ECOLOGICAL NICHE MODELING OF A ZOONOSIS: A CASE STUDY USING ANTHRAX OUTBREAKS AND CLIMATE CHANGE IN KAZAKHSTAN

By

TIMOTHY ANDREW JOYNER

A THESIS PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE UNIVERSITY OF FLORIDA 2010
To my family – thank you for the support that you have given to me all of my life
ACKNOWLEDGMENTS

I thank my family for all of the support and encouragement that they have given to me over my entire life and specifically during the past six years of my academic life. I also thank my fiancée, Jessica Bell, who has endured nearly two long years of being apart while I have completed this degree. I am forever grateful for the members of the Spatial Epidemiology and Ecology Research Lab both at California State University, Fullerton and the University of Florida: Ian Kracalik, Shane Foster, Ailam Truong, Pamela Rittelmeyer, and Ann Espejo. They have provided much-needed friendship and support throughout the entire graduate school process and that has been invaluable.

I would like to thank my thesis advisor, Dr. Jason K. Blackburn, for his constant support of my education over the past six years. He has provided an incredible amount of opportunities to me during both my undergraduate career at Louisiana State University and my graduate career at California State University, Fullerton and the University of Florida. I would also like to thank the members of my committee, Dr. Michael Binford and Dr. Andrew Tatem, for their support and guidance during the formation of this thesis.

I thank the United States Defense Threat Reduction Agency (DTRA) who funded this Cooperative Biological Research project as part of the Biological Threat Reduction Program in Kazakhstan. I thank Yerlan Sansyrbayev and Andrew Curtis who assisted greatly with the development of the original anthrax data set. I also thank Martin Hugh-Jones for his assistance and counsel during my research. Finally, I thank my colleagues in Kazakhstan whom I have worked with extensively over the past several years on multiple projects: Bolat Ospanov, Talgat Tassynov, Vladimir Dubyanskiy, Larissa Lukhnova, Yerlan Pazylov, Gulnara Temiraliyeva, and Alim Aikimbayev.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>7</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>8</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>13</td>
</tr>
<tr>
<td>Anthrax</td>
<td>13</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>19</td>
</tr>
<tr>
<td>Ecological Niche Modeling and Climate Change</td>
<td>22</td>
</tr>
<tr>
<td>The Genetic Algorithm for Rule-Set Prediction (GARP)</td>
<td>26</td>
</tr>
<tr>
<td>Thesis Goals</td>
<td>29</td>
</tr>
<tr>
<td>2 MODELING THE POTENTIAL DISTRIBUTION OF BACILLUS ANTHRACIS</td>
<td>32</td>
</tr>
<tr>
<td>UNDER MULTIPLE CLIMATE CHANGE SCENARIANS FOR KAZAKHSTAN</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>32</td>
</tr>
<tr>
<td>Data and Methods</td>
<td>41</td>
</tr>
<tr>
<td>Anthrax Occurrence Data</td>
<td>41</td>
</tr>
<tr>
<td>Current and Future Climate Datasets</td>
<td>42</td>
</tr>
<tr>
<td>Modeling Scenarios</td>
<td>45</td>
</tr>
<tr>
<td>Implementation and Methodology of Desktop GARP and Accuracy Metrics</td>
<td>46</td>
</tr>
<tr>
<td>Modeling Parameters</td>
<td>48</td>
</tr>
<tr>
<td>Analysis of Habitat Change</td>
<td>50</td>
</tr>
<tr>
<td>Results</td>
<td>51</td>
</tr>
<tr>
<td>Accuracy Metrics</td>
<td>51</td>
</tr>
<tr>
<td>Current and Future Distributions of <em>B. anthracis</em></td>
<td>51</td>
</tr>
<tr>
<td>Discussion</td>
<td>55</td>
</tr>
<tr>
<td>3 EVALUATING ENVIRONMENTAL PARAMETERS OF BACILLUS ANTHRACIS</td>
<td>78</td>
</tr>
<tr>
<td>IN KAZAKHSTAN: AN EXAMINATION OF RULE-SET WRITING AND MAPPING</td>
<td></td>
</tr>
<tr>
<td>WITHIN AN ECOLOGICAL NICHE MODELING TOOL</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>78</td>
</tr>
<tr>
<td>Data and Methods</td>
<td>90</td>
</tr>
<tr>
<td>Anthrax Occurrence Data</td>
<td>90</td>
</tr>
<tr>
<td>Ecological Niche Modeling</td>
<td>91</td>
</tr>
<tr>
<td>Modeling Procedures and Methods</td>
<td>94</td>
</tr>
<tr>
<td>Analysis of Environmental Parameters Established within GARP Rule-Sets</td>
<td>95</td>
</tr>
</tbody>
</table>
4 CONCLUSION AND FUTURE RESEARCH ................................................................. 127
LIST OF REFERENCES .................................................................................. 135
BIOGRAPHICAL SKETCH ........................................................................... 149
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Environmental variables used for Genetic Algorithm for Rule-set Prediction (GARP) models</td>
<td>66</td>
</tr>
<tr>
<td>2-2</td>
<td>Accuracy Metrics for the current predicted distributions</td>
<td>66</td>
</tr>
<tr>
<td>2-3</td>
<td>A comparison of habitat change (%) between Special Report on Emissions Scenarios (SRES) A2 and B2 climate change scenarios</td>
<td>66</td>
</tr>
<tr>
<td>3-1</td>
<td>Accuracy Metrics for the current predicted distributions</td>
<td>111</td>
</tr>
<tr>
<td>3-2</td>
<td>Dominant rules from one of the 10 best subsets created in GARP using current environmental conditions that included measures of precipitation, temperature, and Normalized Difference Vegetation Index (NDVI).</td>
<td>113</td>
</tr>
<tr>
<td>3-3</td>
<td>Commission values for the future GARP model that utilized the A2 climate change scenario</td>
<td>117</td>
</tr>
<tr>
<td>3-4</td>
<td>Commission values for the future GARP model that utilized the B2 climate change scenario</td>
<td>119</td>
</tr>
<tr>
<td>3-5</td>
<td>An example of minimum and maximum ranges of two rule-sets produced in current scenario one</td>
<td>120</td>
</tr>
<tr>
<td>3-6</td>
<td>An example of minimum and maximum ranges of two rule-sets produced in current scenario two</td>
<td>121</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Political map of Kazakhstan with oblasts, topography, major cities, rivers, and lakes</td>
<td>31</td>
</tr>
<tr>
<td>2-1</td>
<td>Map of Kazakhstan with anthrax locality data</td>
<td>68</td>
</tr>
<tr>
<td>2-2</td>
<td>Ten 85/15 random subsets with corresponding model runs showing consistency between testing and training point locations as well as predicted distributions.</td>
<td>69</td>
</tr>
<tr>
<td>2-3</td>
<td>Random 85/15 subsets were created for the northern and southern areas of Kazakhstan (separated at 48°N) independently, then added together to construct ten total testing and training subsets.</td>
<td>70</td>
</tr>
<tr>
<td>2-4</td>
<td>Current predicted distribution of <em>B. anthracis</em> using Current Scenario 1 (A) and future predicted distributions based on the A2 climate change scenario (B) and B2 climate change scenario (C)</td>
<td>71</td>
</tr>
<tr>
<td>2-5</td>
<td>Potential future habitat changes based on the A2 climate change scenario (A) and B2 climate change scenario (B) derived from Current Scenario 1. The differences between these climate change scenarios are shown in (C).</td>
<td>72</td>
</tr>
<tr>
<td>2-6</td>
<td>Current predicted distribution of <em>B. anthracis</em> using Current Scenario 2 (A) and future predicted distributions based on the A2 climate change scenario (B) and B2 climate change scenario (C)</td>
<td>73</td>
</tr>
<tr>
<td>2-7</td>
<td>Potential future habitat changes based on the A2 climate change scenario (A) and B2 climate change scenario (B) derived from Current Scenario 2. The differences between these climate change scenarios are shown in (C).</td>
<td>74</td>
</tr>
<tr>
<td>2-8</td>
<td>Current Scenario 3 (A) and Current Scenario 4 (B) show the current distribution of <em>B. anthracis</em> using different environmental variables. The difference between scenarios 3 and 4 is also examined (C).</td>
<td>75</td>
</tr>
<tr>
<td>2-9</td>
<td>Current predicted distribution of <em>B. anthracis</em> using Current Scenario 4 (A) and future predicted distributions based on the A2 climate change scenario (B) and B2 climate change scenario (C)</td>
<td>76</td>
</tr>
<tr>
<td>2-10</td>
<td>Potential future habitat changes based on the A2 climate change scenario (A) and B2 climate change scenario (B) derived from Current Scenario 4. The differences between these climate change scenarios are shown in (C).</td>
<td>77</td>
</tr>
<tr>
<td>3-1</td>
<td>Genetic Algorithm for Rule-set Prediction (GARP) models showing the summated best subsets for current scenario one (A) and current scenario two (B)</td>
<td>110</td>
</tr>
</tbody>
</table>
3-2 Rules from one of the 10 best subsets in current scenario one (A) and rules form one of the 10 best subsets in current scenario two (B)............................... 112

3-3 Maps showing the dominant rules (presence – red color ramp; absence – blue color ramp) of the 10 best subset tasks projected onto the landscape for current scenario one. ................................................................. 114

3-4 Maps showing the dominant rules (presence – red color ramp; absence – blue color ramp) of the 10 best subset tasks projected onto the landscape for current scenario two. ................................................................. 115

3-5 Maps showing the dominant rules (presence – red color ramp; absence – blue color ramp) of the 10 best subset tasks projected onto the landscape for the A2 climate change scenario................................................................. 116

3-6 Maps showing the dominant rules (presence – red color ramp; absence – blue color ramp) of the 10 best subset tasks projected onto the landscape for the B2 climate change scenario................................................................. 118

3-7 Median range of variables describing both northern and southern distributions using measures of Normalized Difference Vegetation Index (NDVI) (current scenario one)................................................................. 122

3-8 Median range of variables describing both northern and southern distributions without using measures of NDVI (current scenario two)............ 122

3-9 Variable cloud delineated by any area predicted present (light grey) and areas predicted present by all 10 models (dark grey) and visualized in dimensions of wettest month precipitation and temperature range.............. 123

3-10 Variable cloud delineated by location and visualized in dimensions of wettest month precipitation and temperature range................................................................. 124

3-11 Variable cloud delineated by any area predicted present (light grey) and areas predicted present by all 10 models (dark grey) and visualized in dimensions of mean NDVI and mean temperature................................................................. 125

3-12 Variable cloud delineated by location and visualized in dimensions of mean NDVI and mean temperature................................................................. 126
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>BioClim</td>
<td>Bioclimatic</td>
</tr>
<tr>
<td>CGCM</td>
<td>Coupled ocean-atmosphere General Circulation Model</td>
</tr>
<tr>
<td>CSIRO</td>
<td>Commonwealth Scientific and Industrial Research Organisation</td>
</tr>
<tr>
<td>DG</td>
<td>Desktop GARP</td>
</tr>
<tr>
<td>ENFA</td>
<td>Ecological Niche Factor Analysis</td>
</tr>
<tr>
<td>ENM</td>
<td>Ecological Niche Modeling</td>
</tr>
<tr>
<td>FSU</td>
<td>Former Soviet Union</td>
</tr>
<tr>
<td>GARP</td>
<td>Genetic Algorithm for Rule-set Prediction</td>
</tr>
<tr>
<td>GCM</td>
<td>General Circulation Model</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information Systems</td>
</tr>
<tr>
<td>HadCM3</td>
<td>Hadley Coupled Model version 3</td>
</tr>
<tr>
<td>IPCC</td>
<td>Intergovernmental Panel on Climate Change</td>
</tr>
<tr>
<td>KSCQZD</td>
<td>Kazakh Science Center for Quarantine and Zoonotic Diseases</td>
</tr>
<tr>
<td>MaxEnt</td>
<td>Maximum Entropy</td>
</tr>
<tr>
<td>MLVA</td>
<td>Multiple Locus Variable number tandem repeat Analysis</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>PCA</td>
<td>Principle Components Analysis</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>SRES</td>
<td>Special Report on Emissions Scenarios</td>
</tr>
<tr>
<td>STATSGO</td>
<td>State Soil Geographic</td>
</tr>
<tr>
<td>TALA</td>
<td>Trypanosomiasis And Land use in Africa</td>
</tr>
<tr>
<td>USSR</td>
<td>Union of Soviet Socialist Republics</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
</tbody>
</table>
Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

ECOLOGICAL NICHE MODELING OF A ZOONOSIS: A CASE STUDY USING
ANTHRAX OUTBREAKS AND CLIMATE CHANGE IN KAZAKHSTAN

By

Timothy Andrew Joyner

May 2010

Chair: Jason K. Blackburn
Major: Geography

Anthrax, caused by the bacterium Bacillus anthracis, is a zoonotic disease that persists throughout much of the world in livestock, wildlife, and secondarily infects humans. This is true across much of Central Asia, and particularly the Steppe region, including Kazakhstan. This study employed the Genetic Algorithm for Rule-set Prediction (GARP) to model the current and future geographic distribution of B. anthracis in Kazakhstan based on the A2 and B2 Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios (SRES) climate change scenarios using a 5-variable dataset at 55km² and 8km², a 6-variable Bioclimatic (BioClim) dataset at 8km², and an 8-variable dataset using BioClim and measures of vegetation at 8km². Additionally, extracting landscape level ecological ranges of B. anthracis may help to better understand the conditions predicted by the models. Two studies have indicated that GARP produces useful biological information and through examining the rule-sets created by GARP, we can develop a more robust explanation as to why the species is present in some areas and absent in others. Through the use of the rule-set writing and mapping application of GARP, the study also examined the ranges of various parameters and how these ranges differed at varying latitudes.
Future models suggest large areas predicted under current conditions may be reduced by 2050 with the A2 model predicting 14-16% loss across the three spatial resolutions. There was greater variability in the B2 models across scenarios predicting ~15% loss at 55km², ~34% loss at 8km², and ~30% loss with the BioClim variables. Only very small areas of habitat expansion into new areas were predicted by either A2 or B2 in any models. Greater areas of habitat loss are predicted in the southern regions of Kazakhstan by A2 and B2 models, while moderate habitat loss is also predicted in the northern regions by either B2 model at 8km². Lower temperature ranges were observed in the southern region along with wider precipitation ranges as compared to those in the north. Additionally, the distribution of *B. anthracis* was defined by a narrow range of mean Normalized Difference Vegetation Index (NDVI). Though much variation was exhibited between rule types and total number of rules used for each model experiment, a consistent environmental envelope that supports *B. anthracis* survival was identified and spatially visualized. Anthrax disease control relies mainly on livestock vaccination and proper carcass disposal, both of which require adequate surveillance. In many situations, including that of Kazakhstan, vaccine resources are limited, and understanding the geographic distribution of the organism, in tandem with current data on livestock population dynamics, can aid in properly allocating doses. While speculative, contemplating future changes in livestock distributions and suitable environments for *B. anthracis* can be useful for establishing future surveillance priorities. This study may also have broader applications to global public health surveillance relating to other diseases in addition to *B. anthracis*. 
CHAPTER 1
INTRODUCTION

Anthrax

Anthrax is considered a disease of antiquity and is thought to have originated from sub-Saharan Africa and then subsequently spread by humans to Eurasia, North America, and Australia where it became endemic in areas with specific environmental conditions favorable to its survival (Hanson 1959, Hart & Beeching 2002, Keim et al. 1997, Kolonin 1971, Van Ness & Stein 1956, Van Ness 1971). Anthrax is a bacterial disease that is caused by the organism *Bacillus anthracis* and it primarily affects ungulates (i.e., both livestock and wildlife) (Bardell 2002, Thappa and Karthikeyan 2001). Infection begins through ingestion or inoculation (biting flies) and can lead to acute gastrointestinal infections or cutaneous infections. Gastrointestinal infections usually lead to acute septicemia, which often results in death (Lindeque & Turnbull 1994, Wallace et al. 2002). Current research suggests that the route of infection is species specific such as cattle becoming infected through the ingestion of contaminated soil and Kudu becoming infected through the ingestion of contaminated leaves (Braack & de Vos 1990, Thappa & Karthikeyan 2001). Not only does the disease decrease the population of a herd of cattle (an economic loss for a farmer/herder), but the disease can also be transferred to humans through contact with an infected animal (e.g., handling of a carcass, coming in contact with the hides, etc.) or through inhalation (Woods et al. 2004). In humans, anthrax often manifests itself in the cutaneous form, causing skin lesions or other dermatological problems that are usually non-fatal, but can be if not treated (Woods et al. 2004). Anthrax can also be manifested in humans through two other modes: pulmonary and gastrointestinal (Turnbull et al. 1998).
Pulmonary anthrax is contracted through inhalation and almost always results in death, while gastrointestinal anthrax is contracted through ingestion and causes severe inflammation and intestinal difficulty and approximately half of all cases result in death.

To improve surveillance and vaccination efforts of the disease, we must first understand the underlying organism that causes anthrax: *Bacillus anthracis*. The word “Bacillus” has Latin origins and means “rod,” while the word “anthracis” has Greek origins and means “coal” (Bardell 2002). Basically, “Bacillus” explains the rod-shaped structure of the organism and “anthracis” explains the black blood and black skin lesions that are symptoms of an infection. *Bacillus anthracis* is a large, nonmotile, brick-shaped, Gram-positive organism (2.5 x 10 micrometers) that occurs singly or in pairs in tissue (Lalitha & Kumar 1996). The bacterium may be able to survive in both a vegetative cell form and a spore form. A vegetative cell is an actively growing cell while a spore is an endospore that has been released from a cell, but is capable of germinating and producing actively replicating cells. Virulence, the relative ability of a pathogen to cause a disease, must be maintained in order for an infection to occur regardless of the form in which the bacterium survives. *B. anthracis* virulence is due to toxic factors from two plasmids: px01 and px02. Virulence factors include the encoding edema factor (EF), lethal factor (LF), and pathogen (or protective) antigen (PA) genes which are located in the px01 plasmid, while the genes involved in biosynthesis of the capsule are located in the px02 plasmid (Hugh-Jones & Blackburn 2009, Thappa & Karthikeyan 2001). The PA genes carry the EF and LF into the target cell and protect the bacterium, while the capsule causes resistance of vegetative forms of the bacterium to phagocytosis (Giagtzoglou & Bellen 2006). If a certain level of virulence is
maintained, then for an infection to occur some form of *B. anthracis* is ingested and the bacteria is then phagocytosed by macrophages and carried to regional lymph nodes in the body where it can germinate within macrophages and become a vegetative bacterium. The vegetative bacterium is then released from macrophages and finds its way into the blood stream where it multiplies and causes hemorrhagic septicemia (Hugh-Jones & Blackburn 2009, Mongoh et al. 2008). Virulence factors that are caused by *B. anthracis* result in toxemia with systemic effects that cause the host to die (Mongoh et al. 2008, Thappa & Karthikeyan 2001). Symptoms usually include a dark, bloody discharge from the mouth and anus in addition to the swelling of vessels and lymph nodes and an enlarged spleen (Dragon & Elkin 2001, Hugh-Jones & de Vos 2002). Sporulation occurs when the vegetative cells are exposed to oxygen and soil can be re-contaminated if spores re-enter the soil adjacent to a contaminated carcass (Dragon et al. 2005). The spores can then survive for very long periods of time under harsh conditions, but evidence suggests that spores begin to lose virulence if successive outbreaks do not occur (Hugh-Jones & Blackburn 2009, Mongoh et al. 2008, Moynihan 1963, Sharp & Roberts 2006). Most field research has indicated that *B. anthracis* is only maintained in spore-form between outbreaks (Dragon & Elkin 2001, Gates et al. 1995, Turnbull et al. 1996). However, the vegetative cell form of *B. anthracis* may also be able to survive in the soil if it is supported by complex biofilms that help to maintain virulence between outbreaks by allowing for the replication of the bacterium within the biofilms (Schuch & Fischetti 2009).

Regardless of which form the bacteria exist in between outbreaks, it is maintained in specific soil environments. In the mid-20th century, soil and other environmental
conditions were examined and it was discovered that basic mollic or chernozemic soils that are alkaline and rich in organic matter or calcium provide an ideal environment for the organism to survive in spore form (Van Ness & Stein 1956, Van Ness 1971). Van Ness & Stein (1956) outlined favorable soils for anthrax and created one of the first deterministic spatial distributions of where *B. anthracis* was likely to exist in the US. The study examined where soils exist that are favorable for *B. anthracis* survival and where anthrax cases had occurred historically. Areas that matched both criteria were considered to be at a higher risk for anthrax outbreaks. Similar soil mapping efforts were also taking place in the Former Soviet Union (FSU) around the same time (Hugh-Jones & Blackburn 2009).

It has also been suggested through field study that different genetic strains of *B. anthracis* have different soil preferences (Smith et al. 1999, 2000). A study in Kruger National Park revealed that A and B strains of *B. anthracis* were present in the same year and represented an overlap in their distributions and time of occurrence, but isolate B was generally found in soils with different calcium and pH levels than those coinciding with isolate A indicating unique environmental requirements for each strain (Smith et al. 2000). Field evidence also indicates that genetically unique *B. anthracis* groups show different soil affinities. At a local scale, the A group has a greater tolerance for lower pH and calcium levels (Smith et al. 2000). Despite evidence of the increasing likelihood of a highly complex *B. anthracis* life cycle and a broad genetic diversity, based on what we know about the epidemiology of anthrax it has a specific environmental reservoir and still occurs in previously researched soil and environmental conditions (Dragon &

Many theories about the life cycle of *B. anthracis* have been proposed. The incubator area theory proposed in Van Ness (1971) stated that an area may contain micro-environments and climatic conditions which favor germination of spores, subsequent vegetative multiplication and further sporulation. These areas coincided with prime soil conditions and anthrax outbreaks (Van Ness & Stein 1956, Van Ness 1959). The persistent spore theory argued that the vegetative growth cycle of *B. anthracis* can only occur with a host and that the soil and surrounding environment simply store the spores until a new host arrives (Gates et al. 1995). Another theory, known as the concentrator area theory, also argues that vegetative growth does not occur without a host, but that certain high-calcium environments could aid in spore preservation and post-dormancy germination (Dragon & Rennie 1995).

The incubator area theory has been heavily scrutinized since its proposal and evidence has been lacking to confirm the theory, but other research has also alluded to the possibility of incubation areas (Gates et al. 1995, Prins & Weyerhauser 1987, Rees et al. 1977). Prins & Weyerhauser (1987) indicated that “incubation areas” were possible areas of higher risk and noted that soil conditions and flooding were contributing factors in these areas. The study stated that climate may effect and even create “incubator areas.” Another study noted the propensity of anthrax outbreaks to occur after long periods of dormancy and hypothesized that anthrax bacilli may undergo cycles of propagation in the soil, but the appearance of the disease is limited to periods when the environment has been altered by climatic disturbances or other elements.
(Rees et al. 1977). Gates et al. (1995) outlined how the incubator area hypothesis may relate to bison outbreaks in northern Canada. Two alternatives were proposed to describe organism survival between outbreaks. The first hypothesis was that a spore is formed when *B. anthracis* is exposed to oxygen post-infection and that the organism remains dormant in spore form for potentially long periods of time between outbreaks. The second hypothesis was the incubator area theory that *B. anthracis* undergoes potentially multiple cycles of spore germination, vegetative cell outgrowth, and resporulation in the soil environment. After much research about the possibility of a spores-only survival option versus a repetitive cycle of vegetative multiplication and re-sporulation, a recent laboratory study suggested that *B. anthracis* may have the ability to vegetate and multiply in the soil through the use of biofilms which could maintain vegetative reservoirs (Schuch & Fischetti 2009). Through laboratory testing, the study indicated that *B. anthracis* can interact with and colonize invertebrate, soil-borne worm populations and can therefore have a more dynamic alternative lifecycle as opposed to sporulation and extended periods of dormancy. Additionally, Saile & Koehler (2006) suggested with lab experiments that *B. anthracis* can germinate and replicate in certain grassland environments by surviving with a plant host. This indicates that plants may play a role in the life cycle of *B. anthracis* by helping the organism maintain its virulence between outbreak events. Cherkasskiy (1999) also suggested that spore survival alone does not support long-term organism survival because of a loss of virulence over time and that soil infectivity is only maintained through repetitive biotic/abiotic cycles which would indicate that a natural reservoir or “incubator area” exist. However, Cherkasskiy (1999) infers that a mixture of theories is probably true concerning the life cycle of *B.*
*B. anthracis* because one theory over another has not been confirmed. An increasing knowledge of the organism has resulted in the belief that *B. anthracis* has a highly complex life cycle.

**Kazakhstan**

The examination of global patterns of disease dispersal and distribution is imperative to the planning and advancement of public health and disease surveillance. Insufficient disease monitoring subsists in most parts of the world, causing undue strain on the public health systems of many nations because of the inability to observe and track the initial stages of potential disease outbreaks (Cherkasskiy 1999, Hugh-Jones 1990, Hugh-Jones & de Vos 2002). Various diseases create immense stresses on a population through direct (human infections) or indirect (e.g., animal infections) means. Central Asia presents a prime area to study how public health systems have responded over the past several decades to disease outbreaks because of the availability of long records over much of the 20th century (Veatch 1989, Vlassov 2000). Central Asia is primarily composed of the countries of Kazakhstan, Uzbekistan, and Kyrgyzstan which are part of what is collectively called the Former Soviet Union (FSU). The countries were once part of the largest public health surveillance system in the world under the guidance of the Union of Soviet Socialist Republics (USSR). In 1991, the functionality of the system was significantly reduced in coordination with the dissolution of the Soviet Union (Wuhib et al. 2002). Newly independent countries, such as Kazakhstan, were forced to use the fragmented pieces of the former public health system to build a new public health system – a very difficult task when funding is considerably limited (Coker et al. 2004).
The rapid deterioration of public health surveillance in the country of Kazakhstan in the post-USSR period is the cause of a multitude of current health issues and problems (Coker et al. 2004). In recent years, the implementation of new monitoring technologies and techniques in Kazakhstan has lead to moderate improvements in public health surveillance, but there are still many areas of the system that need to be upgraded. Geographic information systems and spatial and temporal analysis techniques have been integrated into many disease surveillance and monitoring projects and consequently have improved database management efforts. Many diseases that are known to exist in Kazakhstan (e.g., brucellosis, Crimean- Congo hemorrhagic fever, hanta virus, anthrax, etc.) are currently being monitored with the use of many of the aforementioned new technologies. Historically, surveillance efforts have been employed in the FSU for many decades and some records for Kazakhstan even contain information about disease locations from the 1930’s (Aikimbayev unpublished manuscript). Most records were kept on paper and data sharing was difficult, but the introduction of computer technology has helped to create more up-to-date digital records and has also improved collaboration efforts.

Records for anthrax outbreaks in Kazakhstan have been well-kept and are quite robust. The records primarily contain reports of livestock infection (i.e., sheep and cattle). Locations of nearly 4,000 cases of anthrax outbreaks have been documented across the country dating back to 1933. The overall size and time-span of the records for Kazakhstan make anthrax an ideal disease to study. Previous and current spatial and temporal patterns can be analyzed to better understand the disease. Coincidentally, its interaction with the environment in Kazakhstan may lead to furthering
the overall knowledge of the disease that can then be applied on a global scale of analysis. However, the FSU and the republics of Central Asia in particular usually have the highest anthrax morbidity in animals and humans compared to other parts of Asia both historically and currently (Cherkasskiy 1999). Anthrax became an even bigger problem after the fall of the USSR because of financial problems and deterioration in the ability to control the disease (Cherkasskiy 1996, Hugh-Jones 1990). The region has a history of epizootics occurring over the past two centuries and many studies are ongoing in an effort to understand various aspects of anthrax that could be integrated into control and management programs (Aikimbayev et al. 2010, Joyner et al. 2010). Recent efforts in Central Asia have addressed major issues concerning anthrax and *B. anthracis* by examining genetic diversity of archival strains (Aikimbayev et al. 2010) and spatial clusters (Aikimbayev et al. unpublished manuscript, Kracalik et al. In Review) and predicted distributions (Aikimbayev et al. unpublished manuscript).

In an effort to understand the distribution of anthrax outbreaks, we must first examine the geography and climate of Kazakhstan and how they may affect livestock populations and movement. Kazakhstan is the ninth largest country in the world (~2,727,300 km²) and its shape has much greater longitudinal variation than latitudinal variation creating a larger differentiation of types of landscape across the country from west to east (refer to Figure 1-1). Insolation decreases from south to north and atmospheric pressure increases, strongly affecting the soil and vegetation covers (Woodward & Geldyeva 2006). The country is composed of seven main landscape types: plains, lowlands, plateaus, high plains, insular mountains, intermountain troughs, and mountain ranges. Most of Kazakhstan is composed of plain landscapes that
include forest steppe regions, semi-arid steppe regions, semi-desert regions, and desert regions. The remaining landscape of Kazakhstan is composed primarily of mountainous regions where a variation of forest, meadow, steppe, and desert landscapes persist. Plains and forest-steppe regions dominate the western and northern areas of Kazakhstan, while semi-arid steppe and desert regions dominate the central, southern, and eastern areas of Kazakhstan. The extreme southeastern areas of Kazakhstan are dominated by mountainous regions. Overall, arid landscapes make up more than 50% of the landscape of Kazakhstan (Woodward & Geldyeva 2006).

Livestock rangeland is predominantly located in the northern and southeastern regions of Kazakhstan and cattle migration is confined in the area because of political limitations placed on nomadic herdsmen over the past century, but livestock populations and migrational patterns are often dependent on climate (Robinson & Milner-Gulland 2003). Robinson & Milner-Gulland (2003) examined factors that regulated domestic livestock numbers and movement over the past century in Kazakhstan and determined that the timing and amount of precipitation are the most crucial factors.

**Ecological Niche Modeling and Climate Change**

To predict where *B. anthracis* may occur on the landscape of Kazakhstan, we must first understand the niche as a theoretical concept, and then apply this concept in a computer modeling environment. Grinnell (1917, 1924) defined the ecological niche as the environmental conditions needed by a species to maintain its population without immigration. The Grinnellian definition of a niche focuses on the range of ecological conditions where a population can be maintained and describes the niche as a limit on geographic distribution (Peterson 2003). Hirzel et al. (2008) outlined the difference between the aforementioned Grinnellian definition and the Eltonian definition of a niche.
The Eltonian definition describes a species relationship with other species. The presence of a species indicates that environmental parameters allow for the population to be maintained and that interactions with other species allow for the species to survive (Hirzel et al. 2008, Soberon 2007). Because inter- and intra-specific interactions are difficult to model, for the purpose of this study the definition of a niche will be expanded and follow the Hutchinsonian theory that the niche is an n-dimensional hypervolume of ecological parameters that allow a species to maintain its population without immigration (Hutchinson 1957). The geographical and ecological footprint is also referred to as the fundamental niche (Hutchinson 1957). A fundamental niche is the potential maximum suitable environment where a species can inhabit, but a species often does not inhabit many areas of its fundamental niche because of other variables that are not environmental or bioclimatic (e.g., competition, urban development, mobility). The resulting area where the species actually does inhabit is called the realized niche and this area is usually established after analyzing the fundamental niche (Hutchinson 1957, MacArthur 1958). This process of post hoc analysis is highly complex and there is often no method to determining the exact realized niche of a species, but identifying the perceived fundamental and realized niches is at least a starting point.

There are many tools available for modeling the potential distribution of a species and ecological niche modeling (ENM) is one such tool (Anderson et al. 2002, Peterson 2001, Soberon & Peterson 2005). ENM predicts the potential geographic distribution of a species on the selected landscape by analyzing the relationships that exist between species locality data and combinations of environmental variables. The Genetic
Algorithm for Rule-set Prediction (GARP) is a popular ENM that uses the Hutchinsonian niche definition of an "n-dimensional hypervolume" to fit a model to the Grinnellian definition. Biological variables (such as competition) described by Hutchinson (1957) and MacArthur (1958) cannot be directly accounted for in the ecological niche modeling process as their input environmental variables are limited to bioclimatic and abiotic variables. In this sense, it is appropriate to suggest that GARP predictions estimate the geographic space that supports the Grinnellian niche. In essence, GARP predicts the maximum occupiable space based on variables used within the model-building process.

According to the theory of niche conservatism, which states that a species maintains the same ecological niche over very long periods of time, the ecological relationships that exist today should be maintained in the near future (e.g., 2050) (Peterson et al. 1999). The niche requirements, being controlled by various biotic and abiotic factors, will not experience rapid changes and the bulk of a species’ population will survive within the niche conserved through time and across the mean phenotype (Holt and Gaines 1992). This has been shown to hold true over long periods of evolutionary time (Peterson et al. 1999) and in the short-term GARP can be used in conjunction with the knowledge of niche conservation to estimate future distributions based on current climatic relationships and future climate predictions (Atzmanstorfer et al. 2007, Huntley et al. 2004, Parra-Olea et al. 2005, Pearson et al. 2006, Peterson et al. 2001, Peterson 2003, Thuiller 2004). Despite the potential for relatively rapid evolutionary change in bacteria, and known genetic diversity across the global distribution of *B. anthracis* (Van Ert et al. 2007), the organism (like all other species) maintains a niche with Grinnellian and MacArthurian controls. Additionally, recent efforts suggest that GARP experiments can
accurately predict the geographic distribution of areas with repeat outbreaks using
landscape-level ecological parameters (Blackburn et al. 2007).

Some ENMs such as GARP can project the future distribution of a species, here
*B. anthracis*, based on the parameters of the current modeled niche requirements and
this may be very useful when implementing surveillance and control strategies since the
ecology of *B. anthracis* appears to be likely unchanged when we examine endemic
anthrax. Many other studies have used a similar approach to predict the future
distribution of a species based on climate change scenarios (Atzmanstorfer et al. 2007,
2004). A comparable study was seen in Parra-Olea et al. (2005). The study examined
the current and future distributions of two species of salamanders found in Mexico. The
Canadian Center’s CGCM2 SRES A2 scenarios for 2045–2055 were used and a
projected distribution was created. Drastic habitat contraction was predicted for both
species of salamanders. Areas where a suitable habitat was found tended to rapidly
recede to higher elevations, causing an increasingly fragmented ecological footprint.
The rapid recession was also alarming because it was at a faster rate than the
salamanders were able to move and adapt to, leading to the possibility of extinction for
each species of salamander. Because the ecological niche of the salamander is
conserved over very long periods of time, the salamander must move to the new areas
where environmental parameters will remain the same or face extinction. Adapting to a
new ecological niche takes a long time, but relatively drastic climatic changes are
expected to take place over the next 50 to 100 years. While similar studies are
becoming more and more common, much uncertainty surrounds the prediction of a
species’ future distribution (Davis et al. 1998, Pearson et al. 2006, Thuiller 2004). Davis et al. (1998) argued appropriately that we have no means of determining the changing interactions between species because of climate change; however the global climate may continue to change in the coming century and the accuracy of future predictions will be immeasurable until we reach the projected year.

**The Genetic Algorithm for Rule-Set Prediction (GARP)**

The Genetic Algorithm for Rule-set Prediction (GARP) was utilized to develop an ENM of *B. anthracis* across Kazakhstan. GARP is a presence-only modeling tool that uses species locality data and environmental coverage sets that can include such variables as temperature, precipitation, altitude, and measures of vegetation (Stockwell & Peters 1999). GARP determines the relationships that exist between the locality data and environmental data through a process of IF/THEN rule types developed in an iterative, stochastic process. A total of 50 rules are created from four main rule types (atomic, range, negated range, and logit rules – assuming the user has selected all rule types within a modeling experiment) for each model run to explain the relationship between locality data and environmental parameters. Once a rule-set (i.e., the combination of all 50 rules in each model run) is created, then the relationship is applied to other areas of the landscape that have similar environmental parameters. Each of the four rule types create IF/THEN statements that describe presence or absence parameters for the landscape (Stockwell & Peters 1999). The ability to use multiple rule types in an iterative process to create each rule-set establishes GARP as a super-set algorithm as opposed to many other modeling approaches that may only use range rules or logistic regression singularly. It is assumed that GARP predicts the fundamental niche of a species due to the lack of biological interactions not modeled,
but if the model is limited by a biased set of occurrence data (or poor fitting relationships between the occurrences and environmental variables) then only a realized portion of the niche is likely to be predicted. GARP can also use rules that were produced for a specific landscape and time period to project distributions onto another landscape or into a different time period based on available and relevant environmental coverage sets (Stockwell & Peters 1999).

GARP has been useful and successful in many studies across a range of taxa (Beard et al. 2002, Blackburn et al. 2007, Brotons et al. 2004, Feria & Peterson 2002, Wiley et al. 2003). It has also been utilized to evaluate the distribution of various disease vectors (Adjemian et al. 2006, Peterson et al. 2002c, Peterson & Shaw 2003, Peterson et al. 2003, Peterson et al. 2004, Peterson et al. 2005, Sweeney et al. 2006) and disease organisms (Aikimbayev et al. unpublished manuscript, Blackburn et al. 2007, Ron 2005). For example, the spatial distribution of various flea species implicated as vectors for plague was effectively modeled in California using GARP in conjunction with limited spatial data for many of the flea species (Adjemian et al. 2006). GARP was also utilized to predict the spatial distribution of triatomine insects in southern Texas in an effort to understand the domestic transmission cycle of chagas disease (Beard et al. 2002). Peterson et al. (2002c) also utilized GARP to model potential reservoirs for chagas disease in Mexico. Additionally, GARP has been used in previous efforts to model disease organisms such as *B. anthracis* (Aikimbayev et al. unpublished manuscript, Blackburn et al. 2007) and the amphibian pathogen *Batrachochytrium dendrobatidis* (Ron 2005).
Potential changes in the future distribution of disease vectors based on climate change scenarios have also been investigated. A study in southern Brazil explored the potential effects of climate change on vectors for cutaneous leishmaniasis using GARP (Peterson & Shaw 2003), while another study examined the potential effects of climate change on the reservoirs for tularemia and plague (Nakazawa et al. 2007).

Many studies have utilized GARP to predict the ecological and geographical distributions of a species or multiple species (e.g., Anderson et al. 2002, Elith et al. 2006, Parra-Olea et al. 2005, Peterson 2001, Peterson et al. 2007), however very few have analyzed the various IF/THEN rule-set statements that are first created by GARP and later projected on to the landscape (Blackburn et al. 2007, McNyset 2005, Wiley et al. 2003). For this reason, GARP is sometimes referred to as a “black box” because of the seemingly inexplicable method that it uses to predict the environmental parameters that support the survival of a species (Stockman et al. 2006). Stockman et al. (2006) stated that “there is no way to analyze the respective contributions of individual predictor variables to the model” in reference to the GARP modeling outputs. Another study that examined multiple modeling approaches determined that GARP “performed poorly” partly because of low accuracy scores and partly because of a conflation of species distribution modeling and ecological niche modeling in the study (Elith et al. 2006). Elith et al. (2006) examined the realized distribution of multiple species on the landscape and therefore constrained the GARP predictions through methodology to perform accordingly.

GARP is a superset algorithm meaning that it has the ability to use multiple rule-types (atomic, range, negated range, and/or logit) in each rule-set (Stockwell & Peters
GARP does not modify each subsequent rule-set based on previous rule-sets, but rather uses an iterative, stochastic process to create rule-sets that produce ecological relationships that match known locality data (i.e., presence-data) (Stockwell & Peters 1999). The rule-set writing and mapping application available within Desktop GARP v. 1.1.3 writes each rule-set to a text file, and then projects each rule-set into geographic space. This feature of GARP provides an opportunity to evaluate the biological information contained in the modeling process. Unlike other efforts (Blackburn 2006, Levine et al. 2007, Peterson et al. 2004, Ron 2005) that attempt to describe only the variable space in which a species exists independent of the modeling process, this thesis will examine spatial patterns in variable space where a species (*B. anthracis*) exists based on the output of the modeling process. Because variable space is examined after the modeling process, the location on the landscape of where rules are applied can also be examined in variable space to identify differences and similarities. The rules will also be examined to determine if a massive list of truly heterogeneous rules were produced or if the rules were conserved within and between models. Blackburn (2006) showed that GARP rule-sets are conservative between individual models within a best subset of models used to define a prediction of *B. anthracis*, with US models requiring only a few rules per model to capture the majority of predicted presence or absence on the landscape.

**Thesis Goals**

Specifically, this thesis focused on several key research questions. First, the current spatial distribution of known anthrax outbreaks may help determine the environment that supports *B. anthracis*. Through the study of anthrax outbreak locations, the potential distribution of *B. anthracis* in Kazakhstan may be able to be
predicted. Subsequently, if the current potential distribution of \textit{B. anthracis} can be predicted, then the future potential distribution of \textit{B. anthracis} may be able to be determined based on various climate change scenarios. It is imperative to predict the future distribution of \textit{B. anthracis} in order to better inform the public health system of the potential movement and expansion/contraction of areas where anthrax outbreaks are most likely to occur. Next, the focus will shift to the underlying environmental parameters that support \textit{B. anthracis} survival which will be examined through the rule-set writing and mapping component of Desktop GARP v. 1.1.3. In general, \textit{B. anthracis} will be used as a case study to outline the utility of extracting important biological information from the modeling process in GARP and to evaluate variable ranges across rules and models to determine if variable ranges are being conserved.

Answering questions about where a disease exists and why is often the centerpiece of medical geography research. Finally, the necessity of a geographical perspective for the many problems introduced by the following questions outline the geography-centric nature of the thesis. The thesis seeks to answer four main questions: 1) what is the current distribution of \textit{B. anthracis} in Kazakhstan? 2) what will be the future distribution of \textit{B. anthracis} in Kazakhstan? 3) what are current environmental parameter combinations (or rule-set combinations) that describe the current distribution of \textit{B. anthracis} across Kazakhstan? and 4) how useful is the rule-set writing and mapping application of GARP in providing important biological information about a species?

(Refer to Figure 1-1 to view a map of the study area: Kazakhstan)
Figure 1-1. Political map of Kazakhstan with oblasts (equivalent to a province or state), topography, major cities, rivers, and lakes.
CHAPTER 2
MODELING THE POTENTIAL DISTRIBUTION OF BACILLUS ANTHRACIS UNDER MULTIPLE CLIMATE CHANGE SCENARIOS FOR KAZAKHSTAN

Introduction

Bacillus anthracis is a spore-forming bacterium that is endemic to specific soil environments and the causative organism for anthrax, an infectious disease primarily found in herbivorous wildlife and livestock species, and secondarily in humans (Van Ness 1971). Limited data are available to define the geographic extent of environmental variables that support long-term B. anthracis survival, but current literature suggests that B. anthracis likely replicates in the animal host and can then survive for long periods in specific soil environments (Gainer & Saunders 1989, Kaufmann 1990, Smith et al. 1999, 2000). However, new evidence on the potential role of bacteriophages and soil-dwelling invertebrates (e.g. worms) suggests a more complicated life cycle for B. anthracis in soil that may or may not require a mammalian host for multiplication and may provide an alternative to a spore-only survival mechanism in soil (Schuch & Fischetti 2009). In either case, it is plausible that these scenarios require similar soil conditions to those described for “spore survival” in the earlier literature. Hugh-Jones & Blackburn (2009) summarize the general soil conditions for B. anthracis survival from a large body of literature as humus-rich, alkaline soils with pH >6.0 and distributed across the steppe and grassland soils.

Until recently, knowledge concerning the distribution of these environments was limited to studies that focused primarily on the distribution of B. anthracis in North America (e.g., Blackburn et al. 2007, Dragon et al. 1999, Van Ness 1971) and parts of Africa (e.g., Smith et al. 2000), but a recent study in Kazakhstan revealed some of the environmental constraints of B. anthracis on the landscape (Aikimbayev et al.)
unpublished manuscript). A second study (Aikimbayev et al. 2010) confirmed that the majority of anthrax cases in Kazakhstan over the last century affected both large (cattle) and small ruminants (sheep and goats). Human anthrax cases in Kazakhstan are primarily caused by exposure to infected animals – usually cattle, sheep, horses, or goats (Woods et al. 2004). Anthrax cases were predominantly cutaneous infections and were most often linked directly to the slaughtering and/or butchering of infected animals and no reports of human to human transmission occurred in the study. People in rural environments were more commonly infected because of a lifestyle that was more involved with livestock management/production. Additionally, insufficient vaccination efforts (lack of access, availability, surveillance, etc.) were the main reason for infection in Kazakhstan and the surrounding central Asian countries (Woods et al. 2004). Since exposure to livestock is a major source of anthrax infections in humans, it is also important to consider the factors that help to regulate domestic livestock numbers. One such study (Robinson & Milner-Gulland 2003) examined factors that regulated domestic livestock numbers over the past century in Kazakhstan and determined that the timing and amount of precipitation are the most crucial factors because these factors affect the amount of vegetation available.

Recent studies have attempted to understand the geographic distribution of *B. anthracis* and anthrax outbreaks in Kazakhstan by employing Geographic Information Systems (GIS), spatial analysis, molecular genotyping techniques (Aikimbayev et al. 2010) and spatial statistics and ecological niche modeling (ENM; Aikimbayev et al. unpublished manuscript). Ecological niche modeling has often been used to model a species’ ecological and geographic distribution. Many different ENM approaches have
been utilized for various studies including the presence-absence approach and the presence-only modeling approach (Brotons et al. 2004). The presence-absence modeling approach requires that presence and absence locality data be provided in order to model the ecological niche of a species. Absence data, however, are often difficult to validate because many areas that may be classified as being absent of a certain species may, in actuality, provide a suitable habitat (Pearce & Boyce 2006). In some situations, a species may not have been observed in an area where it actually does exist. For example, sampling gear biases may limit the successful capture of live specimens (Carlson & Cortes 2003, Remson & Good 1996) or sampling efforts may not exhaustively search all possible areas within the species’ range. In the case of pathogen-based studies, proper diagnostics, test sensitivity, and detection thresholds must all be considered when defining the causative agent as present or absent.

The presence-only modeling approach requires locality data to create a predicted geographic distribution of a species based on environmental parameters that exist where the species is confirmed to be present (Pearce & Boyce 2006). Pseudo-absence data are often generated in this approach to determine areas that do not match the environmental parameters of areas that are known to be present for a particular species (Pearce & Boyce 2006). The presence-only ENM approach has been successfully employed to model the potential geographic distribution of a number of taxa (Parra-Olea et al. 2005, Peterson et al. 2002a, Peterson et al. 2002b, Terribile et al. 2009, Wiley et al. 2003), including disease vectors (Adjemian et al. 2006, Peterson et al. 2002c, Peterson & Shaw 2003, Peterson et al. 2003, Peterson et al. 2005, Sweeney et al. 2006) and disease organisms (Aikimbayev et al. unpublished manuscript, Blackburn et
An ENM constructs a definition of the niche of an individual species in ecological (variable) space and predicts its potential geographic distribution through the analysis of relationships between combinations of environmental variables (e.g., temperature, precipitation, and elevation derived from digital maps or satellite data) and species' locality data (Blackburn et al. 2007).

The ecological niche can be defined as those environmental conditions that allow a species to maintain its population without immigration (Grinell 1917, 1924). That definition was later expanded to state that the presence of a species is correlated to quantifiable environmental and biotic variables that promote its survival, or a region in multi-dimensional space that describes states of the environmental variables which are suitable for the species to exist (i.e. a hypervolume of parameters; Hutchinson 1957).

The complexity of intra- and inter-specific interactions was recognized and niche space was consequently sub-divided into a fundamental niche (maximum extent of environment that can sustain its population) and a realized niche (actual environment that a species inhabits). Theoretically, a species often cannot inhabit its entire fundamental niche because of disturbance (e.g., habitat fragmentation; Parra-Olea et al. 2005), inter-specific competition (MacArthur 1972), or intra-specific limits (e.g. vagility, reproductive success; Peterson & Cohoon 1999).

An ENM known as the Genetic Algorithm for Rule-set Prediction (GARP), that can be broadly defined as a fundamental niche modeling approach (Soberon & Peterson 2005), was recently used to examine the geographic distribution of *B. anthracis* in the United States (US) under current (Blackburn et al. 2007) and future ecological conditions (Blackburn 2010). Another study from Kazakhstan also used GARP to model
the potential geographic distribution of environments that likely support long-term persistence of \textit{B. anthracis} and confirmed that repeat livestock anthrax epizootics occur within that predicted geographic range of the organism (Aikimbayev et al. unpublished manuscript). In that study it was predicted that the northern and southeastern regions of Kazakhstan may provide a suitable habitat for \textit{B. anthracis} survival, while the interior and western regions of the country are potentially unsuitable for \textit{B. anthracis}.

Recent work has advocated for the use of ENM as a method to provide improved surveillance strategies for anthrax across the United States (Blackburn et al. 2007). The same is true for Kazakhstan. The geographic potential of \textit{B. anthracis} covers a very large area in both countries, but vaccination in both cases is usually administered as a reactionary measure in response to outbreaks. However, knowledge of the distribution of \textit{B. anthracis} can allow for better monitoring and control measures in areas where the disease (or its causative agent) is predicted to be present (Blackburn et al. 2007). The use of ENM to model the current distribution of \textit{B. anthracis} in Kazakhstan also produced similar results intended to improve surveillance and target control strategies in an effort to be more proactive in the management of anthrax outbreaks in livestock (Aikimbayev et al. unpublished manuscript).

This current study employed ENM to examine the environments that support \textit{B. anthracis} across Kazakhstan and expands on previous efforts (Aikimbayev et al. 2010 & unpublished manuscript) by predicting the potential impact of climate change on the distribution of \textit{B. anthracis} across the landscape of Kazakhstan, as well as providing an expanded 8-variable niche definition under current environmental conditions. Previously, only a 5-variable niche definition that included total annual precipitation,
average annual temperature, elevation, average Normalized Difference Vegetation Index (NDVI), and NDVI amplitude, which measures the amount of seasonal variation of vegetation, was used to predict the current distribution of *B. anthracis* in Kazakhstan (Aikimbayev et al. unpublished manuscript). The 8-variable niche definition includes the 5 previous variables and adds temperature annual range, precipitation of wettest month, and precipitation of driest month (Figure 2-1). These variables were used because of previously discovered relationships between *B. anthracis* and specific parameters of precipitation, temperature, NDVI, and elevation (Blackburn 2006, Blackburn et al. 2007).

A major advantage of GARP (and other ENMs) is the ability to project the future distribution of a species based on its current relationship to environmental variables and the prediction of climate change that will occur over the geographical area inhabited by the species. The theory of ecological niche conservatism with respect to ENM helps to support this approach (Peterson et al. 1999). It states that a species maintains the same ecological niche over very long periods of time. This allows for the prediction of habitat change for a species based on future climate change scenarios (Atzmanstorfer et al. 2007, Holt et al. 2009, Huntley et al. 2004, Parra-Olea et al. 2005, Pearson et al. 2006, Peterson et al. 2001, Peterson 2003, Thuiller 2004). However, some uncertainty surrounds the prediction of a species’ future distribution (Davis et al. 1998, Holt et al. 2009, Pearson et al. 2006, Thuiller 2004). It has been argued appropriately that we have no means of determining the changing interactions between species because of climate change (Davis et al. 1998). However, Global Climate Models do provide some measures of confidence and speculation through the use of current and future bioclimatic data may help to plan for possible future changes in a species’ distribution.
Since the release of future climate/emissions scenarios by the Intergovernmental Panel on Climate Change (IPCC), many published studies have predicted future climate change patterns that may occur in central Asia over the next 50-100 years (e.g., Giorgi et al. 2001, Kimoto 2005, Lal & Harasawa 2001, Rosenzweig et al. 2008). Multiple studies have concluded that 1) an increase in annual precipitation over most of Asia with 2) an overall rise in temperatures that is most pronounced in the winter months has occurred over the past several decades (Alexander et al. 2006) and may continue to occur in the future (Baettig et al. 2007, Lal & Harasawa 2001). Annual, inter-annual, and decadal trends have also been studied recently to analyze the relationship between atmospheric forcing mechanisms (e.g., teleconnections) and recent Eurasian climate variability (Saito & Cohen 2003, Watanabe & Nitta 1998). The importance of snow cover extent changes and its possible role as an amplifier of regional atmospheric patterns has also been examined (Watanabe & Nitta 1998). Snow season lengths, snow depths, and annual snow accumulation variability have also been studied in coordination with global sea surface temperature (SST) variability, regional atmospheric changes (increased precipitation and increased temperatures overall), and regional atmospheric oscillation patterns over varying periods of time (Ye 2001, Ye et al. 1998, Ye & Ellison 2003). One study concluded that snow cover depth increased across northern Eurasia (>60°N latitude), while a decrease occurred in southern Eurasia (<60°N latitude) suggesting that there has been an increase in precipitation and temperatures across the region related to surface climate warming in the Arctic region (Ye & Ellison 2003). Another study examined recent changes of the onset date of green-up for portions of central Asia and determined that the steppe regions were highly
influenced by spring precipitation (Yu et al. 2003). A higher amount of precipitation in the spring has caused these regions to have earlier green-up dates than they had previously. Areas of the Mongolian steppe that had particular vegetation types and a higher level of spring soil moisture exhibited an overall trend of earlier green-up and an overall temperature increase was observed across much of the region as well as a warming trend at the beginning of the growing season. It is important to note that a significant part of interior Kazakhstan is primarily composed of the Kazakh steppe (refer to figure 2-2), which is an extension of the neighboring Mongolian steppe to the east. Because of the similarity and proximity of the steppe regions, the Kazakh steppe may also exhibit similar green-up patterns.

Other studies conducted at similar latitudes to Kazakhstan have examined the potential expansion and contraction of rangeland (i.e., grasslands used for the grazing of domestic cattle) and changes in phenological phases based on climate change (Baker et al. 1993, Bradley et al. 1999). One study concluded that the northern latitudes of the US rangeland would experience an increase in growing season and an increase in plant production as well as an increase in peak standing crop (Baker et al. 1993). An increase in forage across the northern latitudes resulted in less feed being needed to supplement the winter diet of cattle, potentially resulting in an increase in cattle numbers and an increase in calf weight (Baker et al. 1993). Models used in this study predicted substantial variation in yearly green-up periods indicating an increasing sporadicity related to climate change. Overall, both plant and animal production increased for the northern latitudes according to the study. In addition to being more productive in most locations, rangelands also were predicted to expand into previously more arid locations.
Changes in green-up and precipitation sporadicity in conjunction with rangeland expansion could indicate that some changes in the epidemiology of anthrax could occur such as longer anthrax seasons and an exposure of animals to more areas where *B. anthracis* may exist (Blackburn 2010). Because large anthrax epizootics often appear to occur after specific rain events (in association with overall hot, dry summer conditions; Parkinson et al. 2003, Turner et al. 1999), the increasingly sporadic rate of precipitation may also create some changes in the epidemiology of anthrax in the US as well as potentially in Kazakhstan. Changes in phenological phases that have occurred since the late 1930’s were also studied, and maximal increases in earliness of photosynthetic activity were observed for latitudes between 45º N and 65º N (Bradley et al. 1999). While many plants did experience an overall increase in earliness of photosynthetic activity related to climate change, some plants were unaffected because they were more regulated by photoperiods (Bradley et al. 1999).

Because anthrax remains a problem in livestock in the region and sometimes affects humans, further examination of the spatial ecology and geographic distribution of *B. anthracis* is imperative. Kazakhstan has limited veterinary services and predominantly rural agricultural practices, thus surveillance priorities should be dynamic and readily employed at any moment. The political boundary of Kazakhstan creates a larger amount of longitudinal change than latitudinal change and much of Kazakhstan lies within the upper mid-latitudes (Figure 2-2A). Based on a previous study at similar latitudes (Blackburn 2010) we expect that there will be an overall contraction of *B. anthracis* environments by 2050 in the US with slightly more habitat contraction occurring in the southern latitudes. NDVI is a satellite-derived indicator of vegetation
and measures of NDVI have been important limiting variables in recent studies (Aikimbayev et al. unpublished manuscript, Blackburn et al. 2007). We thus assessed the effect that these measures have on models by comparing two current predictions: one that used measures of NDVI and one that did not. Modeling efforts were also utilized to determine if predicted changes in precipitation and temperature can give an indication of potential changes to the geographic distribution of *B. anthracis* across the landscape. We also created a more robust 8-variable niche definition as opposed to the previously used 5-variable niche definition. The objective of this study is to determine the current and future potential geographic distributions of *B. anthracis* based on the Hadley Coupled Model version 3 climate predictions for 2045-2055 using multiple resolutions.

**Data and Methods**

**Anthrax Occurrence Data**

A database totaling 3,947 outbreaks was constructed from historical records between 1937 and 2006 archived at the Kazakh Science Center for Quarantine and Zoonotic Disease (KSCQZD), Almaty, Kazakhstan. Of those, 3,929 records represented outbreaks in livestock. A total of 1,790 individual locations were reported, with 805 of those reporting repeat outbreaks (Aikimbayev et al. 2010). Outbreak events in domesticated animals, large (cattle) and small (sheep and goats) ruminants, constituted the majority of the dataset. Following a previous ENM effort in Kazakhstan (Aikimbayev et al. unpublished manuscript), this study utilized data from 1960-2000 to most closely reflect the disease situation in the period after broad vaccination and control strategies had been introduced. A total of 1,181 outbreaks were reported in large ruminants and
1,303 outbreaks were reported in small ruminants across the database from 1960–2000 (Figure 2-2A).

A filtering technique was applied to these 2,484 outbreaks to create smaller datasets that contained only spatially unique points for each of two environmental dataset pixel resolutions, 8 and 55km$^2$, respectively (Figure 2-2B-C). Points were considered spatially “unique” when they did not occur within the same pixel. GARP utilizes a single point per grid cell to identify it as present for *B. anthracis*. *Presence* and *absence* are the only two categories that GARP uses to separate grid cells and the presence of more than one point in a grid cell could create inflated accuracy metrics if points from the same grid cells are used to test whether or not GARP predicted a grid cell accurately. It would be the equivalent of using the same data for both the training and testing of a GARP model. Because GARP is a presence-only modeling approach, only species presence data are needed and pseudo-absences are generated from background areas where no species data occur (Stockwell & Peters 1999).

**Current and Future Climate Datasets**

There are four main emissions scenarios produced by the IPCC in its Special Report on Emissions Scenarios (SRES) and Third Assessment Report (IPCC 2000). The first is the A1 scenario which accounts for a low population growth, but very rapid economic growth and globalization. Less focus is placed on sustainability and energy efficiency in this scenario. The second scenario is the B1 scenario which accounts for the same low population growth, but development that is more focused on environmental sustainability and accountability. The third is the A2 scenario and it estimates a very rapid population growth due to less convergence of fertility rates (approximately 15 billion by 2055) and only minor improvements in emission standards
(increase of 1% of CO$_2$) over that same time period. The fourth scenario is the B2 scenario which estimates a smaller global population growth than A2 (approximately 10 billion by 2055), but a higher population growth than both the A1 and B1 scenarios with more improvements in emission standards (increase of 0.5% of CO$_2$) (Arnell 2004, IPCC 2000). We chose to use the HadCM3 (Hadley Coupled Model version 3) ensemble “a” versions of the A2 and B2 climate change scenarios for 2045-2055 (hereafter referred to as 2050) in order to evaluate the effects of both a conservative (B2) and a less conservative (A2) scenario of how climates may change over the next several decades. Other popular general circulation models (GCMs) such as the CGCM and CSIRO models use flux adjustments to offset and reduce significant climate drift, but it is most desirable to eliminate their use in the coupled models that we use for future climate simulations (Flato & Boer 2001, Gordon & O'Farrell 1997, Gordon et al. 2002). The HadCM3 model was chosen over other models because of its ability to produce a simulation without the use of flux adjustments (Collins et al. 2001, Johns et al. 2003).

Current and future climate grid data were freely downloadable (www.worldclim.org) on the WORLDCLIM website (Hijmans et al. 2005). The initial interpolation of the grids was scaled to a relatively coarse resolution (~111 km$^2$) before a thin-plate smoothing spline algorithm was applied to reduce the surfaces to various finer resolutions. Each resolution was validated multiple times against historical weather station data obtained from weather stations around the world to reduce error associated with interpolation (Hijmans et al. 2005). A resolution of 8 km$^2$ was utilized for this study because village latitude and longitude coordinates were occasionally
estimated to be greater than 1 km away from farms where anthrax outbreaks occurred. Current grids describing monthly precipitation values as well as maximum and minimum temperatures were available along with bioclimatic (BioClim) grids that were created through the manipulation of the aforementioned monthly variables in order to create more biologically meaningful variables that represent annual trends, seasonality, and extreme/limiting environmental factors (Hijmans et al. 2005). One apparent advantage of the WORLDCLIM dataset is the availability of BioClim variables which may be biologically more meaningful than annual mean, minimum, and maximum temperature and precipitation because they help to capture more specific climatological patterns.

Future grids (e.g., for 2050 A2 and B2 climate change scenarios) describing monthly maximum and minimum temperatures and precipitation totals were also available, but bioclimatic grids were not available for future scenarios. For this reason, bioclimatic grids were calculated for both the A2 and B2 climate change scenarios. Bioclimatic variables were derived for current and future conditions following calculations provided on the WORLDCLIM website (www.worldclim.org). The calculations were performed with the use of the raster calculator within the Spatial Analyst extension of ArcMap 9.2 (ESRI 2007). BioClim variables have been used in a recent study to develop current and future predictions of Yersinia pestis infected ground squirrels, Spermophilus beecheyi, in California using the approach described here (Holt et al. 2009). Additionally, two measures of NDVI which were calculated based on Advanced Very High Resolution Radiometer satellite-derived data were provided at a scale of 8km² by the Trypanosomiasis and Land Use in Africa (TALA) research group at Oxford University (Oxford, United Kingdom) (Hay et al. 2006). Once calculations were
complete, a total of eight world environmental variable grids were clipped to represent the spatial extent of Kazakhstan because all anthrax locality data were located within its political boundaries (Table 2-1).

Given that the native resolution of climate models is relatively crude, the accuracy of climate data resampled to a high spatial resolution is questionable (Nakazawa et al. 2007). To test for agreement between low and high resolution datasets, we constructed models using near-native resolution climate data directly from the IPCC at 55km$^2$. Without monthly data at low resolution, we did not calculate BioClim variables at 55km$^2$. To compare the resolution of 55km$^2$ and 8km$^2$, we used five variables to construct models at both resolutions: elevation, total annual precipitation, mean temperature, minimum annual temperature, and maximum annual temperature. A model using identical variables from the 8km$^2$ climate dataset was constructed in order to make a fair comparison between the two resolutions. Current and future climate grids were clipped and resampled to represent the spatial extent of Kazakhstan at this resolution.

**Modeling Scenarios**

For this study, four separate modeling scenarios were employed to examine the current geographic distribution of *B. anthracis*. **Current Scenario 1** and **Current Scenario 2** contained five non-bioclimatic environmental variables that described temperature, precipitation, and elevation. **Current Scenario 1** utilized the five variables at a resolution of 55km$^2$, while **Current Scenario 2** utilized the five variables at a resolution of 8km$^2$. The spatial resolution of climate datasets can be problematic when modeling the potential distribution of a species (Nakazawa et al. 2007) so we have utilized a resolution that is closer to the resolution used by continental climate models (i.e., 55km$^2$) to compare to predictions made using a finer resolution (i.e., 8km$^2$).
Current scenario 3 contained eight environmental variables describing temperature, precipitation, elevation, and measures of NDVI to construct a model of the potential current distribution; see Table 2-1. Measures of NDVI were derived through satellite data and were not obtainable for the future. To make a fair comparison of the current distribution of *B. anthracis* and the future distribution, identical environmental variables had to be used; therefore measures of NDVI were excluded from the fourth current distribution model. Current scenario 4 thus contained only six environmental variables that described temperature, precipitation, and elevation to create a model of the potential current distribution of *B. anthracis*. Current scenarios 3 and 4 were used to create a comparison between two models of current distribution that used different environmental variables. The exclusion of measures of NDVI by current scenario 4 was examined to determine the potential limiting ability of a vegetation measure on the predicted distribution of *B. anthracis*. Two models of the future distribution of *B. anthracis* were also created for current scenarios 1, 2, and 4. Temperature and precipitation trends predicted for 2050 by the A2 climate change scenario and B2 climate change scenario were used to construct the models and compare the future potential distributions to the coinciding current distribution.

**Implementation and Methodology of Desktop GARP and Accuracy Metrics**

The specific ENM chosen for this study was the Genetic Algorithm for Rule-set Prediction (GARP; Stockwell & Peters 1999). GARP is a presence-only genetic algorithm that models species’ potential geographic distributions through an iterative process of training and testing that occurs through resampling and replacement of input data (Stockwell & Peters 1999). A pattern matching process is applied that finds non-random relationships between species localities and specific variables that describe the
environment. These relationships are written as a series of if/then logic statements (known as rules) that define whether conditions within the rule are defining presence or absence. A GARP “model” is a combination of up to 50 rules that define the landscape as present or absent and the resulting rules are known as a rule-set. The rules consist of four specific types: range, negated range, atomic, and logistic regression (Stockwell & Peters 1999). GARP is genetic, meaning that rule development is done through an automated process, whereby rules are randomly generated, tested with internal statistical tests, and modified (through the rules of genetics – point mutations, crossovers, deletions, insertions, Stockwell & Peters 1999) to determine which rules to keep and delete based on their accuracy at predicting internal testing data. Data splits occur both internally and externally for the purpose of validation and are established by the user. A best subset of models is usually created during an experiment. A best subset is a group of a user-defined number of models from an experiment that meet omission and commission criteria established by the user as a means of selecting those models that best balance between low omission and median commission values (Anderson et al. 2003).

While GARP has received some criticism as a “black box” (Stockman et al. 2006), or being less precise than more recently developed tools (Elith et al. 2006), recent studies have shown GARP to perform well (McNyset & Blackburn 2006, Blackburn et al. 2007) and it should be noted that this criticism was in part due to evaluations based on an unequal calculation of the accuracy metric used (Peterson et al. 2007, 2008). Part of this confusion is also due to a conflation of ecological niche modeling and species
distribution modeling (Phillips & Dudik 2008). Here we employ the former, while the criticism (Elith et al. 2006) was concerned with the latter.

Spatially unique point data were randomly split once into 85% training and 15% testing data subsets prior to model development using SPSS (version 16.0) (Figure 2-2A). The same 85% training dataset was used within the model-building process for all models, while the 15% testing dataset was withheld completely from the modeling experiments to evaluate the predictive accuracy of the models post hoc. Ten more 85/15 random splits of the data were performed to illustrate the spatial consistency between random splits (Figure 2-3). Additionally, locality data in Kazakhstan were divided into northern and southern sections separated at 48°N latitude. Random 85/15 splits were performed on locality data in each section, then added back together to create training and testing datasets using this alternative partitioning approach (Figure 2-4). The consistency exhibited between each random subsetting method allows us to only use one random split for model-building and validation instead of multiple random splits or spatially-informed random splits. Maps were then created from the model to identify the potential geographic distribution of B. anthracis based on both the 8-variable and 6-variable niche definition. Because GARP is a two-step modeling process, first modeling in variable space and then projecting onto the landscape, it is plausible to project current rule-sets onto the potential future conditions of a landscape. This current study employed the Desktop GARP version 1.1.6 [DG] software application, an open source modeling program (http://www.nhm.ku.edu/desktopgarp/).

**Modeling Parameters**

For all modeling scenarios, the training data were uploaded into DG with a 50/50 internal data split, meaning that 50% of the data were used within GARP to construct
models and the remaining 50% were used for internal accuracy assessment of the rule-
set and model building process. We employed 200 modeling runs using a convergence
limit of .01 and 1000 max iterations using all four rule-types. The best subsets
procedure was implemented to select optimal models for *B. anthracis* using an extrinsic
omission measure and the selection of 20 models under a hard omission threshold of
10% and a commission threshold of 50%. This produces a 10-model best subset,
where the 10 models with an accuracy of 90% or greater and closest to the median
commission value are chosen to represent the potential geographic distribution. These
10 models were imported into ArcGIS and summated using the raster calculator routine
in the Spatial Analyst extension. These maps represent values between 0 and 10, with
0 equally “absent” and values of 1 through 10 representing the number of models from
the best subset that predicted that pixel as present; the greater the number, the higher
the confidence in the model outcome (Ron 2005). A total of four summated maps were
produced for this study. A map of *current scenario 2* was developed to compare against
*current scenario 1* and then two maps of the projected distribution (i.e., A2 and B2
climate change scenarios) were created to show the potential geographic distribution in
2050.

The accuracy of current scenarios 1 and 2 was then quantified through the use of
accuracy metrics, which utilized the 15% testing data that was withheld from both
modeling experiments. A receiver operating characteristic (ROC) analysis was used to
produce area under the curve (AUC) scores. Additionally, two measures of omission
(i.e., total and average), and two measures of commission (i.e., total and average) were
also calculated for the current distribution model output. An AUC score ranges from 0.5
(lowest predictive accuracy – completely random) to 1.0 (perfect score – points were predicted 100% of the time), but AUC measurements are not ideal for validating the accuracy of GARP because they are subject to an area effect (McNyset 2005, Peterson et al. 2008, Wiley et al. 2003). GARP usually only makes predictions across a small portion of the ROC plot, but AUC scores are measured across the entire area, not just the area predicted by GARP (Peterson et al. 2008). Consequently, ROC measurements should be regarded with caution. A recent study noted that the relative poorness of AUC scores is not a failure of GARP to predict an accurate distribution, but rather limitations of the statistics that are currently used to test model accuracy (McNyset 2005). To provide a more robust evaluation of the models we presented AUC scores but along with measures of omission and commission that were based on the 15% testing subset (McNyset 2005).

**Analysis of Habitat Change**

Summated maps from the best subset were reclassified to visualize the habitat changes that occurred between each of the current predicted distributions and future distributions. Grids for the current distributions and the projected A2 and B2 distributions were reclassified as presence (6 or more models agree) or absence (5 or fewer models agree). The six model agreement threshold was selected because this indicates that more than 50% of the models agreed on the areas predicted as present. The raster calculator was then used to subtract the projected distributions from the current distribution. In total, two maps were produced representing habitat change (i.e., habitat expansion, habitat loss, no habitat change, unsuitable environment) occurring for the A2 and B2 climate change scenarios at each resolution and modeling scenario.
The percentages of area occupied for each of the four categories of habitat change were tabulated.

Results

Accuracy Metrics

Accuracy metrics were performed only on the models of current distribution because the location of future outbreak events is unknown and therefore unavailable for validation. The modeling processes for each of the four scenarios reached convergence of accuracy (0.01) prior to the maximum iteration setting of 1,000 meaning that less than a 0.01 increase in accuracy occurred between successive rules. Current scenario 1 received an AUC score of 0.7045 from the ROC analysis and was significantly different from a line of no information (p < 0.01). The model had a total omission of 0.0% and average omission of 5.5% meaning that 100.0% of the independent (testing) locality data were predicted correctly by at least one model and 94.5% of the independent locality data were predicted correctly by all models in the best subset. Current scenario 2 received an AUC score of 0.6502 (p < 0.01). The model had a total omission of 5.1% and average omission of 10.2%. Current scenario 3 had an AUC score of 0.7312 (p < 0.01). The model had a total omission of 2.6% and average omission of 7.3%. Current scenario 4 received an AUC score of 0.6995 (p < 0.01). The model had a total omission of 5.1% and average omission of 10.0%. All accuracy metrics for the current predictions are summarized in Table 2-2.

Current and Future Distributions of B. anthracis

Current and future climate grid data were examined at the near-native resolution to verify if broad agreement occurred between 55km² outputs and the higher resolution 8km² climate data using non-bioclimatic variables. At the 55km² resolution (Current
Scenario 1) areas of northern and southeastern Kazakhstan were predicted to be suitable for *B. anthracis* survival, while the A2 (drastic population increase and 1% increase in CO$_2$) and B2 (average population increase and 0.5% increase in CO$_2$) climate change scenarios predicted smaller geographic distributions in southeastern Kazakhstan as well as slightly smaller geographic distributions in interior and western Kazakhstan (Figure 2-5). Overall the predicted current distribution of *B. anthracis* stretches across the northern tier, eastern quarter, and southeastern regions of Kazakhstan. It is predicted that these areas are potentially maintaining suitable environments for *B. anthracis*. The northern predictions follow a line of latitude approximately 48° N from West Kazakhstan to the eastern area of the Karaganda oblast near Lake Balkhash where the predictions then extend southward to the oblast of Aktobe. Model agreement decreases south of 48° N latitude in the southern half of the Karaganda oblast where no model predicts suitable habitat for *B. anthracis*. From eastern Karaganda oblast, habitat suitability expands farther to the south to encompass the eastern oblasts of Kazakhstan including nearly all of the Pavlodar, Almaty, and East Kazakhstan oblasts with slightly less suitability in the higher altitudes of the Altay Mountains in far eastern East Kazakhstan and the Tian Shan Mountains in the southern and southeastern regions of the Almaty oblast. The southern half of the Zhambyl and South Kazakhstan oblasts are also areas of high suitability with less model agreement in the north closer to their borders with the Karaganda oblasts and the Kazakh Steppe. Only the extreme southeastern areas of the Kyzylorda oblast provide potentially suitable habitat for *B. anthracis* while areas in the Kazakh Steppe and around both the Aral and Caspian Seas are not predicted to support *B. anthracis*. When considering the A2 and
B2 climate change scenarios, a noticeable change occurs in many areas of Kazakhstan including parts of West Kazakhstan and Aktobe where a suitable environment for spore survival recedes to only the northern-most reaches of each oblast. While the southern half of Kostanay exhibits a contracting suitable environment, the northern half of the oblast and most of Akmola, North Kazakhstan, and Pavlodar, which border Siberian Russia, retain a suitable environment for *B. anthracis* spore survival. Contraction also occurs in the southern areas of Almaty, Zhambyl, and South Kazakhstan bordering Kyrgyzstan, China, and Uzbekistan. The predicted changes were more easily discernible in Figure 2-6 where areas of predicted habitat expansion and contraction were delineated for each climate change scenario and the percentages of habitat change were summarized in Table 2-3.

At the 8km$^2$ resolution (Current Scenario 2) areas of northern and southeastern Kazakhstan were predicted to be suitable for *B. anthracis* survival, while the A2 climate change scenario predicted a smaller geographic distribution in southeastern and eastern Kazakhstan and the B2 climate change scenario predicted a smaller geographic distribution in southeastern, northeastern, and central Kazakhstan (Figure 2-7, Figure 2-8). The models suggest that there are significant areas of southeastern and northwestern Kazakhstan where a suitable environment for *B. anthracis* will cease to exist, while most of the habitat will remain intact across the northern tier with marginal habitat losses closer to the interior of the country. Northeastern Kazakhstan may also experience drastic habitat loss, but only the B2 scenario predicts this response. The oblasts of West Kazakhstan, Aktobe, Almaty, Zhambyl, and South Kazakhstan could lose nearly all areas that were previously predicted to be suitable habitats for *B.*
*B. anthracis* under current climatic conditions. There are also several very small areas of expanded habitat scattered across portions of interior and eastern Kazakhstan in Karaganda, East Kazakhstan, and Almaty. The percentages of expanded habitat, unchanged habitat, unsuitable habitat, and contracted habitat occurring across Kazakhstan for each climate change scenario at each resolution were summarized in Table 2-3.

Current geographic distributions were also produced using BioClim variables from WORLDCLIM (www.worldclim.org) (Current Scenarios 3 and 4). The predicted geographic distribution for *B. anthracis* in Kazakhstan using current scenario 3 (which included measures of NDVI) is shown in Figure 2-9a, while the predicted geographic distribution produced by current scenario 4 is shown in Figure 2-9b. The two models of current distribution were then combined to better determine differences and similarities between each of the two predicted current distributions of *B. anthracis* (Figure 2-9c). Current scenario 3 predicted potentially more suitability across portions of the southern extent of the northern predictions, while current scenario 4 predicted potentially more suitability on the northern fringe of the southeastern predictions. The models show minor differences across the landscape, but overall the predicted current distribution of *B. anthracis* stretches across the northern tier, eastern quarter, and southeastern regions of Kazakhstan in both models.

Future potential distributions were also created using Current Scenario 4. Areas of northern and southeastern Kazakhstan were predicted to be currently suitable for *B. anthracis* survival, while the A2 climate change scenario predicted a smaller geographic distribution in southeastern and eastern Kazakhstan and the B2 climate change
scenario predicted a smaller geographic distribution in southeastern, northeastern, and central Kazakhstan (Figure 2-10, Figure 2-11). The environmental parameters that allow for *B. anthracis* survival occur in only the northern-most section of West Kazakhstan and Aktobe in 2050 according to the B2 climate change scenario. Much of Akmola, Pavlodar, and East Kazakhstan are predicted to no longer maintain environments suitable for *B. anthracis*. A smaller geographic distribution is also predicted for the southeastern oblasts of Kazakhstan. The environments of interior Kazakhstan remain unsuitable for *B. anthracis* under the B2 scenario.

**Discussion**

The similarity in accuracy metrics for each of the current scenarios indicates that GARP successfully predicted actual outbreak locations withheld from the model-building process. Very low total and average omission scores for each scenario indicate a high predictive accuracy for each best subset presented. Additionally, an evaluation of individual test locations that were omitted in any of the current modeling scenarios shows that at least some of those are in areas unlikely to support *B. anthracis* in soils anyway based on the low frequency of such cases in a rather extensive time series of anthrax outbreaks. AUC scores were also reasonable for each scenario suggesting that our models are significantly better than random at identifying *B. anthracis* environments. As AUC directly reflects the relationship between omission and commission rates in its calculation (McNyset 2005), current scenario 3 performed best of all in this study. While current scenario 4 had a higher AUC than current scenario 2, it also predicted a smaller geographic extent of presence, so we would expect the AUC score to be higher. Given that both had equal total and average omission rates, it is unrealistic to consider any significant difference in performance of these two scenarios overall. While future
changes in the distribution of *B. anthracis* are purely speculative, current models appear to be accurate regardless of resolution and climate datasets for 2050 show a broad level of overall agreement with habitat expansion in the north and contraction in the south. From this, it is arguable that *B. anthracis* has established a natural ecology across many regions of Kazakhstan, primarily the northern half, eastern quarter, and southeastern regions along the borders with Uzbekistan, Kyrgyzstan, and China.

A comparison of current scenarios 3 and 4 showed that measures of NDVI may be an important limiting variable in modeling the distribution of *B. anthracis*, as was suggested in a previous study in the US (Blackburn et al. 2007). Current scenario 3 predicted a smaller area of potential habitat on the landscape of Kazakhstan than did current scenario 4 (38.21% of the landscape compared to 46.88% of the landscape). Current scenario 3 also received a slightly higher AUC score than scenario two (refer to Table 2-2). A study in the US also assumed that soil moisture and pH would not change substantially between the present day and 2050 and thus used them as variables when projecting future distributions of *B. anthracis* in the continental United States using the B2 climate change scenario (Blackburn 2010). Slower responses in soil content to climate change have also been noted in similar latitudes of Europe (Emmett et al. 2004) supporting the argument that a lack of soils data in Kazakhstan may lead to overly liberal predictions of habitat loss and/or gain. The use of soils data has been shown to create a more conservative estimate of habitat change that may occur by 2050 in the US (Blackburn 2010). However, it is important to note that those estimates may also be related to the geographic setting of the US and the effects of global increases of CO$_2$ and temperature to that region.
To examine the rate that soil conditions may change based on future climate change scenarios, Emmett et al. (2004) specifically analyzed changing soil conditions based on changing temperature and precipitation that were predicted by climate change scenarios. The study determined that a very gradual change in soil content could occur because of an increase in sporadic climate events (i.e., more wet/dry years than “normal” years). An overall increase in “dry” or “wet” years may cause some soil content changes (because of changes in vegetation growth and type), but since climate change often manifest itself with an increase in abnormal or extreme climatic events (or seasons), the soil itself may change at a more gradual pace because one “extreme” year may offset or balance the changes that may have occurred in the previous “extreme” year (Emmett et al. 2004). This suggests that the use of soil variables in future US models was likely an ideal approach (Blackburn 2010). However, Blackburn (In Press) employed the high resolution STATSGO soils database (http://soils.usda.gov/survey/geography/statsgo/) to generate soils variables for the US, for which no contemporary data set was available for Kazakhstan.

While future changes in the distribution of *B. anthracis* are speculative, current models appear to make similar predictions regardless of resolution and all climate datasets for 2050 show a similar level of overall agreement. South-central and southeast regions of Kazakhstan that are now considered suitable environments for *B. anthracis* (and where a significant group of anthrax outbreaks have occurred over the past 70 + years (Aikimbayev et al. 2010 & unpublished manuscript) may no longer have environmental conditions that support the long-term survival of *B. anthracis* according to projections from the A2 and B2 climate change scenarios at both resolutions.
A comparison between 55km$^2$ and 8km$^2$ climate data found that there was broad agreement across modeling experiments for the northern regions of Kazakhstan for the A2 climate change scenario. The southern areas of the Almaty, Zhambyl, and South Kazakhstan oblasts were predicted to experience drastic habitat loss (i.e., near total) at both resolutions, but drastic habitat loss in northern Kazakhstan was only predicted by the B2 climate change scenario at a resolution of 8km$^2$ (both current scenario 2 & 4). The actual reasons for major differences in the predicted distribution by each resolution are uncertain, but a lack of data points, a relatively steep change in elevation, the calculation of bioclimatic variables, and/or the splining technique used to downscale WORLDCLIM data may be possible explanations. There is still a great amount of uncertainty in future climate predictions even at a crude resolution and all future estimates should be regarded with caution. More guidance from climatologists in selecting climate datasets is warranted when considering how various climatic or bioclimatic variables may affect the potential distribution of a species.

Currently, much anthrax surveillance is focused on the south-central and southeast regions of Kazakhstan because many anthrax cases have occurred there in an area of high human population density, i.e. observation bias. Based on future bioclimatic data alone there may be a reduction in anthrax cases reported for this region. Future changes in temperature and precipitation may also cause geographic contraction of rangeland in the southern regions where livestock currently graze, while causing geographic expansion of rangeland in the northern regions. This would subsequently allow more animals to graze in environments that are predicted to be suitable for $B.\ anthracis$ in the north, while less grazing in the south in conjunction with a
less suitable environment for *B. anthracis* may also lead to further reduction in epizootics for this region. While climatic conditions may have changed between 1960 and 2000, current climatic variables were averaged between 1960 and 1990 thus we assume that locality data collected over the past several decades accurately reflect environmental parameters needed for *B. anthracis* presence on the landscape.

Overall, the hypothesis of predicted habitat loss in the south, but gain in the north was partially disproven. While a very small area of expanded habitat was consistently predicted in the northeastern regions of Kazakhstan, habitat loss was predicted in nearly every part of the country except the extreme northern regions bordering Russia. There was far more predicted habitat contraction in the southern regions of Kazakhstan than anticipated. Projected changes may reflect over-predictions of future habitat loss due to a lack of soils data, but nonetheless the southeast region should expect to observe some reduction in *B. anthracis* habitat.

The results of this current study agree with the results of similar continental scale studies where southern habitat reduction was also predicted due to the potential effects of climate change on other bacterial zoonoses (Blackburn 2010, Holt et al. 2009, Nakazawa et al. 2007) and we have documented this pattern in all four climate datasets used at both 55km² and 8km² resolutions. In the US, parts of the southern range of *B. anthracis* were predicted to contract by 2050, while some parts of the northern range were predicted to expand (Blackburn 2010). Nakazawa et al. (2007) investigated the effects of climate change on tularemia and plague in the US with ENM and multiple climate change scenarios and predicted similar trends with more contraction occurring in the southern habitats than in the northern habitats for 2050. Similarly, a recent study
that modeled the future distribution of plague-carrying ground squirrels in California using 1km$^2$ BioClim variables suggested a subtle geographic shift to higher latitudes and altitudes with a limited reduction at lower latitudes (Holt et al. 2009). Collectively, these trends were not as drastic as the trends predicted for Kazakhstan, but contraction of a southern range was suggested for all three diseases. The more extreme changes in predicted distribution for Kazakhstan may be a result of the region potentially experiencing a more severe climatic change between now and 2050. However, it is not implausible that variables, such as soil conditions that were unavailable for this study, might limit the habitat reduction to smaller portions of the Kazakh landscape.

Research over the past several decades has indicated that sporadic vegetation growth occurred from year to year based on rainfall amounts in the desert and steppe regions of Kazakhstan (Robinson & Milner-Gulland 2003). This may infer that an increase in rainfall variability (as predicted in the region of central Asia by climate change scenarios) from year to year in desert and semi-arid steppe climates could equate to a more sporadic occurrence of anthrax outbreaks. While models may have predicted a complete disappearance of habitat for *B. anthracis* in certain regions, anthrax outbreaks may simply become increasingly sporadic, but not disappear altogether in these regions as the A2 and B2 climate change scenarios suggested. Changes in the landscape could limit (if desertification occurs) or increase (if an increase in rangeland occurs) the ability for cattle to migrate (Robinson & Milner-Gulland 2003). These potential changes in migratory patterns could help to spread or limit the range of anthrax outbreaks and subsequent *B. anthracis* introduction and survival. Cattle migration is already confined because of limitations placed on nomadic
herdsmen over the past century (Robinson & Milner-Gulland 2003). Overall, cattle now graze on smaller areas than they did previously (Robinson & Milner-Gulland 2003) and in areas where outbreaks have occurred, we would expect a possible increase in outbreak potential if population densities are high (Dobson 2004).

The current spatial distribution of *B. anthracis* follows similar latitudinal patterns as those predicted by a study in the United States with larger areas of the northern regions predicted to be endemic for *B. anthracis* compared to smaller areas predicted to be endemic for *B. anthracis* in the southern region (Blackburn et al. 2007). This also closely follows the predicted current distribution of *B. anthracis* on the landscape of Kazakhstan (Aikimbayev et al. unpublished manuscript). The predicted areas of southern Kazakhstan traverse the foothills and mountain ranges of the Tian Shan and Altay Mountains, which have climates that are somewhat comparable to climates farther north. In maps of the projected distribution, it can also be determined that the suitable environments for *B. anthracis* (specifically in the southern regions) may move to areas of higher elevation greatly limiting its dispersal based on cattle grazing limitations (Robinson & Milner-Gulland 2003). Sheep, however, may not have similar grazing limitations because they are often transported either by foot or by truck/train to summer grazing areas in more mountainous regions (Wilson 1997). Because of their mobility, sheep may be able to adapt to climate changes in the south more so than cattle and may subsequently remain in environments that continue to be suitable for *B. anthracis*. Rainfall has dictated livestock numbers and migratory patterns over the past several decades (Robinson & Milner-Gulland 2003) so this could in turn limit the contact that cattle may have with an environment where *B. anthracis* exist in the soil. The opposite
may also be true if rainfall increases across many parts of Kazakhstan, more land could be available for grazing (similar to increases in forage in the northern latitudes of the United States (Baker et al. 1993)) thus allowing livestock to possibly move to more areas where they could come in contact with *B. anthracis*. An inverse relationship could potentially be created based on rainfall estimates that allow for livestock range expansion and *B. anthracis* range contraction. It is also important to consider the differences between the climate of Kazakhstan (continental with minimal influence from oceans) and the climate of the United States (surrounded by the Atlantic and Pacific Oceans as well as the Gulf of Mexico) when comparing the distribution of *B. anthracis* across the landscape of each.

Potential changes in seasonal vegetation patterns should also be examined in conjunction with typical seasonal patterns of anthrax outbreaks to determine if these patterns may coincide. Anthrax has a distinct seasonality and is primarily a summertime (May – October in northern latitudes) disease in both wild and domestic ruminants that is usually associated with wet springs and hot, dry summers followed by a rain event (Dragon et al. 2001, Gates et al. 1995). The predicted rise in temperatures and potential for increasingly sporadic rain events across much of central Asia (Lal & Harasawa 2001) could lead to spatial and temporal changes in where and when anthrax outbreaks occur in Kazakhstan. Rangeland expansion and contraction as well as changes in rangeland production in Kazakhstan could lead to a higher population of livestock in the northern regions, where *B. anthracis* is predicted to remain in 2050, and subsequently a potentially greater number of anthrax outbreaks. A rise in temperatures in the southern regions of Kazakhstan could create an environment that *B. anthracis*
and/or livestock may not be able to survive in, thus potentially decreasing the number of anthrax outbreaks there. It has been shown in the US that areas supporting *B. anthracis* survival do overlap with livestock distributions, however they are not identical (Blackburn 2010). Livestock may graze in areas that are unsuitable for *B. anthracis* and likewise, *B. anthracis* may exist in areas that are either unsuitable or not used for livestock grazing.

It is also interesting to consider the possible evolutionary implications of these climate change scenarios. While the genetic understanding of *B. anthracis* in Kazakhstan is incomplete, recent efforts (Aikimbayev et al. 2010) have provided insights into the spatial distribution of Kazakh specific genotypes for the country. Employing the 8-primer Multiple Locus Variable number tandem repeat Analysis (MLVA)-typing developed by Keim et al. (2000), a recent study described 92 culture isolates from several historical outbreaks. The majority of these isolates belong to the A1.a genetic cluster and the majority of that diversity was located in the southern regions of Kazakhstan, predicted to no longer support *B. anthracis* in 2050 by all both resolutions and climate scenarios. This might suggest that a reduction in suitable habitats in southern Kazakhstan may also correspond with a reduction in genetic diversity, but it is difficult to estimate changes in diversity in the northern most extent of Kazakhstan, as no cultures were available for typing (Aikimbayev et al. 2010). However, six of the 92 isolates from the existing data set represented a distinct member of the A3b sublineage. Interestingly, the B2 scenarios derived from current scenarios 2 and 4 suggest the northeastern region where these strains were isolated will no longer support *B. anthracis* in 2050.
When comparing climate change scenarios at a resolution of 8km$^2$, more habitat loss was predicted by the B2 climate change scenario – supposedly the more conservative (or optimistic) of the two scenarios. The B2 scenario delineates that more habitat loss may occur in the northern interior areas of Kazakhstan as well as the northeastern areas of Kazakhstan. Conversely, several small areas in southeastern and northwestern Kazakhstan that were classified as areas of habitat loss actually are predicted to retain their habitats in the B2 climate change scenario. While variations in the predicted precipitation and temperature changes for 2050 may have been the main reasons for distributional differences seen between the A2 and B2 scenarios, GARP used a combination of variables to create rule-sets that determined the environmental parameters that support *B. anthracis*. For example, a warmer and wetter environment in the north may create a more suitable environment for *B. anthracis* survival, but a warmer and drier environment in the south may also create a more suitable environment for *B. anthracis* survival in previously uninhabitable areas (e.g. in the higher elevations of the Tian Shan Mountains). Previous studies allude to the importance of examining specific rules within GARP rule-sets to evaluate changing relationships between variables across the landscape (Blackburn et al. 2007, McNyset 2005) and variable combinations for this study should also be examined to further understand environmental constraints on the habitat of *B. anthracis*. Temperature and precipitation changes will not be uniform across the vast landscape of Kazakhstan. For this reason, the internal rule-sets need to be examined to determine which variables and combination of variables were most important in predicting the ecological niche of *B. anthracis*. A closer examination of individual variables and variable combinations
derived through rule-sets may also help to reveal the potential driving mechanism(s) of
the predicted habitat change for *B. anthracis* across many areas of Kazakhstan.

Population growth and urbanization may also alter future predictions, but land cover use
change may affect future predictions more if rangelands expand/contract in certain
areas. Based on trends during the past century, Kazakhstan is not expected to
experience drastic population growth or urbanization that would greatly modify future
predictions.
Table 2-1. Environmental variables used for Genetic Algorithm for Rule-set Prediction (GARP) models

<table>
<thead>
<tr>
<th>Environmental Variables</th>
<th>Name</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Mean Temperature</td>
<td>BIO1</td>
<td>WorldClim (<a href="http://www.worldclim.org">www.worldclim.org</a>)</td>
</tr>
<tr>
<td>Temperature Annual Range</td>
<td>BIO7</td>
<td>WorldClim (<a href="http://www.worldclim.org">www.worldclim.org</a>)</td>
</tr>
<tr>
<td>Annual Precipitation</td>
<td>BIO12</td>
<td>WorldClim (<a href="http://www.worldclim.org">www.worldclim.org</a>)</td>
</tr>
<tr>
<td>Precipitation of Wettest Month</td>
<td>BIO13</td>
<td>WorldClim (<a href="http://www.worldclim.org">www.worldclim.org</a>)</td>
</tr>
<tr>
<td>Precipitation of Driest Month</td>
<td>BIO14</td>
<td>WorldClim (<a href="http://www.worldclim.org">www.worldclim.org</a>)</td>
</tr>
<tr>
<td>Elevation (Altitude)</td>
<td>ALT</td>
<td>WorldClim (<a href="http://www.worldclim.org">www.worldclim.org</a>)</td>
</tr>
<tr>
<td>Mean NDVI</td>
<td>WD1014A0</td>
<td>TALA (Hay et al. 2006)</td>
</tr>
<tr>
<td>NDVI Annual Amplitude</td>
<td>WD1014A1</td>
<td>TALA (Hay et al. 2006)</td>
</tr>
</tbody>
</table>

Table 2-2. Accuracy Metrics for the current predicted distributions

<table>
<thead>
<tr>
<th>Metric</th>
<th>Scenario One</th>
<th>Scenario Two</th>
<th>Scenario Three</th>
<th>Scenario Four</th>
</tr>
</thead>
<tbody>
<tr>
<td>N to build models</td>
<td>125</td>
<td>218†</td>
<td>218†</td>
<td>218†</td>
</tr>
<tr>
<td>N to test models</td>
<td>22</td>
<td>39</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>Total Omission</td>
<td>0.0</td>
<td>5.1</td>
<td>2.6</td>
<td>5.1</td>
</tr>
<tr>
<td>Average Omission</td>
<td>5.5</td>
<td>10.2</td>
<td>7.3</td>
<td>10.0</td>
</tr>
<tr>
<td>Total Commission</td>
<td>50.27</td>
<td>51.71</td>
<td>40.67</td>
<td>35.91</td>
</tr>
<tr>
<td>Average Commission</td>
<td>59.59</td>
<td>62.33</td>
<td>53.31</td>
<td>53.44</td>
</tr>
<tr>
<td>AUC*</td>
<td>0.7045 (z=7.7§, SE=0.06)</td>
<td>0.6502 (z=9.8§, SE=0.05)</td>
<td>0.7312 (z=9.9§, SE=0.05)</td>
<td>0.6995 (z=9.0§, SE=0.05)</td>
</tr>
</tbody>
</table>

* AUC = area under curve
† N was divided into 50% training/50% testing at each model iteration
§ p < 0.001
Note: Independent data used for accuracy metrics appear in figure 2-2 (yellow points)

Table 2-3. A comparison of habitat change (%) between Special Report on Emissions Scenarios (SRES) A2 and B2 climate change scenarios

<table>
<thead>
<tr>
<th>Habitat Change</th>
<th>A2 Scenario (55km²)</th>
<th>B2 Scenario (55km²)</th>
<th>A2 Scenario (8km²)</th>
<th>B2 Scenario (8km²)</th>
<th>A2 Scenario (BioClim)</th>
<th>B2 Scenario (BioClim)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expanded Habitat</td>
<td>4.15%</td>
<td>3.63%</td>
<td>0.71%</td>
<td>0.89%</td>
<td>0.20%</td>
<td>0.52%</td>
</tr>
<tr>
<td>No Change</td>
<td>43.85%</td>
<td>44.67%</td>
<td>40.94%</td>
<td>29.28%</td>
<td>36.81%</td>
<td>22.54%</td>
</tr>
<tr>
<td>Not Suitable</td>
<td>37.04%</td>
<td>37.56%</td>
<td>36.12%</td>
<td>35.94%</td>
<td>46.72%</td>
<td>46.76%</td>
</tr>
<tr>
<td>Habitat Loss</td>
<td>14.96%</td>
<td>14.15%</td>
<td>22.22%</td>
<td>33.88%</td>
<td>16.27%</td>
<td>30.18%</td>
</tr>
</tbody>
</table>
Figure 2-1. Environmental variables used during model-building process
Figure 2-2. Map of Kazakhstan with anthrax locality data. Training data (green) were used to build models while independent data (yellow) were used to validate the accuracy of models. Inset A shows where all anthrax outbreaks occurred between 1960 and 2000. Inset B shows training and independent data used for building models at a resolution of 8km$^2$. Inset C shows training and independent data used for building models at a resolution of 55km$^2$. 
Figure 2-3. Ten 85/15 random subsets with corresponding model runs showing consistency between testing and training point locations as well as predicted distributions.
Random 85/15 subsets were created for the northern and southern areas of Kazakhstan (separated at 48°N) independently, then added together to construct ten total testing and training subsets. Consistency is shown between point locations and predicted distributions.
Figure 2-5. Current predicted distribution of *B. anthracis* using Current Scenario 1 (A) and future predicted distributions based on the A2 climate change scenario (B) and B2 climate change scenario (C).
Figure 2-6. Potential future habitat changes based on the A2 climate change scenario (A) and B2 climate change scenario (B) derived from Current Scenario 1. The differences between these climate change scenarios are shown in (C).
Figure 2-7. Current predicted distribution of *B. anthracis* using Current Scenario 2 (A) and future predicted distributions based on the A2 climate change scenario (B) and B2 climate change scenario (C).
Figure 2-8. Potential future habitat changes based on the A2 climate change scenario (A) and B2 climate change scenario (B) derived from Current Scenario 2. The differences between these climate change scenarios are shown in (C).
Figure 2-9. Current Scenario 3 (A) and Current Scenario 4 (B) show the current distribution of *B. anthracis* using different environmental variables. The difference between scenarios 3 and 4 is also examined (C).
Figure 2-10. Current predicted distribution of *B. anthracis* using Current Scenario 4 (A) and future predicted distributions based on the A2 climate change scenario (B) and B2 climate change scenario (C)
Figure 2-11. Potential future habitat changes based on the A2 climate change scenario (A) and B2 climate change scenario (B) derived from Current Scenario 4. The differences between these climate change scenarios are shown in (C).
CHAPTER 3
EVALUATING ENVIRONMENTAL PARAMETERS OF BACILLUS ANTHRACIS IN KAZAKHSTAN: AN EXAMINATION OF RULE-SET WRITING AND MAPPING WITHIN AN ECOLOGICAL NICHE MODELING TOOL

Introduction

Understanding a species’ environmental parameters is essential in identifying its ecological and geographic distribution across space (Brotons et al. 2004, Huntley et al. 2004, Parra-Olea et al. 2005, Peterson 2001). Environmental parameters are often studied individually when examining a species’ distribution, but a combination of environmental parameters often produces a more robust explanation of why a species is present in one place, but absent in another (Pearson et al. 2006, Soberon & Peterson 2005). Specifically, the ecological parameters of Bacillus anthracis, the spore-forming bacterium that causes anthrax, have been the focus of many studies (Aikimbayev et al. unpublished manuscript, Blackburn et al. 2007, Cherkasskiy et al. 1999, Joyner et al. 2010, Smith et al. 2000, Van Ness & Stein 1956). B. anthracis has a very specific set of environmental parameters that must be present in order for the organism to be endemic to a landscape (Cherkasskiy 1999, Dragon et al. 1999, Smith et al. 2000, Van Ness 1971, Van Ness & Stein 1956). Van Ness & Stein (1956) outlined favorable soils for anthrax and created one of the first deterministic spatial distributions of where B. anthracis was likely to exist in the United States (US). The study examined where soils exists that are favorable for B. anthracis survival and where anthrax cases had occurred historically. Areas that matched both criterions were considered to be at a higher risk for anthrax outbreaks. It has also been suggested through field study that different genetic strains of B. anthracis have different soil preferences (Smith et al. 1999, 2000). A study in Kruger National Park revealed that A and B strains of B. anthracis were
present in the same year and represented an overlap in their distributions and time of occurrence, but isolate B was generally found in soils with different calcium and pH levels than those coinciding with isolate A indicating unique environmental requirements for each strain (Smith et al. 2000). Blackburn et al. (2007) studied the distribution of *B. anthracis* in the US and suggested that a better understanding of the distribution of the bacteria would increase eradication and prevention efforts, while Aikimbayev et al. (unpublished manuscript) examined the distribution of *B. anthracis* on the landscape of Kazakhstan and conveyed similar conclusions.

Spores are often found in calcium-rich environments where they can be sustained between outbreaks (Gates et al. 1995, Hugh-Jones & de Vos 2002, Hugh-Jones & Blackburn 2009). Van Ness (1971) indicated that basic mollic or chernozemic soils provide an ideal environment for resilient anthrax spores to survive and replicate, while Dragon & Rennie (1995) indicated that climate also plays an important role in spore survival and anthrax outbreak potential. The rapidity of sporulation increases with increasing environmental temperature and spores have been known to survive for more than 20 years in the soil (Thappa & Karthikeyan 2001). The threat of infection is generally more serious during drought conditions when herds must graze on vegetation close to the ground thereby risking accidental ingestion. The vegetation may also be coarser during a drought causing cuts on lips thus making the animals more susceptible to infection (Thappa & Karthikeyan 2001). When abundant rainfall is preceded by a prolonged drought, it is suspected that spores may rise to the surface (Laforce 1994) and many studies in Canada have found that heavy rainfall events often preceded

Many species distribution modeling and spatial modeling techniques have been used to determine the geographic and ecological space where a species can exist. The Bioclimatic Prediction System (BIOCLIM) is one of the most simplistic tools used to create models of ecological niches (Nix 1986). It develops a model by intersecting the ranges where a species exist along environmental axes (e.g., annual temperature range of 39-51°C by total precipitation of 146-680mm by mean Normalized Difference Vegetation Index (NDVI) of 0.04-0.36, etc.). Regression-tree analyses, general linear models, and distance-based algorithms are also common approaches, but each of these approaches and BIOCLIM try to find a single rule or small set of rules that describe the niche of a species instead of a complex set of rules that may more closely resemble a species’ niche (Carpenter et al. 1993, Guisan & Zimmermann 2000, Thuiller et al. 2003). Another common algorithm used to identify a species distribution is the logistic regression model which is a type of generalized linear model that identifies the probability of presence or absence on a landscape by modeling it as a linear function of all possible explanatory environmental variables (Manel et al. 2001).

Ecological Niche Factor Analysis (ENFA) is an example of a modeling algorithm that uses presence-only data to quantify a realized portion of the niche of a species (Hirzel et al. 2002). ENFA uses ecogeographical variables that describe an entire study area and compares locality data to global values in order to determine where the species is most likely to be present on the landscape (Hirzel et al. 2002). Factors are extracted that analyze the distance between optimum conditions for the species and the
mean habitat within the study area as well as the ratio of ecological variance in mean habitat compared to that observed for the species (Basille et al. 2008).

A statistical approach to ecological modeling called discriminant analysis has also been explored (Rogers 2000, Rogers 2006). The approach encompasses a range of methods that develop rules for classifying previously unclassified species to groups that have been defined (Estrada-Peña & Thuiller 2008). The discriminant function uses presence and absence data to assign a species to a group by multiplying a vector of locality data by a vector of coefficients to produce a value that is used to place the species in a group. Discriminant function can also use abundance data if available. To identify areas of absence, random regions that are no closer than 0.5° and no farther than 10° from presence locations are sampled. Both linear and non-linear discriminant functions can be used. A study detailing the distribution of tsetse flies used a non-linear discriminant analysis which identified the covariance characteristics of species presence and absence (Rogers et al. 1996). During discriminant analysis, the corrected Akaike information criterion in conjunction with other criteria is used to identify subsequent variables to add to each model as a form of step-wise inclusion (Rogers 2006). This approach helps to identify how well the current model fits the data.

A habitat modeling method that has recently gained increased attention is the Bayesian modeling approach which can be employed in many platforms. The Bayesian approach has been used in some form for several decades (Williams et al. 1978). It is based on Bayes’ Theorem and combines frequencies of association between the presence of a species and values in each environmental dataset with pre-modeling probabilities of occurrence to estimate post-modeling probabilities of a species being
present on the landscape (Hepinstall & Sader 1997). A recent study used a Bayesian framework to create statistical models that provide details about a species’ niche and distribution as well as the effects of outside disturbances (Latimer et al. 2006). The study argued that this approach may minimize certain problems that have been found with single-level regression models that are more widely used by ecologists such as irregular sampling intensity and spatial dependence. The evolution of a Bayesian approach to species distribution modeling is ongoing, but may hold a promising future.

Maximum Entropy (MaxEnt) is a very common modeling application used today. MaxEnt estimates the potential distribution of a species by finding the distribution of maximum entropy (i.e., close to uniform) that is limited by the expected value of each feature matching its empirical average (Phillips et al. 2004). Basically, the goal of Maxent is to estimate the probability distribution of a species that is the most expansive while at the same time being constrained by actual species presence data. It is a presence-only modeling approach that uses entropy to generalize species locality data and it defines suitability by estimating a probability distribution over every pixel in the study area. A recent study by Phillips & Dudik (2008) inferred that the output produced by MaxEnt most closely resembles a realized niche model or species distribution model.

The Genetic Algorithm for Rule-set Prediction (GARP) is another popular modeling application and it is a fundamental niche model and therefore seeks to model the maximum potential distribution of a species, which consequently usually includes areas of unknown occurrence and areas where the species may not exist because of competition, human influence or other external influences (McNyset 2005, Peterson et al. 2008). GARP is a presence-only modeling tool that uses species locality data and
environmental coverage sets to determine ecological relationships through a process of IF/THEN rule types developed in an iterative, stochastic process. GARP is also considered to be a super-set algorithm because of its ability to use multiple rule-types (i.e., range, negated range, atomic, logit) to create potential geographic distributions as opposed to other modeling approaches that may only use range rules or logistic regression singularly.

Recent studies have employed a spatial version of Principle Components Analysis (PCA) where the probability of harbor porpoise occurrence in each grid cell was calculated by taking total eigen scores for each principal component and then dividing by its eigen value to visualize which areas were most likely to contain a suitable habitat (MacLeod et al. 2008, Mandleberg 2004). Each study used the PCA statistical approach which is normally, but not always, aspatial to predict the habitat suitability of harbor porpoises in Scotland.

The previously mentioned modeling approaches are intended to produce a geographic distribution of presence and absence of a species on a landscape. However, investigators are often interested in which environmental relationships are constructed within models and exploratory methods to visualize the ecological space where a species exist. It is often difficult to gain an understanding of the ecological space or biological information associated with the geographic prediction of presence and absence, thus it is often useful to use PCA to explore potential relationships between variable space and geography. Two studies have used PCA in conjunction with the GARP ecological niche modeling approach (Blackburn 2006, Ron 2005). Ron (2005) used PCA to construct an environmental envelope of where *Batrachochytrium*
*dendrobatidis* occurred in ecological space and to analyze the relative importance of certain variables. For example, principle component I in Ron (2005) was positively correlated with precipitation indicating that it was an important variable in the prediction of *B. dendrobatidis*. Blackburn (2006) used PCA to help visualize where within ecological space anthrax cases occurred in the US and to evaluate clustering of anthrax cases in ecological space.

Another method of examining niche models in ecological space is through the construction of variable clouds (Peterson et al. 2004). Niche models of the Ebola virus in areas of Africa were constructed using GARP and then the values of selected environmental variables where the virus was present were visualized in 2-dimensional space (Peterson et al. 2004). This method was highly useful in identifying the ecological space of a disease because when visualized in dimensions of precipitation and temperature, it was evident that Ebola was concentrated in hot, wet climates.

Jackknifing and bootstrapping are useful approaches that do not necessarily help to visualize the ecological space where a species exists, but help to refine or filter the number of environmental coverages on which predicted distributions are based. When used within GARP, a jackknife manipulation provides all possible examples of analysis with the subtraction of one coverage set for each model run to eliminate each coverage set systematically (Peterson & Cohoon 1999). For example, if ten coverages were used, then ten possible examples would be provided with nine coverages each. Likewise, each coverage is also included systematically in single-coverage analyses. A bootstrap manipulation is used in conjunction with jackknifing so that combinations of coverages can be analyzed through a process of sampling with replacement (Peterson
et al. 2008). Levine et al. (2009) examined the statistical significance of ecological parameters by comparing individual maps that were constructed using the jackknife process. To analyze the relative importance of each ecological parameter in the development of the model, the individual jackknife maps were compared to the comprehensive map pixel-by-pixel and the study found that this statistical technique had a capacity to explore subtle differences among ecological parameters as well as extremes in the importance of individual environmental factors. Another method used to determine the importance of individual variables used within a model-building process extracts results from bootstrap models and sorts them according to their Akaike information criterion values (Rogers 2006). Variables are plotted and color-coded according to their importance as a predictor variable and the most consistently chosen variable(s) in each model are recognizable by a continuous, or near continuous, line.

The examination of both the ecological and geographical space suitable for the survival of a species can be highly useful and GARP provides us with the ability to examine both. GARP has occasionally been described as a “black box” because of the inability of the user to examine the internal functions and methods of the model-building process. Elith et al. (2006) stated that GARP “performed poorly,” but this conclusion was reached because only one accuracy metric (Area Under the Curve) was used to compare GARP with other models. Ecological niche models (ENMs) and species distribution models differ greatly across platforms and are too complex for only one accuracy statistic to be used to make comparisons because the goals of each modeling system may be different (e.g., fundamental vs. realized niche prediction). There are different methods used to measure each tool. The Area Under the Curve (AUC)
statistic has been shown to be imprecise in measuring the true accuracy of GARP models in recent studies because the statistic examines the entire curve instead of the location on the curve where occurrence data exists (Peterson et al. 2007). For this reason, it is more informative to use multiple measures of accuracy in conjunction with AUC scores. A conflation of niche theory also occurred in Elith et al. (2006) and differences between fundamental and realized niche models as well as species distribution models were not indicated. The ability of GARP to describe nonrandom relationships between locality data and environmental variables with the use of multiple rules that can be output as IF/THEN statements and projected onto a map indicates that it is not a “black box” algorithm (Kluza & McNyset 2005, Wiley et al. 2003). Various IF/THEN statements are created to describe the presence or absence of a species on the landscape and these rules can be viewed and analyzed to determine which variables and ranges of variables are indicative of the niche of a species (Blackburn et al. 2007, Kluza & McNyset 2005, McNyset 2005). Other modeling techniques only examine one or two rule types when estimating the distribution of a species (Box et al. 1993, Carpenter et al. 1993, Manel et al. 2001), but GARP is a superset-algorithm that combines multiple types of rules (atomic, range, negated range, and/or logit rules) to construct a prediction of the distribution of a species on the studied landscape (McNyset 2005, Peterson et al. 2002a, Stockwell & Peters 1999). McNyset (2005) was the first paper that listed the rules produced from a single model run to illustrate their structure within a GARP model. McNyset (2005) indicated that the interactions between variables were more important than only examining single variables when predicting the distribution of a species.
Other methods of measuring the influence that single variables may impose on the distribution of a species in ecological space have also been examined (Bauer & Peterson 2005, Costa et al. 2002). Bauer & Peterson (2005) developed the 'Boundary U-test' which explores environmental variables that are correlated to distributional limits across geography. Grid cells that are inside and outside of a predicted geographic distribution are compared for each environmental variable by using a Mann-Whitney U-test. The tool helps to detect and visualize environmental range edges to better understand environmental correlates. Costa et al. (2002) analyzed the ecological space that contained four populations of *Triatoma brasiliensis* by comparing the relationship found between annual mean minimum temperature and annual mean precipitation as described by GARP model outputs. Charts describing the ecological requirements of each species showed narrow environmental ranges of minimum temperature and mean precipitation that represent the niche of the species. Costa et al. (2002) alluded to the complexity of niche requirements when examining the varying environmental ranges that were needed by each species. McNyset (2005) also concluded that niche requirements were complex, but that examining multiple parameters within each rule-set could help to identify a large portion of the ecological niche of a species.

A major problem in ecological modeling is the occurrence of spatial autocorrelation. Spatial autocorrelation is the tendency of nearer objects to be more or less closely related than expected for random groups of observations and it is an inherent species modeling issue because of behavioral processes of a species and the tendency of neighboring locality data to exhibit similar environmental conditions due to spatial proximity (Legendre 1993). Some modeling approaches have tried to reduce the
effects of spatial autocorrelation. For example, in a study that used Bayesian modeling and GAP analysis to predict the distribution of 28 different bird species, measured variables that reached an asymptote at buffer strips between 51-200 meters from an occurrence point indicated that a spatial limit to the autocorrelation of agreement measures had potentially been reached (Hepinstall & Sader 1997). Despite efforts to reduce the effect of spatial autocorrelation, it is a very difficult issue that confronts ecological niche and species distribution modeling.

While spatial autocorrelation most likely does exist amongst variables and locality data to an extent, it does not occur in the model-building process of GARP because GARP is a heuristic pattern matching algorithm and not a traditional statistical modeling approach. GARP creates rules based on relationships between locality data and environmental variables. These rules are then applied to the landscape pixel-by-pixel and therefore surrounding pixels have no impact on each other because the rule application process starts over at each pixel. There are also other approaches to reducing the effect that spatial autocorrelation has when modeling the distribution of a species. A recent study used an autologistic regression approach that approximated the strength of species-habitat relationships and the strength of dependence between neighboring areas (Klute 1999, Klute et al 2000). This helped to describe the factors that influenced the distribution of a specific species. Essentially, spatial dependence does occur in natural ecosystems and spatial autocorrelation may affect model outputs in some approaches more so than in other approaches, but there are methods and approaches that seek to minimize its effect.
Predicting areas of a landscape that may be suitable for a species’ survival is highly useful and the central goal of ecological niche modeling applications. In addition to predicting the potential geographic distribution of a species, the rule-set writing and mapping application of Desktop GARP v 1.1.3 also provides the geographic location of where modeled environmental ranges (i.e., rules) exist on a landscape. McNyset (2005) was the first study to present a complete rule-set that showed each rule that was created in a single model run. The study showed that the relationship between distributions and variables included in the model is intricate and concluded that interactions between variables are commonly more meaningful than a value derived from a single variable. Blackburn (2006) examined the distribution of *B. anthracis* in the US and showed the dominant rules from the GARP 10 best subset projected onto the landscape. The maps produced illustrated that only a few rules dominate a best subset with usually between 2 and 4 presence rules per model and the study was the first illustration of the rule set and resulting rule maps. Blackburn et al. (2007) utilized multiple environmental variables including measures of temperature, precipitation, soil, and vegetation to establish a potential distribution model of *B. anthracis* in the United States based on the relationship between known occurrence data and environmental variables in proximity to the data. The study also produced a rule-set showing the primary presence and absence rules from a single best subset model experiment to demonstrate the value of examining biological information described within the GARP model output. Blackburn et al. (2007) found that specific ranges (or envelopes) of mean NDVI, precipitation, and elevation most often characterized the ecological niche of *B. anthracis* in the United States. The study emphasized the utility of being able to identify
environmental parameters that support \textit{B. anthracis} survival in order to delineate areas that may be at a higher risk of having an anthrax outbreak. Additional research by Blackburn et al. (2009) found that an anthrax outbreak in the fall of 2008 at a ranch near Bozeman, Montana occurred in an area that had never reported an anthrax outbreak, but was predicted to be within the endemic zone of the disease as described in Blackburn et al. (2007).

This study will describe the utility of the rule-set writing and mapping application of GARP in identifying distinct environmental parameters for a species as well as delineating the underlying environmental parameters that determine the predicted current and future potential distributions of the bacterium \textit{B. anthracis} in the central Asian country of Kazakhstan in an effort to expand on previous research by Aikimbayev et al. (unpublished manuscript) and Joyner et al. (2010). More specifically, the study will seek to answer the following questions: 1) What environmental parameters describe the current distribution of \textit{B. anthracis} across Kazakhstan? and 2) How useful is the rule-set writing and mapping application of GARP in providing important biological information about a species?

**Data and Methods**

**Anthrax Occurrence Data**

\textit{B. anthracis} locality data were produced from a historical record of anthrax outbreaks across Kazakhstan that occurred between 1937 and 2006. Most outbreaks occurred in the livestock population and locality data were most often reported as the farm location or nearest village location of the outbreak. Only outbreaks in livestock occurring between 1960 and 2000 were examined because of the implementation of wide-spread vaccination efforts around 1960. To make the data spatially unique at
8km² (i.e., the spatial resolution of the environmental variables used in the model), a further reduction of the data resulted in a small subset that contained 257 outbreaks of the original 3,947. The subset of 257 was then split into training (218 point locations) and testing (39 point locations) subsets (refer to figure 2-2). For modeling purposes, a cell only needs to be identified as present or absent for a species. The training subset was used for model building and the testing subset was used for measuring model accuracy after the model-building process. Bioclimatic (BioClim) data (both current and future) were freely downloadable (www.worldclim.org) on the WorldClim website (Hijmans et al. 2005) and satellite-derived environmental data were provided by the Typanosomiasis and Land Use in Africa (TALA) research group at Oxford University (Oxford, United Kingdom) (Hay et al. 2006). The Hadley Coupled Model Version 3 (HadCM3) A2 and B2 climate change scenarios for 2050 were used to predict the future potential distribution of B. anthracis across Kazakhstan as well as examine variance amongst rule-set combinations (Arnell 2004, Collins et al. 2001, Gordon et al. 2000).

All data were available at a resolution of approximately 8 kilometers and were used to produce environmental variables representing only the spatial extent of Kazakhstan. Resolutions of the TALA data and bioclimatic data were slightly different so a resampling procedure was also applied to produce environmental grids with identical cell sizes.

**Ecological Niche Modeling**

The study utilized the Genetic Algorithm for Rule-Set Prediction (GARP) to generate ecological niche models for B. anthracis in Kazakhstan. The rule-set writing and mapping application of Desktop GARP v 1.1.3 was employed for all GARP model production. Desktop GARP v 1.1.3 allows for the examination of the actual rule-sets
written during the model-building process and the location on the landscape where these rules apply, while Desktop GARP v 1.1.6 does not provide this output. GARP is a presence-only modeling tool that analyzes the relationship between locality data and the parameters of environmental variables in the same location. A total of 50 rules are created from four main rule types (atomic, range, negated range, and logit rules) for each model run to explain the relationship between locality data and environmental parameters. Once a rule-set (i.e., the combination of all 50 rules in each model run) is created, then the relationship is applied to other areas of the landscape that have similar environmental parameters. Each of the four rule types create IF/THEN statements that describe presence or absence parameters for the landscape (Stockwell & Peters 1999). Atomic rules use only single values of each variable to describe presence or absence, e.g. “if the total annual precipitation is 240 millimeters and the average annual NDVI is 0.63 then the species is present.” Range rules identify a specific range of multiple variables in space that need to exist in order for a species to be present or absent; e.g. “if the total annual precipitation is between 220 millimeters and 540 millimeters AND annual average temperature is between 5°C and 15°C AND annual average NDVI is between 0.05 and 0.71 then the species is present.” Negated range rules are similar to range rules except they describe the variable ranges that a species cannot exist in; e.g. “if the total annual precipitation is NOT between 110 millimeters and 200 millimeters AND annual average temperature is NOT between -3°C and 4°C AND annual average NDVI is NOT between -0.98 and -0.53 then the species is present.” Logit rules describe how the locations of a species fit to a logistic regression model that examines the environmental variables (Stockwell & Peters 1999). The
logistic regression gives an output probability ($p$) that verifies if a rule should apply to a particular part of the landscape where $p$ is calculated. Presence is predicted by the logit rule if $p$ is greater than 0.75 (Stockwell et al. 2006). The ability to use multiple rule types in an iterative process to create each rule-set establishes GARP as a super-set algorithm as opposed to many other modeling approaches that may only use range rules or logistic regression singularly.

The GARP modeling approach is stochastic, or random, and consequently produces different outputs with each model run. Because of the variance between each model run output, it is important to produce multiple runs and utilize the best-subset technique of selecting the 10 best models out of the original 50 that meet certain optimization parameters. Omission and commission thresholds are defined by the user to obtain a set of models that find a balance between sensitivity (absence of omission error) and specificity (absence of commission error) (Anderson et al. 2003). Omission is a measure of how much locality data are excluded from the area that is predicted to be present for a species, while commission is a measure of how much of the landscape was predicted present for a species including areas where no locality data exists. The best-subset procedure selects optimal output grids from all model runs and subsequently allows the user to examine the grids individually or simultaneously in a geographic information system (GIS). The grids can be summated to reveal different levels of model certainty. For example, some areas may be predicted present by only one model whereas other areas may be predicted present by all ten models. More model agreement infers more certainty when examining the fundamental niche of a species.
Modeling Procedures and Methods

Environmental variables were combined within the Desktop GARP Dataset Manager to create environmental coverage sets. Four different coverage sets were produced for the study: 1) coverage set for current scenario one (including altitude, bioclimatic variables, and measures of NDVI), 2) coverage set for current scenario two (including altitude and bioclimatic variables, but excluding measures of NDVI), 3) future coverage set (including altitude and bioclimatic variables predicted using the A2 climate change scenario), 4) future coverage set (including altitude and bioclimatic variables predicted using the B2 climate change scenario).

Current scenario one was the first model run and it used a 50/50 training split with a maximum of 200 runs and a convergence limit of 0.01. The “max iterations” was set to 1000 and all rule types were applied. A best subset was selected with an extrinsic omission measure and a hard omission threshold of 10%. The total “models under the hard omission threshold” was set to 20 and the commission threshold was set to 50% of the distribution. Locality data were analyzed using the coverage set created for current scenario one.

The 10 best models subset was output along with rule-set grids that showed the spatial extent of all rule-set combinations on the landscape for each model run produced by the model. A text file detailing rule-set combinations (IF…THEN statements) for each model run was also produced. Rule-set grids that described each of the 10 best models were projected onto maps within ArcMap 9.3 (ESRI 2008) and recoded in accordance with the dominant rule-sets within each model (Blackburn 2006). Dominant rule-sets were determined when a combination of rule-sets covered approximately 90% of the landscape of Kazakhstan (e.g., sometimes only 4 rule-sets
predicted presence/absence across 90% or more of Kazakhstan, while at other times up to 10 rule-sets were needed to predict presence/absence across 90% or more of Kazakhstan). Presence rules were displayed using a red color ramp, while absence rules were displayed using a blue color ramp. Once dominant rules were determined, they were also extracted from the text file that contained rule-set combinations. The rule-sets were organized in a table and delineated by model number (task number) and rule type (i.e., atomic, range, negated range, or logit rule) similar to McNyset (2005).

Current scenario two was the second model run and it used identical parameters that were used previously within the Desktop GARP environment. Locality data were analyzed using the second coverage set created for current scenario two. Both future coverage sets (A2 and B2 climate change scenarios) were also used to project the future potential distribution of *B. anthracis* on the landscape of Kazakhstan in 2050. Best model subsets were created for each of the three projections along with rule-set grids that showed which rules predicted varying parts of the landscape. Ten maps were produced for each of the three projections that showed where the dominant rules predicted presence and absence of *B. anthracis* in Kazakhstan.

**Analysis of Environmental Parameters established within GARP Rule-Sets**

In GARP, minimum and maximum environmental values (i.e., rainfall, temperature, NDVI) of range rules that described presence on the landscape were extracted from the model output and entered into a database. Zonal statistics were applied for areas described as being present by a logit rule to extract minimum and maximum environmental values of the area. These values were also input into a database. Since various rules described different regions of Kazakhstan seemingly based on latitude, the rules were divided into northern and southern rules whereas northern rules
predominantly described the environmental parameters north of 48°N latitude, while southern rules predominantly described the environmental parameters south of 48°N latitude. Some rules that were present in both the northern and southern regions were labeled as indeterminable. Once the database was complete, median values for each minimum and maximum environmental variable were calculated and input into a separate database where a bar chart was created showing the ranges of each environmental variable in the northern region and each environmental variable in the southern region. A chart was created for each current scenario to illustrate potential changes in environmental parameters when measures of NDVI were utilized in the model-building process versus when measures of NDVI were not utilized.

Additionally, centroids were created for each 8 kilometer grid cell across Kazakhstan. The centroids were then clipped by areas that were predicted to be present by one or more models. Next, a random sample of 1500 centroids from the clipped area was created and zonal statistics were used to identify the values of temperature range and wettest month precipitation at these locations.

A second clip of the centroids was then performed for areas where only total model agreement occurred (i.e., all 10 models). Zonal statistics were used again to identify the values of temperature range and wettest month precipitation at these locations. The resulting values for temperature range and wettest month precipitation were then plotted against each other and delineated by areas predicted present by at least one model and areas predicted present by all models. The locations of these variables were then visible in dimensional space.
Centroids were then separated into three categories that represented areas where
1) dominant northern rules occurred, 2) dominant southern rules occurred, and 3) at
least one model predicted presence.  Zonal statistics were utilized to identify the values
of temperature range and wettest month precipitation in each of these three areas.  The
results were plotted against each other in dimensional space and delineated by the
three categories to visualize the differences and similarities in values by location on the
landscape.  The process was repeated for mean temperature and mean NDVI.

Results

Overall, every model created 50 rules, but not all 50 were always used.  When a
best subset of 10 models was chosen for each of the current distribution predictions
(current scenario one and current scenario two) and future distribution predictions (A2
and B2 scenarios), then a total of 500 rules explained presence and absence for each
of the four predictions.  In concurrent GARP research (e.g., Anderson et al. 2003), a
summation of the 10 model best subset is usually created to show where high and low
model agreement occurred on the landscape.  Therefore, a summation of the 10 model
best subset was created for each of the current scenarios (Figure 3-1).  Accuracy
metrics including Area Under the Curve (AUC), total and average omission, and total
and average commission were also calculated for each of the current scenarios using
the testing subset (Table 3-1).  A total omission of 2.6 and average omission of 10.4
were reported, while a total commission of 38.18 and average commission of 54.48
were also reported for current scenario one.  The model also received an AUC score of
0.7046.  For current scenario two, a total omission of 2.6 and average omission of 8.6
were reported, while a total commission of 37.71 and average commission of 53.42
were also reported.  The model also received an AUC score of 0.7148.
The summated 10 model best subset and concomitant accuracy metrics are normally the only GARP outputs that are shown in many studies, but this study chose to delve deeper in order to obtain the maximum amount of information currently possible from a GARP model. A text file was written that contained all 50 rules for each model run and all of these rules were then projected onto the landscape (Figure 3-2). The resulting maps are complicated, but highly informative and illustrative of where the landscape embodies each of the rules created by GARP. To simplify the rule-sets and extract information about the ranges of the most influential variables, “dominant rules” were selected that described at least 90% of the landscape. An example of a dominant rule-set from current scenario one is given in Table 3-2. The rule-set contains 3 different rule types (logit, range, negated range) and a total of 8 dominant rules. Of the 8 dominant rules, 6 are rules that describe presence on the landscape and the remaining 2 describe absence on the landscape. These 8 rules can be visualized in figure 3-3 inset A. Of the 500 rules created for the 10 model best subset in current scenario one, only 72 were needed to explain over 90% of the landscape. These rules were subsequently labeled as dominant rules. Current scenario two needed 75 rules out of the original 500 to explain over 90% of the landscape. The A2 prediction needed 58 rules out of the original 500 to explain over 90% of the landscape, while the B2 prediction needed 60 rules to explain over 90% of the landscape.

A range of three to nine rule-set combinations predicted the presence or absence of *B. anthracis* on 90% or more of the landscape of Kazakhstan in each of the 10 best models for the four different projections. Maps showing the spatial extent of presence and absence rules for current scenario one were produced (Figure 3-3). Overall, the
maps aided in visualizing the dominant rule-set combinations that predicted the presence or absence of *B. anthracis* in specific regions of Kazakhstan. A difference in rules that described the northern limits of *B. anthracis* in Kazakhstan and the southern limits was also evident and a demarcation of approximately 49°N latitude most often separated these varying rules. Between two and four rule-set combinations were used to predict the presence of *B. anthracis* in the northern tier of Kazakhstan, while between one and three different rule-set combinations were used to predict the presence of *B. anthracis* in the southeastern region. The rules were representative of various IF…THEN statements that outlined the ranges and parameters of environmental variables that must be present for *B. anthracis* survival according to Desktop GARP v 1.1.3. In current scenario one a total of 28 range rules were used to predict areas of presence considered to be in the northern regions of Kazakhstan, while only 9 range rules were used to predict areas of presence considered to be in the southern regions. Conversely, 4 logit rules were used to describe areas of presence in the southern regions, while only 3 logit rules were used to describe areas of presence in the northern regions. No negated range rules were considered to be dominant presence rules, but a similar number of negated range (6) and range rules (5) were used to describe absence on the landscape while 16 logit rules were used to describe absence. All rule types were used to describe presence and absence on the landscape, but dominant presence rules were composed of only range and logit rules and dominant absence rules were composed of negated range, range, and logit rules. Several model runs showed more variance in the amount of rules that were needed to predict up to 90% of the landscape. Task 17 (Figure 3-3, Inset B) utilized nine different presence and absence rules, while
Maps were also created that described areas of Kazakhstan that were predicted to currently be present or absent of *B. anthracis* using current scenario two (Figure 3-4). Between one and two rules were used to describe most of the southern limits of *B. anthracis*, while between one and four rules were used to describe most of the northern limits of *B. anthracis*. A total of 21 range rules were used to describe most of the northern areas of Kazakhstan that were predicted to be present, while only 1 logit rule was utilized.

Rules describing the future potential distribution of *B. anthracis* using the A2 climate change scenario were also mapped onto the landscape (Figure 3-5). The prediction determined that a smaller geographic space would be suitable for *B. anthracis* survival based on future climate projections. Generally, fewer rules were used to determine presence and absence on at least 90% of the landscape in most model runs. Individual model commission was calculated for each task in addition to total and average commission for the entire best subset in order to determine how much of the landscape was predicted to be suitable for *B. anthracis* and how much variance in commission existed between each model run (Table 3-3). A wide range of variance in individual model commission existed with task 16 only reporting a commission of 36.7, while task 71 reported a commission of 50.1. A total commission value of 26.60 and average commission value of 42.03 were reported.

The B2 climate change scenario was also used to predict the future potential distribution of *B. anthracis* in Kazakhstan based on climate change predictions that were
dissimilar from those of the A2 climate change scenario and the rules were also mapped onto the landscape (Figure 3-6). Individual model commission was calculated for each task in addition to total and average commission for the entire best subset in order to determine the amount of landscape predicted present for B. anthracis in each model run and how much variance in commission existed between each model run (Table 3-4). A wide range of variance in individual model commission existed with task 32 only reporting a commission of 25.3, while task 20 reported a commission of 50.6. A total commission value of 14.12 and average commission value of 31.92 were reported.

The minimum and maximum values of all dominant presence rules for each of the current scenarios were organized into a database. A subset of rule values from “task 11” and “task 17” from current scenario one was shown to illustrate the similarity between certain minimum and maximum values found in each rule-set (Table 3-5). Each “task” represents a single model run. For example, the minimum and maximum precipitation values were nearly identical between rule number 4 and rule number 24 in “task 17” and the minimum and maximum wettest month values were absolutely identical between rule number 46 and rule number 47 in “task 17.” The mean NDVI values were also very similar in six out of the seven rules that the variable was used in and the driest month values were nearly equal in each of the seven rules that it appeared in. A subset of rule values from “task 66” and “task 70” from current scenario two was also shown to disclose similarities in minimum and maximum values found in each rule-set (Table 3-6). Precipitation values were similar in each rule and identical between rule number 38 and rule number 46 in “task 66,” while average temperature values were comparable in five out of the six rules that the variable was used in. In
particular, the maximum average temperature value varied less than half a degree between all six rules. The driest month values were uniform in each of the three rules that used the variable. The rules were also ascribed as being predominantly found in either the northern tier of Kazakhstan or the southern tier of Kazakhstan and values of rules describing *B. anthracis* presence in the each tier were also similar. Specifically, temperature range maximum values in the northern tier were mostly higher than maximum values found in the southern tier.

A bar chart was created to better visualize the overall environmental parameters in the northern and southern ranges of predicted *B. anthracis* distribution in current scenario one (Figure 3-7). Ranges of most environmental variables were similar with the exception of a lower temperature range and wider precipitation ranges (total and wettest month) in the southern region. A narrow envelope of “mean NDVI” was also identified in the chart.

A bar chart was again created to better visualize the overall environmental parameters in the northern and southern ranges of predicted *B. anthracis* distribution in current scenario two (Figure 3-8). Ranges of most environmental variables were similar with “temperature range” revealing the largest disparity between the parameters of each region. Lower temperature ranges were again observed in the southern region along with wider precipitation ranges (total and wettest month).

Temperature range and wettest month precipitation totals showed the widest variance between northern and southern rule types so these variables were plotted against each other in dimensional space as another way to visualize the disparity in variable ranges for the two regions (Figure 3-9). The variables were initially delineated
by whether they were in areas predicted by at least one model or only areas predicted by all 10 models (Figure 3-9). The variables were then delineated by northern and southern regions in the areas that were predicted by all 10 models (Figure 3-10). Variable ranges for each of the two variables within the best subset covered the entire spectrum of environmental parameters that were suitable for *B. anthracis* survival (shown in grey). The environmental parameters in the north exhibited a narrow range of wettest month precipitation and a narrow and high temperature range when compared to maximum range found in the best subset (shown in red). The environmental parameters in the south exhibited a wide range of wettest month precipitation and a wide and relatively low temperature range when compared to the maximum range found in the best subset (shown in orange). Mean NDVI and mean temperature were also plotted against each other in dimensional space and initially delineated by whether they were in areas predicted by at least one model or only areas predicted by all 10 models (Figure 3-11). The variables were then delineated by northern and southern regions in the areas that were predicted by all 10 models (Figure 3-12). Variable ranges for each of the two variables within the best subset covered the entire spectrum of environmental parameters that were suitable for *B. anthracis* survival (shown in grey). The environmental parameters in the north exhibited a very narrow and mostly compact range of mean temperature and mean NDVI (shown in red). The environmental parameters in the south exhibited a narrow range of NDVI, but scattered range of mean temperature.

**Discussion**

This study performed an in-depth analysis on the actual rules and rule-sets written by GARP during the modeling process in an effort to quantify the environmental
variables that were used to model the potential geographic distribution of *B. anthracis* in Kazakhstan. The study also revealed the importance of being able to use rule-sets to discern between models that use different variables and models that are projected into the future. An examination of the complex rule-sets written by GARP affirmed the usefulness of obtaining biological data from the GARP modeling process as described in previous research efforts (Blackburn et al. 2007, Kluza and McNyset 2005, McNyset 2005). GARP adapted by changing the range of some variables (e.g., annual precipitation total) to account for the loss of others (e.g., measures of NDVI) when comparing models that used different variables. It was also useful to project the current environmental parameters onto the future landscape to determine the geography of future rules.

In GARP, atomic rules were not considered to be dominant presence or absence rules in any model run indicating that they can only be used to predict a very small, specific area of the landscape and are not very useful in identifying a robust “envelope” of species distribution. A range of parameters in variable was a better descriptor of habitat suitability, which has also been supported by Holt and Gaines (1992), which suggested that the ecological niche of a species’ represents that environment that supports the mean phenotype of the population. The model of current distribution that used measures of NDVI (i.e., current scenario one) produced a slightly more constrained distribution than the model of current distribution that did not use measures of NDVI (i.e., current scenario two) suggesting that NDVI was a limiting variable in the *B. anthracis* models for Kazakhstan. Blackburn et al. (2007) described mean NDVI as one of the main limiting variables to the predicted distribution of *B. anthracis* in the
United States and this study confirms that the same is true of the predicted distribution in Kazakhstan. A slightly less constrained distribution was produced by current scenario two possibly because many land surface vegetation nuances were not as evident when only examining measures of precipitation and temperature although these factors are often associated with concurrent NDVI values.

Different rule sets for the northern and southern/southeastern regions of Kazakhstan indicated that there was apparent variation in relationships between environmental parameters (rules) written for the predicted northern and southern regions. Differences in annual temperature range between the northern and southern regions seemed to be one of the most apparent parameter variations between the rules.

It is important to remember that the goal of GARP (and other ENMs) is to produce a robust and accurate prediction of the spatial distribution of a target species. Because of this, the 10-best subset is still highly useful in displaying the spatial distribution of model agreement. However, when using the geographic area predicted by any given level of model agreement to display that region in variable space, it homogenized the landscape and did not show the heterogeneity that actually existed in the northern and southern rules. The disparity in ranges between the northern and southern regions showed that the environmental envelope for \textit{B. anthracis} shifted across latitude. This did not indicate that different niche requirements exist or infer that two different sub-species of \textit{B. anthracis} exists on the landscape, but rather that the northern and southern regions fulfilled different parts of the niche requirements of \textit{B. anthracis}.

Modeling at the regional level allowed us to examine realized portions of a species’ niche.
The use of NDVI in current scenario one showed that a narrow envelope of mean NDVI exists where *B. anthracis* is predicted as being present on the landscape, while a fairly large envelope of NDVI amplitude also described the presence of *B. anthracis* (Figure 3-8). The narrower envelope of mean NDVI may help to explain some of the variation in commission between the two models of current distribution, but much of the area that was predicted using ranges of NDVI was also predicted in the subsequent model of current distribution that did not use measures of NDVI indicating that measures of precipitation and temperature captured most of the ecological variability that measures of NDVI also captured. Similar areas (although not identical) of the northern and southern regions of Kazakhstan were predicted to be present for *B. anthracis*, while interior and western regions were again predicted to be absent in the second model of current distribution. In current scenario one, areas that were predicted to be present tended to expand slightly more into interior Kazakhstan from the southern and northern regions. A wider range of annual precipitation totals may have been used to account for the lack of NDVI variables. Also, when comparing ranges of variables in the northern and southern regions, noticeably narrower wettest month precipitation values were observed for the north as well as narrower and overall higher temperature range values (Figure 3-9). The ability to visualize this disparity through examining bar charts that were created using median maximum and minimum values as well as creating a variable cloud that showed the disparity in dimensional space was a major advantage of the GARP rule-set writing and mapping application.

Individual commission of each task in the A2 and B2 climate change scenarios showed a wide range of variance, indicating that there was a high amount of
disagreement between the future prediction models (Table 3-3 and 3-4). Individual commission varied between 36.7 and 50.1 for the A2 prediction, whereas individual commission varied between 25.3 and 50.6 for the B2 prediction. This indicates a high degree of uncertainty in future predictions of the geographic distribution of *B. anthracis* although both models reported lower total and average commissions than are reported in present conditions. A contraction of suitable habitat continues to be suggested by both the A2 and B2 predictions, but variance among commission values should not be ignored. Commission values also indicate that the B2 climate change scenario predicts a greater amount of habitat contraction than does the A2 climate change scenario. Various rules predicted onto the future landscape also indicate where current ranges of environmental parameters that are suitable for *B. anthracis* survival will exist in the future. The same variable ranges that were identified in the current distribution were projected into the future to show where these ranges expand and contract.

Generally, range rules were predominantly used to describe presence in the northern regions (above 48°N latitude), while logit rules were predominantly used to describe presence in the southern regions (below 48°N latitude) in all four models (current (2), A2 prediction, and B2 prediction). However, some overlap where presence rules were predicted on the landscape did occur between the two regions. In both models of current distribution, a combination of range and logit rules were used to describe most of the southern range of *B. anthracis*, while almost no logit rules were used to explain the northern range. An even combination of negated range rules and range rules described absence in current scenario one, while more negated range rules described absence in current scenario two.
The disparity in rule types used to predict the southern and northern regions was also evident in both future prediction models. Only range rules were used to describe most of the presence in the northern regions in both the A2 and B2 predictions, while only logit rules were used to describe most of the presence in the southern regions predicted by the A2 and B2 scenarios. Rules that predicted absence in both future prediction models were almost entirely negated range rules while few range rules predicted absence in each model.

Though much variation was exhibited between rule types and total number of rules used for each model experiment, a consistent environmental envelope was identified and spatially visualized. Additionally, the actual variation between rules within models and even between models is minimal (Refer to Table 3-5 & Table 3-6). Many rules produced in each model showed similar variable ranges thereby reducing the actual number of unique rules to less than 10 in most model subsets from an original total of 500 rules. The ability of GARP to visualize changes in variable relationships as defined by geography is a major advantage along with its ability to apply rules to the landscape pixel-by-pixel although it is difficult and labor-intensive to extract this information from a GARP output.

Recent studies conducted by McNyset (2005) and Blackburn et al. (2007) outlined the importance of examining rule-sets to extract vital biological data that delineate the range of a species in ecological space and this study further confirms the utility of identifying important rule-set combinations that predict areas of a landscape to be present or absent of a species. The study also produced a complete rule-set for the entire best model subset along with corresponding maps that showed where the rules
were applied to the landscape consequently concluding that GARP is not a black box, but rather a useful and explanatory ecological niche modeling tool. The ability of GARP to describe complex environmental requirements makes it very useful for a multitude of applications including modeling the current and future potential distributions of invasive species (Arriaga et al. 2004) and targeting conservation endeavors for endangered species (Peterson & Robins 2003). More research on modeling techniques is needed to expand on the utility of their outputs and to unlock even more biological data that may potentially be found by modeling the ecological niche of a species.
Figure 3-1. Genetic Algorithm for Rule-set Prediction (GARP) models showing the summated best subsets for current scenario one (A) and current scenario two (B)
Table 3-1. Accuracy Metrics for the current predicted distributions

<table>
<thead>
<tr>
<th>Metric</th>
<th>Scenario One</th>
<th>Scenario Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>N to build models</td>
<td>218†</td>
<td>218†</td>
</tr>
<tr>
<td>N to test models</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>Total Omission</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Average Omission</td>
<td>10.4</td>
<td>8.6</td>
</tr>
<tr>
<td>Total Commission</td>
<td>38.18</td>
<td>37.71</td>
</tr>
<tr>
<td>Average Commission</td>
<td>54.48</td>
<td>53.42</td>
</tr>
<tr>
<td>AUC*</td>
<td>0.7046 (z=9.232§, SE=0.047)</td>
<td>0.7148 (z=9.399§, SE=0.047)</td>
</tr>
</tbody>
</table>

* AUC = area under curve
† N was divided into 50% training/50% testing at each model iteration
§ p < 0.001
Note: Independent data used for accuracy metrics appear in figure 2-2 (yellow points)
Figure 3-2. Rules from one of the 10 best subsets in current scenario one (A) and rules form one of the 10 best subsets in current scenario two (B)
**Table 3-2.** Dominant rules from one of the 10 best subsets created in GARP using current environmental conditions that included measures of precipitation, temperature, and Normalized Difference Vegetation Index (NDVI).

*****Task 11 (Figure 3-4 Insert A)*****

1 negated range rule
IF NOT altitude=[-34.85,2821.49] AND wettest month=[21.21,127.55] AND driest month=[0.00,23.91] AND temperature range=[37.59,50.75] AND mean NDVI=[-1.00,0.42] THEN sp = ABSENCE

2 range rule
IF altitude=[4.02,1480.76] AND mean temperature=[2.82,15.26] AND precipitation=[322.85,687.99] AND wettest month=[23.46,110.88] AND driest month=[0.00,23.91] AND NDVI amplitude=[0.06,0.33] THEN sp = PRESENCE

10 logit rule
IF - mean temperature*0.0000 + precipitation*0.0078 - driest month*0.0039 - temperature range*0.0039 + NDVI amplitude*0.0039 THEN sp = PRESENCE

15 range rule
IF precipitation=[166.78,658.54] AND wettest month=[20.31,103.67] AND temperature range=[37.73,45.42] AND mean NDVI=[0.10,0.34] AND NDVI amplitude=[0.03,0.36] THEN sp = PRESENCE

30 logit rule
IF - mean temperature*0.0039 + precipitation*0.0039 + wettest month*0.0039 - driest month*0.0000 - temperature range*0.0039 + mean NDVI*0.0039 THEN sp = PRESENCE

33 logit rule
IF + altitude*0.0039 - mean temperature*0.0000 - precipitation*0.0273 + wettest month*0.0039 + mean NDVI*0.0039 - NDVI amplitude*0.0000 THEN sp = PRESENCE

41 range rule
IF altitude=[586.94,1500.19] AND mean temperature=[1.07,15.26] AND driest month=[0.89,22.89] AND temperature range=[39.23,51.46] AND mean NDVI=[0.10,0.34] AND NDVI amplitude=[0.06,0.33] THEN sp = PRESENCE

46 range rule
IF mean temperature=[1.07,15.36] AND precipitation=[146.17,658.54] AND wettest month=[19.86,102.77] AND driest month=[0.89,22.89] AND temperature range=[39.44,50.18] AND NDVI amplitude=[0.06,0.33] THEN sp = PRESENCE
Figure 3-3. Maps showing the dominant rules (presence – red color ramp; absence – blue color ramp) of the 10 best subset tasks projected onto the landscape for current scenario one. A) Task 11; B) Task 17; C) Task 21; D) Task 30; E) Task 35; F) Task 44; G) Task 49; H) Task 54; I) Task 56; J) Task 62. An inset of each task showing presence (red) and absence (grey) is also shown within each map.
Figure 3-4. Maps showing the dominant rules (presence – red color ramp; absence – blue color ramp) of the 10 best subset tasks projected onto the landscape for current scenario two. A) Task 1; B) Task 16; C) Task 20; D) Task 21; E) Task 32; F) Task 56; G) Task 65; H) Task 66; I) Task 70; J) Task 71. An inset of each task showing presence (red) and absence (grey) is also shown within each map.
Figure 3-5. Maps showing the dominant rules (presence – red color ramp; absence – blue color ramp) of the 10 best subset tasks projected onto the landscape for the A2 climate change scenario. A) Task 1; B) Task 16; C) Task 20; D) Task 21; E) Task 32; F) Task 56; G)Task 65; H) Task 66; I) Task 70; J) Task 71. An inset of each task showing presence (red) and absence (grey) is also shown within each map.
Table 3-3. Commission values for the future GARP model that utilized the A2 climate change scenario

<table>
<thead>
<tr>
<th>Model (Task)</th>
<th>Individual Commission</th>
<th>Total Commission</th>
<th>Average Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 01</td>
<td>49.2</td>
<td>26.60</td>
<td>42.03</td>
</tr>
<tr>
<td>Task 16</td>
<td>36.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 20</td>
<td>48.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 21</td>
<td>37.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 32</td>
<td>38.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 56</td>
<td>39.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 65</td>
<td>37.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 66</td>
<td>39.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 70</td>
<td>44.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 71</td>
<td>50.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3-6. Maps showing the dominant rules (presence – red color ramp; absence – blue color ramp) of the 10 best subset tasks projected onto the landscape for the B2 climate change scenario. A) Task 1; B) Task 16; C) Task 20; D) Task 21; E) Task 32; F) Task 56; G) Task 65; H) Task 66; I) Task 70; J) Task 71. An inset of each task showing presence (red) and absence (grey) is also shown within each map.
Table 3-4. Commission values for the future GARP model that utilized the B2 climate change scenario

<table>
<thead>
<tr>
<th>Model (Task)</th>
<th>Individual Commission</th>
<th>Total Commission</th>
<th>Average Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 01</td>
<td>37.9</td>
<td>14.12</td>
<td>31.92</td>
</tr>
<tr>
<td>Task 16</td>
<td>25.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 20</td>
<td>50.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 21</td>
<td>26.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 32</td>
<td>25.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 56</td>
<td>27.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 65</td>
<td>26.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 66</td>
<td>29.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 70</td>
<td>31.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 71</td>
<td>37.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3-5. An example of minimum and maximum ranges of two rule-sets produced in current scenario one. Precipitation is in millimeters, temperature is in degrees Celsius, and altitude is in meters.

<table>
<thead>
<tr>
<th>Model Number</th>
<th>11</th>
<th>11</th>
<th>11</th>
<th>11</th>
<th>11</th>
<th>17</th>
<th>17</th>
<th>17</th>
<th>17</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule Number</td>
<td>2</td>
<td>10</td>
<td>15</td>
<td>30</td>
<td>41</td>
<td>4</td>
<td>24</td>
<td>25</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>Rule Type</td>
<td>range</td>
<td>logit</td>
<td>range</td>
<td>logit</td>
<td>range</td>
<td>range</td>
<td>range</td>
<td>range</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean NDVI Min</td>
<td>0.10</td>
<td>0.19</td>
<td>0.10</td>
<td>0.14</td>
<td>-0.96</td>
<td>0.13</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean NDVI Max</td>
<td>0.34</td>
<td>0.42</td>
<td>0.34</td>
<td>0.34</td>
<td>0.40</td>
<td>0.33</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDVI Amplitude Min</td>
<td>0.06</td>
<td>0.05</td>
<td>0.03</td>
<td>0.06</td>
<td>0.01</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>NDVI Amplitude Max</td>
<td>0.33</td>
<td>0.44</td>
<td>0.36</td>
<td>0.33</td>
<td>0.42</td>
<td>0.33</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Precipitation Min</td>
<td>322.85</td>
<td>241.00</td>
<td>166.78</td>
<td>288.00</td>
<td>146.17</td>
<td>146.17</td>
<td>296.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipitation Max</td>
<td>687.99</td>
<td>796.00</td>
<td>658.54</td>
<td>487.00</td>
<td>646.76</td>
<td>652.65</td>
<td>749.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wettest Month Min</td>
<td>23.46</td>
<td>20.31</td>
<td>49.00</td>
<td>14.45</td>
<td>16.70</td>
<td>20.76</td>
<td>20.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wettest Month Max</td>
<td>110.88</td>
<td>103.67</td>
<td>74.00</td>
<td>127.10</td>
<td>103.67</td>
<td>98.71</td>
<td>98.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driest Month Min</td>
<td>0.00</td>
<td>1.00</td>
<td>0.89</td>
<td>0.00</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driest Month Max</td>
<td>23.91</td>
<td>24.00</td>
<td>22.89</td>
<td>31.87</td>
<td>21.00</td>
<td>21.00</td>
<td>21.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Temperature Min</td>
<td>2.82</td>
<td>-1.20</td>
<td>1.07</td>
<td>1.29</td>
<td>1.29</td>
<td>1.07</td>
<td>-11.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Temperature Max</td>
<td>15.26</td>
<td>2.70</td>
<td>15.26</td>
<td>14.82</td>
<td>14.82</td>
<td>15.26</td>
<td>15.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Range Min</td>
<td>39.40</td>
<td>37.73</td>
<td>45.90</td>
<td>39.23</td>
<td>35.67</td>
<td>35.17</td>
<td>35.17</td>
<td>39.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Range Max</td>
<td>47.30</td>
<td>45.42</td>
<td>48.90</td>
<td>51.46</td>
<td>42.29</td>
<td>53.10</td>
<td>53.17</td>
<td>50.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude Min</td>
<td>4.02</td>
<td>586.94</td>
<td>4.02</td>
<td>664.66</td>
<td>-132.00</td>
<td>4.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude Max</td>
<td>1480.76</td>
<td>1500.19</td>
<td>1422.47</td>
<td>1985.96</td>
<td>4784.00</td>
<td>1208.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>South</td>
<td>South</td>
<td>South</td>
<td>North</td>
<td>North</td>
<td>South</td>
<td>North</td>
<td>North</td>
<td>North</td>
<td>North</td>
</tr>
</tbody>
</table>
Table 3-6. An example of minimum and maximum ranges of two rule-sets produced in current scenario two. Precipitation is in millimeters, temperature is in degrees Celsius, and altitude is in meters.

<table>
<thead>
<tr>
<th>Model Number</th>
<th>66</th>
<th>66</th>
<th>66</th>
<th>66</th>
<th>66</th>
<th>70</th>
<th>70</th>
<th>70</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule Number</td>
<td>1</td>
<td>2</td>
<td>30</td>
<td>38</td>
<td>46</td>
<td>5</td>
<td>6</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Rule Type</td>
<td>range</td>
<td>logit</td>
<td>range</td>
<td>Range</td>
<td>range</td>
<td>logit</td>
<td>logit</td>
<td>range</td>
<td>range</td>
</tr>
<tr>
<td>Precipitation Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>281.62</td>
<td>178.56</td>
<td>178.56</td>
<td>284.00</td>
<td>287.51</td>
<td>184.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipitation Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>682.10</td>
<td>620.26</td>
<td>620.26</td>
<td>617.00</td>
<td>667.38</td>
<td>646.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wettest Month Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.45</td>
<td>17.00</td>
<td>14.45</td>
<td>30.67</td>
<td>18.00</td>
<td>31.00</td>
<td>18.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wettest Month Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>128.45</td>
<td>128.00</td>
<td>110.88</td>
<td>96.91</td>
<td>128.00</td>
<td>81.00</td>
<td>105.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driest Month Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td></td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Driest Month Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20.87</td>
<td>21.00</td>
<td>21.00</td>
<td></td>
<td>21.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Temperature Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-12.58</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Temperature Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Range Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>39.70</td>
<td>35.40</td>
<td>39.48</td>
<td>39.27</td>
<td>35.30</td>
<td>42.60</td>
<td>39.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Range Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>41.86</td>
<td>46.20</td>
<td>51.28</td>
<td>51.28</td>
<td>46.80</td>
<td>51.60</td>
<td>50.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28.19</td>
<td>-30.00</td>
<td>508.76</td>
<td>8.17</td>
<td>6.00</td>
<td>8.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2611.25</td>
<td>3721.00</td>
<td>4493.48</td>
<td>2491.11</td>
<td>2141.00</td>
<td>2491.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>South</td>
<td>South</td>
<td>North</td>
<td>North</td>
<td>North</td>
<td>South</td>
<td>North</td>
<td>North</td>
<td>North</td>
</tr>
</tbody>
</table>
Figure 3-7. Median range of variables describing both northern and southern distributions using measures of Normalized Difference Vegetation Index (NDVI) (current scenario one)

Figure 3-8. Median range of variables describing both northern and southern distributions without using measures of NDVI (current scenario two)
Figure 3-9. Variable cloud delineated by any area predicted present (light grey) and areas predicted present by all 10 models (dark grey) and visualized in dimensions of wettest month precipitation and temperature range.
Figure 3-10. Variable cloud delineated by location and visualized in dimensions of wettest month precipitation and temperature range
Figure 3-11. Variable cloud delineated by any area predicted present (light grey) and areas predicted present by all 10 models (dark grey) and visualized in dimensions of mean NDVI and mean temperature.
Figure 3-12. Variable cloud delineated by location and visualized in dimensions of mean NDVI and mean temperature
CHAPTER 4
CONCLUSION AND FUTURE RESEARCH

The goals of this thesis were to predict the current and future potential distributions of *Bacillus anthracis* in Kazakhstan and to examine the environmental parameters produced by the Genetic Algorithm for Rule-set Prediction (GARP) in the model-building process. It is expected that the potential habitat for *B. anthracis* will contract in most regions of the country, especially the southeastern regions where the majority of anthrax outbreaks have occurred. Information on the future potential changes in distribution is of particular importance when identifying areas where there is an increased or decreased possibility that *B. anthracis* could be present on the landscape especially when supplies and resources are limited. Additionally, the output produced by GARP provided a better understanding of how GARP constructs rule-sets and projects presence and absence onto the landscape and the output was important in understanding the geographical and ecological space where *B. anthracis* exist in Kazakhstan.

A comparison of models in chapter 2 (i.e., *current scenario one* and *current scenario two*) concluded visually that measures of vegetation did effect the modeled ecological niche of *B. anthracis* (Figure 2-9). Less interior areas of Kazakhstan were predicted to provide suitable habitat for *B. anthracis* when modeled using measures of Normalized Difference Vegetation Index (NDVI) and chapter 3 later confirmed that mean NDVI was the most limiting variable for *B. anthracis* in Kazakhstan with a very narrow envelope indicating the range of mean NDVI where *B. anthracis* can survive (Figure 3-7).
There were also slightly different regional environmental parameter preferences that delineated the northern and southern distributions of *B. anthracis*. Different genetic strains of *B. anthracis* exist across the landscape of Kazakhstan and the areas in the south appear to be dominated by the A1a and A3b strains (Aikimbayev et al. 2010). The predicted contraction of suitable habitat in the southern regions of Kazakhstan may lead to the eventual disappearance of these strains or at least a significant contraction of their current habitat. Much of the predicted contraction in *B. anthracis* habitat may also depend on where cattle production may shift in the future based on climate change in the region. Climate change is expected to affect precipitation and temperature patterns in central Asia and these changes may expand rangeland in some areas and reduce rangeland in other areas. More importantly from an anthrax control and management standpoint, since the disease is density dependent it will be vitally important to monitor changes in cattle distribution as cattle may be moved to new areas that have not previously reported anthrax outbreaks but where *B. anthracis* is present in the soil. Conversely, cattle may also be moved to areas that previously exhibited a high amount of anthrax outbreaks, but the area may no longer be suitable for the organism’s survival.

Predictions based on future climate change simulations should be repeatedly tested and monitored to insure that they most accurately reflect expected changes. Future climate change scenarios are hypothetical and are merely predictions based on our current knowledge of climatic processes. The models and algorithms used to predict future changes in climate can and most likely will change as the technology used to create climate change scenarios evolves. Actual climate change will also fluctuate
and evolve based on an intricate combination of atmospheric processes as well as political and cultural changes, however advanced planning that seeks to identify areas of potentially expanded or contracted *B. anthracis* habitat may serve to start a discussion about potential changes that researchers may need to make about how surveillance can adapt to geographic changes of the distribution of the disease.

Anthrax outbreaks must be continually monitored to distinguish whether or not a shift is occurring. If a trend of fewer outbreaks in the southern region and more outbreaks in the northern region appears to be transpiring, then control and surveillance measures must be enacted accordingly. The results of current modeling endeavors also indicate that further modeling efforts should be employed to improve upon our current knowledge of environmental parameters for *B. anthracis* and also to confirm and/or update future climate change predictions. Similar studies could also be implemented in other regions at various latitudes around the world to examine if comparable reductions on the landscape are predicted for the future in additional areas where *B. anthracis* is endemic. More research on the environmental requirements of *B. anthracis* will only add to recent findings about what is needed for *B. anthracis* survival and GARP has provided a valuable window into these niche constraints.

The Genetic Algorithm for Rule-set Prediction was utilized extensively in both studies, but the second study was as much a foray into the prediction of environmental parameters that support *B. anthracis* survival as it was an examination of the inner-workings of an ecological niche modeling (ENM) system. GARP is a popular niche modeling approach, but the actual process that is used to define the landscape where a species may or may not exist is not often discussed.
GARP is a superset algorithm that uses a stochastic and iterative approach during the model-building process, therefore repetition or similarity in rule-set outcomes indicates if a set of environmental ranges are conserved within and across models. The similarity in variable ranges across models also made it less difficult to define the approximate environmental parameters that GARP established to predict the ecological and geographical distribution of *B. anthracis*. Through the use of the rule-set writing and mapping application of GARP v. 1.1.3 a repetition in similar environmental ranges was observed and the rules created by GARP were projected on the landscape and spatially visualized. The study showed that GARP has the ability to identify environmental parameters, write rules about these parameters, and then project the rules onto the appropriate area of the landscape. The ability to deconstruct the output produced by GARP in a step-by-step approach was a major advantage that facilitated the capacity to obtain biologically useful information about *B. anthracis* and explain why it is predicted in certain parts of Kazakhstan and not in others. This advantage also alludes to the ability of GARP to obtain biologically useful information about other species while at the same time addressing the issue of being a “black box” algorithm (Elith et al. 2006, Stockman et al. 2006).

Both of these current studies concluded that the GARP modeling process is highly useful and can be used to model countless other flora, fauna, and diseases, but understanding the process that GARP uses and the outputs that GARP produces is the key to interpreting the distributions that GARP produces. Many other modeling approaches also identify environmental ranges and should be explored more
extensively in future studies, but this thesis has emphasized that a great amount of biological data can potentially be obtained from GARP if used appropriately.

There are many avenues for future research that not only involve expanded modeling efforts, but also further exploration of the robust anthrax dataset from Kazakhstan. The dataset not only contains the locations of anthrax outbreaks over a 70 year period, but also the monthly occurrences of outbreaks in conjunction with the total number of animals that were affected during each occurrence. The temporal component of the dataset was not explored in previous studies, but anthrax is a seasonal disease that normally occurs in spring and summer months and depends on many preceding climatic events to align. Many studies have concluded that conditions resulting from a wet spring followed by a dry summer and than a heavy rain event have historically coincided with anthrax outbreaks (Dragon et al. 1999, Hugh-Jones & Blackburn 2009, Parkinson et al. 2003, Smith et al. 2000, Van Ness 1971). Weather patterns that preceded an anthrax outbreak that occurred in Alberta, Canada in 1999 were examined and it was determined that the outbreak occurred after a prolonged period of warm, dry weather followed by a heavy rain event (Parkinson et al. 2003). Parkinson et al. (2003) revealed that abnormally high temperatures in May and June may have facilitated spore growth and multiplication and that the heavy rain event potentially made the spores more accessible to livestock that incidentally ingested contaminated soil during the period.

Potential patterns between outbreaks and atmospheric forcing mechanisms should also be examined to determine if anthrax outbreaks in Kazakhstan have been historically affected by one or more teleconnections. In other areas of the world,
atmospheric patterns such as the El-Nino Southern Oscillation have been significantly correlated to disease outbreaks (Harvell et al. 2002, Rodo et al. 2002, Stapp et al. 2004). Specifically, the changes in the intensity and location of the Siberian High should be investigated more. Panagiotopoulos et al. (2005) examined the intensity of the Siberian High over multiple decades and concluded that the strength of the high pressure system has been in decline since the late 1970’s. However, the Siberian High continues to be the most dominant pressure system in central Asia and has a far-reaching influence on the regions climate (Sahsamanoglou et al. 1991). It has the greatest intensity and densest air masses of any northern hemispheric pressure system (Ding & Krishnamurti 1987). Multiple relationships between the Siberian High and central Asian climate have been identified with the pressure system having a far-reaching impact on the East Asian Winter Monsoon, Aleutian Low, and temperature patterns in south Asia (D’Arrigo et al. 2005). The impact of the Siberian High on disease outbreak variability in the region is rarely discussed, but could be explored as a part of future research efforts.

While preceding regional climatic conditions have often been associated with anthrax outbreaks, other climatic conditions may help to minimize or even end an outbreak altogether. The relationship between cold air arrival and the slowing of an epidemic has been noted in previous studies. Dragon & Rennie (1995) hypothesized that an influx of colder and more humid air could decrease the amount of sporulation in the soil and slow down and eventually end an anthrax outbreak epidemic. Consequently, knowledge of climatic drivers that could potentially signal an end to epidemics could be equally as useful as knowledge of climatic drivers that potentially
signal a beginning to epidemics. Further research of both signals would be useful to public health agencies as they plan and prepare for potential outbreaks.

It is important to understand the multiple aspects of climatic patterns, changes, phases, and modes in order to better understand the underlying epidemiology of anthrax outbreaks not only in Kazakhstan, but in all regions of the world. Suprayogi et al. (2007) emphasized the need for more research on climatic changes and patterns that impact anthrax and other disease outbreaks. Knowledge of the driving mechanisms behind anthrax outbreaks is paramount to forecasting an increase or decrease in risks for disease outbreaks. The ability to forecast an increased likelihood of outbreaks is crucial to disease prevention and vaccination. Blackburn et al. (2007) indicated that further understanding of anthrax will aide in efforts to prevent and potentially eradicate the disease from specific landscapes and Suprayogi et al. (2007) advocated for the implementation of an early warning system. Further research on outbreaks and related climatic patterns is warranted to possibly create a reliable warning system for Kazakhstan that could be efficient and economically feasible and that could reduce the risk of anthrax infection in both the livestock and human populations.

To effectively implement many of the previously discussed disease surveillance and warning systems in Kazakhstan and elsewhere in central Asia applicable training and resources must be provided to improve the current public health system. The dissolution of the Union of Soviet Socialist Republics (USSR) left much of the regions public health infrastructure in disrepair, but many technological disadvantages are currently being eroded. Over the past several years, Geographic Information System
(GIS) technology has been implemented in disease surveillance stations throughout central Asia to help create digital spatial databases and streamline monitoring and tracking efforts (Aikimbayev et al. 2010 & unpublished manuscript). Scientists around the globe have assisted in GIS training efforts in central Asia and this has increased the ability to obtain, process, analyze, and share data and results. Newly emerging GIS, modeling, and remote sensing technologies will provide much-needed tools that may aid the public health systems potentially develop and sustain near real time disease surveillance. Additionally, the implementation of new technologies may increase the ability of the Kazakhstan public health system to monitor and respond to outbreak situations. Through the exploration of ecoenvelopes, evolutionary patterns, and spatial distributions of *B. anthracis*, GARP may aid in providing important information to public health officials. This thesis may also lay the groundwork for future research with GARP and other modeling tools on *B. anthracis*, which may subsequently provide more information to public health officials about the disease.
LIST OF REFERENCES


Blackburn JK, Hugh-Jones ME, Hunter D, McNyset KM, Hadfield T (2009) Field validation of fundamental niche predictions of *Bacillus anthracis* for the U.S. Bacillus ACT 2009, 30 August - 3 September 2009, Santa Fe, New Mexico, USA.


BIOGRAPHICAL SKETCH

Timothy Andrew Joyner was born in 1985 in Waxhaw, North Carolina. He graduated *summa cum laude* from Parkwood High School in May 2004. In the fall of 2004, Joyner began undergraduate studies at Louisiana State University where he was a member of the Honors College and the College of Arts and Sciences. In August 2004, he was employed as a student research assistant in the World Health Organization Collaborating Center (WHOCC) for GIS and Remote Sensing for Public Health within the Department of Geography and Anthropology. While with the lab, Joyner collaborated on multiple projects including a study of mosquitoes transmitting West Nile in East Baton Rouge Parish, animating a historical yellow fever outbreak in New Orleans, database management of the Kazakhstan anthrax dataset, database management of Red Cross data for Hurricane Katrina, and media research for Hurricane Katrina. Joyner also provided technical support for a GIS teaching seminar in Almaty, Kazakhstan in July 2006. He graduated *cum laude* in spring 2008 with a Bachelor of Science in geography.

In the fall of 2008, Joyner enrolled in California State University, Fullerton to pursue his Master's degree in Geography. While at Fullerton, he became a graduate research assistant in the Spatial Epidemiology and Ecology Research (SEER) Laboratory where he collaborated on numerous projects in Kazakhstan pertaining to multiple diseases and potential disease reservoirs/vectors, developed training materials and conducted several GIS and ecological niche modeling training seminars in Kazakhstan and Rhode Island, and presented research at multiple conferences. In the summer of 2009, Joyner transferred to the University of Florida to continue working in
the SEER Lab which had also moved to the University of Florida. He received his
Master of Science from the University of Florida in spring 2010.