STATGO U.S. Soil database</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>STATGO U.S. Soil database</td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td>(Hay et al. 2006 in press)</td>
</tr>
</tbody>
</table>
Figure 2.1. Environmental coverages used to develop the GARP models for anthrax in the Continental United States. A- annual mean temperature, B- annual precipitation, C- elevation above mean sea level, D- soil liquid limit (as % of weight), E- soil pH, F- mean NDVI (temporal Fourier processed – see Hay et al. 2006 in press)
with 1000 iterations each and convergence set at 0.01. The best subsets procedure was used to optimize final selection with extrinsic omission set at a 10% threshold and a commission threshold of 50%, model output was set at omission at 20 models, therefore when commission was evaluated, each experiment was left with the 10 best models. Model 1 was designed to maximize the number of training points and represent the entire outbreak data set. A 50% training/testing partition was used with a point file with all anthrax locations (minus the pre-selected hold-out). Because the outbreak point data had a large sample size in the Dakotas regions (due to detailed reporting from the 2005 outbreak), a haphazardly selected sub-sample of 46 locations was derived to more equally represent those outbreak regions with fewer data points (Oklahoma and Nevada) to test whether GARP was sensitive to under-represented areas of known outbreaks. Random sampling was not employed for sub-sampling in Model 2, as this was as likely to select points close together from the center of outbreak regions and would not have represented the geographic distribution of anthrax cases. Model 2 was designed to test GARP’s sensitivity to over-represented sample areas and random sampling was not necessary. Model 2 parameters matched model 1, with the exception of the training/testing data. A 91% training, 9% testing partition was used for model development to maximize the number of sub-sampled points in each model. The raster calculator routine in ArcGIS 9.0 Spatial Analyst was used to summate the 10 models in each model experiment to visualize the geographic areas of presence/absence predicted across the best subsets.

The rule-set writing application was used on Model 1 to allow for evaluation of rule-sets in each of the 10 best subset models and to determine individual rule types and environmental parameters important in each model. Additionally, the rule-set mapping application was used to
map the spatial assignment of dominant presence/absence rules for each of the 10 models in the best subset.

2.2.4 GARP Model Metrics

An area under the curve (AUC) in a Receiver Operating Characteristic (ROC) analysis was used to evaluate the prediction performance of each of the 10 best model subsets (Model 1 and Model 2) using measures of specificity (absence of commission error) and sensitivity (absence of omission error) following Wiley et al. (2003) and later McNyset (2005). The ROC was derived from the independent test data points withheld from the original GARP model building data sets. The greater the number of test points predicted present by the model set, the higher the AUC. A maximum AUC is achieved if all independent test points are predicted by all 10 of the best models within a subset (McNyset 2005). AUCs were calculated following Hadley and McNeil (1982).

Two measures of omission were calculated from the 10 best model subsets and the independent test data from both models following McNyset (2005). First, total omission was calculated as the total number of independent test points predicted as absent (pixel value = 0) by the summated grid surface of all 10 best models. The zonal statistics routine in ArcGIS 9.0 Spatial Analyst Extension was used to determine the total number of models (0 – 10) that predicted presence or absence at locations of independent test points. Secondly, a weighted omission was calculated as the average omission (number of independent known localities predicted as absent) across each of the individual 10 best models. Two commission indices were also developed. First, total commission was calculated as the total number of pixels predicted present across all ten models divided by the total number of pixels in the study area. Pixel counts were extracted from a summated grid of the 10 best models. Second, an average
commission was calculated as the total number of cells predicted present divided by the total number of pixels within the study area on a model-by-model basis for each of the 10 individual models within the best subset.

2.2.5 Disease Persistence

The county-level data set was used to evaluate persistence of the disease between 1915 and 2005. These data were overlain on the summated GARP surface from Model 1 to evaluate: 1) the overlap between areas of persistent outbreaks and the predicted fundamental niche of anthrax, and 2) to evaluate how well GARP predicted areas of known recent outbreaks with no point location data available. Persistence maps were developed for each decade in the data series and for pre- and post-vaccination introduction.

2.2.6 Defining a Multi-Variate Environmental Envelope

In addition to GARP modeling, an environmental envelope was developed for anthrax in the U.S. using a principal components analysis (PCA) to evaluate the explanatory power of the environmental coverages in multi-variate ecological space. Following Ron (2005), 5,000 randomly selected point locations were selected from the entire study area using the random point generator in the Hawth’s Analysis Tools extension for ArcGIS (www.spatialecology.com; Figure 2.2). The random points are generated from the entire study site to ensure that areas of both disease presence and absence are captured and to provide a measure of all ecological space in the U.S. to compare to the ecological space of anthrax cases. The zonal statistics routine in ArcGIS Spatial Analyst extension was used to assign the environmental parameters from each environmental coverage to each unique point location in the anthrax locality data and the randomly generated points. The zonal statistic assigns data to points based on the pixel that each point falls within. In cases where multiple points from the same file (anthrax or random) fall
within a single grid cell, zonal statistics returns a single value for that grid cell to each of the point layers that had data. In other words, if 5 points from the random data set fell in a single cell, the zonal routine returns a single cell value. This reduced the total number of points for both point files for inclusion in the PCA.

Two analyses were constructed to derive the environmental envelope for anthrax. First, a PCA was constructed from a combination of the anthrax case locations and the random point locations. This analysis was designed to visualize ecological space in the U.S. and indicate the position within ecological space where anthrax cases occurred. A second analysis was conducted solely on the anthrax cases to evaluate clustering of anthrax cases in ecological space. PCA was performed in SPSS v11 (SPSS, Inc, Chicago, IL). The “eigenvalue one criterion” (Norman and Streiner 2000) and a scree plot were used to select only PC’s with eigenvalues of 1 or greater for final selection. Kaiser-Meyer-Olkin Measure of Sample Adequacy, Bartlett Test of Sphericity and the SPSS anti-image matrix were used to test the appropriateness of PCA for the variables selected. The varimax rotation routine was used to rotate factors, and environmental envelopes were visualized by plotting PC I vs. PC II to evaluate the relative position of anthrax cases in ecological space.

2.3 Results

2.3.1 GARP Distribution Models

The GARP distributions for Model 1 and Model 2 were very similar, suggesting that GARP is capable of detecting environmental signals for anthrax presence across geographic space despite a greater number of cases locations in the Dakotas Region (Figure 2.2; illustrated as summated surfaces of the 10 best subsets for each). In both modeling experiments,
models were generated with convergence (0.01) achieved prior to reaching the total number of iterations (1,000). AUC’s were high for both models and both were significantly different from a line of no information (p < 0.01), indicating that both models performed better than random model generation (Table 2.4). Figure 2.4 illustrates the output of Model 1, which was the model used to write out the rule-set and the model used to map the dominant rules. Model sample sizes, AUC’s, and omission/commission indices are summarized in Table 2.3.3.1.

Table 2.5 lists the dominant rules for the 10 best subset models. Figure 2.5 illustrates the geographic distribution of the dominant presence and absence rules for each of the 10 best subset models for Model 1. Figure 2.6 displays a single model output from the best subset (Model 1 Task 9) to illustrate the dominant rules.
Table 2.4. Model sample sizes and accuracy metrics for GARP model development and validation.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$ to build models</td>
<td>296†</td>
<td>46‡</td>
</tr>
<tr>
<td>$N$ to test models (independent)</td>
<td>39</td>
<td>326</td>
</tr>
<tr>
<td>Total Omission</td>
<td>5.40%</td>
<td>0.74%</td>
</tr>
<tr>
<td>Weighted Omission</td>
<td>12.20%</td>
<td>5.72%</td>
</tr>
<tr>
<td>Total Comission</td>
<td>54.60%</td>
<td>68.58%</td>
</tr>
<tr>
<td>Average Comission</td>
<td>32.60%</td>
<td>42.53%</td>
</tr>
<tr>
<td>AUC</td>
<td>0.8994*</td>
<td>0.8818**</td>
</tr>
</tbody>
</table>

†$N$ was divided into 50% training / 50% testing at each model iteration
‡$N$ was divided into 91% training / 8% testing at each model iteration
* $z=10.503$ (p<0.01), ** $z = 19.545$ (p < 0.01)

2.3.2 Persistence Mapping

County-level persistence is mapped by decade and then for the 2005 anthrax season in figure 2.7. Figure 2.8 displays the counties positive for anthrax between 1915 and 1955 over the GARP output from Model 1. Figure 2.9 displays the counties positive for anthrax between 1956 and 2005 overlain on the GARP output from Model 1. Figure 2.10 displays the most recent period of anthrax outbreaks (1996 – 2005) overlain on the GARP output from Model 1.

2.3.3 Environmental Envelope Analysis

The zonal statistics routine returned 156 unique anthrax localities and 4,666 random locations (based on 0.01 degree grid cell resolution). Despite having a broad geographic distribution across six states, the ecological distribution of anthrax locations was narrow in comparison to the random locations (Figure 2.11). Table 2.6 summarizes the character loadings and percentage of variance explained for PC’s I and II for the first analysis. Figure 2.12 illustrates the envelope
Figure 2.3. GARP outputs for anthrax in the U.S. A) GARP Model 1 output derived from 296 anthrax localities and independently validated with 39 localities. B) GARP Model 2 output derived from 46 haphazardly chosen locations to more equally represent all geographic areas (independently tested with 136 localities). Red dots indicate data used for model building; yellow dots indicated independent test data.
Figure 2.4. GARP prediction for potential distribution of anthrax, *Bacillus anthracis*, in the continental United States based on outbreak locations. Legend colors from green to red indicate increased confidence in model accuracy of predicted disease presence (based on agreement between 10 best subset models). Red dots indicate outbreak locations used to develop GARP predictions (n = 296). Yellow dots indicate outbreak locations used to independently test GARP outputs (n = 39 unique points).
Table 2.5. Dominant rules from the 10 models in the final model best subset (Model 1). Tasks represent model numbers selected for best subset out of 100 models. Tasks are projected onto rule-set maps in Figure 2.5 (A – J).

****Task 1 (Figure 2.5 Insert A)

1 range rule
IF  
   temperature=[-3.90,24.31] AND precipitation=[283.80,694.11] AND elevation=[-201.58,778.71]  
   AND soil moisture=[0.00 ,0.60 ] AND soil pH1=[0.00 ,7.90 ]  
THEN  sp = PRESENCE

3 logit rule
IF  
   - precipitation*0.0078 - elevation*0.0078 - soil pH1*0.0000 + NDVI*0.0078  
THEN  sp = PRESENCE

4 negate range rule
IF NOT  
   temperature=[-2.90,22.65] AND precipitation=[129.93,1758.34] AND elevation=[67.52,3796.46]  
   AND soil moisture=[-0.00,0.90 ] AND NDVI=[-0.61,0.51 ]  
THEN  sp = ABSENCE

7 range rule
IF  
   temperature=[2.54 ,21.87] AND precipitation=[437.66,758.22] AND soil moisture=[0.00 ,0.60 ]  
   AND NDVI=[0.19 ,0.51 ]  
THEN  sp = PRESENCE

9 range rule
IF  
THEN  sp = ABSENCE

16 logit rule
IF  
   - temperature*0.0039 + elevation*0.0234 - soil moisture*0.0000 - soil pH1*0.0000 - NDVI*0.0039  
THEN  sp = ABSENCE

*****Task 9 (Figure 2.5 Insert B)

1 range rule
IF  
   precipitation=[219.69,1078.77] AND elevation=[221.29,682.60] AND soil pH1=[0.00 ,7.90 ]  
   AND NDVI=[0.24,0.38 ]  
THEN  sp = PRESENCE

4 range rule
IF  
   - temperature*0.0039 - precipitation*0.0156 - elevation*0.0078 - soil pH1*0.0000 + NDVI*0.0117  
THEN  sp = PRESENCE

6 range rule
IF  
   precipitation=[117.11,2784.11] AND elevation=[778.71,3777.24] AND soil moisture=[0.00 ,0.90 ]  
   AND NDVI=[0.01,0.65 ]  
THEN  sp = ABSENCE

12 negated range rule
IF NOT  
   temperature=[-2.90,23.20] AND elevation=[144.40,2277.97] AND NDVI=[-0.99,0.70 ]  
THEN  sp = ABSENCE
13 negated range rule  
IF NOT temperature=[0.32,22.20] AND precipitation=[53.00,2271.23] AND elevation=[-9.37,2969.94]  
AND soil moisture=[0.00,0.90] AND NDVI=[0.12,0.55]  
THEN sp = ABSENCE

14 range rule  
IF temperature=[1.21,21.42] AND precipitation=[1091.59,2335.34] AND elevation=[-566.79,3815.68]  
AND soil moisture=[0.00,0.90] AND soil pH1=[0.00,9.04] AND NDVI=[-0.92,0.69]  
THEN sp = ABSENCE

16 range rule  
IF temperature=[2.32,22.09] AND elevation=[221.29,1432.23] AND soil pH1=[0.00,7.90]  
AND NDVI=[0.27,0.46]  
THEN sp = PRESENCE

*****Task 15 (Figure 2.5 Insert C)*****

1 logit rule  
IF - precipitation*0.0078 - elevation*0.0078 - soil pH1*0.0000 + NDVI*0.0078  
THEN sp = PRESENCE

2 range rule  
IF temperature=[-3.79,24.20] AND elevation=[67.52,1682.11] AND soil moisture=[-0.00,0.90]  
AND soil pH1=[4.48,7.90] AND NDVI=[0.29,0.36]  
THEN sp = PRESENCE

8 negated range rule  
IF NOT temperature=[-3.01,23.20] AND precipitation=[104.29,2463.56] AND elevation=[-9.37,2643.18]  
AND soil pH1=[0.00,9.04] AND NDVI=[-0.98,0.51]  
THEN sp = ABSENCE

10 logit rule  
IF - temperature*0.0039 - elevation*0.1484 - soil pH1*0.0000 - NDVI*0.0000  
THEN sp = ABSENCE

12 negated range rule  
IF NOT precipitation=[360.73,1258.28] AND elevation=[-739.78,3969.45] AND soil moisture=[0.00,0.90]  
AND soil pH1=[0.00,9.04]  
THEN sp = ABSENCE

27 range rule  
IF precipitation=[245.33,1078.77] AND elevation=[278.95,990.14] AND soil pH1=[-0.04,9.00]  
AND NDVI=[0.29,0.43]  
THEN sp = PRESENCE

*****Task 22 (Figure 2.5 Insert D) *****

1 range rule  
IF precipitation=[437.66,706.93] AND elevation=[-489.90,4084.78] AND soil moisture=[0.00,0.60]  
AND NDVI=[0.19,0.54]  
THEN sp = PRESENCE

3 range rule  
IF temperature=[2.21,22.09] AND precipitation=[206.87,3297.00] AND elevation=[125.18,1643.67]  
AND soil moisture=[0.00,0.60] AND soil pH1=[-0.04,9.00] AND NDVI=[0.29,0.43]  
THEN sp = PRESENCE  

(TABLE 2.5 Continued)
4 range rule
IF temperature=[-2.79,23.53] AND elevation=[605.72,4027.11] AND soil pH1=[0.00 ,9.04 ]
   AND NDVI=[0.08 ,0.68 ]
THEN sp = ABSENCE

13 negated range rule
IF NOT precipitation=[117.11,1604.48] AND elevation=[-47.81,2181.87] AND soil moisture=[0.00 ,0.90 ]
   AND soil pH1=[0.00 ,9.04 ] AND NDVI=[-0.67,0.41 ]
THEN sp = ABSENCE

*****Task 27 (Figure 2.5 Insert E)

1 range rule
IF temperature=[2.43 ,22.09] AND precipitation=[373.55,617.17] AND elevation=[144.40,1701.33]
   AND soil moisture=[0.00 ,0.60 ] AND soil pH1=[0.00 ,7.97 ]
THEN sp = PRESENCE

3 negated range rule
IF NOT temperature=[1.10 ,22.20] AND precipitation=[437.66,3130.31] AND soil pH1=[0.00 ,9.04 ]
   AND NDVI=[0.02 ,0.69 ]
THEN sp = ABSENCE

4 negated range rule
IF NOT temperature=[-0.01,23.09] AND soil moisture=[0.00 ,0.90 ] AND soil pH1=[0.00 ,9.04 ]
   AND NDVI=[-0.76,0.51 ]
THEN sp = ABSENCE

8 negated range rules
IF NOT temperature=[-0.57,23.31] AND precipitation=[219.69,1617.30] AND elevation=[125.18,3508.14]
   AND soil pH1=[0.00 ,8.96 ] AND NDVI=[-1.00,0.78 ]
THEN sp = ABSENCE

9 logit rule
IF - temperature*0.0039 + precipitation*0.0352 - soil moisture*0.0000 - NDVI*0.0039
THEN sp = ABSENCE

30 range rule
IF precipitation=[219.69,1078.77] AND elevation=[221.29,1739.77] AND soil pH1=[0.00 ,7.90 ]
   AND NDVI=[0.29 ,0.50 ]
THEN sp = PRESENCE

*****Task 37 (Figure 2.5 Insert F)

1 range rule
IF temperature=[-4.01,24.09] AND precipitation=[219.69,1078.77] AND elevation=[-662.89,586.49]
   AND soil moisture=[0.00 ,0.60 ] AND soil pH1=[0.00 ,7.90 ] AND NDVI=[-0.19 ,0.40 ]
THEN sp = PRESENCE

3 negated range rule
4
IF NOT elevation=[144.40,1951.21] AND soil pH1=[0.00 ,9.00 ] AND NDVI=[-1.00,0.69 ]
THEN sp = ABSENCE

(TABLE 2.5 Continued)
4 logit rule
IF \(- \text{temperature} \cdot 0.0039 - \text{precipitation} \cdot 0.0078 + \text{elevation} \cdot 0.0313 - \text{soil moisture} \cdot 0.0039 - \text{NDVI} \cdot 0.0039\) THEN \(sp = \text{ABSENCE}\)

5 range rule
IF \(\text{temperature} = [-1.57,6.10] \text{ AND precipitation} = [194.04,1053.13] \text{ AND elevation} = [163.62,1701.33] \text{ AND soil pH} = [0.00,7.90] \text{ AND NDVI} = [0.23,0.51]\) THEN \(sp = \text{PRESENCE}\)

6 range rule
IF \(\text{precipitation} = [219.69,1078.77] \text{ AND elevation} = [-605.23,624.94] \text{ AND soil pH} = [0.00,7.90] \text{ AND NDVI} = [0.22,0.50]\) THEN \(sp = \text{PRESENCE}\)

7 negated range rule
IF \(\not \text{precipitation} = [168.40,1681.41] \text{ AND elevation} = [-28.59,3066.05] \text{ AND soil pH} = [0.00,8.93] \text{ AND NDVI} = [-0.28,0.52]\) THEN \(sp = \text{ABSENCE}\)

28 range rule
IF \(\text{temperature} = [3.43,5.10] \text{ AND precipitation} = [258.15,1104.42] \text{ AND elevation} = [163.62,1451.45] \text{ AND soil moisture} = [0.04,0.59]\) THEN \(sp = \text{PRESENCE}\)

*****Task 40 (Figure 2.5 Insert G)*****

1 logit rule
IF \(- \text{precipitation} \cdot 0.0078 - \text{elevation} \cdot 0.0078 - \text{soil pH} \cdot 0.0000 + \text{NDVI} \cdot 0.0078\) THEN \(sp = \text{PRESENCE}\)

2 negated range rule
IF \(\not \text{temperature} = [-0.35,23.09] \text{ AND precipitation} = [181.22,3271.36] \text{ AND elevation} = [86.74,624.94] \text{ AND soil moisture} = [0.00,0.90] \text{ AND soil pH} = [0.00,9.04] \text{ AND NDVI} = [-0.99,0.76]\) THEN \(sp = \text{ABSENCE}\)

6 negated range rule
IF \(\not \text{temperature} = [-0.35,23.20] \text{ AND soil moisture} = [-0.00,0.89] \text{ AND soil pH} = [0.00,9.04] \text{ AND NDVI} = [0.13,0.50]\) THEN \(sp = \text{ABSENCE}\)

27 range rule
IF \(\text{precipitation} = [219.69,1078.77] \text{ AND elevation} = [125.18,1701.33] \text{ AND soil moisture} = [0.00,0.60] \text{ AND NDVI} = [0.10,0.38]\) THEN \(sp = \text{PRESENCE}\)

39 range rule
IF \(\text{precipitation} = [168.40,1053.13] \text{ AND elevation} = [125.18,1701.33] \text{ AND NDVI} = [0.14,0.41]\) THEN \(sp = \text{PRESENCE}\)

40 range rule
IF \(\text{temperature} = [2.32,22.09] \text{ AND precipitation} = [360.73,963.37] \text{ AND elevation} = [240.51,1470.68] \text{ AND soil moisture} = [0.00,0.60] \text{ AND soil pH} = [-0.04,9.00]\) THEN \(sp = \text{PRESENCE}\)

(TABLE 2.5 Continued)
****Task 49 (Figure 2.5 Insert H)

1 range rule
IF temperature=[-2.57, 5.10] AND precipitation=[219.69, 1078.77] AND elevation=[259.73, 1662.89]
   AND soil pH1=[0.00, 7.90] AND NDVI=[0.29, 0.42]
THEN sp = PRESENCE

3 negated range rule
IF NOT temperature=[-4.01, 24.09] AND precipitation=[129.93, 2155.83] AND elevation=[202.07, 2796.95]
   AND soil pH1=[-0.04, 9.00] AND NDVI=[-0.98, 0.51]
THEN sp = ABSENCE

4 range rule
IF precipitation=[412.02, 1040.30] AND elevation=[-701.34, 4123.22] AND soil pH1=[0.00, 7.90]
   AND NDVI=[0.16, 0.38]
THEN sp = PRESENCE

5 range rule
IF temperature=[-2.90, 23.53] AND precipitation=[53.00, 3297.00] AND elevation=[740.26, 3738.79]
   AND soil moisture=[0.00, 0.90] AND NDVI=[-1.01, 0.72]
THEN sp = ABSENCE

21 negated range rule
IF precipitation=[437.66, 873.62] AND elevation=[86.74, 1778.22] AND soil moisture=[0.00, 0.60]
   AND soil pH1=[-0.04, 7.93]
THEN sp = PRESENCE

25 range rule
IF temperature=[2.32, 21.76] AND elevation=[182.85, 1451.45] AND soil moisture=[0.00, 0.60]
   AND soil pH1=[4.48, 7.90] AND NDVI=[0.19, 0.51]
THEN sp = PRESENCE

****Task 81 (Figure 2.5 Insert I)

1 negated range rule
IF NOT temperature=[-3.90, 24.20] AND precipitation=[399.20, 1694.23] AND soil moisture=[0.00, 0.90]
   AND soil pH1=[0.00, 9.04] AND NDVI=[-0.00, 0.50]
THEN sp = ABSENCE

3 range rule
IF precipitation=[424.84, 655.64] AND elevation=[48.30, 1682.11] AND soil moisture=[0.00, 0.60]
   AND soil pH1=[0.00, 9.04] AND NDVI=[-0.99, 0.78]
THEN sp = PRESENCE

7 negated range rule
   AND soil moisture=[0.00, 0.90] AND soil pH1=[0.00, 9.04]
THEN sp = ABSENCE

15 range rule
IF temperature=[-2.57, 23.31] AND precipitation=[104.29, 2181.47] AND elevation=[1067.03, 3604.25]
   AND soil pH1=[0.00, 9.04] AND NDVI=[-1.00, 0.74]
THEN sp = ABSENCE

(TABLE 2.5 Continued)
18 logit rule
IF - temperature*0.0039 - precipitation*0.0078 - elevation*0.0078 - soil moisture*0.0039 - soil pH1*0.0000 + NDVI*0.0117
THEN sp = PRESENCE

24 range rule
IF precipitation=[219.69,1078.77] AND elevation=[-739.78,4084.78] AND soil pH1=[-0.04,7.97 ]
   AND NDVI=[0.25 ,0.40 ]
THEN sp = PRESENCE

30 range rule
IF temperature=[2.21 ,22.65] AND precipitation=[232.51,1078.77] AND elevation=[182.85,1739.77]
   AND NDVI=[0.26 ,0.51 ]
THEN sp = PRESENCE

*****Task 88 (Figure 2.5 Insert J)*****

1 logit rule
IF - precipitation*0.0078 - elevation*0.0078 - soil moisture*0.0000 - soil pH1*0.0000 + NDVI*0.0078
THEN sp = PRESENCE

2 range rule
IF precipitation=[219.69,1078.77] AND elevation=[144.40,1701.33] AND soil moisture=[0.00 ,0.60 ]
   AND soil pH1=[0.00 ,7.90 ] AND NDVI=[0.26 ,0.37 ]
THEN sp = PRESENCE

3 negated range rule
IF NOT temperature=[1.10 ,22.09] AND precipitation=[117.11,1617.30] AND soil moisture=[0.00 ,0.90 ]
   AND soil pH1=[0.07 ,9.00 ] AND NDVI=[0.26 ,0.52 ]
THEN sp = ABSENCE

4 negated range rule
IF NOT temperature=[2.32 ,22.09] AND precipitation=[232.51,1078.77] AND elevation=[-778.22,4104.00]
   AND soil moisture=[0.00 ,0.60 ]
THEN sp = ABSENCE

5 logit rule
IF + precipitation*0.0234 - elevation*0.0078 - soil moisture*0.0000 - NDVI*0.0000
THEN sp = ABSENCE

7 negated range rule
IF NOT temperature=[0.76 ,22.20] AND precipitation=[412.02,1796.81] AND elevation=[-9.37,2835.39]
   AND soil moisture=[0.00 ,0.90 ] AND soil pH1=[0.00 ,9.00 ]
THEN sp = ABSENCE

8 range rule
IF temperature=[-3.90,24.31] AND precipitation=[386.38,809.51] AND elevation=[182.85,1451.45]
   AND soil moisture=[0.01 ,0.58 ] AND soil pH1=[4.48 ,7.90 ]
THEN sp = PRESENCE

10 logit rule
IF - temperature*0.0039 - precipitation*0.0156 + elevation*0.0547 - soil pH1*0.0000 - NDVI*0.3984
THEN sp = ABSENCE

(TABLE 2.5 Continued)
for the 156 anthrax locations in ecological space. Table 2.7 summarizes the character loadings and percentage of variance explained for the 156 anthrax locations. Additionally descriptive statistics were produced for the 156 locations based on the six environmental variables (Table 2.8). Due to the broad geography of anthrax cases, the descriptive statistics for the anthrax cases were sub-divided the by state (Table 2.9).

2.4 Discussion

These results suggest that GARP is a suitable and accurate tool for modeling the ecological niche for anthrax in the U.S. at the continental scale using outbreak locations ranging from carcass locations to ranch locations. Both Models 1 and 2 showed similar distributions of presence (Figure 2.3A,B) indicating that GARP is sensitive to geographic variability in environmental parameters despite a bias towards the Dakotas region from a greater number of anthrax case locations (as compared to Oklahoma or Nevada). Model 1 was slightly more conservative in the spatial distribution of anthrax (based on the number of subset models that agree), but this could also be due to the stochastic nature of the modeling approach. Rule-set analysis and rule-set mapping were performed on Model 1, as a greater number of anthrax locations were used to produce the models. The high AUC scores indicate that both models performed with high accuracy for this data. This also suggests that the environmental coverage set used explained the distribution of anthrax cases well.

The predicted distribution of anthrax (areas of greatest model agreement) is geographically isolated to a narrow corridor from southwest Texas northward through central Oklahoma, central Kansas, central Nebraska, and into the Dakotas and Minnesota. The areas of highest certainty (10 models agree) in Model 1 indicate that the area from west Texas through Nebraska is quite narrow up through to Southern South Dakota, expanding eastward and
westward through central South Dakota, into Minnesota and North Dakota. The northwestern most portion of South Dakota and southwestern most portion of North Dakota were predicted as potential habitat with far less certainty, with some parts of this area predicted by only one of the 10 best subset models. The northwest corner of Montana (which saw anthrax cases on at least two ranches in 2005) was predicted by six or more models with no point locations used for model building from this area. GARP also predicted areas of high certainty east and southeast of Minnesota along the Great Lakes and into portions of Ohio. This region of the U.S. has not reported anthrax outbreaks since the 1960s, and may indicate that GARP is over-predicting this region, or more likely that this area is geographically situated within the fundamental niche of anthrax but either lacks livestock populations or the land has been utilized for other purposes, such as grain crops, where livestock do not graze. This also suggests more work should be done to evaluate disease control and eradication programs for this region during the last large outbreak period to determine if practices were more successful here than elsewhere in the fundamental niche. This presents a research question for future work on the role of land use practices and disease management strategies.

Three separate disjunct geographic areas were also predicted as anthrax habitat with high certainty. First, the southwestern corner of Arizona was predicted with relatively high certainty. Second, central California was predicted as a habitat from the western coast from San Diego northward to central California on both the eastern and western slopes of the Sierra Nevada Mountains. A third disjunct area was predicted with high certainty centered on the eastern border of Oregon and Washington, where the two states meet Idaho.
Figure 2.5. Maps of the dominant presence (red color ramps – P) and absence (blue color ramps - A) rules for each of the top 10 best models in the subset (model 1). A) Task 1; B) Task 9; C) Task 15; D) Task 22; E) Task 27; F) Task 37; G) Task 40; H) Task 49; I) Task 81; J) Task 88. Inserts show the presence/absence (1 red / 0 gray) maps for each Task. Task numbers correspond to individual rule-sets in the results section and in Appendix 1; numbers in the legend correspond to rules.
Figure 2.6. Map displaying a selection of the dominant rule-set assignments for a single task (Task 09) selected from the 10 best model subset (Model 1). Red colors indicate presence rules (P) and blue colors indicate absence rules (A). Insert A) Task 09 model output (red = presence / gray = absence). Insert B) Final GARP model based on the 10 best subset model.
Figure 2.7. Decadal persistence of anthrax by county. Black outlines indicate positive counties.
Figure 2.8. Counties positive for anthrax from 1915 – 1955 as summarized by Stein (1945) and Stein and van Ness (1955) representing the anthrax situation in the lower 48 states prior to the use of systematic vaccination.
Figure 2.9. Counties positive for anthrax from 1956 - 2005. Note the high association of model agreement with positive counties west of the Mississippi River, supporting the hypothesis that the FN for anthrax spans a corridor through the central states, and has a disjunct distribution in the western U.S.
Figure 2.10. Overlay of positive anthrax counties 1996 – 2005 (black outlines) on the GARP prediction based on six environmental coverages. Legend colors from green to red indicate increased confidence in model accuracy of predicted disease presence (based on agreement between 10 best subset models). With the exception of Oklahoma data, all anthrax localities used in the model building were post-1995.
Figure 2.11. Principal components (PC) I and II based on six environmental variables for 4,622 locations randomly derived from across the lower 48 states (light gray symbols) and 156 unique anthrax foci (black symbols). The light gray distribution represents the environmental space of the Continental US, as defined by the six variables chosen. The black symbols represent the environmental space for anthrax, *Bacillus anthracis*. The distribution of anthrax foci, while showing some geographic separation, is concentrated in relatively narrow ranged of soil pH, soil moisture, vegetation coverage, and precipitation.

Table 2.6. Character loadings and percentage of variance explained by principal components (PC) I and II for six environmental variables at 156 known anthrax localities and 4,666 random localities within the continental United States. Bold text represents variables with the highest loadings.

<table>
<thead>
<tr>
<th>Environmental Variable</th>
<th>PC I</th>
<th>PC II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean annual temperature (°C)</td>
<td>0.669</td>
<td>0.333</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td><strong>0.831</strong></td>
<td>-0.389</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>-0.845</td>
<td>-7.68E-02</td>
</tr>
<tr>
<td>Soil Moisture (lowest limit represented as %)</td>
<td>0.146</td>
<td><strong>0.662</strong></td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>-0.348</td>
<td><strong>0.777</strong></td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td><strong>0.775</strong></td>
<td>-0.399</td>
</tr>
<tr>
<td>Percentage of variance explained</td>
<td>46.62</td>
<td>21.11</td>
</tr>
</tbody>
</table>
Figure 2.12. Principal components (PC) I and II based on six environmental variables for 156 anthrax locations (calculated for anthrax foci, excluding random locations). Different symbology represent individual states used in the analysis.

Table 2.7. Character loadings and percentage of variance explained by principal components (PC) I and II for six environmental variables at 156 known anthrax locations. Bold text represents variables with the highest loadings.

<table>
<thead>
<tr>
<th>Environmental Variable</th>
<th>PC I</th>
<th>PC II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean annual temperature (ºC)</td>
<td>0.516</td>
<td><strong>0.720</strong></td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td><strong>0.934</strong></td>
<td>-8.68E-02</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>-0.656</td>
<td>2.30E-01</td>
</tr>
<tr>
<td>Soil Moisture (lowest limit represented as %)</td>
<td>-0.164</td>
<td>0.467</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>-0.253</td>
<td><strong>0.819</strong></td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td><strong>0.926</strong></td>
<td>-6.88E-02</td>
</tr>
<tr>
<td>Percentage of variance explained</td>
<td>42.69</td>
<td>23.87</td>
</tr>
</tbody>
</table>

Table 2.8. Descriptive statistics for environmental variables at 156 known anthrax sites

<table>
<thead>
<tr>
<th>Environmental Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean annual temperature (ºC)</td>
<td>0</td>
<td>22.5</td>
<td>9.0</td>
<td>6.3</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td>201</td>
<td>1111</td>
<td>547</td>
<td>169</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>109</td>
<td>1767</td>
<td>454</td>
<td>235</td>
</tr>
<tr>
<td>Soil Moisture (lowest limit represented as %)</td>
<td>0.00</td>
<td>0.60</td>
<td>0.24</td>
<td>0.15</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>3.6</td>
<td>7.9</td>
<td>6.4</td>
<td>0.85</td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td>0.17</td>
<td>0.56</td>
<td>0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Table 2.9. Descriptive statistics grouped by individual state for 156 unique anthrax locations.

<table>
<thead>
<tr>
<th>Environmental Variable</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dakotas region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean annual temperature (°C)</td>
<td>21</td>
<td>2.4</td>
<td>4.2</td>
<td>3.2</td>
<td>0.63</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td>21</td>
<td>491</td>
<td>599</td>
<td>535</td>
<td>37</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>21</td>
<td>295</td>
<td>428</td>
<td>343</td>
<td>37</td>
</tr>
<tr>
<td>Soil Moisture (lowest liquid limit as % of weigh)</td>
<td>21</td>
<td>0.00</td>
<td>0.50</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>21</td>
<td>4.5</td>
<td>7.4</td>
<td>6.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td>21</td>
<td>0.34</td>
<td>0.43</td>
<td>0.36</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>North Dakota</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean annual temperature (°C)</td>
<td>48</td>
<td>3.1</td>
<td>5.6</td>
<td>4.8</td>
<td>0.51</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td>48</td>
<td>436</td>
<td>506</td>
<td>489</td>
<td>16</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>48</td>
<td>276</td>
<td>624</td>
<td>413</td>
<td>74</td>
</tr>
<tr>
<td>Soil Moisture (lowest liquid limit as % of weigh)</td>
<td>48</td>
<td>0.00</td>
<td>0.50</td>
<td>0.19</td>
<td>0.07</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>48</td>
<td>5.6</td>
<td>7.4</td>
<td>6.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td>48</td>
<td>0.30</td>
<td>0.37</td>
<td>0.33</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>South Dakota</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean annual temperature (°C)</td>
<td>40</td>
<td>5.5</td>
<td>8.9</td>
<td>7.0</td>
<td>0.85</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td>40</td>
<td>427</td>
<td>551</td>
<td>470</td>
<td>28</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>40</td>
<td>385</td>
<td>647</td>
<td>513</td>
<td>71</td>
</tr>
<tr>
<td>Soil Moisture (lowest liquid limit as % of weigh)</td>
<td>40</td>
<td>0.00</td>
<td>0.60</td>
<td>0.35</td>
<td>0.18</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>40</td>
<td>5.6</td>
<td>6.6</td>
<td>6.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td>40</td>
<td>0.28</td>
<td>0.37</td>
<td>0.33</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Southern Region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean annual temperature (°C)</td>
<td>12</td>
<td>14.2</td>
<td>15.3</td>
<td>14.6</td>
<td>0.38</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td>12</td>
<td>1053</td>
<td>1111</td>
<td>1080</td>
<td>16</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>12</td>
<td>180</td>
<td>281</td>
<td>243</td>
<td>27</td>
</tr>
<tr>
<td>Soil Moisture (lowest liquid limit as % of weigh)</td>
<td>12</td>
<td>0.00</td>
<td>0.50</td>
<td>0.17</td>
<td>0.20</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>12</td>
<td>3.6</td>
<td>5.6</td>
<td>5.0</td>
<td>0.6</td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td>12</td>
<td>0.47</td>
<td>0.56</td>
<td>0.50</td>
<td>0.03</td>
</tr>
<tr>
<td>Texas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean annual temperature (°C)</td>
<td>31</td>
<td>18.1</td>
<td>22.5</td>
<td>19.6</td>
<td>3.85</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td>31</td>
<td>484</td>
<td>699</td>
<td>581</td>
<td>58</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>31</td>
<td>109</td>
<td>699</td>
<td>422</td>
<td>194</td>
</tr>
<tr>
<td>Soil Moisture (lowest liquid limit as % of weigh)</td>
<td>31</td>
<td>0.00</td>
<td>0.41</td>
<td>0.24</td>
<td>0.11</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>31</td>
<td>5.6</td>
<td>7.9</td>
<td>7.3</td>
<td>0.8</td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td>31</td>
<td>0.30</td>
<td>0.53</td>
<td>0.38</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Western Region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean annual temperature (°C)</td>
<td>4</td>
<td>8.9</td>
<td>9.9</td>
<td>9.4</td>
<td>0.44</td>
</tr>
<tr>
<td>Annual precipitation (mm)</td>
<td>4</td>
<td>201</td>
<td>240</td>
<td>227</td>
<td>18</td>
</tr>
<tr>
<td>Elevation (m above mean sea level)</td>
<td>4</td>
<td>1469</td>
<td>1767</td>
<td>1653</td>
<td>131</td>
</tr>
<tr>
<td>Soil Moisture (lowest liquid limit as % of weigh)</td>
<td>4</td>
<td>0.15</td>
<td>0.30</td>
<td>0.20</td>
<td>0.07</td>
</tr>
<tr>
<td>Soil pH (lowest reaction -- no units)</td>
<td>4</td>
<td>6.1</td>
<td>6.6</td>
<td>6.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Mean NDVI (no units)</td>
<td>4</td>
<td>0.17</td>
<td>0.24</td>
<td>0.20</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The habitat in Arizona may identify an important corridor for the historical movement of cattle from Texas to California (van Ness 1971). Likewise, there appears to be patches of anthrax habitat west from North Dakota to the large area of anthrax habitat predicted in the Pacific
Northwest. Hugh-Jones (pers. comm.\textsuperscript{1}) has mapped the historical cattle trails for the U.S. and confirms that routes of cattle movement most likely traversed this region of Arizona on their way to the northwest.

The rule-sets indicate a relatively narrow range of environmental parameters used to define anthrax presence across all 10 models in the best subset. Additionally, the rule-sets indicate relatively few rules are used to predict the majority of the geographic areas of presence or absence (Figure 2.5). This suggests that the point distribution of anthrax used here represents the true distribution of anthrax without a large number of points representing sink populations (Soberon and Peterson 2005). Additionally, this indicates that despite the potential for great complexity in GARP rule selection due to the genetic behavior of rule development, GARP detected a strong environmental signal in the environmental coverages and was able to explain the distribution of anthrax in relatively few rules. This challenges the argument made by Rogers (2006 \textit{in press}) that GARP produces statistically significant maps with limited biological information. While rule-sets may need to be more complex to describe other taxa, this is not the case for anthrax. A review of the rule-sets in Table 2.5 indicates that the majority of presence data are defined by range rules of narrow bands of precipitation, NDVI, and to a lesser degree, soil pH. The narrow band of NDVI was the single most important variable across rules and models, suggesting it was the most important variable in the coverage set. Additionally, this agrees with preliminary attempts to model anthrax in the U.S. using the Rogers model (Rogers and Blackburn, unpublished results), where NDVI was the dominant variable. This also agrees with the PCA, as PC I is defined by high character loadings for NDVI in both analyses. In most of the maps showing the geography of rule-sets (Figure 2.5), there is a general pattern of a dominant western presence rule and a dominant eastern presence rule to define the anthrax niche.

\textsuperscript{1} Martin E. Hugh-Jones, Department of Environmental Sciences, LSU, Baton Rouge, LA 70803
In general terms, GARP assigned a broad absence rule that defined the Rocky Mountains and the western deserts. While a second rule generally defined absence east of the anthrax corridor that runs through the center of the continent. Overall, the dominant rules in the rule-sets were range rules and negated range rules, with few logit rules. While a few atomic rules are present in the appendix, none of the dominant rules were atomic. This makes sense biologically, as the fundamental niche of a species is defined by a population, not an individual (Holt and Gaines 1992). It is unlikely then that a single value from a single parameter will often define more than an isolated percentage of a population. GARP also detected subtle changes in the environmental conditions that support the species across geographic space. For example, in Task 9 (Figure 2.6) the NDVI envelope expands from west to east, as illustrated by rules 1 and 16, but still captures the distribution of the disease. This illustrates a powerful advantage of the GARP modeling approach over other modeling approaches. The development of heterogeneous rules allows GARP to explore variation in environmental parameters across the landscape. The assignment of rules to pixels (spatial assignment) occurs after the rule set is developed; therefore each pixel can be evaluated using different rules until an appropriate rule is selected. In the case of other modeling approaches, such as logistic regression, a single descriptor is used to describe variance in the species’ locality data. In this way modeling approaches such as discriminant function assume that the environmental variables that describe data points are uniform across ecological (and therefore geographic) space, which is not biologically plausible. Species range across geographic space and environments vary through space, especially across areas as vast the U.S, which spans from the tropics (south Florida) to boreal forest.

Soil pH has long been defined as an important limiting factor for anthrax spore survival (van Ness and Stein 1956, van Ness 1971, Dragon and Rennie 1995, Smith et al. 2000);
specifically that anthrax requires alkaline soils (relatively high pH) and high calcium levels (Gates et al. 1995). Soil pH and calcium are thought to control the maintenance of the exosporium and regulate the timing of sporulation, respectively (Gates et al. 1995). The models developed in this study indicated that NDVI was a better explanatory variable for anthrax predictions than pH, but this most likely indicates either 1) a lack of spatial precision for pH at a continental scale (an artifact of the statsgo soils database), or 2) that NDVI may serve a surrogate for soil conditions that promote spore survival. Additionally, this study did not differentiate genetic strains of *B. anthracis* for model development. Smith et al (2000) documented variability in the physiological tolerances of different genetic groups within the Kruger National Park in Africa, a far smaller geographic area than the continental United States. This opens up further research for improving our knowledge of strain specific environmental tolerances in North America and determining the biological significance of satellite-based vegetation data sets. Again, this serves to promote the use of GARP as an exploratory tool both for predicting distributions and also for understanding the biological relationships between species and the environment.

### 2.4.2 Anthrax Persistence in the U.S.

The geographic distribution of disease persistence expanded from 1915 to 1955. However, after the introduction of the anthrax vaccine, the distribution of case persistence was drastically limited, with only Texas and California consistently showing counties with positive cases from 1956 thru 1995. Between 1996 and 2005, anthrax persisted in Texas and re-emerged in the Dakotas Region. These areas were predicted with very high certainty by GARP. Additionally, the area of western central California was predicted by GARP, despite a lack of point data for model development. Outbreaks in the eastern states (e.g. Arkansas, Mississippi)
documented between 1915 and 1965 were areas not predicted by the models. This could indicate that either the environmental coverages used fail to capture a signal for presence in the eastern states, or that disease persistence in those areas was more likely related to continued introduction of spores from elsewhere (e.g. contaminated feed). Future work is needed to further explore this issue, though a solution is limited by the lack of bacterial cultures collected during those outbreaks.

The area of eastern Oklahoma used for model development (1957 outbreak; van Ness 1959) was not predicted by all of the subset models. A review of environmental conditions in that area suggest that northeastern Oklahoma is on the edge of the anthrax corridor and not suitable for long-term spore survival (e.g. acidic soils). This is partially substantiated by the lack out outbreaks in this part of the state since the 1957 outbreak.

West Texas and the Dakotas region remain areas of consistent outbreaks, with 2000, 2001, and 2005 all having large outbreaks in both wildlife and livestock in both regions (see also Table 1 in Coker 2002 for a list of outbreaks). Both of these regions are important economic centers for wildlife ranching and livestock agriculture. Likewise, both regions are predicted as highly suitable for anthrax spore survival. This suggests that long-term surveillance, reporting, and livestock vaccination are critical in these areas and should remain priorities for veterinary extension and animal health agencies.

2.5 References


CHAPTER 3 PREDICTING THE DISTRIBUTION OF ANTHRAX IN MEXICO

3.1 Introduction

Anthrax remains a problem for wildlife and livestock throughout North America (Hugh-Jones 1999, Hugh-Jones and De Vos 2002). Despite continued outbreaks in white-tailed deer, *Odocoileus virginianus*, and domestic livestock (horses, cattle, goats) along the Texas-Mexico border in recent years (2001 – 2005; Coker 2002, Blackburn, Chapter 2 – this dissertation) no published records are available on the geographic distribution of anthrax cases in Mexico. Additionally, the ecology of anthrax in the environment remains poorly understood (Smith et al. 2000, Coker 2002). Hugh-Jones (1999) defined Mexico as endemic for the disease based on annual country-level reports to the World Organization for Animal Health (OIE), but with the additional comment that higher resolution data would be useful to improve our understanding of the disease in Central America. Given this lack of known occurrence data, ecological niche modeling can provide an exploratory tool for predicting the potential distribution of *Bacillus anthracis*, the etiological agent for anthrax, in the context of its fundamental ecological niche (Hutchinson 1957, Raxworthy 2003). The use of predictive modeling in disease studies remains under-investigated, but studies that have employed such approaches to other disease systems have successfully provided new insights into their geography and ecology. For example, Peterson et al. (2004) developed a predictive model of filoviruses (e.g. ebola) in Africa and the Philippines using published case records and the Genetic Algorithm for Rule-set Production (GARP), a machine learning algorithm designed to relate species localities with environmental data. While these viral diseases can inflict severe mortality during outbreaks, little is known about the ecological niche or epidemiology of these viruses. Peterson et al. (2004) provided new distribution predictions, which may allow for targeted surveillance and collection of previously
under-evaluated reservoir and vector species. Likewise, Gilbert et al. (2005) employed a logitistic regression-based predictive model to predict the distribution and spread of bovine tuberculosis, another poorly understood disease, in Great Britain.

The purpose of this study was to demonstrate the use of the GARP modeling system to project the distribution of anthrax in Mexico, based on environmental relationships between anthrax case locations and environmental coverages modeled for the continental United States (U.S.). The capability to project modeled relationships from one geographic region to another is a powerful advantage of the GARP modeling system (Kluza and McNyset 2005), and it is has been successfully employed on a variety of taxa (Raxworthy et al. 2003, Kluza and McNyset 2005). The results of Chapter 2 showed the usefulness of the GARP system for modeling anthrax in the U.S. and this paper expands the methodology of that study as the first attempt at predicting anthrax in Mexico. This information will be used to identify areas of potential risk and suggest areas for future surveillance and ecological research to increase our understanding of this disease in the new world.

3.2 Materials and Methods

A GIS database of anthrax locations was developed for the continental U.S. from a series of outbreak investigations, official reports, and historical maps (for a complete list sources, see Chapter 2, Table 2.2, page 27). A data set of 332 confirmed anthrax locations representing carcass sites, pastures, and farm locations were derived for this analysis. The locations represented data from U.S. outbreaks from 2001 – 2005, with the additional inclusion of a single large outbreak from Oklahoma in 1957. Prior to model development, a 20% sub-sample (n=39) of randomly selected locations was excluded from model development to allow for independent post-hoc model accuracy assessment. This left 296 data points for model building.
Chapter 2 describes the methodology used to systematically sample environmental coverages for modeling anthrax distributions. The results of that study confirmed that satellite-derived temperature, precipitation, elevation, vegetation (NDVI), and interpolated ground-observations of soil pH were important ecological parameters for describing anthrax. Additionally, those modeling efforts identified mean NDVI as the single most important descriptor for anthrax in the U.S. For this study, mean annual temperature, annual precipitation, elevation, and four measures of NDVI were used. The availability of soil pH data was not available for Mexico at the same resolution as the U.S. For this study, temporal fourier processed NDVI provided four measures: 1) mean NDVI, 2) annual minimum NDVI, 3) annual maximum NDVI, and 4) annual amplitude NDVI. All environmental coverages were processed to develop a coverage set for the continental U.S. for model development and for Mexico for model projection. All coverages were standardized to 0.10 degrees$^2$ (approximately 8km$^2$) spatial resolution.

The GARP modeling system is defined elsewhere in great detail in Chapter 2 of this dissertation. Briefly, GARP is a machine learning algorithm that relates point location data to environmental coverages that describe ecological space (Stockwell and Peters 1999; www.lifemapper.org/desktopgarp). The modeling system iteratively develops a set of heterogeneous rules (logic strings) from a super-set of modeling approaches (e.g. range rules, logit regression, BIOCLIM; Stockwell and Peters 1999) that define the relationships between ecological parameters and known localities for anthrax. Initially, point locations are partitioned into training/testing sets that are used intrinsically to test and refine rule development. GARP is genetic in the sense that random mutations can occur between rules to explore a wide range of ecological possibilities. Like evolution, mutations that lead to rules with high predictive
accuracy and statistical significance will lead to rule inclusion in the model and further rule
development. Rule modification and selection will cease when a maximum iteration is
completed for convergence between rules (the point where predictive accuracy is no longer
improved between rules). One misconception about GARP is that all rule modifications are
logistic regression rules; however, logit regression is only one of the multiple rule types with rule
sets being developed. It is therefore entirely plausible to have a complete rule-set with few or no
logistic regression rules in the final model set. As GARP is a stochastic modeling approach, and
as with many modeling approaches (e.g. Rogers 2006 in press), variation can exist between
models (and outputs). To account for this stochasticity, a series of models are produced, errors
are evaluated to optimize final model selection, and a series of models are then summated to
produce a prediction surface. Since the environmental parameters are GIS derived, model outputs
can be plotted in geographic space, and maps can be developed to evaluate the predicted
distributions of anthrax.

Initially 100 models of anthrax were developed with 1000 iterations per model run and
convergence set at 0.01. A 50% training/testing partition was used for internal model testing and
validation. The best subset procedure was used to optimize the final models selected (Anderson
et al. 2003). Within the subsets routine, extrinsic omission was set at 10%, meaning models with
greater than 10% of the testing points omitted by models would be excluded from the final model
set. Additionally, optimal models compromise between omission (known locations being falsely
predicted as absence points – over-fitted models) and commission (areas of absence falsely
predicted as presence – over-predicting models). The commission threshold was set as 50%,
truncating the upper and lower 25% of models selected under the omission threshold. A final
subset of 10 models was developed for the U.S.
Mexico predictions were developed by projecting the U.S. rule sets onto the environmental coverages for Mexico. GARP applies the rule-sets to define areas of presence and absence from the U.S. to areas with no known localities in Mexico. This differs from simply modeling the U.S. and Mexico in a single model run, as this would consider all of Mexico as being initially absent. This could obviously skew rule-sets against anthrax promoting environments in Mexico. Therefore, projects were produced for Mexico using the U.S. modeling parameters with the same modeling parameters and a 10 best model subset.

The 10 best model subset for the U.S. was evaluated using the Area Under the Curve (AUC) in a Receiver Operating Characteristic (ROC) analysis to evaluate predictive accuracy (McNyset 2005). The U.S. model set was independently evaluated with the 20% hold out sample (Fielding and Bell 1997). The AUC is a threshold-independent statistic, which plots specificity (absence of commission error) versus sensitivity (absence of omission error) from error components derived from a confusion matrix (Fielding and Bell 1997). In this framework, each independent test point is scored based on the number of best subset models that correctly predict presence for the given pixel that the test point lies in. A line is plotted and the AUC is calculated. As AUC increases (between 0.5 and 1.0), so does the certainty that the model is accurate, or the higher the sensitivity and specificity of the model (Hanley and McNeil 1982). The AUC of the model is compared to the AUC of a “line of no information” (AUC = 0.5) using a z-test. Values near 0.5 indicate that a model is predicting no better than a random model, where as an AUC of 1.0 would represent a perfect model (Centor 1991). A weighted omission was also calculated as the average proportion of the independent test points predicted by 0 – 10 of the models in the best subset (McNyset 2005). As no independent test data are available for
Mexico, accuracy metrics were produced for the U.S. and evaluated for usefulness prior to evaluating the results for Mexico.

GARP outputs binary grid surfaces of presence and absence that match the geography of the study areas. Presence is assigned a score of 1 and absences are defined by 0. To evaluate the certainty of model outputs, the 10 best models are summated to a single grid surface using the raster calculator routine in ArcGIS 9.0 (ESRI, Redlands, CA). A score of 0 indicates that no models predicted the area as present, and 10 models indicate that in all models the best subset predicted presence. Maps were produced for the U.S. and Mexico to evaluate the predicted geographic distribution of *B. anthracis*. Additionally, the rule-set writing and mapping application was used to evaluate the individual rules used to define anthrax presence or absence in the model runs and to map the geographic distribution of individual rules across Mexico.

### 3.3 Results and Discussion

The accuracy metrics indicated that GARP models of anthrax in the U.S. are robust (AUC = 0.88, SE = 0.035, $z = 10.43$, $p<0.001$). The weighted omission was 6.89%, indicating that the model was not over-fitting the predicted niche of anthrax. These accuracies have been substantiated across a wide range of taxa (Peterson 2001, Kluza and McNyset 2005, McNyset 2005), including disease studies (Peterson et al. 2001, Peterson et al. 2004, Adjemian et al. 2006).

The spatial distribution of anthrax in the U.S. is limited to a narrow corridor through the central states of the U.S. In general, anthrax is predicted to occur from central southwest Texas north to the Canadian border through the Dakotas. Three disjunct areas are also predicted to occur in central Arizona, California, and the eastern corners of Oregon and Washington (Figure 3.1). The distribution of Mexico is predicted to expand the area in Texas south into the
northwestern corner of Coahuila de Zargoza, most of Nueva Leon, and the western central portion Tamaulipas (Figure 3.2). A second disjunct area was predicted for central Sonora. The northwestern corner of Baja California Norte was predicted as an extension of the disjunct region of California. The southern tip of the Baja peninsula was also predicted as anthrax habitat.

The rule-set map in Figure 3.3 illustrates the conservative nature of rule development in GARP. Despite the potential for stochasticity and a large number of rules per model, the environmental signal for anthrax presence and absence was captured in only a few rules. The rule-sets for this modeling experiment indicate that anthrax in Mexico is primarily defined by two range rules limiting the distribution to a narrow band of NDVI. Figure 3 illustrates the geographic distribution of the dominant rules from a single model output and demonstrates that the eastern portion of presence is predicted by a narrow band of mean NDVI values, while the western region is defined by this rule and a rule limiting the NDVI amplitude. Absence rules are generally broader, as they represented a greater geographic area. The east and western coasts of Mexico were predicted absent by a single rule, while a second rule defined absence in north central Mexico. The east coast of Baja was primarily defined absent by the same rule that defined central Mexico, while the west coast was primarily defined absent by the mainland coastal rule. This indicates that the relationship between NDVI and anthrax is strong and clearly detected by the model. GARP is conservative in defining anthrax presence with few rules and limited number of environmental coverages, primarily a narrow range of NDVI (vegetation), temperature, and precipitation. The rule-sets from Mexico agree with those of the U.S. model development stage. Likewise, these models agree with those of Chapter 2, assigning NDVI as the most important descriptor variable, despite having slightly different coverage sets between the two studies.
Figure 3.1. Predicted distribution of anthrax in the U.S. and Mexico. Color ramp indicates model agreement across the 10 best model subset. Green points indicate anthrax localities used in model building, yellow points indicate independent test data.
Figure 3.2. Close up of the GARP predictions for anthrax in Mexico.
Figure 3.3. Rule set map of dominant presence rules (red) and absences rules (blue) for Mexico. A- Prediction map for this rule set. B – GARP prediction for Mexico. P indicates presence rules and A indicates absence rules.
This study presents the first attempts to model the potential distribution of anthrax in Mexico. Though no independent records are available to verify these predictions, the high accuracy metrics for the U.S. study justify its subsequent use to project the distribution in Mexico. While these results are preliminary, they do suggest that the most likely areas of anthrax presence in Mexico occur in close proximity to the U.S. Given the persistence of outbreaks in wildlife and livestock in Texas along the border, it is important that U.S. and Mexican officials and researchers work together to ensure that neither country lacks surveillance or disease control measures in these areas, as outbreaks in either country could easily cross the border and impact farms or ranches in both countries. This study provides a new potential distribution of anthrax in Mexico and opens the door for new international research and surveillance efforts. While data are required for validating the models presented here, this study certainly identify areas of potential anthrax risk that should be of interest to public health officials in both countries.

3.4 References


4.1 Introduction

Anthrax remains a problem for both wildlife and livestock in parts of North America, particularly in Texas (Hugh-Jones and De Vos 2002). Despite being a disease of antiquity, the ecology of *Bacillus anthracis*, the etiological agent of anthrax, remains poorly understood (Smith et al. 2000). While anthrax is primarily defined as a soil-borne bacterium (van Ness 1971, Smith et al. 2000), there is limited evidence for the potential role of both necrophilic and biting insects, primarily hematophagous flies, as potential vectors for the disease (Gates et al. 1995). De Vos (1990) confirmed the role of necrophilic flies in case multiplication in the Kruger National Park in Africa during an outbreak in greater kudu, *Tragelaphus strepsiceros*, by the movement and deposition of anthrax through fly defecation on vegetation regularly fed on by the affected population. Likewise, a series of studies have confirmed the successful isolation and transmission of anthrax by biting flies under both field and laboratory conditions. Ganeva (2004) summarized a number of fly studies specific to anthrax and documented at least 21 species representing five genera from the Tabanidae family that have been found experimentally to mechanically transmit anthrax bacilli (on body parts – e.g. legs and mouth parts).

Hypotheses on the role of biting flies in the transmission of anthrax date back to the early part of the last century. Morris (1918) conducted a series of experiments confirming that lethal concentrations of bacilli could be transmitted to guinea pigs by flies of the genus *Tabanus*. This study indicated that transmission was most successful within the first 4 hours following a blood meal. Turrell and Knudson (1987) found similar results for *Stomoxys calcitrans*, another species of biting fly, and two species of mosquitoes under experimental conditions. This latter study confirmed that flies allowed to feed on infected guinea pigs could successfully inflict a lethal
infection to disease-free guinea pigs through mechanical transmission. Turrell and Knudson (1987) indicated that transmission was most successful within the first four hours of a blood meal, but indicated that sample sizes were not large enough to definitely limit transmission potential to the first several hours after a blood meal.

While the work of Turrell and Knudson (1987) was limited to laboratory experiments, there is evidence for transmission of bacilli during an outbreak. Mohiyuddeen and Krishna Rao (1958) isolated bacilli in the blood smears of cutaneous lesion on affected bovines during an outbreak in India. Additionally, biting flies collected on those anthrax positive cows tested positive for bacilli. Other field studies have implicated biting flies in wildlife outbreaks, though no active sampling was completed to confirm the presence of bacilli in flies. Broughton (1987) and Gates et al. (1995, 2000) anecdotally documented the association of high biting fly numbers and high numbers of anthrax deaths in wood bison, *Bison bison athabascae*, in northern Canada. Additionally, Olsufjev and Lelep (1935 cited by Ganeva 2004) associated high anthrax case numbers in livestock with high biting fly numbers. This study indicated that the peak in anthrax cases followed a 10-day delay in fly numbers. The anthrax outbreak reported by Mohiyuddeen and Krishna Rao (1958) was suggested to be a rapidly spreading outbreak with a large geographic distribution by its completion. Braack and De Vos (1990) confirmed that the movement of anthrax bacilli by necrophilic flies was limited to vegetation in close proximity to carcasses (e.g. several meters), despite the potential for movements of greater than 30 km by individual flies. Hugh-Jones and De Vos (2002) acknowledged that many outbreaks in wildlife, or outbreaks that may involve biting flies, travel geographically in a “wave” with high cases early on in the outbreak and declining numbers of infected animals towards the end of the
outbreak. Hugh-Jones and DeVos (2002) speculated that transmission by biting flies could increase the spatial footprint of an outbreak.

Despite a number of studies implicating flies in outbreaks and a high diversity in the species capable of transmission, no known studies have quantified the spatial distribution of biting fly populations or densities during an anthrax season. With the exception of the work by Braack and De Vos (1990), few studies have characterized the spatial patterns of flies in relationship to anthrax outbreaks; however, this study was limited to necrophilic flies. This lack of spatial analysis has also been noted in the study of other arthropod vectors (Jeffery et al. 2002). However, the growing field of spatial analyses and geospatial statistics (see O’Sullivan and Unwin 2002) lends itself to spatially explicit analyses in epidemiological studies. For example, Jeffery et al. (2004) employed kriging, a probabilistic interpolation method, and a measure of local spatial autocorrelation to define the distribution of mosquito vectors for Ross River and Barmah Forest viruses in Australia. Sciarretta et al. (2005) employed similar kriging methods and spatio-temporal mapping to model the seasonal variation in tsetse flies (Glossinia spp.) in Ethiopia.

This current paper presents a pilot study to determine the usefulness of local spatial statistics in the analysis of biting fly collections within the endemic anthrax region of Texas. The goal of this study was to test the null hypothesis that biting flies exhibit a uniform distribution across the study area. Kernel density analysis was used to interpolate the density of fly populations across a ranch with well-documented anthrax cases in recent years. Additionally, the Getis’ $G_{i}^{*}$ statistic was employed to determine if fly catch rates were spatially clustered, and if so to determine the spatial scale at which clustering occurred and in what environments. Both analyses were conducted over multiple time periods to determine the temporality of clustering.
throughout the anthrax season. In addition, this study aims to identify spatial overlap between biting flies and known anthrax-positive carcass locations.

4.2 Materials and Methods

Biting fly collections were conducted on a white-tailed deer, *Odocoileus virginianus*, ranch located approximately 75 km north of Del Rio, Val Verde County, Texas. The ranch is comprised of nearly 7,406 hectares and is situated within the North American endemic zone for anthrax (Hugh-Jones and De Vos 2002; Figure 4.2 insert). Well documented anthrax outbreaks have occurred on the ranch in recent years, and active surveillance for dead deer has been conducted by ranch staff and researchers annually during the anthrax season (~May – October) since 2001.

Biting fly sampling was conducted over three sampling periods between 5 June and 21 August 2005 using seven un-baited Nzi fly traps (www.nzitrap.com; Figure 4.1). The Nzi is a nylon/polyester net trap with counter-shaded blue and black panels designed to attract flies through black body attraction (Mihok 2002). The top of the trap is translucent net mesh that allows natural light in. Once a fly enters the trap, it flies upwards to a small opening in the net and into a series of 2 liter trap bottles, eventually ending up in a plastic catch bag (Figure 4.1). Traps were setup for approximately 24 hours at each sampling site. Forty sampling sites were selected prior to the collection period to represent the geographic area of the ranch and the variety of habitats across the study area. The order of sites was randomly selected through the first setup period and then repeated throughout the sampling season to ensure that the time periods and spatial locations were represented near equally. Individual sample bags of flies were frozen at -30°C until they could be sorted and counted. Temperature, humidity, and wind speed were collected at the start and end of each sampling event using a handheld digital meter (Kestrel...
Instruments Model 4100, Neilsen-Kellerman, Boothwyn, PA). Wind speed readings were averaged for 1 minute and recorded as meters/second. Wind direction was measured with a handheld wind vein and recorded in degrees between 0 and 360 using a handheld compass.

In addition to environmental conditions collected at the time of trap setup and recollection, elevation for each site was derived from a USGS 30 m Digital Elevation Model. A simple six class vegetation classification was developed from a LandSat 7 Thematic Mapper image captured in March 2003. A 20-class unsupervised classification was conducted in Erdas Imagine 8.7 (Leica Geosystems, St. Gallen, Switzerland), and was reduced to six classes using local knowledge from ranch management staff, 5m DOQQ aerial photos, and GPS-based ground truthing surveys conducted in January and June 2005. Habitat type and elevation values were extracted from raster imagery in ArcGIS 9.0 using the zonal statistics routine in the Spatial Analyst Extension (ESRI, Redlands, CA).

Given the great diversity in species that can transmit anthrax, total fly counts were pooled for this pilot study. Flies were thawed, sorted into hematophagous flies, necrophilic flies, and other insects, based on head and body morphology (e.g. eye shape, mouth part shape) following Goodwin and Drees (1991) and counted as total numbers. To account for the variation in fly numbers that could be associated with differences in sampling event lengths, each fly count was divided by the total time the trap was set out uninterrupted. Therefore, fly catch rates were defined as:

\[
\text{Catch rate} = \frac{\text{Total biting flies}}{\text{decimal hours of trapping event}}
\]

Descriptive statistics were calculated for each of the continuous environmental variables for the forty-two sampling sites for each setup period. A total of four setup periods were
completed during the sampling season; however, Setup 4 was an additional set of traps within time Setup 3, therefore both represent a single time period.

*Spatial Analysis*

Kernel density analysis was used to plot the spatial distribution of fly densities (standardized catch rate values) for each of the setup periods. Kernel density analysis is a deterministic interpolation technique for calculating weighted densities of events over a gridded surface within a kernel, or spatial bin (Fotheringham et al. 2000). The most commonly applied version of the analysis approximates the standard power-based distance decay curve. Whereby, events occurring closer to the centroid (trap sites with catch rates) exert greater influence on the density calculation than those closer to the kernel edge. Kernel density analysis was performed with the Spatial Analyst Extension for ArcGIS 9.0. The output grid cell size was set at 100 m and a fixed bandwidth of 1000 m was used for the search radius. Hotspots for hematophagous flies were defined as those areas having high catch rates for that given setup period expressed in a distribution for all sampling sites for that time period. Therefore, a hotspot in any one setup period is only reflective of that single period and not reflective of fly occurrence in pre- or post-setup periods. Surface maps were produced using the standard deviation (SD) of the density values. Only those regions that were greater than or equal to 2SD from the mean were considered as hotspots. Anthrax positive deer carcass locations from 2002-2004 and suspect cases from 2005 were overlain on each density surface. Additionally, density surfaces were overlain on the habitat map and the DEM to illustrate the environments where fly concentrations were the greatest.

The Getis-Ord statistic $G_i^*(d)$ (Getis and Ord 1992; Ord and Getis 1995) was used to determine if significant local spatial clustering occurred for any of the first four setup periods.
The $G_i^*(d)$ statistic tests for local spatial clusters in group-level data and assesses the association of the variable of interest within a set distance of each observation in the data set tested (Durbeck et al. 2002). The $G_i^*(d)$ is useful for identifying individual members of local cluster events (Getis et al. 2003). The $G_i^*$ statistic is written as:

$$G_i^*(d) = \frac{\sum_{j} w_{ij}(d)x_j}{\sum_{j} x_j}$$

in Clusterseer2, where $x_j$ are the catch rates of flies, $w_{ij}$ is a binary (0, 1), symmetric weights matrix with 1 for all $j$ within distance $d$ of point $i$ with all other values being 0.

Two $G_i$ statistics can be calculated, $G_i(d)$, where all values within $d$ of the $i^{th}$ observation excluding $i$, and $G_i^*(d)$, where the $i^{th}$ observation is included in the calculation. $G_i^*(d)$ statistics were calculated for the hematophagous fly catch data set using the ClusterSeer2™ software application (www.terraseer.com). Clusterseer2 calculates the statistic for spatial data in a shapefile format meaning analysis can be performed on data created directly in ESRI GIS software. Probability is calculated in two ways in ClusterSeer2: 1) the $G_i^*(d)$ statistic is calculated and output as the standard normal variant with an associated probability from the z-score distribution (Getis and Ord 1992), and 2) a Monte Carlo randomization procedure can be performed to calculate probability from a generated distribution that simulates complete spatial randomness (CSR), where the user sets the number of simulation runs (1000 for this study).

Four distances ($d$) were selected for this study. The smallest distance, 500m, was selected to capture localized densities, representative of individual trap sites and nearest neighbors. Distances were then set in 500 m increments to 2,000 m. The largest distance, 2000m, was selected to capture a greater percentage of the study area in each calculation. As the $G_i^*$ values are normal variants of the z-distribution, only those $G_i^*$ values greater than 1.96 and significant...
at \( p < 0.05 \) in the Monte Carlo simulations were considered significant. Following Getis et al. (2003) cluster membership was defined by significant \( G_i^* \) values at the distance with the maximum \( G_i^* \) score. For a trap site to remain a member of a statistically significant cluster from one distance to another, the \( G_i^* \) value must increase from test distance size to test distance size. If the \( G_i^* \) value did not increase with distances, though \( G_i^* \) values may be statistically significant, they were not considered members of clusters. Getis and Aldstadt (2004) defined this as the critical distance, \( d_c \). All clusters presented in this study are defined at \( d_c \). The \( G_i^* \) can be evaluated as a two-tailed test with negative values representing clusters of low values (Getis and Ord 1992). Maps were produced to illustrate the spatial distribution of biting fly clusters. Anthrax positive and suspect carcass locations were overlain to show the relative distribution of disease cases relative to fly clusters.

Figure 4.1. Nzi fly trap setup. Flies enter through the opening between the black patches, fly upwards towards the light and get trapped in the bottles. Insert shows flies trapped in the bottles and in the catch bag.
4.3 Results

A total of 5,114 biting flies were collected during 113 uninterrupted trapping events. Setup 1 spanned the period 5 – 11 June 2005 and included 40 sampling sites. Setup 2 spanned 10 – 16 June 2005 and included 40 sampling sites. Setup 3 spanned 12 – 21 August and included 23 sampling sites. Setup 4 represents a sub-sample of sites that were sampled again on 15 August 2005 (a time period when multiple deer carcasses were being found). Descriptive statistics for the four setup periods are listed in Table 4.1.

Table 4.1. Descriptive statistics for the three time periods analyzed in this study. Note that Setup 3 and Setup 4 occurred in the same time period.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Setup 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td>40</td>
<td>388</td>
<td>564</td>
<td>477</td>
<td>43</td>
</tr>
<tr>
<td>wind speed (m/s)</td>
<td>40</td>
<td>1.1</td>
<td>5.7</td>
<td>2.8</td>
<td>1.3</td>
</tr>
<tr>
<td>wind direction</td>
<td>40</td>
<td>76</td>
<td>200</td>
<td>141</td>
<td>26</td>
</tr>
<tr>
<td>temperature C</td>
<td>40</td>
<td>22.5</td>
<td>37.5</td>
<td>29.3</td>
<td>3.9</td>
</tr>
<tr>
<td>humidity (no units)</td>
<td>40</td>
<td>35.6</td>
<td>84.6</td>
<td>59.8</td>
<td>15.0</td>
</tr>
<tr>
<td>catch rate (flies/hr)</td>
<td>40</td>
<td>0.00</td>
<td>9.26</td>
<td>1.32</td>
<td>1.68</td>
</tr>
<tr>
<td><strong>Setup 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td>38</td>
<td>417</td>
<td>564</td>
<td>477</td>
<td>41</td>
</tr>
<tr>
<td>wind speed (m/s)</td>
<td>38</td>
<td>0.4</td>
<td>4.1</td>
<td>2.1</td>
<td>0.9</td>
</tr>
<tr>
<td>wind direction</td>
<td>38</td>
<td>62</td>
<td>240</td>
<td>145</td>
<td>39</td>
</tr>
<tr>
<td>temperature C</td>
<td>38</td>
<td>1.7</td>
<td>33.5</td>
<td>29.9</td>
<td>5.4</td>
</tr>
<tr>
<td>humidity (no units)</td>
<td>38</td>
<td>28.5</td>
<td>83.1</td>
<td>53.3</td>
<td>12.8</td>
</tr>
<tr>
<td>catch rate (flies/hr)</td>
<td>38</td>
<td>0.00</td>
<td>10.21</td>
<td>2.42</td>
<td>2.67</td>
</tr>
<tr>
<td><strong>Setup 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td>23</td>
<td>425</td>
<td>549</td>
<td>480</td>
<td>37</td>
</tr>
<tr>
<td>wind speed (m/s)</td>
<td>23</td>
<td>0.7</td>
<td>5.1</td>
<td>2.3</td>
<td>1.0</td>
</tr>
<tr>
<td>wind direction</td>
<td>23</td>
<td>73</td>
<td>198</td>
<td>143</td>
<td>29</td>
</tr>
<tr>
<td>temperature C</td>
<td>23</td>
<td>27.6</td>
<td>36.7</td>
<td>32.5</td>
<td>2.4</td>
</tr>
<tr>
<td>humidity (no units)</td>
<td>23</td>
<td>32.8</td>
<td>65.2</td>
<td>49.0</td>
<td>8.9</td>
</tr>
<tr>
<td>catch rate (flies/hr)</td>
<td>23</td>
<td>0.10</td>
<td>4.88</td>
<td>1.78</td>
<td>1.43</td>
</tr>
<tr>
<td><strong>Setup 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td>8</td>
<td>451</td>
<td>536</td>
<td>479</td>
<td>28</td>
</tr>
<tr>
<td>wind speed (m/s)</td>
<td>8</td>
<td>1.3</td>
<td>2.9</td>
<td>1.9</td>
<td>0.5</td>
</tr>
<tr>
<td>wind direction</td>
<td>8</td>
<td>114</td>
<td>242</td>
<td>165</td>
<td>37</td>
</tr>
<tr>
<td>temperature C</td>
<td>8</td>
<td>24.8</td>
<td>32.4</td>
<td>28.7</td>
<td>2.7</td>
</tr>
<tr>
<td>humidity (no units)</td>
<td>8</td>
<td>49.4</td>
<td>81.4</td>
<td>63.2</td>
<td>12.2</td>
</tr>
<tr>
<td>catch rate (flies/hr)</td>
<td>8</td>
<td>0.04</td>
<td>8.49</td>
<td>2.38</td>
<td>2.94</td>
</tr>
</tbody>
</table>
The kernel density analysis identified a pattern of hotspots across time periods in the northeast central portion of the ranch (Figure 3.2). A single hotspot was stable across all four setup periods in this area, while other smaller hotspots fluctuated in size and position around this central hotspot. The areas of highest density overlap with the distribution of both positive and suspect anthrax cases (dead deer carcass locations), with both hotspots and carcass locations dominant on the eastern half of the ranch. With the exception of a single hotspot in the southwestern area of the ranch in setup period 3, no hotspots of high fly catch rates occurred on the western side of the ranch.

Figure 4.3 (Insert A) displays the kernel density hotspots overlain on the vegetation classification derived from the LandSat data. Figure 3.3 (Insert B) illustrates the kernel density hotspots over the DEM. In general, hotspots occurred in low-lying areas with dense vegetation centered on a large riverbed (dark green habitat class in Insert A and black values in Insert B). These two maps suggest that high elevation areas, such as steep slopes (the areas that divide the hotspots in the eastern portion of the ranch) and large upland areas (the large plateau with little green vegetation in the northwest region of the ranch) do not support fly habitat. Though all trap locations had flies during at least one single sampling event in this study, these highland areas typically had fewer than 10 totals flies per trapping event. In contrast, total catches on the eastern side of the ranch were often greater than 100 flies. In addition to having less vegetation coverage in the upland areas, wind speeds were generally higher in these areas. Lowland canyons did provide shelter during periods of high winds, often with wind speeds nearly an order of magnitude higher upslope. Figure 4.4 illustrates an interpolated wind surface of the ranch. There is clear delineation between the ridgeline along the northwest border of the ranch and the
Figure 4.2. Kernel density derived hotspots for biting flies during the anthrax season. Lighter colors indicate areas standard deviations between 2 and 3, darker colors indicate areas greater than 3 standard deviations. Insert indicates location of ranch in Texas. Stars indicate 2003 carcass locations, squares indicate 2004 locations, X’s indicate 2005 positive carcasses, open circles indicate anthrax suspect cases from 2005. Closed circles indicate the fly sampling locations used in each kernel density analysis.
Figure 4.3. Density surfaces over base maps. A- Kernel density hotspots (red outlines) overlain on habitat map. Gray colors represent non-vegetated areas; yellow to green color ramp represents a continuum from grasses to dense vegetation. B- Kernel density hotspots overlain on the DEM. Light colors indicate high areas (upslope) and dark areas indicate low areas (river valley, canyons). Stars indicate 2003 carcass locations, squares indicate 2004 locations, X’s indicate 2005 positive carcasses, and circles indicate anthrax suspect cases from 2005.
eastern river valley. Kernel density estimates are overlain on Figure 4.4 to demonstrate the relationship between wind speeds and fly catches.

The $G_i^*$ statistic identified local clusters of high catch rates in each of the four setup periods. The highest $G_i^*$ values for high catch rates at each of the four periods was at $d_c = 500$ m. Clusters of low values were detected at 2000 m for Setup 1 and Setup 2, but not for the other two periods. Figure 4.5 illustrates the distribution of clusters for both low and high values at both critical distance thresholds. Distances 1000 m and 1500 m are not displayed, as they were did not achieve the greatest $G_i^*$ values for any of the four setup’s. The $G_i^*$ clusters agree with the kernel density results and both indicate that the northeast central region (low lying areas)
Figure 4.5. $G_{k}^*$ cluster locations at 500m and 2000m. Red circles indicate clusters of high values; blue circles indicate clusters of low values. Gray circles identify the distance band around each trap site (gray symbols). Red symbols indicate confirmed anthrax carcass locations on the ranch between 2003 and 2005. Black symbols represent suspect anthrax sites from the 2005 outbreak. Note Setup 3 and Setup 4 were collected during the same time period.
promote high fly numbers, while the northwestern (upland) areas do not. In fact, the highland areas were identified as clusters of significantly low catch rate values at larger distances. The $G_i^*$ results agree with the kernel density hotspot analysis. Both analyses identified hotspots of high fly catches in the eastern low lying areas, with smaller scale low value clustering in the west. Additionally, both indicated localized clusters of high fly numbers around trap sites.

4.4 Discussion

The results from this preliminary study indicate that the distribution of biting flies is not spatially uniform across the study area and that clustering of biting flies occurs on the eastern end of the ranch. This coincides with the locations of anthrax positive and suspect deer cases between 2003 and 2005. Likewise, the distribution of suspect cases closely mirrors the location of deer carcasses found and burned in a large 2001 outbreak (greater than 100 individual deer) on the ranch that was not mapped with the precision of this current study (B. Culp, pers. comm.). This provides quantitative evidence that the fly/carcass associations observed by Gates et al. (1995b) for bison in Canada, may have an important role for deer in Texas.

The spatial techniques used here have been useful in other disease systems for identifying areas of high vector density and the scale of clustering. Jeffery et al. (2002) used a combination of kriging and the $G_i^*$ statistic to identify areas of abundant mosquito vectors for Ross River and Barmah Forest viruses in Australia. Like the results of this current study, Jeffery et al. (2002) found agreement between both techniques for identifying spatio-temporal hotspots of insect populations and used the Gi* statistic to identify clusters of both high and low catch values across weekly sampling intervals. Getis et al. (2003) also found spatio-temporal clustering at local scales for mosquitoes associated with Dengue Fever in Peru. Vazquez-Prokopec et al.

---

2 B. Culp is the current ranch manager of the study ranch and was primarily responsible for carcass disposal in the 2001 and subsequent outbreaks.
Cecere et al. (2005) identified significant clusters of high Triamtomine insect populations in rural Argentina using the $G_i^*$. Similar to this current study, maximum $G_i^*$ values were at local distances, suggestive of limited individual dispersal despite species-specific capacity for long-distance movements (over several kilometers). In this current study clustering or high fly catches were greatest (highest $G_i^*$ values) at a critical distance of 500m, suggesting that single traps and their immediate neighboring sites had a greater relationship to one another than distant sampling sites during any given sampling period. This may be an artifact of the sampling scheme, with some trap locations spaced farther apart than others (Figure 4.2A). However, this also makes biological sense, as Foil (1989) provided strong evidence that despite the potential for long-distance movements in biting flies (upwards of several hundred meters to several kilometers; tabanids are strong fliers), most individuals in mark-recapture studies traveled less than 50 m and often returned to the same host for multiple feedings. Though these results do not preclude flies from traveling several kilometers, it is plausible that flies will remain within as short a distance as possible to deer. Deer counts by ranch staff also indicate a greater number of deer on the eastern end of the ranch, which may support shorter travel distances between meals for individual flies. In this current study, the kernel density analysis confirmed hotspots of fly locations on the eastern side of the ranch in the three time periods. While there was some shifting of small hotspots around individual trap sites, a core area was identified along the large dry river bed that runs through the north central portion of the ranch (Figure 3.3A). This area had the densest vegetation (dark green habitat class in Figure 3.3A) and was at low altitude, providing refugia from high winds found in the northern upland areas (Figure 3.3B).

The $G_i^*$ values indicated clusters of low catch values along the northern boundary of the ranch (Figure 4.3) at a large critical distance (2000 m). A review of Figure 4.3B indicates that
the northern and western boundaries of the ranch are dominated by a high, sparsely vegetated ridgeline that provides minimal shelter for flies from direct summer sunlight and higher average wind speeds upslope. Figure 4.4 confirms that winds along this ridgeline are very high. Additionally, deer on this portion of the ranch would likely move vertically and horizontally to find patches of vegetation for shade, which could increase the search distance between blood meals. The combination of these factors makes the western and northern along the ridgeline less suitable for flies than the eastern river valley.

The spatial distribution of flies on this study ranch indicates a strong relationship between biting fly densities and known sites of anthrax deaths in deer. Like flies, anthrax spores are sensitive to direct heat from sunlight and more often associated with canyon bottoms in this part of Texas because of a higher percentage of organic material and calcium (van Ness 1971). The scenario described above indicates that the geography of the ranch lends itself to greater potential for deer/fly interactions and anthrax spore survival along the dry river valley and associated habitats. While it can not be ruled out that biting fly numbers and anthrax cases occur contemporaneously through mutually exclusive ecological mechanisms, the high overlap between biting fly clusters and anthrax does suggest a potential relationship that warrants future research.

While the clustering patterns of this study indicate greater fly numbers at small distances, no evidence is available on the flight distances of flies on this ranch. While Foil (1989) indicated that flies return to individual host animals for repeat blood meals, his study did indicate that flies can travel upwards of 200 m in search of a meal. Likewise, wind speeds did show diel variability and high valley winds could potentially force flies great distances. While Chapter 6 confirmed that individual deer have limited home ranges on this ranch, no data are available on
potential avoidance behaviors in this deer population. A combination of environmental conditions, deer avoidance behaviors, and 50 m or greater search patterns could quickly expand the spatial footprint of an outbreak if flies were exposed to anthrax bacilli during blood meals. This can be defined as the *space-multiplier hypothesis*.

This study confirms a positive spatial relationship between confirmed anthrax deaths and biting flies, no diagnostic tests were run to confirm *B. anthracis* in biting flies. Likewise, collections were not identified to species, given a lack of knowledge of which flies in Texas may carry bacilli. The summary of species in Ganeva (2004) suggests that mouth parts of several species are appropriate, and since transmission is mechanical (Turrell and Knudson 1987) the majority of fly species are likely capable of moving bacilli. This study indicates that the role of biting flies is under-appreciated and under-investigated in relation to anthrax. The spatial techniques applied here support a positive relationship between disease cases and fly densities and suggests a direction for future research in this area.

4.5 References


CHAPTER 5   THE CASE-MULTIPLIER HYPOTHESIS: POSITIVE CONFIRMATION OF ANTHRAX SPORES IN NECROPHILIC FLIES IN WEST TEXAS

5.1   Introduction

Necrophilic flies have been implicated as Bacillus anthracis spore carriers for more than a century (Buchanan 1907, Graham-Smith 1914). Graham-Smith (1914) experimentally fed viable spores to multiple life stages of flies and successfully re-isolated spores from the gut content and fecal deposits of all life stages, suggesting not only that flies could move anthrax spores, but that spores could survive in the gut of flies through life stages (e.g. larvae to adult). Additionally, Graham-Smith (1914) confirmed that spores can survive on body parts in the gut and feces of flies for extended periods of time (several to 24 hours on the body parts and upwards of 20 days in fecal matter). In an outbreak situation, flies feed on the body fluids and tissue of dead animals and pick up the spores on body parts or through ingestion. Once satiated or disturbed, flies travel to nearby vegetation and defecate or regurgitate and deposit spores, which may then be ingested by browsing ungulates (Hugh-Jones and De Vos 2002). These flies complete digestion by eating their regurgitated food and thus can lead to extensive contamination of the browse surrounding an infected carcass. De Vos (1990) confirmed that blowflies of the genus Chrysoma contaminated vegetation important to the diet of greater kudu, Tragelaphus strepsiceros, with spores and promoted anthrax outbreaks within the Kruger National Park in Africa. Likewise, the study confirmed that fly defecation near carcasses usually occurred on vegetation at heights preferential for greater kudu browsing feeding behavior. Hugh-Jones and De Vos (2002) suggested a similar anthrax transmission cycle of necrophilic fly contamination/infection for white-tailed deer, Odocoileus virginianus, in North America. However, no published data are available to confirm the presence of anthrax spores in necrophilic flies in the U.S. in recent outbreaks. This paper introduces preliminary findings on
the presence of anthrax spores in necrophic flies collected during the 2005 anthrax season (summer months) in West Texas and presents the *Case-Multiplier Hypothesis* to explain the role of non-blood feeding flies in the anthrax transmission cycle.

### 5.2 Methods and Results

Necrophilic flies were collected from the carcasses of eight dead deer between 5 June and 15 October 2005 on a ranch in west Texas with well documented previous anthrax exposure. This ranch, approximately 50 miles north of Del Rio, has seen repeated cases of anthrax in deer between 2001 and 2005. Large outbreaks were documented in 2001 (Coker 2002) and 2005, with smaller intermittent incidents of a single or few (2 or 3) cases in non-outbreak years (see Chapter 6 for 2005 case data). Flies were collected by holding a sterile whirlpak® bag over a natural skin tear in the axillary region of the carcass. As flies would leave the body cavity, they would exit the wound and fly upwards into the bag. The bag was closed after several individuals (e.g. 5 – 10 count) entered. Flies were frozen at -30°C in the field and transported to the diagnostic laboratory at Midwestern Research Institute for diagnostic work-up. Prior to laboratory analysis, flies were identified as members of the Calliphoridae and Sarcophagidae (Diptera) families by the Department of Entomology at Louisiana State University.

Flies from a single sampling bag were placed in a large glass test tube and ground with a tissue grinder. 100 µL samples from each bag were then transferred to blood agar plates and incubated at 30°C for 24 hours. A single colony was isolated from one of the eight representative samples of flies collected. The colony was non-hemolytic and showed atypical morphology on blood agar, lacking the typical ground glass appearance (colony appeared smooth). The colony was sub-cultured and re-plated on tryptacase soy agar. The sub-culture showed typical morphology and was subjected to microbial biochemical characterization
(Microlog, BiOLOG, Hayward, CA) and confirmatory PCR (PathAlert, Invitrogen Corporation, Carlsbad, CA).

This preliminary result confirms that necrophilic flies do interact with anthrax-positive deer carcasses in West Texas and may therefore play a role in outbreak promotion. Whatever the role of necrophilic flies, they probably do little to expand the spatial extent of anthrax outbreaks, given that these species tend to deposit the greatest amount of fecal material within the immediate vicinity of carcasses, despite the potential to move over 30 km (Braack and De Vos 1990). It is more plausible that necrophilic flies move viable spores from carcasses on the ground to vegetation preferred by deer, likely increasing the number of deer that will be infected by spores from a single index carcass. This can be defined as the Case-Multiplier Hypothesis. This hypothesis, might partially explain the rapid onset of cases in a wildlife outbreak discussed by Hugh-Jones and De Vos (2002). As many of the wildlife species infected in Africa and North America are browsers, it could be that browsing spores is more a realistic transfer mechanism than grazing spores and ingesting soil. This suggests that further research is warranted to quantify the role of flies in North American wildlife outbreaks.

5.3 References


CHAPTER 6  HOME RANGE CALCULATIONS FOR WHITE-TAILED DEER IN AN ANTHRAX AREA

6.1 Introduction

Anthrax disease remains a problem for both wildlife and livestock in parts of North America, particularly wildlife in Texas (Hugh-Jones and De Vos 2002). Prin and Weyerhaeuser (1987) considered anthrax a major component in population control and natural selection in small game reserves in Africa. This study also suggested that an in-depth understanding of the ecology of anthrax is critical for successful game reserve management. Despite these suggestions, the ecology of Bacillus anthracis, the etiological agent of anthrax, remains poorly understood (Smith et al. 2000). This is particularly true with respect to transmission cycles and pathways of infection in wildlife species. There is little direct evidence available to determine the specific routes of anthrax infection in free ranging animals. Gates et al. (2001) suggest that spore ingestion or inhalation are the most likely sources of infection for ungulates. However, this suggests that species feed on vegetation close to the ground. This is certainly plausible for grazing feeders – individuals that feed on grasses and near the soil, such as bison, Bison bison, or domestic cattle. Cattle, for example, use their tongue to pull vegetation out of the ground and ingest soil coated roots. However, the most common wild species affected in North America is the white-tailed deer, Odocoileus virginianus, a preferential browser with a potentially limited appetite for grasses (Halls 1984, Whittaker and Lindzey 2004). This suggests that other ecological factors may be involved with disease transmission in non-grazers.

Gates et al. (1995) suggests that individual behavior may play a role in infection. For example, bison outbreaks in the northern plains of Canada are often gender-biased with a greater number of males dying than females. While no direct evidence has been presented to date, Gates et al. (1995) suggest that male infection may occur due to territorial behaviors, where males
defend a large wallow – soil bath – and actively dig and kick up dust. The hypothesis suggests
that males interact with spores in the wallows. However, attempts to isolate spores in wallows
have proven difficult. Lindeque and Turnbull (1994) found similar results for a number of
African ungulate species in Namibia; again suggesting that some behavioral mechanism or
physiological tolerance triggers a greater number of infections in males. However, Gates et al.
(2001) suggest that gender-specific physiology is unlikely to affect infective dosages, as anthrax
occurs in summer when most species are out of breeding season and both genders are under near
equal environmental stress from heat and lack of rain. Likewise, a review of carcasses surveyed
in Texas from 2001 – 2005 indicated that deer sex ratios are near equal during anthrax outbreaks.

While *B. anthracis* is primarily defined as a soil-borne bacterium (van Ness 1971, Smith
et al. 2000), there is limited evidence on the potential role of both necrophilic and biting insects
as potential vectors for the disease, primarily hematophagous flies (Gates et al. 2001). This
poses a second pathway of infection that could impact wildlife, again with minimal field
observation or behavioral research. De Vos (1990) confirmed the role of necrophilic flies in case
multiplication in the Kruger National Park in Africa during an outbreak in greater kudu,
*Tragelaphus strepsiceros*, through the movement and deposition of anthrax in defecation
deposited on vegetation regularly fed on by the affected population. Hugh-Jones and De Vos
(2002) theoretically extended this transmission mechanism to North America, as white-tailed
deer are browsers. Chapter 5 of this dissertation supports this hypothesis in Texas, with the
positive isolation of *B. anthracis* collected on necrophilic flies feeding on a dead deer carcass.
However, no work has been done to determine the infective dose for deer, nor the spore potential
in individual flies or their feces or regurgitation.
Mohiyuddeen and Krishna Rao (1958) did isolate bacilli in the blood smears of cutaneous lesions on affected bovine during an outbreak in India. Additionally, biting flies collected on anthrax positive cows tested positive for bacilli. Other field studies have implicated biting flies in anthrax outbreaks, though no active sampling was completed to confirm the presence of bacilli in flies. Broughton (1987) and Gates et al. (1995) both documented the association of high biting fly numbers and high numbers of anthrax deaths in wood bison in northern Canada. No known work has been done on the interactions of biting flies and deer in the endemic zone of Texas (Hugh-Jones and De Vos 2002). Additionally, no studies are available on the spatial relationships between biting flies and wildlife with respect to anthrax for North America. Likewise, no work is available on the movement patterns of wildlife with respect to anthrax in North America.

Gates et al. (1995) suggest that animal behavior patterns are poorly understood with respect to anthrax infection and should be addressed more thoroughly. Realistically, behavior studies need to be coupled with studies that focus on the infection pathways and potential vectors. This is especially true in West Texas, where no work has been done to quantify the movement patterns of deer on ranches with well documented anthrax problems. Long et al. (2005) recognized the importance of animal movement behavior in disease transmission, and suggested that an understanding of deer movements and dispersal were important to understanding disease spread within populations.

One such mechanism for evaluating behavior space is the construction of individual animal home ranges. Burt (1943) defined the home range of an animal as the “area traversed by an individual in its normal activities of food gathering, mating, and caring for young.” This definition is still applied in modern movement studies and serves as a means of defining the
activity space of an individual, or if sampled across enough animals, the average activity space of the study population. Wildlife telemetry has become a popular tool for determining animal movements and can be powerful for exploring a variety of movement-based behaviors, such as habitat selection, migration patterns, sexual segregation, etc (White and Garrott 1990).

This current study follows the suggestions of Gates et al. (1995a) and Long et al. (2005) and presents the first home range estimates for white-tailed deer from a ranch in West Texas regularly impacted by anthrax. This study was a pilot study to evaluate the usefulness of radio telemetry to advance the current knowledge on the importance of behavior in anthrax outbreaks. The goal of the study was to describe the population structure of deer on the ranch, describe the sex ratios of affected animals, and to collect preliminary data on the home ranges of individual deer of both sexes during the summer months. The telemetry results are viewed in the context of anthrax carcass locations from 2001 – 2005 and the distribution of biting flies collected during the tracking period to evaluate movements in relation to other epidemiological components. By developing an understanding of deer movement, along with known carcass and fly distributions, two components of the meso-scale anthrax dispersal mechanism are examined.

6.2 Materials and Methods

Radio telemetry experiments were conducted on a 7,406 hectare deer ranch in Del Rio, Val Verde County, Texas (Figure 6.3 insert) from June 30 to 22 August 2005 to determine the summer home range of white-tailed deer with respect to the anthrax season. Ranch management staff provided deer population estimates derived from state regulated spotlight counts and helicopter surveys. Total population estimates were provided from 1995 to 2005. Additionally, anthrax surveys were available from 2003 – 2005. Population numbers and carcass totals were plotted to illustrate the relationship between the two. Based on the locations of disease cases
available, it appears the central and eastern region of the ranch are more prone to infection, while the northwestern and western regions appear to remain disease free. This provided an excellent research opportunity to compare these regions and determine the differences across space that might promote the disease in one region and deter the disease in another. Telemetry experiments were designed to evaluate whether or not the summer home ranges of deer were large enough to allow geographic overlap across the ranch, or if they were limited to smaller areas. Smaller home ranges could indicate that deer on the eastern side of the ranch may be at greater risk of anthrax infection than those on the western half.

Nine white-tailed deer were immobilized, radio collared, ear tagged, and released on the ranch between 30 June and 6 July 2005 (Figure 6.2). Deer were anesthetized with a combination of 5 mg/kg Xylazine (Rompun; Bayer Health Care, Germany) and 2.45 mg/kg Telazol\(^3\) (Ft. Dodge Animal Health, Ft. Dodge, Iowa) using a cartridge-powered dart rifle (http://www.pneudart.com). Telazol was reversed with an injection of Tolazine 45 minutes post-darting after the xylazine had worn off. Animals were handled at night to minimize capture stress and reduce the likelihood of stress-induced myopathy. Deer were monitored for several hours prior to their release in the same area of original capture. Deer were collared with 150 MHz Very High Frequency (VHF) radio collars with a 2 second pulse rate and mortality sensors (Advanced Telemetry Systems, Isanti, Minnesota). Individual collars were programmed to different radio frequencies and individuals were re-located using a 16-channel handheld radio receiver and a 3-element directional aluminum yagi. Deer tracking experiments were conducted simultaneously and in tandem with a tabanid fly study (see Chapter 4). The goal of this pilot study was to collect an adequate sample for calculating home range, so a target of 25 sample points was

\(^3\) Dosages were provided by Dr. Mike Vickers, a large animal veterinarian in South Texas. Dr. Vickers regularly immobilizes Texas white-tailed deer with this drug combination.
collected for each deer (White and Garrott 1990, Lesage et al. 2002). Animal re-locations were conducted on foot, by ATV, and by vehicle. Efforts were made to re-locate each animal at least once every 24 hours. Time of day was randomly selected for each animal to insure that spatial autocorrelation was avoided (Reynolds and Laundre 1990).

Animal positions were estimated using standard triangulation techniques (White and Garrott 1990). Briefly, collared animals were approached from the last known position. Once the signal was re-located, a GPS position and compass bearing were recorded. The user’s position was changed with respect to the signal and a series of position/bearing pairs were collected. A minimum of two positions are required to triangulate an animal, however, when possible, three or more positions were collected. When possible, animals were observed to determine body condition, collar position, and to glean insight into daily behavior patterns. Position error was estimated by collecting GPS locations of known positions (animals were spotted and exact locations were recorded) after a triangulation event and comparing estimates versus actual positions.

Animal triangulations and position error were analyzed using the GIS application Location of a Signal (LOAS v3.1.4 - www.ecostats.com). Triangulations were estimated using Maximum Likelihood estimators, when three or more bearing vectors were available for a position. When only two points were available, the best biangulation routine was employed. Animal position estimates were coupled with actual positions obtained in instances when animals were startled or when animal locations could be determined with a laser range finder and a compass bearing and offset in ArcView 3.2 software using the Distance/Azimuth Tools Extension v1.6 (www.jenessent.com). Additionally, positions were obtained from individual
animals from 1 – 22 August 2005 using motion-activated cameras set out by ranch management staff for an annual population census.

The Minimum Convex Polygon (MCP) estimator was used to determine the home range of collared individuals (Mohr 1947). The MCP defines the home range area by joining the outermost points in the distribution of animal positions. While the area of this measurement is likely to increase with sample size, it is an estimate that can be calculated with minimal samples sizes. MCP estimates were derived using Biotas v1.0.3 (www.ecostats.com), a GIS application specific to animal movement statistics. Home ranges were calculated in hectares and acres. MCP’s were overlain on an elevation surface derived from a USGS 30 m Digital Elevation Model. Likewise, MCP’s were overlain on a simple six class vegetation classification developed from a LandSat 7 Thematic Mapper image captured in March 2003. A 20 class unsupervised classification was conducted in Erdas Imagine 8.7 (Leica Geosystems, St. Gallen, Switzerland) and reduced to six classes using local knowledge from ranch management staff, 5m DOQQ aerial photos, and GPS-based ground truthing surveys conducted in January and June 2005.

Additionally, home range estimates were overlain on the results of a hotspot analysis for biting flies conducted during the same sampling period to evaluate the potential for spatial interaction between deer and possible anthrax vectors. MCP’s were also mapped in relation to known anthrax carcass locations from 2003 – 2005.

6.3 Results

The deer population estimates show a steady increase from approximately 640 animals to approximately 1360 between 1995 and 2001 (Figure 6.1). A large anthrax outbreak occurred in 2001 and ranch management estimated over 100 dead animals. Ranch management imported deer between the 2001 outbreak and the 2002 hunting season. A second large outbreak occurred
across west Texas in June - July and again in August – October 2005. The June – July outbreak was centered on the Town of Sonora, nearly 200 miles north of Del Rio. While no large outbreaks were reported from Val Verde County during the June/July outbreak, one positive anthrax case was recorded on the study ranch (see Chapter 5, this dissertation). A large outbreak did occur on the Ranch late in the summer, with cases first being found in early September (first case found 6 September 2005) and the last case being found in mid-October. While lab results are still tentative for anthrax, this outbreak coincided with laboratory confirmed cases throughout Val Verde County. No large outbreaks occurred between 2001 and 2005 on the ranch; however, ranch managers and researchers confirmed between one and seven cases per year in the interim periods (Figure 6.2). Carcass locations were recorded with GPS units and available for mapping.

Figure 6.1. Deer population estimates between 1995 and 2005. Arrows indicate periods of anthrax outbreaks and sporadic cases.
the distribution of cases for deer found between 2003 and 2005. The distribution of carcasses indicates that the eastern end of the ranch was more heavily impacted between 2003 and 2005 (Figure 6.3A). Though the 2001 outbreak was not GPS-mapped, the 2003 – 2005 distribution coincides with the distribution of those cases according to the ranch staff that disposed of carcasses.

Figure 6.2. Anthrax case numbers between 2001 and 2005 on a study ranch in west Texas.

Two collared animals (both males) died during this study prior to achieving the minimum number of telemetry positions and were removed from this study. Home ranges (MCP’s) were calculated for five does (female) and two bucks (males) during this study. Table 6.1 summarizes the gender, ages, time at liberty with collars and home range sizes for each individual. The mean
MCP was 72.8 hectares (179.8 acres) for all deer in this study. The MCP’s varied between individuals, but there was similarity between males and females. Deer 1 (2.5 year old buck) and Deer 5 (1.5 year old doe) had overlapping MCP’s throughout the study period (Figure 6.3A). In general, MCP’s were similar in size across the ranch, regardless of biting fly densities (Figure 6.3B). Fly densities were greatest along the large dry river valley and surrounding vegetation on the western side of the ranch. However, this appears to have little effect on home ranges.

Habitats across the ranch vary with elevation (Figure 6.4A, B). The western end of the ranch is dominated by higher elevations (a large ridgeline), sparsely vegetated slopes, and patchy canopy. In contrast the eastern end of the ranch is a dry river valley with more continuous dense canopy.

Ranch staff also supplements protein diet throughout the summer period to improve the health of the population. Figure 6.5 illustrates the relationship between home ranges and protein feeders. Motion cameras were used between 1 and 22 August and four of the seven deer were captured on film feeding at protein feeders (Figure 6.6). Deer 8 was photographed feeding at two feeders in a single night.

<table>
<thead>
<tr>
<th>Deer ID</th>
<th>Gender</th>
<th>Age</th>
<th>Collar Date</th>
<th>Last Location</th>
<th>Days at Liberty</th>
<th>Locations (n)</th>
<th>MCP (hectares)</th>
<th>MCP (acres)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>2.5</td>
<td>30-Jun-05</td>
<td>22-Aug-05</td>
<td>54</td>
<td>40</td>
<td>99.81</td>
<td>246.54</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>1.5</td>
<td>1-Jul-05</td>
<td>22-Aug-05</td>
<td>53</td>
<td>27</td>
<td>105.68</td>
<td>261.04</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>2.5</td>
<td>1-Jul-05</td>
<td>22-Aug-05</td>
<td>53</td>
<td>26</td>
<td>81.68</td>
<td>201.74</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>1.5</td>
<td>2-Jul-05</td>
<td>22-Aug-05</td>
<td>52</td>
<td>40</td>
<td>24.62</td>
<td>60.82</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>1.5</td>
<td>3-Jul-05</td>
<td>22-Aug-05</td>
<td>51</td>
<td>26</td>
<td>109.96</td>
<td>271.60</td>
</tr>
<tr>
<td>8</td>
<td>Female</td>
<td>2.5</td>
<td>6-Jul-05</td>
<td>22-Aug-05</td>
<td>47</td>
<td>29</td>
<td>29.81</td>
<td>73.63</td>
</tr>
<tr>
<td>9</td>
<td>Female</td>
<td>1.5</td>
<td>6-Jul-05</td>
<td>22-Aug-05</td>
<td>47</td>
<td>28</td>
<td>58.21</td>
<td>143.77</td>
</tr>
</tbody>
</table>
Figure 6.3. MCP’s (red lines) for seven deer tracked between 30 June and 22 August 2005. Insert – the location of the ranch in Texas. A – MCP’s in relation to known anthrax carcasses. Black dots indicate deer positions, open circles represent suspect carcass locations from 2005, red symbols represent confirmed cases between 2003 and 2005. B – MCP’s in relation to fly hotspots (see Chapter 4).
Figure 6.4. MCP’s (red lines) for seven deer tracked between 30 June and 22 August 2005. A – MCP’s in relation to habitat types. Black dots indicate deer positions, open circles represent suspect carcass locations from 2005, red symbols represent confirmed cases between 2003 and 2005. (gray represents hardpan and rock, light shades of green represent grasses while darker greens represent increasing vegetation height and density); B – MCP’s in relation to elevation. (dark colors represent low-lying areas, light colors represent increases in elevation).
Figure 6.5. Location of home ranges relative to protein feeders (green dots) and corn feeders (blue dots). Note corn feeders are only used during the hunting season and are not filled during the summer season. Protein feeders are filled continuously from May to August.
Figure 6.6. A collared deer captured by a motion-sensitive camera. This is a 1.5 year old female tracked on the western end of the ranch.

6.4 Discussion

The evidence from this study suggests that anthrax cases are present nearly every year in low numbers, with large outbreaks in years with appropriate climatic conditions (Hugh-Jones and De Vos 2002). This suggests that continued surveillance is necessary on an annual basis, and not just in response to large outbreaks. The annual infection of even a few animals means that new and/or recycled spores are introduced into the environment every year, increasing the likelihood of persistence. The sex ratio of the carcasses surveyed in the September 2005 outbreak were near equal (17 male:21 female, of those sampled). This is similar to the few cases reported from 2003 to 2005 and matches reports from ranch management staff on the 2001 outbreak that was not mapped.

The summer home ranges in this study are similar to other studies across the geographic distribution of the species. Lesage et al. (2000) suggests that this species generally exhibits strong fidelity to its home range and that deer are likely to use the same home range over several years. Ranch management confirms from motion cameras that bucks are often seen in the same
area or using the same feeders from year to year. Halls (1984) reported an average home range of 176 acres from white-tailed deer in west Texas. The average home range in this current study was 179 acres. Campbell et al. (2004) reported home ranges of 98.0 hectares for males and 79.0 hectares for females on a study site in the central Appalachian Mountains. These home ranges expanded to over 200 hectares in the fall during the breeding season. The ranch staff has recorded similar expansions from this current study from camera and field surveys, with some males being identified over 7 km from their summer time feeders. While the specific location of anthrax spores is unknown on ranch, the limited home range of white-tailed deer in summer suggests that the source of infection occurs within a small geographic area.

Evidence from this study and Chapters 4 and 5 suggests that if biting fly transmission occurs on the ranch it is more likely in the east. Additionally, the isolation of anthrax bacilli from necrophilic flies also occurred on the eastern side of the ranch. The limited home ranges of this study suggest that even if inundated by flies, deer are not likely to travel to the western side of the ranch to seek refuge. Likewise, these home ranges suggest that deer on the western end of the ranch are unlikely to interact with the western region during the summer period.

It is likely that deer on the eastern portion of the ranch may be susceptible to spore interaction from necrophilic flies, supporting the case-multiplier hypothesis. While these results do not confirm that deer ingest contaminated browse, the collection of spores from flies suggests this. Additionally, the limited home ranges presented here suggest that deer will remain in close proximity to carcasses once they fall. Figure 6.3A illustrates the location of deer carcasses found during the 2005 anthrax outbreak. There is no evidence from this study that home ranges shifted during the outbreak.
Figure 6.4A and B illustrate relatively little calculated difference in home range sizes between the eastern and western ends of the ranch, despite habitat and drastic elevation differences from west to east. The western end of the ranch is higher in elevation with less dense vegetation. While limited behavioral observations are available from this study, animals were as likely to be located up or down slope during this study period. More work is needed to determine the diel movement patterns within these home ranges to determine what affect this may have on interaction between flies and deer. However, the limited home ranges suggest that animals become infected within close proximity to daily activities and most likely die within that home range. Hugh-Jones and De Vos (2002) noted that wildlife outbreaks travel in waves with rapid onset and a fast increase in the spatial footprint of the outbreak. The home ranges presented here suggest that this is unlikely due to drastic shifts in movement patterns, especially given that home ranges did not change once the 2005 outbreak began. The overlap between biting fly hotspots and deer home ranges in the east may provide an interesting direction for future research to determine what specific ecological mechanism drives this “wave-like” spatial pattern of outbreaks. While these results are preliminary, the lack of deer movement across the ranch suggests that an alternative spatial transfer mechanism, such as possible fly transmission, results in the relocation diffusion of anthrax outbreaks.

6.5 References


CHAPTER 7       DISCUSSION, CONCLUSIONS, FUTURE DIRECTIONS

7.1       Discussion, Conclusions, and Future Directions

This dissertation encompasses a broad range of topics that together comprise several important components of the spatial ecology and epidemiology of anthrax disease in wildlife. Figure 7.1 provides an expanded version of figure 1.1 and lists in detail the relationships between each of the multi-scale approaches.

At the continental scale, this project has lead to the first predicted distribution of anthrax in the U.S. using the GARP modeling system. Likewise, this is the first study to report the GARP rule-set in detail to illustrate the application of rules and to define the environmental parameters that are most descriptive within the modeling system. The GARP system has received criticism for lacking biologically important information and has been considered a black box. (e.g. Rogers 2006 in press, Stockman et al. 2006). GARP has also been misinterpreted as a genetic algorithm that modified logistic regression based rules. However, the results of this study directly refute these arguments by clearly showing that GARP identified a relatively narrow band of values from only a few important variables for anthrax. This suggests that the decision process within GARP is conservative, even when an allotment of 50 potential rules per model is available for defining the ecological space of the target species. The rule-set for anthrax also indicates that logistic regression was not a dominant rule type for describing anthrax. Instead, range rules and negated range rules made up over 90% of the rules in the 10 model best subset in Chapter 2 and the majority of rules in the Mexico projections in Chapter 3. This makes sense from an evolutionary ecology perspective as well. Holt and Gaines (1992) define the ecological niche of a species as the ecological space utilized by the mean phenotype of the population. It should be rare to find a species that is defined by a single value, especially
when modeling packages are searching both ecological and geographic space. Space varies. The heterogeneous rule-set in GARP promotes an unbiased search of this ecological space and should be considered advantageous. Stockman et al. (2006) defined GARP as a black box, however, a detailed review of that paper clearly illustrates that the authors misused the application and lacked an understanding of the fundamental principles of ecological niches. It is hoped that this dissertation will illustrate the rule-set and processes in GARP clearly and negate some of these unfounded criticisms. Likewise, it is hoped that modelers will explore multiple possibilities in approaching geographic distribution studies and allow for more objective reviews of all techniques.

While this study provides the first estimates for anthrax distributions in the U.S. and Mexico, it is important to validate these models with field data. Internal validation metrics are extremely important to the modeler, but it is also critical to validate these models against biological data. With that stated, it is important to open communication between U.S. and Mexican scientists and public health officials and determine the true distribution of anthrax. As the predicted distributions straddle the U.S./Mexico border it is important to determine if outbreaks in either country lead to outbreaks in the other. Coker (2002) did report on the isolation of *B. anthracis* from samples collected across the Border from Del Rio, Texas in 2001. However, there are no data available on Mexico from 2002 to present. In light of the large outbreak in Del Rio in 2005 and the confirmation of small numbers of background infections in wild deer populations, it is likely that Mexico saw a similar scenario.

At the meso-scale, the combination of data from the positive *B. anthracis* in necrophilic flies (Chapter 5), the annual cases in wildlife (Chapter 6), and the positive spatial relationships between biting fly populations and anthrax cases (Chapter 4) it is confirmed that the case-
multiplier hypothesis is active in the North American anthrax transmission cycle and that the role of biting flies requires more scientific attention. Likewise, background infections, which were suggested to promote persistence by Hugh-Jones and De Vos (2002), have been confirmed in the U.S. However, during a large outbreak in the Dakota Region (Chapter 2) in 2005, there was no surveillance of wildlife beyond the affected farmed animals. The results of this study confirm that anthrax remains a problem in wildlife and that the disease is active in nearly all years, and not just in large sporadic outbreaks. This strongly suggests that wildlife surveillance should be considered an essential part of anthrax surveillance in future years, especially 2006 and 2007 (in direct response to the large outbreaks in North America in 2005).

Anthrax vaccination has proven useful in combating the disease in livestock (Hugh-Jones and De Vos 2002); however, this is not feasible in wildlife and again supports the need for increased wildlife surveillance for the disease. From an economic standpoint, Mörner et al. (2002) provided evidence that wildlife surveillance often leads to preemptive identification of zoonotic outbreaks and rapid response. Likewise, Bengis et al. (2002) argue that as wildlife farming and wild game hunting or eco-toursim expand, the interface between wildlife and livestock disease will increase. This is evident in west Texas where deer ranching has rapidly increased in recent years and become a larger component of local economies than it was a short time ago.

Surveillance is expensive. Therefore, efforts such as ecological modeling and disease mapping should be employed to target areas most likely to promote disease survival and persistence. Likewise, these modeling approaches can lead to new areas of research and surveillance that were not previously considered (Peterson et al. 2004). The ecological modeling in Chapters 2 and 3 were completed with a modeling package developed in the open source
Figure 7.1. An expanded version of figure 1.1 illustrating in detail the position of each study in this dissertation in the spatial ecology of anthrax.
software community and is available for no charge. Likewise, there were no charges incurred for the environmental coverages used. This is an attractive feature and it should be clearly stated that these approaches are available to the research community. This also means that costs should not prohibit developing nations from exploring similar techniques to identify new potential areas for surveillance and research where funding is even more limited.

The results of the GARP have identified geographic areas where surveillance should be considered. While the corridor between Texas and the Dakotas is well-known as an endemic area in the literature, the eastern expansion of this corridor into parts of Michigan, Illinois, and Ohio are poorly investigated. These results suggest that new research should be initiated with wildlife researchers and managers in these areas to determine if the disease is present in deer and either under- or not reported. Likewise, if the disease is not present in these areas, the GARP results suggest that the introduction of the disease may lead to long-term survivability in these areas. While determining the exact reasons for commission in GARP requires attention, commission provides insights into potential geographic areas suitable for anthrax that have been neglected.

There is also a need to determine how other environmental coverages, such as wildlife densities or land use, could improve model accuracy and possibly explain areas of commission and omission in the models. Likewise, the most recent version of the DG application allows for the user to define the geographic region where GARP can derive the 1,250 pseudo-absence points per model run. This has research potential for directly comparing pseudo-absence generation approaches in GARP and other modeling approaches. Likewise, this may serve as a potential means of reducing the asymmetry in the confusion matrix.
The results in this dissertation present a series of spatially explicit patterns that should be evaluated more closely. The movements of deer are limited, yet the disease tends to spread rapidly over great distances, suggesting a long-distance transmission mechanism may be active. The spatial relationship of biting flies and anthrax cases suggests that the role of biting insect requires more attention. Specifically, experiments are needed to: 1) confirm the presence of anthrax bacilli in the biting fly populations during outbreaks, 2) define the potential spore counts present on biting and necrophilic flies, or how capable species are of trans-locating and spreading the disease, 3) determine the lethal spore dose for white-tailed deer (LD 50), or how many spores are required to infect and kill, and 4) confirm the specific infection pathway(s) for white-tailed deer. Answering each of these questions will further our understanding of the disease and improve our efficiency at surveying and controlling the disease.

GIS has proven a useful tool in epidemiology (Pfeiffer and Hugh-Jones 2002) and in this dissertation. While the primary use of GIS in this dissertation was to execute advanced spatial techniques, GIS is also a powerful databasing tool. Database packages such as Oracle are continually increasing the ability to manage spatial data. Likewise, products like Google Earth (www.earth.google.com) are increasing the availability of user-friendly GIS packages. Surveillance work should take advantage of these tools. For example, the predicted surfaces from this study could be pushed out to Google Earth or another web-based mapping server to allow public health and veterinary officers across the U.S. and Mexico to review the predictions without the need for expensive software (Google Earth is free) or extensive training.

This dissertation has presented as many questions on the geography of anthrax, the role of flies in transmission, and the role of behavior in deer infection as it has attempted to answer. However, this dissertation has provided evidence that these leads are worth pursuing and that the
spatial relationships between anthrax cases and possible vectors are more than speculative and should be researched further. These further studies must be inter-disciplinary in nature, should include spatial analyses, and should include mechanisms for surveillance. For example, if fly spore carrying potential were available, coupled with the fly bite rates on deer, and a lethal dose for deer known, a spatially-explicit Individual Base Model (IBM) could be developed to model the daily movement potential of deer versus fly bites to simulate infection pathways. Likewise, other modeling systems, such as MaxEnt or discriminant function, could be improved by masking agricultural areas, including fly collection data, and cases from other countries.

7.2 References


VITA

Jason Kenna Blackburn was born March 9, 1977, in San Diego, California, to Scott and Joanne Blackburn. He grew up in the rural community of Boulevard, California, with his parents and his older brother (4 years to the day), Eric. Jason attended Mountain Empire High School in Pine Valley, California. Upon graduation in 1995, Jason attended Grossmont Community College part-time and worked as a tour guide and a marine mammal keeper at SeaWorld, San Diego. Jason attended Grossmont College and the University of California, San Diego part-time until August of 1999, at which time he moved to Baton Rouge, Louisiana, to complete his bachelor’s degree at LSU. Jason transferred to LSU as a junior and began taking classes in zoology and geography. Upon the completion of biogeography in the fall of 1999, Jason changed his undergraduate major to physical geography and completed that degree in the summer of 2001. Jason continued his education at LSU in geography and completed a master’s degree in geography under Drs. Andrew Curtis and Bruce Thompson in 2003. Jason continued to pursue geography and work for Dr. Curtis as doctoral student immediately following the completion of his master’s. Between his master’s and doctoral programs, Jason began to work extensively with Dr. Martin Hugh-Jones and changed his doctoral research topic to the study of spatial ecology of anthrax disease. Jason will complete his doctoral degree in May of 2006 and continue his work with Drs. Curtis and Hugh-Jones as a research associate in the World Health Organization Collaborating Center for Remote Sensing and GIS for Public Health, where he will expand his research on ecological modeling and wildlife diseases.