

ENVIRONMENTAL AND CLIMATIC PREDICTORS OF PRESENCE OF ADULT
AMBLYOMMA AMERICANUM IN FLORIDA

By

WILLIAM H. KESSLER

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

2017

© 2017 William H. Kessler

To my mother and father for their support, wisdom, and encouragement

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Gregory Glass for his astute observations, subtle guidance, and shared suffering during endless hours of fieldwork in the Florida sun. I would also like to thank Dr. Jason Blackburn for his assistance with the analytical methodologies used in this study and for playing the role of reviewer #2. Thanks also goes to Dr. Katherine Saylor for her expertise in all things ticks and for assistance in identifying specimens.

I would also like to acknowledge the contributions of my geography cohort, especially Joe Andreoli, Morgan Ridler, Brittany Hodik, and James Richardson in keeping this endeavor in perspective and providing the motivation to finish.

Lastly, I would like to thank my family for their support and encouragement.

TABLE OF CONTENTS

| | <u>page</u> |
|--|-------------|
| ACKNOWLEDGMENTS | 4 |
| LIST OF TABLES | 6 |
| LIST OF FIGURES | 7 |
| LIST OF ABBREVIATIONS..... | 8 |
| ABSTRACT..... | 9 |
| CHAPTER | |
| 1 INTRODUCTION | 10 |
| 2 CHARACTERIZATION OF ENVIRONMENTAL AND CLIMATIC PREDICTORS OF ADULT <i>AMBLYOMMA AMERICANUM</i> IN FLORIDA | 16 |
| 2.1 Background..... | 16 |
| 2.2 <i>Amblyomma americanum</i> : Species of Public Health Concern | 17 |
| 2.3 Species Distribution Models..... | 20 |
| 2.3.1 Generalized Linear Model Framework for SDMs..... | 21 |
| 2.3.2 Estimating Geographic Distributions of Ixodid Ticks..... | 22 |
| 2.4 Materials and Methods | 26 |
| 2.4.1 Study Area | 26 |
| 2.4.2 Tick Collection and Environmental Data Sources | 27 |
| 2.4.3 Methodology to Account for Spatial Autocorrelation..... | 33 |
| 2.4.4 Multivariate Logistic Regression | 35 |
| 2.4.5 Model Evaluation | 37 |
| 2.4.6 Spatial Predictions and Gradients..... | 42 |
| 2.5 Results..... | 44 |
| 2.6 Discussion..... | 60 |
| 3 CONCLUSIONS | 67 |
| APPENDIX | |
| A LOGISTIC REGRESSION IN R..... | 70 |
| B CONSIDERATION OF COVARIATE ANOMOLIES ON SPATIAL PREDICTIONS | 73 |
| LIST OF REFERENCES | 77 |
| BIOGRAPHICAL SKETCH | 87 |

LIST OF TABLES

| <u>Table</u> | <u>page</u> |
|--|-------------|
| 1-1 Yearly incidence of reportable tick-borne diseases in Florida from 2010-2016. | 13 |
| 2-1 Characteristics of variables | 29 |
| 2-2 Reclassification scheme for land cover variables | 31 |
| 2-3 Specimen totals by site..... | 45 |
| 2-4 Environmental and climatic variables used to model the spatial distribution of <i>Amblyomma americanum</i> in Florida. | 48 |
| 2-5 Environmental and climatic variables with low collinearity determined by Variance Inflation Factor..... | 49 |
| 2-6 Coefficients and variables of final model. | 50 |
| 2-7 Model validation metrics. | 51 |
| 2-8 Presence-absence discrimination for testing data. | 52 |

LIST OF FIGURES

| <u>Figure</u> | <u>page</u> |
|---|-------------|
| 1-1 Posted safety alert in San Felasco state park warning of exposure risk to ticks..... | 14 |
| 2-1 Distribution of collection sites | 46 |
| 2-2 Tick collections by month..... | 47 |
| 2-3 Spatial distribution of model residuals.. .. | 53 |
| 2-4 Probabilities of occurrence of <i>Amblyomma americanum</i> in Florida.. .. | 56 |
| 2-5 Estimated distribution of <i>Amblyomma americanum</i> in Florida | 57 |
| 2-6 Latitudinal trend in percentage of suitable area | 58 |
| 2-7 Latitudinal trend in high suitability areas | 59 |
| B-1 Examples of potential interpolation error in WorldClim bioclimatic variables. | 75 |
| B-2 Impact of interpolation error in bioclimatic variable Bio13 on the spatial prediction produced by the logistic regression model..... | 76 |

LIST OF ABBREVIATIONS

| | |
|----------------------|---|
| <i>A. americanum</i> | <i>Amblyomma americanum</i> , the lone star tick |
| <i>D. variabilis</i> | <i>Dermacentor variabilis</i> , the American dog tick |
| ENM | Ecological Niche Model |
| <i>I. scapularis</i> | <i>Ixodes scapularis</i> , the black-legged tick |
| NDVI | Normalized Difference Vegetation Index |
| SDM | Species Distribution Model |

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

ENVIRONMENTAL AND CLIMATIC PREDICTORS OF PRESENCE OF ADULT
AMBLYOMMA AMERICANUM IN FLORIDA

By

William H. Kessler

December 2017

Chair: Gregory Glass

Major: Geography

The lone star tick, *Amblyomma americanum*, is the most commonly reported tick in the southern United States and the primary vector of several human and livestock pathogens of public health concern. Although this species is endemic to Florida, little is known about the ecological preferences and spatial distribution of the species in the state. Using occurrence records of adult *A. americanum* collected between September 2015 and September 2016, a logistic regression model was developed to estimate environmental and climatic associations, as well as predict the fine-scale distribution, of the tick in Florida. Occurrence of adult lone star ticks was found to be associated with a combination of habitat and bioclimatic variables, namely forested land cover and seasonality of precipitation. The estimated spatial distribution, at a resolution of 1 hectare, indicated that probable occurrence decreases from North to South with very little area deemed suitable in the far southern reaches of the state. This trend reaffirms findings in the literature that prevalence of *A. americanum* on wildlife increases with latitude in Florida. The 1-hectare resolution of the estimated distribution is a significant improvement over distributions currently published in the literature and will better inform the public and state or federal agencies of potential risk of exposure to *A. americanum*.

CHAPTER 1 INTRODUCTION

In the United States, tick-borne diseases are the most commonly reported vector-borne diseases, outpacing reports of mosquito-borne illness by an average factor of 9:1 from 2004-2013 (USGCRP 2016). The species of ticks that can transmit these pathogens are not homogeneously distributed which results in disparate areas of exposure risk and disease incidence. In Florida, five tick species are commonly reported; the black-legged tick (*Ixodes scapularis*), the American dog tick (*Dermacentor variabilis*), the lone star tick (*Amblyomma americanum*), the gulf coast tick (*Amblyomma maculatum*), and the brown dog tick (*Rhipicephalus sanguineus*). Each of these species are considered endemic in the state (Bishopp and Trembley 1945, Yabsley, et al. 2009). In addition, many counties in Florida have reported other tick species including *Ixodes affinis* and *Ixodes minor* (Clark 2004, Forrester, et al. 1996, Wehinger, Roelke and Greiner 1995, Durden, Klompen and Keirans 1993). Collectively, these species are considered well established across much of Florida. Furthermore, reports indicate successful establishment of the exotic species *Amblyomma auricularium* and *A. rotundatum* in southern Florida (Oliver Jr, Hayes, et al. 1993, Mertins, Vigil and Corn 2017).

Coincidentally, the five most commonly reported tick species in Florida are known to transmit a number of pathogens capable of causing debilitating disease in humans and livestock. *Ixodes scapularis* is most commonly associated with tick-borne illnesses as it is the primary North American vector for Lyme disease- the most commonly diagnosed vector-borne disease in the United States (Armstrong, et al. 2001, Gray 1998). However, it is also known to transmit the pathogens causing anaplasmosis and several other diseases (Adelson, et al. 2004). *Dermacentor variabilis* and *Rhipicephalus sanguineus* are both known to vector the causative agent of Rocky Mountain spotted fever (RMSF); *D. variabilis* also vectors *Francisella tularensis* while *R.*

sanguineus additionally vectors canine ehrlichiosis (Bell 1944, Green 1931, Demma, et al. 2005, Smith, et al. 1976). The gulf coast tick, *A. maculatum* transmits *Rickettsia parkeri*, a spotted fever rickettsiosis (Sumner, et al. 2007). *Amblyomma americanum* is the most commonly encountered tick in Florida and other places in the southeastern United States and is capable of vectoring a number of disease causing pathogens (Armstrong, et al. 2001, Childs and Paddock 2003, Felz, Durden and Oliver Jr. 1996). While the lone star tick does not vector the Lyme disease pathogen, it can transmit *Ehrlichia chaffeensis* and *Ehrlichia ewingii* (which cause human ehrlichiosis), *Francisella tularensis*, and the presently unknown agent of Southern Tick Associated Rash Illness known as STARI (Hopla 1955, Hopla 1953, Masters, Grigery and Masters 2008, Goddard and Varela-Stokes 2009, Anderson, Greene, et al. 1992, Anderson, Sims, et al. 1993). In places where *A. americanum* and *I. scapularis* co-occur perception of Lyme disease may be exaggerated while *A. americanum* vectored diseases (namely STARI) go unreported; this could be due to growing public awareness of Lyme disease and the association of the disease with ticks in general, and the nuisance biting habits of *A. americanum* (Armstrong, et al. 2001). Expansion of the lone star tick's distribution into areas dominated by *I. scapularis* will further confuse diagnoses of Lyme disease and STARI as the presentation of symptoms are similar (Sayler, Boyce and Wisely 2016).

The multitude of tick species occurring in the state suggest that climatic and environmental conditions in Florida are broadly suitable for the persistence of many tick species and the pathogens they carry. Temperatures and humidity levels, which contribute greatly to tick activity and life cycle, rarely stray beyond what is considered highly suitable for many species (Heath 1981, Yoder and Tank 2006, Yoder and Benoit 2003). High temperatures coupled with low relative humidity limit ticks' ability to retain moisture leading to reduced host-seeking

activity and ultimately, desiccation. The monthly average relative humidity for much of Florida ranges from greater than 68% to nearly 90% while average monthly temperatures range from 10-29°C (Henry, Portier and Coyne 1994). A study by Yoder and Tank (2006) indicated that five common North American species of veterinary importance, including the dry-adapted *Rhipicephalus sanguineus* and the hydrophilic species *Amblyomma americanum*, share similar upper temperature limits (between 30-35°C) before rapid moisture loss occurs. These findings indicate that Florida's climate is well within the suitable temperature range for many North American and exotic tick species.

Understanding the distribution of arthropod vectors can provide insights into at-risk areas for vector-borne pathogens although there are some caveats (Ostfeld, Glass and Keesing 2005). For example, although high predicted exposure risk (based on habitat suitability) to the arthropod vector (*Ixodes pacificus*) has been positively associated with disease incidence as is the case of Lyme incidence in California, concordance between Lyme disease incidence and habitat suitability is not always high (Eisen, et al. 2006, Atkinson, et al. 2014). Furthermore, cases of Lyme borreliosis disproportionately cluster around northwestern Wisconsin and on the northeastern seaboard around New Jersey and east-central Pennsylvania even though the primary vector for the disease in the eastern United States is found in 37 states, primarily east of the Mississippi river (Kugeler, et al. 2015, Eisen, Eisen and Beard 2016). Still, studies have shown that presence of the primary disease vector and environmental factors associated with high populations of the species increase the risk of some vector-borne diseases, for example tick-borne Lyme disease (Glass, et al. 1995, Eisen, et al. 2006). Based on this assertion, models that estimate the distribution of a vector species can provide a broad estimate of disease risk (Ostfeld, Glass and Keesing 2005, Eisen, et al. 2006).

Ticks are common pests in Florida- as indicated by signs and postings in state parks and other public lands warning of the risk of tick exposure (Figure 1-1). Several dozen cases of locally acquired tick-borne illness are reported every year in Florida (Table 1-1) however little is known about the statewide distributions of the vector species.

Table 1-1. Yearly incidence of suspected and confirmed locally acquired reportable tick-borne diseases in Florida from 2010-2016. (Reportable Diseases Frequency Report 2017)

| Diseases | Year | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|----------------------|-------|--------|--------|--------|--------|--------|--------|--------|
| | Total | Counts |
| | 433 | 41 | 50 | 80 | 53 | 79 | 63 | 67 |
| Anaplasmosis | 11 | 3 | 6 | 1 | 0 | 0 | 0 | 1 |
| Ehrlichiosis | 120 | 8 | 13 | 20 | 17 | 23 | 16 | 23 |
| Lyme Disease | 210 | 22 | 22 | 39 | 21 | 35 | 35 | 36 |
| Spotted Fever | 92 | 8 | 9 | 20 | 15 | 21 | 12 | 7 |
| Rickettsiosis (RMSF) | | | | | | | | |

There have been few studies characterizing the geographic distributions and ecological preferences of ticks in the state at sub-county scales. Estrada-Pena (1998) estimated the probable distribution of *Ixodes scapularis* across North America at a resolution of 8km using a cokriging interpolation. The consideration of temperature and vegetation conditions improved model fit (sensitivity and specificity) and found that much of Florida was likely habitat for the species. Similarly, James, et al. (2015) modeled the distribution of *Dermacentor variabilis* in the United States using biologically relevant climatic variables, primarily summer vegetation greenness, at a resolution of 1km and found most of Florida to have high habitat suitability for this species as well. Another important species for public health, *Amblyomma americanum*, is the most commonly encountered tick in the southeastern United States and, as noted previously, a species of capable of vectoring multiple pathogens including *Ehrlichia chaffeensis*, *Ehrlichia ewingii*, the Panola Mountain *Ehrlichia*, among others (Estrada-Pena and Jongejan 1999, Childs and Paddock 2003, Reeves, et al. 2008). A 2015 analysis by Springer and colleagues of the

distribution of this species in the United States used a series of ecological niche modeling methods at the county level, and indicated a potential meridional gradient of decreasing suitability across Florida. A national study of *I. scapularis* nymphs also reported no collections of *A. americanum* from study sites south of roughly 28°N in Florida (Diuk-Wasser, et al. 2006). The results of these two studies indicate the potential for high heterogeneity in suitability for *A. americanum* across Florida and has implications for the distribution of *A. americanum*-borne disease risk.

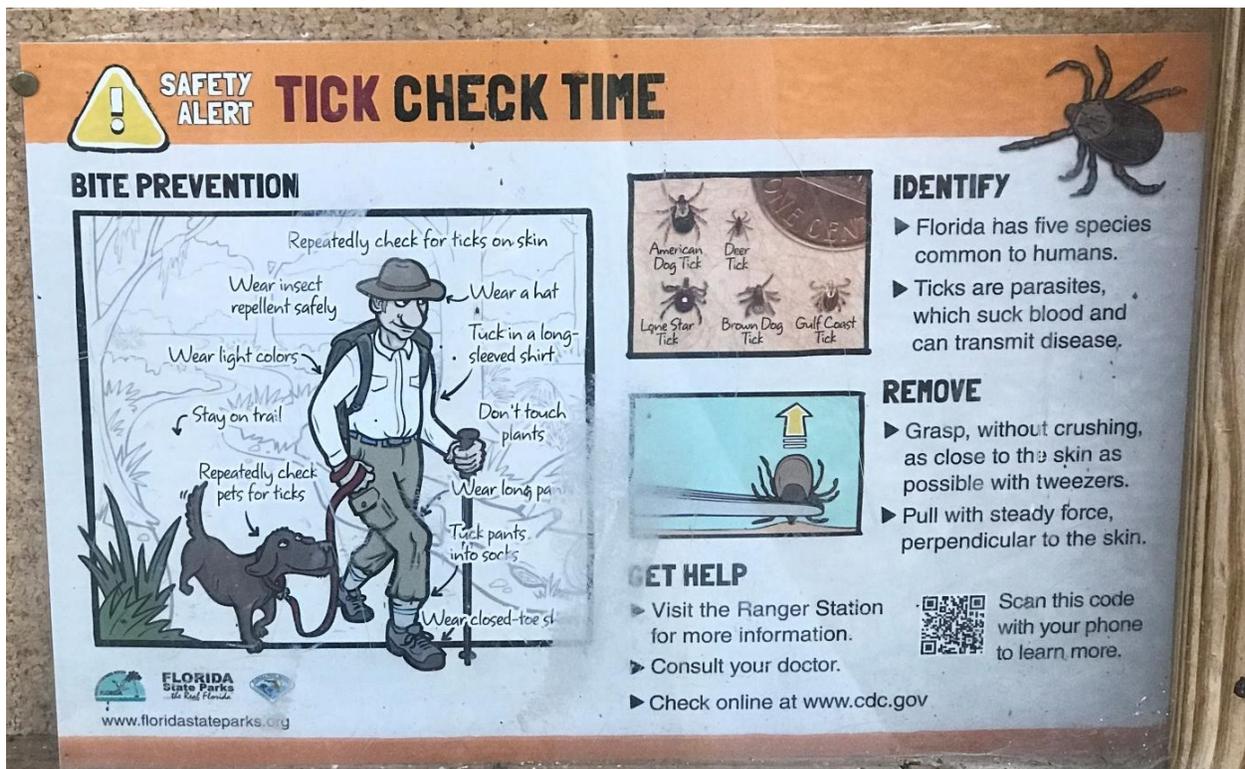


Figure 1-1: Posted safety alert in San Felasco state park warning of elevated exposure risk to ticks, highlighting risk mitigation techniques, and treatment of tick bites.

The current thesis research will develop a state-wide spatial model to answer the following two research questions:

1. What environmental and/or climatic factors are associated with the current distribution of adult *Amblyomma americanum* in mainland Florida? Several recent studies have linked local distributions of *A. americanum* to environmental and land cover factors. This study considers statewide associations between species occurrence and environmental and climatic conditions at a sub-county resolution.
2. Where are the areas of high probability of occurrence of *A. americanum* in Florida? Areas that are likely to harbor *A. americanum* may also be areas of increased human risk of specific tick-borne diseases. Understanding the distribution of this species will improve decisions about alerting the public to the risk of exposure to ticks and tick-borne disease.

The remainder of this thesis includes two additional chapters. The second chapter will examine the two research questions described above and provide further background, methods, results, and conclusions. The third and final chapter will examine the implications and limitations of this study.

CHAPTER 2
CHARACTERIZATION OF ENVIRONMENTAL AND CLIMATIC PREDICTORS OF
ADULT *AMBLYOMMA AMERICANUM* IN FLORIDA

2.1 Background

Hard-bodied ticks (family: Ixodidae) are obligate, blood feeding, ectoparasitic arthropods that are found worldwide (Jongejan and Uilenberg 2004). All Ixodid lifecycles are composed of four stages: eggs, larvae, nymphs, and adults. Oviposition results in a single egg mass in the leaf litter or in animal burrows. Each life stage requires a blood meal before detachment from the host and undergoing ecdysis (Anderson and Magnarelli 2008). The time requirement for ticks to complete a life cycle depends on a number of factors. The seasonal regulation of generational times may result in northern populations of a species taking significantly longer than in southern populations (1-3 years as opposed to one or more generations per year) (Oliver Jr 1989, Strickland, et al. 1976). The photoperiod, availability of hosts, temperature, and moisture all impact the diapause period of eggs and molting of instars (Oliver Jr 1989).

Sufficient moisture availability modulates both the general times for many tick species as well as daily behavior patterns. Ticks are prone to dehydration due to their infrequent intake of fluids (during feeding), large surface to volume ratio, and the amount of time spent in low moisture environments while questing (Anderson and Magnarelli 2008, Knulle and Rudolph 1981). To maintain an adequate water balance, most ticks retreat from questing locations on exposed plant surfaces back into moist leaf litter or other locations when ambient temperatures climb or relative humidity decreases (Anderson and Magnarelli 2008, Vail and Smith 2002).

Many tick species are host generalists that feed on various species of animals as adults. Ixodid ticks most commonly parasitize mammals, although at least one species in each genus also can utilize birds or reptiles during one or more life stages (Oliver Jr 1989). Immature life

stages may preferentially take blood meals from a single vertebrate class. Larval or nymphal ticks from more than 300 species feed on rodents (Oliver Jr 1989).

Most hard-bodied ticks are three-host species meaning that they feed on three different individual hosts, one during each life stage. Many three-host species such as *Ixodes scapularis*, *Dermacentor variabilis*, and *Amblyomma americanum* are important vectors of wildlife and human diseases because the multi-host feeding characteristics increase the likelihood of transmitting a pathogen from an infected individual to an uninfected individual via co-feeding, transstadial, or interstadial transmission (Randolph, Gern and Nuttall 1996, Jaworski, Bowen and Wasala 2013, Bremer, et al. 2005).

2.2 *Amblyomma americanum*: Species of Public Health Concern

Amblyomma americanum, commonly referred to as the lone star tick, is a non-nidicolous (non nest-dwelling) species whose known regional distribution extends from west-central Texas to the eastern seaboard and as far north as Maine (Anderson and Magnarelli 2008). Adult and nymphal ticks are generally most active from April through June with activity declining in late summer. Larval ticks frequently become active later in the season than adults or nymphs (Davidson, Siefken and Creekmore 1994, Cilek and Olson 2000). The abundance of *A. americanum* is highly dependent on the availability of suitable host species (Childs and Paddock 2003). Immatures (both larvae and nymphs) of the lone star tick feed on various small, medium and large mammals as well as birds, whereas in the adult life stage this species frequently parasitizes white-tailed deer, cattle, and feral swine (Anderson and Magnarelli 2008).

The general preference for adults to feed on deer and the non-specific biting habits of other life stages contribute to the species' significance as a vector for wildlife and human diseases (Childs and Paddock 2003). White-tailed deer (*Odocoileus virginianus*) are a primary reservoir of several *Ehrlichia* species making infection of feeding ticks likely (Childs and

Paddock 2003). Furthermore, a survey of emerging wildlife pathogens by Dobson and Foufopoulos (2001) found that one common trait of many emerging pathogens is the ability to infect a wide range of hosts. A non-discriminatory vector such as *A. americanum* would similarly be advantageous for promoting the spread of a pathogen to previously uninfected hosts (Childs and Paddock 2003).

Among other human pathogens, this species is known to transmit *Ehrlichia chaffeensis*, the currently undefined agent of southern tick-associated rash illness (STARI), and *Francisella tularensis* (Estrada-Pena and Jongejan 1999, Anderson and Magnarelli 2008, Hopla 1955, Hopla 1953, Masters, Grigery and Masters 2008, Anderson, Greene, et al. 1992, Anderson, Sims, et al. 1993). The public health burden of lone star tick associated illness is likely to increase in the future, however the noted rise of *A. americanum* as a species of major public health concern is confounded by improvements in surveillance and diagnostics for tick-borne diseases (Childs and Paddock 2003). Nevertheless, changes in vector-reservoir-host dynamics and human interactions will potentially increase human risk of tick-borne diseases. Increases in white-tailed deer densities, expansion of the geographic distribution of *A. americanum* and increased human contact with vectors can all influence the landscape of lone star tick transmitted pathogens (Childs and Paddock 2003).

Already in many areas in the southeastern United States, the lone star tick is the most abundant nuisance tick, and the most frequently encountered by humans (Childs and Paddock 2003). It is notorious for its voracious biting of humans. Studies on parasitism of humans by ticks in the southern US show that bites are overwhelmingly (83-95% in some instances) attributed to *A. americanum* (Childs and Paddock 2003, Felz, Durden and Oliver Jr. 1996).

The abundance of adult *A. americanum* in Florida is seasonal with conflicting reports about when peak adult activity occurs. In the northwestern region of the state adult ticks are highly abundant from April through August (Cilek and Olson 2000), although the peak is reported as early as March for Florida (Allan, Simmons and Burrridge 2001). Studies on lone star tick abundance in nearby states found similar peaks in adult abundance from April to June and from May to July in Georgia and Missouri, respectively (Semtner and Hair 1973, Kollars Jr, et al. 2000). In central Florida, recent studies have found that temporal patterns of nymphal and larval activity mirror that of northern populations with peak activity in the fall (Allan, Simmons and Burrridge 2001). However, no studies were found that have collected adult or immature *A. americanum* in numbers great enough to describe temporal activities of southern Florida populations.

The local geographic distribution of *A. americanum* in Florida is unknown beyond crude county-level maps. The seminal work on North American tick distributions by Bishopp and Trembley (1945) currently provides the basis for The Centers for Disease Control and Prevention (CDC) published map of the regional distribution of the lone star tick and indicates all of Florida falls within the species' potential range. Springer et al. (2014) describe establishment of the species at the county level across the United States using historic data and report occurrence in nearly all counties in northern and central Florida by the 1960s. Subsequent records between 1960 and the 2010s include additional counties in south-central Florida, however, the authors caution that apparent recent establishment in counties not previously reported may be due increased surveillance rather than invasion of previously unoccupied areas (Springer et al. 2014). This is important to note as previous studies indicate that the prevalence of the species diminishes from north to south in the state (Allan, Simmons and Burrridge 2001), with the

southernmost reports of adult individuals occurring in Palm Beach Co., Glades Co., and Miami-Dade County (Taylor 1951, Greiner, et al. 1984). These findings however appear to be host-specific as the southernmost collection of *A. americanum* from big cats was from central Florida in Highlands Co. (Wehinger, Roelke and Greiner 1995), while collections from feral swine and white-tailed deer in the southern portions of the state were inconsistent. The reports from Palm Beach, Glades, and Miami-Dade Counties constitute a single adult collected by Allan et al. (2001), 3 by Greiner (1984) and those by Taylor (1951) from deer or swine. Other collections from south Florida deer have not recovered *A. americanum* (Smith Jr. 1981). The reasons for presumptively rare occurrences of *A. americanum* in southern Florida is unclear although Allan et al. (2001) posited that climate, vegetation, and host abundance may play a roll.

2.3 Species Distribution Models

There are several statistical and pattern matching approaches to estimating the distribution of organisms in the environment. These techniques can range from purely theoretical to mechanistic or empirical (Guisan and Zimmermann 2000). Predictive species distribution models (SDMs), are designed to detect non-random associations between location-specific occurrence observations and environmental conditions (often as raster-based predictors) to estimate a response at unobserved locations (Guisan and Zimmermann 2000, Blackburn 2010). Broadly speaking, they are classed as either estimating occurrence of a species or abundance of a species in relation to the environment and rely on presence-only, presence-absence, or count data.

Species distribution models are a commonly employed technique in ecology, biogeography, and other fields to estimate the habitat suitability for many species and to describe distributions of disease or disease causing agents (Austin, Nicholls and Margules 1990, Pearson, Raxworthy, et al. 2007, Guisan and Thuiller 2005, Springer, et al. 2015). Species distribution

model are frequently divided into two approaches: presence-only models, and presence-absence models. The primary distinction between these two approaches stems from the structure of the data available (Elith, et al. 2011). Commonly used presence-only methods include genetic algorithms, and maximum entropy models (Stockwell 1999, Phillips, Anderson and Schapire 2006). A presence-absence data structure- the combination of occurrence and (thorough) absence data lend themselves to regression techniques such as generalized linear models (GLM). Other methods such as classification trees are also used to fit presence-absence data, but will not be discussed here (Elith, et al. 2011, Guisan, Edwards and Hastie 2002).

2.3.1 Generalized Linear Model Framework for SDMs

Generalized linear models include the well-known logistic regression models, for describing simple presence/absence, and Poisson or negative-binomial structured count models describing frequency or abundance of occurrence (Guisan, Edwards and Hastie 2002). Broadly, GLMs attempt to apply a single, linear combination of coefficients and corresponding predictor variables to describe a response variable while minimizing unexplainable error (Guisan, Edwards and Hastie 2002, Potts and Elith 2006, Guisan and Zimmermann, *Predictive Habitat Distribution Models in Ecology* 2000). In ecological terms, this means that the observed response, that is presence or absence, is limited to a fixed directional relationship with each predictor as indicated by the magnitude of the regression coefficient. Conceptually the output of a SDM, a habitat suitability surface, is an estimation of where the combination of environmental conditions are closest to optimum within the species' niche conditioned on the considered environment (Elith, et al. 2011). In other words, the suitability surface is a probability estimate of presence or absence in a pixel.

The geographic distributions of both vagile and non-vagile species have been estimated by GLM methods successfully. Guisan et al. (1998) used logistic regression to model the

distribution of an alpine plant species; Zimmermann and Breitenmoser (2002) modeled the distribution of the Eurasian Lynx; the relationship between landscape pattern and south Australian bird species was evaluated by Westphal et al. (2003); Brown et al. (2011) estimated high risk areas for exposure to the lone star tick (*A. americanum*) in Missouri. Additional applications for GLMs include distributions of plant and animal communities (see Guisan et al., 2002 for examples).

2.3.2 Estimating Geographic Distributions of Ixodid Ticks

Many studies have sought to estimate the potential distribution and ecological preferences of Ixodid ticks using climatic or environmental variables at a variety of spatial scales (Springer, et al. 2015, Estrada-Pena 1998, James, et al. 2015, Raghavan, et al. 2016). Advances in modeling techniques has spurred the application of a number of methodological approaches or ensembles of multiple modeling approaches (Guisan and Thuiller 2005, Elith, et al. 2011). The general agreement upon variables associated with tick presence and the resulting spatial predictions across many of these studies supports the notion that the models are capturing biologically relevant phenomena (Springer, et al. 2015, Ostfeld and Brunner 2015).

Due to the direct influence of climatic conditions on the physiological processes and behavior of ticks, variables characterizing temperature and precipitation patterns are frequently included in SDMs for ticks. For example, Porretta et al. (2013) estimated the current and future distribution of *Ixodes ricinus* in Europe using the MaxEnt modeling algorithm. The presence-only data was composed of occurrence records from across the known range of the species. Bioclimatic variables from the WorldClim database (Hijmans, Cameron, et al. 2005) were used providing a resolution of roughly 5km. These variables are derived from average monthly precipitation and temperature data and capture seasonal and annual trends. The minimum temperature of the coldest period and precipitation of the driest quarter provided the greatest

contribution to the model. The predicted area of climatic suitability encompassed the distributions inferred from other studies lending credence to the appropriateness of using climatic variables to determine geographic ranges for Ixodid ticks.

Springer et al. (2015) applied an ensemble modeling approach to examine the present and future distribution of *A. americanum* at the county level for the contiguous United States under both current and projected future climate conditions. An ensemble model approach utilizes multiple modeling algorithms to identify areas of agreement or disagreement (Araujo and New 2006, Springer, et al. 2015). Published presence records for *A. americanum* in the contiguous US going back to 1898 were aggregated to the county level, representing establishment of the species in 651 counties across 32 states. A total of 66 climate variables were generated to represent the average climatic conditions in each county. Selection of present day models by each modeling approach: general linear model, boosted regression trees, maximum entropy, random forest, and multivariate adaptive regression spline, resulted in between one and nine climate variables showing significant predictive ability. Overwhelmingly, in each model mean vapor pressure in July had the greatest explanatory power. Other included variables were annual precipitation, mean growing degree days in October, mean temperature of driest quarter, mean snow days in October, precipitation seasonality, mean temperature of wettest quarter, mean growing degree days in February, and mean diurnal range. The authors concluded that summertime humidity is highly associated with determining habitat suitability. In addition, the inclusion and varied relationships of additional variables in one or more models suggest that at the continental scale, different ecological pressures may drive the distribution of *A. americanum* outside their core range in the south central and southeastern region of the US.

The geographic distribution of the American dog tick (ADT), *Dermacentor variabilis*, has also been estimated using the MaxEnt method at the continental scale. The study by James et al. (2015) utilized the same bioclimate variables dataset as Porretta et al. (2013) as well as additional habitat metrics including monthly average measures of vegetation greenness (Normalized Difference Vegetation Index, NDVI), and elevation. Presence data were drawn from 317 observations of *D. variabilis*. Several possible distributions were estimated using combinations of uncorrelated variables. The set of considered MaxEnt models indicated that while several climate variables were still important predictors of suitable habitat, other environmental variables including elevation and summer (July) NDVI produced the greatest contribution to the models. Furthermore, while climate features may drive the broad distribution patterns, the inclusion of habitat specific characteristics may improve the resolution of the analysis (1km x 1km pixels). For example, the ADT is associated with habitat conditions found at lower elevations (Bishopp and Trembley 1945) so the inclusion of climate and elevation variables could improve the model's ability to discern habitat suitability in areas with significant topographical variation. One limitation of this study results from the mismatch between the spatial resolution of the geographic prediction and the accuracy or precision of observation locations; observation precision ranged from GPS coordinates to city centroids. The results are potentially biased associations between observations and environmental variables.

An autologistic framework has been used to estimate the distribution of *Ixodes scapularis* for the coterminous United States (Brownstein, Holford and Fish 2003). Known occurrences of *I. scapularis* at the county level were regressed against climatological data for minimum, mean, max and standard deviations of monthly minimum, maximum and average temperature averages, and vapor pressure at a resolution of 0.5°. Higher order polynomial terms for these variables

were also considered. Brownstein et al. found *I. scapularis* was associated with average maximum temperature up to a 4th order polynomial, and vapor pressure. These findings indicate that temperature extremes likely play a part in determining the distribution of the black legged tick in the United States.

At the state level, MaxEnt has been used to estimate the distribution of *A. americanum* in Kansas at a resolution of 10' (~18x18 km) (Raghavan, et al. 2016). Climate and environmental characteristics were evaluated with principal component analysis before being considered in the model. At this scale, variables representing temperature and soil moisture provided the largest contribution to the model although variables for precipitation were also important (Raghavan, et al. 2016). Notably, the estimated distribution extended further west than has been suggested by the CDC, and is in general agreement with the prediction by Springer et al. (2015) (Raghavan, et al. 2016).

Elevated risk to *A. americanum* adults and nymphs (separately) has also been predicted in Missouri using logistic regression (Brown, et al. 2011) at a resolution of 30m x 30m. Elevated risk was defined as collection rate of ticks above the median rate across the study area. Relative humidity and habitat type (forest vs grassland) were significant predictors *A. americanum* nymphs and adults. Forested area was found to be highly favorable to both nymphs and adult lone star ticks.

The purpose of this current study was to determine the relative importance of environmental and climatic variables in determining the presence of adult *Amblyomma americanum* in Florida. A species distribution modeling approach developed using a logistic regression framework to estimate the probable occurrence of *A. americanum* was implemented to

predict the spatial potential distribution of the species across Florida at a resolution of one hectare.

2.4 Materials and Methods

2.4.1 Study Area

The surveyed area for this study encompassed mainland Florida and excluded the Florida Keys. The state includes a wide range of environmental conditions with high climatic variability from north to south, which supports numerous habitats. The National Oceanic and Atmospheric Administration (NOAA) divides the state into seven distinct regions considered climatically homogeneous (Keim, Fischer and Wilson 2005). Both the El Niño Southern Oscillation (ENSO), and the Atlantic Multidecadal Oscillation (AMO) impact long-term climate trends in Florida however the dominant drivers include latitude (as reflected in the NOAA climate divisions), prevailing winds and pressure systems, as well as ocean currents (Tsai, Southworth and Waylen 2014). Northerly regions of Florida experience greater seasonal variability and receive a larger proportion of rainfall during the winter months than southern regions. The seasonal precipitation patterns are reflected in patterns of changing greenness (using the Normalized Difference Vegetation Index or NDVI) wherein greenness tends to increase in the fall and winter and decline in the spring (Tsai, Southworth and Waylen 2014).

Sampled habitats ranged from monoculture agricultural crop fields, seasonal wetlands, hardwood forests and pine scrub. Central and southern regions of the state contained a high percentage of seasonal fields and orchards, primarily citrus crops.

A set of 33 locations (henceforth called “sites”) were selected from various Florida State Parks, county extension service locations, and University of Florida affiliated Institute of Food and Agricultural Science research and education centers (IFAS-REC) (Figure 2-1). These sites

were selected to be widely distributed geographically and to incorporate variability in local environmental and climatic conditions.

2.4.2 Tick Collection and Environmental Data Sources

Observed presence or absence of adult *Amblyomma americanum* was determined from collections performed on transects from each site between September 2015 and September 2016 using flagging. Flags consisted of a 1 m² white flannel cloth attached to a pole which could be detached for storage and cleaning. All tick collections were performed in accordance with the Florida Department of Environmental Protection under Research and Collection Permit #03111610. Each site was visited a minimum of twice during the collection period, with most sites being repeatedly sampled approximately once per month between April and September. At each of the 33 sites pairs of 100m-200m transects were performed through primary areas of homogeneous land cover types.

Transects were located by latitude and longitude with the World Geodetic System 1984 (WGS84) using a GPS-enabled Nexus tablet at the start, middle, and end points of the transects using the DoForms application for Android to input data. Each GPS tagged location was time stamped and accompanied by a photograph of the surrounding land cover including canopy and understory for ground-truthing of land cover coverages. A minimum of one set of paired transects were run at any given site although sites with multiple dominant land cover types included upwards of five pairs of transects (10 total transects at the site). Each transect was given a unique identifier by county, site, date, and transect number. As sites were sampled multiple times throughout the study period transects repeated at the same location were given the same transect number although these transects may not exactly include the same start and stop points. Because care was taken to limit each transect to a single land cover type there was variability in the length of transects between sites.

The flags were checked for ticks approximately every 10m along a transect and specimens were removed and placed into vials containing 100% EtOH. Flagging material was placed into plastic bags after each transect so that any remaining tick larvae and nymphs could be removed later without cross contaminating subsequent transects. Ticks were identified morphologically by species and life stage based on standard taxonomic keys (Keirans and Litwick 1989, Strickland, et al. 1976) and adults and nymphs were stored at -80°C. Presence of larval ticks were noted and identified to suspected genus although no specimens were retained.

The environmental and climatic variables included in this analysis were selected primarily based on reported associations with tick presence/absence in the literature (Table 2-1) (Springer, et al. 2015, James, et al. 2015, Fryxell, et al. 2015). Variables describing environmental parameters were land cover, Normalized Difference Vegetation Index (NDVI), elevation, and distance to waterbodies. These datasets were retrieved from the Florida Geographic Data Library (FGDL) and derived from the Florida Cooperative Land Cover Database (CLC), MODIS NDVI composites, the ASTER Global DEM, and the national hydrography dataset, respectively. Variables describing climate parameters are the 19 “bioclimatic variables” (BioClim) calculated by Hijmans et al. (2005) from the WorldClim Version 1.4 climate database and use the same naming schema.

The CLC Database includes statewide classifications of all major land cover types at a native resolution of 10m. Classification of land cover types is determined at both the state level and at site level which includes additional classifications for certain areas. Using the existing hierarchical classification scheme major state-level land cover types were reclassified as one of five primary types: Forest, which includes pine and hardwoods; Shrub, encompassed shrub and

brush lands; Grasses; Wetlands; and a final, general category including all other land types such as water bodies and urban areas (Table 2-2).

Table 2-1. Characteristics of variables

| Variable | Dataset | Resolution (m) | Measure |
|---|--------------------------|----------------|------------|
| Forest Cover | CLC v3.0 | 10 | binary |
| Shrub Cover | CLC v3.0 | 10 | binary |
| Grass Cover | CLC v3.0 | 10 | binary |
| Wetland Cover | CLC v3.0 | 10 | binary |
| Other Cover | CLC v3.0 | 10 | binary |
| Elevation | ASTER global DEM | 30 | Continuous |
| Minimum annual NDVI | Modis | 250 | Continuous |
| Maximum annual NDVI | Modis | 250 | Continuous |
| Average annual NDVI | Modis | 250 | Continuous |
| Mean Temperature [Bio1] | WORLDCLIM | 1000 | Continuous |
| Mean diurnal temperature range (mean(period max-min)) (°C) [Bio2] | WORLDCLIM | 1000 | Continuous |
| Isothermality (Bio02 ÷ Bio07) [Bio3] | WORLDCLIM | 1000 | Continuous |
| Temperature seasonality (Coeff of Variation) [Bio4] | WORLDCLIM | 1000 | Continuous |
| Max temperature of warmest month (°C) [Bio5] | WORLDCLIM | 1000 | Continuous |
| Min temperature of coldest month (°C) [Bio6] | WORLDCLIM | 1000 | Continuous |
| Temperature annual range (Bio05-Bio06) (°C) [Bio7] | WORLDCLIM | 1000 | Continuous |
| Mean temperature of wettest quarter (°C) [Bio8] | WORLDCLIM | 1000 | Continuous |
| Mean temperature of driest quarter (°C) [Bio9] | WORLDCLIM | 1000 | Continuous |
| Mean temperature of warmest quarter (°C) [Bio10] | WORLDCLIM | 1000 | Continuous |
| Mean temperature of coldest quarter (°C) [Bio11] | WORLDCLIM | 1000 | Continuous |
| Annual Precipitation [Bio12] | WORLDCLIM | 1000 | Continuous |
| Precipitation of Wettest Month [Bio13] | WORLDCLIM | 1000 | Continuous |
| Precipitation of Driest Month [Bio14] | WORLDCLIM | 1000 | Continuous |
| Precipitation seasonality (Coeff of Variation) [Bio15] | WORLDCLIM | 1000 | Continuous |
| Precipitation of wettest quarter (mm) [Bio16] | WORLDCLIM | 1000 | Continuous |
| Precipitation of driest quarter (mm) [Bio17] | WORLDCLIM | 1000 | Continuous |
| Precipitation of warmest quarter (mm) [Bio18] | WORLDCLIM | 1000 | Continuous |
| Precipitation of coldest quarter (mm) [Bio19] | WORLDCLIM | 1000 | Continuous |
| Distance to water | Natl Hydrography Dataset | 1000 | Continuous |

The five selected categories potentially capture biologically relevant relationships between ticks and their ecological preferences. Areas dominated by forested landscapes provide a multitude of conditions beneficial for ticks; forests provide habitat and protection for potential hosts and increased leaf litter and canopy cover produce microclimates with increased humidity and cooler midday temperatures. Similarly, shrub dominated landscapes provide food sources and suitable microclimates for ticks. A study by Ostfeld et al. (1995) showed that shrubby areas

supported high *I. scapularis* density surpassed only by oak and maple forests. The authors posit that hosts using these areas for food and shelter played a role in tick numbers. Furthermore, a study by Allan et al. (2010) indicates that areas dominated by brushy vegetation (the invasive honeysuckle species *Lonicera maackii*) had higher counts of nymphal *A. americanum* compared to more sparsely growing native vegetation, and that removal of the species reduced overall nymph numbers. Grassland areas likely provide some contrast with areas of moist microclimates. *I. scapularis* are unlikely to inhabit open grasslands even though it provides a thick layer of moisture providing chaff (Ostfeld, Cepeda, et al. 1995). *A. americanum* are generally considered fairly drought tolerant, which may permit colonization of grassland ecosystems that would be otherwise too xeric to support other tick species. Wetlands in Florida consist of areas temporarily or permanently inundated with water throughout the year. Permanently submerged areas are unlikely to harbor permanent tick populations however; areas with transient water or seasonal flooding may allow populations to persist during times of no water. Weiler et al. (2017) found that after a flood event of the Danube River, *Ixodes ricinus*, *Haemaphysalis concinna*, and *Dermacentor reticulatus* were recovered after approximately 2 weeks of submersion albeit in lower numbers compared to pre-flood collections. Under laboratory conditions adult lone star tick has been shown to survive submersion for extended periods (LT50: 3.3-3.5 weeks) with nymphs survival extending far beyond that time (Koch 1986). The final land cover classification considered in the study encompassed all remaining areas that would be inhospitable to lone star tick populations. These areas consisted primarily of water bodies and urban areas. Several miscellaneous land cover classifications that encompassed negligible area were included in this category as they were not obvious fits for other categories.

Table 2-2. Reclassification scheme for land cover variables

| CLC Land Cover Type | CLC State code | Reclassified LC |
|----------------------------------|----------------|-----------------|
| Hardwood Forested Upland | 11 | |
| High Pine and Scrub | 12 | Forest |
| Pine Flatwoods and Dry Prairie | 13 | |
| Mixed Hardwood-Coniferous | 14 | |
| Shrub and Brushland | 15 | |
| Coastal Uplands | 16 | Shrub |
| Cultural-Terrestrial | 180 | |
| mowed grass | 181 | Grass |
| rural | 183 | |
| Palustrine | 20 | |
| Freshwater Non-forested Wetlands | 21 | |
| Freshwater Forested Wetlands | 22 | Wetlands |
| Non-vegetated Wetland | 23 | |
| Altered Wetlands | 24 | |
| Lacustrine | 30 | |
| Natural Lakes/Ponds | 31 | |
| Artificial Lakes/Ponds | 32 | |
| Riverine | 40 | |
| Natural Rivers and Streams | 41 | |
| Cultural-Riverine | 42 | |
| Estuarine | 50 | |
| Subtidal | 51 | |
| Intertidal | 52 | |
| Cultural-Estuarine | 53 | |
| Marine | 60 | |
| Surf Zone | 61 | |
| Exotic Plants | 70 | Other |
| Australian Pine | 71 | |
| Melaleuca | 72 | |
| Brazilian Pepper | 73 | |
| Exotic Wetland Hardwoods | 74 | |
| Unconsolidated Substrate | 91 | |
| Barren, Sinkhole, Outcrop | 17 | |
| urban | 182 | |
| transportation | 184 | |
| communication | 185 | |
| utilities | 186 | |
| extractive | 187 | |
| Bare Soil/Clear Cut | 188 | |

Maximum, minimum, and average normalized difference vegetation index (NDVI) were calculated for the study period from MODIS NDVI 16-Day composites. NDVI is a frequently considered metric when estimating the distribution or ecological niche of ticks. It is considered a proxy for moisture availability i.e. high humidity microclimates produced by high vegetation density or leaf litter. However, there is some dispute about the ability of NDVI to produce meaningful predictions of tick presence or population density. Bisanzio (2008) found that the correlation between NDVI and *Ixodes ricinus* was generally positive although the strength of the correlation varied throughout the year and between years. This variability indicates that the usefulness and specification of NDVI metrics should be considered carefully. Studies of mosquito (*Aedes* and *Culex spp.*) populations have also found variable correlations with NDVI, however due to the very different biology between these invertebrates, the underlying reasons are likely very different (Britch, et al. 2008, Linthicum, et al. 1990). Furthermore, the geographic extent of the study area produces significant variability in green-up time and NDVI patterns as noted by Tsai et. al. (2014). These factors precludes the use of a single statewide NDVI 16-day composite to represent maximum or minimum NDVI. In order to address potential spurious geographic or temporal intra-annual and inter-annual variation or correlation in this study, the absolute maximum, minimum and average NDVI values were calculated by pixel for a five-year period preceding and including the study period.

The bioclimatic variables from the WorldClim dataset were calculated from interpolated surfaces of average maximum and minimum temperature and precipitation data from 1960-1990 (Hijmans, Cameron, et al. 2005). This dataset is widely used in ENM and SDM applications. Porretta et al. (2013) and James et al. (2015) considered these variables in estimating the distributions of *I. ricinus* and *D. variabilis*, respectively. The base variables of maximum, and

minimum temperatures and precipitation can have direct biological impacts on biology and physiology (Randolph, et al. 2002, Ogden, et al. 2005). High and low temperatures may reduce questing activity and survivability. High precipitation, especially in low-lying areas can impact survivability of eggs (Weiler, et al. 2017, Koch 1986, James, et al. 2015). The bioclimatic variables derived from these base variables may have other biologically significant relationships with tick occurrence (James, et al. 2015).

2.4.3 Methodology to Account for Spatial Autocorrelation

The sampling procedure used to collect ticks at each of the 33 collection sites results in a high density of unique transects within a very small area (<1km). The clustered nature of the presented study design lent itself to two issues that are common in species distribution models: sampling, or observation, bias and spatial autocorrelation (Guisan and Zimmermann 2000). Spatial autocorrelation is a statistical property inherent in many ecological processes in which paired observations in geographic space are more or less similar than expected for randomly associated observations (Legendre 1993, Dormann, et al. 2007). In regression analyses and statistical testing the presence of positive or negative spatial autocorrelation can violate the assumption of independence among observations (Legendre 1993).

The procedure of repeatedly sampling each site over time results in multiple, proximal points with similar counts of adult *A. americanum* and therefore a high degree of spatial autocorrelation of tick presence or absence between transects. To evaluate the impact of this sampling scheme, and to evaluate spatial autocorrelation, semivariogram analysis of *A. americanum* abundance was performed to establish the distance at which the number of collected adult specimens was independent of findings at nearby transects (the range of the semivariogram) (Legendre 1993). The average nearest neighbor distance was calculated from the midpoints of all transects and used to set the bin width in the semivariogram. The resulting range

of the semivariogram indicates independence of observations from neighboring observations i.e. where the semivariance between observations stops increasing (Karl and Maurer 2010).

To account for spatial autocorrelation at distances less than the calculated range, all transects for which the GPS designated midpoint were located closer together than the semivariogram range distance were collapsed to a single central point in ArcGIS. The resulting point, was designated as having presence or absence (1 or 0) of *A. americanum* if adult specimens were collected from one or more transects represented by that point. Transects for which the midpoint was greater than 111m from another transect were retained as individual points.

All environmental and climatic data layers were resampled to 100m x100m resolution to match the spatial scale of the aggregated transects and to preserve a large enough sample size for modeling. Resampling raster data to smaller cell sizes than the native resolution does not provide additional information from the raster. For example, resampling a 1 km x 1km cell with a value of two to a resolution of 100m x 100m will simply result in 100, 100m x 100m cells each with a resolution of two, which occupy the same extent as the original 1km² cell. The discrete values of the land cover data set were resampled using a nearest-neighbor method in ArcGIS. This method is well suited for discrete data as the original cell values are retained (Richards 2013). All other data were resampled to the extent and resolution of the land cover layer using bilinear interpolation in R (Richards 2013). All layers and data were then projected to an Albers Conic Equal Area projected coordinate system to minimize linear scale and areal distortion (U.S. Geological Survey 1987). ArcGIS and R were used to process spatial data, project the data in projected coordinate systems, perform interpolations and resample raster layers.

2.4.4 Multivariate Logistic Regression

In this study, a multivariate logistic regression was used to estimate the relationship between the presence or absence of *A. americanum* and the set of environmental covariates. The standard form of the multivariate logistic regression is expressed in the form of Equation 2-1 (Hosmer and Lemeshow 1989):

$$P(Y = 1|x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

Where (2-1)

$$g(x) = \beta_0 + \sum_{j=1}^{k-1} \beta_j X_j$$

The dependent variable, Y, is the conditional probability of the outcome being present. In the context of this study, Y is the probability of *A. americanum* being present at a specific location. The logit function, g(x), is the linear combination of the variables and their coefficients, X_j and β_j, respectively (Guisan, Edwards and Hastie 2002).

Once the associated variables are determined and their covariates are estimated, the probability of species presence or absence can be predicted for any location on the landscape using Equation 2-2:

$$Y = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}} \quad (2-2)$$

Due to the relatively small sample size, the logistic regression model was validated using k-fold cross validation. Procedures for model building, selection, and validation are described below.

The model building and selection processes for the logistic regression were performed on centered and standardized continuous variables. The centering and Z-score standardization of continuous variables provided a uniform basis by which the magnitude and direction of a

covariate's influence on the dependent variable can be compared across models (Bring 1994).

Each continuous variable considered was centered using equation 2-3:

$$X_{j,centered} = x_j - \frac{\sum_{j=1}^k x_j}{k} \quad (2-3)$$

Where the centered values of covariate X are the difference of the column mean from each element of covariate X. The result is a distribution of values centered on zero. Similarly, the resulting centered variables are scaled using equation 2-4 (Bring 1994):

$$X_{j,scaled} = \frac{x_{j,centered}}{\sqrt{\left(\frac{\sum_j x_j^2}{k-1}\right)}} \quad (2-4)$$

When performed on centered data, as done here, the scaling function results in data centered on zero with a standard deviation of one (Bring 1994). This allows for greater interpretability of regression coefficients (Schielzeth 2010). Selection of variables for consideration in the final model search was performed in two parts. The R code written for model building is included in Appendix A.

As part of an initial exploratory data analysis, univariate relationships were evaluated to identify the set of covariates with a significant ($p < 0.05$) associations with the dependent variable (presence/absence of *A. americanum*) (Springer, et al. 2015). The restriction of considered variables during this step reduced the computation time of model searching while also limiting the possibility of misspecification of the final model that included unimportant variables (Chatterjee and Hadi 2013).

Many of the remaining variables considered in this analysis are derived from precipitation and temperature measurements. As a result, many of those spatial patterns described by these variables were highly correlated (Springer, et al. 2015). To reduce the likelihood of introducing collinearity into the model, a stepwise procedure was implemented in R

using the 'usdm' package to calculate the variance inflation factor (VIF) for the remaining predictors (Naimi, et al. 2014). The VIF is calculated from the square of the multiple correlation coefficient determined by regressing a given variable against all other variables (Chatterjee and Hadi 2013). The higher the VIF, the greater the correlation between the selected response variable and all other covariates (Graham 2003, Craney and Surles 2002, Chatterjee and Hadi 2013). A VIF with a value greater than 10 was considered a sign that the model had a collinearity problem (Chatterjee and Hadi 2013), i.e. the selected response variable was highly collinear with the other predictors. The variable with the highest VIF was excluded and the VIF was recalculated for all remaining variables. This process was repeated until no variables remained with a $VIF > 10$ (Naimi, et al. 2014, Craney and Surles 2002). A major criticism of using stepwise regression for selecting model variables is that the order of parameter entry or deletion can affect the parameterization of the final model (Whittingham, et al. 2006, Derksen and Keselman 1992). However, this VIF stepwise procedure is largely unconcerned with the predictor variables' relationship with the dependent variable as VIF relates to how much variability in each independent variable is explained by the other independent variables (Craney and Surles 2002, Graham 2003). As a result, the sole purpose of this procedure is the minimization of collinearity among predictors, and not specifically the development of a variable set that best fits the response variable (Graham 2003). The remaining uncorrelated variables were evaluated with an exhaustive logistic regression model search to select a 'best' model.

2.4.5 Model Evaluation

Evaluation of model performance is the most important step in developing a useful model (Symonds and Moussalli 2011). Both the statistical assumptions and predictive performance must be assessed to ensure that model is suitable and well characterized for its intended purpose. Although logistic regression relaxes many of the statistical assumptions of linear regression,

remaining assumptions such as absence of collinearity among predictors and independence of error structure must be met (Hosmer and Lemeshow 1989). In addition, the predictive performance must also be evaluated to ensure that it is reliable and discriminatory.

The identification of significant, non-collinear variables resulted in a subset of continuous variables and binary categorical variables used for final model selection. The final, “best model” was determined using the “glmulti” package in R to perform an exhaustive model search of first order variable interactions. Exhaustive model searches fit a set of models using every combination of covariates including a model with every covariate included, and one with no covariates. With n covariates, the result is 2^n possible models (Chatterjee and Hadi 2013). Exhaustive model searches can be time, and computationally intensive: in selecting a best model from a set of 6 possible covariates, 2^6 (64) models must be estimated; if another two covariates are considered, the number of possible model combinations balloons to 256 (2^8). The use of standardized variables allowed for evaluation of model quality based on the corrected Aikake information criterion (AICc). This measure compares a model relative to a suite of other potential models and a ‘best’ model selected based on the minimum AICc among the set of models (Chatterjee and Hadi 2013, Symonds and Moussalli 2011). These models represent the set of possible variable combinations that pass the statistical assumption of non-collinear predictors.

The regression residuals of the selected best model were evaluated for spatial autocorrelation using Global Moran’s I. One assumption of logistic regression is the independent and identical distribution of residuals. To determine if spatial dependence in the underlying data structure was adequately controlled by aggregating the data points the spatial distribution of model residuals was assessed for clustering (Franklin, et al. 2009). The equation for the Global

Moran's I statistic is given by equation 2-5 and can be evaluated by z-score and associated p-value, provided in equation 2-6 (Anselin 1995).

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2} \quad (2-5)$$

The associated z-score is computed as:

$$Z_I = \frac{I - E[I]}{\sqrt{V[I]}} \quad (2-6)$$

The null hypothesis for the Global Moran's I states that the distribution of values is randomly distributed. Interpretation of the p-value returned by this statistic using a two-sided approach infers negative, positive, or no spatial autocorrelation. A non-significant p-value indicates the null hypothesis cannot be rejected whereas a significant p-value and positive or negative z-score indicates positive or negative spatial autocorrelation, respectively (Anselin 1995). If there is positive spatial dependence remaining in the model, as indicated by significant p-value and positive z-score, it could be indicative of a poorly defined model (missing explanatory variables) or that a spatial lag or an autologistic error term should be included (Legendre 1993, Miller, Franklin and Aspinall 2007).

The Hosmer-Lemeshow test was used to evaluate the goodness of fit for the logistic regression. This goodness of fit test estimates fit based on the values of estimated probabilities (Hosmer and Lemeshow 1989). Grouping values based on percentiles is a common strategy, employed here. The test statistic \hat{C} is obtained by calculating the Pearson chi-square statistic from the observed and expected frequencies. The calculation for \hat{C} is given in equation 2-7:

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \pi_k)^2}{n'_k \pi_k (1 - \pi_k)} \quad (2-7)$$

Where n'_k is the number of covariate patterns in the k^{th} group, o_k is the number of responses among the n'_k covariate patterns, and π_k is the average estimated probability. The

distribution of the Hosmer-Lemeshow statistic approximates that of the chi-square with $(g-2)$ degrees of freedom where g is the number of groups or deciles. If the p-value determined from the chi-square table is large we can accept the assumption of good fit in the model. If the p-value is very small e.g. less than 0.1 or 0.05, the model is poorly fit (Hosmer and Lemeshow 1989).

Due to the relatively small number of positive points in the dataset ($n=23$), k-fold cross validation was selected to estimate the model's predictive performance as an alternative to further reducing the effective size of the dataset by dividing into training and testing sets. Cross validation (CV) and its variations are hailed for their applicability to a variety of algorithms and frameworks because they only assume an i.i.d. of data, and that the partitioned data are independent (Arlot and Celisse 2010). To improve accuracy measures, repeated k-fold CV was performed using $k=10$ folds with 100 repetitions. In this form of CV, the procedure follows that of simple k-fold CV in which data was divided into k subsets (folds). Model training uses $k-1$ folds, and predictions are performed on the fold that was withheld (Arlot and Celisse 2010). This procedure was repeated until each of the 10 folds had been withheld and probability estimates had been determined. This procedure was repeated 100 times in which the folds were resampled each repetition to ensure unbiased training and testing folds (Arlot and Celisse 2010). It should be noted that when available, the use of a truly independent dataset (i.e. a unique set of locations sampled post model development) is preferable to splitting or folding data, especially in cases where the training data may not be representative of the entire study region (Pearce and Ferrier 2000). The predictions from each fold are compared to the known values and the model is assessed using a measure of overall accuracy and the Kappa statistic. The Kappa statistic calculated using Equation 2-8 measured the agreement between the observed data and those predicted by the model.

$$\hat{K} = \frac{\text{Overall Accuracy} - \text{Chance Agreement}}{1 - \text{Chance Agreement}} \quad (2-8)$$

The statistic is standardized with a range of -1 to 1 where 1 indicates perfect agreement between the observed and predicted data and values less than zero indicate systematic disagreement. A value of zero is what would be expected by random chance (Allouche, Tsoar and Kadmon 2006). Fielding and Bell (1997) modify the Kappa statistic ranges proposed by Landis and Koch (1977) such that poor agreement: $K < 0.4$; moderate, $0.4 < K < 0.75$; and excellent agreement are $K > 0.75$.

The model was further assessed against an external dataset to which the model was naïve to check the model's generalizability spatially and temporally. The testing dataset consisted of 316 transects performed during a subsequent field season between October 2016 and May 2017 using the same collection procedures and sites (with an additional 4 sites) detailed in section 2.3.2. The model was used to generate probability estimates for this testing dataset and a variation of the binning scheme for the Hosmer-Lemeshow GOF test was performed as well as model accuracy metrics (sensitivity, specificity, accuracy), and Kappa values were determined.

The variation of on the Hosmer-Lemeshow groupings was used to determine if presence or absence could be easily discerned from the calculated probability estimates. The bins for Equation 2-7 are defined such that $g=10$ groups can be based on fixed values corresponding to deciles of estimated probabilities $k=1, \dots, 9$, or as done here, determined based on percentiles of estimated probabilities such that $n_1=n/10$ observations have the smallest estimated probabilities and the $n_{10}=n/10$ observations have the largest estimated probabilities (Hosmer and Lemeshow 1989). The ratio of positive and negative observations across the bins should ideally follow the curve of the logistic function, however a distribution tending towards dichotomous separation of observations is also indicative of high discrimination by the model (Pearce and Ferrier 2000).

The calculation of these metrics on an independent dataset evaluates how reliably the model can make predictions for unseen data and how well it can discern presence and absence in unseen data.

2.4.6 Spatial Predictions and Gradients

The geographic distribution of *A. americanum* was estimated by generating a map of predicted probabilities for occurrence based on the output of the logistic regression model. The raster layers for each variable included in the logistic regression model were stacked in R and output values of the logistic regression were estimated for every location on the landscape using the ‘predict’ function. The values of each pixel in the stack of raster layers at a given location provide the inputs for the logistic regression and outputs a probability value for that location with a range from 0-1. The closer the predicted probability is to one, the more likely that the environmental conditions at that pixel would be suitable for *A. americanum*. This method produced a statewide probability surface representing probable occurrence of *A. americanum* at a resolution of one hectare.

The potential distribution of *A. americanum* in Florida was then estimated by reclassifying the map of occurrence probabilities based on a specified cutoff criterion. The cutoff value used in this study was the probability value where sensitivity and specificity are equivalent. Sensitivity and specificity are measures of the percentage of actual presence values correctly predicted by the model, and the percentage of actual absence values correctly predicted by the model, respectively (Fielding and Bell 1997). By selecting a cutpoint where these two measures are equivalent, implies that positive and negative observations are likely to be predicted correctly with equal chance (Freeman and Moisen 2008, Fielding and Bell 1997). All probability values greater than the sensitivity/specificity cutoff are deemed suitable for presence

of adult *A. americanum*, and all values less than the cutoff are deemed unsuitable for adult *A. americanum*.

The existence of latitudinal gradients in the predicted distribution of occurrence was evaluated by both changes in the proportion of suitable habitat for *A. americanum* and by a reduction in the percentage of ‘highly suitable’ areas from north to south. Highly suitable was defined as a more conservative estimate of suitable habitat. Whereas suitable habitat was defined based on the LPT, ‘highly suitable’ was determined as probability estimates greater than or equal to the average probability estimate for the entire state. Although these two measures are similar, they provide information on two different questions about the spatial distribution of *A.*

americanum in Florida:

1. Is existence of suitable habitat distributed uniformly across the state?
2. Is suitability of predicted habitat distributed uniformly across the state?

The answers to these questions provide additional information about the risk of exposure to ticks in Florida.

The proportion of area predicted to be suitable for *A. americanum* using the LPT cutoff was determined in one-hectare wide bands extending east to west across Florida. However, the trend is presented here as a 10km wide average solely to smooth local variability for visualization purposes. The proportional area for each band was calculated from the summation of the number of cells above the cutoff (suitable areas) divided by the total number of cells within that band. Similarly, the trends in predicted probabilities were assessed by determining the percentage suitable area containing high predicted probability values. Using the same 10km bands, the total area with high probabilities, greater than the mean occurrence probability of ‘suitable area,’ as defined by areas with probability values greater than the LPT, was determined and divided by total suitable area within that band.

2.5 Results

A. americanum, *D. variabilis*, and *I. scapularis* which comprise three of the five most common pest ticks in Florida, were collected during this study (Table 2-3). The target species, *A. americanum* was the most commonly collected species followed by *D. variabilis*, and *I. scapularis*. The gulf coast tick, *Amblyomma maculatum*, and the brown dog tick, *Rhipicephalus sanguineus*, were not found. Adult ticks belonging to any of the identified three species were collected from 23 of the 33 (70%) sites. Adult *A. americanum* were collected from 33 of 328 unique transects (10%) spread across 13 of 33 (40%) sites, (Figure 2-1). Nymphal ticks were collected in far greater numbers than adults, accounting for 63% of all collected ticks. Large numbers of larval ticks were collected at several sites, however they were not counted or included in the results of this study. One site, Faver-Dyke state park accounted for approximately 50% of all adult *A. americanum* collected in this study. San Felasco state park produced the second highest percentage of adults (13%). Anecdotal evidence from correspondences with state park staff and officials at other sites support that these are hot spot areas; officials from other sites indicated ticks were far less of a problem.

Table 2-3. Specimen totals by site.

| Site Name | <i>Ixodes scapularis</i> | <i>Dermacentor variabilis</i> | <i>Amblyomma americanum</i> | <i>Amblyomma maculatum</i> | <i>Rhipicephalus sanguineus</i> | Nymphs |
|-----------------|--------------------------|-------------------------------|-----------------------------|----------------------------|---------------------------------|--------|
| Apopka | 0 | 0 | 0 | 0 | 0 | 0 |
| Astor | 1 | 1 | 4 | 0 | 0 | 9 |
| Big Talbot | 7 | 0 | 2 | 0 | 0 | 15 |
| Colt Creek | 0 | 1 | 0 | 0 | 0 | 0 |
| Everglades REC | 0 | 0 | 0 | 0 | 0 | 0 |
| Farles Lake | 0 | 2 | 6 | 0 | 0 | 3 |
| FaverDyke | 0 | 0 | 50 | 0 | 0 | 93 |
| Fore Lake | 0 | 1 | 8 | 0 | 0 | 6 |
| Ft Pierce | 0 | 0 | 0 | 0 | 0 | 0 |
| Gainesville | 0 | 0 | 0 | 0 | 0 | 0 |
| Jay | 0 | 0 | 0 | 0 | 0 | 2 |
| JD MacArthur | 0 | 0 | 0 | 0 | 0 | 0 |
| Juniper Prairie | 0 | 0 | 0 | 0 | 0 | 1 |
| Lake Kerr | 0 | 0 | 0 | 0 | 0 | 1 |
| Lovers Key | 0 | 0 | 0 | 0 | 0 | 0 |
| Marianna | 0 | 0 | 0 | 0 | 0 | 0 |
| Monticello | 0 | 0 | 3 | 0 | 0 | 21 |
| Myakka | 0 | 0 | 0 | 0 | 0 | 0 |
| Oleta | 0 | 0 | 0 | 0 | 0 | 0 |
| Ona | 0 | 1 | 0 | 0 | 0 | 0 |
| Paleo Hammock | 0 | 0 | 0 | 0 | 0 | 0 |
| Patrick AFB | 0 | 0 | 0 | 0 | 0 | 0 |
| Pine Castle | 0 | 4 | 3 | 0 | 0 | 4 |
| Quincy | 0 | 0 | 1 | 0 | 0 | 0 |
| Rodman | 0 | 0 | 0 | 0 | 0 | 0 |
| San Felasco | 0 | 0 | 13 | 0 | 0 | 35 |
| Satsuma | 0 | 0 | 5 | 0 | 0 | 17 |
| Sellers Lake | 0 | 0 | 1 | 0 | 0 | 1 |
| St Sebastian | 0 | 0 | 0 | 0 | 0 | 0 |
| SW REC | 0 | 0 | 0 | 0 | 0 | 0 |
| Tropical Rec | 0 | 0 | 0 | 0 | 0 | 0 |
| Tyndall | 0 | 0 | 1 | 0 | 0 | 0 |
| Yulee | 3 | 4 | 1 | 0 | 0 | 1 |
| Totals | 11 | 14 | 98 | 0 | 0 | 209 |

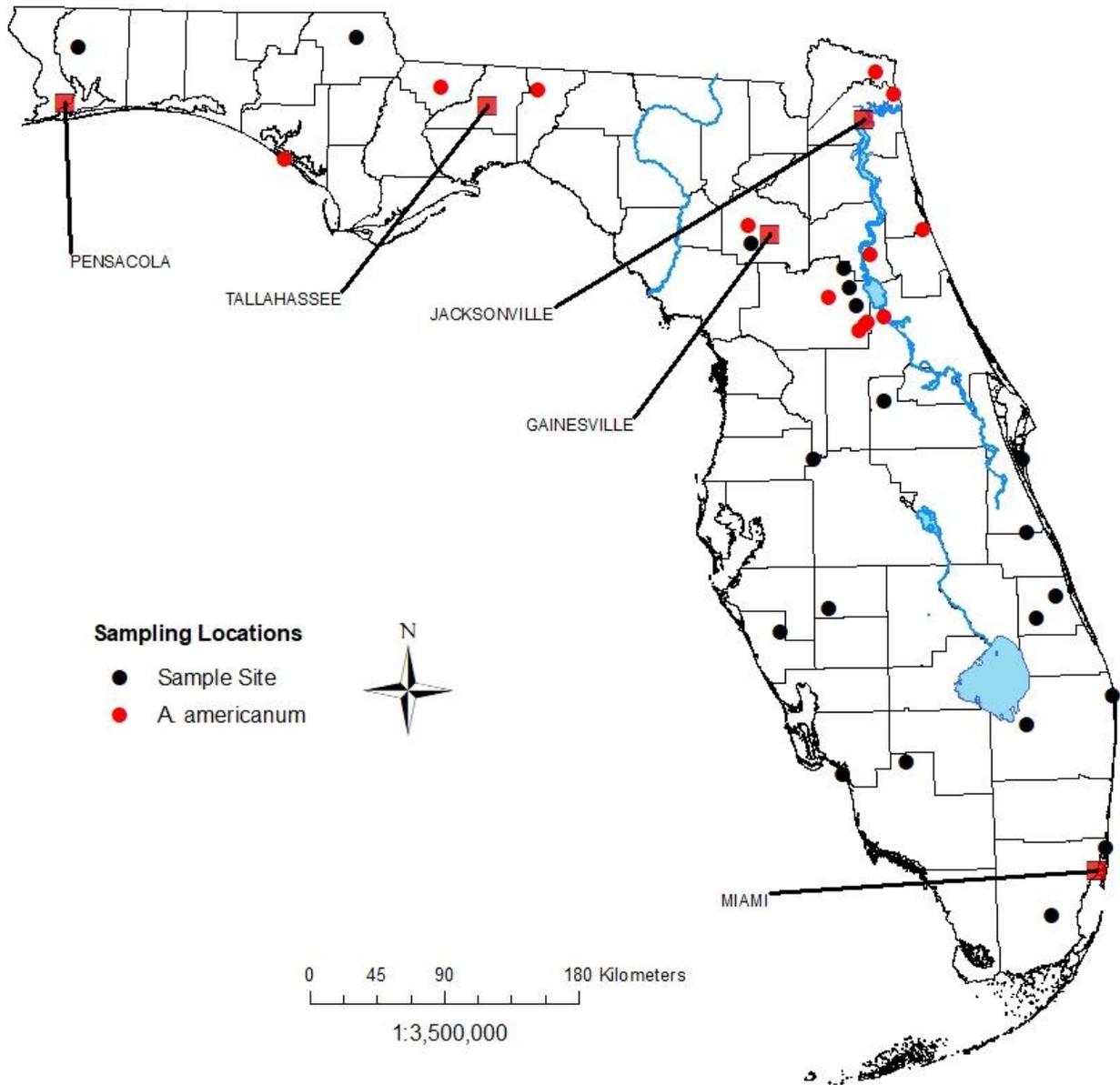


Figure 2-1: Distribution of sample collection sites from August 2015-August 2016 indicating where adult *A. americanum* were collected (red), and locations that returned no adult *A. americanum* (black).

The majority of nymphs were identified as belonging to the genus *Amblyomma* but were not included in the study as a specific identification could not be made. *Amblyomma* nymphs were collected from 14 sites including three locations from which no adult *A. americanum* were reported. In total, data from 328 individual transects surveyed from September 2015 through August 2016 were used in this study.

Results from the collection effort for this study indicate questing adult *A. americanum* occurred from April through September with a peak number of adults collected per sampling effort in June (Figure 2-2). The largest numbers of adult *A. americanum* were collected in April.

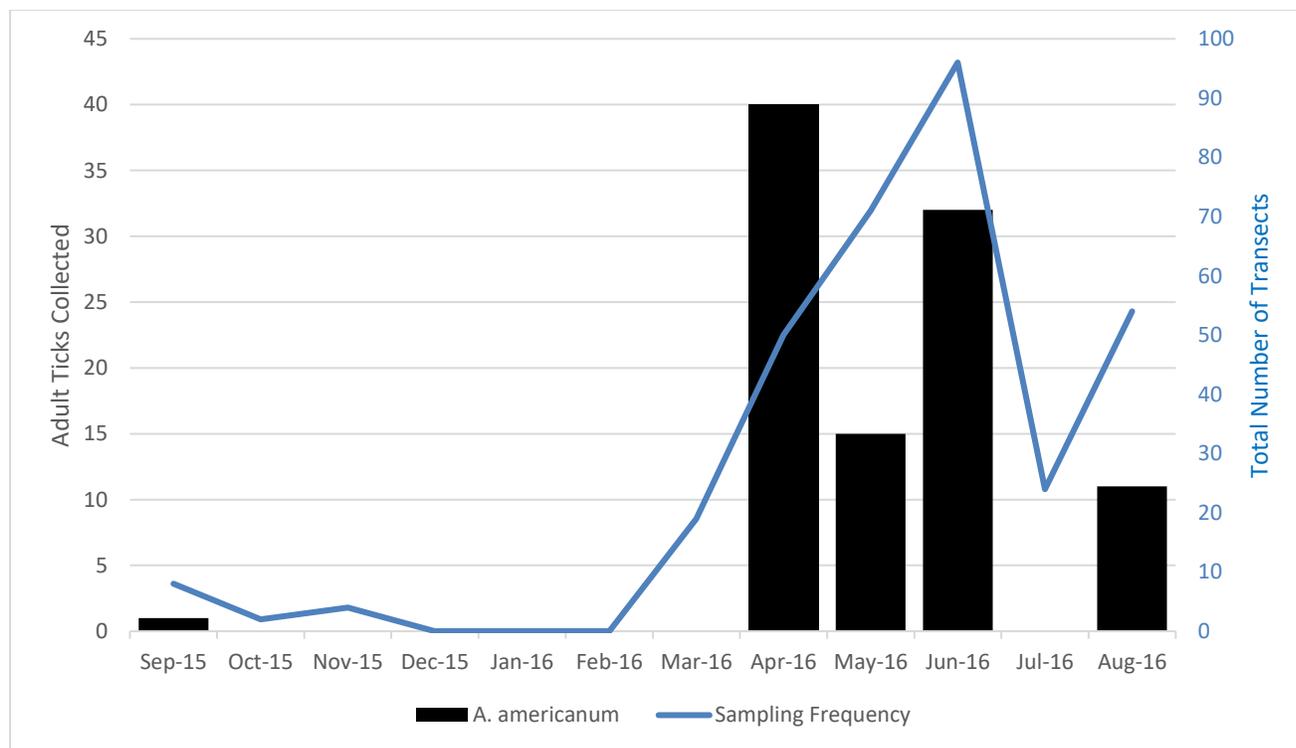


Figure 2-2: Monthly totals of adult *Amblyomma americanum* and collection intensity during the study period of September 2015 through August 2016. No collections took place during December, January, and February. The highest total number of adult ticks was collected during April, however the highest catch-per-unit-effort occurred in June.

Mean transect length was 134.2 meters with an interquartile range of 70.3 meters. Adult *A. americanum* were absent from all sites south of Ocala National Forest with the highest numbers collected from north central Florida and the central Atlantic coast.

The results from the semivariogram analysis indicated that variance between *A. americanum* counts increased up to approximately 110m, giving the effective range at which the number of ticks found on neighboring transect are independent of each other. This result was used to dictate the aggregation distance for the 328 transects. After collapsing neighboring transects within 110m of one another, the resulting dataset was composed of 130 points, with 23 points containing positively identified adult *A. americanum*.

Due to the relatively small number of positive locations, all points were used to build the logistic regression model. Results from the univariate analysis of significant relationships between tick presence and the environmental covariates indicated that 12 of the 28 considered variables provided no significant predictive ability ($p > 0.05$). These variables were removed from consideration for the remainder of the model building process.

Table 2-4. Environmental and climatic variables used to model the spatial distribution of *Amblyomma americanum* in Florida.

| Variable | P-Value |
|-----------|----------|
| Forest | 0.008204 |
| Bio1 | 0.004293 |
| Bio3 | 0.01259 |
| Bio4 | 0.006253 |
| Bio6 | 0.002727 |
| Bio7 | 0.002762 |
| Bio8 | 0.020901 |
| Bio9 | 0.001816 |
| Bio10 | 0.001376 |
| Bio11 | 0.004418 |
| Bio13 | 0.002655 |
| Bio15 | 0.045435 |
| Bio16 | 0.011945 |
| Bio18 | 0.009604 |
| Max NDVI | 0.000421 |
| Mean NDVI | 0.001037 |

The remaining 16 variables, which significantly predicted *Amblyomma americanum* presence in the logistic regression model, are listed in (Table 2-4).

The results of the VIF procedure further reduced the number of variables under consideration from 16 to 10 by limiting the potential for collinearity in the final model (excluding variables with a VIF>10). After excluding collinear variables, the minimum and maximum correlations between variables are 0.003 and 0.8, respectively. The final set of variables used to build the model (Table 2-5) consisted of one land cover variable (forest), a suite of bioclimatic measures pertaining primarily to precipitation, and two measures of NDVI.

Table 2-5. Environmental and climatic variables with low collinearity determined by Variance Inflation Factor, highly collinear variables not shown. Variables listed constitute final variable set considered in logistic regression model selection.

| Variable | VIF |
|-----------|----------|
| Forest | 1.104511 |
| Bio3_fl | 4.587343 |
| Bio7_fl | 4.507246 |
| Bio8_fl | 2.272024 |
| Bio9_fl | 4.087867 |
| Bio13_fl | 6.758279 |
| Bio15_fl | 7.523549 |
| Bio18_fl | 8.014768 |
| NDVI_max | 1.982174 |
| NDVI_mean | 2.06272 |

Selection of the final multivariate logistic regression model utilized an exhaustive model search on the remaining 10 variables resulting in 1024 models being considered. The set of potential ‘best’ models with the smallest AICc values differed by less than 2.0 units, indicating they are of essentially equal quality (Symonds and Moussalli 2011). After careful consideration of variable combinations within the set of ‘best’ models, ultimately the model with the lowest AICc was selected. This model (Table 2-6) included 6 of the 10 considered variables, each with a significance of $p < 0.01$.

Table 2-6. Coefficients and variables included in the best logistic regression model to predict presence of *A. americanum* across Florida.

| Coefficients: | | | | | |
|---------------|---------|------------|----------|----------|--------|
| | Beta | Std. Error | Z-Value | Pr(> z) | |
| Intercept | -3.1486 | 0.5878 | -5.356 | 8.49E-08 | *** |
| Forest | 1.3328 | 0.6366 | 2.094 | 0.034006 | * |
| Bio3 | -1.1053 | 0.5214 | -2.12 | 0.034006 | * |
| Bio8 | -1.0896 | 0.4937 | -2.207 | 0.027313 | * |
| Bio13 | -1.4302 | 0.6082 | -2.352 | 0.018692 | * |
| Bio15 | 1.6852 | 0.6718 | 2.508 | 0.012126 | * |
| Max NDVI | 1.4121 | 0.387 | 3.648 | 0.000264 | *** |
| Sign. Codes | 0 '***' | 0.001 '**' | 0.01 '*' | 0.05 '.' | 0.1 '' |

The variables included in the final model were Forest cover, maximum NDVI, Bio3 (temperature stability or isothermality), Bio8 (mean temperature of wettest quarter), Bio13 (precipitation of wettest month), and Bio15 (precipitation seasonality). Both forested land cover (which includes predominantly coniferous forest, deciduous forest, and mixed forests) and maximum NDVI (a proxy for vegetation health) were both positively associated with tick presence. This indicates that *A. americanum* is more likely to occur in areas with higher vegetation health (greenness) and in forested areas. Three of the four included bioclimatic variables: Bio3, Bio8, and Bio13 were all negatively associated with *A. americanum* presence. Thus, as isothermality increases (Bio3) occurrence of *A. americanum* is likely to decrease. Similarly, areas that receive the highest amounts of rainfall during the wettest months (Bio8) or are the warmest during the wettest quarter, are associated with a lower probability of occurrence for *A. americanum*. The final bioclimatic variable included in the model, precipitation seasonality (Bio15) was positively associated with tick presence indicating that as intra-annual variation in precipitation increases (i.e. areas with more pronounced wet and dry seasons), probability of occurrence for *A. americanum* increases as well.

To test the assumptions of statistical independence and adequate control of spatial autocorrelation in the final model, the spatial distribution of the standardized residuals were assessed using Moran's I. The null hypothesis of Moran's I is complete spatial randomness in the distribution of point values. The test returned an insignificant ($p > 0.05$) p value of 0.83 which indicates the null hypothesis should not be rejected. This result indicates that one of the main assumptions of logistic regression, independence of model residuals, was met and that the model adequately controlled for spatial autocorrelation. The spatial distribution of residuals indicates that presence locations frequently had larger positive residuals (Figure 2-3).

The assessment of model accuracy via the Hosmer-Lemeshow and cross validation (Table 2-7) indicated that the model did an adequate job of fitting the training data. The model successfully fits the training data as indicated by a non-significant p-value for the Hosmer-Lemeshow GOF test. Further validation via k-fold cross validation indicated that the model did an exceptional job of identifying true positive points (sensitivity), however it did a much poorer job of discerning true negativity points (specificity). The Kappa metric is considered a better measure than overall accuracy for determining how well a model performs as it is a comparison against what could be expected by random chance. The calculated Kappa statistic of 0.43 indicated a solid improvement in model performance over what is expected by chance (Fielding and Bell 1997).

Table 2-7. Model validation metrics indicating model fit against training and testing datasets indicating a reduction in overall accuracy and sensitivity.

| | Hosmer-Lemeshow GOF (p-value) | Sensitivity | Specificity | Accuracy | Kappa |
|----------|----------------------------------|-------------|-------------|----------|-------|
| Training | 0.087 | 0.94 | 0.48 | 0.86 | 0.43 |
| Testing | 3.27E-12 | 0.63 | 0.83 | 0.81 | 0.278 |

By examining the ratio of positive and negative observations in each decile of estimated probabilities it was found that the model produces a nearly dichotomous discrimination between presence and absence (Table 2-8). Positive observations are absent until reaching the top 40% of observations and correspond to a minimum estimated probability of 0.1.

Table 2-8. Distribution of presence to absence points for each range of estimated probabilities indicating good discrimination by the model. Each bin corresponds to the range of estimated probabilities for 10% of observations in the testing dataset.

| Expected Probabilities | Absence | Presence | Proportion of Presence |
|------------------------|---------|----------|------------------------|
| (1.3e-06,6.9e-03) | 31 | 0 | 0 |
| (7.1e-03,1.5e-02) | 31 | 0 | 0 |
| (1.6e-02,3.0e-02) | 31 | 0 | 0 |
| (3.0e-02,4.4e-02) | 31 | 0 | 0 |
| (4.4e-02,9.3e-02) | 31 | 0 | 0 |
| (9.4e-02,1.8e-01) | 28 | 3 | 0.11 |
| (1.8e-01,3.2e-01) | 28 | 3 | 0.11 |
| (3.4e-01,4.9e-01) | 27 | 4 | 0.15 |
| (5.0e-01,6.0e-01) | 27 | 4 | 0.15 |
| (6.0e-01,9.4e-01) | 24 | 13 | 0.54 |

Further, calculations of model accuracy (sensitivity, specificity, accuracy, and Kappa) indicated large reductions in Kappa and sensitivity, although the model continued to perform better than random. Specificity of the model was greatly improved against the testing dataset however; the reduced sensitivity produced a slight decrease in overall lower model accuracy.

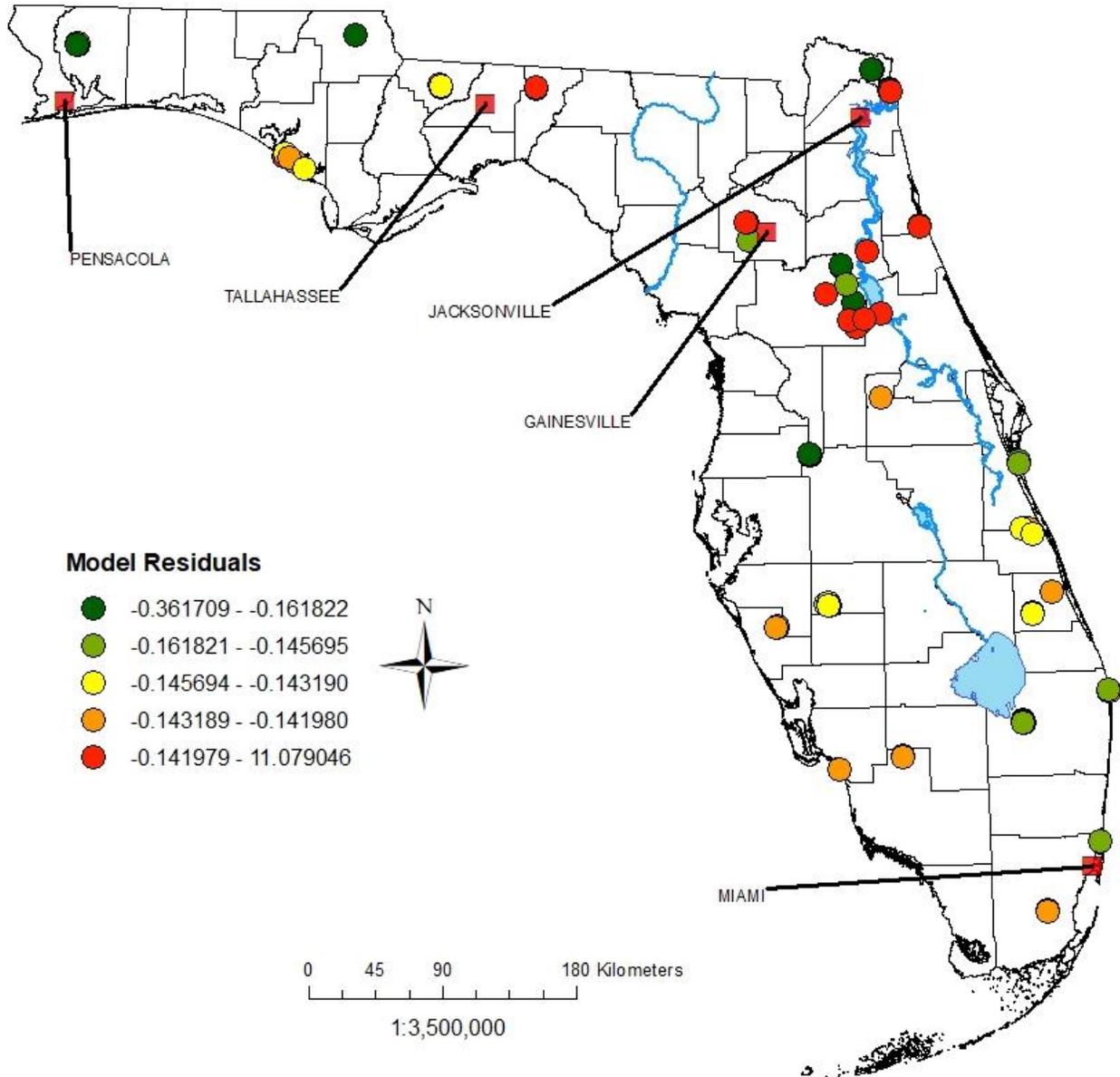


Figure 2-3: Spatial distribution of model residuals. Legend values indicate the five quantiles: minimum, 25%, median, 75%, maximum. Positive and negative values indicated by green and red, respectively.

The resulting spatial prediction based on the logistic regression model (Figure 2-4) indicates that large portions of northern and central Florida have moderate to high predicted occurrence probabilities for *A. americanum*. The areas with the highest predicted probabilities for occurrence are the far western panhandle, gulf coast, and the far northeastern corner extending south along the coast to the Melbourne area and inland to the Ocala National Forest. The spatial prediction indicates that most of southern Florida is not suitable for adult *Amblyomma americanum*.

The cutoff value for delineating presence from absence in the model was 0.186, which is where the model correctly predicts approximately 78% of presence and absence values correctly. This value was used as the cutoff value to distinguish what could be suitable conditions (presence location) from an unsuitable (absence) location. In James et al. (2015) this method of determining a cutoff to delineate medium from high suitability was used and produced a value of 0.1. The average probability value of suitable areas based on the LPT in the present study was determined to be 0.42. The spatial prediction was reclassified based on this cutoff to produce a map of the estimated spatial distribution of adult *A. americanum* (Figure 2-5) in Florida.

An examination of the spatial gradients in predicted occurrence of *A. americanum* (Figure 2-6) produced obvious directionality with the percentage of land suitable for the species based on the LPT threshold. Habitat suitable for adult *A. americanum* comprised 30-50% of all areas from the Big Bend northwards. Less than 10% of land in southern Florida was predicted to be suitable for the species. The trend of estimated probability values within the potential distribution area was also examined (Figure 2-7). This measure examines if areas of high estimated suitability are uniformly distributed within the areas of predicted occurrence. High

suitability was defined as an area with probability values greater than the mean value of suitable areas (areas with probability values greater than the LPT). In this study the average probability value of 'suitable' habitat was 0.42. Therefore 'high suitability' areas were all areas with a probability value $P(x \geq 0.42)$. The trend in highly suitable areas roughly mirrors that for suitable areas in general. There is a distinct reduction in the suitability of areas within the estimated distribution in southern areas of the state. Between 30-50% of the estimated distribution in north Florida is highly suitable, whereas in the south, 0-10% of the estimated distribution would be considered highly suitable for the species. The far southern regions of the state are not highly suitable for the species at all.

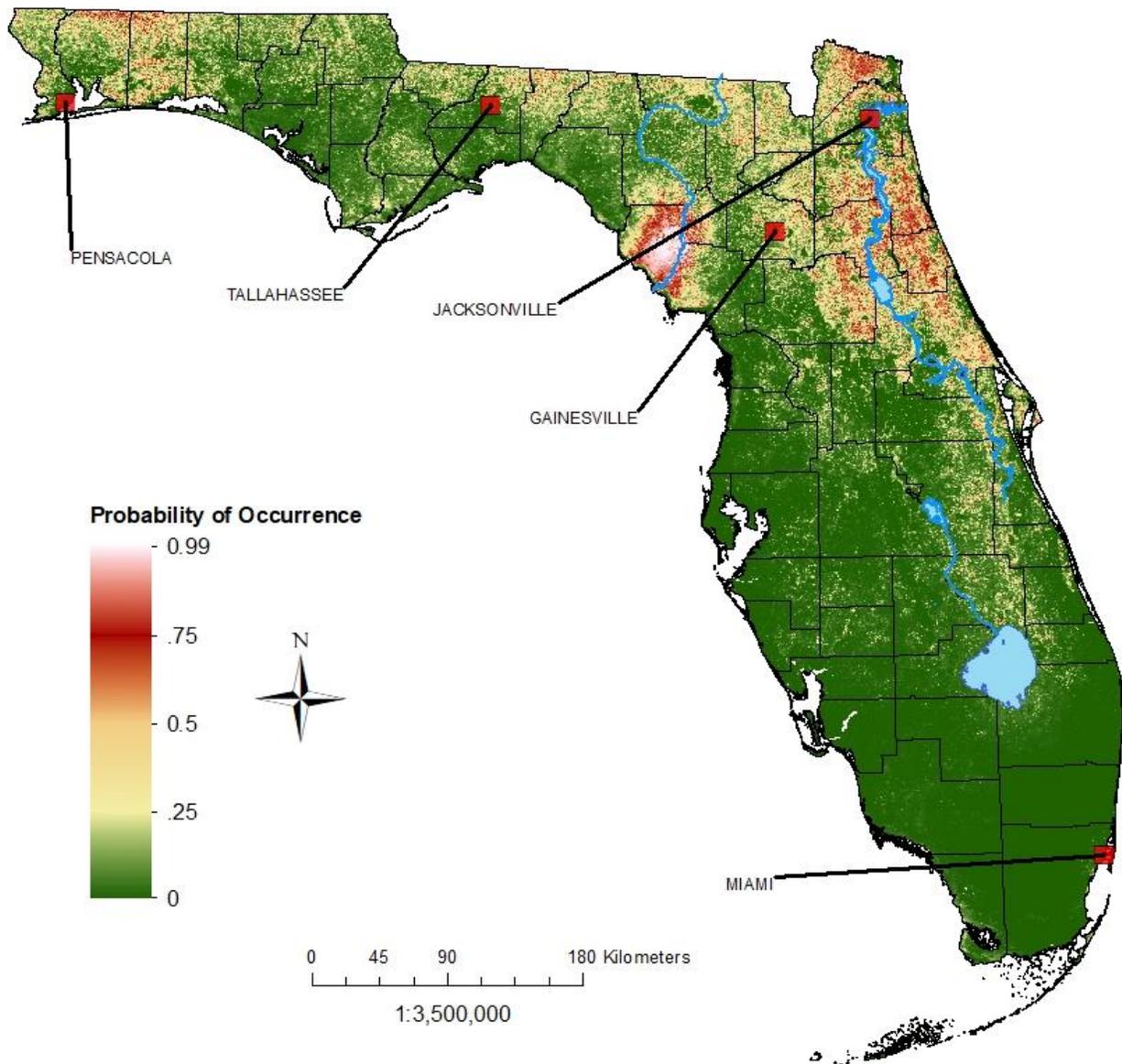


Figure 2-4: Probabilities of occurrence of *Amblyomma americanum* in Florida. Green shading and red shading indicates low and high probability of occurrence, respectively.

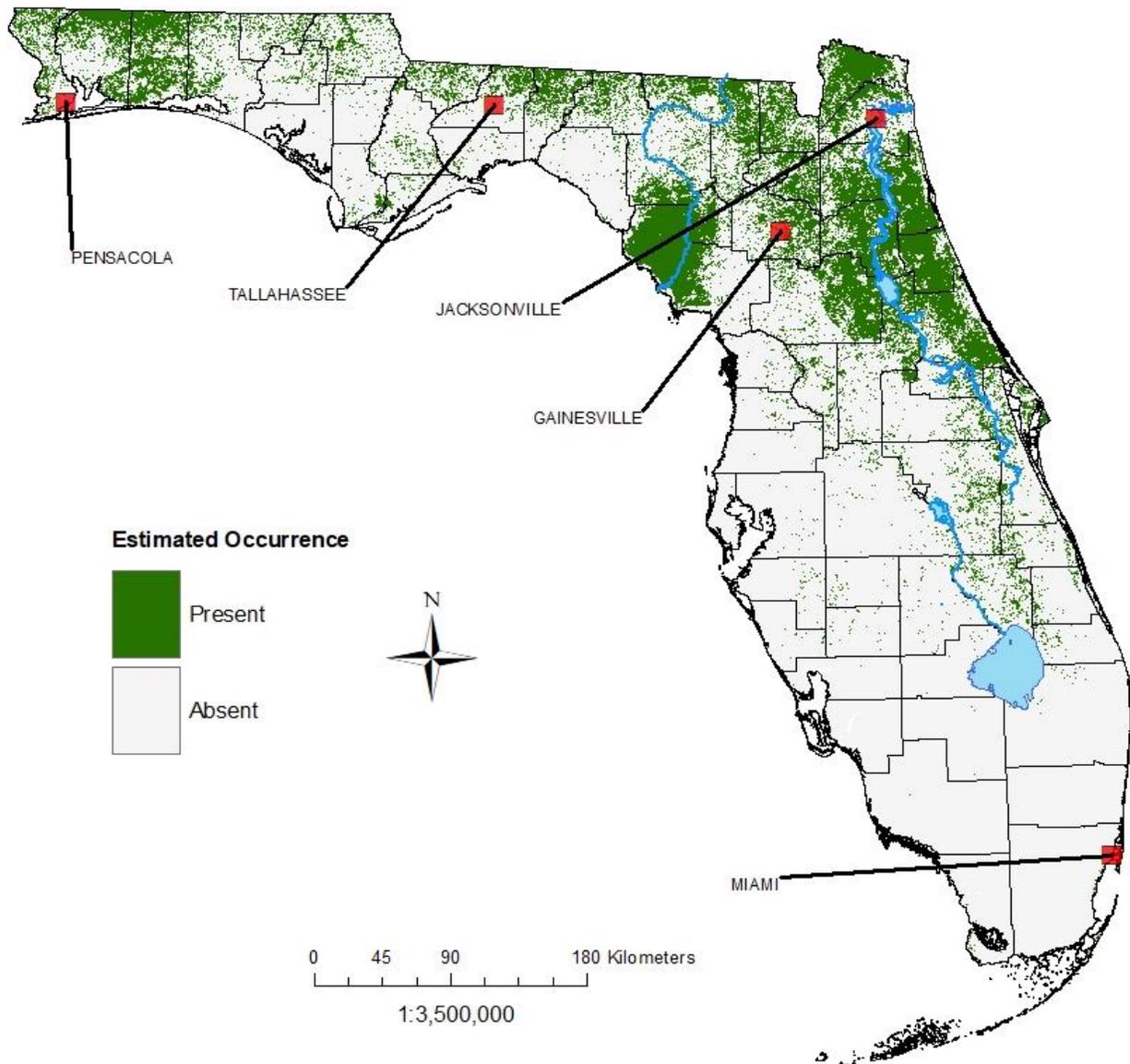


Figure 2-5: Estimated distribution of *Amblyomma americanum* based on a lowest presence threshold (LPT) indicating substantial areas of suitable habitat in northern Florida and increased unsuitable areas in southern Florida.

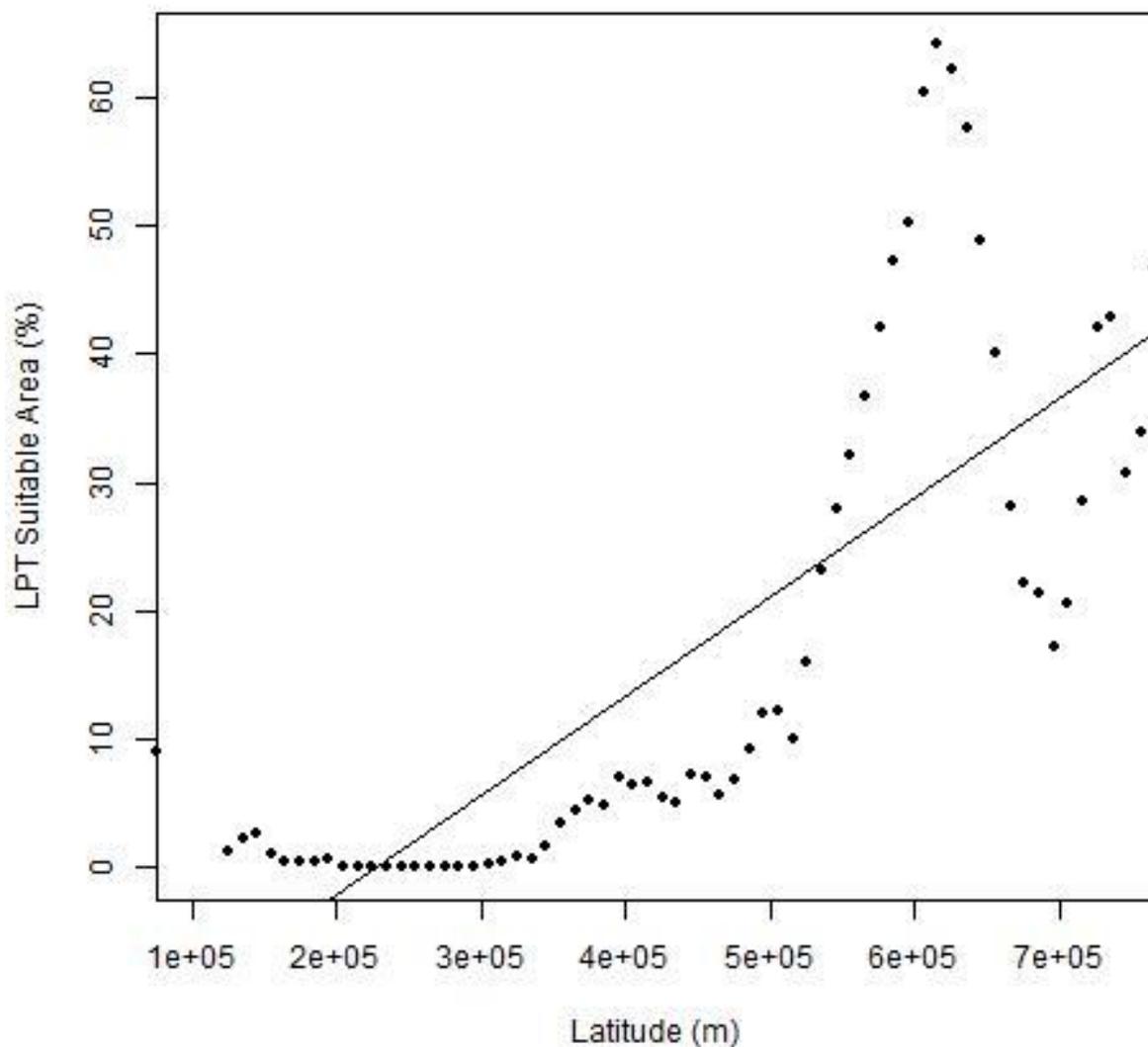


Figure 2-6: Latitudinal trend indicating an increase in percent of total area suitable for *A. americanum* from South to North in Florida. Points indicate totals in 10km increments. Latitude (km) from South (L) to North (R), expressed in meters, shown on the X-axis. Percentage of land area deemed suitable for *A. americanum* based on LPT is shown on the Y-axis.

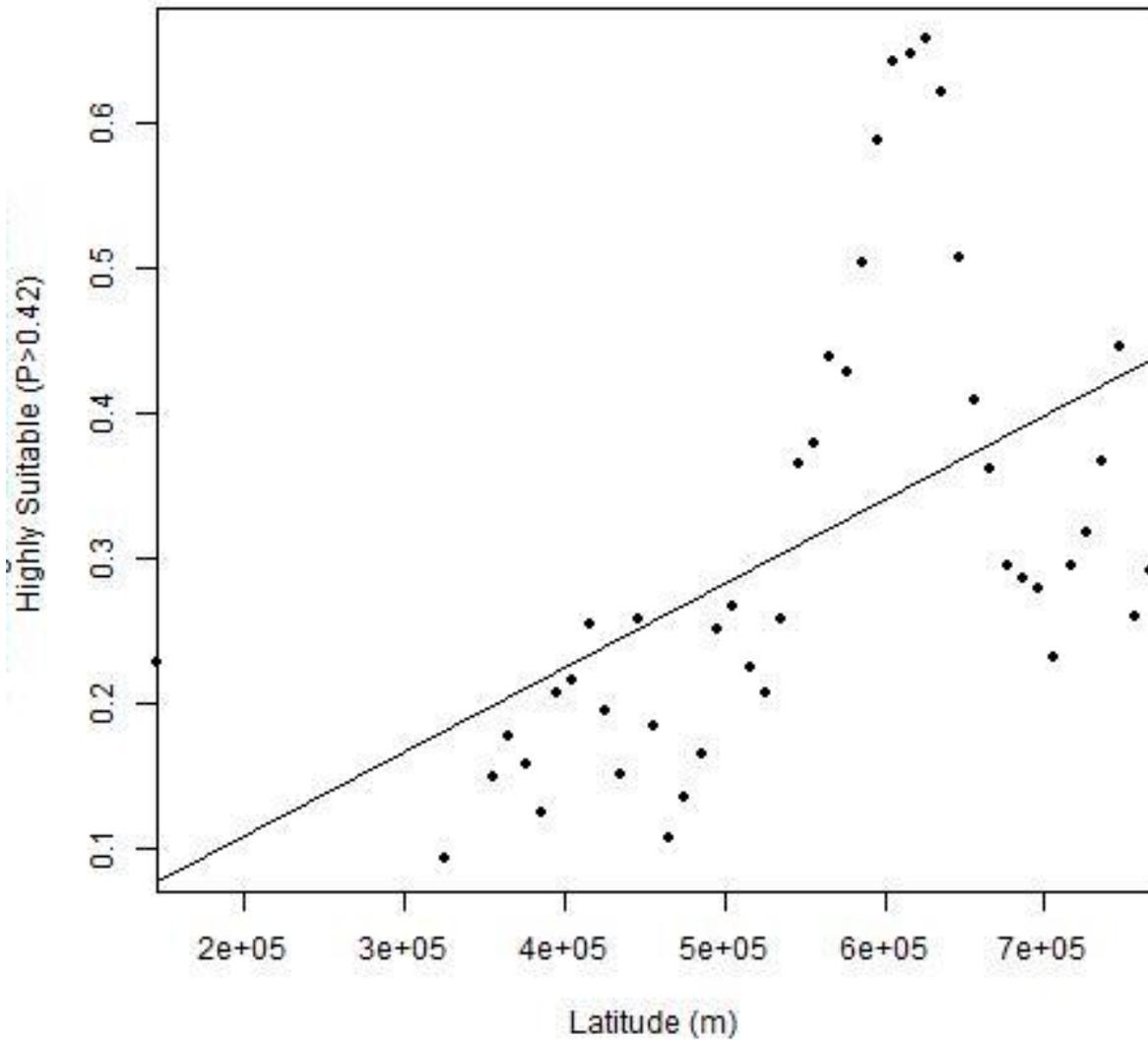


Figure 2-7: Latitudinal trend indicating an increase in high predicted suitability in relation to total suitable area from South to North in Florida. Points show totals in 10km increments. Latitude (km) from South (L) to North (R), expressed in meters, shown on the X-axis. Proportion of area within the estimated distribution of *Amblyomma americanum* predicted to be highly suitable for *A. americanum* is shown on the Y-axis.

2.6 Discussion

The estimated spatial distribution of adult *A. americanum* presented here is the first effort to produce a well-resolved spatial distribution at the sub-county level for Florida. The spatial prediction from the logistic regression model indicates that a large portion of the state is environmentally favorable to this tick species; however, the favorable areas are not uniformly distributed.

Areas of suitable habitat drop off significantly south of Lake Okeechobee (27°N). A meridional gradient for both the probability of presence and for percent area of predicted habitat indicates that habitat suitable for the lone star tick decreases from North to South. During the study period adult *A. americanum* were not collected from any sites south of Ocala national forest, which would support the model prediction of limited habitat suitability farther south in the state. However, additional collection efforts performed between September 2016 and May 2017 have returned adult *A. americanum* as far south as Lake Okeechobee in areas predicted to be suitable by the model.

The ensemble model of county level habitat suitability for *A. americanum* by Springer et al. (2015) similarly shows reduced model agreement in southern Florida counties indicating lower certainty in predicted habitat suitability. These results reflect reports in the literature (Springer, et al. 2014) and the findings from collections for this study that adult lone star ticks in southern Florida are scarce; it is still unclear if this is due solely to lower sampling intensity in the southern portion of the state or because of the underlying distribution of ticks.

The results of this study also are supported by findings in Allan et al. (2001) showing regional proportions of *A. americanum* from ticks collected from deer and swine in Florida dropped markedly between northern, central, and southern regions (deer: 56.9%, 47.9%, 0%; swine: 27.3%, 0.6%, 1.5%). In this study, it was shown that the percentage of area suitable for *A.*

americanum decreased from north to south. While collections by flagging more closely resemble what could be expected of human exposure during outdoor activities, estimates of tick presence or density derived from free ranging wildlife potentially provide a more accurate representation of the true tick burden on the landscape due to the continuous environmental exposure and wide geographic movements of the animals (Ginsberg and Ewing 1989). It is therefore assumed that the significant reduction in *A. americanum* prevalence found by Allan et al. (2001) across the three ecoregions is result of underlying ecological conditions and not simply local variability between sampling locations.

The divisions used by Allan et al. (2001) to classify the state into three regions based on the dominant ecosystem types relies heavily on differences in dominant vegetation types, many of which overlap multiple regions. This approach provides a distinct boundary underpinned by biologically relevant parameters by which differences are compared. In this study however, land cover classifications were insignificant predictors of *A. americanum* presence (forest areas being the one exception) and the application of such an approach would therefore be arbitrary. By considering latitudinal differences in suitable area based on the resolution of the data a more continuous trend in habitat suitability becomes apparent. Both the reduction in suitability and proportional area of potential occurrence from north to south at this scale supports the notion that ecological conditions become less favorable for *A. americanum* in southern Florida.

One region in the ‘big bend’ area of Florida, along the gulf coast was predicted to be especially favorable to *A. americanum*. This region was not sampled during the course of this study and it is unknown if the high probability estimates in this area are substantiated by occurrence of ticks or if the results in this area are anomalous. Springer et al (2015) do not note particularly high habitat suitability in this region likely due to its prominence being masked by

the county level scale of analysis. Collections by Sayler et al (2014), returned large numbers of adult *A. americanum* from several Florida state parks bordering this area. However, visualization of the BioClim climate coverages used in this study reveal sharp transitions in temperature and precipitation trends at this location indicative of possible interpolation errors. Further consideration of this area is contained in Appendix B.

The variables included in the final model were Forest cover, maximum NDVI, Bio3 (temperature stability or isothermality), Bio8 (mean temperature of wettest quarter), Bio13 (precipitation of wettest month), and Bio15 (precipitation seasonality). Each of these variables can be interpreted as biologically relevant information about the ecological preferences of *A. americanum*. Each variable included in the model was statistically significant with greater than 95% confidence (Table 2-7).

Both forested land cover (which includes predominantly coniferous forest, deciduous forest, and mixed forests) and maximum NDVI (a proxy for vegetation health) were positively associated with tick presence. This indicates that *A. americanum* is more likely to occur in areas with higher vegetation health (greenness) and in forested areas. Presence of *A. americanum* was positively correlated with the presence of forested landscapes ($\beta= 1.33, \alpha<0.05$). Previous studies indicate ticks utilize the microclimate provided by dense leaf litter and brushy understory to prevent desiccation when environmental conditions become unfavorable and to lay eggs (Oliver Jr 1989). Anderson et al. (2008) state that the highest numbers of adult *A. americanum* are found on brushy vegetation due to the relatively high humidity. NDVI is a proxy for vegetation health and productivity. The inclusion of a positive association with maximum NDVI ($\beta= 1.41, \alpha<0.001$) speaks to the importance of healthy plant biomass for sustaining tick populations. Healthy vegetation produces necessary high humidity microclimates utilized by questing ticks

(Anderson and Magnarelli 2008, Vail and Smith 2002). Vegetation growth will generally be greatest during summer months resulting in maximum NDVI values during a time of the year when temperatures are highest and ticks are most prone to desiccation. Areas with high maximum NDVI are significantly more likely to have adult *A. americanum* present.

The determination of significant associations between *A. americanum* and both Forest land cover and maximum NDVI is noteworthy. During the variable selection process environmental variables were checked against each other for collinearity and the correlation between Forest and maximum NDVI was deemed to be satisfactorily low. And while other land cover types such as shrubs or grasslands alone yielded no significant association with tick presence, maximum NDVI values outside of forested areas must play a role in increasing the probability of tick presence.

Three of the four included bioclimatic variables: Bio3, Bio8, and Bio13 were all negatively associated with *A. americanum* presence. Thus, as isothermality increases (Bio3) occurrence of *A. americanum* is likely to decrease. Isothermality is a quantification of how large day-night temperature oscillations relate to summer-winter oscillation. High isothermality is indicative of temperature evenness over the course of the year, whereas a low isothermality indicates that the size of temperature swings vary between summer and winter. Tick life cycles are regulated by seasons and temperatures can play a role in development and reproduction. James et al. (2015) posits that isothermality impacts immature development for *D. variabilis* and Oliver (1989) states that lone star tick development rate can be accelerated under laboratory conditions by modulating temperature. The negative relationship reported in this study between *A. americanum* presence and isothermality ($\beta = -1.10$, $\alpha < 0.05$), that even temperatures reduce the probability of *A. americanum* occurrence, appears to contradict the general biology of the

species. Previous studies of the environmental preferences of ticks at various geographic extents have considered isothermality but found it to be non-significant (Springer, et al. 2015, James, et al. 2015). It may be that the isothermality has a very indirect effect not related to the biology of the species.

Areas that receive the highest amounts of rainfall during the wettest month (Bio13) or are the warmest during the wettest quarter (Bio8), are associated with a lower probability of occurrence for *A. americanum*. The mean temperature of the wettest quarter (Bio8) showed a negative association with tick presence ($\beta = -1.08$, $\alpha < 0.05$). High temperatures can decrease survivorship in ticks by increasing the likelihood of experiencing desiccation during host seeking activities (Knulle and Rudolph 1981, Vail and Smith 2002, Anderson and Magnarelli 2008). While the availability of moisture can reduce this risk by increasing relative humidity, laboratory studies on survivorship among multiple larval *Boophilus (Rhipicephalus)* species show that even at high relative humidity (97%) survival is significantly impacted by temperature increases from 20°C to 25°C (Davey, Cooksey and Despins 1991). In the current study, no ticks were found in southern Florida, which coincidentally experiences the highest average temperatures in the state during the wettest months (>25°C). Springer et al. (2015) found mean temperature of the wettest quarter to be a significant predictor of *A. americanum* presence.

James et al. (2015) cites precipitation of the wettest month (Bio13) as impacting egg development and larval survivability. In this analysis greater precipitation during the wettest month was negatively associated with adult tick presence ($\beta = -1.43$, $\alpha < 0.05$). The combination of high annual precipitation and a great deal of low lying areas and flat topography makes many areas in Florida prone to flooding or standing water. During the peak of Florida's monsoon season, large precipitation events may result in standing water long enough to reduce

survivability of eggs or larvae and ultimately impacting adult tick populations. Weiler et al. (2017) showed that major flood events can reduce the abundance of questing ticks of several species. Adejinmi (2011) also found that prolonged submersion in water decreased hatchability of *R. sanguineus* and *H. leachi* eggs. Similarly, the final bioclimatic variable included in the model, precipitation seasonality (Bio15) was positively associated with tick presence indicating that as intra-annual variation in precipitation increases (i.e. areas with more pronounced wet and dry seasons), probability of occurrence for *A. americanum* increases as well. The positive association between *A. americanum* and the seasonality of precipitation ($\beta = 1.68$, $\alpha < 0.05$) implies that areas of North Florida with a more pronounced wet and dry season are more likely to harbor adult ticks. The shift between wet and dry seasons may allow areas to dry out enough for eggs and larvae to survive. Koch (1986) showed in the laboratory *A. americanum* could survive submersion for several weeks so periodic flooding would have little long-term effect on the species. In south Florida where the precipitation is less variable, constant moisture in areas otherwise appropriate for sustaining tick populations result in unsuitable conditions. Springer et al. (2015) and James et al. (2015) found precipitation seasonality to be a significant predictor of *A. americanum* and *D. variabilis* distributions, respectively.

The preliminary examination of significant relationships (via univariate logistic regression during the variable selection process) between the considered environmental variables and the presence of *A. americanum* revealed that only a fraction of the considered habitat and vegetation characteristics (Forest cover, and two measures of NDVI) had any significant association. Distance to water bodies, all other land cover types, and a third measure of NDVI had non-significant associations with tick presence. Of the considered climate variables, six of 19 considered variables were also found to be not significant predictors. Tick physiology is

notably modulated by abiotic conditions so the dominance of climate measures is not surprising. Survivorship of all life stages, reproductive success, and development are affected by temperature and moisture availability (Anderson and Magnarelli 2008, Yoder and Tank 2006, Yoder and Benoit 2003, Randolph, et al. 2002). For this reason, many attempts to estimate the geographic distributions of ticks have focused on associating presence or abundance with temperature and moisture measurements (Atkinson, et al. 2014, Estrada-Pena 1998, Porretta, et al. 2013, Springer, et al. 2015, Raghavan, et al. 2016, James, et al. 2015). Furthermore, limited findings (76 sample sites across 74.5km²) have indicated that habitat and vegetation characteristics including land cover type, NDVI, and distance to water, are not adequate predictors of tick presence or abundance in the southeastern United States (Fryxell, et al. 2015). The observed relationships between presence of *A. americanum* and the variables considered in this study appear to support the notion that climatic conditions (temperature and moisture) are most important in determining the distribution of the species in Florida.

CHAPTER 3 CONCLUSIONS

This thesis attempts to identify environmental and climatic factors which contribute to the spatial distribution of adult *A. americanum*, and identify areas with high probability of infestation. Environmental and climatic data was used to generate a species distribution model using a logistic regression framework. The high resolution spatial prediction generated by this study can better inform the public and state and federal agencies about areas with high risk of exposure to *A. americanum*. This study constitutes the first attempt at estimating the state-wide geographic distribution of this this pest at resolutions below the county level. The following are the major findings drawn from this study:

3. The distribution of *A. americanum* in Florida is significantly associated with a number of climatic variables and land cover characteristics. The general exclusion of most variables associated with habitat and vegetation is in agreement with other studies predicting ecological preferences of ticks in the southeastern United States.
4. Habitat suitability for *A. americanum* and total area of predicted presence for *A. americanum* decreases meridionally. While *A. americanum* is commonly collected from northern and central Florida, there are limited reports, both in counts and frequency, of the species from southern counties. The findings in this study support the notion that the species is uncommon in South Florida.

There are several limitations to this study. This study was limited to estimating the distribution of adult *A. americanum* as larval and nymphal ticks were excluded. However, all life stages are important to the transmission of disease causing pathogens to humans and an estimated distribution based solely on adults may underestimate risk areas in places where other life stages occur. There were few sites in this study where the presence of nymphal *A. americanum* was noted and adults were absent although their inclusion would improve the interpretation of the model.

The size of the study area was a major challenge in determining the locations, number of sites sampled, and intensity of sampling at each site. The selection of sites was determined based on geographic distribution and variability in local environmental and climatic conditions; however, the available habitat types sampled at various latitudes might contribute to the north-south distribution trend indicated by the model. The location of sampled sites may have also contributed to the anomalous habitat suitability in the Big Bend area as no sites were sampled there. Future improvements on this work should seek to reduce bias in sampling locations. A spatially stratified random design modified from the site selection procedure used in Diuk-Wasser (2006) would ensure un-biased statewide coverage. Selection of additional independent sites for model validation would also improve estimates of model accuracy.

Another limitation stems from differences between otherwise similar sites that was not captured by the considered variables. Differences in active management strategies occurs at many of the sites for conservation or other purposes. This active management may bias the variability among where ticks are present by influencing land cover, understory, or other habitat characteristics. For example, active management (controlled burning or canopy thinning), results in significantly reduced long-term tick counts for *A. americanum* and other species (Gleim, et al. 2014, Stafford, Ward and Magnarelli 1998). However, in some locations recent burning may increase tick abundance due to reintroduction by large herbivores grazing on re-emergent understory vegetation (Cilek and Olson 2000). Inclusion of additional sites, including private properties where active habitat management is less common, would reducing this source of sampling bias.

This thesis intends to identify the geographic distribution of *A. americanum* in Florida using environmental and climatic factors. The dominance of climatic variables in governing the

distribution of adult lone star ticks in Florida reinforces previous reports in the literature that habitat related variables alone are insufficient in predicting tick presence or density. Historically, tick collections efforts have under sampled many counties in Florida potentially resulting in under estimates of the species distribution and the human risk of exposure to this species and its pathogens. The spatial prediction of *A. americanum*' s distribution in Florida produced by this research contributes to the current understanding of potential risk of exposure to ticks and may inform control measures by state and federal agencies.

The importance of the lone star tick as a vector of multiple human pathogens should guide future work on estimating human exposure risk to this vector and its pathogens. The incorporation of other aspects of tick ecology, such as host distribution or density, and additional climate variables (such as vapor pressure) or variable interactions should be considered to improve estimated distribution. Consideration of available pathogen prevalence or tick-borne disease incidence can further identify tick-borne disease risk. Patterns of disease risk have already been described from the prevalence of the Lyme disease pathogen and its primary vector, *Ixodes pacificus*, in California (Eisen, et al. 2006). Future efforts to produce better-resolved distribution maps of *A. americanum* in Florida should attempt to reduce the limitations of a GLM modeling approach. The use of ensemble predictions can reduce the bias of any single model by leveraging agreement between models.

APPENDIX A LOGISTIC REGRESSION IN R

The model building process for the logistic regression was performed on aggregate data to account for spatial autocorrelation among the sampled points. The following script was written in R and uses true presence/absence data that has been aggregated to 100m. Continuous Variables have all been centered and scaled around a mean=0, std dev=1. Code for generating summary statistics, raster manipulation, model validation or figures is not included. The structure of the input data is included and must be followed for the code to work as shown. This code corresponds to Methods section 2.3.4

```
#####
#           Build Logistic Regression Model
#####

str(agggregatedata_presabs_std)                    #Examine Structure of Dataframe for Modeling
data.frame':      130 obs. of  30 variables:
 $ presabs  : num  0 0 1 0 0 0 1 0 0 0 ...
 $ forest   : num  1 0 0 0 0 0 0 1 1 0 ...
 $ grass    : num  0 0 0 0 0 0 0 0 0 0 ...
 $ other    : num  0 1 1 1 1 1 1 0 0 1 ...
 $ shrub    : num  0 0 0 0 0 0 0 0 0 0 ...
 $ wetlands : num  0 0 0 0 0 0 0 0 0 0 ...
 $ DEM      : num  0.498 0.607 2.031 -0.981 -0.981 ...
 $ bio1_fl  : num  -0.67 -0.67 -0.67 -1.1 -1.1 ...
 $ bio2_fl  : num  1.271 1.255 1.354 -0.715 -0.715 ...
 $ bio3_fl  : num  0.183 0.183 0.183 -1.54 -1.54 ...
 $ bio4_fl  : num  0.584 0.586 0.6 1.263 1.263 ...
 $ bio5_fl  : num  0.64 0.64 0.64 -1.58 -1.58 ...
 $ bio6_fl  : num  -0.823 -0.823 -0.928 -0.913 -0.913 ...
 $ bio7_fl  : num  0.893 0.893 0.996 0.673 0.673 ...
 $ bio8_fl  : num  -0.536 -0.536 -0.536 -0.536 -0.536 ...
 $ bio9_fl  : num  0.268 0.257 0.237 -0.366 -0.366 ...
 $ bio10_fl : num  -0.905 -0.905 -0.905 -0.18 -0.18 ...
 $ bio11_fl : num  -0.641 -0.644 -0.66 -1.133 -1.133 ...
 $ bio12_fl : num  -0.115 -0.114 -0.245 1.433 1.433 ...
 $ bio13_fl : num  0.464 0.464 0.418 0.556 0.556 ...
 $ bio14_fl : num  0.0814 0.0892 -0.0757 1.5775 1.5775 ...
 $ bio15_fl : num  -0.0546 -0.0698 -0.0782 -0.8928 -0.8928 ...
 $ bio16_fl : num  0.248 0.244 0.178 0.179 0.179 ...
 $ bio17_fl : num  -0.229 -0.227 -0.246 1.765 1.765 ...
 $ bio18_fl : num  0.496 0.492 0.419 0.284 0.284 ...
 $ bio19_fl : num  0.569 0.57 0.57 1.272 1.272 ...
 $ dist2water: num  -0.4 -0.303 -0.282 -0.144 -0.225 ...
 $ NDVI_max : num  1.0176 1.2166 1.6686 0.0272 0.388 ...
 $ NDVI_mean : num  1.251 1.251 1.327 0.334 0.432 ...
 $ NDVI_min : num  0.57 0.242 -0.168 1.128 1.176 ...

##2.1 Create Base Model with all variables included
# Create list of variables names
stdcov<-c(colnames(agggregatedata_presabs_std))
stdcov<-stdcov[-1]                    # Remove 'presence' column
stdcov<-stdcov[-3]                   # Remove 'Other' Landcover classification

stdvars<-paste(stdcov, collapse="+")
names
```

```

# Run Binomial GLM with Logit Link
stdaggnave_model<-glm(paste("presabs~", stdvars,sep=""),
  data=aggregatedata_presabs_std,
  family=binomial(link = "logit"))

summary(stdaggnave_model) # Examine Model Summary,

##2.2 Begin Model Selection to Generate Valid Model
#2.2a Determine Variables with significant predictive capacity
#of dependent variable

dep_vars<-names(aggregatedata_presabs_std[1]) # Specify the Dependent Variable

# pairwise combinations with dep_vars:
var_comb <- expand.grid(dep_vars, stdcov) # Create Matrix of 1 var combos of dep/ind

# formulas for all combinations
formula_vec <- sprintf("%s ~ %s",
  var_comb$Var1,
  var_comb$Var2)

# Run each logistic model combination
glm_res <- lapply( formula_vec, function(f){
  fit1 <- glm( f, data = aggregatedata_presabs_std,
    family = binomial("logit"))
  fit1$coefficients <- coef( summary(fit1))
  return(fit1)
})

names(glm_res) <- formula_vec # Name each model based on variable
combination

# Extract p-values for covariates of each model in a data.frame
p_values <-
cbind(stdcov,formula_vec, as.data.frame ( do.call(rbind,
  lapply(glm_res, function(x) {
    coefs <- coef(x)
    rbind(c(coefs[,4]))
  } )))

names(p_values)[1:4]<-c("var_names", "formula_vec",
  "Intercept" ,"Covariate") # Name Columns of P-value dataframe

p_values$var_names<-as.character(p_values$var_names)
p_values # View dataframe of P-values for each model

# Variables that have a significant univariate relationship
#with presence/absence of A. americanum
signvar<-subset(p_values,p_values$Covariate<0.05) # Create vector of sig Variables p<0.05
nonsignvar<-subset(p_values,p_values$Covariate>=0.05) # Create vector of NONsign Variables p>=0.05

myvars<-paste(signvar[[1]], sep=",")
corr_var<-aggregatedata_presabs_std[myvars] # Parse the tick data to only list significant
variables

```

```

#2.2b Select uncorrelated variables using VIF,
      #cutoff for inclusion is VIF=<10.0
var.cor<-vifstep(corr_var, th=10)
overfitting/redundancy
var.cor

# Extract variable names for use in Logit Model
sig.var<-as.character(var.cor@results$Variables)

##2.3 Identify 'best' logistic regression model
#2.3a Identify n best models from set of all possible models
      #using exhaustive search
model.sigvar.all <- glmulti("presabs",xr=sig.var,
                           data=aggregatedata_presabs_std,
                           family=binomial(link="logit"), level=1,
                           method="h", crit="aicc",
                           maxsize=length(sig.var))
#create table of best models
tbm <- weightable(model.sigvar.all)

#subset table of best models within a range of minimum AIC+2
tbms <- tbm[tbm$aicc <= min(tbm$aicc) +2,]
tbms

#2.3b Best Model
sigvarmodelminaicc<-glm(paste(tbms[1,1]),
                        data=aggregatedata_presabs_std,
                        family = binomial(link = "logit"))
summary(sigvarmodelminaicc)

#2.3c Check statistical assumptions are met
# Spatial Distribution of Residuals
scaledresid<-scale(sigvarmodelminaicc$residuals,
                   center = TRUE, scale=TRUE)
sigvarmodelminaicc_spat<-cbind(aggregatedata_presabs[2:3],
                               sigvarmodelminaicc$residuals,
                               scaledresid)

coordinates(sigvarmodelminaicc_spat)<-
  sigvarmodelminaicc_spat[1:2]

# Moran's I to test for clustering of residuals
points<-coordinates(sigvarmodelminaicc_spat)
x<-sigvarmodelminaicc_spat[[3]]
dists<-as.matrix(dist(points, method="euclidean"))
listw<-1/dists
diag(listw)<-0
Moran.I(x,weight=listw,alternative = "two.sided",
        na.rm = TRUE)
# GOODNESS OF FIT TEST
anova(stdaggaive_model,sigvarmodelminaicc,test="Chisq")

h1_sigvar<-hoslem.test(sigvarmodelminaicc$y,
                       fitted(sigvarmodelminaicc),g=10)
h1_sigvar

```

```

# Consider collinearity to help limit

# View collinearity results

# Create vector of names for logistic model

# First order, evaluated by AICc

# Table of all models considered

# Table of Candidate 'best' Models
# View table of best models

# Best Model=min(AICc)
# view Model Summary

# Extract residuals from best model

# Determine coordinates for each residual

# Calculate Global Moran's I on residuals

# IF p<0.05 reject the reduced model

#If p>0.05 accept null hypothesis
# View Results of HS GOF test

```

APPENDIX B CONSIDERATION OF COVARIATE ANOMOLIES ON SPATIAL PREDICTIONS

Gridded climate surfaces such as those generated for this study are produced by any number of interpolation algorithms. They are frequently used in environmental and biological studies where fine spatial scales are desired in order to local environmental variability. There are severally recently developed datasets commonly used in species distribution models that capture precipitation and temperature measures including the Daymet and WorldClim databases (Fick and Hijmans 2017, Hijmans, Cameron, et al. 2005, Thornton, et al. 2017).

The WorldClim dataset implements the ANUSPLIN algorithm to interpolate desired variables from land based weather stations and high-resolution elevation data. Due to distribution bias in station locations, the authors note that uncertainty is not uniform; low station density and low elevation variation result in areas of high uncertainty (Hijmans, Cameron, et al. 2005). The second version of the WorldClim dataset attempts to reduce the uncertainty by using covariates in addition to elevation. The authors note that most surfaces were not significantly improved but this addition are areas of increased uncertainty remain (Fick and Hijmans 2017). The WorldClim web portal lists a number of known issues, errors, and problems with earlier versions of the dataset however it is unclear this known errors in older versions have been corrected in version 1.4 or 2.0 (Hijmans, WorldClim Issues n.d.). One type of error reported in version 1.2 are where changes in interpolation domains result is sudden climate changes.

The region surrounding the Perry-Foley airfield in the Big Bend area of Florida appears to be an area of high uncertainty or change in interpolation domain. A circular area of local maxima or minima in this location results in a veritable mountains and valleys of values in several biovariable layers (Figure B-1). The region is near the coast, and has relatively little

variation in its topography. Hijmans et al. (2005) cites each both these conditions as potential sources of uncertainty in the interpolation.

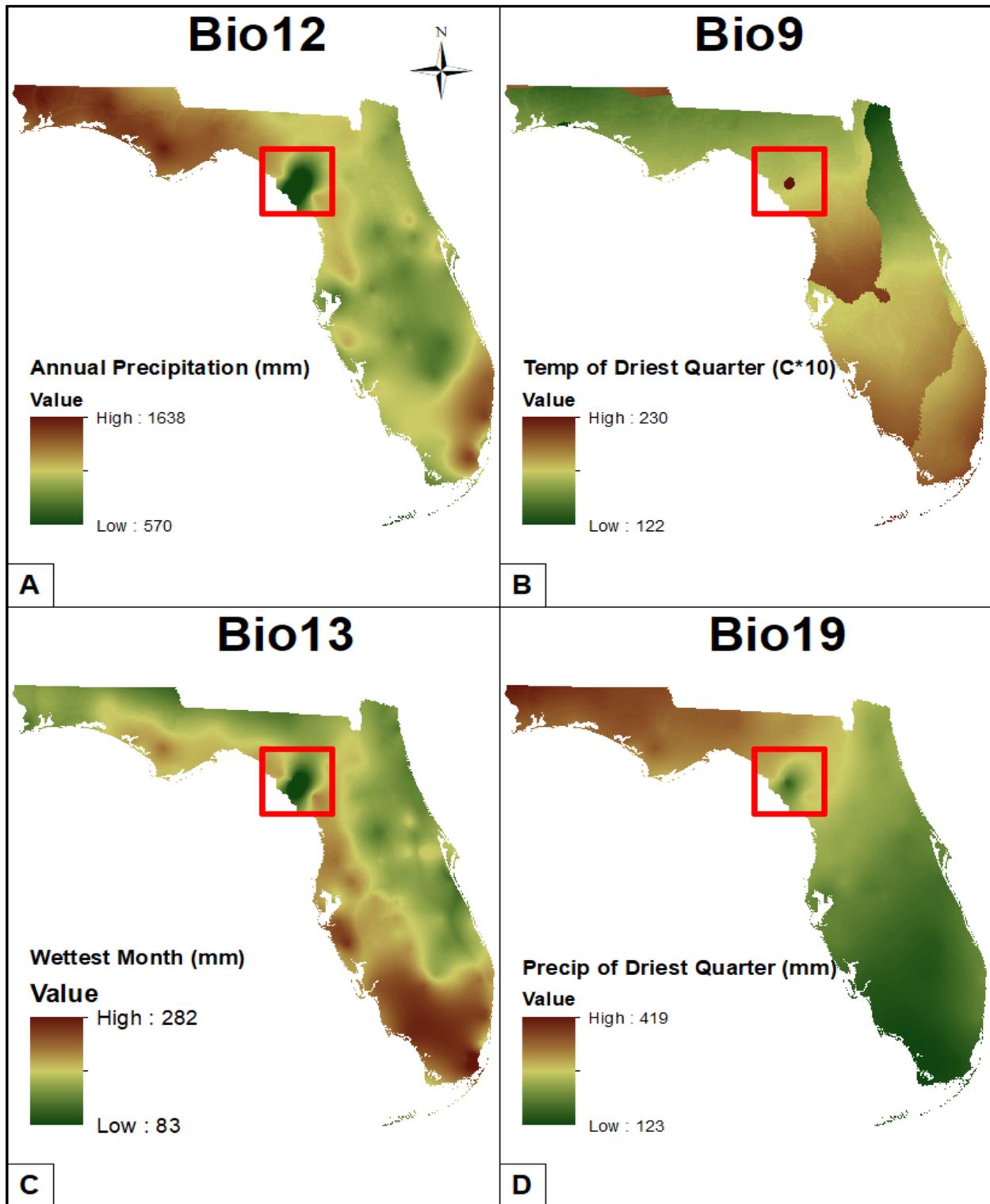


Figure B-1: Examples of potential interpolation error in WorldClim bioclimatic variables. Error appears to originate the annual measured precipitation(B-1a). The error impacts derived variables for precipitation and temperature including: average temperature of driest quarter (B-1b), wettest month (B-1c), and precipitation of driest quarter (B-1d).

This abnormality appears to influence the estimated distribution of *A. americanum* presented in this study. The final logistic regression model includes the bioclimate variables Bio3, Bio8, Bio13 and Bio15. Of these, Bio13 (precipitation of wettest month) contains the aberration much more strongly than any of the other included variables and results in a pocket of extremely low values. The negative association between occurrence and precipitation during the wettest month determined by the logistic regression results in extremely high probability values in this area (Figure B-2).

Affect of Input Data on Spatial Prediction

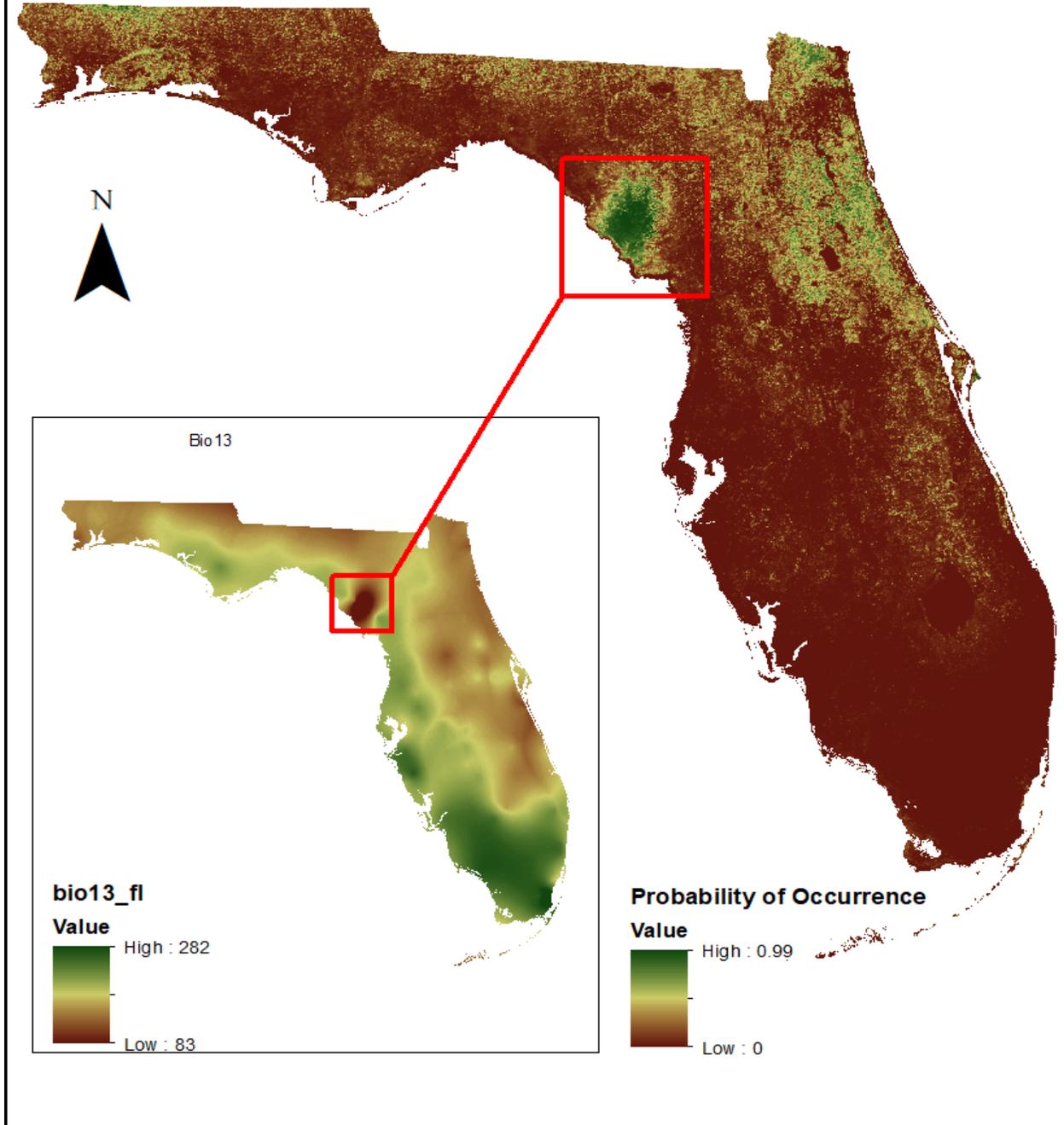


Figure B-2: Impact of interpolation error in bioclimatic variable Bio13 on the spatial prediction produced by the logistic regression model. The low values in the Bio13 data and the negative association ($\beta=-1.43$) between Bio13 and occurrence results in extremely high probability values for occurrence of *A. americanum*.

LIST OF REFERENCES

- Adejinmi, Johnson O. 2011. "Effect of Water Flooding on the Oviposition Capacity of Engorged Adult Females and Hatchability of Eggs of Dog Ticks: *Rhipicephalus sanguineus* and *Haemaphysalis leachi leachi*." *Journal of Parasitology Research*.
- Adelson, Martin E., Raja-Venkitesh S. Rao, Richard C. Tilton, Kimberly Cabets, Eugene Eskow, Lesley Fein, James L. Occi, and Eli Mordechai. 2004. "Prevalence of *Borrelia burgdorferi*, *Bartonella* spp., *Babesia microti*, and *Anaplasma phagocytophila* in *Ixodes scapularis* Ticks Collected in Northern New Jersey." *Journal of Clinical Microbiology* 2799-2801.
- Allan, Brian F., Humberto P. Dutra, Lisa S. Goessling, Kirk Barnett, Jonathan M. Chase, Robert J. Marquis, Genevieve Pang, Gregory A. Storch, Robert E. Thach, and John L. Orrock. 2010. "Invasive Honeysuckle eradication reduces tick-borne disease risk by altering host dynamics." *PNAS* 18523-18527.
- Allan, Sandra A., Leigh-Anne Simmons, and Michael J. Burridge. 2001. "Ixodid ticks on white-tailed deer and feral swine in Florida." *Journal of Vector Ecology* 93-102.
- Allouche, Omri, Asaf Tsoar, and Ronen Kadmon. 2006. "Assessing the Accuracy of Species Distribution Models: Prevalence, Kappa and the True Skill Statistic (TSS)." *Journal of Applied Ecology* 1223-1232.
- Anderson, B. E., C. E. Greene, D. C. Jones, and J. E. Dawson. 1992. "*Ehrlichia ewingii* sp. nov., the etiologica agent of canine granulocytic ehrlichiosis." *International Journal of Systematics and Evolutionary Microbiology* 299-302.
- Anderson, B. E., K. G. Sims, J. G. Olson, J. E. Childs, J. F. Piesman, C. M. Happ, G. O. Maupin, and B. J. Johnson. 1993. "*Amblyomma americanum*: A Potential Vector of Human Ehrlichiosis." *American Journal of Tropical Medicine and Hygiene* 239-244.
- Anderson, John F, and Louis A Magnarelli. 2008. "Biology of Ticks." *Infectious Disease Clinics of North America* 195-215.
- Anselin, Luc. 1995. "Local Indicators of Spatial Association- LISA." *Geographical Analysis* 93-115.
- Araujo, Miguel B., and Mark New. 2006. "Ensemble Forecasting of Species Distributions." *Trends in Ecology and Evolution* 42-47.
- Arlot, Sylvain, and Alain Celisse. 2010. "A Survey of Cross-Validation Procedures for Model Selection." *Statistical Surveys* 40-79.
- Armstrong, Philip M., Laura Rosa Brunet, Andrew Spielman, and Sam R. Telford III. 2001. "Risk of Lyme Disease: Perceptions of Residents of a Lone Star Tick-Infested Community." *Bulletin of the World Health Organization* 916-925.
- Atkinson, Samuel F., Sahotra Sarkar, Aldo Avina, Jim A. Schuermann, and Phillip Williamson. 2014. "A Determination of the Spatial Concordance Between Lyme Disease Incidence and Habitat Probability of its Primary Vector *Ixodes Scapularis* (black-legged tick)." *Geospatial Health* 203-212.
- Austin, M. P. 2002. "Spatial Prediction of Species Distribution: An Interface Between Ecological Theory and Statistical Modelling." *Ecological Modeling* 101-118.

- Austin, M. P., A. O. Nicholls, and C. R. Margules. 1990. "Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species." *Ecological Monographs* 161-177.
- Bell, J. Frederick. 1944. "The Infection of Ticks (*Dermacentor Variabilis*) with *pasteurella Tularensis*." *Journal of Infectious Disease* 83-95.
- Bisanzio, Donal, Giuseppina Amore, Charlotte Ragagli, Laura Tomassone, Luigi Bertolotti, and Alessandro Mannelli. 2008. "Temporal Variations in the Usefulness of Normalized Difference Vegetation Index as a Predictor for *Ixodes ricinus* (Acari: Ixodidae) in a *Borrelia lusitaniae* Focus in Tuscany, Central Italy." *Journal of Medical Entomology* 547-555.
- Bishopp, F. C., and Helen Louise Trembley. 1945. "Distribution and Hosts of Certain North American Ticks." *The Journal of Parasitology* 1-54.
- Blackburn, Jason K. 2010. "Integrating Geographic Information Systems and Ecological Niche Modeling into Disease Ecology: A Case Study of *Bacillus anthracis* in the United States and Mexico." In *Emerging and Endemic Pathogens: Advances in Surveillance, Detection, and Identification*, by K. P. O'Connell, A. Sulakvelidze and L., ed. Bakanidze, 59-88. New York: Springer.
- Bremer, William G., John J. Schaefer, Elizabeth R. Wagner, S.A. Ewing, Yasuko Rikihisa, Glen R. Needham, Sathaporn Jittapalapong, Debra L. Moore, and Roger W. Stich. 2005. "Transstadial and Intrastadial experimental transmission of *Ehrlichia canis* by male *Rhipicephalus sanguineus*." *Veterinary Parasitology* 95-105.
- Bring, Johan. 1994. "How to Standardize Regression Coefficients." *The American Statistician* 209-213.
- Britch, Seth C., Kenneth J. Linthicum, Assaf Anyamba, Compton J. Tucker, Edwin W. Pak, Francis A. Maloney Jr., Kristin Cobb, et al. 2008. "Satellite Vegetation Index Data as a Tool to Forecast Population Dynamics of Medically Important Mosquitoes at Military Installations in the Continental United States." *Military Medicine* 677-683.
- Brown, Heidi E., Karen F. Yates, Gabrielle Dietrich, Katherine MacMillan, Christine B. graham, Sara M. Reese, Wm. Steve Helderbrand, et al. 2011. "An Acarologica Survey and *Amblyomma americanum* Distribution Map with Implications for Tularemia Risk in Missouri." *American Journal of Tropical Medicine and Hygiene* 411-419.
- Brownstein, John S., Theodore R. Holford, and Durland Fish. 2003. "A Climate-Based Model Predicts the Spatial Distribution of the Lyme Disease Vector *Ixodes scapularis* in the United States." *Environmental Health Perspectives* 1152-1157.
- Chatterjee, Samprit, and Ali S. Hadi. 2013. *Regression Analysis by Example*. Somerset: Wiley .
- Childs, James E., and Christopher D. Paddock. 2003. "The Ascendancy of *Amblyomma Americanum* as a Vector of Pathogens Affecting Humans in the United States." *Annual Review of Entomology* 307-337.
- Cilek, J. E., and M. A. Olson. 2000. "Seasonal Distribution and Abundance of Ticks (Acari: Ixodidae) in Northwestern Florida." *Journal of Medical Entomology* 439-444.
- Clark, Kerry. 2004. "Borrelia Species in Host-Seeking Ticks and Small Mammals in Northern Florida." *Journal of Clinical Microbiology* 5076-5086.

- Craney, Trevor A., and James G. Surlis. 2002. "Model-Dependent Variance Inflation Factor Cutoff Values." *Quality Engineering* 391-403.
- Davey, Ronald B., Lynita M. Cooksey, and Joseph L. Despins. 1991. "Survival of larvae of *Boophilus annulatus*, *Boophilus microplus*, and *Boophilus hybrids* (Acari: Ixodidae) in Different Temperature and Humidity Regimes in the Laboratory." *Veterinary Parasitology* 305-313.
- Davidson, William R., Debra A. Siefken, and Lynn H. Creekmore. 1994. "Seasonal and Annual Abundance of *Amblyomma americanum* (Acari: Ixodidae) in Central Georgia." *Journal of Medical Entomology* 67-71.
- Demma, Linda J., Marc S. Traeger, William L. Nicholson, Christopher D. Paddock, Dianna M. Blau, Marina E. Ereemeeva, Gregory A. Dasch, et al. 2005. "Rocky Mountain Spotted Fever from an Unexpected Tick Vector in Arizona." *The New England Journal of Medicine* 587-594.
- Derksen, Shelley, and H. J. Keselman. 1992. "Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables." *British Journal of Mathematical and Statistical Psychology* 265-282.
- Diuk-Wasser, M. A., A. G. Gatewood, M. R. Cortinas, S. Yaremych-Hamer, J. Tsao, U. Kitron, G. Hickling, et al. 2006. "Spatiotemporal Patterns of Host-Seeking *Ixodes scapularis* Nymphs (Acari: Ixodidae) in the United States." *Journal of Medical Entomology* 166-176.
- Dobson, A., and J. Foufopoulos. 2001. "Emerging Infectious Pathogens of Wildlife." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 1001-1012.
- Dormann, Carsten F., Jana M. McPherson, Miguel B. Araujo, Roger Bivand, Janine Bolliger, Gudrun Carl, Richard G. Davies, et al. 2007. "methods to account for spatial autocorrelation in the analysis of species distributional data: a review." *Ecography* 609-628.
- Durden, L. A., J. S. H. Klompen, and J. E. Keirans. 1993. "Parasitic Arthropods of Sympatric Opossums, Cotton Rats, and Cotton Mice from Merritt Island, Florida." *The Journal of Parasitology* 283-286.
- Eisen, Rebecca J, Lars Eisen, and Charles B Beard. 2016. "County-Scale Distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the Continental United States." *Journal of Medical Entomology* 349-386.
- Eisen, Rebecca J., Robert S. Lane, Curtis L. Fritz, and Lars Eisen. 2006. "Spatial Patterns of Lyme Disease Risk in California Based on Disease Incidence Data and Modeling of Vector-Tick Exposure." *American Journal of Tropical Medicine and Hygiene* 669-676.
- Elith, Jane, Steven J. Phillips, Trevor Hastie, Miroslav Dudik, Yung En Chee, and Colin J. Yates. 2011. "A Statistical Explanation of MaxEnt for Ecologists." *Diversity and Distributions* 43-57.
- Estrada-Pena, Agustin. 1998. "Geostatistics and Remote Sensing as Predictive Tools of Tick Distribution: a Cokriging System to Estimate *Ixodes scapularis* (Acari: Ixodidae) Habitat Suitability in the United States and Canada from Advanced Very High Resolution Radiometer Satellite Image." *Journal of Medical Entomology* 989-995.

- Estrada-Pena, Agustin, and Frans Jongejan. 1999. "Ticks feeding on humans: a review of records on human-biting Ixodoidea with special reference to pathogen transmission." *Experimental and Applied Acarology* 685-715.
- Felz, Michael W., Lance A. Durden, and James H. Oliver Jr. 1996. "Ticks Parasitizing Humans in Georgia and South Carolina." *The Journal of Parasitology* 505-508.
- Fick, S.E., and R.J. Hijmans. 2017. "Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas." *International Journal of Climatology*.
- Fielding, Alan H., and John F. Bell. 1997. "A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models." *Environmental Conservation* 38-49.
- Forrester, Donald J., Grace S. McLaughlin, Sam R. Telford Jr., Garry W. Foster, and James W. McCown. 1996. "Ectoparasites (Acari, Mallophaga, Anoplura, Diptera) of White-Tailed Deer, *Odocoileus virginianus*, from Southern Florida." *Journal of Medical Entomology* 96-101.
- Franklin, Janet, Katherine E. Wejnert, Stacie A. Hathaway, Carlton J. Rochester, and Robert N. Fisher. 2009. "Effect of Species Rarity on the Accuracy of Species Distribution Models for Reptiles and Amphibians in Southern California." *Diversity and Distributions* 167-177.
- Freeman, Elizabeth A., and Gretchen G. Moisen. 2008. "A Comparison of the Performance of Threshold Criteria for Binary Classification in terms of Predicted Prevalence and Kappa." *Ecological Modeling* 48-58.
- Fryxell, R.T. Trout, J.E. Moore, M.D. Collins, Y. Kwon, S.R. Jean-Philippe, S.M. Schaeffer, A. Odoi, M. Kennedy, and A.E. Houston. 2015. "habitat and Vegetation Variables Are Not Enough When Predicting Tick Populations in the Southeastern United States." *PLoS ONE*.
- Ginsberg, Howard S., and Curtis P. Ewing. 1989. "Comparison of Flagging, Walking, Trapping, and Collecting from Hosts as Sampling Methods for Northern Deer Ticks, *Ixodes dammini*, and Lone-Star Ticks, *Amblyomma americanum* (Acari: Ixodidae)." *Experimental and Applied Acarology* 313-322.
- Glass, Gregory E., Brian S. Schwartz, John M. Morgan III, Dale T. Johnson, Peter M. Noy, and Ebenezer Israel. 1995. "Environmental Risk Factors for Lyme Disease Identified with Geographic Information Systems." *American Journal of Public Health* 944-948.
- Gleim, Elizabeth R., Mike L. Conner, Roy D. Berghaus, Michael L. Levin, Galina E. Zemtsova, and Michael J. Yabsley. 2014. "The Phenology of Ticks and the Effects of Long-Term Prescribed Burning on Tick Population Dynamics in Southwestern Georgia and Northwestern Florida." *PLoS ONE*.
- Goddard, Jerome, and Andrea S. Varela-Stokes. 2009. "Role of the Lone Star Tick, *Amblyomma americanum*, in Human and Animal Disease." *Veterinary Parasitology*.
- Graham, Michael H. 2003. "Confronting Multicollinearity in Ecological Multiple Regression." *Ecology* 2809-2815.
- Gray, J. S. 1998. "The Ecology of Ticks Transmitting Lyme Borreliosis." *Experimental and Applied Acarology* 249-258.

- Green, R. G. 1931. "The Occurrence of *Bacterium tularense* in the Eastern Wood Tick, *Dermacentor variabilis* ." *American Journal of Hygiene* 600-613.
- Greiner, Ellis C., Pamela P. Humphrey, Robert C. Belden, William B. Frankenberger, David H. Austin, and E. Paul J. Gibbs. 1984. "Ixodid Ticks on Feral Swine in Florida." *Journal of Wildlife Diseases* 114-119.
- Guisan, Antoine, and Niklaus E. Zimmermann. 2000. "Predictive Habitat Distribution Models in Ecology." *Ecological Modeling* 147-186.
- Guisan, Antoine, and Wilfried Thuiller. 2005. "Predicting species distribution: offering more than simple habitat models." *Ecology Letters* 993-1009.
- Guisan, Antoine, Jean-Paul Theurillat, and Felix Kienast. 1998. "Predicting the Potential Distribution of Plant Species in an Alpine Environment." *Journal of Vegetation Science* 65-74.
- Guisan, Antoine, Jr, Thomas C. Edwards, and Trevor Hastie. 2002. "Generalized Linear and Generalized Additive Models in Studies of Species Distributions: Setting the Scene." *Ecological Modeling* 89-100.
- Heath, A.C.G. 1981. "The Temperature and Humidity Preferences of *Haemaphysalis longicornis*, *Ixodes holocyclus* and *Rhipicephalus sanguineus* (Ixodidae): Studies on Engorged Larvae." *International Journal for Parasitology* 169-175.
- Henry, James A, Kenneth M Portier, and Jan Coyne. 1994. *The Climate and Weather of Florida*. Sarasota: Pineapple Press.
- Hijmans, Robert J. n.d. *WorldClim Issues*. Accessed June 18, 2017. <http://www.worldclim.org/issues>.
- Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces For Global Land Areas." *International Journal of Climatology* 1965-1978.
- Hopla, C. E. 1953. "Experimental Studies on Tick Transmission of Tularemia." *American Journal of Hygiene* 101-118.
- Hopla, C. E. 1955. "The Multiplication of Tularemia Organisms in the Lone Star Tick." *American Journal of Hygiene* 371-380.
- Hosmer, David W., and Stanley Lemeshow. 1989. *Applied Logistic Regression*. New York: John Wiley & Sons.
- James, A. M, C Burdett, M. J McCool, A Fox, and P Riggs. 2015. "The Geographic Distribution and Ecological Preferences of the American Dog Tick, *Dermacentor variabilis* (Say), in the U.S.A." *Medical and Veterinary Entomology* 178-188.
- Jaworski, Deborah C., C. Jerry Bowen, and Nalinda B. Wasala. 2013. "A White-Tailed Deer/Lone Star Tick Model for Studying Transmission of *Ehrlichia chaffeensis*." *Vector-Borne and Zoonotic Diseases* 193-195.
- Jongejan, F., and G. Uilenberg. 2004. "The Global Importance of Ticks." *Parasitology* S3-S14.

- Joyner, Timothy Andrew, Larissa Lukhnova, Yerlan Pazilov, Gulnara Temiralyeva, Martin E. Hugh-Jones, Alim Aikimbayev, and Jason K. Blackburn. 2010. "Modeling the Potential Distribution of *Bacillus anthracis* under Multiple Climate Change Scenarios for Kazakhstan." *PLoS ONE*.
- Karl, Jason W., and Brian A. Maurer. 2010. "Spatial Dependence of Predictions from Image Segmentation: A Variogram-based Method to Determine Appropriate Scales for Producing Land-Management Information." *Ecological Informatics* 194-202.
- Keim, Barry D, Michell Russo Fischer, and Adam M Wilson. 2005. "Are there spurious precipitation trends in the United States Climate Division database?" *Geophysical Research Letters*.
- Keirans, James E., and Taina R. Litwick. 1989. "Pictorial Key to the Adulats of Hard Ticks, Family Ixodidae (Ixodida: Ixodoidea), East of the Mississippi River." *Journal of Medical Entomology* 435-448.
- Knulle, W., and D. Rudolph. 1981. "Humidity Relationships and Water Balance of Ticks." In *Physiology of Ticks*, by FD. Obenchain and R. Galun, 43-50. Oxford (UK): Pergamon Press.
- Koch, Henry G. 1986. "Survival of the Lone Star Tick (Acari: Ixodidae) Under Flooding Conditions: A Laboratory Evaluation." *Journal of Economic Entomology* 1555-1557.
- Kollars Jr, Thomas M., James H. Oliver Jr, Lance A. Durden, and Peggy G. Kollars. 2000. "Host Associations and Seasonal Activity of *Amblyomma americanum* (Acari: Ixodidae) in Missouri." *Journal of Parasitology* 1156-1159.
- Kugeler, Kiersten J, Grace M Farley, Joseph D Forrester, and Paul S Mead. 2015. "Geographic Distribution and Expansion of Human Lyme Disease, United States." *Emerging Infectious Diseases* 1455-1457.
- Kumar, Sunil, and Thomas J. Stohlgren. 2009. "MaxEnt Modeling for Predicting Suitable Habitat for Threatened and Endangered Tree *Canacomyrica monticola* in New Caledonia." *Journal of Ecology and Natural Environment* 94-98.
- Landis, J. Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 159-174.
- Legendre, Pierre. 1993. "Spatial Autocorrelation: Trouble or New Paradigm." *Ecology* 1659-1673.
- Linthicum, K. J., C. L. Bailey, C. J. Tucker, K. D. Mitchell, T. M. Logan, F. G. Davies, C. W. Kamau, P. C. Thande, and J. N. Wagatoh. 1990. "Application of Polar-Orbiting, Meteorological Satellite Data to Detect Flooding of Rift Valley Fever Virus Vector Mosquito Habitats in Kenya." *Medical and Veterinary Entomology* 433-438.
- Masters, Edwin J., Chelsea N. Grigery, and Reid W. Masters. 2008. "STARI, or Masters Disease: Lone Star Tick-Vectored Lyme-like Illness." *Infectious Disease Clinics of North America* 361-376.
- Mertins, James W, Stacey L Vigil, and Joseph L Corn. 2017. "*Amblyomma auricularium* (Ixodida: Ixodidae) in Florida: New Hosts and Distribution Records." *Journal of Medical Entomology* 132-141.
- Miller, A., D. Cayan, T. Barnett, N. Graham, and J. Oberhuber. 1994. "The 1976-77 Climate Shift of the Pacific Ocean. ." *Oceanography* 21-26.

- Miller, Jennifer, Janet Franklin, and Richard Aspinall. 2007. "Incorporating Spaaatial Dependence in Predictive Vegetation Models." *Ecological Modeling* 225-242.
- Naimi, B., N. A. S. Hamm, T. A. Groen, A. K. Skidmore, and A. G. Toxopeus. 2014. "Where is positional uncertainty a problem for species distribution modelling?" *Ecography* 191-203.
- Ogden, N. H., M. Bigras-Poulin, C. J. O'Callaghan, I. K. Barker, A. Maarouf, K. E. Smoyer-Tomic, D. Waltner-Toews, and D. Charron. 2005. "A Dynamic Population Model to Investigate Effects of Climate on Geographic Range and Seasonality of the Tick *Ixodes Scapularis*." *International Journal of Parasitology* 375-389.
- Oliver Jr, James H. 1989. "Biology and Systematics of Ticks (Acari: Ixodida)." *Annual Review of Ecology and Systematics* 397-430.
- Oliver Jr, James H, Marc P Hayes, James E Keirans, and David R Lavender. 1993. "Establishment of the Foreign Parthenogenetic Tick *Amblyomma rotundatum* (Acari: Ixodidae) in Florida." *Journal of Parasitology* 786-790.
- Ostfeld, Richard S., and Jesse L. Brunner. 2015. "Climate Change and Ixodes Tick-Borne Diseases of Humans." *Philosophical Transactions of the Royal Society B*.
- Ostfeld, Richard S., Gregory E. Glass, and Felicia Keesing. 2005. "Spatial Epidemiology: An Emerging (or Re-Emerging) Discipline." *TRENDS in Ecology and Evolution* 328-336.
- Ostfeld, Richard S., Obed M. Cepeda, Kirsten R. Hazler, and Michael C. Miller. 1995. "Ecology of Lyme Disease: Habitat Associations of Ticks (*Ixodes Scapularis*) In a Rural Landscape." *Ecological Applications* 353-361.
- Pearce, Jennie, and Simon Ferrier. 2000. "Evaluating the Predictive Performance of Habitat Models Developed using Logistic Regression." *Ecological Modeling* 225-245.
- Pearson, Richard G., Christopher J. Raxworthy, Miguel Nakamura, and A. Townsend Peterson. 2007. "Predicting Species Distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar." *Journal of Biogeography* 102-117.
- Pearson, Richard G., Terence P. Dawson, and Canran Liu. 2004. "Modelling Species Distributions in Britain: A Hierarchical Integration of Climate and Land-cover data." *Ecography* 285-298.
- Peterson, A. Townsend, Monica Papes, and Muir Eaton. 2007. "Transferability and Model Evaluation in Ecological Niche Modeling: A Comparison of GARP and MaxEnt." *Ecography* 550-560.
- Phillips, Steven J., Miroslav Dudik, and Robert E. Schapire. 2004. "A Maximum Entropy Approach to Species Distribution Modeling." *Proceedings of the 21st International Conference on Machine Learning* 655-662.
- Phillips, Steven J., Miroslav Dudik, Jane Elith, Catherine H. Graham, Anthony Lehmann, John Leathwick, and Simon Ferrier. 2009. "Sample Selection Bias and Presence-only Distribution Models: Implications for background and pseudo-absence data." *Ecological Applications* 181-197.
- Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modeling* 231-259.

- Porretta, Daniele, Valentina Mastrantonio, Sara Amendolia, Stefano Gaiarsa, Sara Epis, Claudio Genchi, Claudio Bandi, Domenico Otranto, and Sandra Urbanelli. 2013. "Effects of global changes on the climatic niche of the tick *Ixodes ricinus* inferred by species distribution modeling." *Parasites & Vectors* 1-8.
- Potts, Joanne, M., and Jane Elith. 2006. "Comparing Species Abundance Models." *Ecological Modelling* 153-163.
- Raghavan, Ram K., Douglas G. Goodin, Gregg A. Hanzlicek, Gregory Zolnerowich, Michael W. Dryden, Gary A. Anderson, and Roman R. Ganta. 2016. "Maximum Entropy-Based Ecological Niche Model and Bio-Climatic Determinants of Lone Star Tick (*Amblyomma americanum*) Niche." *Vector-Borne and Zoonotic Diseases* 205-211.
- Randolph, S. E., R. M. Green, A. N. Hoodless, and M. F. Peacey. 2002. "An empirical quantitative framework for the seasonal population dynamics of the tick *Ixodes ricinus*." *International Journal of Parasitology* 979-989.
- Randolph, S.E., L. Gern, and P.A. Nuttall. 1996. "Co-feeding Ticks: Epidemiological Significance for Tick-borne Pathogen Transmission." *Parasitology Today* 472-479.
- Reeves, Will K., Amanda D. Loftis, William L. Nicholson, and Alan G. Czarkowski. 2008. "The first report of human illness associated with the Panola Mountain Ehrlichia species: a case report." *Journal of Medical Case Reports*.
2017. "Reportable Diseases Frequency Report." Tallahassee: Florida Department of Health.
- Richards, John A. 2013. *Remote Sensing Digital Image Analysis: An Introduction*. New York: Springer.
- Sallam, Mohamed F., Azzam M. Al Ahmed, Mahmoud S. Abdel-Dayem, and Mohamed A. R. Abdullah. 2013. "Ecological Niche Modeling and Land Cover Risk Areas for Rift Valley Fever Vector, *Culex tritaeniorhynchus* Giles in Jazan, Saudi Arabia." *PLoS ONE*.
- Sayler, Katherine A., Anthony F. Barbet, Casey Chamberlain, William L. Clapp, Rick Alleman, Julia C. Loeb, and John A. Lednicky. 2014. "Isolation of Tacaribe Virus, a Caribbean Arenavirus, from Host-Seeking *Amblyomma americanum* Ticks in Florida." *PLoS ONE* 1-16.
- Sayler, Katherine, Carisa Boyce, and Samantha Wisely. 2016. *Facts about Wildlife Diseases: South Tick Associated Rash Illness of "STARI"*. WEC376, Gainesville: University of Florida Institute of Food and Agricultural Sciences.
- Schielzeth, Holger. 2010. "Simple means to improve the interpretability of regression coefficients." *Methods in Ecology and Evolution* 103-113.
- Semtner, P. J., and J. A. Hair. 1973. "The Ecology and Behavior of the lone star tick (Acarina: Ixodidae)." *Journal of Medical Entomology* 618-628.
- Smith Jr., H. M. 1981. *Parasites of wild swine (Sus scrofa) in the southern United States*. University of Georgia.
- Smith, R. D., D. M. Sells, E. H. Stephenson, M. R. Ristic, and D. L. Huxsoll. 1976. "Development of *Ehrlichia canis*, causative agent of canine ehrlichiosis, in the tick *Rhipicephalus sanguineus* and its differentiation from a symbiotic Rickettsia." *American journal of Veterinary Research* 119-126.

- Springer, Yuri P, Catherine S Jarnevich, David T Barnett, Andrew J Monaghan, and Rebecca J Eisen. 2015. "Modeling the Present and Future Geographic Distribution of the Lone Star Tick *Amblyomma americanum* (Ixodida: Ixodidae), in the Continental United States." *American Journal of Tropical Medicine and Hygiene* 875-890.
- Springer, Yuri P., Lars Eisen, Lorenza, James, Angela M. Beati, and Rebecca J. Eisen. 2014. "Spatial Distribution of Counties in the Continental United States with Records of Occurrence of *Amblyomma americanum* (Ixodida: Ixodidae)." *Journal of Medical Entomology* 342-351.
- Stafford, Kirby C., Jeffrey S. Ward, and Louis A. Magnarelli. 1998. "Impact of Controlled Burns on the Abundance of *Ixodes scapularis*." *Journal of Medical Entomology* 510-513.
- Stockwell, David. 1999. "The GARP Modeling System: Problems and Solutions to Automated Spatial Prediction." *International Journal of Geographical Information Science* 143-158.
- Strickland, Robert K., Robert R. Gerrish, James L. Hourrigan, and Glen O. Schubert. 1976. *Ticks of Veterinary Importance*. Animal and Plant Health Inspection Service, U.S. Dept of Agriculture.
- Sumner, John W., Lance A. Durden, Jerome Goddard, Ellen Y. Stromdahl, Kerry L. Clark, Will K. Reeves, and Christopher D. Paddock. 2007. "Gulf Coast Ticks (*Amblyomma maculatum*) and *Rickettsia parkeri*, United States." *Emerging Infectious Diseases* 751-753.
- Symonds, Matthew R. E., and Adnan Moussalli. 2011. "A Brief guide to Model Selection, Multimodel Inference and Model Averaging in Behavioural Ecology using Akaike's Information Criterion." *Behavioral Ecology and Sociobiology* 13-21.
- Taylor, Doyle Jackson. 1951. "The Distribution of Ticks in Florida." *University of Florida, M.S. Thesis*.
- Thornton, P.E., M.M. Thornton, B.W. Mayer, Y. Wei, R. Devarakonda, R.S. Vose, and R.B. Cook. 2017. "Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3." Oak Ridge, Tennessee: ORNL DAAC.
- Tsai, HuiPing, Jane Southworth, and Peter Waylen. 2014. "Spatial persistence and temporal patterns in vegetation cover across Florida, 1982-2006." *Physical Geography* 151-180.
- U.S. Geological Survey. 1987. *Map projections: A working manual*. USGS Numbered Series, Washington, D.C.: U.S. Government Printing Office.
- USGCRP. 2016. *The Impacts of Climate Change on Human Health in the United States: A Scientific Assessment*. Washington, DC.: U.S. Global Change Research Program.
- Vail, Stephen G., and Gary Smith. 2002. "Vertical Movement and Posture of Blacklegged Tick (Acari: Ixodidae) Nymphs as a Function of Temperature and Relative Humidity in Laboratory Experiments." *Journal of Medical Entomology* 842-846.
- Wehinger, K. A., M. E. Roelke, and E. C. Greiner. 1995. "Ixodid ticks from panthers and bobcats in Florida." *Journal of Wildlife Disease* 480-485.
- Weiler, Martin, Georg Gerhard Duscher, Monika Wetcher, and Julia Walochnick. 2017. "Tick Abundance: A One Year Study on the Impact of Flood Events Along the Banks of the River Danube, Austria." *Experimental and Applied Acarology* 151-157.

- Westphal, Michael I., S. A. Field, A. J. Tyre, D. Paton, and H. P. Possingham. 2003. "Effects of Landscape Pattern on Bird Species Distribution in the Mt. Lofty Ranges, South Australia." *Landscape Ecology* 413-426.
- Whittingham, Mark J., Philip A. Stephens, Richard B. Bradbury, and Robert P. Freckleton. 2006. "Why Do We Still Use Stepwise Modelling in Ecology and Behaviour." *Journal of Animal Ecology* 1182-1189.
- Wintle, Brendan, A., Jane Elith, and Joanne M. Potts. 2005. "Fauna Habitat Modelling and Mapping: A Review and Case Study in the Lower Hunter Central Coast Region of NSW." *Austral Ecology* 719-738.
- Yabsley, Michael J, Todd N Nims, Mason Y Savage, and Lance A Durden. 2009. "Ticks and Tick-Borne Pathogens and Putative Symbionts of Black Bears (*Ursus americanus floridanus*) from Georgia and Florida." *Journal of Parasitology* 1125-1128.
- Yoder, Jay A, and Justin L Tank. 2006. "Similarity in Critical Transition Temperature for Ticks Adapted for Different Environments: Studies on the Water Balance of Unfed Ixodid Larvae." *International Journal of Acarology* 323-329.
- Yoder, Jay A., and Joshua B. Benoit. 2003. "Water Vapor Absorption by Nymphal Lone Star Tick, *Amblyomma americanum* (Acari: Ixodidae), and its Ecological Significance." *International Journal of Acarology* 259-264.
- Zimmermann, Fridolin, and Urs Breitenmoser. 2002. "A Distribution Model for the Eurasian Lynx (*Lynx lynx*) in the Jura Mountains, Switzerland." *Predicting Species Occurrences*.

BIOGRAPHICAL SKETCH

William Kessler received his undergraduate education at the New Mexico Institute of Mining and Technology in 2010. He pursued a bachelor of science in biology and graduated cum laude in 2014. In the fall of 2015 William began his master's degree in geography at the University of Florida. His research interests include the study of vector-borne diseases, infectious disease modeling, and public health. William's goals upon graduation include continuing his academic career as a PhD student studying epidemiology.