Lyme Disease and Forest Fragmentation in the Peridomestic Environment

Pyrros Alexander Telionis

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Geography

Korine N. Kolivras, Chair
Yang Shao
Bryan L. Lewis

February 21, 2020
Blacksburg, Virginia

Keywords: Autocovariate, Black-legged Tick, Boosted Regression Trees, *Borrelia burgdorferi*, Edge Habitat, Forest-Herbaceous, Forest Fragmentation, Forest Patch, Geographic Information Systems, GIS, Incidence, Infectious Disease, *Ixodes scapularis*, Land Cover, Lyme Disease, Maximum Entropy, Niche Model, Spatial Autocorrelation.

Lyme Disease and Forest Fragmentation in the Peridomestic Environment

Pyrros Alexander Telionis

ABSTRACT

Over the last 20 years, Lyme disease has grown to become the most common vector-borne disease affecting Americans. Spread in the eastern U.S. primarily by the bite of *Ixodes scapularis*, the black-legged tick, the disease affects an estimated 329,000 Americans per year. Originally confined to New England, it has since spread across much of the east coast and has become endemic in Virginia. Since 2010 the state has averaged 1200 cases per year, with 200 annually in the New River Health District (NRHD), the location of our study.

Efforts to geographically model Lyme disease primarily focus on landscape and climatic variables. The disease depends highly on the survival of the tick vector, and white-footed mouse, the primary reservoir. Both depend on the existence of forest-herbaceous edge-habitats, as well as warm summer temperatures, mild winter lows, and summer wetness. While many studies have investigated the effect of forest fragmentation on Lyme, none have made use of high-resolution land cover data to do so at the peridomestic level.

To fill this knowledge gap, we made use of the Virginia Geographic Information Network's 1-meter land cover dataset and identified forest-herbaceous edge-habitats for the NRHD. We then calculated the density of these edge-habitats at 100, 200 and 300-meter radii, representing the peridomestic environment. We also calculated the density of <2-hectare forest patches at the same distance thresholds. To avoid confounding from climatic variation, we also calculated mean summer temperatures, total summer rainfall, and number of consecutive days below freezing of the prior winters. Adding to these data, elevation, terrain shape index, slope, and aspect, and including lags on each of our climatic variables, we created environmental niche models of Lyme in the NRHD. We did so using both Boosted Regression Trees (BRT) and Maximum Entropy (MaxEnt) modeling, the two most common niche modeling algorithms in the field today.

We found that Lyme is strongly associated with higher density of developed-herbaceous edges within 100-meters from the home. Forest patch density was also significant at both 100-meter and 300-meter levels. This supports the notion that the fine scale peridomestic environment is significant to Lyme outcomes, and must be considered even if one were to account for fragmentation at a wider scale, as well as variations in climate and terrain.

Lyme Disease and Forest Fragmentation in the Peridomestic Environment

Pyrros Alexander Telionis

GENERAL AUDIENCE ABSTRACT

Lyme disease is the most common vector-borne disease in the United States today. Infecting about 330,000 Americans per year, the disease continues to spread geographically. Originally found only in New England, the disease is now common in Virginia. The New River Health District, where we did our study, sees over 200 cases per year.

Lyme disease is mostly spread by the bite of the black-legged tick. As such we can predict where Lyme cases might be found if we understand the environmental needs of these ticks. The ticks themselves depend on warm summer temperatures, mild winter lows, and summer wetness. But they are also affected by forest fragmentation which drives up the population of white-footed mice, the tick's primary host. The mice are particularly fond of the interface between forests and open fields. These edge habitats provide food and cover for the mice, and in turn support a large population of ticks.

Many existing studies have demonstrated this link, but all have done so across broad scales such as counties or census tracts. To our knowledge, no such studies have investigated forest fragmentation near the home of known Lyme cases. To fill this gap in our knowledge, we made use of high-resolution forest cover data to identify forest-field edge habitats and small isolated forest patches. We then calculated the total density of both within 100, 200 and 300 meters of the homes of known Lyme cases, and compared these to values from non-cases using statistical modeling. We also included winter and summer temperatures, rainfall, elevation, slope, aspect, and terrain shape.

We found that a large amount of forest-field edges within 100 meters of a home increases the risk of Lyme disease to residents of that home. The same can be said for isolated forest patches. Even after accounting for all other variables, this effect was still significant.

This information can be used by health departments to predict which neighborhoods may be most at risk for Lyme. They can then increase surveillance in those areas, warn local doctors, or send out educational materials.

Dedication

To my wife and father, for believing when I didn’t... and to ο Μεγάλος and ο Πρόεδρος, to Zettaki and Lulaki, and to my mom Vasso, who should have been here… μας λείπεις.

Acknowledgements

I would like to acknowledge my committee members, Drs. Kolivas, Shao, Lewis, and Abbas for their patience and time. Indeed, it was a wonderful opportunity to join the Department of Geography and to further my understanding of the field beyond simply understanding how to employ ArcGIS. I would also like to acknowledge my cohort and friends in the Kolivras Lab, namely Danielle, Elizabeth, Logan, Marissa, Megan, Molly and Shreejana. Additionally, I thank the Virginia Department of Health for trusting us with these data and allowing us to help with such a significant problem. I also thank Ms. Karen Bland, and Ms. Dennie Munson for the great help in navigating the graduate school's requirements. Finally, I'd like to acknowledge ArcGIS for behaving long enough for me to complete this work, which was far from guaranteed at the onset.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AUC	“Area Under the [ROC] Curve” (see ROC)
BRT	“Boosted Regression Trees” (a machine-learning algorithm)
CDC	“US Centers for Disease Control and Prevention”
DIN	“Density of Infected Nymphs”
Esri	“Environmental Systems Research Institute”
FDA	“Food and Drug Administration”
MaxEnt	“Maximum Entropy” (a machine-learning algorithm)
NLCD	“National Land Cover Database”
NRHD	“New River Health District (Virginia)”
OspA	“Outer Surface Protein A”
OspC	“Outer Surface Protein C”
PHDI	“Palmer Hydrological Drought Index”
ROC	“Receiver Operating Characteristic”
TSI	“Terrain Shape Index”
USGS	“United States Geological Survey”
VDEQ	“Virginia Department of Environmental Quality”
VDH	“Virginia Department of Health”
VGIN	“Virginia Geographic Information Network”

Chapter

1. Statement of Problem

1.1 Statement of Problem

While the news focuses on more sensational and frightening tropical pathogens like Ebola, Lyme disease has quietly grown to become America's most impactful vector-borne disease. With approximately 33,000 infections per year (CDC 2019), and an underreporting rate of roughly 10:1 (Coyle et al. 1996), the disease is responsible for significant morbidity across the United States. Though generally limited to arthritic degradation, it is also capable of causing a variety of neurological and cardiovascular issues and is suspected of causing some mucosal cancers. It is also a source of significant economic losses, as an estimated 25% of Lyme patients leave work and file for disability insurance. Though a vaccine, LYMErix™, was approved by the Food and Drug Administration in 1998, it was recalled only three years later (Nigrovic and Thompson 2007), leaving tick-habitat avoidance and judicial tick removal as the only effective means to prevent Lyme disease. Firstline treatment involves a month-long course of doxycycline or amoxicillin (Hansmann 2009), which can cause other side effects.

Though ancient in roots, Lyme was first identified in Connecticut in 1975 when physicians noticed an unusual cluster of what was thought to be juvenile arthritis. It has since spread across New England, the Mid-Atlantic, and the Midwest. Virtually unheard-of in Virginia prior to 2000, Lyme has since become endemic across the commonwealth. Once established Lyme cases numbered around 250 per year until an increase in 2007. Since then the commonwealth averages about 1200 cases per year. The New River Health District (NRHD) in specific had almost no cases prior to 2010, but now suffers about 250 cases per year. Indeed, spatial scan studies have found the NRHD to be one of the two major hotspots in the commonwealth, with the other being in the Northern Virginia area (Li et al. 2014).

Geographic studies of Lyme disease are common and generally focus on the landscape and climate required to support the lifecycle of the disease. Elevation, forest fragmentation, mild winters, and wet warm summers are required for the survival of the primary vectors and reservoirs, the blacklegged tick and white-footed mouse, respectively. Forest fragmentation is considered critical as the forest-herbaceous edges provide the ideal habitat for the survival of the mice. The fragmentation also reduces species diversity, eliminating the presence of other mammalian dilution hosts which drive down Lyme prevalence. The whitetail deer that serve as reproductive hosts and aid in the distribution of the ticks are also fond of forest-edge habitats, which provide them access to nutritious ornamental plants.

Though these associations between Lyme and fragmentation have been studied at broad scales such as the census-tract level (Seukep et al. 2015) or with 800-meter buffers (Moon et al. 2019), a knowledge gap remains in our understanding of the role of peridomestic forest fragmentation in the Lyme disease cycle near the home. Indeed, such studies are difficult to implement as land cover data often lacks the resolution needed to identify such forest fragments. The National Land Cover Dataset (NLCD), commonly used in such studies, has a resolution of 30-meters. This makes it virtually impossible to use the product to find small forest patches, or accurately calculate the length of edge-habitat in an individual's back yard.

The creation of the Virginia Geographic Information Network's (VGIN) high resolution land cover dataset gives us a unique opportunity to investigate Lyme disease in the peridomestic environment. With a 1-meter resolution, 900 times the resolution of the NLCD, we can accurately extract forest-edge habitats, as well as identify forest patches a few square-meters in size. This allows us to address the gap in our understanding of Lyme disease in peridomestic environments and investigate the effects of fine scale forest fragmentation on Lyme disease outcomes. Specifically, we seek to answer the following research questions:

1. How does forest fragmentation in the peridomestic environment affect Lyme disease?
2. Does high resolution land cover data improve Lyme models?

If we demonstrate that these effects are indeed significant, and that one cannot accurately estimate Lyme risk without considering the peridomestic environment, health departments could accurately identify areas of Lyme risk and better target education campaigns to the public. It would also inform future land use efforts, allowing city planners to avoid creating neighborhood landscapes that encourage the propagation of Lyme disease. Finally, it may encourage the development of similar high-resolution land cover data for other states afflicted with Lyme.

Chapter

2. Literature Review

2.1 Introduction

Once confined to the coastal forests on the northern shores of Long Island Sound, Lyme disease has spread far beyond the eponymous Connecticut town in which it was first identified. Now a threat to the entire Northeast and Mid-Atlantic, Lyme disease has become endemic in temperate Virginia. With 744 confirmed cases and 395 probable cases in 2018 (CDC 2015a), the commonwealth faces a significant and growing burden from this bacterial interloper.

Lyme cannot be transmitted by person-to-person contact, and accordingly is only found in areas where both the vector and reservoir species thrive. Consequently, the spatial distribution of Lyme disease is exceedingly heterogeneous. Combining spatial analysis with medical ecology, the field of medical geography is uniquely qualified to investigate the factors that drive this heterogeneity. A better understanding of this distribution, and the causal factors driving it, will allow the Department of Health to forecast future incidence rates, improve the targeting of anti-Lyme interventions, and uncover areas of systemic underreporting.

As we will demonstrate shortly, the consensus among the medical and scientific community is that Lyme disease is strongly associated with the edge-habitats preferred by the disease's reservoirs and vectors. Studies investigating habitat fragmentation and the resulting effects on the aforementioned reservoirs and vectors are also well known. However, fine scale analyses into the peridomestic environment, needed to properly quantify the risks of Lyme in rural and suburban environments, are rare and underrepresented in the literature. Furthermore, spatiotemporal analyses investigating the impact of land cover change on the incidence of Lyme is not present in the literature. The latter is particularly relevant as the Great Recession of 2007 halted many residential development projects, leaving behind numerous unfinished lots and potentially creating the ideal habitat for the proliferation of Lyme disease. A fine scale analysis may also allow the health department to identify specific neighborhoods under threat from the disease and mail educational materials directly to the homeowners most at risk. The same work could be used to better inform physicians in the area as to the risks posed to their patients, particularly those living in high risk regions.

In this study, we will perform a fine scale examination of Lyme disease incidence in Virginia's New River Health District to determine if there an observable and quantifiable association between Lyme disease presence and the peridomestic environment. Specifically, we seek to determine if we can objectively demonstrate that the risk of Lyme disease in the community is a function of forest fragmentation, and in conjunction with other factors, if we can reliably predict presence in similar areas.

2.2 Lyme Borreliosis

First identified as “erythema chronicum migrans” in Old Lyme Connecticut in 1975 (Mast and Burrows 1976), Lyme disease is today found across North America and Continental Europe (Habálek and Halouzka 1997; CDC 2015b). On the American continent, the primary etiological agent is the gram-negative spirochete *Borrelia burgdorferi* (Burgdorfer et al. 1982), though at least eight other members of the *Borrelia* genus are known to cause similar symptoms in Europe and Asia (Stanek et al. 2012). Included in this list is relative newcomer, *Borrelia mayoii*, which also seems to cause Lyme-like illness (Pritt et al. 2016).

The primary vector in Virginia is the blacklegged tick, *Ixodes scapularis*, colloquially called the “deer tick” or “seed tick”, that is infected in turn by feeding upon the reservoir species (Donahue et al. 1987). Note that some older literature refers to the tick as *Ixodes dammini*. Originally thought to be a separate species, *I. dammini* has since been proven to be a morph of *I. scapularis* (Oliver et al. 1993). For the sake of simplicity, we will henceforth use the latter name to refer to both morphs. Despite the common name of the tick, the primary host of the nymphs and primary reservoir of the pathogen in question is the white-footed mouse, *Peromyscus leucopus*, a ubiquitous rodent that can be found across the eastern seaboard. The competence of the mice as a reservoir species was demonstrated effectively by Donahue et al. (1987) who allowed lab-reared uninfected ticks to feed on infected mice, before confirming the presence of the pathogen in the midgut of the ticks via immunofluorescence. Bacteremia and resulting infectivity in the mice peaked after one week of exposure, but the mice remained a source of the pathogen for the entire nine-week study. On the other hand, the whitetail deer, *Odocoileus virginianus*, implicated by the colloquial name of the tick, is not a competent host for the pathogen (Telford III et al. 1988). Deer were originally thought to be noteworthy hosts, and indeed spirochetes have been isolated from blood drawn from whitetail deer (Bosler et al. 1984), but the Telford III et al. (1988) study showed that the deer do not develop bacteremia sufficient to infect uninfected ticks, and are accordingly incompetent hosts.

Nevertheless, several secondary reservoir hosts are present in North America. Most notable are chipmunk species of the *Sciuridae* family (Anderson et al. 1985) and cottontail rabbits, *Sylvilagus floridanus* (Anderson et al. 1989). Spirochetes have also been isolated from mid-sized mammals such as raccoons and opossums (Anderson et al. 1983), and several bird species (Anderson et al. 1986). Still, the white-footed mouse is considered the primary reservoir of concern and the most consequential driver of Lyme disease cases (Lane et al. 1991).

Though exonerated of serving as reservoir hosts by molecular evidence, the whitetail deer do play a role in the propagation of Lyme by maintaining the adult tick population and transporting the ticks across relatively large distances (Telford III et al. 1988). Whitetail deer density is strongly associated with the density of larval ticks (Wilson et al. 1985), and yearling dispersal provides a means by which ticks may find themselves transported across relatively large distances. Though the home range of whitetail deer is normally limited to a few hundred hectares, approximately 20% of yearling females will leave their natal ranges between late May and late June, potentially traveling more than 100 kilometers, and averaging 3.7 km per day (Barton et al. 1995). The males are even more likely to seek independence, with 32% setting out on their own and traveling on

average 4.4 km per day (Barton et al. 1995). When one considers that the average *Ixodes* tick will feed for four days before detaching itself of its own volition (Piesman et al. 1987), it is entirely possible for a particularly fortuitous tick to travel great distances in a single feeding. Since the white-footed mouse is particularly averse to travel and is generally contained in an area measuring a few hundred meters across (Ormiston 1984), it seems reasonable to assume that the deer are culpable when it comes to the geographic spread of Lyme. *Ixodes* ticks also feed on several bird species (Battaly and Fish 1993), and may be moved around by their migrations.

The lifecycle of the bacteria themselves is rather unusual. While circulating within the bloodstream of the reservoir, the spirochetes express outer surface protein A (OspA) that allows them to colonize the midgut of the ticks which imbibe them (de Silva et al. 1996). After colonization, a second blood meal stimulates the bacteria to express outer surface protein C (OspC), allowing colonization of vertebrates, as well as migrate from the midgut through the hemolymph to the salivary glands of the tick (Schwan and Piesman 2000). This process takes approximately 24-48 hours, explaining the unusually long delay for a tick to pass Lyme disease on to a host (Schwan and Piesman 2000). Removal of the tick within the first 24 hours is extremely effective at preventing transmission; unfortunately, the small size of the *Ixodes* species makes them very difficult to detect or identify, especially for patients who are unaware of the risk.

Of note, an effective Lyme disease vaccine was available to the general public between 1998 and 2002. Originally developed by SmithKline Beecham, the vaccine had an efficacy of 76% for adults and nearly 100% for children (Poland and Jacobson 2001). The vaccine itself was rather unusual in that it targeted OspA, the surface protein found on the spirochetes residing in the tick's midgut (de Silva et al. 1996). It therefore prevented the bacteria from ever expressing OspC, infective to humans, or moving from the midgut to the salivary glands. The antibodies targeted the bacteria within the tick itself, preventing the human host from ever being exposed to a single *B. burgdorferi* microbe. Lamentably, the growth of the then fledgling anti-vaccine movement, the failure of insurance companies to reimburse patients for the vaccine, and unproven claims that it caused dangerous autoimmune disorders (Abbott 2006), forced the vaccine from the market (Nigrovic and Thompson 2007). As there is no FDA approved replacement vaccine, targeted educational interventions are crucial to combatting the spread of this disease.

2.3 Burden of Lyme Disease

The incidence rate of Lyme disease has increased significantly in recent years. Though the cause of this increase is still up for debate, and indeed we seek to examine the potential role of forest fragmentation in this process, the burden of Lyme disease is undeniable. In the last fifteen years, the incidence in Virginia has increased by an order of magnitude, from 1.65 to 11.7 cases per 100,000 person-years, in 2001 and 2014 respectively (CDC 2015a). The same study showed a nearly threefold jump between 2006 and 2007, with 357 confirmed cases in the former year versus 959 cases in the latter. Similar changes were observed in neighboring states such as West Virginia, but not in New England or the Midwest, implying that the change is likely not the result of improved reporting, or a change in the national case definition that occurred in 2008.

While it is easy to recognize the burden posed by the disease, measuring that burden is a considerably more difficult proposition. There is substantial incongruity within the medical community concerning the very existence of chronic Lyme disease, with some claiming that it has become an unscientific catchall for difficult-to-diagnose symptoms (Feder et al. 2007), while others claim to have identified and successfully treated it with a year-long course of tetracycline (Donta 1997). Accordingly, we will avoid attempting to quantify the encumbrance posed by chronic Lyme disease. On the other hand, the impact of late-stage Lyme disease, which includes musculoskeletal and neurological sequelae, is well-understood and can be quantified. Noting that the average treatment costs of early-stage and late-stage symptomatic Lyme disease are $464 and $1380 respectively (Zhang et al. 2006), and that Lyme is underreported by a factor of between six to twelve (Coyle et al. 1996) we can generate a very rough estimate of the burden on Virginia. Taking the best case scenario, all patients are treated in early stages, unreported cases are only six times the reported cases, and discarding the 370 probable cases as type I errors, we are still left with 5,856 cases in 2014 at a direct treatment cost of $2.7 million. Indirect costs are likely considerably higher considering that an estimated 25% of patients received state disability benefits and 37% of those awarded benefits continued on disability programs for more than five years after treatment (Johnson et al. 2011). We are not prepared to estimate the indirect costs associated with loss of productivity from missed work and loss of employment, though this must be substantial.

While the financial burden is frightening, the physical burden on infected individuals is far more personal. While Lyme rarely kills, it is capable of causing significant cardiovascular and neurological disorders, including life-threatening heart arrhythmias and heart blockages, as well as seizures, potentially fatal meningitis, debilitating facial palsies and even serious liver damage and cirrhosis (CDC 1997). In chronic form, it can cause critical and permanent arthritic degradation of joints, and occasionally, in neuropsychiatric form, can lead to significant cognitive impairment including paranoia, mania and even auditory and visual hallucinations (Fallon and Nields 1994). Lyme has even been implicated as the causative agent of certain gastric cancers such as cutaneous mucosa-associated lymphoid tissue lymphoma (Guidoboni et al. 2006). Though rare, such lymphomas have a five year survival rate of between 20 and 50% (Moore and Wright 1984).

Though potentially suffering from significant self-selection bias, surveys of the public paint a dire picture concerning the diagnosis and treatment of Lyme. In one study, 84% of respondents reported that proper diagnosis took in excess of four months and required visiting several specialists (Johnson et al. 2011). With a spike in incidence in the early 2000s, Lyme is now considered the most burdensome vector-borne disease in the United States (Bacon et al. 2008). With no available vaccine, and most cases going undiagnosed, we feel that the value of a study of the risk factors driving this disease is self-evident.

2.4 Medical Geography

Fortunately, medical geography is perfectly suited to helping us understand the processes that drive Lyme disease. The theoretical basis for our work lies in the triad of landscape epidemiology, which posits that disease emergence is a function of interactions between the host, the pathogen and the environment that they share (Snieszko 1974). Our work is also informed by the triangle of human ecology, which specifies that one cannot fully understand health without considering population, behavior, and environment, as well as the interactions between the three (Meade and Emch 2010).

Though the field's nascent beginnings lie with Hippocrates, who noted spatiotemporal variations in malaria in Periclean era Greece, modern efforts began in earnest when Dr. John Snow used spatial analysis to demonstrate the culpability of the Broad Street pump in the London cholera outbreak of 1854 (Snow 1855). Since that time, medical geography has continually provided invaluable tools to epidemiologists and physicians fighting disease outbreaks. In a world of rapid computerized spatial modeling, medical geography even informs outbreak responses as they are ongoing. A good example of such work was when researchers at Oxford University were able to generate an ecological niche model to predict the range of ebolavirus reservoirs during the 2014 West African Ebola outbreak (Pigott et al. 2014). Prior to the study there was considerable debate within the medical community concerning the possibility of zoonotic reintroductions and the need for interventions to take these into account. It was theorized that the epidemic was occurring outside the natural range of the reservoirs, but the geographic study demonstrated that the primary reservoirs were indeed endemic to the area and that zoonotic transmission remained a possibility. The study, published less than two months after a state of emergency was declared by the president of Liberia, is an elegant example of the power of medical geography when applied to problems of zoonotic disease spread.

Medical geographers are quite familiar with Lyme disease as well. A literature search found over sixty refereed journal articles concerning spatial modeling of Lyme, with more than two thirds published within the last fifteen years. Researchers have recognized the power of GIS modeling to supplement exploratory field work since the mid-1990s (Glass et al. 1995a). Though not a replacement for ground truth, spatial analyses have since proven accurate enough to replace randomized tick dragging in risk modeling. In fact, modern GIS techniques are accurate enough to predict the presence or absence of ticks in areas never before surveyed; such was the case when researchers predicted the existence of *Ixodes* ticks on Queen Charlotte Island, where they had never been observed, and later confirmed these predictions with traditional tick dragging techniques (Mak et al. 2010).

In our study area, the omnipresent nature of both reservoirs and vectors makes it quite difficult to predict the attack rate of the disease at a local level without extremely detailed spatial analyses. Since the deer, mice, and ticks are all well distributed across the entire state, one may be tempted to expect a uniform distribution of the disease. This is clearly not the case. On the contrary, Lyme has spread in a southwestern direction since the early 2000s and now is found in distinct clusters across the entire commonwealth (Brinkerhoff et al. 2014). Research making use of the novel space-time scan statistic have demonstrated that multiple spatiotemporal clusters of the disease

are found in Virginia (Li et al. 2014). Moreover, quite fortuitously, our study area overlaps a disease cluster that first appeared after 2003 in the Li et al. (2014) study and was not observed by Brinkerhoff et al. (2014) study until after 2005. This makes the area ideal for investigating the factors driving increased Lyme incidence as it is literally at the center of the phenomenon. The study area also benefits from being within a single health district of the Virginia Department of Health. As this district is managed by single epidemiologist, who has held the position since the early 2000s, we expect little confounding from reporting and filing inconsistencies.

2.5 Environmental Factors of Lyme

Though our intention is to investigate the impact of land use change and forest fragmentation, we must first acknowledge the role played by climate change. In Europe, increasing average temperatures are allowing the vectors to survive in higher elevations and colonize northern regions previously beyond their range (Lindgren et al. 2006). In the northern climates of New England, the density of *I. scapularis* larva and nymphs in an area can be almost entirely explained by temperature alone (Ogden et al. 2004), specifically the number of days above zero centigrade (Ogden et al. 2006). The density of leafy vegetation has also been implicated in the changing range of Lyme disease as both the vector and reservoir favor wooded transition zones (Dister et al. 1997), but one must note that this vegetation itself is also affected by climate change (Estrada-Peña 2002).

The association between Lyme disease incidence and other climactic variables such as winter harshness and drought conditions are also well studied. Tick density can be modeled as a function of the Palmer Hydrological Drought Indices over the last three years and harshness of the prior two years, explaining variability in otherwise homogeneous geographic ranges (Subak 2003). Again, one must note the influence of climate change that contribute to these drought conditions and winter harshness. We must conclude that the change in Lyme disease is inexorably tied with climate change, but we intend to show that human driven land use change also plays a consequential role as land cover conditions must support the Lyme disease transmission cycle in order for climate conditions to play a role.

2.6 Forest Fragmentation

Significant associations between Lyme incidence and underlying geology, elevation, slope, and soil type, are well known in the literature (Glass et al. 1995a). However, among environmental factors, the presence of edge-habitats seems to be the most consequential driver of the disease.

Forest fragmentation is strongly associated with a reduction of species diversity, and has been shown to negatively affect birds and mammals alike (Andren 1994). This is one of the most basic principles of freshmen ecology courses, but there is one consequential exception. Perhaps unique among the species of North America, the white-footed mouse actually prefers areas of forest fragmentation; favoring areas smaller than half a hectare in size, mouse density is inversely proportional to patch size (Nupp and Swihart 1996). This fragmentation has three effects on Lyme density: first it increases the density of white-footed mice who find themselves nearly without competition, second it increases the density of the *I. scapularis* ticks whose larva find the white-footed mouse an ideal host, and third, it increases the prevalence of zoonotic Lyme disease by

increasing the likelihood that the ticks will feed on the mice, rather than incompetent reservoirs unlikely to be present in a forest fragment. The whitetail deer, that play host to adult ticks and assist in their distribution, are also fond of edge-habitats, particularly those with ornamental plants which the deer feed upon (Maupin et al. 1991). This confluence of factors allows the adult ticks to thrive in edge-habitats and brings their nymphs in closer contact with humans.

As a result of the affinity of reservoirs and vectors to fragmented forest patches and edge-habitats, these factors are particularly notable among environmental drivers of Lyme disease incidence. Forest patch size itself has a considerable impact on the density of nymph and adult ticks. Multiple studies have shown a significant inverse relationship between patch size and nymph density, as well as patch size and nymph infection prevalence (Allan et al. 2003; Brownstein et al. 2005b). Brownstein et al. (2005) also noted an association between patch isolation and vector density. Accordingly, residence near fragmented forest areas poses a significant risk, while residence within a compact suburban neighborhood minimizes exposure to such habitats and is protective (Cromley et al. 1998a). Indeed, incidence is strongly associated with the density of peridomestic land-cover edges and forest area (Jackson et al. 2006a), with forest-herbaceous edges being the most significant contributors (Seukep et al. 2015). These factors must be accounted for in any model intended to predict Lyme incidence.

2.7 Conclusion

Multiple studies have explored the association between forest fragmentation and Lyme disease in fine scale areas. Often these investigations are done at a resolution no higher than that of census tracts, which is more than sufficient for their purposes. Even adjusting for climate change and static contributors such as elevation and drainage, it is evident that forest fragmentation is driving Lyme incidence. The question we seek to answer is whether it is possible to quantify this effect by examining the peridomestic environment via ultra-fine scale 1-meter analysis. We do not seek to merely confirm an association, but rather increase the state of knowledge by quantifying the strength of this association. Furthermore, as we have seen that the entire life cycle of Lyme disease is benefited by forest fragmentation and edge-habitats, it is entirely plausible that the recession of 2007, which left countless undeveloped lots, has inadvertently created the ideal habitat for Lyme propagation. Finally, we intend to inform future Health Department interventions targeting Lyme disease by allowing the accurate spatiotemporal forecasting of future Lyme incidence, and by identifying areas of particular risk within the New River Health District.

Chapter

3. Lyme Disease as a Function of Peridomestic Forest Fragmentation

Attribution

The following chapter derives from a manuscript we intend to submit to EcoHealth. Co-authors include Dr. Korine Kolivras who supervised this work, as well as Dr. David Gaines, chief entomologist of the Virginia Department of Health, who supplied data and expertise on the matter.

3.1 Abstract

The most common vector-borne disease in the United States, Lyme disease poses a significant risk to the health of the American public. Though most cases are limited to arthralgias, Lyme is also capable of causing significant neurological and cardiovascular illness. As the disease range continues to expand, Virginia has become a hot spot for Lyme cases.

Modeling efforts focused on Lyme are extensive and typically relate the disease incidence to abiotic factors such as temperature, elevation, rainfall, and forest fragmentation. These abiotic factors are the primary drivers of the density of the white-footed mouse and blacklegged tick, the primary host and vector of Lyme respectively. However, these efforts are generally limited to wide scale geographic areas. In this study we examine their effects in the peridomestic environment using high-resolution land cover data.

Using 1-meter land cover data we extracted habitat edges then calculated the density of these edges across our study area at different scales. Combined with forest patch density, summer temperature, rainfall, landscape metrics, and winter severity, we created Lyme suitability models using Boosted Regression Trees and Maximum Entropy modeling. We find that Lyme is highly associated with high densities of forest-herbaceous edges within 100 meters of the home. We also find a significant association between Lyme and the density of isolated forest patches at 100- and 300-meter levels. These results suggest that the peridomestic environment is significant and must be considered even if one accounts for confounding variables.

3.2 Introduction

3.2.1 The Burden of Lyme

First identified in Connecticut in 1975, Lyme disease has spread to afflict a significant portion of the American Northeast, Mid-Atlantic, and Midwest (CDC 2015a). With a total of 33,666 confirmed or probable cases in 2018 (CDC 2019) and an underreporting rate between 6:1 and 12:1 (Coyle et al. 1996), Lyme disease is the most common vector-borne illness in the United States today. Lyme continues to make inroads in Virginia as well. Lyme was almost unknown prior to 2000. After becoming endemic, it caused around 250 cases per year until a major upsurge in 2007. The cause of this surge remains a point of inquiry, but it occurred a year before the change in CDC case definition. Today Virginia sees on average 1200 cases per year, and the New River Valley, the focus of our study, averages 250 cases despite almost none detected prior to 2010.

Though Lyme is not associated with significant mortality, it can cause substantial morbidity and disability, specifically long-lasting arthritic degradations and neurological sequelae. In rare cases, Lyme can also cause seizures, arrhythmias, meningitis, facial palsy, and cirrhosis, as well as cognitive impairment (Fallon and Nields 1994), and mucosal lymphomas (Guidoboni et al. 2006). As the range of Lyme continues to expand (Lindgren et al. 2006), further investigation of the environmental determinants of this disease is critical.

3.2.2 The Geography of Lyme

With a very distinct ecological niche, Lyme disease has been studied extensively by medical geographers and landscape epidemiologists. In Virginia, space-time scan studies have demonstrated the existence of multiple disease clusters (Brinkerhoff et al. 2014; Li et al. 2014). Geographers have also noted significant relationships between Lyme incidence or tick density and a variety of environmental factors including: the Palmer Hydrological Drought Index (PHDI) (Subak 2003), winter and summer temperatures (Ogden et al. 2006), vegetation density (Dister et al. 1997), as well as geology, elevation, slope, and soil type (Glass et al. 1995a; Ferrell and Brinkerhoff 2018).

Multiple studies have also found that various metrics of forest fragmentation are significant drivers of Lyme. Others have noted that tick density is inversely proportional to the degree of isolation of small forest patches in the surrounding area (Allan et al. 2003; Brownstein et al. 2005b). Linear edge-habitat length, particularly at the interface between forest and herbaceous land cover types, is also a strong predictor of Lyme (Jackson et al. 2006a; Seukep et al. 2015).

While these factors have been studied extensively at the census tract level and at similar scales, a knowledge gap exists in our understanding of the interaction between Lyme and these factors at the fine scale peridomestic level. Although recent studies have attempted to investigate this (Moon et al. 2019), none have done so at scales smaller than 800-meters, or with the use of high resolution land cover data. As such, the purpose of this study is to examine the relationship of forest fragmentation metrics, in the fine scale peridomestic environment.

3.2.3 The Modeling of Lyme

At first glance, modeling Lyme incidence is a complex endeavor. Human Lyme itself is a function of the density of infected nymphs (DIN) of the tick *Ixodes scapularis* (Ostfeld et al. 2001). The nymphs are bound by the interactions of their reservoir, reproductive, and dilution hosts, which themselves are affected by predators at multiple scales (Levi et al. 2012). The hosts are also heavily dependent upon resource availability. In fact, one of the best predictors of Lyme incidence is the acorn crop density from two years prior (Ostfeld et al. 2001). A surplus of acorns leads to population increase of the white-footed mouse, *Peromyscus leucopus,* the primary reservoir for the nymphs (Lane et al. 1991). One must also account for the presence of reproductive hosts such as Whitetail deer, as well as dilution hosts and predator interference (Levi et al. 2012). To model all these interactions at a fine scale is a seemingly intractable problem.

We bypass these issues by citing the Eltonian noise hypothesis: so long as the underlying ecology remains stable, the specific interactions of these hosts are sufficiently random, and occur at such a fine scale, as to not adversely affect distributions across large study areas (Peterson 2014). Moreover, each contributing vector and host species is bound by abiotic environmental conditions. One could hypothetically generate a species distribution model for each, use the outputs to build the DIN-model, and then relate DIN to Lyme incidence. Indeed, even acorn density is effectively predicted by rainfall and temperature of the prior year. Instead of modeling these biotic factors, we employ a black box model (Peterson 2014), which relates Lyme disease incidence directly to the physical environment alone and assumes that the biotic factors are themselves modulated by said physical environment. Such models have been shown to be accurate even when projected

across continents (Peterson 2003). The primary limitation of this approach is in our inability to detect major changes in underlying ecology, such as the introduction of coyotes to an area.

3.2.4 Climatic Factors of Lyme

Multiple studies have found a strong association with temperature lows and tick survival (Brownstein et al. 2003), specifically the number of days below freezing with a one and two year lag times (Subak 2003; Ogden et al. 2005; Berger et al. 2014). The reproductive rate of the ticks is also affected by the same (Ogden et al. 2014). Temperature highs are also predictive of nymph density (Brownstein et al. 2005a), and is associated with faster development and increased questing activity to a point (Ogden and Lindsay 2016). Temperature extremes can cause mortality and reduce activity (Vail and Smith 1998; Ogden and Lindsay 2016). The ticks are also affected by mean temperatures (Brownstein et al. 2003; Ogden et al. 2004) and isothermality (Feria-Arroyo et al. 2014). One should note that acorn mast is also affected by temperature at a one-year lag (Koenig et al. 1996), which is significant in this case as it affects both reservoir and reproductive host (Feldhamer et al. 1989; Levi et al. 2012). Temperature is also predictive of the onset of the Lyme season, measured as degree-growing days from last freeze (Monaghan et al. 2015).

Tick survival and Lyme incidence is also strongly affected by moisture levels. Multiple studies have found an association with May-June rainfall and precipitation in the wettest quarter and wettest months (McCabe and Bunnell 2004; Ostfeld et al. 2006; Feria-Arroyo et al. 2014). The ticks themselves require humid environments, and mortality increases significantly in dry seasons, though extreme rainfall also inhibits tick activity (Ogden and Lindsay 2016). Acorn density is strongly associated with rainfall on a one-year lag (Koenig et al. 1996) and models predicting Lyme onset also require spring rainfall data for the same year (Monaghan et al. 2015).

Humidity values are also associated with Lyme incidence. The survival of the ticks themselves depends directly on humidity (Bertrand and Wilson 1996; Ogden and Lindsay 2016), and high leaf-level humidity is associated with increased nymph activity (Vail and Smith 1998). Lyme incidence has been significantly related to the PHDI with a two year lag (Subak 2003). Nymph abundance was also found to be highly correlated with leaf-litter humidity levels with a two year lag (Berger et al. 2014), as well as monthly vapor pressure averaged across many years (Brownstein et al. 2003). Moreover, studies predicting the onset of Lyme cases at the beginning of the season found saturation deficits over the prior five weeks (Monaghan et al. 2015) and the two-year lagged PHDI values (Subak 2003) to be highly predictive.

3.2.4 Land Cover and Lyme

The effect of forest fragmentation on Lyme disease is so substantial that we cannot make any forecasts without it. The consensus is that forest fragmentation reduces mammalian diversity, but favors the white-footed mouse, which thrives in edge-habitats (Yahner 1992) and is a particularly competent reservoir host for *Borrelia burgdorferi* (Donahue et al. 1987). Several studies have found an inverse association between species richness diversity and Lyme disease (Ostfeld and Keesing 2001; Werden et al. 2014; Turney et al. 2018). Effectively, the loss of diversity drives Lyme. Associations between forest patch size and nymph density are also well established (Allan et al. 2003; Brownstein et al. 2005b), and the same can be said of forest patch size and mouse density (Nupp and Swihart 1996). Forest patch isolation is also a useful predictor of nymph density

(Brownstein et al. 2005b), as is patch frequency in a given area (Frank et al. 1998). The effects of edge-habitats have also been a focus of studies. The density of forest-edges within a given area is a good predictor for tick abundance (Das et al. 2002; Khatchikian and Prusinski 2012), and distance from forest edges is a useful predictor of Lyme incidence (Glass et al. 1995b; Horobik et al. 2006). The greatest effect is observed when forested areas interact with herbaceous terrain such as grasslands or open fields. Areal Lyme incidence has been directly linked to the percentage of herbaceous terrain edges bordering forest (Jackson et al. 2006a), as well as the total edge-contrast index of forest-herbaceous edges (Seukep et al. 2015). Though some studies in northern regions have successfully incorporated normalized difference vegetation index values into models of Lyme incidence (Kitron and Kazmierczak 1997), in Virginia specifically, the total forested area has limited predictive power (Seukep et al. 2015).

Despite the strong contribution from forest areas, we cannot ignore the remaining classes of land cover. Well-maintained lawns are negatively correlated with Lyme incidence (Frank et al. 1998) and tick density (Maupin et al. 1991). The ticks are negatively affected by agricultural lands and highly developed areas (Das et al. 2002). Even medium density development reduces tick abundance by a factor of ten when compared to low density residential settings (Cromley et al. 1998b). The odds ratio of someone acquiring Lyme in a high-density residential area is less than a third that of someone in a neighboring low-density area (Glass et al. 1995b). Population density is also negatively correlated with Lyme incidence (Seukep et al. 2015). On the other hand, excluding well maintained lawns, herbaceous areas are positively correlated with Lyme incidence (Jackson et al. 2006a; Seukep et al. 2015).

3.2.5 Purpose of Study

Despite the thorough understanding of the environmental factors driving Lyme disease, the majority of literature on the subject focuses on broad scale analysis. GIS based studies of Lyme at the census tract level (Seukep et al. 2015), in 800-meter buffers (Moon et al. 2019), or in 10-km^2 cells (Jackson et al. 2006b) are well known. But none focus on fine scale analysis of the peridomestic environment. Furthermore, most studies use the National Land Cover Dataset (NLCD) to identify forest edges. The 30x30-meter resolution of this product does not suffice for fine scale analysis. To this end, the purpose of this study is to make use of high-resolution land cover data, to examine the effect of forest fragmentation at the peridomestic level. We propose that the inclusion of fine scale fragmentation metrics are critical to properly calculating Lyme risk for an individual.

3.3 Methods and Materials

3.3.1 Study Area

The study area in question is the New River Health District (NRHD) of western Virginia, a constituent of Virginia Department of Health's (VDH) "Health Planning Region III". Nestled in the foothills of the Appalachian Mountains, the area is predominantly rural with a total population of about 180,000 spread across Montgomery, Pulaski, Giles and Floyd counties, as well as the independent city of Radford. With elevation ranging from 350 to 1329 meters, land cover is dominated by deciduous forest and managed pastures with small towns peppered in. The NRHD is a significant hotspot for Lyme disease, which was largely unknown to the area prior to 2007, but has since become a significant burden with over 200 cases per year since 2015. The area also exhibits a significant variation in forest fragmentation, making it ideal for our study design.

3.3.2 Lyme Case Data

In this work, we make use of Lyme disease case data from 2007 to 2016 collected by the VDH. Lyme is reportable disease in Virginia, and as such, diagnostic laboratories and physician offices have a duty to submit documentation to VDH should a patient meet the positive case criteria as established by the US Centers for Disease Control and Prevention (CDC). Accordingly, we feel these data should represent the majority of all detected cases in the study area. Our study area falls within the jurisdiction of a single health district, and the same district epidemiologist was responsible for the area during the entire study period. Though underreporting is an issue for all disease surveillance, we do not anticipate any spatial or temporal biases in the underreporting rate.

The VDH prepared these case data and sanitized them of all personal identifying information (PII) under the direction of the institutional review boards of both Virginia Tech and the VDH. We received only geocoordinates of the home address of each case, as well as year of diagnosis. VDH uses commercial geocoding software to resolve these geocoordinates. Though we cannot confirm the accuracy of these geocoordinates without personally identifying address data, this is the same procedure used by VDH for all internal work, and we are confident their internal procedures are robust and well validated.

Pseudo-absence points were created to represent the underlying structure of the study area. Following standard guidelines (Barbet-Massin et al. 2012; Phillips et al. 2016) we intended to have at least a ten to one ratio of pseudo-absence to presence points for any given year. As the maximum number of annual cases we recorded was 281, we created 2810 pseudoabsence points across the study area for each year of the study. Random point generation was biased by census tract population, but the final product was not exactly proportional to population. Rather, points were dropped at random, with the probability of being assigned to a census tract being proportional to that tract's population as a percentage of the total study area population.

3.3.3 Land Cover Data

To calculate forest fragmentation statistics, we made use of the high-resolution land cover data provided by the Virginia Geographic Information Network (VGIN) and the Virginia Department of Environmental Quality (VDEQ). With a spatial resolution of 1x1 meters, the VGIN dataset has 900

times the resolution of the more commonly used 30-meter NLCD. As with the NLCD, the VGIN dataset assigns a land cover code, such as “open water” or “forest”, to each cell in the raster. VGIN data are based on four-band orthoimagery taken between 2011-2015 and have an accuracy greater than 95% for forest and impervious surface tiles, as well as a minimum of 85% accuracy for all other land cover type. An example of VGIN data is seen in **Figure 1**.

VGIN data are confined to the geographic boundaries of Virginia. Though our study area is also limited to Virginia, we are concerned with forest fragmentation within a specified radius of homes. As such, some of our analysis footprint intruded a few hundred meters into parts of West Virginia. As there was no competing product to use for West Virginia, we manually classified these regions using Google Earth Pro and Esri ArcGIS. Note that these manually classified regions represent less than 0.1% of the total study area.

We reclassified the VGIN land cover data from the supplied 12-categories to five categories, merging similar variables such as “impervious extracted” and “impervious local data”. Using Esri ArcGIS we then converted the raster data to polygons, continuous areas of the same land cover type, then converted these polygons to lines representing the perimeter of each polygon. The result of this effort was a series of vector files representing the edges of each land cover type. Taking the intersection of any two such files allowed us to isolate the edge-habitat at the interface between any two land cover types in the study area. In this case, we focused on forest-herbaceous edges.

As the VGIN data used to create these edges come in a raster format, the resulting edge-habitat vectors are rectangular, limited to following a Manhattan-like grid. This artificially increases the length of linear edge-habitats in the study area. To correct this, we generalized our edge-habitats using a series of Simplify Line functions with varying distance thresholds from one to ten meters, using both the Douglas-Peucker and Wang-Müller algorithms via Esri ArcGIS. We then used Google Earth Pro to manually trace out 20 patch-edges of roughly 200-1200 meters in length. For use in our study, we selected the simplified edge-habitat lines that most closely matched the manually drawn lines as measured by Pearson correlation and mean-squared edge-length error. This process was followed by a Smooth Line operation with the same threshold.

With our edge-habitats in hand, we use focal statistics operations to create three 10x10 meter raster surfaces where the value of each cell represents the total density of linear forest-herbaceous edges within 100-, 200- and 300-meter radii, respectively. These values were chosen to represent the roughly 200-meter peridomestic threshold common in other epidemiological studies (Martínez-Vega et al. 2015).

With the polygonal data representing VGIN derived forest patches in hand, we also identified forest fragments in the study area. This included any isolated patch of forest with a total surface area of less than two hectares as is commonly used in forest fragmentation analysis (Jackson et al. 2006b; Seukep et al. 2015). After deleting all fragments larger than two hectares, we again used a 10x10 meter focal statistics operation to calculate the density of these forest fragments within the same 100-, 200-, and 300-meter radii from the cell center.

3.3.4 Complementary Data

Though our focus is on forest fragmentation, we sought to include other metrics of environmental and climatic data to reduce the possibility of confounding.

To this end, we sourced a 10-meter digital elevation model (DEM) from the United States Geological Survey (USGS). Using Esri ArcGIS we then derived slope, terrain shape index (TSI) (McNab 1989), and aspect from this DEM. With then used focal statistics to find the mean of the three in a 100-meter radius, again representing the peridomestic environment. Aspect values were transformed (Beers et al. 1966), then also subject to a 100-meter mean focal statistics operation.

PRISM climatic data at a 4x4 km resolution were obtained from Oregon State University. Using daily temperature grids, we calculated the maximum number of continuous days below zero for the winter preceding each case diagnosis on zero, one, and two-year lags. To ensure we captured all periods of sub-zero temperatures, we considered any cold spell between July 1 and the following June 31 to be part of the latter year's winter. A zero-year lag implies that the date of diagnosis of the case matched the year of the winter period ending in June 31.

Using PRISM monthly mean temperature and mean precipitation data, we also calculated the mean temperature and total precipitation between May 1 and August 31 of each year, including these data again with zero, one, and two-year lags. Finally, we extracted the values of each of these variables, seen in **Table 1** with descriptions, at each point in both space and time.

3.3.5 Variable Selection

The resulting data included 19 variables across 29,276 observations. We expected significant multicollinearity and sought to reduce this by using Boosted Regression Trees (BRT) (Elith et al. 2008) via the Dismo package in R (Ihaka and Gentleman 1996). BRT is particularly useful in variable selection even in the presence of colinear factors (Elith et al. 2008; Elith and Leathwick 2016). In our case we made five BRT models, according to the procedures laid out in Elith et al. (2016). For the sake of model parsimony, we then identified the ten variables that contributed most consistently and significantly to these five models, eliminating the remaining variables. A correlation matrix was then created for the raster layers representing these ten remaining variables.

3.3.6 Final Models

With our variables selected, we employed the same method to create five more parsimonious BRT models. We then projected these models on our study area to create a map of BRT estimated fine scale Lyme suitability. In order to avoid the possibility of overfitting to our data, or being biased by the algorithm in question, we also created five parsimonious models using Maximum Entropy (MaxEnt) modeling (Phillips and Dudík 2008). MaxEnt is the primary competitor to BRT and is similarly well-known in the field. MaxEnt also allows for jackknifing to measure the permutation importance of specific variables. In this case we ran five MaxEnt replicates, then again projected the final model onto our study area to create a map of fine scale Lyme suitability. Random seeded cross validation was used to evaluate the predictive power of both series of BRT and MaxEnt models.

Though spatial data often displays spatial autocorrelation, models with excessively autocorrelated residuals are problematic and violate the “assumption of normally distributed independent errors” (Crase et al. 2012). To ensure that our residuals are not spatially autocorrelated to a problematic degree, we used the final BRT and MaxEnt ensemble models to predict the Lyme status of our presence and pseudo-absence points. We then calculated model residuals for each model at each point and ran a global univariate Moran’s I function for each set of residual points using GeoDa version 1.14 (Anselin et al. 2010). We used an inverse-distance spatial conceptualization and a distance threshold of 7,000 meters, roughly one-tenth the width of the study area. We also used GeoDa to plot spatial correlograms for each set of residuals to examine autocorrelation at differing distance bands.

3.4 Results

3.4.1 Edge-Habitat Simplification

In our edge-habitat reduction efforts, we found that the Douglas and Peucker line simplification algorithm using a 10-meter tolerance produced the most accurate results. As per **Table 2**, this effort was able to reduce the mean-squared error of the original extracted edges by almost 90% as compared to the manually traced edges. The resulting output also exhibited a Pearson correlation of 0.970 between the pre- and post-simplified edge-habitat lengths. Upon visual inspection, we feel the simplified edge-habitats are significantly more realistic as is evident in **Figure 2**.

3.4.2 Model Outputs

Table 3 shows the variable contributions of all constituents of our comprehensive models, visualized in **Figure 3**. Though all 19 initial members contributed to some degree, we selected the top ten to include in later parsimonious models. **Table 4** shows the correlation between these final 10 variables. Forest-Herbaceous edge density at 100m and 200m exhibited a strong correlation exceeding 0.900, as did forest fragment density 200m and 300m. Reducing the dimensionality from 19 to ten reduced the average cross validation AUC by approximately three percent, from 0.822 to 0.796.

Figure 4 and **Table 5** show the variable contributions in the final parsimonious BRT models, which exhibited a mean AUC of 0.796. In all five models, elevation was the most significant contributor, with a mean contribution of 16.6%. This was followed by forest-herbaceous edge density at 100m, mean summer temperature lagged by one year, and forest fragment density at 100m, with mean contributions of 13.4%, 11.1%, and 9.7%, respectively. The cumulative contributions of all forest fragmentation and edge-habitat variables was 46.6%. Response curves for these models are shown in **Figure 5**, and a map of BRT predicted suitability for the NRHD is shown in **Figure 6**.

Variable contributions and permutation importance derived from the MaxEnt models are seen in **Figure 7** and **Table 6**. Mean summer temperature lagged by one year was the most significant contributor with a mean contribution of 26.4%. This was followed by forest-herbaceous edge density at 100m, elevation, and forest fragment density at 300m with mean contributions of 26.1%, 23.7%, and 9.3%, respectively. The cumulative contributions of all forest fragmentation and edge-

habitat variables was 43.3%, and the cumulative permutation importance for the same layers was 53.78%. Response curves are seen in **Figure 8**. The MaxEnt models returned a mean AUC of 0.762 as seen in **Figure 9**, and jackknifed permutation importance is plotted in **Figure 10**. The map of MaxEnt predicted suitability for the NRHD is shown in **Figure 11**.

3.4.3 Residual Autocorrelation

Though autocorrelation is almost inevitable in spatial data, we found our BRT model residuals exhibited a minor Moran's I score of 0.017. MaxEnt model residuals exhibited a Moran's I of 0.023. A trend surface created by inverse-distance weighted interpolation of these point residuals is seen in **Figure 12**. It shows that for both models the residuals are quite random, without any immediately obvious clustering or bias. Spatial correlograms are shown in **Figure 13** showing that the residuals from both models exhibited minimal spatial autocorrelation at all distance bands.

3.5 Discussion

3.5.1 Major Findings

The most consequential finding of this work is that peridomestic forest-herbaceous edge density and peridomestic forest fragment density are strong contributors to our Lyme distribution models. This was the case for both the BRT and MaxEnt models and was common across all replicates. In terms of distance threshold, both algorithms preferred the 100-meter radius for the forest-herbaceous edge density. This suggests that, not only are forest-herbaceous edges consequential drivers of Lyme, but that edges immediately around a home are the most significant when considering this effect. Both algorithms also considered fragment density at both 100- and 300-meter radii significant, with BRT prioritizing the former, and MaxEnt the latter. Note that these layers have a 0.780 correlation coefficient, and as such it is difficult to disentangle the two.

Furthermore, upon inspection, the response curves generated by both modeling algorithms, seen in **Figures 5** and **8**, show a strong positive correlation between forest-herbaceous edge-density and Lyme response. This suggests that edge-habitats drive Lyme disease incidence, likely as a result of these habitats supporting an excess of white-footed mice and blacklegged ticks. The response curves generated by both algorithms also showed a positive association between forest fragment density at 100-meters and Lyme disease, with response curves plateauing at about 5,000 square-meters and dropping off after 17,000 square-meters. For forest fragment density at a 300-meter radius, Lyme response curves showed a weakly negative effect. This suggests that while an extreme excess of forest fragments may be deleterious across broad scales, in the peridomestic environment, Lyme is promoted by the presence of small isolated patches. Again, this may be explained by the habitat preferences of the white-footed mice and blacklegged ticks.

Collectively, these findings support our hypothesis that the fine scale forest structure of the peridomestic environment must be considered when calculating Lyme risk. In fact, the cumulative MaxEnt permutation importance for all edge-habitat and fragment density variables was 53.78%, suggesting that ignoring these variables would reduce the predictive power of these models by more than half.

Among landscape metrics, elevation was most consequential, while slope, aspect, and TSI remained modest contributors. Lyme cases strongly favored higher elevations above 700 meters and had minor associations with lower grade slopes and terrain-shape indices slightly above 0 (concave slopes). Neither rainfall nor duration of sub-zero cold spells were very consequential contributors to the model. We suspect this may be the result of low-resolution climatic data, and the correlation with elevation, which was a major contributor. Finally, mean summer temperatures of the preceding summer were considered significant, with Lyme presence most closely associated with temperatures between 14 and 16 C.

3.5.2 Corroborating Studies

To our knowledge, no studies have made use of high-resolution land cover data to investigate the response of Lyme to the fine scale peridomestic environment. Nevertheless, several have studied the effects of edge-habitats and fragmentation on Lyme and found similar results. Seukep et al. (2015), who studied Lyme at the census tract level, also found a strong positive association between Lyme and forest-herbaceous edges. These findings were confirmed by Moon et al. (2019) who investigated edge-habitat density within an 800-meter radius of the home.

Moon et al. (2019) also found that total forest patch density was significant, though as with our own study was not as consequential as edge-habitat density. The studies differed however in that Moon et al (2019) was tracking total forest patch size, while we limited our research to isolated forest patches of <2-hectares. Seukep et al. (2015) and Jackson et al. (2006) also investigated <2-hectare forest patches. The former found no significant association with Lyme, while the latter found a weakly negative relationship, corroborating the effect that we observed in our study area with the 300-meter fragment density data. We do not feel this contradicts the positive association we noticed at lower densities in the 100-meter fragment density data. These effects could not have been isolated or detected at the census-tract level by Seukep et al. (2015).

The findings of positive Lyme response to increasing elevation corroborate the work of Ferrell and Brinkerhoff (2018) and the response to warmer summer climate and milder winters was expected given the work of Ogden et al. (2016).

3.5.3 An Exploration of Alternatives

Though the strong responses and statistically significant model contributions appear promising, we must consider alternative explanations for these results. Indeed, one is immediately drawn to the fact that the Lyme risk maps seem to highlight rural neighborhoods. Though these results are plausible, and indeed we would expect to find the highest incidence in such areas, this may also be a result of how the presence and pseudo-absence points were created. The former derived from a geocoding service used by the VDH, and likely were dropped near a road. Pseudoabsence points were dropped randomly, in a manner approximating the density of census tracts in the region. As such, the model may be overinflating the importance of living near a road, and underestimating Lyme in remote areas. One could explore this in more detail by dropping pseudoabsence points within a set distance from a road network such as the one provided by OpenStreetMap. That said, the unusual vegetation and leaf cover found near rural homes makes these areas ideal habitats for the white-footed mouse. Whitetail deer are also quite fond of grazing

along these edges. As such, we feel it is plausible that rural and suburban plots are indeed the areas of the highest risk.

3.5.4 Study Limitations

Geocoding accuracy remains a limitation in any such study. Given that the case data we obtained were geocoded and sanitized by the VDH before arriving, we have no way of gaging the accuracy of the VDH's methodology. This can pose a problem as some services geocode by finding a position along a street then assuming the home is immediately adjacent, whereas in rural areas homes may be offset by hundreds of meters. The VDH makes use of the same geocoding procedures for all their work, and as a well-respected institution we are confident that their methods are reliable. Still we felt it prudent to assume an error in the 10s of meters might be possible. For this reason, we made use of 100-meter radii when calculating mean landscape characteristics. We also made the 100-meter radius the smallest value for edge-habitat and fragmentation statistics. Given that the corresponding explanatory variables were the most significant contributors to the niche model, a reader may be tempted to suggest attempting even smaller values for density radii. We simply lack the geocoding resolution to confidently do so.

One must also note that we are assuming Lyme was acquired peridomestically, when in fact these individuals may have acquired the disease while hiking far from home, or even visiting another state. Still, studies have supported the notion that most such cases are acquired near the home (Falco and Fish 1988; Dister et al. 1997), and we suspect this is the case in our study. Furthermore, allochthonous cases should be randomly distributed within the study area, and we do not expect confounding from this to be a problem.

Raster registration point and resolution conflicts are another common issue to such studies. In this case the VGIN land cover data came in a 1x1 meter resolution and was reduced to 10-meter forest fragmentation rasters, then paired with matching 10-meter digital elevation models, and 4-km climatic data. The latter poses the most consequential problem. Though we could have interpolated these data using spatial kriging, going from 4-km to 10-meters introduces significant error and was ultimately unnecessary. Aside from mean summer temperature with one-year lag, none of the climatic variables we included were present in the final model, likely as a result of the inclusion of elevation, which itself is highly correlated with temperature. Given that the remaining data were all averaged across a minimum 100-meter radius from each presence or pseudo-absence point, we feel that any issues from <1-meter misalignments are immaterial.

Another limitation we faced was the lack of diagnosis date for each case. Without these data, it is possible that our winter lags may be off by a year. Consider for example a case identified in late January, while the longest stretch of sub-zero days occurs in February of the same year. In such a case, our model would mistakenly use the same year's values for zero-lag, rather than the preceding year's. We do not consider this a significant limitation as lagged winter colds were not significant contributors to the final models. Moreover, as per VDH case data, over 90.7% of Lyme cases in Virginia over the last 20 years were diagnosed after May 1, with a significant peak in June and July. Fewer than 3.6% of cases were diagnosed in January and February.

Uncertainty in mean time to diagnosis is another concern. Although a discussion on the validity of so called "chronic Lyme" is beyond the scope of this study, it is certainly possible that some cases

were not diagnosed for months after exposure. In fact, diagnostic tests will not correctly identify Lyme antibodies until at least four weeks after initial infection. This may potentially confound our climatic responses by binning cases into the wrong year. Given that the peak of Lyme diagnoses occurs in June and July, which corresponds to peak nymph density, we are confident that the majority of our cases are assigned to the correct year. Moreover, the only variables that change in time are the climatic data, which represented approximately 11.1% of our final BRT model.

A final point of concern is the need to remove so many variables from the final models. With fragmentation statistics calculated at every 100-meters we had a thorough understanding of the composition of the entire area surrounding our cases. By discarding nearly half of our variables, we invariably lose some signal. Moreover, we faced significant multicollinearity between the fragmentation variables of similar radii. It is possible to reduce dimensionality using a principal component analysis (PCA) but doing so results in raster layers with unknown or meaningless units. As our desire was to directly investigate the response of Lyme disease to forest using comprehensible units, we chose to avoid the PCA procedure.

3.5.5 Further research

Considering the importance of the 300-meter fragment density variable in the model, a reader may also be tempted to ask why we did not make use of even larger radii. We feel this is beyond the scope of this study that seeks to explicitly investigate the peridomestic environment. Other studies have corroborated the importance of forest fragmentation at scales as large as the census tract level (Seukep et al. 2015). This finding is also to be expected given that one would not expect to find an abundance of Lyme bearing ticks in a forest patch if the surrounding area was entirely devoid of suitable habitats. Nevertheless, a more thorough investigation of the effects of fine scale fragmentation at larger range may be warranted.

To our surprise, Lyme did not have a strong response to any of the elevation-derived terrain metrics, though all three contributed to model fitness. It is entirely plausible that the ticks and mice are just as tolerant of sloping, north-facing hills as their human neighbors. Given the requirements of modern construction, there is quite likely at least some suitable habitat near any given home in the study area. Further investigation of terrain indices was considered beyond the scope of this work. Nevertheless, it may be worth investigating alternative terrain metrics such as landform curvature (Troeh 1964) and topographic wetness index (Beven et al. 1984), both of which may affect the survival of ticks that depend on humid leaf litter and soil wetness.

Although we did not find significant residual autocorrelation, it may be advantageous to include spatial and temporal autocovariate terms in future studies as it increases predictive power of the models (Crase et al. 2012). One would expect that the case load of an area would be highly dependent on the case load of the prior year as well as that of surrounding areas. These terms were not included in our study because Lyme was not well established in the area until a few years into our study period. Those investigating areas with significant Lyme history or with a longer study period may consider including such terms in their work.

3.5.6 Conclusion

After accounting for climatic and terrain metrics, forest-herbaceous edge-habitat density and forest fragment density remained significant contributors of models fit by two disparate niche

modeling algorithms. We feel that these results support our hypotheses that forest structure at peridomestic scales is a consequential driver of Lyme disease. While broad scale habitat cannot be ignored, and in fact is often sufficient for census tract level studies of Lyme, it is not enough to predict fine scale responses. Moreover, land cover data in more common formats, such as NLCD, do not have the resolution to allow accurate measurement of edge-habitat density. As such we feel that future studies into Lyme at the household level should incorporate edge-habitat density metrics. Creating high resolution land cover data should also be a priority for state governments.

3.6 Acknowledgements

We would like to acknowledge Dr. David Gaines and James Broyhill at the Virginia Department of Health for kindly providing us with geocoded case data. We also acknowledge the VGIN for providing high resolution land cover data.

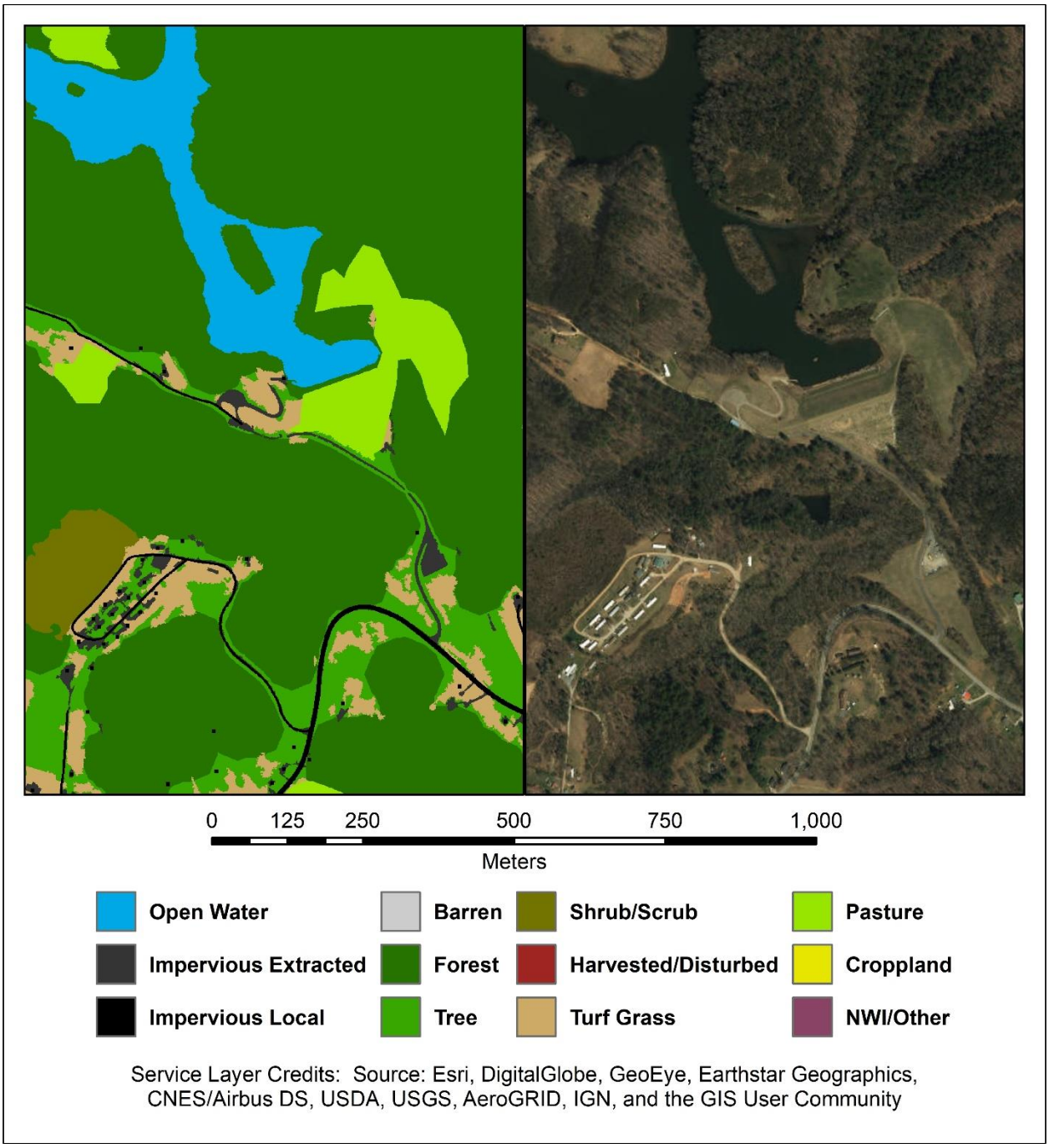

Figure 1: The high resolution VGIN derived land cover data (left) compare favorably to orthometric aerial photography (right). The VGIN product accurately captures the land cover of the study area.

Figure 2: The extraction of edge-habitats from VGIN land cover data produce jagged outputs as seen here (upper). Smoothing these edge-habitats (lower) produce more realistic edge-habitat lengths, as well as more realistic outputs.

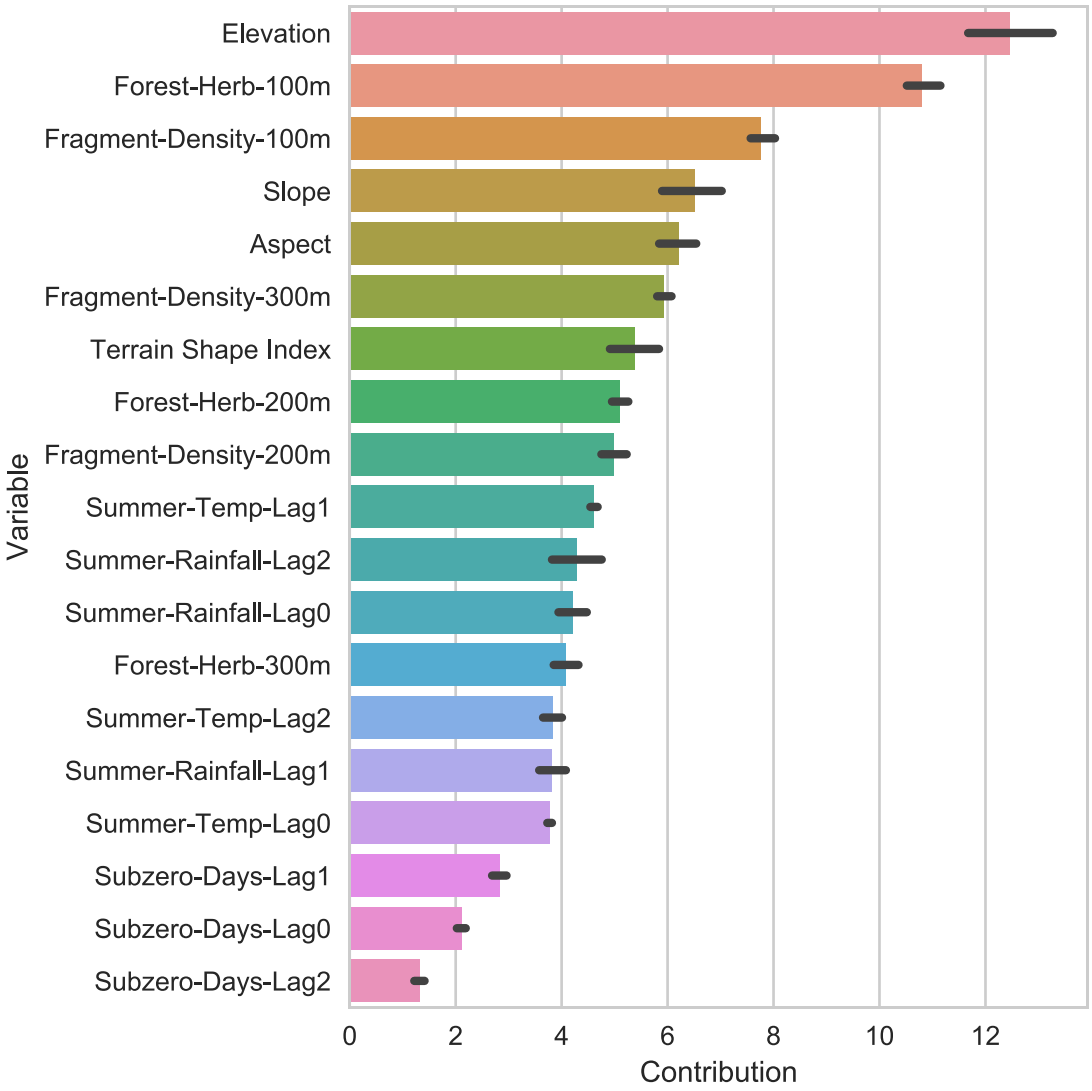

Figure 3: BRT derived mean, minimum, and maximum variable importance for the comprehensive models including all 19 potential regressors.

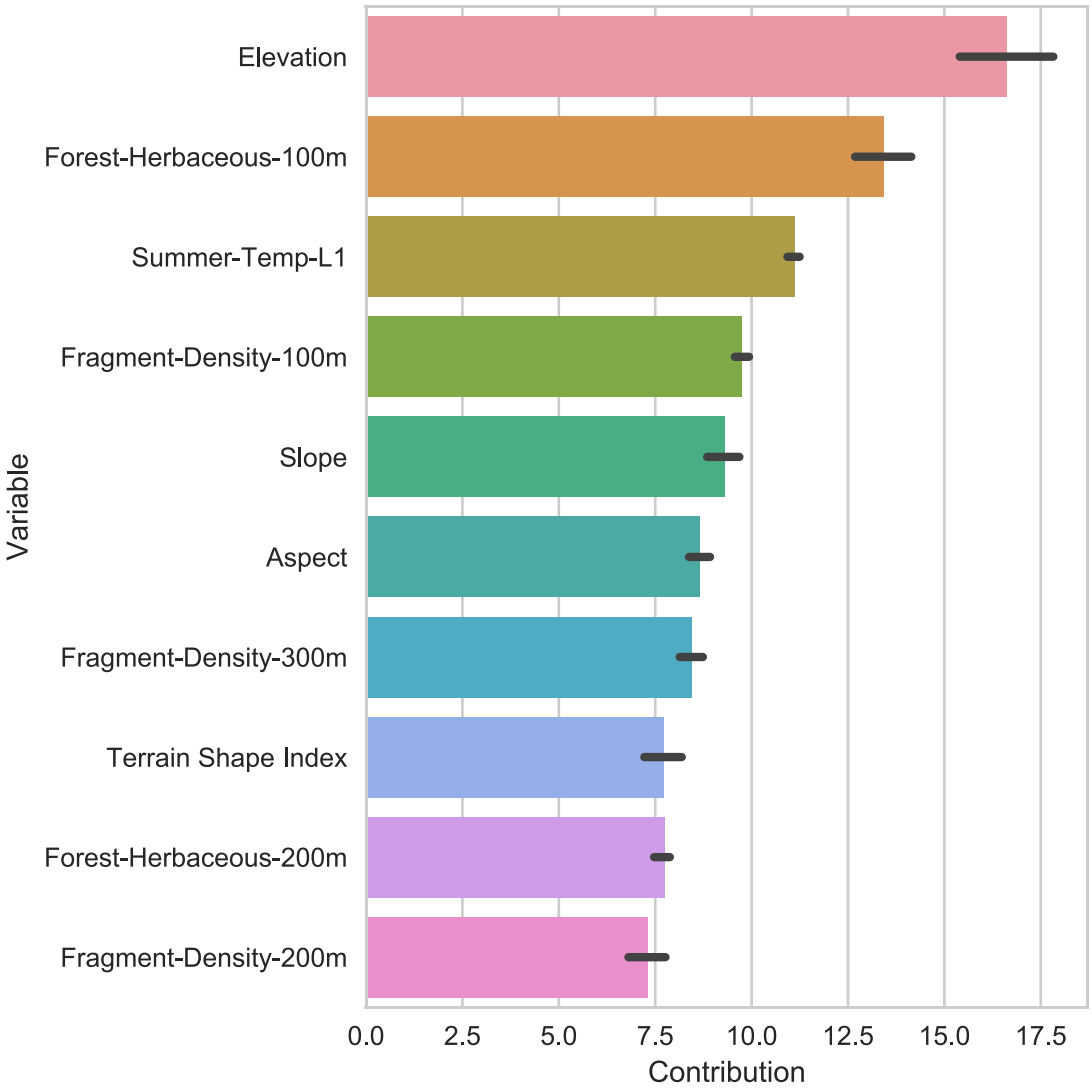

Figure 4: BRT derived mean, minimum, and maximum variable contribution for the parsimonious models including only the ten regressors chosen for our final model fit.

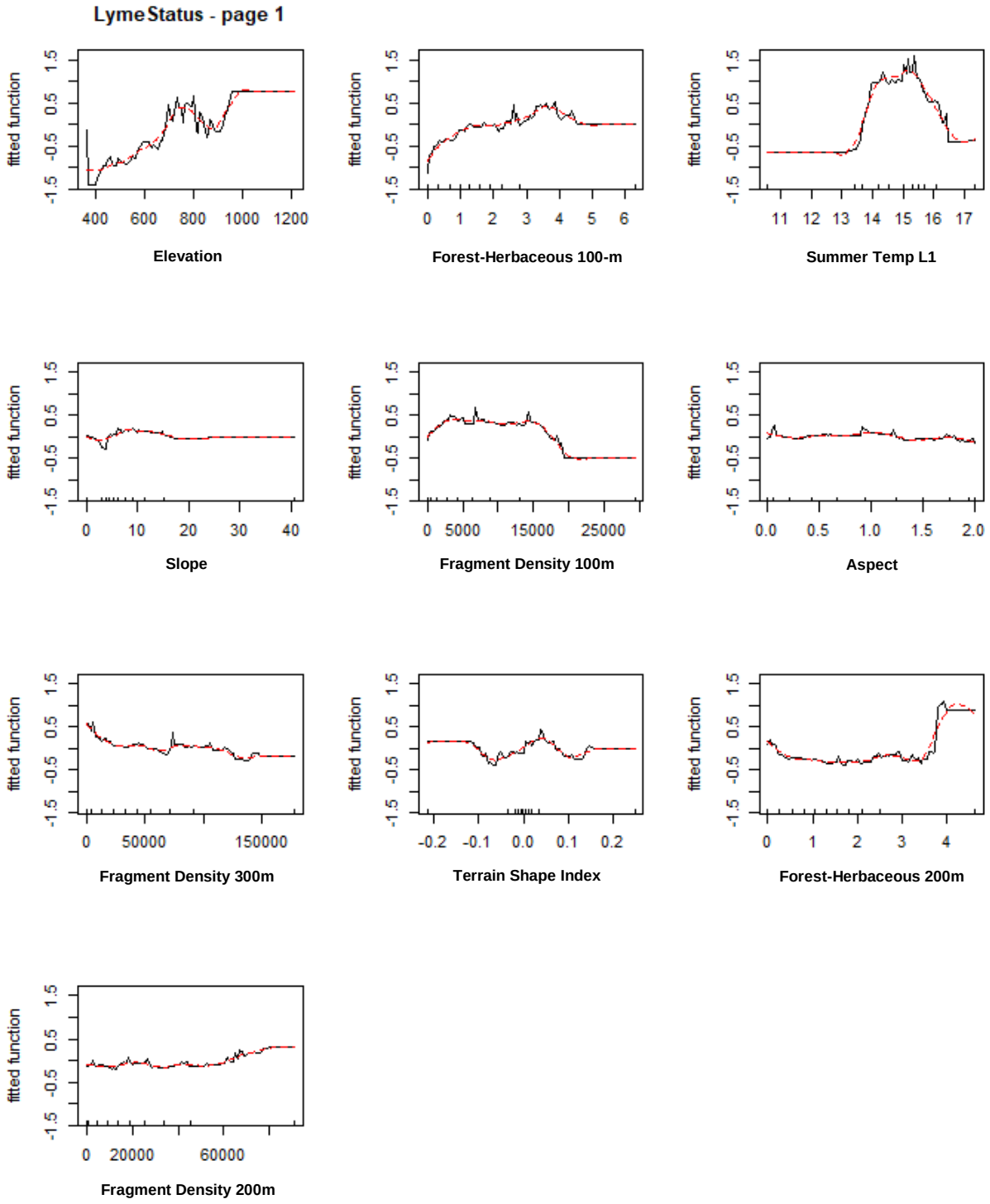

Figure 5: A sample of the BRT derived fitted functions for the parsimonious models from Model 1. The fitted function output of our five models were virtually identical and as such only one is shown here. Though BRT produces decision trees that cannot easily be plotted, these figures approximate the response curves one might obtain from a more traditional regression analysis.

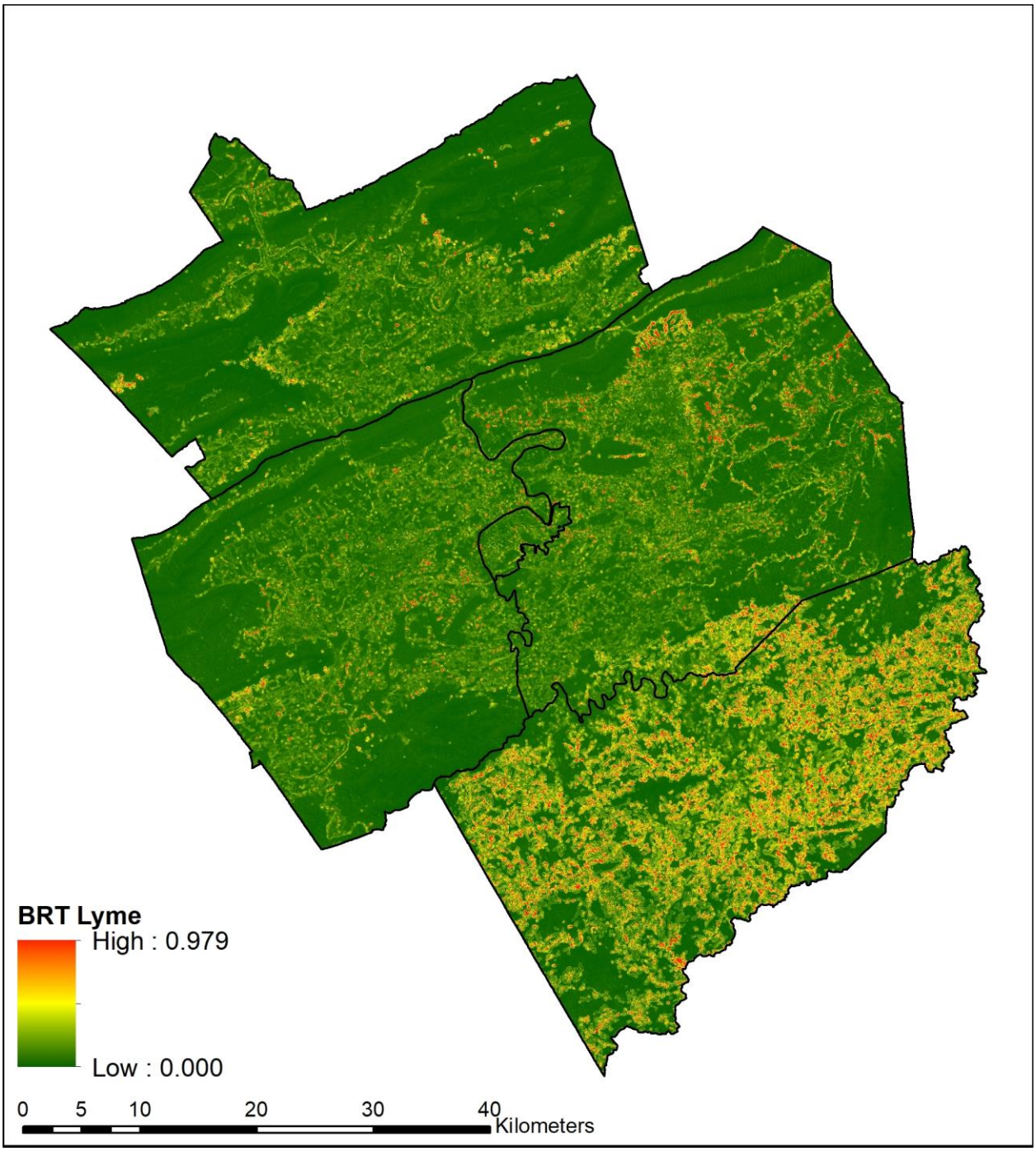

Figure 6: The BRT predicted suitability of Lyme disease in the New River Health District.

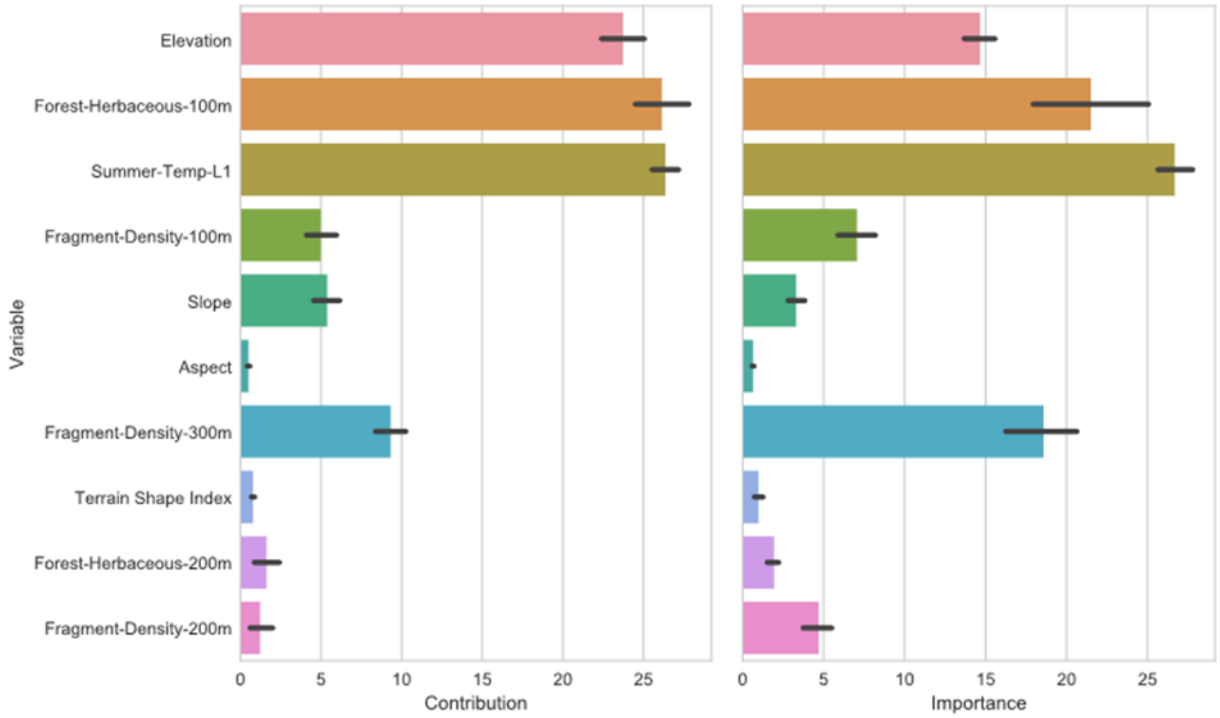

Figure 7: MaxEnt derived mean, minimum, and maximum variable contribution (left) to the parsimonious models fits. Also included are mean, minimum, and maximum permutation importance (right) derived via jackknifing.

Note that the order of these variables matches that of Figure 4, allowing for the direct comparison of model contribution by variable.

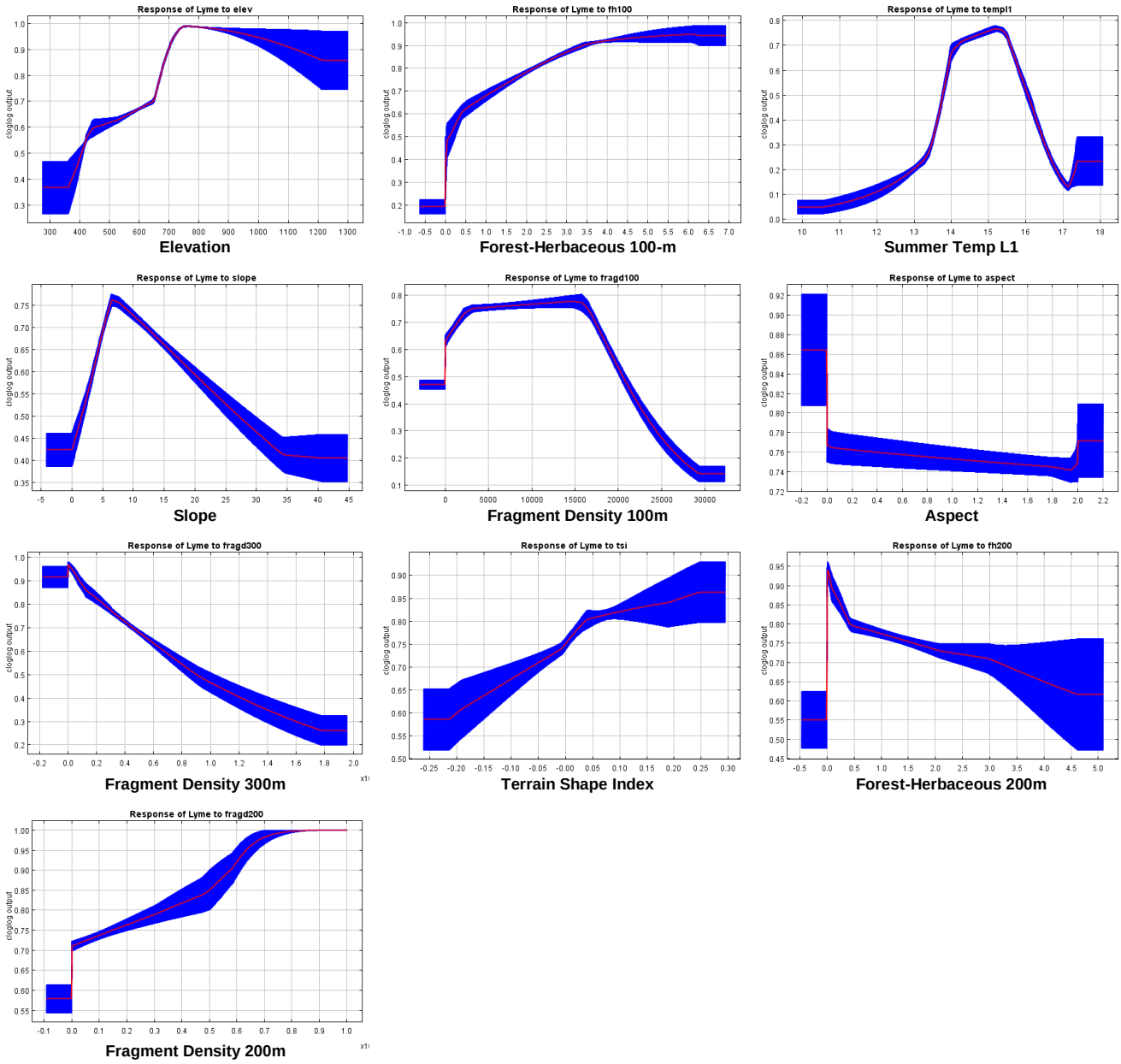

Figure 8: MaxEnt derived marginal variable response curves for our final parsimonious models. These responses are adjusted and already account for the interaction of competing variables. The red curve indicates the mean of the variable response curve across all five models, while the blue shadow sweeps an area the width of one standard deviation.

Note that the order of the layout of this graph matches Figure 5, allowing for the direct comparison of response curves by variable.

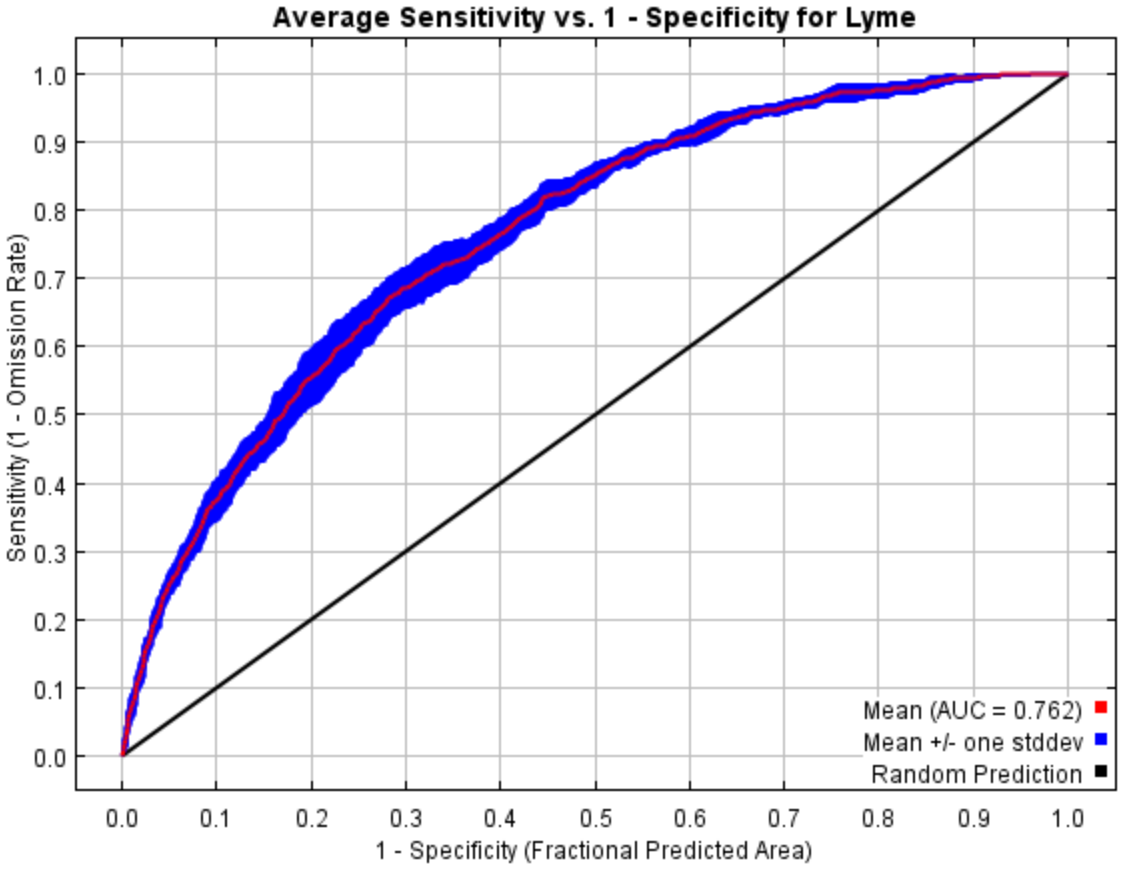

Figure 9: The area under the receiver operating characteristic curve (AUC) for our final MaxEnt models. The red curve indicates the mean of the AUC across all five models, while the blue shadow sweeps an area the width of one standard deviation.

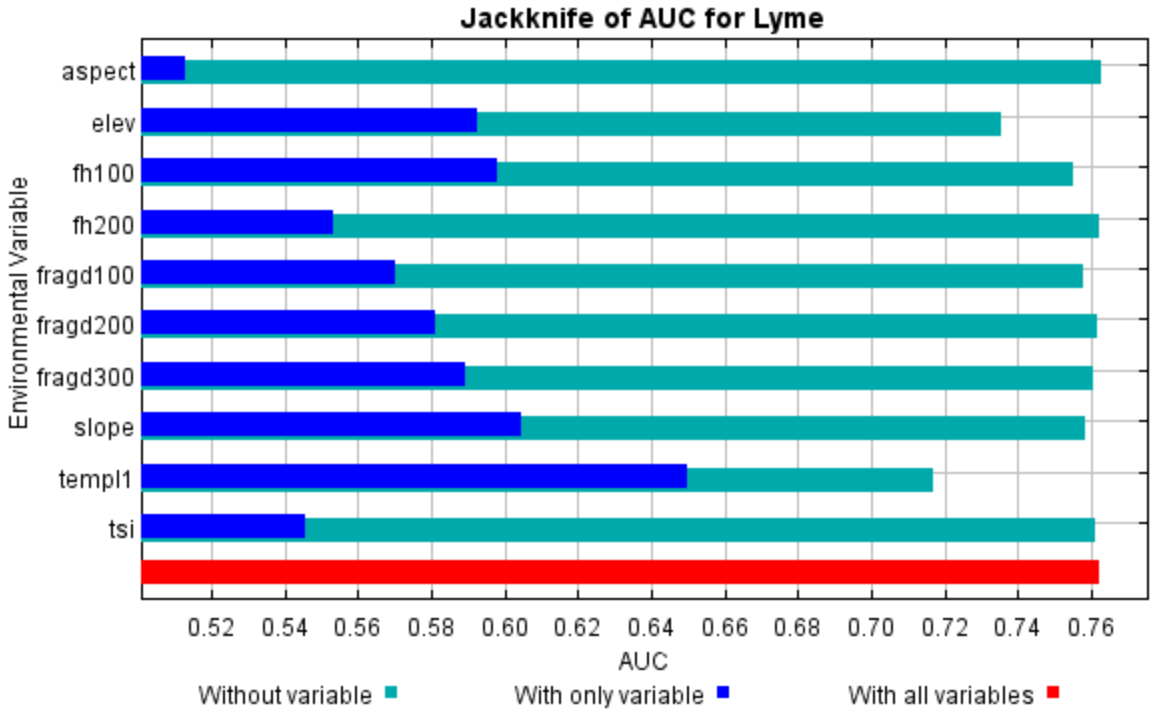

Figure 10: MaxEnt derived variable jackknifing of the AUC for our final parsimonious models. Here we see that the AUC is most heavily affected by the removal of summer temperatures on a one-year lag, forest-herbaceous edges at 100 meters, and elevation.

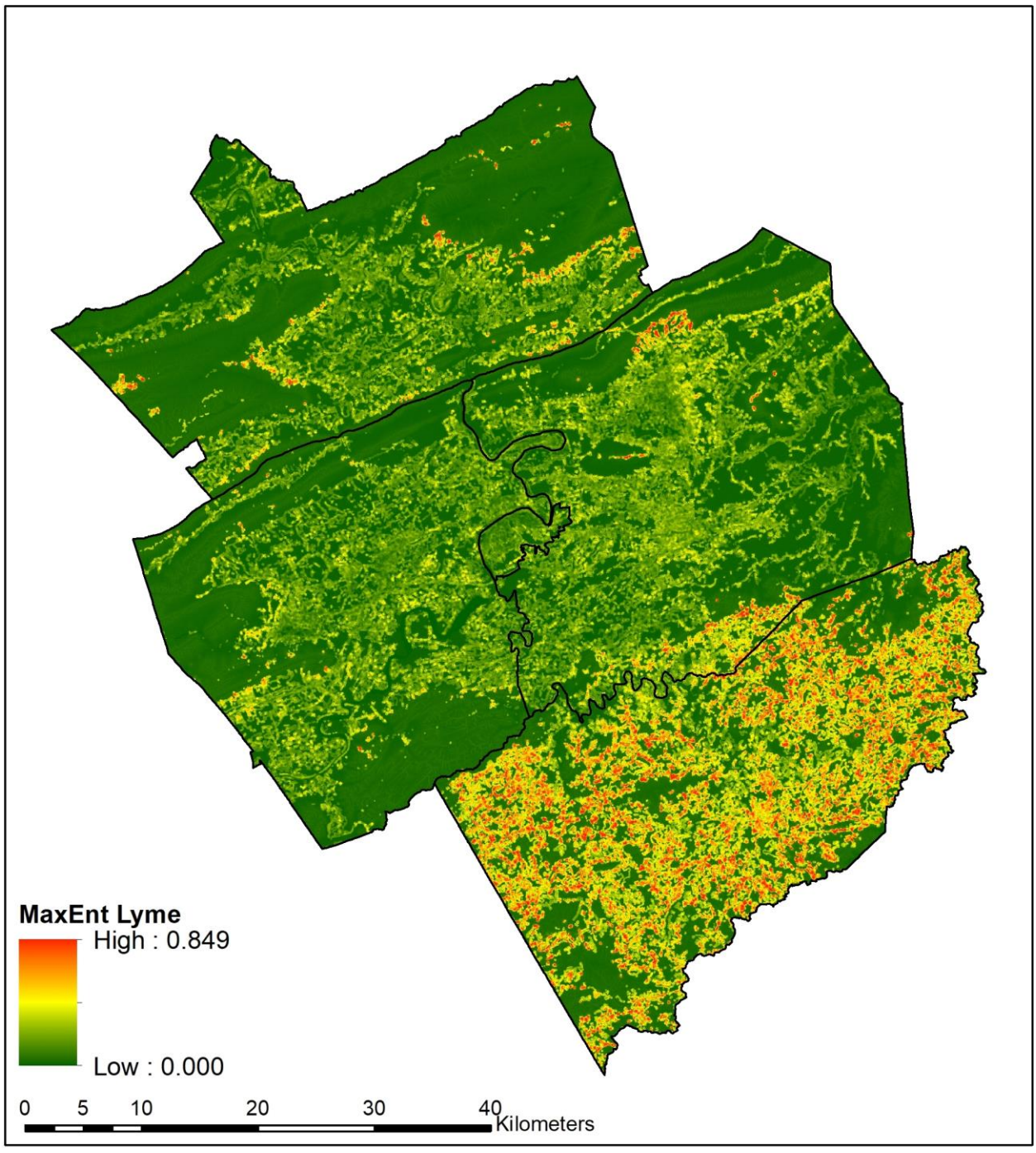

Figure 11: The MaxEnt predicted suitability of Lyme disease in the New River Health District.

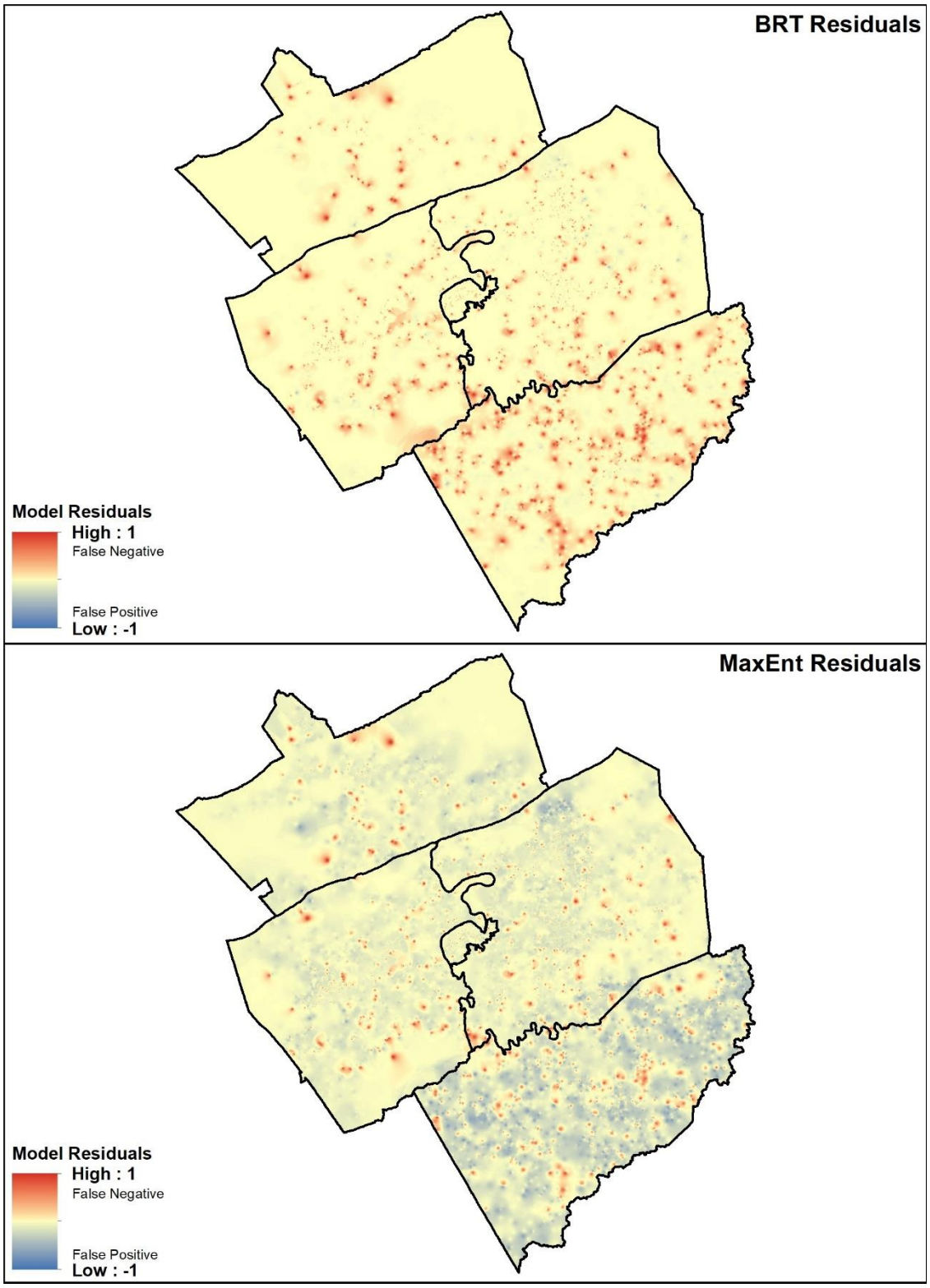

Figure 12: The interpolated surface created from our BRT (upper) and MaxEnt (lower) model residuals does not seem to show any significant spatial gradients or broad scale clustering.

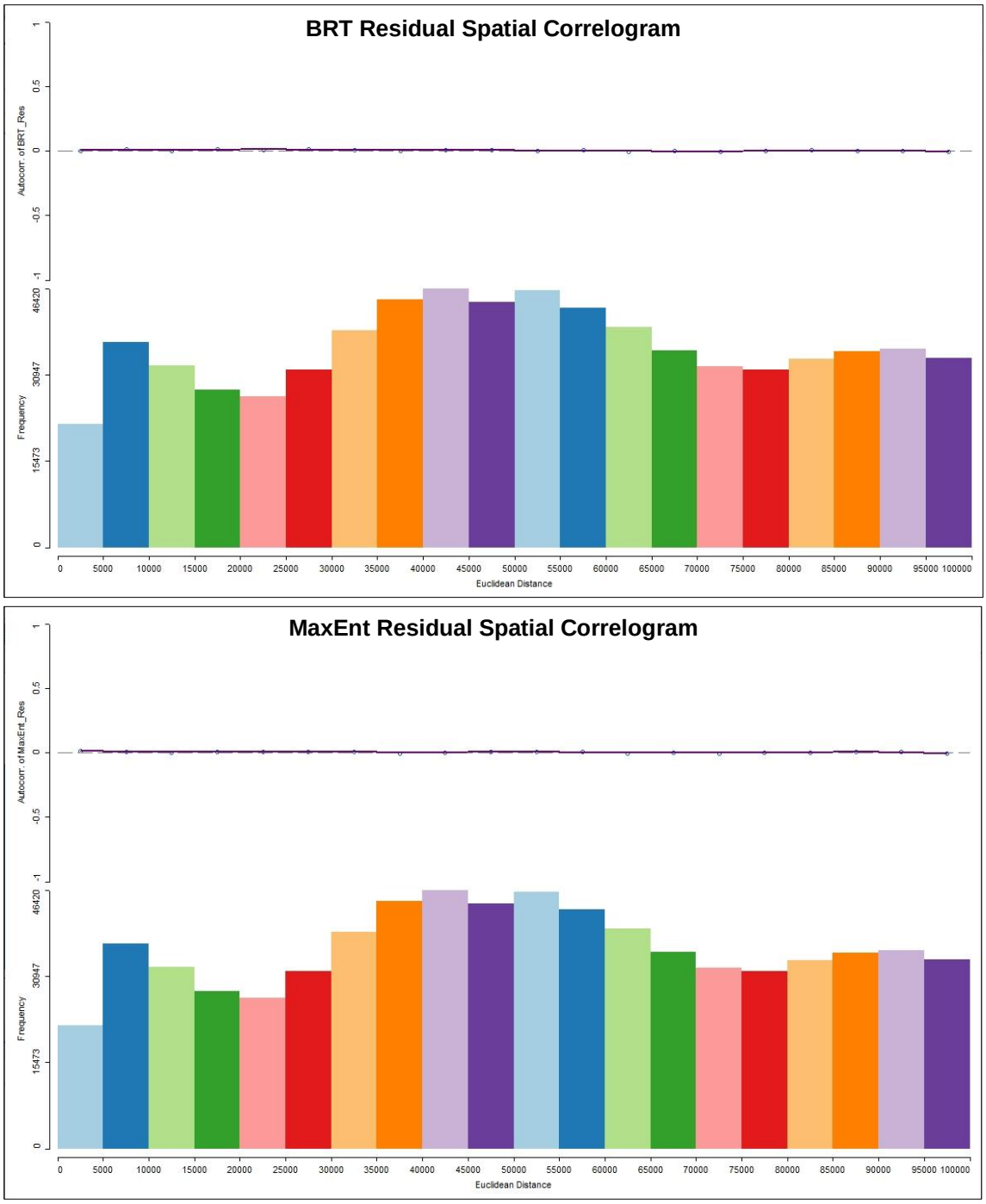

Figure 13: The spatial correlograms for our BRT (upper) and MaxEnt (lower) model residuals show that residual autocorrelation is not a significant problem regardless of distance band.

Table 1: A list of all variables and their units used in initial variable selection phase.

Code	Description	Units
Aspect	Aspect (transformed)	degrees
Elevation	Elevation	meters
Forest-Herbaceous-100m	Density of linear forest-herbaceous edge habitat within a 100-meter radius	1 / meters
Forest-Herbaceous-200m	Density of linear forest-herbaceous edge habitat within a 200-meter radius	1 / meters
Forest-Herbaceous-300m	Density of linear forest-herbaceous edge habitat within a 300-meter radius	1 / meters
Fragment-Density-100m	Surface area of >2-ha forest patches within a 100-meter radius	NaN
Fragment-Density-200m	Surface area of >2-ha forest patches within a 200-meter radius	NaN
Fragment-Density-300m	Surface area of >2-ha forest patches within a 300-meter radius	NaN
Summer-Rainfall-Lag0	Total rainfall between May 1 and August 31 with lag = 0 years	mm
Summer-Rainfall-Lag1	Total rainfall between May 1 and August 31 with lag = 1 years	mm
Summer-Rainfall-Lag2	Total rainfall between May 1 and August 31 with lag = 2 years	mm
Slope	Slope	degrees
Summer-Temp-Lag0	Average daily temperature between May 1 and August 31 with lag = 0 years	C
Summer-Temp-Lag1	Average daily temperature between May 1 and August 31 with lag = 1 years	C
Summer-Temp-Lag2	Average daily temperature between May 1 and August 31 with lag = 2 years	C
Terrain Shape Index	Terrain Shape Index	NaN
Subzero-Days-Lag0	Maximum number of successive days below freezing with lag = 0 years	days
Subzero-Days-Lag1	Maximum number of successive days below freezing with lag = 1 years	days
Subzero-Days-Lag2	Maximum number of successive days below freezing with lag = 2 years	days

Table 2: We tested a variety of algorithms and distance thresholds to simplify the jagged edge-habitats extracted from VGIN data. We then compared both the original extracted edges, and our simplified edges, to edges which were manually traced on Google Earth Pro. We then selected the method which had the greatest reduction in mean squared error when compared to the traced edges. In this case, the Douglas-Peucker algorithm with a 10-meter threshold (bottom left) was most effective, reducing MSE by 89.77% while maintaining a 0.9701 correlation with the original extracted edge data.

Algorithm and Distance Threshold	Reduction in MSE	Correlation w/ Original	Algorithm and Distance Threshold	Reduction in MSE	Correlation w/ Original
Douglas-Peucker 1.4 m	0.6294	0.9994	Wang-Muller 1.4 m	0.6988	0.9959
Douglas-Peucker 2.0 m	0.7172	0.9977	Wang-Muller 2.0 m	0.7667	0.9929
Douglas-Peucker 2.8 m	0.7948	0.9934	Wang-Muller 2.8 m	0.8375	0.9860
Douglas-Peucker 3.0 m	0.8043	0.9924	Wang-Muller 3.0 m	0.8395	0.9858
Douglas-Peucker 4.0 m	0.8500	0.9862	Wang-Muller 4.0 m	0.8650	0.9807
Douglas-Peucker 4.3 m	0.8569	0.9850	Wang-Muller 4.3 m	0.8693	0.9788
Douglas-Peucker 5.0 m	0.8704	0.9822	Wang-Muller 5.0 m	0.8810	0.9767
Douglas-Peucker 5.7 m	0.8767	0.9791	Wang-Muller 5.7 m	0.8845	0.9753
Douglas-Peucker 6.0 m	0.8802	0.9782	Wang-Muller 6.0 m	0.8853	0.9747
Douglas-Peucker 7.0 m	0.8870	0.9757	Wang-Muller 7.0 m	0.8881	0.9722
Douglas-Peucker 7.1 m	0.8869	0.9752	Wang-Muller 7.1 m	0.8887	0.9721
Douglas-Peucker 8.0 m	0.8922	0.9725	Wang-Muller 8.0 m	0.8909	0.9712
Douglas-Peucker 8.5 m	0.8941	0.9718	Wang-Muller 8.5 m	0.8931	0.9698
Douglas-Peucker 9.0 m	0.8956	0.9715	Wang-Muller 9.0 m	0.8938	0.9695
Douglas-Peucker 10 m	0.8977	0.9701	Wang-Muller 10 m	0.8962	0.9674

Table 3: BRT model variable contributions for the comprehensive models. The top ten contributors, here shown in maroon, were selected for use in our parsimonious models. The remaining variables, below the dashed line, were discarded.

Variable	Mean of Contribution	Contribution Lower 95% CI	Contribution Upper 95% CI
Elevation	12.454	11.151	13.758
Forest-Herbaceous-100m	10.805	10.306	11.305
Fragment-Density-100m	7.764	7.382	8.147
Slope	6.504	5.612	7.397
Aspect	6.204	5.624	6.785
Fragment-Density-300m	5.924	5.699	6.150
Terrain Shape Index	5.375	4.643	6.107
Forest-Herbaceous-200m	5.102	4.860	5.343
Fragment-Density-200m	4.987	4.604	5.370
Summer-Temp-Lag1	4.611	4.508	4.713
Summer-Rainfall-Lag2	4.287	3.529	5.046
Summer-Rainfall-Lag0	4.208	3.769	4.647
Forest-Herbaceous-300m	4.084	3.718	4.450
Summer-Temp-Lag2	3.829	3.528	4.130
Summer-Rainfall-Lag1	3.823	3.405	4.241
Summer-Temp-Lag0	3.772	3.706	3.838
Subzero-Days-Lag1	2.827	2.601	3.052
Subzero-Days-Lag0	2.115	1.983	2.246
Subzero-Days-Lag2	1.324	1.170	1.478

Table 4: Correlation coefficients between the ten raster layers chosen during our variable selection process. Highly correlated values (exceeding 0.8) are highlighted in maroon.

Variable	Aspect	Elevation	Forest-Herbaceous-100m	Forest-Herbaceous-200m	Fragment-Density-100m	Fragment-Density-200m	Fragment-Density-300m	Slope	Summer-Temp-L1	Terrain Shape Index
Aspect	1.000	-0.036	0.002	0.002	0.006	0.001	0.000	0.067	0.057	0.002
Elevation	-0.036	1.000	-0.244	-0.214	-0.166	-0.210	-0.241	0.142	-0.751	-0.076
Forest-Herbaceous-100m	0.002	-0.244	1.000	0.941	0.449	0.463	0.464	-0.399	0.171	0.016
Forest-Herbaceous-200m	0.002	-0.214	0.941	1.000	0.410	0.409	0.409	-0.356	0.149	0.011
Fragment-Density-100m	0.006	-0.166	0.449	0.410	1.000	0.891	0.780	-0.254	0.128	0.016
Fragment-Density-200m	0.001	-0.210	0.463	0.409	0.891	1.000	0.945	-0.322	0.163	0.010
Fragment-Density-300m	0.000	-0.241	0.464	0.409	0.780	0.945	1.000	-0.365	0.189	0.009
Slope	0.067	0.142	-0.399	-0.356	-0.254	-0.322	-0.365	1.000	-0.023	-0.001
Summer-Temp-L1	0.057	-0.751	0.171	0.149	0.128	0.163	0.189	-0.023	1.000	0.002
Terrain Shape Index	0.002	-0.076	0.016	0.011	0.016	0.010	0.009	-0.001	0.002	1.000

Table 5: BRT derived mean variable contributions for our final parsimonious models.

Variable	Mean of Contribution	Contribution Lower 95% CI	Contribution Upper 95% CI
Elevation	16.615	14.585	18.645
Forest-Herbaceous-100m	13.413	12.213	14.612
Summer-Temp-L1	11.119	10.864	11.374
Fragment-Density-100m	9.741	9.439	10.043
Slope	9.294	8.616	9.973
Aspect	8.642	8.215	9.069
Fragment-Density-300m	8.449	7.969	8.93
Forest-Herbaceous-200m	7.731	7.374	8.088
Terrain Shape Index	7.703	6.902	8.503
Fragment-Density-200m	7.292	6.507	8.077

Table 6: MaxEnt derived mean variable importance (upper) and jackknife derived permutation importance (lower) for our final parsimonious models.

Variable	Contribution	Contribution Lower 95% CI	Contribution Upper 95% CI
Summer-Temp-L1	26.360	25.035	27.685
Forest-Herbaceous-100m	26.140	23.442	28.838
Elevation	23.740	21.614	25.866
Fragment-Density-300m	9.320	7.845	10.795
Slope	5.380	4.143	6.617
Fragment-Density-100m	4.980	3.421	6.539
Forest-Herbaceous-200m	1.620	0.331	2.909
Fragment-Density-200m	1.220	0.091	2.349
Terrain Shape Index	0.780	0.618	0.942
Aspect	0.500	0.376	0.624

Variable	Importance	Contribution Lower 95% CI	Contribution Upper 95% CI
Summer-Temp-L1	26.680	24.912	28.448
Forest-Herbaceous-100m	21.500	15.807	27.193
Fragment-Density-300m	18.580	15.051	22.109
Elevation	14.660	13.021	16.299
Fragment-Density-100m	7.060	5.109	9.011
Fragment-Density-200m	4.700	3.241	6.159
Slope	3.300	2.462	4.138
Forest-Herbaceous-200m	1.940	1.341	2.539
Terrain Shape Index	0.980	0.538	1.422
Aspect	0.640	0.529	0.751

Bibliography

Abbott A (2006) Lyme disease: uphill struggle. Nature 439:524–525

Allan BF, Keesing F, Ostfeld RS (2003) Effect of forest fragmentation on lyme disease risk. Conserv Biol 17:267–272

Anderson JF, Johnson RC, Magnarelli LA, Hyde FW (1985) Identification of endemic foci of Lyme disease: isolation of Borrelia burgdorferi from feral rodents and ticks (Dermacentor variabilis). J Clin Microbiol 22:36–38

Anderson JF, Johnson RC, Magnarelli LA, Hyde FW (1986) Involvement of birds in the epidemiology of the Lyme disease agent Borrelia burgdorferi. Infect Immun 51:394–396

Anderson JF, Magnarelli LA, Burgdorfer W, Barbour AG (1983) Spirochetes in Ixodes dammini and mammals from Connecticut. Am J Trop Med Hyg 32:818–824

Anderson JF, Magnarelli LA, LeFebvre RB, Andreadis TG, McAninch JB, Perng G-C, Johnson RC (1989) Antigenically variable Borrelia burgdorferi isolated from cottontail rabbits and Ixodes dentatus in rural and urban areas. J Clin Microbiol 27:13–20

Andren H (1994) Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. Oikos 355–366

Anselin L, Syabri I, Kho Y (2010) GeoDa: an introduction to spatial data analysis. In: Handbook of applied spatial analysis. Springer, pp 73–89

Bacon RM, Kugeler KJ, Mead PS, others (2008) Surveillance for Lyme disease--United States, 1992-2006. Department of Health & Human Services, Centers for Disease Control and Prevention

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3:327–338

Barton TR, Harris MP, Wanless S (1995) Natural attachment duration of nymphs of the tick Ixodes uriae (Acari: Ixodidae) on kittiwake Rissa tridactyla nestlings. Exp Appl Acarol 19:499–509 .

Battaly GR, Fish D (1993) Relative importance of bird species as hosts for immature Ixodes dammini (Acari: Ixodidae) in a suburban residential landscape of southern New York State. J Med Entomol 30:740–747

Beers TW, Dress PE, Wensel LC (1966) Notes and observations: aspect transformation in site productivity research. J For 64:691–692

Berger KA, Ginsberg HS, Dugas KD, Hamel LH, Mather TN (2014) Adverse moisture events predict seasonal abundance of Lyme disease vector ticks (Ixodes scapularis). Parasites and Vectors 7:1–8 .

Bertrand MR, Wilson ML (1996) Microclimate-dependent survival of undef adult Isodes scapularis (Acari: Ixodidae) in nature: life cycle and study design implications. J Med Entomol 33:619–627

Beven KJ, Kirkby MJ, Schofield N, Tagg AF (1984) Testing a physically-based flood forecasting model (TOPMODEL) for three UK catchments. J Hydrol 69:119–143

Bosler EM, Ormiston BG, Coleman JL, Hanrahan JP, Benach JL (1984) Prevalence of the Lyme disease spirochete in populations of white-tailed deer and white-footed mice. Yale J Biol Med 57:651–659

Brinkerhoff RJ, Gilliam WF, Gaines DN (2014) Lyme Disease, Virginia, USA, 2000–2011. Emerg Infect Dis 20:1661

Brownstein JS, Holford TR, Fish D (2003) A climate-based model predicts the spatial distribution of the Lyme disease vector Ixodes scapularis in the United States. Environ Health Perspect 111:1152–1157 .

Brownstein JS, Holford TR, Fish D (2005a) Effect of climate change on lyme disease risk in North America. Ecohealth 2:38–46

Brownstein JS, Skelly DK, Holford TR, Fish D (2005b) Forest fragmentation predicts local scale heterogeneity of Lyme disease risk. Oecologia 146:469–475

Burgdorfer W, Barbour AG, Hayes SF, Benach JL, Davis JP (1982) Lyme Disease - A Tick-Borne Spirochetosis? Science (80-) 216:1317–1319

CDC (2019) Annual data for 2018. In: Natl. Notif. Infect. Dis. Cond. https://wonder.cdc.gov/nndss/nndss_annual_tables_menu.asp. Accessed 2 Jan 2020

CDC (2015a) Reported cases of Lyme disease by state or locality, 2005-2014. http://www.cdc.gov/lyme/stats/chartstables/reportedcases_statelocality.html. Accessed 27 Oct 2015

CDC (2015b) Interactive Lyme Disease Map. http://www.cdc.gov/lyme/stats/maps/interactiveMaps.html. Accessed 1 Nov 2015

CDC (1997) Lyme disease--United States, 1996. MMWR Morb Mortal Wkly Rep 46:531

Coyle BS, Strickland GT, Liang YY, Peña C, McCarter R, Israel E (1996) The public health impact of Lyme disease in Maryland. J Infect Dis 173:1260–2

Crase B, Liedloff AC, Wintle BA (2012) A new method for dealing with residual spatial autocorrelation in species distribution models. Ecography (Cop) 35:879–888

Cromley EK, Cartter ML, Mrozinski RD, Ertel S-H (1998a) Residential Setting as a Risk Factor for Lyme Disease in a Hyperendemic Region. Am J Epidemiol 147:472–477

Cromley EK, Cartter ML, Mrozinski RD, Ertel S-H (1998b) Residential Setting as a Risk Factor for Lyme Disease in a Hyperendemic Region. Am J Epidemiol 147:472–477

Das A, Lele SR, Glass GE, Shields T, Patz J (2002) Modelling a discrete spatial response using generalized linear mixed models: Application to Lyme disease vectors. Int J Geogr Inf Sci 16:151–166 .

de Silva AM, Telford Sr, Brunet LR, Barthold SW, Fikrig E (1996) Borrelia burgdorferi OspA is an arthropod-specific transmission-blocking Lyme disease vaccine. J Exp Med 183:271–275

Dister SW, Fish D, Bros SM, Frank DH, Wood BL (1997) Landscape characterization of peridomestic risk for Lyme disease using satellite imagery. Am J Trop Med Hyg 57:687–692

Donahue JG, Piesman J, Spielman A (1987) Reservoir competence of white-footed mice for Lyme disease spirochetes. Am J Trop Med Hyg 36:92–96

Donta ST (1997) Tetracycline therapy for chronic Lyme disease. Clin Infect Dis 25 Suppl 1:S52–S56

Elith J, Leathwick J (2016) Boosted regression trees for ecological modeling

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802–813 .

Estrada-Peña A (2002) Increasing habitat suitability in the United States for the tick that transmits Lyme disease: a remote sensing approach. Environ Health Perspect 110:635

Falco RC, Fish D (1988) Prevalence of Ixodes dammini near the homes of Lyme disease patients in Westchester County, New York. Am J Epidemiol 127:826–830

Fallon BA, Nields JA (1994) Lyme disease: a neuropsychiatric illness. Am J Psychiatry 151:1571–1583

Feder HM, Johnson BJB, O'Connell S, Shapiro ED, Steere AC, Wormser GP (2007) A Critical Appraisal of "Chronic Lyme Disease." N Engl J Med 357:1422–1430

Feldhamer GA, Kilbane TP, Sharp DW (1989) Cumulative Effect of Winter on Acorn Yield and Deer Body Weight. J Wildl Manage 53:292–295

Feria-Arroyo TP, Castro-Arellano I, Gordillo-Perez G, Cavazos AL, Vargas-Sandoval M, Grover A, Torres J, Medina RF, De León AAP, Esteve-Gassent MD (2014) Implications of climate change on the distribution of the tick vector Ixodes scapularis and risk for Lyme disease in the Texas-Mexico transboundary region. Parasites and Vectors 7:1–16.

Ferrell AM, Brinkerhoff RJ (2018) Using landscape analysis to test hypotheses about drivers of tick abundance and infection prevalence with Borrelia burgdorferi. Int J Environ Res Public Health 15:18–20 .

Frank D, Fish D, Moy F (1998) Landscape features associated with Lyme disease risk in a suburban residential environment. Landsc Ecol 13:27–36

Glass GE, Schwartz BS, Morgan JM, Johnson DT, Noy PM, Israel E (1995a) Environmental risk factors for Lyme disease identified with geographic information systems. Am J Public Health 85:944–8

Glass GE, Schwartz BS, Morgan JM, Johnson DT, Noy PM, Israel E (1995b) Environmental risk factors for Lyme disease identified with geographic information systems. Am J Public Health 85:944–8

Guidoboni M, Ferreri AJM, Ponzoni M, Doglioni C, Dolcetti R (2006) Infectious Agents in Mucosa-Associated Lymphoid Tissue--Type Lymphomas: Pathogenic Role and Therapeutic Perspectives. Clin Lymphoma Myeloma 6:289–300

Habálek Z, Halouzka J (1997) Distribution of Borrelia burgdorferi sensu lato genomic groups in Europe, a review. Eur J Epidemiol 13:951–957

Hansmann Y (2009) Treatment and prevention of Lyme disease. In: Lyme Borreliosis. Karger Publishers, pp 111–129

Horobik V, Keesing F, Ostfeld RS (2006) Abundance and Borrelia burgdorferi-infection prevalence of nymphal Ixodes scapularis ticks along forest-field edges. Ecohealth 3:262–268

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph Stat 5:299–314

Jackson L, Levine J, Hilborn E (2006a) A comparison of analysis units for associating Lyme disease with forest-edge habitat. Community Ecol 7:189–197

Jackson LE, Hilborn ED, Thomas JC (2006b) Towards landscape design guidelines for reducing Lyme disease risk. Int J Epidemiol 35:315–322

Johnson L, Aylward A, Stricker RB (2011) Healthcare access and burden of care for patients with Lyme disease: a large United States survey. Health Policy (New York) 102:64–71

Khatchikian C, Prusinski M (2012) Geographical and environmental factors driving the increase in the Lyme disease vector Ixodes scapularis. Ecosphere 3:1–18

Kitron U, Kazmierczak JJ (1997) Spatial Analysis of the Distribution of Lyme Disease in Wisconsin. Am J Epidemiol 145:558–566

Koenig WD, Knops JMH, Carmen WJ, Stanback MT, Mumme RL (1996) Acorn production by oaks in central coastal California: influence of weather at three levels. Can J For Res 26:1677–1683

Lane RS, Piesman J, Burgdofer W (1991) LYME BORRELIOSIS: Relation of Its Causative Agent to Its Vectors and Hosts in North America and Europe. Annu Rev Entomol 36:587–509

Levi T, Kilpatrick AM, Mangel M, Wilmers CC (2012) Deer, predators, and the emergence of Lyme disease. Proc Natl Acad Sci 109:10942–10947

Li J, Kolivras KN, Hong Y, Duan Y, Seukep SE, Prisley SP, Campbell JB, Gaines DN (2014) Spatial and Temporal Emergence Pattern of Lyme Disease in Virginia. Am J Trop Med Hyg 91:1166–1172

Lindgren E, Jaenson TGT, others (2006) Lyme borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and adaptation measures

Mak S, Morshed M, Henry B (2010) Ecological niche modeling of Lyme disease in British Columbia, Canada. J Med Entomol 47:99–105

Martínez-Vega RA, Danis-Lozano R, Díaz-Quijano FA, Velasco-Hernández J, Santos-Luna R, Román-Pérez S, Kuri-Morales P, Ramos-Castañeda J (2015) Peridomestic Infection as a Determining Factor of Dengue Transmission. PLoS Negl Trop Dis 9:1–18

Mast WE, Burrows WM (1976) Erythema chronicum migrans in the United States. JAMA 236:859–60

Maupin GO, Fish D, Zultowsky J, Campos EG, Piesman J (1991) Landscape ecology of Lyme disease in a residential area of Westchester County, New York. Am J Epidemiol 133:1105–1113

McCabe GJ, Bunnell JE (2004) Precipitation and the Occurrence of Lyme Disease in the Northeastern United States. Vector-Borne Zoonotic Dis 4:143–148

McNab WH (1989) Terrain shape index: quantifying effect of minor landforms on tree height. For Sci 35:91–104

Meade MS, Emch M (2010) Medical Geography. Guilford Press

Monaghan AJ, Moore SM, Sampson KM, Beard CB, Eisen RJ (2015) Climate change influences on the annual onset of Lyme disease in the United States. Ticks Tick Borne Dis 6:615–622 .

Moon KA, Pollak J, Poulsen MN, Hirsch AG, DeWalle J, Heaney CD, Aucott JN, Schwartz BS (2019) Peridomestic and community-wide landscape risk factors for Lyme disease across a range of community contexts in Pennsylvania. Environ Res 178:108649

Moore I, Wright DH (1984) Primary gastric lymphoma—a tumour of mucosa-associated lymphoid tissue. A histological and immunohistochemical study of 36 cases. Histopathology 8:1025–1039

Nigrovic LE, Thompson KM (2007) The Lyme vaccine: a cautionary tale. Epidemiol Infect 135:1–8

Nupp TE, Swihart RK (1996) Effect of forest patch area on population attributes of white-footed mice (Peromyscus leucopus) in fragmented landscapes. Can J Zool 74:467–472

Ogden NH, Bigras-Poulin M, O'Callaghan CJ, Barker IK, Lindsay LR, Maarouf A, Smoyer-Tomic KE, Waltner-Toews D, Charron D (2005) A dynamic population model to investigate effects of climate on geographic range and seasonality of the tick Ixodes scapularis. Int J Parasitol 35:375–389

Ogden NH, Lindsay LR (2016) Effects of Climate and Climate Change on Vectors and Vector-Borne Diseases: Ticks Are Different. Trends Parasitol 32:646–656

Ogden NH, Lindsay LR, Beauchamp G, Charron D, Maarouf A, O'Callaghan CJ, Waltner-Toews D, Barker IK (2004) Investigation of relationships between temperature and developmental rates of tick Ixodes scapularis (Acari: Ixodidae) in the laboratory and field. J Med Entomol 41:622–633

Ogden NH, Maarouf A, Barker IK, Bigras-Poulin M, Lindsay LR, Morshed MG, O'callaghan CJ, Ramay F, Waltner-Toews D, Charron DF (2006) Climate change and the potential for range expansion of the Lyme disease vector Ixodes scapularis in Canada. Int J Parasitol 36:63–70

Ogden NH, Radojevic M, Wu X, Duvvuri VR, Leighton PA, Wu J (2014) Estimated Effects of Projected Climate Change on the Basic Reproductive Number of the Lyme Disease Vector Ixodes scapularis. Environ Health Perspect 122:631–638

Oliver JH, Owsley MR, Hutcheson HJ, James AM, Chen C, Irby WS, Dotson EM, Mclain DK (1993) Conspecificity of the ticks Ixodes scapularis and I. dammini (Acari: Ixodidae). J Med Entomol 30:54

Ormiston BG (1984) Population and habitat dynamics of the white-footed mouse (Peromyscus leucopus) in a heterogeneous forest

Ostfeld R, Keesing F (2001) Biodiversity and disease risk: the case of Lyme disease. Conserv Biol 14:722–728

Ostfeld RS, Canham CD, Oggenfuss K, Winchcombe RJ, Keesing F (2006) Climate, deer, rodents, and acorns as determinants of variation in Lyme-disease risk. PLoS Biol 4:1058–1068

Ostfeld RS, Schauber EM, Canham CD, Keesing F, Jones CG, Wolff JO (2001) Effects of acorn production and mouse abundance on abundance and Borrelia burgdorferi infection prevalence of nymphal Ixodes scapularis ticks. Vector Borne Zoonotic Dis 1:55–63

Peterson AT (2003) Predicting the Geography of Species' Invasions via Ecological Niche Modeling. Q Rev Biol 78:419–433

Peterson AT (2014) Mapping disease transmission risk: enriching models using biogeography and ecology. JHU Press

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography (Cop) 31:161–175

Phillips SJ, Dudik M, Elith J, Graham CH, Leathwick J, Ferrier S, Applications SE, Jan N, Phillips SJ, Dud M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2016) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19:181–197

Piesman J, Mather TN, Sinsky RJ, Spielman A (1987) Duration of tick attachment and Borrelia burgdorferi transmission. J Clin Microbiol 25:557–8

Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, Brady OJ, Kraemer MUG, Smith DL, Moyes CL, others (2014) Mapping the zoonotic niche of Ebola virus disease in Africa. Elife 3:e04395

Poland GA, Jacobson RM (2001) The prevention of Lyme disease with vaccine. Vaccine 19:2303–2308

Pritt BS, Mead PS, Johnson DKH, Neitzel DF, Respicio-Kingry LB, Davis JP, Schiffman E, Sloan LM, Schriefer ME, Replogle AJ, Paskewitz SM, Ray JA, Bjork J, Steward CR, Deedon A, Lee X, Kingry LC, Miller TK, Feist MA, Theel ES, Patel R, Irish CL, Petersen JM (2016) Identification of a novel pathogenic Borrelia species causing Lyme borreliosis with unusually high spirochaetaemia: a descriptive study. Lancet Infect Dis 3099:1–9

Schwan TG, Piesman J (2000) Temporal changes in outer surface proteins A and C of the Lyme disease-associated spirochete, Borrelia burgdorferi, during the chain of infection in ticks and mice. J Clin Microbiol 38:382–388

Seukep SE, Kolivras KN, Hong Y, Li J, Prisley SP, Campbell JB, Gaines DN, Dymond RL (2015) An Examination of the Demographic and Environmental Variables Correlated with Lyme Disease Emergence in Virginia. Ecohealth 2011:

Snieszko SF (1974) The effects of environmental stress on outbreaks of infectious diseases of fishes*. J Fish Biol 6:197–208

Snow J (1855) On the mode of communication of cholera. John Churchill

Stanek G, Wormser GP, Gray J, Strle F (2012) Lyme borreliosis. Lancet 379:461–473

Subak S (2003) Effects of Climate on Variability in Lyme Disease Incidence in the Northeastern United States. Am J Epidemiol 157:531–538

Telford III SR, Mather TN, Moore SI, Wilson ML, Spielman A (1988) Incompetence of deer as reservoirs of the Lyme disease spirochete. Am J Trop Med Hyg 39:105–109

Troeh FR (1964) Landform Parameters Correlated to Soil Drainage. Soil Sci Soc Am J 28:808–812

Turney S, Gonzalez A, Millien V, Turney S, Gonzalez A, Millien V (2018) The negative relationship between mammal host diversity and Lyme disease incidence strengthens through time Published by : Wiley on behalf of the Ecological Society of America Stable URL : http://www.jstor.org/stable/43495279 The negative relationship bet. 95:3244–3250

Vail SG, Smith G (1998) Air temperature and relative humidity effects on behavioural activity of blacklegged tick (Acari: Ixodidae) nymphs in New Jersey. J Med Entomol 35:1025–1028

Werden L, Barker IK, Bowman J, Gonzales EK, Leighton PA, Lindsay LR, Jardine CM (2014) Geography, deer, and host biodiversity shape the pattern of lyme disease emergence in the thousand islands archipelago of Ontario, Canada. PLoS One 9

Wilson ML, Adlera GH, Spielman A (1985) Correlation Between Abundance of Deer and That of the Deer Tick, Ixodes dammini (Acari: Ixodidae). Ann Entomol Soc Am 78:172–176

Yahner RH (1992) Dynamics of a small mammal community in a fragmented forest. Am Midl Nat 381–391

Zhang X, Meltzer MI, Peña CA, Hopkins AB, Wroth L, Fix AD (2006) Economic impact of Lyme disease. Emerg Infect Dis 12:653