Integrating Geographic Information Systems and Ecological Niche Modeling into Disease Ecology: A Case Study of *Bacillus anthracis* in the United States and Mexico

Jason K. BLACKBURN

Spatial Epidemiology and Ecology Research Laboratory, Department of Geography, California State University, Fullerton, Fullerton, California

Abstract. This chapter provides an overview of geographic information systems, spatial analysis and spatial statistics, and predictive ecological niche modeling as they apply to disease ecology. I provide a conceptual model of the epidemiology and outbreak ecology of anthrax and the landscape ecology of the pathogen *Bacillus anthracis*. I apply Anselin's exploratory spatial data analysis process to these two components of the anthrax-transmission and spore-survival model. Spatial clustering statistics are reviewed in the context of outbreak epidemiology and potential mechanical vector transmission. I then provide a primer on ecological niche theory and apply ecological niche modeling to estimate the potential geographic distribution of *B. anthracis* on the landscape of the contiguous United States under current and future climate scenarios and to estimate the unknown distribution of *B. anthracis* in Mexico.

1. Introduction

This chapter will briefly introduce geographic information systems (GIS), geographic information science (GISc), data development with GIS and remote sensing, and predictive ecological niche modeling (ENM) and then illustrate their uses in investigating the spatial distribution of *Bacillus anthracis,* the causative agent of anthrax. Secondarily, I will define the epidemiology of this disease in wildlife and livestock and indicate uses of GIS and ENM that can enhance our understanding of the disease. Although this chapter is limited to the spatial ecology of anthrax, readers are encouraged to think of parallel applications to other disease systems and public health issues.

2. Epidemiology of Anthrax and Landscape Ecology of *Bacillus anthracis*

To illustrate the application of GIS, spatial analyses and statistics, and ENM to disease ecology, it is first necessary to review the ecology and transmission of *B. anthracis*, the

causative agent of anthrax and the model disease system for this chapter. It is also important to distinguish between the epidemiology of the disease and the ecology of the disease agent.

Anthrax disease is a continuing zoonosis in livestock and wildlife throughout many countries of the world [2, 3]. Although the disease still affects animal and human populations, its ecology is complex and its distribution remains poorly understood [2] despite recent efforts to model the spatial distribution of the agent [4]. The causative agent, *B. anthracis,* is a ubiquitous, spore-forming, Gram-positive bacterium known to survive in soils for long periods of time, resulting in some areas experiencing regularly occurring outbreaks [2]. Although a number of studies have addressed the physiological and ecological constraints on the species [2, 5–7], few studies have evaluated the geographic potential of the species and the distribution of environmental parameters that can promote spore survival and subsequent outbreaks. The study of the disease and the agent can be divided into two discrete bodies of research, which then can be further subdivided. Figure 1 illustrates a conceptual model of the transmission cycle for *B. anthracis* and recognizes these two research components. Similar work has been developed by Hugh-Jones and De Vos [3] for wildlife transmission in Africa with an emphasis on outbreak ecology once an animal has become infected. In this chapter, I attempt to expand this conceptual model with a discussion on the landscape ecology of the bacterium and the geographic limitations on *B. anthracis* in the contiguous United States.
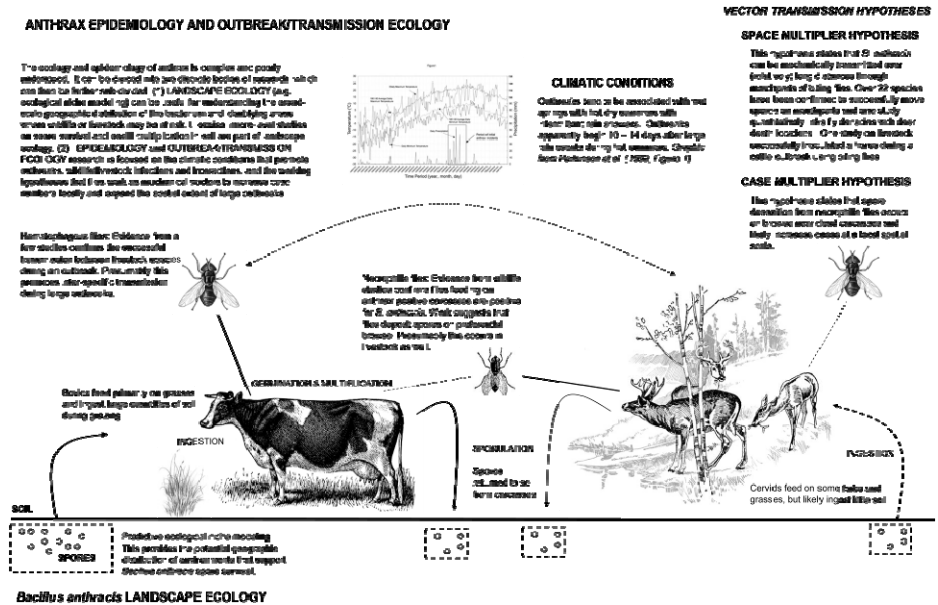

Figure 1. A working model of the landscape ecology of *Bacillus anthracis* and the hypotheses of transmission pathways. *Solid arrows* indicate established transmission pathways between bacilli and animals. *Dashed arrows* represent hypothesized or poorly understood transmission pathways. This chapter is focused primarily on *B. anthracis* landscape ecology so spores are outlined.

2.1. Epidemiology and Outbreak Ecology

First, the study of anthrax can be defined by the body of work on the epidemiology and ecology of outbreaks and the disease transmission cycle. This research is focused on a variety of climatic and local weather events that promote outbreaks and the ecology of affected species, which promotes outbreaks or larger epizootics. Although still lacking a full body of literature, a number of studies have documented the climatic conditions (and associated local weather patterns) that promote livestock outbreaks [8, 9] and wildlife outbreaks [3, 10–12]. In a synthesis on anthrax and wildlife, Hugh-Jones and De Vos [3] define the anthrax season as approximately late spring to early fall. There is a need to expand quantitative analyses of seasonality and outbreak periodicity, particularly outside of African wildlife populations. This area of research also is focused on the movement ecology of wildlife species in relation to outbreaks and anthrax seasonality [12].

Likewise, studies on the potential role of necrophilic flies to increase the number of browsing wildlife cases locally [13] would fit into this category of anthrax research. Hugh-Jones and De Vos [3] reviewed this phenomenon in African wildlife and suggest that a similar mode of transmission could be at work in the United States with deer. Blackburn [12] confirmed the presence of *B. anthracis* in necrophilic flies collected on and around disease-positive, dead white-tailed *Odocoileus virginianus* in west Texas and named the Hugh-Jones and De Vos [3] concept the "case multiplier hypothesis" of anthrax transmission. This states that necrophilic fly species likely increase cases in browsers by feeding on disease-positive carcasses and then depositing *B. anthracis* spores on nearby preferential browse; other animals will feed on the contaminated vegetation and contract the disease. Although data on this mode of transmission are quite limited from field investigations [12–14] given our understanding of deer feeding preferences in the United States, this is a plausible and testable hypothesis.

There is a second working hypothesis on mechanical transmission by hematophagous flies that also was reviewed by Hugh-Jones and De Vos [3] and was evident in the literature for much of the previous century [15–18]. In recent work, at least 21 different species from five genera of the Tabanidae family of flies have been confirmed under experimental conditions to transmit anthrax bacilli on their body parts [19]. This potential mode of transmission suggests that biting flies can pick up spores on their large mouth parts or legs during a blood meal from a bacterimic animal and then transmit them to other animals during subsequent blood meals. In an early study by Rao and Mohiyadeen [16], the researchers were able to isolate bacilli from the edema of bacterimic cattle and biting flies feeding on these animals. Given that hematophagous flies have greater flight strength and potential to travel farther distances, Hugh-Jones and De Vos [3] suggest that this mode of transmission may explain the wave-like pattern of large and fast-moving outbreaks, particularly those that expand beyond individual herds or fenced pastures. This also might explain interspecific infections during outbreaks in which species with different feeding ecologies are infected during the same outbreak. In North America, for example, infections in deer of the Cervidae family and cattle or bison in the Bovidae family have been documented in single outbreaks [12]. Although cervids may feed on some low-growing herbs and forbs, Texas deer primarily feed on browse during the summer period when anthrax outbreaks

are most likely [20]. Gates et al. [21] suggest that there might be a relationship between high biting-fly populations and large outbreaks in Canadian wood bison (*Bison bison athabascae*) but researchers have yet to quantify this phenomenon or isolate bacilli from biting flies during an outbreak.

Blackburn [12] provided limited evidence that there is a positive, quantitative relationship in biting fly densities and the spatial pattern of positive deer cases on a study ranch in west Texas. This study directly employed spatial statistics and geographic information to quantify spatial clusters of high fly counts during trapping periods in relation to individual deer movements (from VHF telemetry) and current and historical case locations of dead deer. Although this study was limited to spatial relationships and lacks the "smoking gun" of spore-positive flies, it does suggest that a "spatial multiplier hypothesis" is plausible and worth further investigation. This hypothesis suggests that interspecific transmission between wildlife species, or wildlife and livestock, during large outbreaks may be due to mechanical transmission by biting flies and that fly movement patterns may expand the spatial footprint of an outbreak. A third testable hypothesis during multispecies epizootics is that both the case multiplier and spatial multiplier hypotheses are working in tandem.

Studies on each of the hypotheses presented in this section can directly employ spatio-temporal analyses of climate data [9] and spatial analyses, such as spatial cluster statistics [12, 22], for fly–animal relationships.

2.2. Landscape Ecology of Bacillus anthracis

Each of the patterns and processes in the section above aim to identify or describe anthrax outbreaks in relation to transmission pathways and species interactions. However, for any of the interactions above to take place and promote the transmission and subsequent infection of anthrax bacilli, the bacilli have to be present. This means that, for naturally occurring outbreaks to occur, we first must understand the landscape-level patterns that promote spore survival and subsequent exposure to populations. I define this area of anthrax research as the landscape ecology of *B. anthracis*. To understand the ecology of outbreaks presented above, and given that most research on *B. anthracis* currently suggests that germination and multiplication occurs in the host while spore survival occurs in the soil [2], it is necessary to identify the geographic area where bacilli spores can thrive for long periods of time. Landscape ecology provides a useful perspective of scale for such analyses. Haines-Young et al. [23] provide an overview of landscape ecology and the role that GIS can play to test hypotheses within this theoretical framework. For this chapter, we can define landscape ecology as the study of relationships between the biological requirements of the bacilli in spore form and the ecological conditions that support spore survival and the geographic areas where these requirements are met.

This landscape perspective can be useful for understanding the broad-scale geographic distribution of the bacterium [4, 5] and identifying areas where wildlife or livestock may be at risk. Likewise, micro-level studies on spore survival and bacilli multiplication in soil are key parts of this approach [2, 7, 24].

Understanding the geographic areas and ecological characteristics that promote *B. anthracis* emergence, propagation, and subsequent exposure to populations at risk is critical to developing or improving disease surveillance and control programs. This is especially true of anthrax given the complexity of transmission and spore survival described in this chapter. Oftentimes, determining the complete spatial distribution of a disease agent or its host, reservoir, or vector requires the use of spatially explicit predictive models and any available data on the known occurrences of outbreak foci or host/reservoir or vector populations. Because of this, GIS and spatial analyses have become important tools in the study of anthrax epidemiology and the landscape ecology of *B. anthracis.*

The bulk of the analyses presented in this chapter focuses on the landscape ecology perspective and aims to define the potential geographic distribution of *B. anthracis* in the lower 48 states of the United States, under both current ecological conditions and a future climate-change scenario, and across the country of Mexico, where surveillance data on this disease are lacking.

3. Geographic Information Systems and Geographic Information Science

Today, GIS and GISc play an ever-increasing role in international research that focuses on disease distributions [25], ecology [4], and epidemiology [26]. The application of spatial data management and spatial analyses intuitively fits into research agendas that focus on a wide variety of topics, such as resource management [27], health and disease surveillance [28, 29], basic ecology [30], and socio-environmental [31] and socio-economic patterns [32]. As the application of GIS technology and the GISc paradigm grows, so, too, does the number of GIS techniques, computer applications and hardware, and GISc-trained personnel.

3.1. Geographic Information Systems

3.1.1. Geographic Information Systems for Disease Studies

Although GIS has been defined in great detail [33, 34] and has a growing body of literature, I will define it briefly in the context of disease surveillance and the modeling of disease ecology. GIS is a combination of computer hardware, computer software, and database technologies that allows for the storage, management/editing, visualization (mapping), and analysis of spatial data [35]. The primary component of a GIS is the ability to establish relationships within data sets and to analyze them spatially. Spatial here is defined as geographic relationships between data sets (such as disease outbreak locations and environmental parameters such as temperature). Disease studies readily lend themselves to GIS-based analyses because, simply stated, diseases are spatial in nature. Pathogen biology and transmission are linked to specific ecologies that promote the long-term survival and fitness of the pathogen, its reservoir or host populations or environment that sustain its population (e.g., soil foci for *B. anthracis* [2] or populations

of small mammals that circulate *Yersinia pestis* [36]), or vector species that transmit disease agents from host to host (e.g., tabanid flies that move *B. anthracis* spores [12], mosquitoes that vector malaria [25] or ticks that move *Nairovirus* species for Crimean-Congo hemorrhagic fever [37]). In these examples, we can evaluate geographic phenomena, such as habitats, that allow for the interaction of disease agents, vectors, and hosts. This is done through the integration of multiple disparate (and often idiosyncratic) data sets linked through relational databases, dynamic maps, and spatially explicit modeling techniques and spatial statistics. In GIS, the visualization component equates to digital map development or graphics that depict spatial relationships (such as histograms). The disparate data sets used to develop a spatial visualization (map) of habitats might include satellite-derived or GIS-interpolated precipitation maps, elevation maps, soil types, or forest cover (see Fig. 2, inset 1, for an example of environmental maps). Likewise, the data sets used to map a disease distribution might be known case locations (perhaps based on serology, laboratory diagnostics, or case definitions), known reservoir species' ranges, or agent-positive vector sampling sites.
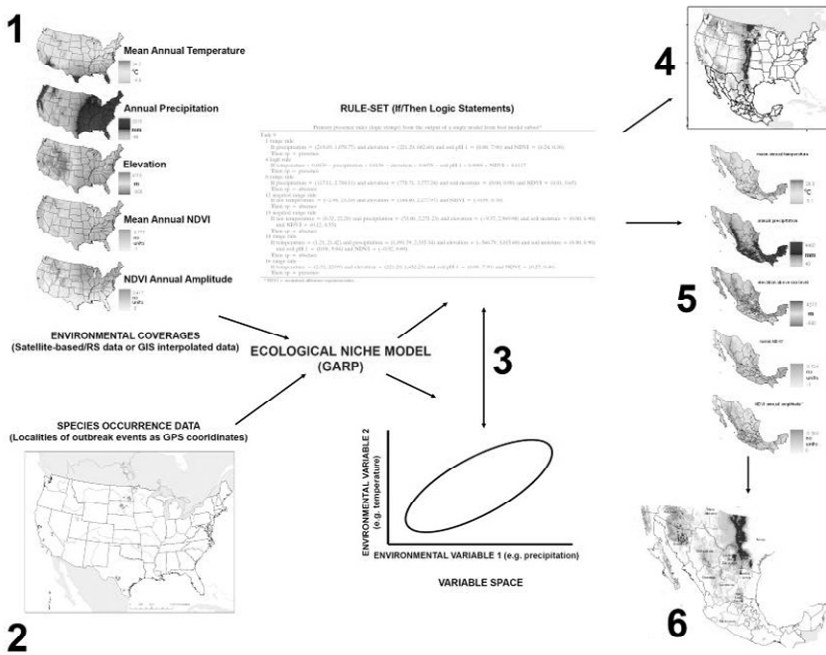

Figure 2. The six steps of the predictive ecological niche modeling [ENM] process for modeling *Bacillus anthracis* in the contiguous U.S. and Mexico: (1) environmental variables are processed in GIS/remote sensing and used in the ENM process to develop the Hutchinsonian N dimensional hypervolume; (2) Locality data from culture-positive *B. anthracis* are input into the ENM application; (3) the modeling system iteratively develops a rule-set of logic strings that describes the ecological space that supports the species; (4) the rule-set is applied to the landscape to develop a map of presence/absence; (5) the rule-set from the known locations can be applied to a environmental coverage set for an area lacking occurrence data; and (6) the projected rule-set is applied to the unknown landscape for a first prediction.

By evaluating each of these data sets in a multivariate paradigm, a single map can be produced to evaluate the intersections in geographic space where appropriate environmental conditions exist to promote the target species or disease-transmission cycle [4, 25, 38]. In short, GIS is a methodology and digital infrastructure for evaluating data in space. This is done through integrating databases with map visualizations. With today's technology, databases are limited only by intellectual imagination. As long as we invest effort and forethought into our database infrastructure, we will be able to expand and scale our GIS applications to capture data on multiple diseases, disease ecosystems, and study designs. This can be done across scales from the local or micro environment to the national or regional macro scale.

Employing GIS requires two infrastructures. The first, a physical infrastructure of computers, servers, network capabilities, and software, is necessary to employ the technology. In the ever-growing computer market, the availability of high-end computers at low prices has made access to high-powered computers tenable in developed and developing countries. Secondly (and more importantly), a data infrastructure of spatial data sets, disease data sets, and environmental data sets is required to construct GIS data layers and build multilayer maps that reflect meaningful epidemiological relationships.

Within the context of how GIS is employed in disease studies, it is important to distinguish how data sets are organized, particularly as this chapter is focused on more sophisticated applications of GIS for constructing predictive ENM scenarios for *B. anthracis*. Data within a GIS can be organized into three data formats. First, there are two basic GIS data types: vector and raster. The main difference between these two formats is how the geographic data values are stored within a database. Vector GIS defines geographic features into three major geometric objects: points (e.g., a disease case location or sampling site), lines (e.g., road networks, streams, railroad lines), and polygons (e.g., political boundaries, building footprints, water bodies; Fig. 2, inset 2, illustrates a map made entirely of vector features). Descriptive data for each point, line, or polygon are stored as attributes in an accompanying data table that links respective vector geometry with appropriate data values. A raster GIS uses a spatial grid with symmetrical cells to store the data. With raster data, attributes are assigned to each grid cell in the database and numerical values are used to represent various features (Fig. 2, insets 1 and 5, illustrate raster layers). For example, a 1-km^2 raster file of temperature data would represent equally sized square grid cells each with a numerical value representing the temperature for that 1×1-km pixel. In this way, a variety of raster files, such as soil pH, vegetation, or elevation, all could be stored in a GIS and represent these values with equal pixel sizes. Because each cell occupies the same space, multiple spatial layers can be compared within each cell, making raster GIS useful for layering multiple environmental data sets and finding associations between them. Vector GIS is well-suited for identifying clusters [22] and diffusion patterns [39] across a landscape or study area.

The third data format managed within the GIS is the aspatial database table (those tables without spatial data assigned). For example, data developed by a collaborating laboratory, such as molecular genetic data about the disease agent, can be managed and stored within the GIS for later data joins or assignment to spatial locations. For a detailed review of GIS methods for developing data in all three of these formats, see Curtis et al. [31].

3.1.2. Visualizing Data within a Geographic Information System – Maps

The primary visualization outputs from GIS analyses are maps (Fig. 2). Simply stated, maps are easy to read. Although GIS is more dynamic than static map production, it is important to realize the value of maps for improving our understanding of complex systems such as disease-transmission cycles [37] or the relationship between human settlement patterns and infectious diseases [28, 40]. We need to remember that maps are only as good as the data we invest into making them, but maps provide easy tools for visualizing disease distributions. Imagine trying to describe the distribution of Crimean-Congo hemorrhagic fever across a country as large as Kazakhstan (the ninth largest country by land mass) without a map to keep track of all the areas you are trying to define. That would be quite difficult compared with showing a single map of the disease distribution. Now imagine trying to relate that message to health managers who might control eradication strategies and trying to sort out which areas should be prioritized first. Managers are likely to communicate and make decisions more clearly through maps. It is important that these maps have the best data available, including precise spatial locality data and accurate data attributes. That is where GIS and data management become important.

Although maps are an important output from GIS analyses, it is important to realize that map development is only one of several tools in the GIS toolbox for improving our understanding of disease distribution, ecology, and epidemiology. GIS provides a set of tools for developing data sets for advanced spatial analyses and statistics to test explicit hypotheses (often completed outside of the primary GIS software environment, such as in a statistics program or programming environment).

3.2. Geographic Information Science and Spatial Data Analysis

GISc can be summarized as the paradigm for formulating geospatial hypotheses and proper employment of the tools of GIS to solve spatially driven research problems systematically. Goodchild [41] first introduced the term, and the field has grown since. Mark [42] provides a detailed review of the evolution of naming the term, Goodchild [43] discusses the terminology further, and Goodchild [44] reviews the roles of GISc as split between the study of GIS to advance its technology and the use of GIS technology to advance scientific fields of study within a spatial framework. In the increasingly fast-growing world of computer technology and push-button analytical tools, GISc provides a systems approach to compartmentalizing, evaluating, and organizing these tools. GISc could be considered the academic discipline for organizing GIS and promoting a proper educational structure to prepare researchers for sustaining GIS and for employing it within a hypothesis-testing framework focused on addressing scientific research. In addition to training basic GIS techniques and data manipulation, GISc can integrate formal training in spatial analysis (such as point-pattern analysis, clustering algorithms, and probability-based testing), geographic processes, biogeography, and spatial ecology to examine spatial phenomena.

In the context of spatial epidemiology, and this chapter, GISc should be defined as that body of work that advances our knowledge of disease ecology through an improved understanding of the spatial patterns and processes that promote disease

propagation and transmission. With that in mind, we can consider a three-step process for (1) evaluating the distribution of a disease visually, (2) exploring patterns of disease (such as outbreaks or vector population dynamics) using spatial statistics, and (3) modeling geographic distributions of disease with spatially explicit predictive models. Anselin [1] provides a generic overview of this process and a review of the classical literature from the field. Here I attempt to integrate those steps into the study of *B. anthracis* ecology and anthrax to establish a framework for advancing our knowledge of this pathogen and its natural ecology and epidemiology.

3.2.1. A Framework for Integrating Geographic Information Systems and Spatial Analysis into Disease Ecology

As an expansion of section 2.1.2 above, the development of maps that relate the spatial position of anthrax outbreaks to the landscape illustrates Anselin's concept of exploratory spatial data analysis (ESDA) [1]. Briefly, data visualization can be considered a first step in evaluating a data set (or data sets) to determine the spatial nature of the data and evaluate the spatial patterns. This often leads to the development of more than maps, including histrograms, graphs, box plots, or animations, that allow the researcher to explore the data and identify possible outliers (spatial or attribute data) or visual clusters (aggregations of data observations in close spatial proximity). Figure 2, inset 2, illustrates a simple point map of the distribution of anthrax outbreaks in the contiguous United States from 1957 to 2005. Although there is no statistical analysis applied to the data in Fig. 2, inset 2, the map alone is informative for determining the location of outbreaks across the lower 48 states over a relatively long time period. Although we will discuss more sophisticated analytical tools in later sections of this chapter, this first step does provide useful information. For example, it is clear from this map that outbreaks are concentrated in two areas (southwestern Texas and the Dakotas), with smaller numbers of outbreaks in the western-most states and eastern Oklahoma. Knowing this, one now can think about how this pattern may have developed. How did *B. anthracis* get from the southern states to the northern states? How did it move from east to west? Is there some environmental gradient in the central portion of the land-scape that prohibits outbreaks in the eastern states during this period? These questions are easier to pose with the map in hand, illustrating an important part of the ESDA process: the development of hypotheses to test on this spatial distribution.

Anselin's [1] second step in the ESDA process is the use of measures of spatial autocorrelation to identify statistically significant patterns within data sets. Here these patterns refer to disease outbreaks. There is a growing body of spatial statistics that employs local measures of spatial autocorrelation to determine whether statistically significant spatial patterns exist within a data distribution. Geographic areas in which more outbreaks occur than would be expected by random chance are defined as hot-spots; those areas with significantly fewer outbreaks than would be expected randomly are defined as cold spots [45]. These statistics are performed through iterative algorithms that evaluate the relationship between outbreak occurrences (at some aggregated level, such as a polygon grid surface or a political boundary) and neighboring occurrences. These neighbors can be defined through a neighbor relationship, in which some boundary of these polygons is shared [1] or based on proximity within a distance

threshold [22]. Two known studies have employed spatial autocorrelation in the study of anthrax epidemiology.

3.2.2. Employing Local Indicators of Spatial Autocorrelation to Anthrax Data

Blackburn [12] employed the Getis statistic to determine whether significant local spatial clusters occurred for any biting fly-trapping events during the 2005 anthrax season on a study ranch in west Texas. The goal of the study was to determine whether there were areas within the outbreak zone that had significantly high tabanid fly catches. 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 [46]. $G_i^*(d)$ is useful for identifying individual members of local cluster events [22]. The Blackburn [12] study identified both cold spots of low fly-catch rates and hotspots of high fly-catch rates. This then was coupled with the spatial distribution of anthrax-positive deer carcasses for that anthrax season plus historical cases from 2001 to 2005 (Fig. 3). Although not confirming any direct relationship between outbreaks and fly densities, the spatial association between clusters of biting flies and the location of carcasses certainly warrants further investigation into possible
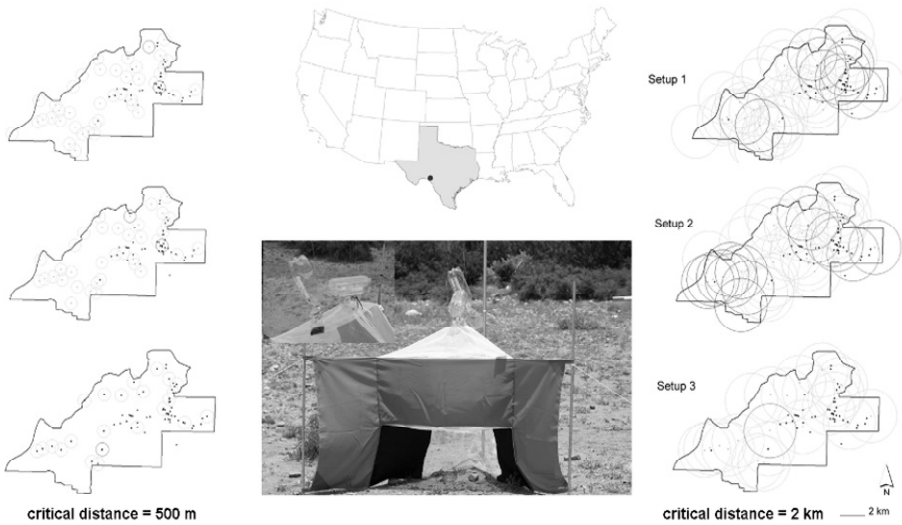

Figure 3. Spatial clusters of tabanid flies on a study ranch during the 2005 Texas anthrax season. Setups equal three time periods that divide the sampling season. Critical distances represent the spatial scale at which any given trapping location was part of a spatial cluster. *Gray circles* represent sampling areas that were not significant, *red circles* indicate spatial hotspots of significantly high catch rates for tabanid flies, and *blue circles* indicate significantly lower than expected catch rates. *Black dots* across the eastern portion of the study represent carcass locations of anthrax-positive deer. Notice the spatial overlap between fly hotspots and anthrax locations. Photograph insert: a Nzi fly trap setup used to collect flies. Adapted from Blackburn [12].

causal relationships. Even in the absence of a direct diagnostic link between flies and disease transmission, there is direct overlap between high catch rates of biting flies and diseased carcasses. There may be an indirect effect between immune suppression in deer and seasonal inundation with flies. This alone may serve to increase the number of susceptible hosts in a given outbreak season.

A second study employed the same Getis statistic to the spatial distribution of anthrax outbreaks in Kazakhstan for four separate decades from 1960 to 1990 [47]. In the second study, only hotspots, or significantly high outbreak counts, were evaluated for statistical significance. In this latter study, cattle and sheep outbreaks were aggregated to fixed-width grid cells for four separate decades and the spatial statistic was calculated for each decade separately. This provided a time series of four maps to evaluate areas of long-term outbreak persistence across the country and the time period. Figure 4 illustrates the decadal hotspots of cattle outbreaks in Kazakhstan. Both of these studies illustrate the use of ESDA for evaluating disease ecology within a hypothesis-testing framework. However, these measures of spatial autocorrelation are inferential and do not provide any causality of high fly-catch rates or counts of cattle outbreaks but rather where these events cluster in space.
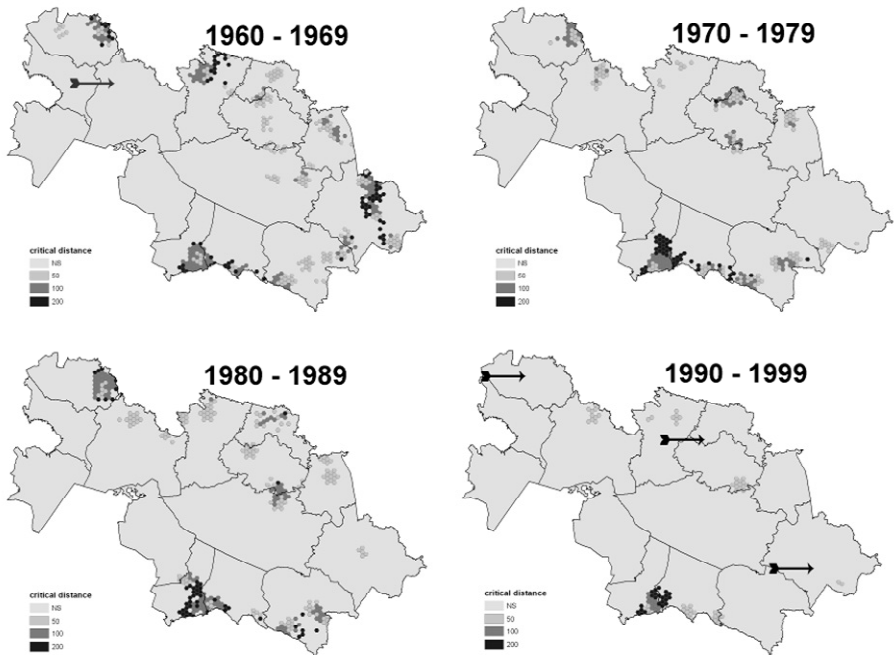

Figure 4. Decadal spatio-temporal clusters of anthrax outbreaks in cattle from Kazakhstan during the period 1960–1999. *Grayscale ramp* indicates the spatial scale of the cluster as determined by the critical distance for that g-score. *Black arrows* indicate areas in which significant clusters of outbreaks disappeared during the study period; the *red arrow* indicates an area in which a cluster developed in the latter decades of the study period. Cluster data adapted from Sagiyev et al. [47].

It is also important to note that both of these examples of cluster analysis focus on the spatial patterns of mechanical vectors or individual outbreaks. Both of these studies represent the use of GIS and spatial analysis for evaluating components of disease transmission and ecology and do not employ the landscape approach or evaluate direct spatial relationships between the environment and *B. anthracis.*

The third approach presented in Anselin [1] is to employ a predictive modeling approach to relate parameters to spatial patterns. As mentioned in previous sections, spatial patterns identified during a naturally occurring anthrax outbreak must be occurring in geographic regions in which bacilli in spore form can be maintained in the environment. To identify these areas, it is necessary to merge Anselin's third concept with a modeling approach that can address the landscape ecology of *B. anthracis* and relate culture-positive outbreak locations to an ecological signature that defines the range limits of the disease agent. ENM provides an ideal tool for predicting species' distributions. In the remainder of the chapter, I will define ENM and present examples of its application to *B. anthracis* ecology.

4. Evaluating Landscape Ecology with Predictive Ecological Niche Modeling

4.1. A Primer on Ecological Niche Theory

To understand how ENM can be useful for predicting the geographic space in which anthrax outbreaks may occur naturally, it is first important to define the concept of the ecological niche and provide a conceptual framework for the modeling process.

Hutchinson [48] provides an excellent overview of ecological niche theory and provides a reference to the original introduction of the term "niche" in the ecological literature. Hutchinson argues that Johnson [49] was the first to use the term, though a more rigorous and technical definition was provided by Grinnell [50], which often serves as a key reference in many niche-modeling articles. Grinnell defines the ecological niche as a limited range of ecological variables that could maintain a population without immigration. As part of this definition, Grinnell [50] states that no two species could occupy a single niche. This definition later was expanded into a quantifiable ecological space by Hutchinson in two articles [51, 52].

While examining relationships between phytoplankton and chemical properties within a lake system, Hutchinson [51, page 20, footnote 5] proposed that the ecological niche can be "…defined as the sum of all environmental factors acting on the organism; the niche thus defined is a region of an n-dimensional hyper-space…" Hutchinson [52] then expanded this definition to an n-dimensional hyper-volume of parameters that could be ordered linearly to define the fundamental niche or the potential ecological space that could maintain a species. Hutchinson [52], and later MacArthur [53], also define the realized niche as that limited portion of ecological space that is actually used by a species owing to biological interactions (such as competition, dispersal limits, and historical events like local extirpation). Under the Hutchinsonian definition, there is acknowledgement that biological interactions play a role in limiting the available ecological space and actual geographic space in which a species can survive. Morrison and Hall [54] argue that the Grinellian niche definition can be considered a dimension

of the Hutchinsonian niche. For a further review of ecological niche theory, see Morrison and Hall [54] and Chase and Leibold [55].

In the context of this chapter, the ecological niche is used as a construct to evaluate spatially explicit modeling approaches to define the potential geographic distribution of a species. This of course is not limited to disease studies and, in fact, has been tested rigorously across a wide range of taxa (see Blackburn [12] for a partial review of taxa modeled). When attempting to understand geographic patterns of disease distribution (or any species for that matter), it is important to define clearly the ecological space being evaluated [56] and the ecological theory that is being tested [54].

4.2. Predictive Ecological Niche Modeling

Predictive ENM provides an ideal tool for determining disease agent or vector distributions using GIS data and remotely sensed data to model the environments that might support disease outbreaks [56, 57]. Currently, several modeling approaches are available to estimate species' distributions (e.g., Stockwell and Peters [58] – genetic algorithm for rule-set prediction [GARP]; Rogers [25, 59] – Discriminant Function; Phillips et al. [60] – MaxEnt). In all of these models, the goal is to identify the geographic region and ecological environments that can sustain the disease agent or promote interaction/infection through host–vector interaction. Identifying these regions is done through computationally intensive iterative algorithms that pattern match disease case or vector species' localities with environmental data layers. The goal is to identify nonrandom relationships between case locations and the environment, either through pattern matching with post hoc statistical evaluation of those patterns or through direct statistical relationships. Once these relationships are identified, the user can data mine the modeling outputs for biological information. Several recent articles have employed ENM successfully to describe the distribution of diseases or as predictions of host/ reservoir populations [38], vector populations [61], or disease-agent presence from outbreak data [4, 62].

Although the software application is described in some detail below, it is useful to provide a conceptual overview of the modeling process employed. Figure 2 provides a diagram of the modeling system and the input data sets required to produce potential geographic distributions for *B. anthracis.* I have divided the modeling process into six parts. First, an environmental coverage set of ecological variables is input into the modeling software (Fig. 2, inset 1). Next, input data of disease occurrence are input into the modeling software (Fig. 2, inset 2). In step three, the rule set (see below) that describes the relationships between *B. anthracis* presence and absence is developed. Figure 2, inset 3, illustrates this with a set of if/then logic strings (the actual rule set from the modeling software) and with a simplified visualization of ecological space with two variables. In the examples used in this figure, the n-dimensional hyper-volume would be constructed from five ecological variables. Figure 2, inset 4, illustrates the next step in the modeling process, in which the rule set is applied to the landscape to produce a potential geographic distribution for the species. One advantage of the GARP modeling system is the opportunity to project the rule set onto landscapes that lack occurrence data. Step five illustrates this, with the rule set from the U.S. model being applied to the landscape (and variable) space of Mexico. Notice it is an identical data

set to the U.S. coverage set in Fig. 2, inset 1, but for the geographic space of Mexico. In step six, the rule-set relationships for Mexico are applied to the landscape to create a potential geographic distribution for *B. anthracis*. Steps five and six could be replaced here with data from future climate scenarios to illustrate the concept for modeling climate change.

4.3. The Genetic Algorithm for Rule-Set Prediction

In this chapter, I present three ENM scenarios to describe the potential geographic distribution of *B. anthracis* on the landscape of the lower 48 contiguous United States and Mexico under current climatic conditions and provide a first estimate of the geographic potential for the species in 2050. For all three modeling scenarios, I employ the GARP modeling algorithm. The GARP modeling system has been explained in detail elsewhere [58], as have examples of the application of GARP to disease systems [4, 40].

In brief, this study employed the DesktopGARP version 1.1.6 [DG] application to develop all GARP models (available from www.lifemapper.org/desktopgarp). GARP is a presence-only modeling technique that determines nonrandom associations between point localities (anthrax outbreak locations) and environmental parameters (environmental "coverages") [58, 63]. Results are in the form of presence/absence predictions based on a set of heterogeneous rules.

GARP modeling is stochastic in nature, owing in part to both the genetic algorithm for building models and the random partitioning of input-locality data. In other words, GARP is a random walk through variable space. Because of this, GARP can generate multiple solutions across multiple model runs. To evaluate this potential intermodel variation, it is critical to develop multiple models. Optimal models are those that compromise between omission (exclusion of known locations from the model) and commission (inclusion of areas with no known cases) [64]. DG employs a "best subset" procedure to optimize model outputs by selecting models with user-defined omission and commission thresholds.

The modeling approach is a two-step process wherein rules are generated to describe presence and absence in variable space. This rule set then is applied to the landscape pixel by pixel to create a spatially explicit prediction of presence and absence.

GARP outputs are rasterized coverages of the study area representing presence and absence pixels that can be manipulated in a GIS. These individual models can be summed to identify geographic areas in which none, some, or all of the models predict presence or absence [65]. The greater the number of models that agree, the more certainty there is in the prediction classification [62]. Likewise, similarity across models indicates stability in the modeling system. For an in-depth review of employing DG, see McNyset [66].

4.3.1. Input Data: Occurrences of *Bacillus anthracis*

A GIS database of specific anthrax outbreak localities within the 48 contiguous United States was developed from a variety of data sources for the period 2000 to 2005, with the exception of a 1957 outbreak report that could be mapped at the point level for

Oklahoma and a 1968 outbreak for east central California. To be included in the final modeling data set, laboratory diagnostics had to confirm the presence of bacilli. Suspect cases were removed from this modeling process. Point data representing confirmed wildlife and livestock outbreaks 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, California; Fig. 2, inset 2). Table 1 summarizes the sample sizes and methods of data collection for each of the states used in this analysis. These occurrence data were used in all three modeling scenarios included in this chapter.

TABLE 1. Data Sources for Anthrax-Outbreak Localities (1957–2005) Used to Develop Ecological Niche Models of *Bacillus anthracis* in the Contiguous United States and to Project the Distribution in Mexico

Data	Data source
Outbreak locality data	
Minnesota	Minnesota Board of Animal Health
North Dakota	North Dakota State University Veterinary Diagnostic Laboratory
South Dakota	South Dakota State University Agriculture Extension and GIS Center for Excellence
Oklahoma	Oklahoma Department of Agriculture
Texas	U.S. Centers for Disease Control and Prevention
	Louisiana State University Field Investigations
Nevada	U.S. Department of Agriculture Animal and Plant Health Inspection Service
California	U.S. Centers for Disease Control and Prevention
	California Department of Food and Agriculture
	Dr. Frank Paterson

4.3.2. Input Data: Environmental Coverages

4.3.2.1. Current-day conditions (1950–2000)

For the two current-day modeling scenarios (*B. anthracis* in the United States and the distribution in Mexico), a set of environmental coverages was constructed from publicly available climatic and biophysical parameters. Nineteen variables were downloaded from the WorldClim data set representative of various temperature and precipitation measurements (www.worldclim.org) [67]. Thirteen additional environmental variables, including temperature and vegetation measures (e.g., normalized difference vegetation index [NDVI]), were provided by the TALA research group at Oxford University [68]. All environmental coverages were resampled to 0.10 degree^2 (~8 × 8 km) and clipped to the boundary of the 48 contiguous United States and again separately for Mexico. All data sets were prepared using ERDAS Imagine version 8.7 (Leica GeoSystems, St. Gallen, Switzerland), ArcGIS 9.2, and ArcView 3.2a (ESRI, Redlands, CA). To select the variables used in any modeling scenario, a rigorous culling methodology was applied. This is described in detail in Blackburn et al. [4]. Variables in the coverage sets are presented in Table 2.

TABLE 2. Environmental Coverages Used to Develop Predictive Ecological Niche Models of *Bacillus anthracis* for the Contiguous United States and to Project the Potential Geographic Distribution of the Pathogen in Mexico and the United States in the Year 2050

Data	Data Sources
Current-day coverages (United States and Mexico study)	
Mean annual temperature (°C)	WorldClim data set; Hijmans et al. [67]
Annual precipitation (mm)	WorldClim data set; Hijmans et al. [67]
Elevation (m above mean sea level)	TALA Research Group, Hay et al. [68]
TFP mean NDVI	TALA Research Group, Hay et al. [68]
TFP NDVI annual amplitude	TALA Research Group, Hay et al. [68]
Current-day/future-scenario coverages (HADCM3 B2 scenario)	
Annual temperature (°C)	IPCC (2001)
Maximum annual temperature	IPCC (2001)
Minimum annual temperature	IPCC (2001)
Annual precipitation (mm)	IPCC (2001)
Annual solar radiation	IPCC (2001)
Annual wind	IPCC (2001)
Elevation (m above sea level)	TALA Research Group, Hay et al. [68]
Soil pH	Blackburn et al. [4]; STATSGO database
Soil moisture	Blackburn et al. [4]; STATSGO database

TFP, temporal Fourier processed [68]; NDVI, normalized difference vegetation index.

4.3.2.2. Future data set (2050)

A third modeling scenario was constructed for this chapter to illustrate the potential changes in *B. anthracis* distribution on the landscape in 2050 under a climate-change scenario. For this modeling experiment, I used the HADCM3 B2 climate data set for 2050 [69]. This scenario is constructed from a 30-year average based around 2055, excluding effects of potential increased climate variability [70]. The B2 scenario is a conservative estimate of overall temperature increase, ranging from 2.1–3.9°C, defined as "a world in which emphasis is on local solutions to economic, social, and environmental sustainability. It is…a heterogeneous world with less rapid and more diverse technological change but a strong emphasis on community initiative and social innovation to find local, rather than global solutions" [69]. For this modeling scenario, a selection of temperature, precipitation, wind, and solar radiation variables were selected from the B2 data set and combined with two continuous soil parameters from the STATSGO data set (soil moisture, soil pH; www.ncgc.nrcs.usda.gov/products/ datasets/statsgo). These were used to incorporate measures of known ecological factors that promote spore survival. Elevation also was included in this coverage set. It was assumed for this study that these soil values would not change significantly in the 50-year period. Both soil variables were rasterized for inclusion in the ENM. All data for this coverage set were resampled to 0.5 degrees (the resolution of the HADCM3 data) and clipped to the contiguous United States.

4.3.3. Modeling Parameters

For modeling the current distribution scenarios, 177 spatially unique anthrax-outbreak locations were available for model building (Table 1). Before initiating DG, a randomly selected, independent hold-out sample of ~25% (n = 47) of the original data was withheld for later calculation of accuracy metrics. The remaining ~75% of the data (n = 130) was used for model building. A training/testing partition (50%/50%, respectively) internal to DG was used for model building. To maximize DG performance, 1,000 models were developed and the best subset procedure was employed to select the 20 best models under a 10% hard omission threshold and a 50% commission threshold for a final 10-model best subset. The final 10 models were summated within the GIS to visualize the geographic areas of presence/absence predicted across the best subsets.

The same parameters were selected for the future modeling scenario. However, given the larger pixel size of the ecological coverages, only 57 sampling sites were used to build the model. A post hoc sample of 27 points was used to calculate accuracy metrics for the future scenario.

4.3.4. Model Accuracy Metrics

An area under the curve (AUC) in a receiver operating characteristic (ROC) analysis was used to evaluate the predictive performance of the 10-best model subset using measures of specificity (absence of commission error) and sensitivity (absence of omission error) following other GARP studies [66, 71]. 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 the best subset to determine whether models are predicting better than random [72, 73]. Likewise, AUCs, as employed here, are based on all pixels of presence and all pixels of absence. The AUC of a given model set is compared with that 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 score, the better the model is predicting presence/absence. Models predicting no better than random will have an AUC approaching 0.5 [74]. The ROC was derived from the 25% independent test data points withheld from the original GARP model-building data sets [66].

Two measures of omission were calculated from the 10-best model subset and the independent test data [66]. First, total omission was calculated as the total number of independent test points predicted absent by the summated grid of all ten best models. Second, an average omission was calculated as the average omission across each of the ten best models. Omission indices are useful for evaluating the success of GARP at predicting known localities not included in model building. Two commission indices also were 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. Second, an average commission was calculated as the average of 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 ten models in the best subset. Little difference between these two measures indicates little variation in the rule sets across the models, whereas a large difference indicates high variation across the models.

5. Ecological Niche Modeling Modeling Scenarios of *Bacillus anthracis* in North America

5.1. Modeling Scenario 1: Modeling the Current Distribution of Bacillus anthracis *in the United States*

Blackburn et al. [4] provided the first ENM-based geographic predictions for the potential distribution of *B. anthracis* for the lower 48 states. That modeling scenario was based on a six-variable niche definition and included two soil parameters (soil moisture content and soil pH). These parameters are not available for Mexico, so to project the geographic distribution of an unknown data set, it is necessary to construct models for both countries with the same ecological parameters. This required the development of a modified coverage set that included variables available for both countries.

Results from the first modeling effort for *B. anthracis* [4] indicated that mean NDVI was the most limiting variable in the rule set. For this chapter, I provide a five-variable coverage set to define the ecological niche for *B. anthracis*. Soil parameters were removed from the coverage set and replaced with an additional NDVI variable – annual amplitude [68].

5.1.1. Model Results

The U.S. model predicted the known distribution of outbreaks with high predictive accuracy (Fig. 5), with 95.6% of the independent test points predicted correctly and a statistically significant AUC score of 0.823 (Table 3), both indicating that the ecological
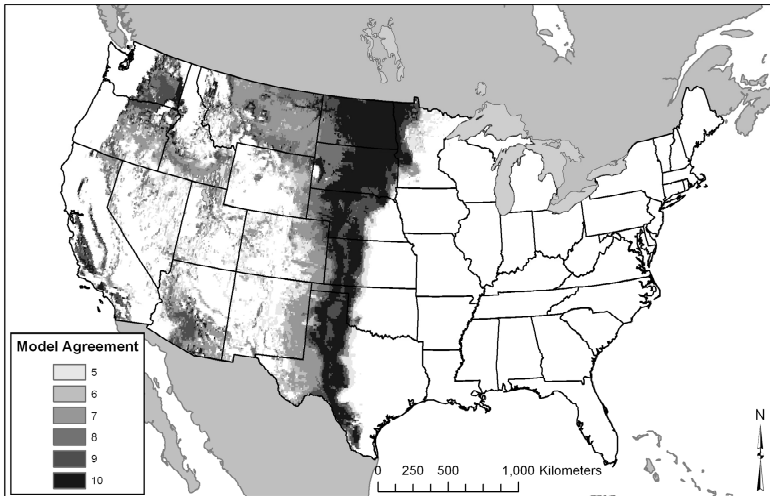

Figure 5. The potential geographic distribution of *Bacillus anthracis* for the contiguous United States based on a five-variable ecological niche modeling experiment using the GARP modeling system. The grayscale ramp indicates model agreement among the ten best models in the subset. Minimum threshold for the acceptable model agreement was set at five models.

niche models are accurate [66]. The geographic distribution was similar to that of Blackburn et al. [4], with a south-to-north corridor from southwest Texas to north-western Minnesota, then westward across the Dakotas, through much of Montana, and through the Snake River drainage. A second east-to-west corridor, though patchier, is present from west Texas through southern Arizona, with a disjunct area predicted in the Californian Central Valley. The results of this geographic prediction based on a five-variable niche definition are consistent with that of Blackburn et al. [4].

TABLE 3. Accuracy Metrics for the U.S. GARP Models Based on Current Climatic Conditions

Metric	Model specifications
Mexico scenario	
N to build models	130*
N to test models (independent)	47
Total omission	4.40%
AUC	0.832[†‡]
HADCM3 B2 scenario (present day)	
N to build models	57*
N to test models (independent)	27
Total omission	0.0%
AUC	0.846[§‖]

AUC, area under the curve.
*N was divided into 50% training/50% testing at each model iteration.
[†]$z = 10.31$ ($p < 0.01$).
[‡]$SE = 0.0367$.
[§]$z = 8.55$ ($p < 0.01$).
[‖]$SE = 0.0471$.

For this study, models were developed with point occurrences and coverages that represent the contiguous United States in the current time period (~1950–2000). Accuracy metrics are used to test these current models with post hoc validation points of known outbreaks in the U.S. to evaluate the quality of the projections onto the Mexican and future U.S. climate scenarios.

5.1.2. Evaluating Scenario 1 Relative to White-Tailed Deer, a Primary Wildlife Host

Following the Grinnellian definition of the niche, no two species can occupy the same niche. Because of this, it is important to conceptualize how the niche for the disease agent is defined when occurrence data were dependent on either livestock or wildlife outbreaks. In one such disease study, Peterson et al. [38] evaluated the potential geographic distribution of sylvatic *Trypanosoma cruzi*, the parasite that causes Chagas disease, across Mexico. In that study, Peterson et al. [38] suggests that the potential geographic distribution was likely at the intersection of the reservoir species (Neotoma wood rats) and the Triatomine insects that vector the parasite. To develop models, Peterson et al. constructed individual niche modeling experiments for each rat species and each insect species and then overlaid them in a GIS to identify potential areas of overlap where reservoirs and vectors might interact. Although this process was laborious, it adheres to the Grinnellian niche definition.

In the study of *B. anthracis*, where a specific vector is not required to transmit the disease (flies most likely serve as secondary vectors once an outbreak has started), it is important to determine whether or not the models reflect the potential distribution of bacterium or the host species. To test this hypothesis, I compared the geographic area predicted by the *B. anthracis* model to the known distribution of the white-tailed deer, *O. virginianus*. The published range distribution for deer was downloaded from the NatureServe Web site (www.natureserve.org) [75] and clipped to the extent of the lower 48 states to match the extent of the modeling experiment. I then converted the range limit for deer to a raster file and recoded those pixels representing the deer range to a value of 1. I then set a threshold limit of five or more models for the *B. anthracis* best subset and recoded presence to a value of 1 and used map algebra to add the two raster files together. The overlap of the two species is presented in Fig. 6. The deer range accounted for 81.9% of the total landscape of the lower 48 states. In contrast, the area where *B. anthracis* and white-tailed deer overlap accounts for only 35.9% of the total deer range. This suggests that the GARP modeling process is not biasing the *B. anthracis* distribution to the larger extent of its host species. Likewise, a comparison of Figs. 5 and 6 shows that, had only the deer distribution been used as a proxy for *B. anthracis*, it would have excluded the areas successfully predicted (and validated) across Nevada and California, suggesting that the deer distribution was not limiting the *B. anthracis* model. This suggests that the modeled geographic potential of
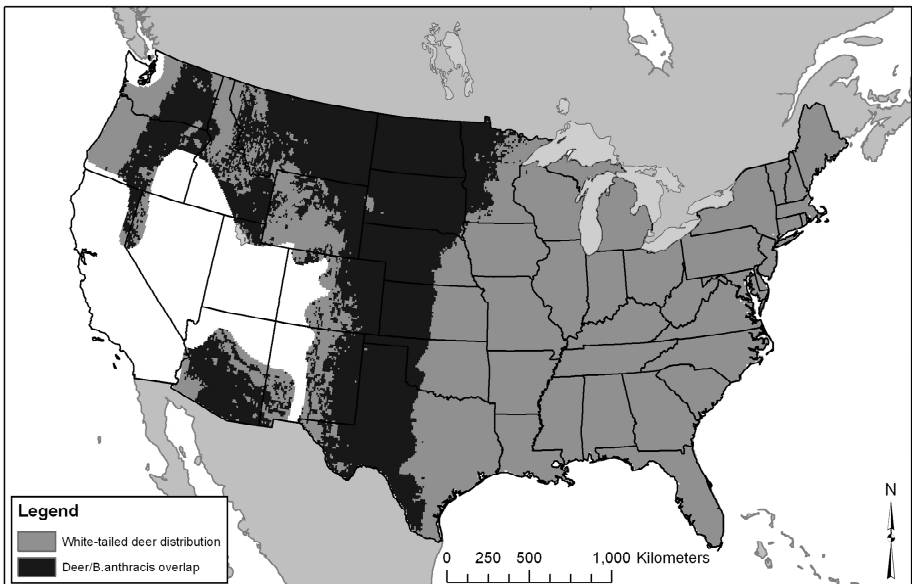

Figure 6. The known distribution of white-tailed deer (*Odocoileus virginianus*) for the contiguous United States is shown in *gray*. In total, this represents 81.9% of the total U.S. landscape. The *black* shading indicates the portion of white-tailed deer distribution that overlaps with the potential distribution of *Bacillus anthracis* (~35.9% of the total deer landscape), illustrating the drastically reduced portion of the landscape where the pathogen and host species might naturally overlap.

B. anthracis is not simply capturing the distribution of the host species and more likely illustrates environments that support spore survival. This also supports the suggestion by Blackburn et al. [4] that these *B. anthracis* models likely indicate surveillance priorities for deer that are meaningful (i.e., the area of overlap in Fig. 6).

5.2. Modeling Scenario 2: Projecting the Distribution of B. anthracis to Unknown Landscapes

Although the specific timing and location of the introduction of anthrax into the United States is still not fully understood [76], recent genetic data suggest a relationship between the North American sublineage and the dominant European subgroup, which supports a European introduction [77]. There are data on the distribution of anthrax in Canada [9, 21] and the contiguous United States [4]; however, comparable data on anthrax outbreaks are especially lacking for Mexico despite historical reports of anthrax as far back as 1923 [78]. Although Mexico participates in country-level livestock reporting to the Organisation Mondiale de la Santé Animale on an annual basis, the specific geographic distribution of *B. anthracis* throughout the country remains unknown, under-reported, and poorly understood [79]. With the exception of limited reports from the state of Zacatecas in 1981 and 1983 [80], a report from ProMED Mail [81], and two municipalities in the state of Nuevo Leon, even aggregated data at regional levels are lacking in Mexico's reporting efforts [82]. The border area between Mexico and the United States is already known to be susceptible to anthrax, with the disease being endemic in western and central Texas [3] and predicted to occur along the Arizona and California borders (see Fig. 5). In Texas, the greatest numbers of cases are associated with wild populations of white-tailed deer, *O. virginianus* [12]. These deer have the potential to move freely throughout the cross-border Tamaulipan/Mexquital ecoregion. Given the shared U.S./Mexico border and the high likelihood of international cross-border transmission, it is critical to determine the distribution of anthrax in Mexico.

To develop model results for Mexico, I modeled the distribution of *B. anthracis* for the contiguous United States using Modeling Scenario 1 above. I elected to use the GARP modeling system specifically because it allows the user to develop a model for an area with known occurrence data and then project those model rule sets onto the geography and environmental layers of another region (see Fig. 2). This is important because it prevents Mexican localities from being included explicitly as pseudo-absences, as would be the case if both countries were modeled in a single experiment using only U.S. outbreak data.

5.2.1. Model Results

The projected distribution of *B. anthracis* in Mexico (Fig. 7B) suggests that the disease is present along the U.S. border near central and western Texas, central Arizona, and south central California. In each of these locations, the distribution in Mexico is predicted to be a southward continuation of predicted areas in the United States Beyond the border regions, the prediction reaches south into central Mexico as far as the state of Puebla. Although limited data were available to validate the model results for Mexico, the model accurately predicted the area surrounding a farm with laboratory-confirmed

livestock cases in the year 2000 (Hugh-Jones, unpublished data), the region of Nuevo Leon described in Siefert et al. [82], and much of the state of Zacatecas [80]. In this latter account, no specific data are provided on the region of the state that reported the outbreaks. The only published report not predicted by five or more models in the best subset is the south central state of Michoacan reported on ProMED-mail [81]. As with the reports from Zacatecas, there is limited information available on the specific region within the state where the outbreaks occurred.
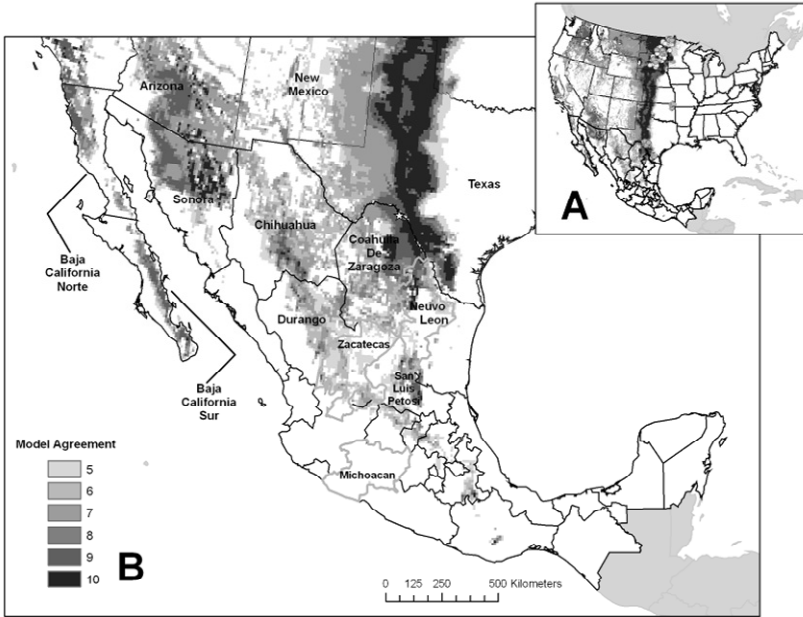
Figure 7. **A.** GARP prediction for the distribution of *Bacillus anthracis* in the U.S. based on outbreak data from 1957 to 2005. *Open circles* represent training data used for model building; *gray circles* represent the independent post hoc data for calculating accuracy metrics. Includes the projection onto Mexico. **B.** Close-up of the predicted distribution of *B. antrhacis* in Mexico. Star indicates the location of a laboratory-confirmed anthrax outbreak in 2000. *Gray-outlined* states indicate limited data from the literature that identify areas with known outbreak histories. *Grayscale ramp* indicates model agreement from the ten-best model subset. A five-model thres-hold is used to visualize predicted presence to balance between overfitting and overpredicting of the model set. This five-model threshold for mapping limits the between-model variability resulting from stochastic effects of the modeling process.

The predicted distribution of anthrax in Mexico suggests that *B. anthracis* is present in adjoining areas where the United States has reported outbreaks in recent years [3, 4]. For example, the northern region of Coahuila De Zaragoza and Nuevo Leon [82]) joins with the southwestern Texas border, where wide-spread epidemics in white-tailed deer, farmed exotic wildlife, and livestock have been frequent [12]. Both sides of the border are predicted to sustain the disease. This suggests that anthrax control efforts in the United States alone are not sufficient to prevent future outbreaks or protect either Mexican or U.S. livestock or wildlife interests.

5.3. Modeling Scenario 3: Predicting the Distribution of Bacillus anthracis in the United States in 2050

Climate change has become an important topic in the ecology community and the field of spatial epidemiology. This is evident in a recent Scientific and Technical Review by the Organisation Mondiale de la Santé Animale [83] dedicated to the topic of climate change and animal diseases, complete with several article on the effect of climate change on disease distributions. As a preliminary effort to evaluate the potential changes in the geographic distribution of *B. anthracis* in the contiguous United States, I developed a model of the organism under current climatic conditions and future climatic conditions using the HADCM3 B2 scenario.

A large body of literature supports that species have the potential to conserve ecological niches over evolutionary time scales [84–86]. Although it would be ideal to develop the U.S. models of *B. anthracis* by genetic lineage [4], the modeling efforts of Blackburn et al. [4] and illustrated here suggest that *B. anthracis* has established a natural ecology in the United States, and the modeling results show a high degree of predictive accuracy, including the ability to exclude some spurious data in the modeling phase. Blackburn et al. [4] show that GARP was sensitive to an outbreak in Oklahoma in 1957 that Van Ness [87] reports was likely from road maintenance that disturbed soils or more likely a food-borne outbreak in animal feed [4]. Given the conservative nature of the GARP rule sets across models for *B. anthracis* and that anthrax remains a reccurring disease on the American landscape despite control and vaccination efforts [6], I feel it is safe to assume *B. anthracis* will have niche requirements o similar to today's in 2050.

5.3.1. Model Results

Similar to the efforts for projecting the Mexican distribution, I used the current U.S. environmental variables and point data from recent outbreaks to build and test a model set to project onto the future climate data set. Accuracy metrics for this current-day model indicate models with low omission and high AUC scores (Table 3). Figure 8A shows the distribution of *B. anthracis* using the current climatic conditions from the HADCM3 data set. Although pixel sizes were quite large relative to the U.S. model presented in scenario 1, the overall geographic distribution is fairly similar, with a dominant south-to-north corridor through western Texas into the Dakotas and Minnesota and westward to the Snake River Drainage. A second east-to-west corridor exists across southern New Mexico and Arizona, with a disjunct portion of habitat in the California Central Valley.

Figure 8B illustrates the potential distribution of *B. anthracis* in 2050. Although the south-to-north corridor is still visible, there is an apparent gap in northern Texas, with more loss of spore-promoting habitat in southernmost Texas. To illustrate this better, I recoded presence and absence for both the current-day prediction and the future projection using the classification. I selected a threshold of six or more model agreement from the best subset of each time period and recoded all values of six or more to a score of 1 to define presence in the current-day models and a score of 4 in the future models. In the current-day model, values of five or less were scored as a zero;
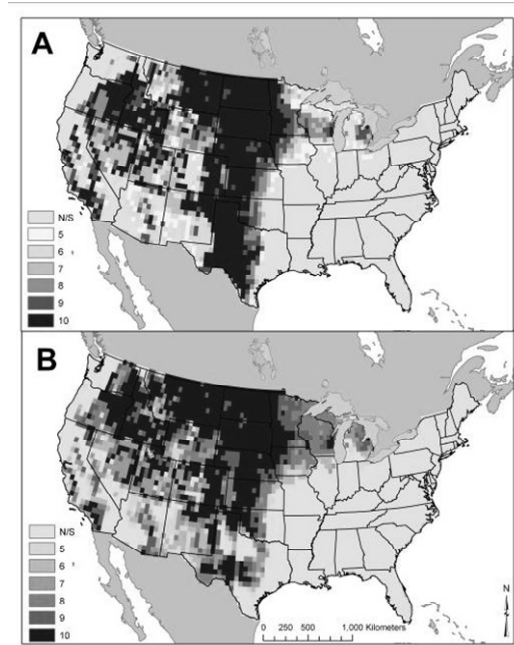
Figure 8. **A.** The potential geographic distribution of *Bacillus anthracis* in the contiguous United States based on a nine-variable ecological niche modeling experiment based on current climatic conditions from the HADCM3 B2 future climate scenario. **B.** The potential geographic distribution of *B. anthracis* in 2050 using the future climate data from the HADCM3 B2 climate scenario. A threshold of six models or better was selected to visualize these analyses.

in the future model, agreement of five or less was scored as a 1. I used raster algebra in the GIS to subtract the future scenario from the present scenario to evaluate areas of predicted overlap, habitat loss, and habitat expansion over the next 50 years. This provides four possible scores. A score of –1 indicates that both time periods predict those pixels as absent. A score of –4 indicates a habitat expansion, where *B. anthracis* is predicted in areas not predicted under the current-day model. A score of 0 indicates that *B. anthracis* is present in the current-day model and that habitat that supports spores has been lost. A score of –3 indicates that the species is predicted present under both climate scenarios. Figure 9 illustrates the differences in the two time periods.

5.3.2. Regional Changes

Although the HADCM3 is based on a global temperature increase of 2.1–3.9°C, changes in the species' distribution shown in Fig. 9 are regional. Overall, 50.7% of the predicted distribution of *B. anthracis* for the present day does not change in 2050. Another 39.6% of the landscape was not predicted as suitable in either of the predictions. There was a 6.09% loss of suitable environment from these experiments from present day to the future, with only a 3.57% expansion of habitat in the future data

set, for a total loss of 2.52% of the landscape no longer suitable for *B. anthracis* survival. Because these projections cannot be validated with field data for another 47 years, speculation must be tempered with caution. However, there are some interesting spatial patterns on the map in Fig. 9.
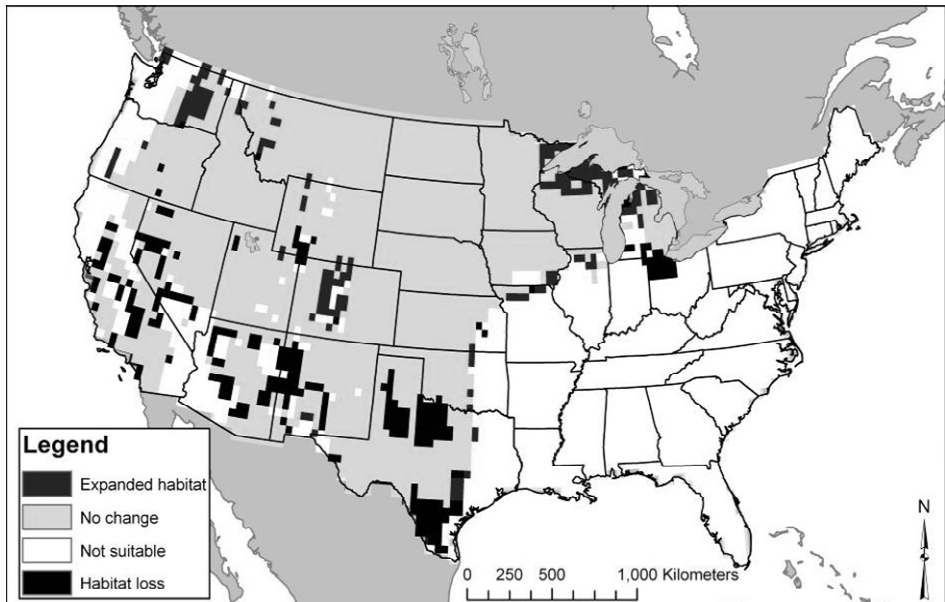

Figure 9. Spatial patterns of habitat expansion and loss between the current time period and 2050 for *Bacillus anthracis* from a GARP modeling experiment. Expanded habitat is present in 2050 but not in the current time period. No change indicates predicted presence in both the current and future model outputs. Not suitable indicates absence of *B. anthracis* habitat in both modeling experiments. Habitat loss indicates habitat that was predicted as present in the current scenario and excluded from the future scenario.

There is minimal geographic change between the two time periods in the Dakotas region. In contrast, much of the southern and southwestern parts of Texas are reduced from the present day to 2050. This shows a geographic change in the environment and the potential extirpation of *B. anthracis* from the southern portion of its U.S. range. This is interesting because the largest number of outbreaks and large numbers of individual animals have been reported in Texas [4]. At the relatively low latitude of southern Texas, it is likely that the increase in temperature and solar radiation may increase soil temperature beyond the physiological limits of the spores.

It is interesting to consider what might happen in the northern end of the predicted range, where any spatial expansion would be minimal. A number of studies on climate change in the Midwestern states have documented increases in winter air temperature [88], early arrival of spring temperatures [89], and earlier greenup periods in spring [90]. In the case of *B. anthracis* across the southern states, this scenario could suggest a reduction of the geographic range and a subsequent reduction in outbreaks as the

environment becomes unsuitable for spore survival. In the northern states, the models suggest a minimum amount of spatial expansion or contraction on the landscape, but earlier springs with longer growing seasons and warm periods might increase the number of outbreaks or the length of anthrax season. Isard et al. [88] also notes colder soil temperatures during the relatively shorter winters, but, given the frequently recurring outbreaks of anthrax in wood bison at much higher latitude [21], it is difficult to imagine that cooler soil temperatures would have a drastic effect on spore survival.

6. Conclusions and Future Directions

This chapter attempts to present a number of spatial analytical techniques within a larger theoretical framework of how to employ these techniques to advance spatial epidemiology and disease ecology. The techniques presented here are relatively easy to execute in the software applications introduced in the pages of this chapter. However, it is necessary to understand the limitations and underlying ecological or statistical theory that supports the results of an analysis. As an example, the body of literature on ESDA is only in its second decade. There are a number of research opportunities in determining the usefulness of several local measures of autocorrelation and their application to real-world epidemiological data sets such as those presented here looking at fly distributions and clusters or anthrax outbreaks. Likewise, the definition of the ecological niche is in no way static. Although those definitions first posed by Johnson [49], Grinnell [50], Hutchinson [51, 52], and MacArthur [53] and updated through a large body of literature (see Chase and Leibold [55]) have been circulating in the literature for nearly a century, there is a great deal of work to be done to better define niche variability, niche seasonality and niche competition and on how to incorporate more information on biotic interactions into the spatial-modeling process reviewed here, particularly as it pertains to pathogens and disease transmission.

Acknowledgements

References

1.   Anselin, L. Spatial statistical modeling in a GIS environment. In Maguire, D., Batty, M., Goodchild, M., editors. GIS, spatial analysis, and modeling. Redlands, CA: ESRI Press; 2005. p. 498.

2.   Smith, K.L., De Vos, V., Price, L.B., Hugh-Jones, M.E., Keim, P. 2000. Bacillus anthracis diversity in Kruger National Park. J. Clin. Microbiol. **38**:3780–3784.

3.   Hugh-Jones, M.E., De Vos, V. 2002. Anthrax and wildlife. Rev. Sci. Tech. **21**:359–383.

4.   Blackburn, J.K., McNyset, K.M., Hugh-Jones, M.E., Curtis, A. 2007. Modeling the geographic distribution of Bacillus anthracis, the causative agent of anthrax disease, for the contiguous United States using predictive ecological niche modeling. Am. J. Trop. Med. Hyg. **77**:1103–1110.

5.   Van Ness, G., Stein, C.D. 1956. Soils of the United States favorable for anthrax. J. Am. Vet. Med. Assoc. **128**:7–9.

6.   Van Ness, G.B. 1971. Ecology of anthrax. Science **172**:1303–1307.

7.   Dragon, D.C., Rennie, R.P. 1995. The ecology of anthrax spores: tough but not invincible. Can. Vet. J. **36**:295–301.

8.   Turner, A.J., Galvin, J.W., Rubira, R.J., Miller, G.T. 2001. Anthrax explodes in an Australian summer. J. Appl. Microbiol. **87**:196–199.

9.   Parkinson, R. Andrijana, R., Jenson, C. 2003. Investigation of an anthrax outbreak in Alberta in 1999 using a geographic information system. Can. Vet. J. **44**:315–318.

10. Lindeque, P.M., Turnbull, P.C.B. 1994. Ecology and epidemiology of anthrax in the Etosha National Park, Namibia. Onderstepoort J. Vet. Res. **61**:71–83.

11. Smith, K.L., De Vos, V., Bryden, H.B., Hugh-Jones, M.E., Klevytska, A., Price, L.B., Keim, P., Scholl, D.T. 1999. Meso-scale ecology of anthrax in southern Africa: a pilot study of diversity and clustering. J. Appl. Microbiol. **87**:204–207.

12. Blackburn, J.K. Evaluating the spatial ecology of anthrax in North America: examining epidemiological components across multiple geographic scales using a GIS-based approach. Doctoral Dissertation, Louisiana State University, Department of Geography and Anthropology, 2006.

13. Braack, L.E.O., De Vos, V. 1990. Feeding habits and flight ranges of blow-flies (Chryosoma spp.) in relation to anthrax transmission in the Kruger National Park, South Africa. Onderstepoort J. Vet. Res. **57**:141–142.

14. De Vos, V., Bryden, H.B. 1996. Anthrax in the Kruger National Park: temporal and spatial patterns of disease occurrence. Salisbury Med. Bull. **87**(suppl.):26–30.

15. Kraneveld F.C., Djaenoedin, R. 1940. Test on the dissemination of anthrax by Tabanus rubidus in horses and buffalo. Overgedrukt uit de Nederlands-Indische Bladen Voor Diergeneeskunde **52**:339–380.

16. Rao, N.S., Mohiyudeen, S. 1958. Tabanus flies as transmitters of anthrax − a field experience. Indian Vet. J. **35**:348–353

17. Davies, J.C. 1983. A major epidemic of anthrax in Zimbabwe. Part II. Cent. Afr. J. Med. **29**:8–12.

18. Turell, M.J., Knudson, G.B. 1987. Mechanical transmission of Bacillus anthrasis by stable flies (Stomoxys calcitrans) and mosquitoes (Aedes aegypti and Aedes taeniorhynchus). Infect. Immun. **55**:1859–1861.

19. Ganeva, D.J. 2004. Analysis of the Bulgarian tabanid fauna with regard to its potential for epidemiological involvement. Bulg. J. Vet. Med. **7**:1–8.

20. Fullbright, T.E., Ortega-S., J.A. White-tailed deer habitat ecology and management on rangelands. College Station, TX: Texas A&M University Press; 2006.

21. 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. Can. J. Vet. Res. **59**:256–264.

22. 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. Am. J. Trop. Med. Hyg. **69**:494–505.

23. Haines-Young, R., Green, D.R., Cousins, S. Landscape ecology and spatial information systems. In: Haines-Young, R., Green, D.R., Cousins, S, editors. Landscape ecology and geographic information systems. Bristol, UK: Taylor & Francis, Inc.; 1994. pp. 3–8.

24. Dragon D.C., Bader D.E., Mitchell J., Wollen N. 2005. Natural dissemination of Bacillus anthracis spores in northern Canada. Appl. Environ. Microbiol. **71**:1610–1615.

25. Rogers, D.J. 2006. Models for vectors and vector-borne diseases. In: Hay, S., Graham, A.J., Rogers, D.J., editors. Global mapping of infectious diseases: methods, examples, and emerging application. London: Academic Press; 2006.

26. Clarke, K.C., McLafferty, S.L., Tempalski, B.J. 1996. On epidemiology and geographic information systems: a review and discussion of future directions. Emerg. Infect. Dis. **2**:85–92.

27. Smith, A.P., Horning, N., Moore, D. 1997. Regional biodiversity planning and lemur conservation with GIS in western Madagascar. Conserv. Biol. **11**:498–512.

28. Blackburn, J.K., Curtis, A., Mujia, F.C., Jones, F., Dorn, P., Coates, R. 2008. The development of the Chagas' Online Data Entry System (CODES-GIS). Trans. GIS **12**:249–265.

29. Kulldorff, M. 2001. Prospective time periodic geographical disease surveillance using a scan statistic. J. Royal Stat. Soc. Series A (Stat. Soc.) **164**:61–72.

30. Guisan, A., Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. **135**:147–186.

31. Curtis, A.C., Blackburn, J.K., Sansyzbayev, Y. Using a geographic information system to spatially investigate infectious disease. In: Tibayrenc, M., editor. Encyclopedia of infectious diseases: modern methodologies. London: John Wiley & Sons, Inc; 2007.

32. Cutter, S.L., Boruff, B.J., Shirley, W.L. 2003. Social vulnerability to environmental hazards. Soc. Sci. Q. **84**:242–261.

33. Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. Geographic information systems and science, second edition. New York: Wiley; 2005.

34. Maguire, D.J., Batty, M., Goodchild, M.F., editors. GIS, spatial analysis, and modeling. Redlands, CA: ESRI Press; 2005.

35. Curtis, A., Mills, J.W., Leitner, M. 2007. Katrina and vulnerability: the geography of stress. J. Health Care Poor Underserved **18**:315–330.

36. Eisen, R.J., Bearden, S.W., Wilder, A.P., Montenieri, J.A., Antolin, M.F., Gage, K.L. 2006. Early-phase transmission of Yersinia pestis by unblocked fleas as a mechanism explaining rapidly spreading plague epizootics. Proc. Natl. Acad. Sci. U S A **103**:15380–15385.

37. Randolph, S., Rogers, D.J. Ecology of tick-borne disease and the role of climate. In: Ergonul, O., Whitehouse, C.A., editors. Crimean-Congo hemorrhagic fever: a global perspective. New York: Springer;2007. pp. 167–186.

38. Peterson, A.T., Sanchez-Cordero, V., Beard, C.B., Ramsey, J.M. 2002. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. Emerg. Infect. Dis. **8**:662–667.

39. Lam N.S.N., Liu, K.B. 1994. Spread of AIDS in rural America, 1982-1990. J. Acquir. Immune Defic. Syndr. **7**:485–490.

40. Peterson, A.T., Bauer, J.T., Mills, J.N. 2004. Ecologic and geographic distribution of Filovirus disease. Emerg. Infect. Dis. **10**:40–47.

41. Goodchild, M.F. 1992. Geographical information science. Int. J. Geo. Inform. Sys. **6**:31–45.

42. Mark, D.M. Geographic information science: defining the field. In: Duckham, M., Goodchild, M.F., Worboys, M.F., editors. Foundations of geographic information science. New York: Taylor and Francis; 2003. pp. 3–18.

43. Goodchild, M.F. 2004. GIScience: geography, form, and process. Ann. Assoc. Am. Geographers **94**:709–714.

44. Goodchild, M.F. Geographical information science: fifteen years later. In Fisher, P.F., editor. *Classics from IJGIS: twenty years of the International Journal of Geographical Information Science and Systems*. Boca Raton, FL: CRC Press; 2006. pp. 199–204.

45. Getis, A., Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. Geo. Anal. **24**:189–260.

46. Durbeck, H., Greiling, D., Estberg, L., Long, A., Jacquez, G. ClusterSeer™ software for identifying event clusters: user guide 2. Crytsal Lake, IL: TerraSeer, Inc.; 2002. p. 316.

47. Sagiyev, Z., Pazilov, Y., Lukhnova, L., Temiraliyeva, G., Meka-Menchenko, T., Sansyzbayev, Y., Joyner, T.A., Curtis, A., Hugh-Jones, M.E., Blackburn, J.K. Spatial hotspots of anthrax cases in Kazakh livestock: identifying control strategy needs. Oral Presentation. URISA's GIS in Public Health Conference, May 20–23, 2007, New Orleans, Louisiana.

48. Hutchinson, G.E. 1978. An introduction to population ecology. New Haven, CT: Yale University Press; 1978.

49. Johnson, R.H. Determinate evolution in the color pattern of the lady-beetles. Publication No. 122. Washington, DC: Carnegie Institute of Washington; 1910.

50. Grinnell, J. 1917. The niche-relationships of the California Thrasher. Auk. **34**:427–433.

51. Hutchinson, G.E. 1944. Limnological studies in Connecticut. VII. A critical examination of the supposed relationship between phytoplakton periodicity and chemical changes in lake waters. Ecology **25**:3–26.

52. Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbour Symposium on Quantitative Biology **22**:415–427.

53. MacArthur, R.H. 1958. Population ecology of some warblers of northeastern coniferous forests. Ecology **39**:599–619.

54. Morrison, M.L., Hall, L.S. Standard terminology: toward a common language to advance ecological understanding and application. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B., editors. Predicting species occurrences: issues of accuracy and scale. Washington, DC: Island Press; 2002. pp. 43–52.

55. Chase, J.M., Leibold, M.A. Ecological niches: linking classical and contemporary approaches. Chicago: University of Chicago Press; 2003.

56. Peterson, A.T. 2008. Biogeography of diseases: a framework for analysis. Naturwissenschaften **95**:483–491.

57. Peterson, A.T. 2006. Ecologic niche modeling and spatial patterns of disease transmission. Emerg. Infect. Dis. **12**:1822–1826.

58. Stockwell, D., Peters, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. Int. J. Geo. Inform. Sci. **13**:143–158.

59. Rogers, D.J. Satellites, space, time and the African trypanosomiases. In: Hay, S.I., Randolph, S.E., Rogers, D.J., editors. Remote sensing and geographical information systems in epidemiology. London: Academic Press; 2000.

60. Phillips, S.J., Anderson, R.P., Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. **190**:231–259.

61. Adjemian, J.C.Z., Girvetz, E.H., Beckett, L., Foley, J.E. 2006. Analysis of genetic algorithm for rule-set prodution (GARP) modeling approach for predicting distributions of fleas implicated as vectors of plague, Yersinia pestis, in California. J. Med. Entomol. **43**:93–103.

62. Ron, R.S. 2005. Predicting the distribution of the amphibian pathogen Batrachochytrium dendrobatidis in the New World. Biotropica **37**:209–221.

63. Stockwell D.R.B., Peterson A.T. 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. **148**:1–13.

64. Anderson, R.P., Lew, D., Peterson, A.T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol. Model. **162**:211–232.

65. Kluza, D.A., McNyset, K.M. 2005. Ecological niche modeling of aquatic invasion species. Aquat. Invad. **16**:1–7.

66. McNyset, K.M. 2005. Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas. Ecol. Freshwater Fish **14**:243–255.

67. Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones P.G., Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**:1965–1978.

68. Hay, S.I., Tatem, A.J., Graham, A.J., Goetz, S.J., Rogers, D.J. Global environmental data for mapping infectious disease distribution. In: Hay, S., Graham, A.J., Rogers, D.J., editors. Global mapping of infectious diseases: methods, examples, and emerging application. London: Academic Press; 2006.

69. Nakicenovic, N., Swart, R., editors. Emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press; 2000.

70. Peterson, A.T., Martínez-Meyer, E., González-Salazar, C., Hall, P. 2004. Modeled climate change effects on distributions of Canadian butterfly species. Can. J. Zool. **82**:851–858.

71. Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R., Stewart, A.M. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography **16**:120–127.

72. Centor, R.M. 1991. Signal detectability: the use of ROC curves and their analyses. Med. Decis. Mak. **11**:102–106.

73. Zweig, M.H., Campbell, G. 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. **39**:561–577.

74. Hanley, J.A., McNeil, B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology **143**:29–36.

75. Patterson, B.D., Ceballos, G., Sechrest, W., Tognelli, M.F., Brooks, T., Luna, L., Ortega, P., Salazar, I., Young, B.E. Digital distribution maps of the mammals of the western hemisphere, version 3.0. Arlington, VA: NatureServe; 2007.

76. Stein, C.D. 1945. The history and distribution of anthrax in livestock in the United States. Vet. Med. **40**:340–349.

77. Van Ert, M.N., Easterday, W.R., Huynh, L.Y., Okinaka, R.T., Hugh-Jones, M.E., Ravel, J., Zanecki, S.R., Pearson, T., Simonson, T., Uren, J.M., Kachur, S.M., Leadem-Dougherty, R.R., Rhoton, S.D., Zinser, G., Farlow, J., Coker, P.R., Smith, K.L., Wang, B., Kenefic, L.J., Fraser-Liggett, C.M., Wagner, D.M., Keim, P. 2007. Global genetic population structure of Bacillus anthracis. PLoS ONE **2**:e461.

78. Machado, M.A. 1976. An industry in limbo: the Mexican cattle industry 1920–1924. Ag. His. **50**:615–625.

79. Hugh-Jones, M. 1999. 1996–97 global anthrax report. J. Appl. Microbiol. **87**:189–191.

80. Fragoso Uribe, R., Villicana Fuentes, H. 1984. Antrax en dos communidades de Zacatecas, Mexico. Bol. Oficina Sanit. Panam. **97**:526–533.

81. Anthrax-cattle, human, livestock, Mexico (Michoacan). ProMED-mail, June 22, 2003. 20030622.1543. Available at http://www.promedmail.org. Accessed August 27, 2007.

82. Siefert, H.S., Bader, K., Cyplik, J., González Salinas, J., Roth, F., Salinas Meléndez, J.A., Sukop, U. 1996. Environment, incidence, aetiology, epizootiology and immunoprophylaxis of soil-borne diseases in north-east Mexico. Zentralbl. Veterinarmed B. **43**:593–606.

83. de la Rocque, S., Hendrickx, G., Morand, S., editors. 2008. Climate change: impact on epidemiology and control in animal diseases. Revue Scientifique et Technique, OIE, 27(2).

84. Holt, R.D., Gaines, M.S. 1992. Analysis and adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evolution. Ecol. **6**:433–337.

85. Peterson, A.T., Soberon, J., Sanchez-Cordero, V. 1999. Conservatism of ecological niches in evolutionary time. Science **285**:1265–1267.

86. Peterson, A.T. 2003. Predicting the geography of species' invasions via ecological niche modeling. Q. Rev. Biol. **78**:419–433.

87. Van Ness, G.B. 1959. Soil relationship in the Oklahoma-Kansas anthrax outbreak of 1957. J. Soil Water Conserv. **14**:70–71.

88. Isard, S.A., Schaetzl, R.J., Andresen, J.A. 2007. Soils cool as climate warms in the Great Lakes region: 1951–2000. Ann. Assoc. Am. Geog. **97**:467–476.

89. Strode, P.K. 2003. Implications of climate change for North American wood warblers (Parulidae). Global Change Biol. **9**:1137–1144.

90. Bradley, N.L., Leopold, A.C., Ross, J., Huffaker, W. 1999. Phenological changes reflect climate change in Wisconsin. Proc. Natl. Acad. Sci. U S A **96**:9701–9704.