ERDC/GRL TR-20-2

Geospatial Research Laboratory

US Army Corps of Engineers®
Engineer Research and Development Center

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Species Distribution Modeling of *Ixodes scapularis* and Associated Pathogens in States East of the Mississippi River

Kathleen V. Payne, Sean P. Griffin, Susan L. Lyon, Robin E. Lopez, and Nicole M. Wayant

February 2020

ERDC/GRL TR-20-2
February 2020

Species Distribution Modeling of *Ixodes scapularis* and Associated Pathogens in States East of the Mississippi River

Kathleen V. Payne, Sean P. Griffin, Susan L. Lyon,
Robin E. Lopez, and Nicole M. Wayant

Geospatial Research Laboratory
U.S. Army Engineer Research and Development Center
7701 Telegraph Road
Alexandria, VA 22315-3864

Final report

Prepared for U.S. Army Corps of Engineers
Washington, DC 20314-1000

Under Program Element Number 62784
Project Number 855
Task Number 22

Abstract

The purpose of this technical report is to present results of an investigation of the spatial distribution of the deer tick, *Ixodes scapularis,* and the three parasites it carries that cause serious diseases (Lyme disease, anaplasmosis, and babesiosis) in humans. The study used the maximum entropy (MaxEnt) species niche modeling technique to produce maps predicting the probability of the presence of *Ixodes scapularis* in the eastern United States. The model makes predictions based on tick and disease surveillance data from the Army Public Health Center, and environmental data collected from satellite remote sensing platforms. Geospatial analysis was also used to locate patterns between the disease-causing parasites. The resulting prediction maps of deer tick location can be used to inform vector interception planning, which attempts to lower the risk of disease-carrying ticks from infecting humans. The maps comparing the spatial distribution of the diseases related to deer ticks can be used as a launch point for further public health study into the drivers behind parasite spread, or to direct treatment resources.

Contents

Figures

Preface

This study was conducted for the Engineering Research and Development Center (ERDC) under Project 855, "New and Enhanced Tools for Civil Military Operations (NET-CMO)."

The work was performed by the Information Generation and Management Branch of the Research Division, U.S. Army Engineer Research and Development Center, Geospatial Research Laboratory (ERDC-GRL). At the time of publication, Mr. Jeffrey B. Murphy was Branch Chief; Ms. Martha Kiene was Division Chief; and Mr. Ritchie Rodeburgh was the Technical Director for Geospatial Research and Engineering (GRE). The Deputy Director of ERDC-GRL was Ms. Valerie Carney and the Director was Mr. Gary Blohm.

COL Teresa A. Schlosser was the Commander of ERDC, and Dr. David W. Pittman was the Director.

1 Introduction

1.1 Background

This technical report (TR) describes an effort to investigate the distribution of *Ixodes scapularis* and its associated pathogens. Also known as the Blacklegged Tick or Deer Tick, this species is a vector for several pathogen species of particular interest to public health researchers: *Borrelia burgdorferi*, which causes Lyme disease (LD); *Anaplasma phagocytophilum*, which causes the disease, anaplasmosis (HGA); and for pathogens of the genus *Babesia*, which cause the disease, babesiosis (BAB).

The diseases vectored by *Ixodes scapularis* pose threats to both civilian and military communities. Previous research is not clear whether military personnel are at a higher risk than civilians, although training requirements often compel service members to spend a significant amount of time outdoors (Schubert and Melanson, 2019). Among the civilian population, Lyme disease alone "causes more than 300,000 illnesses each year in the United States," and is the most commonly occurring vector-borne disease and the sixth most commonly reported notifiable infectious disease" (Centers for Disease Control and Prevention 2017). This disease is also very costly. A study done by Hincklet et al. (2014), found that approximately 3.4 million LD tests were conducted by participating laboratories in 2008, at an estimated cost for laboratory services of $492 million. Anaplasmosis and babesiosis are much less prevalent than Lyme disease among the population, but since all three share a vector, they can be transmitted together and cause more serious illnesses when co-infected (National Institute of Health 2018). Knowing the spatial distribution of *Ixodes scapularis* and its associated pathogens could help agencies concerned about tick-borne diseases identify areas of potential disease expansion or co-infection.

1.2 Objective

This report documents the results of a study that was undertaken to better understand the distribution of *Ixodes scapularis* and its associated infectious pathogens that are a threat to humans. The Tick-Borne Disease Laboratory (TBDL) of the Army Public Health Center requested that the

Engineer Research and Development Center – Geospatial Research Laboratory (ERDC-GRL) analyze tick observation data and model the presence of the vector, *I. scapularis,* and the associated pathogens carried by this vector, using a modeling methodology similar to the one used by St. John et al. 2016. Ecological niche models were developed using the maximum entropy (Maxent) methodology (Phillips, 2006), which creates a probability map predicting where a target species will be found based on a dataset of presence observations of the species, combined with multiple environmental datasets believed to describe habitat suitability.

1.3 Approach

Maxent is a well-documented and widely-accepted ecological niche model. "[I]t has been utilized extensively for modelling species distributions… Government and non-government organizations have also adopted Maxent for large-scale, real-world biodiversity mapping applications" (Elith et al. 2011). Using a set of environmental variables and estimated presence locations, the Maxent models employed in this study generated probability distributions for *Ixodes scapularis* and its associated pathogens in the states east of the Mississippi River, with each pixel representing "a predicted suitability of conditions for the species" (Phillips, 2006).

2 Data and Methods

A .csv dataset containing approximate GPS locations, collected from 1994-2017, for multiple tick species and their associated pathogens, were provided from the Walter Reed Biosystematics Unit (WRBU) via the VectorMap data portal (http://www.vectormap.org). This dataset contained information from multiple sources, using multiple collection methods. The data encompassed all four seasons and included reports from throughout the United States.

The dataset was reduced to only the data entries for the species *Ixodes scapularis* that were located within the region of interest (ROI). The ROI was defined as those states east of the Mississippi River, as this waterway serves as a biophysical constraint and because most of the source data for the target species was restricted to that ROI. Fifty-two entries were eliminated from the dataset because the reliability of the location information for these entries was questionable, leaving a remaining sample set containing 5,283 presence entries for *I. scapularis,* which were used to create the Maxent model.

Maxent (Maximum Entropy) is a widely-accepted species niche model, based on a machine learning algorithm that uses presence-only field data to make a prediction from incomplete information. Maxent generates a probability distribution product in the form of a raster image, with each pixel representing “a predicted suitability of conditions for the species” (Phillips, 2006).

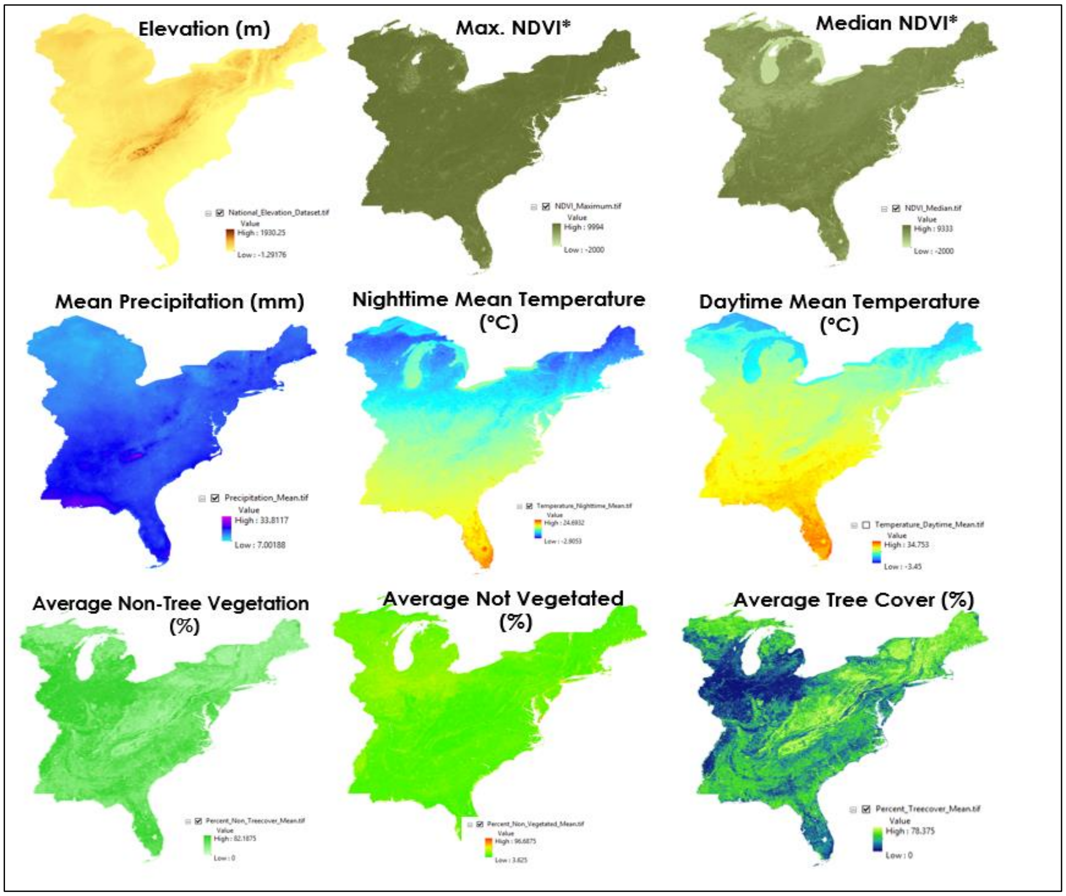
The Maxent model also requires the use of environmental data sets for habitat suitability prediction. These environmental variables should cover phenomena that would be expected by the researchers to influence how suitable an area is to the presence of a species. The variables used in this study, pictured in Figure 1, were: elevation (m), maximum normalized difference vegetation index (NDVI)*, median NDVI, average non-tree vegetation (%), average tree cover (%), average not vegetated pixels (%), average daytime temperature (°C), average nighttime temperature (°C), and average precipitation (mm). These datasets were chosen based on existing literature and conversations with entomologists and researchers

---

* A description of the NDVI can be found here: https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD13Q1

with subject matter expertise on the habitat of *I. scapularis*. All datasets were accessed, processed, and exported from Google Earth Engine at a resolution of 1000 m.

Figure 1. Environmental variables used (1000 m).



Using location and species information for *I. scapularis* and environmental input raster data, a Maxent model was generated using a toolbox and a standard operating procedure (SOP) was developed for use in ArcGIS (Griffin, 2017). Results are discussed in section 3.1. The method of collection (tick drags, box trapping of hosts, or individual collections) was not a concern in this initial model since Maxent is a presence-only model. It does not take into account the number of individual observations collected at a single location.

After the initial presence model was created, the dataset was further reduced to only those presence entries collected through the Human Tick Test Kit Program, because those were the specimens tested for disease-causing pathogens. The Department of Defense (DoD) Human Tick Test

Kit Program (HTTKP) is a free tick identification and testing service provided by the Laboratory Sciences Directorate at the Army Public Health Center, serving all DOD personnel, as well as DOD civilians, reservists, retirees, and their family members (Army Public Health Center 2018). Pathogen screening on submitted specimens is conducted by the TBDL using polymerase chain reaction (PCR) chemistries to analyze ticks for evidence of infection with the agents of several established and emerging tick-borne diseases, including those for Lyme disease (*Borrelia burgdorferi*), human granulocytic anaplasmosis (*Anaplasma phagocytophilum*), and human babesiosis (*Babesia microti*) (Army Public Health Center, n.d.). The dataset was reduced to include only those entries that were assumed to have been tested for all three diseases, as some entries were marked “Not Testable” or contained specific data entries in comments indicating that they were not tested for all three pathogens.

Of the remaining 5,256 *I. scapularis* presence entries, 1,265 were positive for *Borrelia burgdorferi*, 914 were positive for *Anaplasma phagocytophilum*, and 908 were positive for *Babesia spp*. The longitude, latitude, and species information for these positive entries, along with the same environmental input raster data used for the overall species niche model, were used to generate Maxent species niche models for each of the three pathogens. Results are discussed in sections 3.2, 3.3, and 3.4.

The Maxent methodology generates measures to assess the suitability of the model’s prediction in the form of a sensitivity plot, as shown in Figures 3, 6, 10, and 13. The model criteria used to assess the Maxent prediction is area under the curve (AUC). As described in Hahn et al. 2016, which also used the Maxent methodology, “The AUC is a measure of the accuracy of the habitat suitability model. The AUC ranges from 0.5 to 1, where a value of 0.5 indicates that the model is not useful for distinguishing suitable from unsuitable habitat while a value of 1 indicates a perfect model.”

Once the pathogen presence probability models were complete, the models were compared using the Raster Calculator tool in the Map Algebra toolbox under Spatial Analyst Tools in ArcGIS Pro 2.3.3. One raster was created using the expression:

“Borrelia_burgdorferi.asc” – “Babesia_spp.asc” (1)

Another raster was created using the expression:

"Borrelia_burgdorferi.asc" – "Anaplasma_phagocytophilum.asc" (2)

The results from Equation 1 generated a raster that displays a representation of the areas where the probability of the presence of Borrelia *burgdorferi* is greater than the probability of the presence of *Babesia spp;* and the areas where the probability of the presence of *Babesia spp*. is greater than the probability of the presence of *Borrelia burgdorferi*.

The results from Equation 2 generated a raster that displays a representation of the areas where the probability of the presence of *Borrelia burgdorferi* is greater than the probability of the presence of *Anaplasma phagocytophilum;* and areas where the probability of the presence of *Anaplasma phagocytophilum* is greater than the probability of the presence of *Borrelia burgdorferi*.

Areas where overlapping probabilities are shown could indicate locations where co-infections are likely to occur. As *I. scapularis* initially acquires pathogens from previous host species, such as rodents or deer, before passing them on to humans; probability distributions could also indicate regions where higher occurrences of host species carrying disease pathogens may exist.

The resulting models from these equations are discussed in section 3.5.

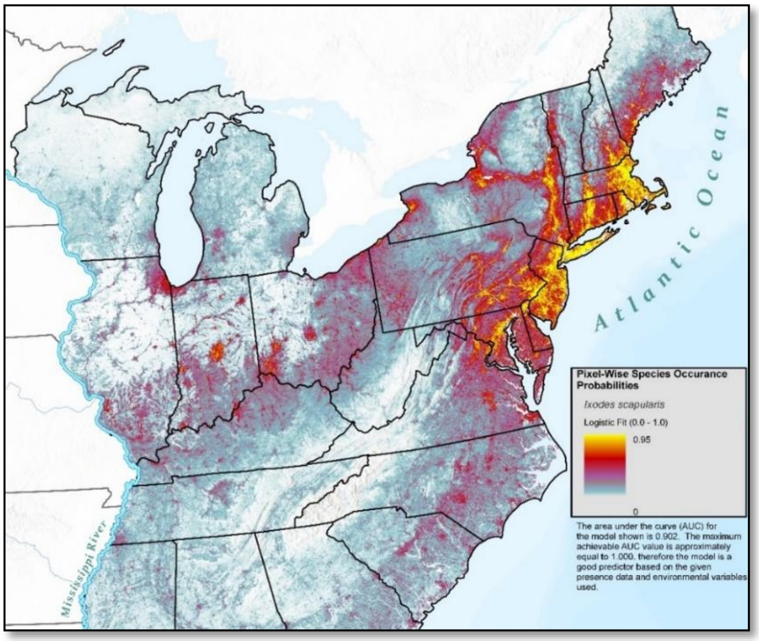
3 Results and Discussion

3.1 Habitat suitability model for *Ixodes scapularis*

Results for the *Ixodes scapularis* habitat prediction model are shown in Figure 2. The color ramp indicates the probability that *Ixodes scapularis* is predicted to be found in that pixel, derived from the presence locations, and the environmental variables.

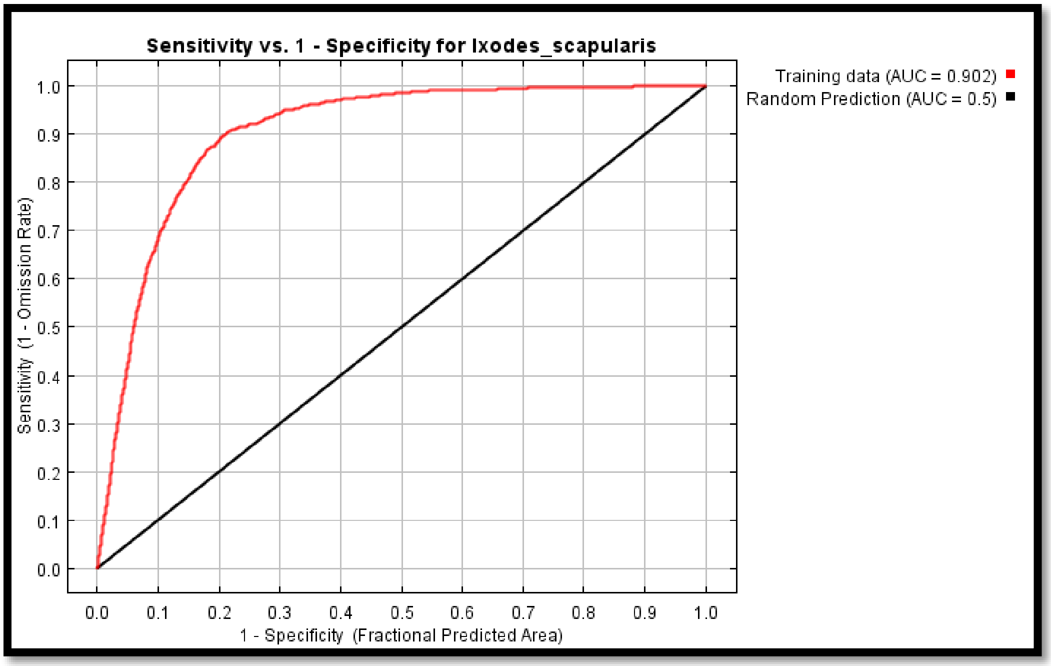
As shown in Figure 3, the AUC for the Maxent habitat suitability model for *I. scapularis*, is 0.902. The maximum achievable AUC value is approximately equal to 1.000, which suggests that the model is a good predictor of *I. scapularis* habitat suitability based on the given presence data and the environmental variables used.

Figure 2. Portion of Habitat Suitability Model for *Ixodes scapularis.* The full model can be found in Appendix A.1.



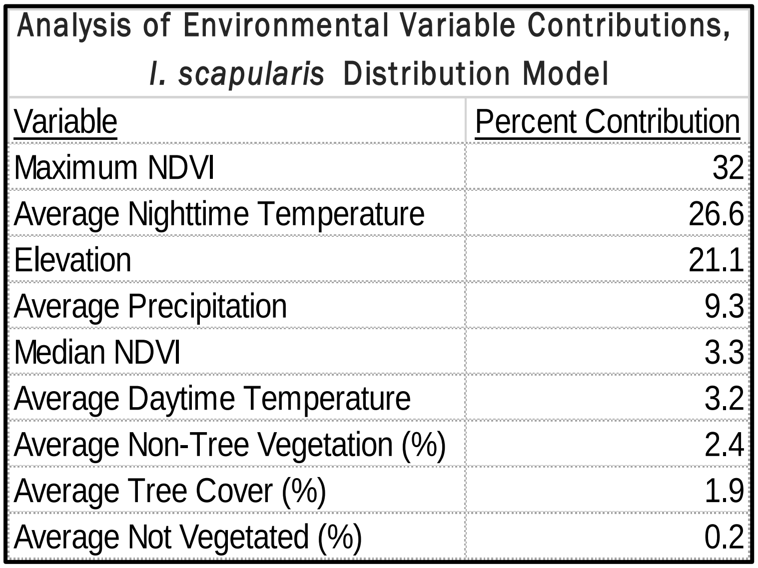
The area under the curve (AUC) for the model shown is 0.902. The maximum achievable AUC value is approximately equal to 1.000, therefore the model is a good predictor based on the given presence data and environmental variables used.

Figure 3. Receiver operating characteristic (ROC) curve for the *I. scapularis* Maxent model and related AUC values.



The Maxent methodology also examines the contribution of each of the environmental variables to the prediction model. Out of the environmental variables used, maximum NDVI, average nighttime temperature, elevation, and average precipitation had the highest percent contribution to the model, as shown in Figure 4.

Figure 4. Percentage of environmental variable contribution for the *I. scapularis* habitat distribution model.



Analysis of Environmental Variable Contributions, *I. scapularis* Distribution Model

Variable	Percent Contribution
Maximum NDVI	32
Average Nighttime Temperature	26.6
Elevation	21.1
Average Precipitation	9.3
Median NDVI	3.3
Average Daytime Temperature	3.2
Average Non-Tree Vegetation (%)	2.4
Average Tree Cover (%)	1.9
Average Not Vegetated (%)	0.2

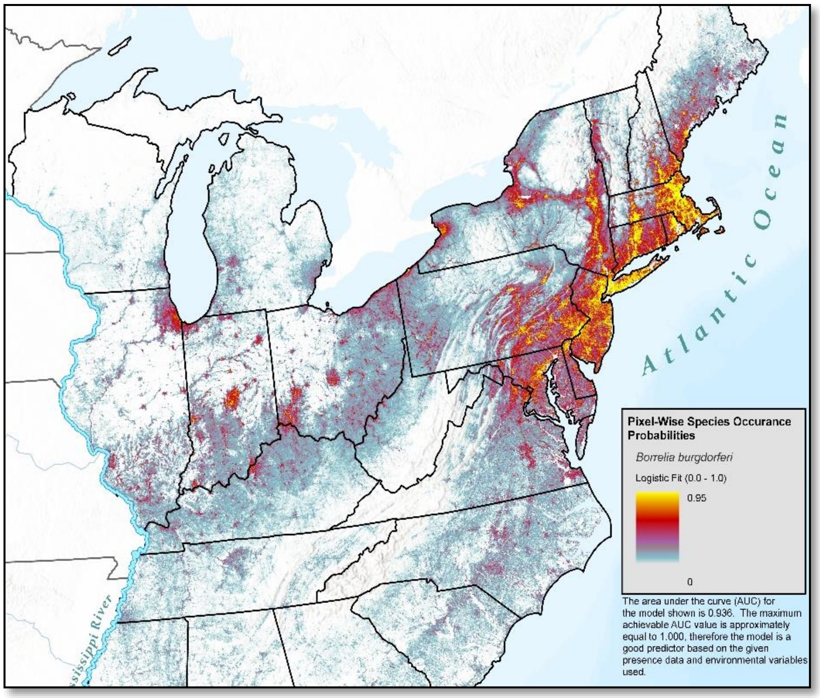
Figure 2 displays predicted habitat suitability of *Ixodes scapularis* in areas similar to other research that has been performed (Diuk-Wasser et al. 2010). The highest species occurrence probability is predicted along the northeast coast, from northern Virginia to Maine, affecting all Mid-Atlantic and New England states. Pockets of high species habitat suitability were also predicted in Illinois, Indiana, Ohio, northern Kentucky, the Ohio Valley section of West Virginia, and the coastal plains of North and South Carolina, as well as in parts of southern Michigan and southern Wisconsin. While some presence was predicted in all southern states, the occurrence probability was lower. This aligns with a previous study, which found that “mortality due to desiccation stress is greater under southern than under northern conditions” (Ginsberg et al. 2017). Desiccation tolerance refers to the species’ ability to endure removal of moisture. Southern black-legged ticks tend to hide under a layer of leaves to retain moisture, whereas northern ticks tend to climb plant stems and wait for their victims to pass (USGS 2017). There was also a noticeable absence of species occurrence shown in regions corresponding with some high-altitude areas in the Appalachian Mountain Range.

When compared to similar research, (Diuk-Wasser et al. 2010) the areas of Wisconsin and Northern Illinois should display a higher predicted species occurrence probability than shown in this model. Differences between this model and similar studies could be attributed to bias in the data, for a number of reasons. First, the samples submitted through the HTTKP were submitted by individuals who may not have known the exact location where a specimen was first acquired. Since the observations are the result of health intervention, the data is biased toward areas with higher populations, better transportation networks, and better health care accessibility. VectorMap data comes from contributions from a heterogeneous mix of providers, and so the data are focused on the areas where the HTTKP is implemented, or from areas close to where contributors operate.

3.2 Pathogen Presence Model for *Borrelia burgdorferi*

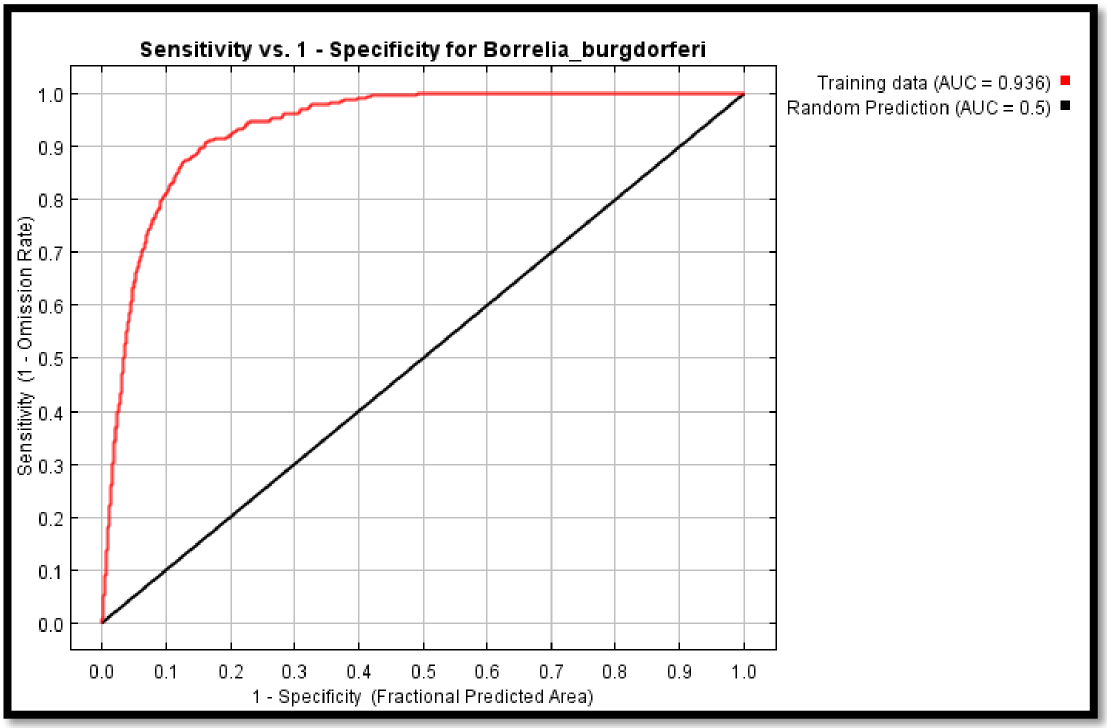
Results for the *Borrelia burgdorferi* presence prediction model are shown in Figure 5. The AUC for the model, shown in Figure 6, is 0.936, which suggests that the model is a good predictor based on the given presence data and environmental variables used.

Figure 5. Portion of pathogen distribution model for *Borrelia burgdorferi*. The full model can be found in AppendixA.2.



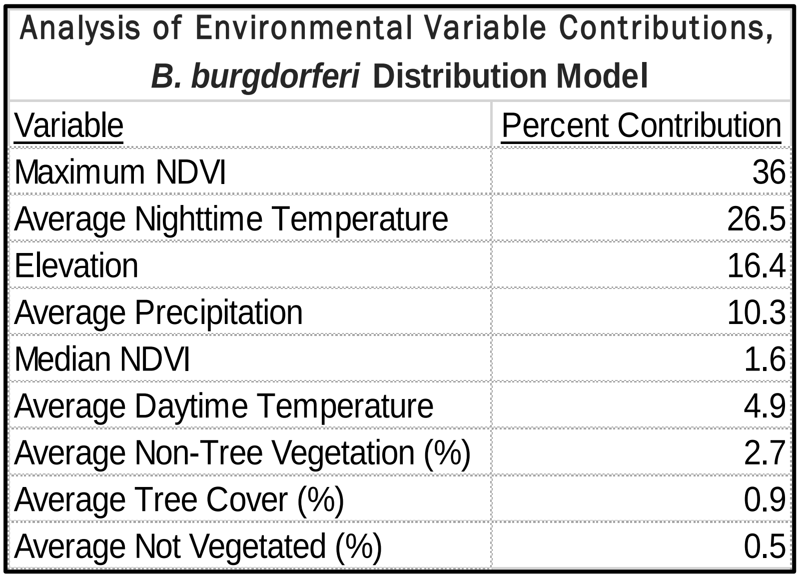
The area under the curve (AUC) for the model shown is 0.936. The maximum achievable AUC value is approximately equal to 1.000, therefore the model is a good predictor based on the given presence data and environmental variables used.

Figure 6. ROC curve for the *Borrelia burgdorferi* Maxent model and related AUC values.



Out of the environmental variables used, maximum NDVI, average nighttime temperature, elevation, and average precipitation had the highest percent contribution to the model, as shown in Figure 7.

Figure 7. Percentage of environmental variable contribution for the *B. burgdorferi* distribution model.

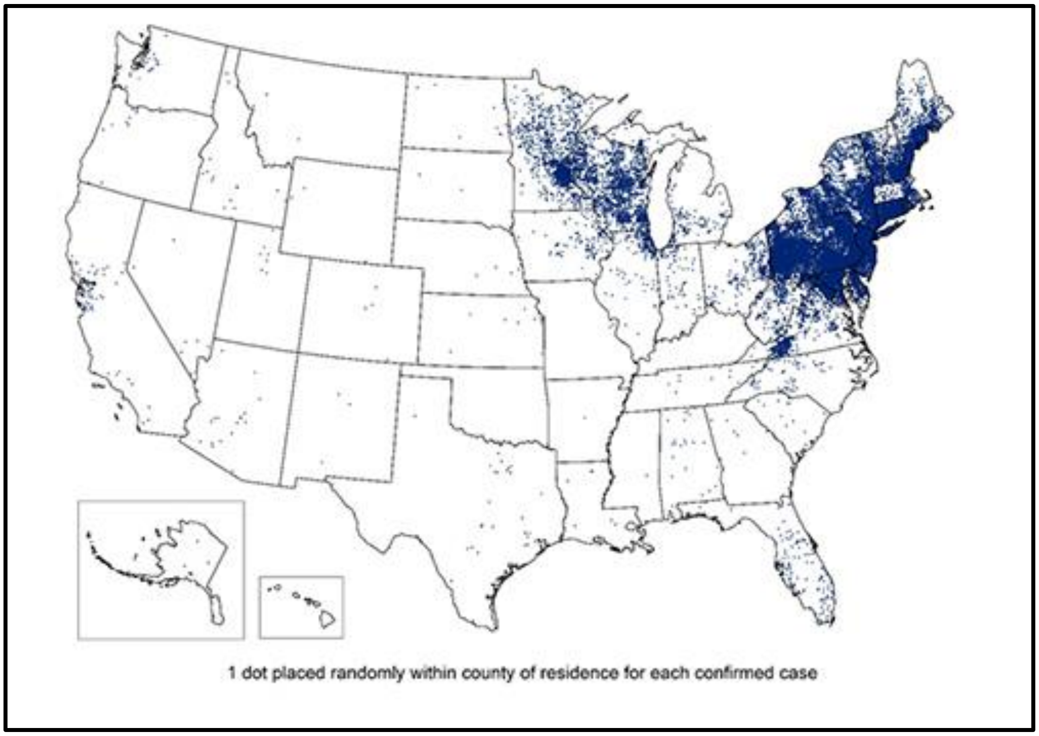


Analysis of Environmental Variable Contributions,
B. burgdorferi **Distribution Model**

Variable	Percent Contribution
Maximum NDVI	36
Average Nighttime Temperature	26.5
Elevation	16.4
Average Precipitation	10.3
Median NDVI	1.6
Average Daytime Temperature	4.9
Average Non-Tree Vegetation (%)	2.7
Average Tree Cover (%)	0.9
Average Not Vegetated (%)	0.5

The model, shown in Figure 5, displays a high predicted presence of *Borrelia burgdorferi* in areas similar to those shown in the Centers for Disease Control (CDC) maps of reported cases of Lyme disease during the years that the data encompasses, shown in Figure 8 (Centers for Disease Control and Prevention, 2017). It displays a high level of predicted pathogen presence in many of the same regions corresponding with the highest species occurrence probability of the vector. However, there are several areas that differ between the two Maxent models, indicating zones where higher occurrences of host species carrying *Borrelia burgdorferi* may exist.

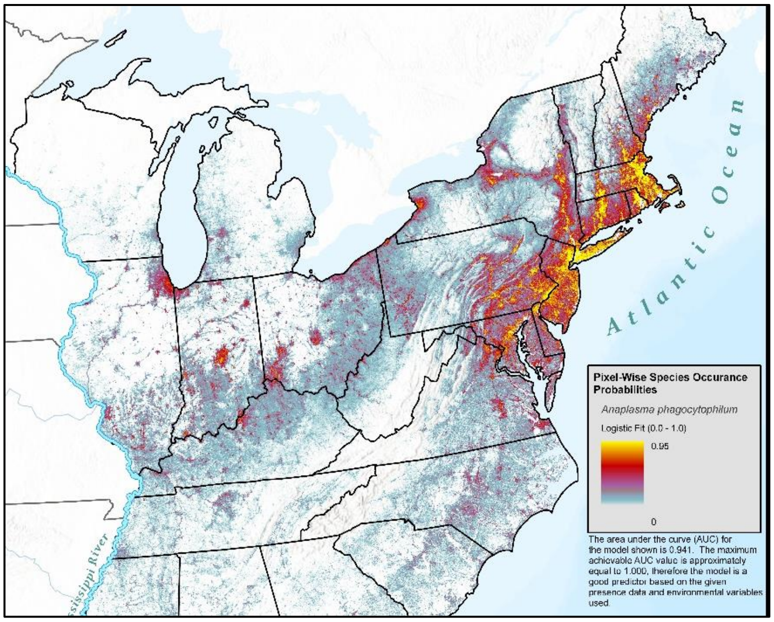
Figure 8. Reported cases of Lyme disease – United States, 2017 (Centers for Disease Control and Prevention, 2017).



3.3 Pathogen presence model for *Anaplasma phagocytophilum*

Results for the *Anaplasma phagocytophilum* presence prediction model are shown in Figure 9. The AUC for the model shown in Figure 10 is 0.941, which suggests that the model is a good predictor based on the given presence data and environmental variables used. Out of the environmental variables used, maximum NDVI, average nighttime temperature, elevation, and average precipitation had the highest percent contribution to the model, as shown in Figure 11.

Figure 9. Portion of the pathogen presence model for *Anaplasma phagocytophilum.* The full model can be found in AppendixA.3.



The area under the curve (AUC) for the model shown is 0.941. The maximum achievable AUC value is approximately equal to 1.000, therefore the model is a good predictor based on the given presence data and environmental variables used.

Figure 10. ROC curve for the *Anaplasma phagocytophilum* Maxent model and related AUC values.

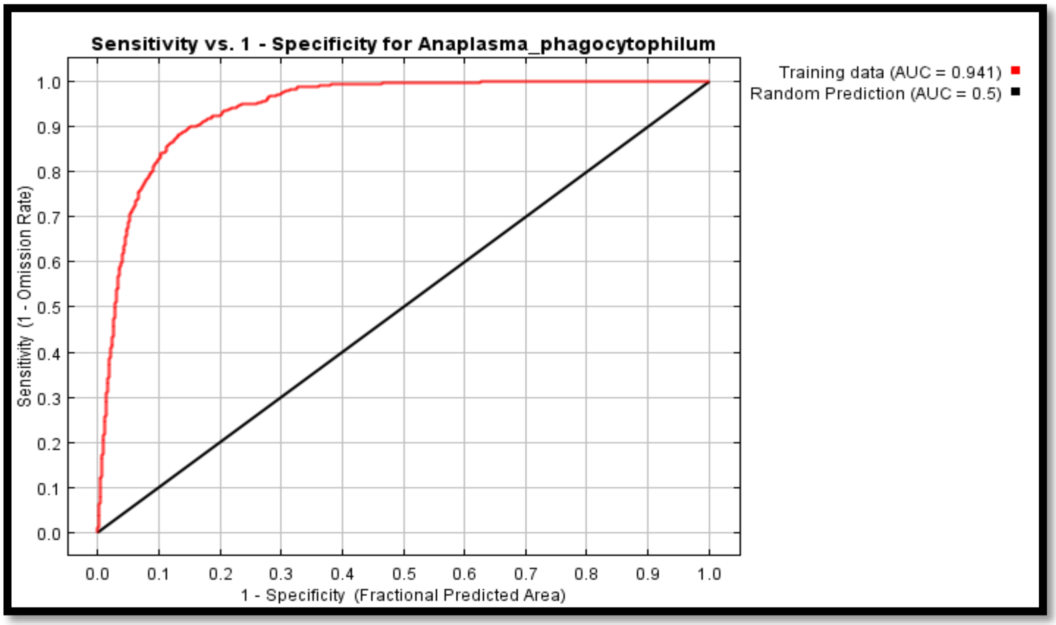
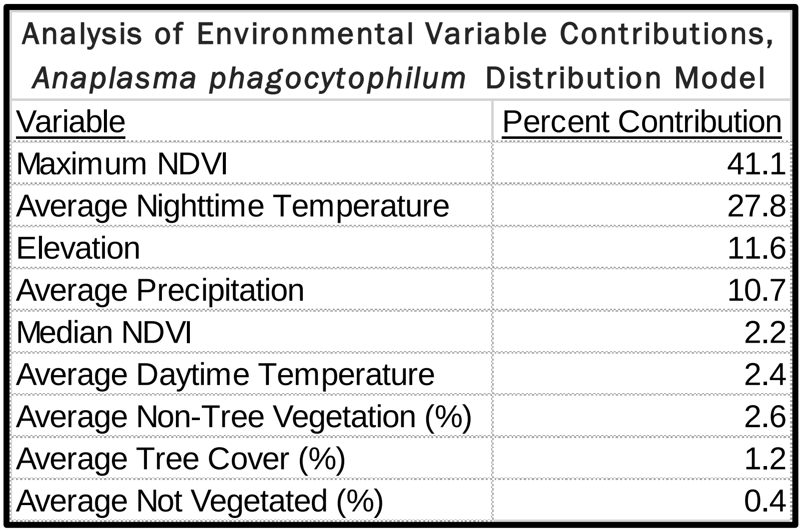


Figure 11. Percentage of environmental variable contribution for the *A. phagocytophilum* distribution model.



Analysis of Environmental Variable Contributions,
Anaplasma phagocytophilum Distribution Model

Variable	Percent Contribution
Maximum NDVI	41.1
Average Nighttime Temperature	27.8
Elevation	11.6
Average Precipitation	10.7
Median NDVI	2.2
Average Daytime Temperature	2.4
Average Non-Tree Vegetation (%)	2.6
Average Tree Cover (%)	1.2
Average Not Vegetated (%)	0.4

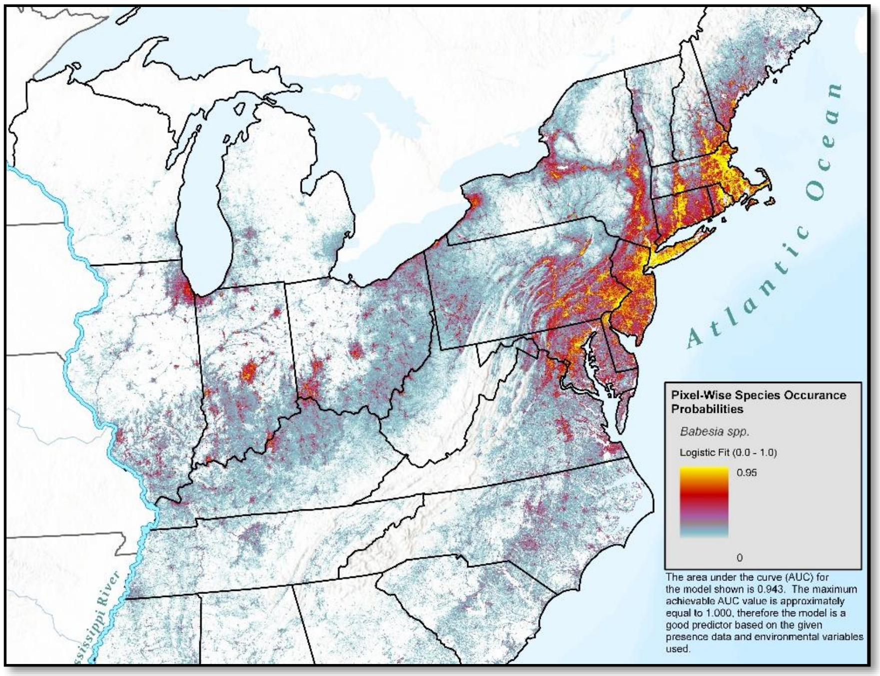
The model, shown in Figure 9, illustrates predicted presence of *Anaplasma phagocytophilum* in many of the same regions corresponding with the highest species occurrence probability of the vector. However, there are several areas that differ between the two models, indicating zones where higher occurrences of host species carrying *Anaplasma phagocytophilum* may exist.

3.4 Pathogen presence model for *Babesia spp.*

Results for the *Babesia spp.* presence prediction model are shown in Figure 12.

The AUC for the model shown in Figure 13 is 0.943, which suggests the model is a good predictor based on the given presence data and environmental variables used. Out of the environmental variables used, maximum NDVI, average nighttime temperature, elevation, and average precipitation had the highest percent contribution to the model, as shown in Figure 14.

Figure 12. Portion of the pathogen presence model for *Babesia spp.* The full model can be found in AppendixA.4.



The area under the curve (AUC) for the model shown is 0.943. The maximum achievable AUC value is approximately equal to 1.000, therefore the model is a good predictor based on the given presence data and environmental variables used.

Figure 13. ROC curve for the *Babesia spp.* Maxent model and related AUC values.

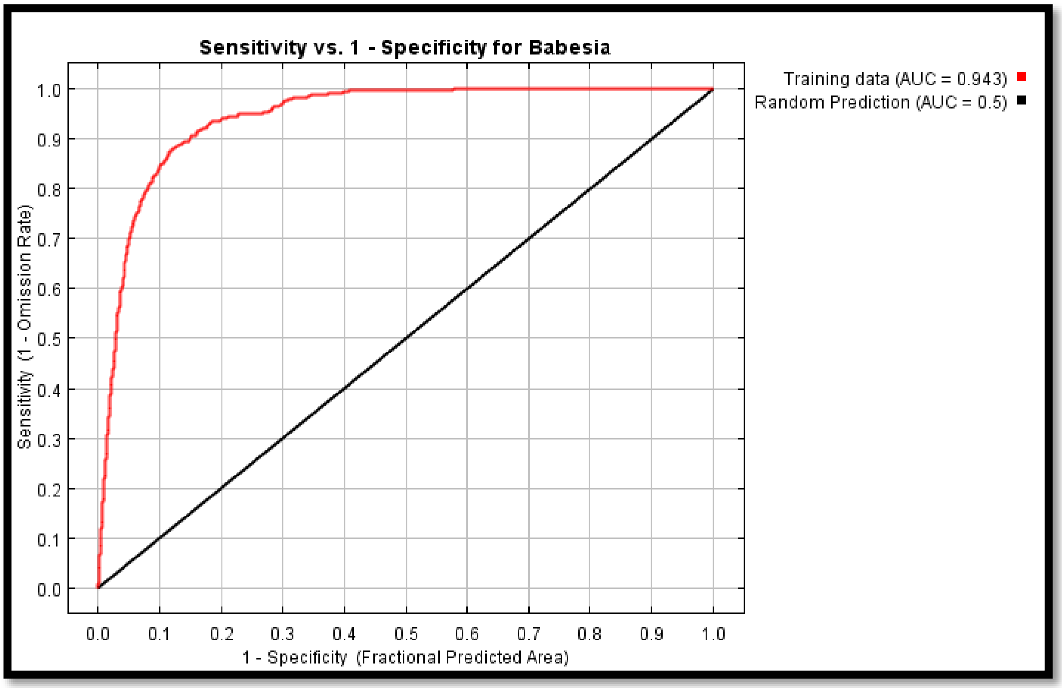
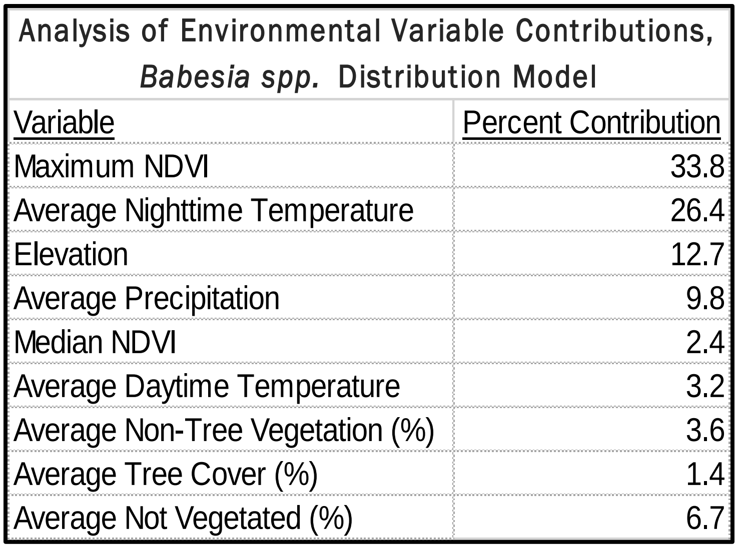


Figure 14. Percentage of environmental variable contribution for the *Babesia spp.* distribution model.



Analysis of Environmental Variable Contributions, *Babesia spp.* Distribution Model

Variable	Percent Contribution
Maximum NDVI	33.8
Average Nighttime Temperature	26.4
Elevation	12.7
Average Precipitation	9.8
Median NDVI	2.4
Average Daytime Temperature	3.2
Average Non-Tree Vegetation (%)	3.6
Average Tree Cover (%)	1.4
Average Not Vegetated (%)	6.7

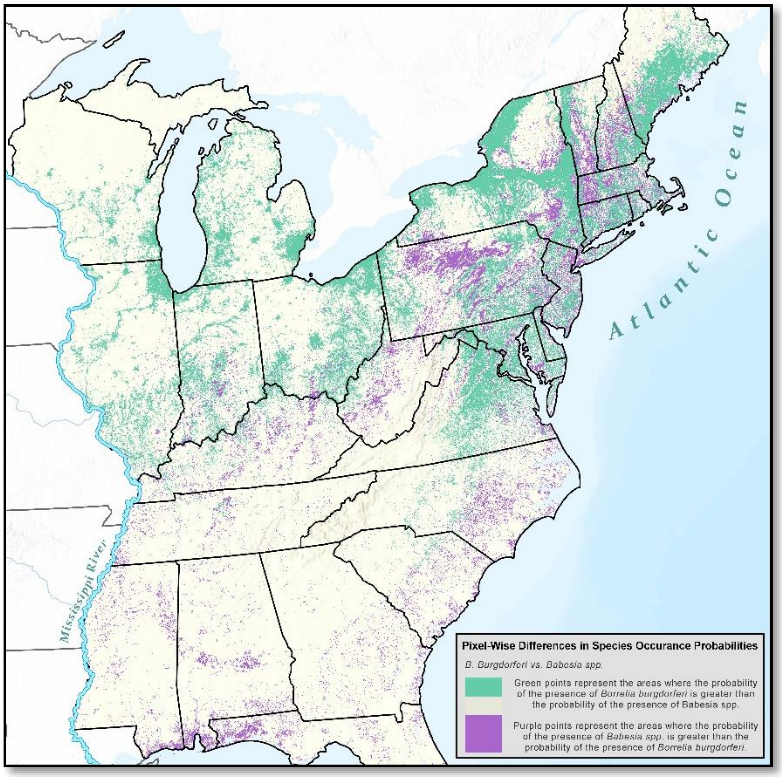
The model, shown in Figure 12, illustrates predicted presence of *Babesia spp.* in many of the same regions corresponding with the highest species occurrence probability of the vector. However, there are several areas that differ between the two models, indicating zones where higher occurrences of host species carrying *Babesia spp.* may exist.

3.5 Discussion of compared pathogen models

3.5.1 Probability of *Borrelia burgdorferi* presence vs. probability of *Babesia spp*. presence

The resulting raster model is shown in Figure 15. In this model, the green points represent the areas where the probability of the presence of *Borrelia burgdorferi* is greater than the probability of the presence of *Babesia spp*. The purple points represent the areas where the probability of the presence of *Babesia spp*. is greater than the probability of the presence of *Borrelia burgdorferi*.

Figure 15. Portion of the probability model of *Borrelia burgdorferi* presence vs. *Babesia spp*. Presence. The full model can be found in AppendixA.5.

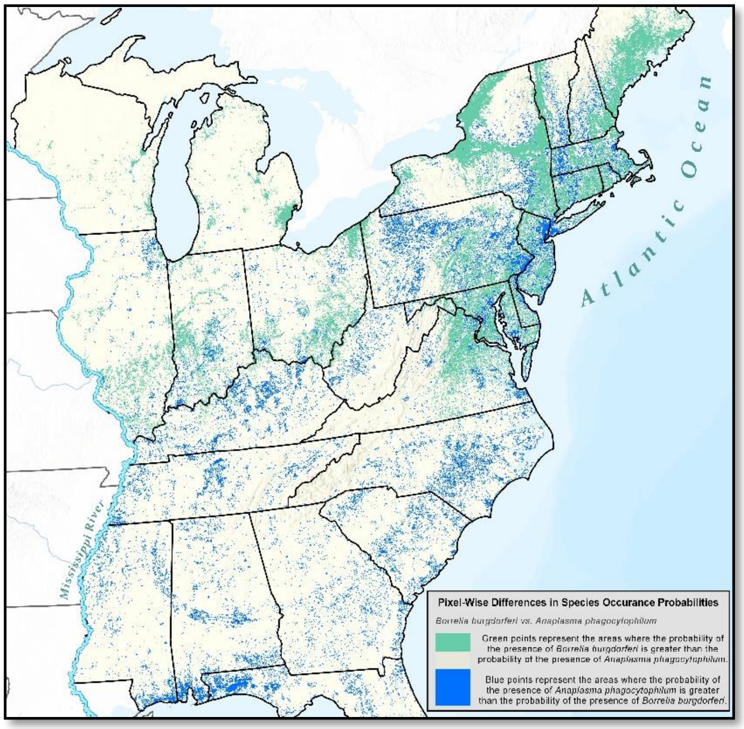


Results show that the probability of the presence of *Borrelia burgdorferi* is greater than the probability of the presence of *Babesia spp*. primarily in the Midwest region, such as Wisconsin, Michigan and Northern Illinois. Probabilities are mixed in the Mid-Atlantic and New England states with no large regions of one or the other species, which indicates that these are areas where co-infections could likely occur. Areas in the south mainly showed a higher probability of the presence of *Babesia spp*. than the probability of the presence of *Borrelia burgdorferi,* although exceptions existed in Northern Virginia, the Ohio Valley section of West Virginia, parts of North and South Carolina, and in Northern Kentucky.

3.5.2 Probability of *Borrelia burgdorferi* presence vs. probability of *Anaplasma phagocytophilum* presence

The resulting raster model is shown in Figure 16. In this model, the green points represent the areas where the probability of the presence of *Borrelia burgdorferi* is greater than the probability of the presence of *Anaplasma phagocytophilum*. The blue points represent the areas where the probability of the presence of *Anaplasma phagocytophilum* is greater than the probability of the presence of *Borrelia burgdorferi*.

Figure 16. Portion of the probability model of *Borrelia burgdorferi* presence vs. probability of *Anaplasma phagocytophilum* presence. The complete model can be found in Appendix A.6.



Results show that the probability of the presence of *Borrelia burgdorferi* is greater than the probability of the presence of *Anaplasma phagocytophilum* in Wisconsin and Michigan. Probabilities are mixed in the Mid-Atlantic, New England, and in Northern Virginia, indicating general areas where co-infections could occur, with some pockets of higher likelihood of one pathogen over the other. Most areas in the south mainly showed a higher probability of the presence of *Anaplasma phagocytophilum* than the probability of the presence of *Borrelia burgdorferi,* although mixed areas existed in Northern Virginia, the Ohio Valley section of West Virginia, and in Northern Kentucky.

3.6 Limitations and Areas for Further Study

While Maxent is a recognized tool for generating species distribution models, all modeling methods have limitations. The primary limitations for the Maxent modeling technique are: (1) the prevalence of species cannot be exactly determined, regardless of sample size; and (2) sample selection bias has a stronger effect on presence-only models (Elith et al. 2011).

The models resulting from this study likely contain some sample selection bias, as a high concentration of the tick specimens were collected in or submitted from densely populated areas, areas with ease of access to roads, or areas where the HTTKP is used more commonly by individuals and health care providers. This is a result of bias in the source data, not resulting from the model itself.

Additional research and exploration could include modeling distribution changes over time, analysis and mapping of species distribution by county, and use of other machine learning modeling methods, some of which can incorporate abundance data or presence and absence data to predict species habitat suitability. Additionally, epidemiologists and entomologists may be particularly interested in further exploring the differences this study identified in the spatial distribution of the disease pathogens associated with *Ixodes scapularis*. These differences have implications for disease intervention and prevention planning in civilian and military contexts.

4 Summary

Species distribution modeling was conducted utilizing the established ecological niche modeling method known as maximum entropy (Maxent), in order to determine areas of ideal habitat for the vector, *Ixodes scapularis*, and potential areas of infection or co-infection for diseases associated with this vector. This study found that the highest species occurrence probability for *Ixodes scapularis* was predicted along the northeast coast, from northern Virginia to Maine, but affecting all Mid-Atlantic, New England, and Midwestern states within the ROI. While some presence was found in all southern states, the occurrence probability was lower. There was also a noticeable absence of species occurrence shown in regions corresponding with some areas in the Appalachian Mountain Range. Out of the environmental variables used, maximum NDVI, average nighttime temperature, elevation, and average precipitation had the highest percent contribution to all models.

The models for pathogens associated with *Ixodes scapularis* all display a high predicted presence of each pathogen in many of the same regions corresponding with the highest species occurrence probability of the vector. However, there are several areas that differ between the models, indicating zones where a higher occurrence of host species carrying disease pathogens may exist.

When comparing the probability of the presence of *Borrelia burgdorferi* to the probability of the presence of *Babesia spp.* and *Anaplasma phagocytophilum,* it was found that presence probabilities are mixed in the Mid-Atlantic and New England states, indicating areas where co-infections could occur. Most areas in the south showed a higher probability of the presence of *Anaplasma phagocytophilum* and *Babesia spp.* than the probability of the presence of *Borrelia burgdorferi*.

Despite potential bias derived from data sources, these results can be useful to the Army Public Health Center and other entities concerned about tick-borne diseases. The results can be used to help pinpoint areas in need of more prevention efforts, surveillance, and additional data collection. The results are also important for the identification of areas of potential disease expansion or co-infection.

References

Army Public Health Center. 2018. *Human Tick Test Kit Program*. Accessed July 29, 2019. https://phc.amedd.army.mil/topics/envirohealth/epm/Pages/HumanTickTestKitProgram.aspx.

____. n.d. *Organization*. Accessed July 25, 2019. https://phc.amedd.army.mil/organization/Pages/default.aspx.

____. n.d. "The Department of Defense Human Tick Test Kit Program, FACT SHEET 18-077-0317." Accessed July 29, 2019. https://phc.amedd.army.mil/PHC%20Resource%20Library/Human%20Tick%20Test%20Kit%20FS_18_077_0317%20FINAL%20(2).pdf.

Centers for Disease Control and Prevention (CDC), DPDx - Laboratory Identification of Parasites of Public Health Concern. 2017. *Babesiosis*. October 30. Accessed July 25, 2019. https://www.cdc.gov/dpdx/babesiosis/index.html.

Centers for Disease Control and Prevention. 2017. *Why is CDC concerned about Lyme disease?* Accessed August 8, 2019. https://www.cdc.gov/lyme/why-is-cdc-concerned-about-lyme-disease.html.

Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), Division of Vector-Borne Diseases (DVBD). 2019. *Anaplasmosis*. Accessed July 25, 2019. https://www.cdc.gov/anaplasmosis/index.html.

____. 2019. *Lyme Disease*. Accessed July 25, 2019. https://www.cdc.gov/lyme/index.html.

Diuk-Wasser, M. A., G. Vourc'h, P. Cislo, A. G. Hoen, F. Melton, S. A. Hamer, and M. Rowland.. 2010. "Field and climate-based model for predicting the density of host-seeking nymphal Ixodes scapularis, an important vector of tick-borne disease agents in the eastern United States." *Global Ecology and Biogeography* (Blackwell Publishing Ltd) 19: 504-514. doi:10.1111/j.1466-8238.2010.00526.x.

Elith, J., S. J. Phillips, T. Hastie, M. Dudı´k, Y. En Chee, and C. J. Yates. 2011. "A statistical explanation of MaxEnt for ecologists." *Diversity and Distributions* 43–57. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1472-4642.2010.00725.x.

Ginsberg, H. S, M. Albert, L. Acevedo, M. C. Dyer, I. M. Arsnoe, J. I. Tsao, T. N. Mather, and R. A. LeBrun. 2017. "Environmental factors affecting survival of immature ixodes scapularis and implications for geographical distribution of lyme disease: The climate/behavior hypothesis." Edited by University of Maryland, College Park, United States Utpal Pal. *PLoS ONE* 12 (1). Accessed August 8, 2019. doi:https://doi.org/10.1371/journal.pone.0168723.

Griffin, S. P. 2017. "Tactical-level disease vector hazard mapping using the maximum entropy (maxent) algorithm and ArcGIS standard operating procedures (SOP)." SOP, Geospatial Research Laboratory, Engineer Research and Development Center. Accessed 2019.

Hahn, M.h B., C. S. Jarnevich, A. J. Monaghan, and R. J. Eisen. 2016. "Modeling the geographic distribution of ixodes scapularis and ixodes pacificus (Acari: Ixodidae) in the contiguous United States." *Journal of Medical Entomology* 53 (5): 1176–1191. Accessed June 24, 2019. doi:https://doi.org/10.1093/jme/tjw076.

Hincklet, A. F., N. P. Connally, J. I. Meek, B. J. Johnson, M. M. Kemperman, K. A. Feldman, J. L. White, and P. S Mead. 2014. "Lyme disease testing by large commercial laboratories in the United States." *Clin Infect Dis*. 59 (5): 676-681. Accessed August 8, 2019. doi:10.1093/cid/ciu397.

National Institute of Health. 2018. *Lyme Disease Co-Infection*. November 16. Accessed August 13, 2019. https://www.niaid.nih.gov/diseases-conditions/lyme-disease-co-infection.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum entropy modeling of species geographic distributions." *Ecological Modelling* 190 (3-4): 231-259. Accessed July 29, 2019. https://www.cs.princeton.edu/~schapire/papers/ecolmod.pdf.

Schubert, CPT S. L., and LTC V. R. Melanson. 2019. "Prevalence of Lyme Disease Attributable to Military Service at the USMA, West Point NY: FY2016–2018." *Military Medicine*. Accessed August 21, 2019. doi:10.1093/milmed/usz156.

St. John, H. K., M. L. Adams, P. M. Masuoka, J. J. Flyer-Adams, P. J. Rozmajzl, E. Y. Stromdahl, and A. L. Richards. 2016. "Prevalence, distibution, and development of an ecological niche model of dermacentor variabilis ticks positive for rickettsia montanensis." *Vector-Borne and Zoonotic Diseases* 16 (4). Accessed March 15, 2019. doi:http://doi.org/10.1089/vbz.2015.1856.

USGS. 2017. "It's the Heat and the Humidity, New Study Finds: Why Lyme Disease is Common in the North, Rare in the South." January 11. Accessed April 9, 2019. doi:https://doi.org/10.1371/journal.pone.0168723 .

Software Documentation

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment.

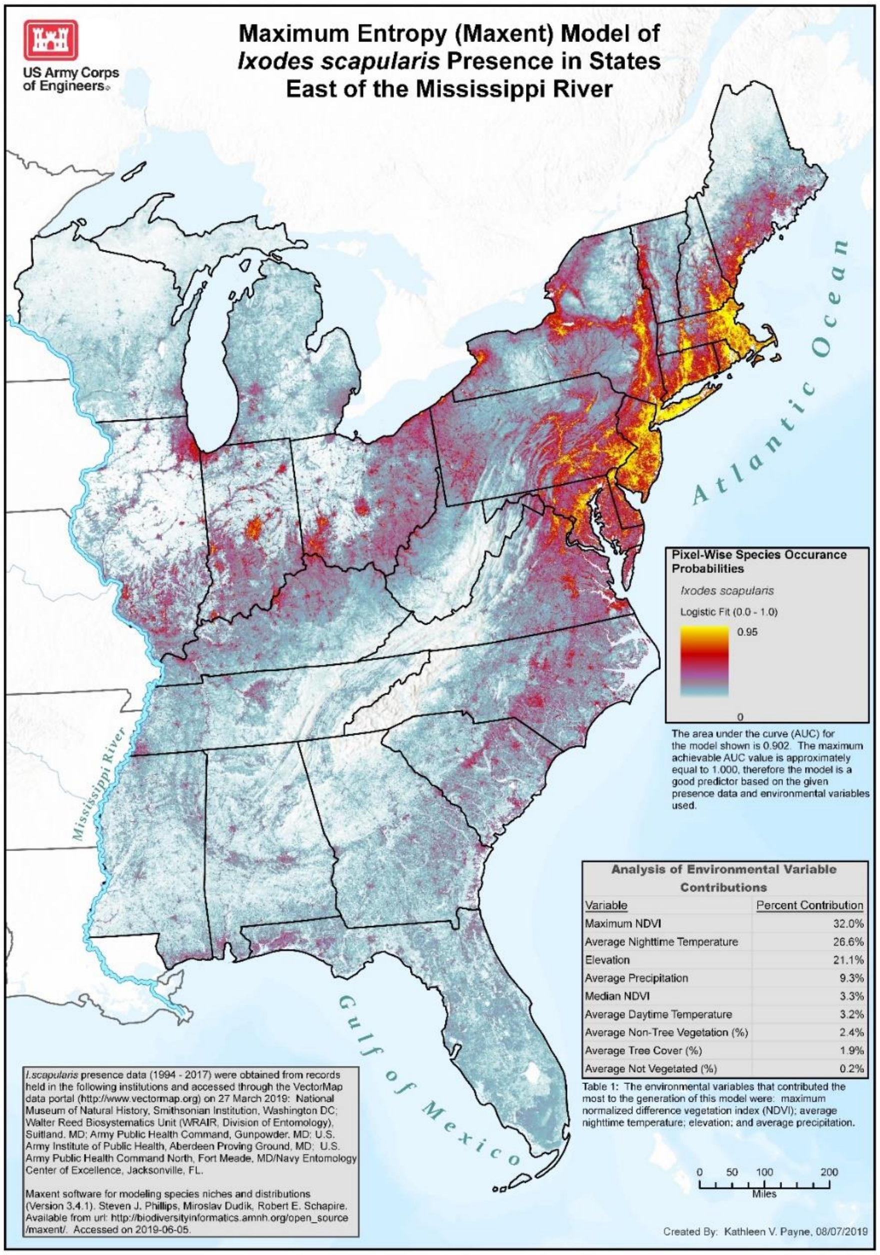
Google News Initiative. n.d. *OpenRefine*. Accessed March-August, 2019. http://openrefine.org/.

Phillips, S. J., M. Dudík, and R. E. Schapire. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from URL: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2019-06-05.

“Maps throughout this technical note were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.”

Appendix A: Distribution Maps

A.1 Habitat Suitability Model of *Ixodes scapularis*



The area under the curve (AUC) for the model shown is 0.902. The maximum achievable AUC value is approximately equal to 1.000, therefore the model is a good predictor based on the given presence data and environmental variables used.

Analysis of Environmental Variable Contributions

Variable	Percent Contribution
Maximum NDVI	32.0%
Average Nighttime Temperature	26.6%
Elevation	21.1%
Average Precipitation	9.3%
Median NDVI	3.3%
Average Daytime Temperature	3.2%
Average Non-Tree Vegetation (%)	2.4%
Average Tree Cover (%)	1.9%
Average Not Vegetated (%)	0.2%

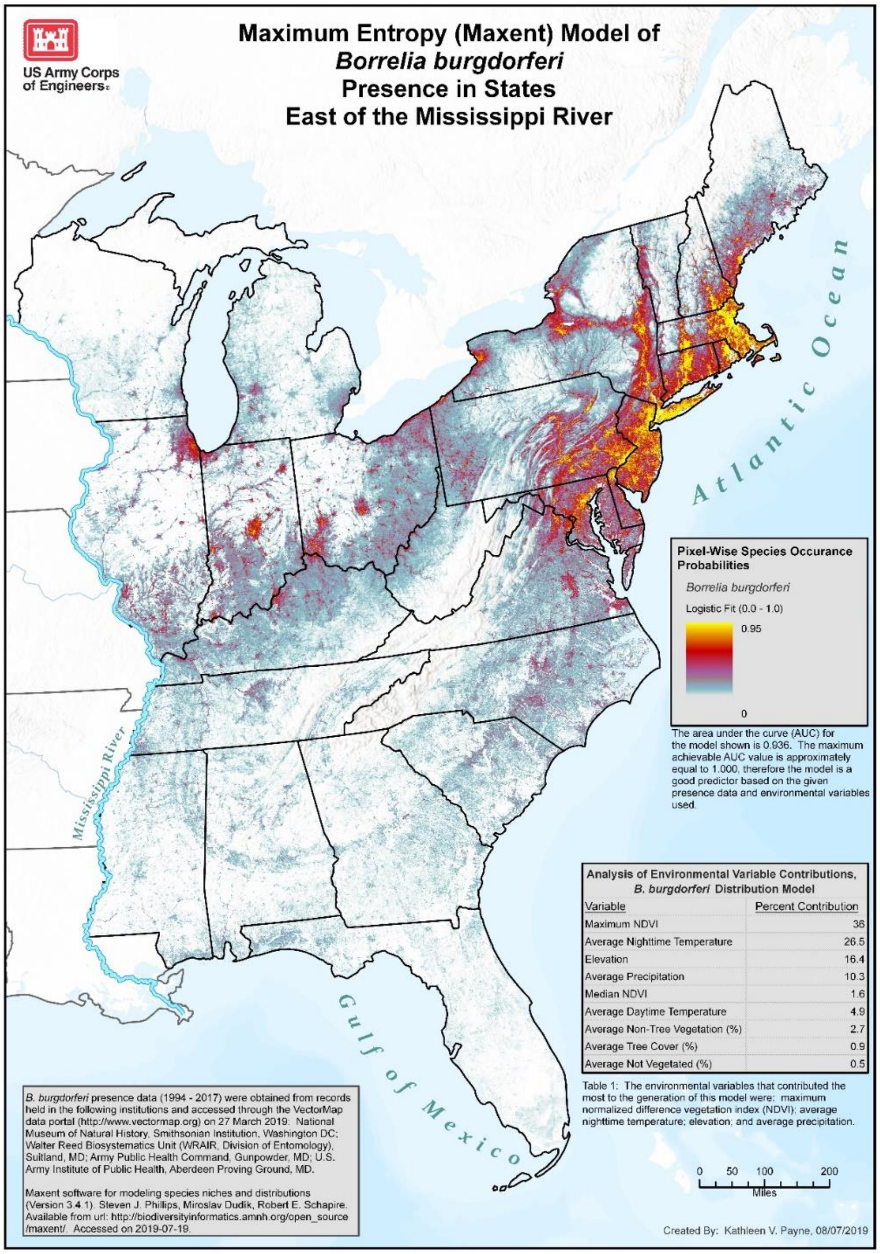
Table 1: The environmental variables that contributed the most to the generation of this model were: maximum normalized difference vegetation index (NDVI); average nighttime temperature; elevation; and average precipitation.

I.scapularis presence data (1994 - 2017) were obtained from records held in the following institutions and accessed through the VectorMap data portal (http://www.vectormap.org) on 27 March 2019: National Museum of Natural History, Smithsonian Institution, Washington DC; Walter Reed Biosystematics Unit (WRAIR, Division of Entomology), Suitland, MD; Army Public Health Command, Gunpowder, MD; U.S. Army Institute of Public Health, Aberdeen Proving Ground, MD; U.S. Army Public Health Command North, Fort Meade, MD/Navy Entomology Center of Excellence, Jacksonville, FL.

Maxent software for modeling species niches and distributions (Version 3.4.1). Steven J. Phillips, Miroslav Dudik, Robert E. Schapire. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2019-06-05.

Created By: Kathleen V. Payne, 08/07/2019

A.2 Pathogen Distribution Model for *Borrelia burgdorferi*



Maximum Entropy (Maxent) Model of *Borrelia burgdorferi* Presence in States East of the Mississippi River

The area under the curve (AUC) for the model shown is 0.936. The maximum achievable AUC value is approximately equal to 1.000, therefore the model is a good predictor based on the given presence data and environmental variables used.

Analysis of Environmental Variable Contributions, *B. burgdorferi* Distribution Model

Variable	Percent Contribution
Maximum NDVI	36
Average Nighttime Temperature	26.5
Elevation	16.4
Average Precipitation	10.3
Median NDVI	1.6
Average Daytime Temperature	4.9
Average Non-Tree Vegetation (%)	2.7
Average Tree Cover (%)	0.9
Average Not Vegetated (%)	0.5

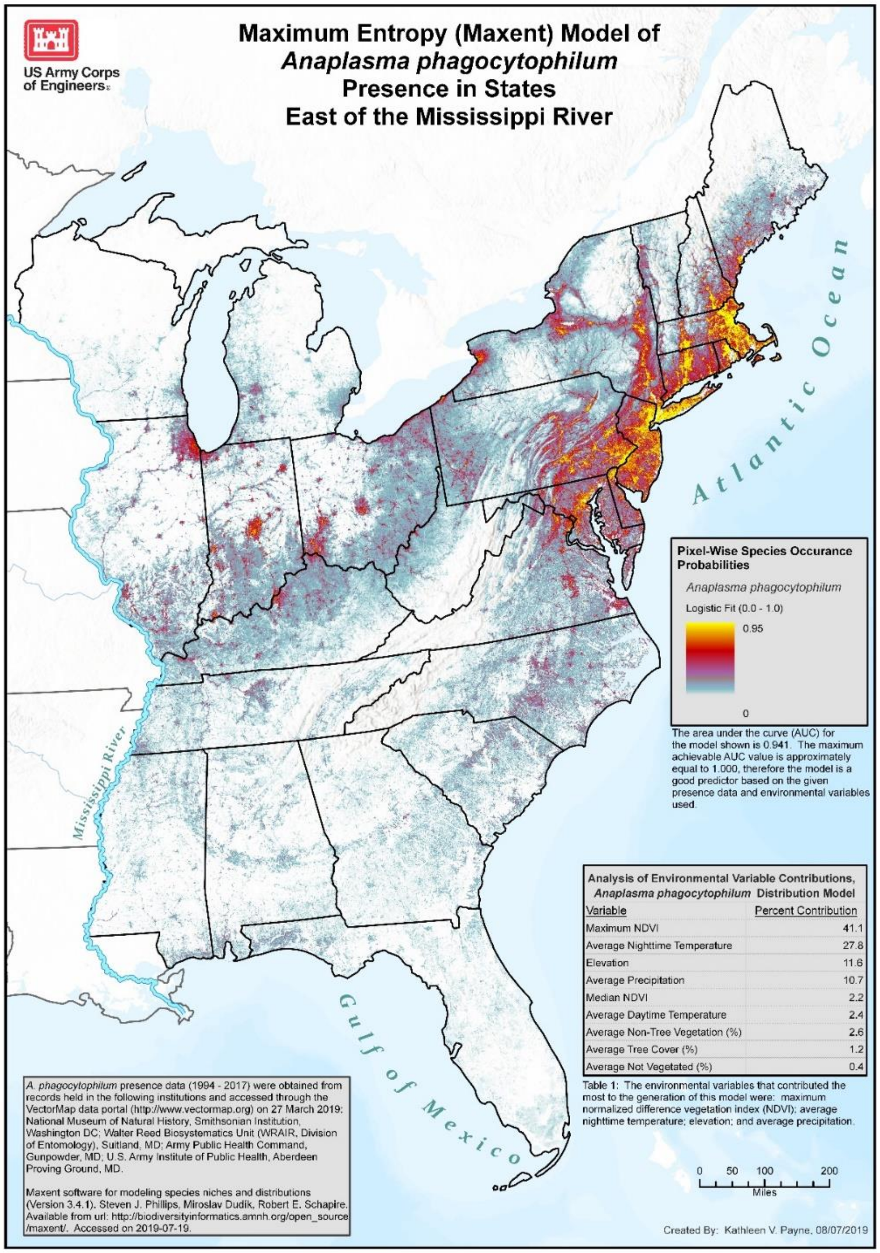
Table 1: The environmental variables that contributed the most to the generation of this model were: maximum normalized difference vegetation index (NDVI); average nighttime temperature; elevation; and average precipitation.

B. burgdorferi presence data (1994 - 2017) were obtained from records held in the following institutions and accessed through the VectorMap data portal (http://www.vectormap.org) on 27 March 2019: National Museum of Natural History, Smithsonian Institution, Washington DC; Walter Reed Biosystematics Unit (WRAIR, Division of Entomology), Suitland, MD; Army Public Health Command, Gunpowder, MD; U.S. Army Institute of Public Health, Aberdeen Proving Ground, MD.

Maxent software for modeling species niches and distributions (Version 3.4.1). Steven J. Phillips, Miroslav Dudik, Robert E. Schapire. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2019-07-19.

Created By: Kathleen V. Payne, 08/07/2019

A.3 Pathogen Distribution Model for *Anaplasma phagocytophilum*



Maximum Entropy (Maxent) Model of *Anaplasma phagocytophilum* Presence in States East of the Mississippi River

The area under the curve (AUC) for the model shown is 0.941. The maximum achievable AUC value is approximately equal to 1.000, therefore the model is a good predictor based on the given presence data and environmental variables used.

Analysis of Environmental Variable Contributions, *Anaplasma phagocytophilum* Distribution Model

Variable	Percent Contribution
Maximum NDVI	41.1
Average Nighttime Temperature	27.8
Elevation	11.6
Average Precipitation	10.7
Median NDVI	2.2
Average Daytime Temperature	2.4
Average Non-Tree Vegetation (%)	2.6
Average Tree Cover (%)	1.2
Average Not Vegetated (%)	0.4

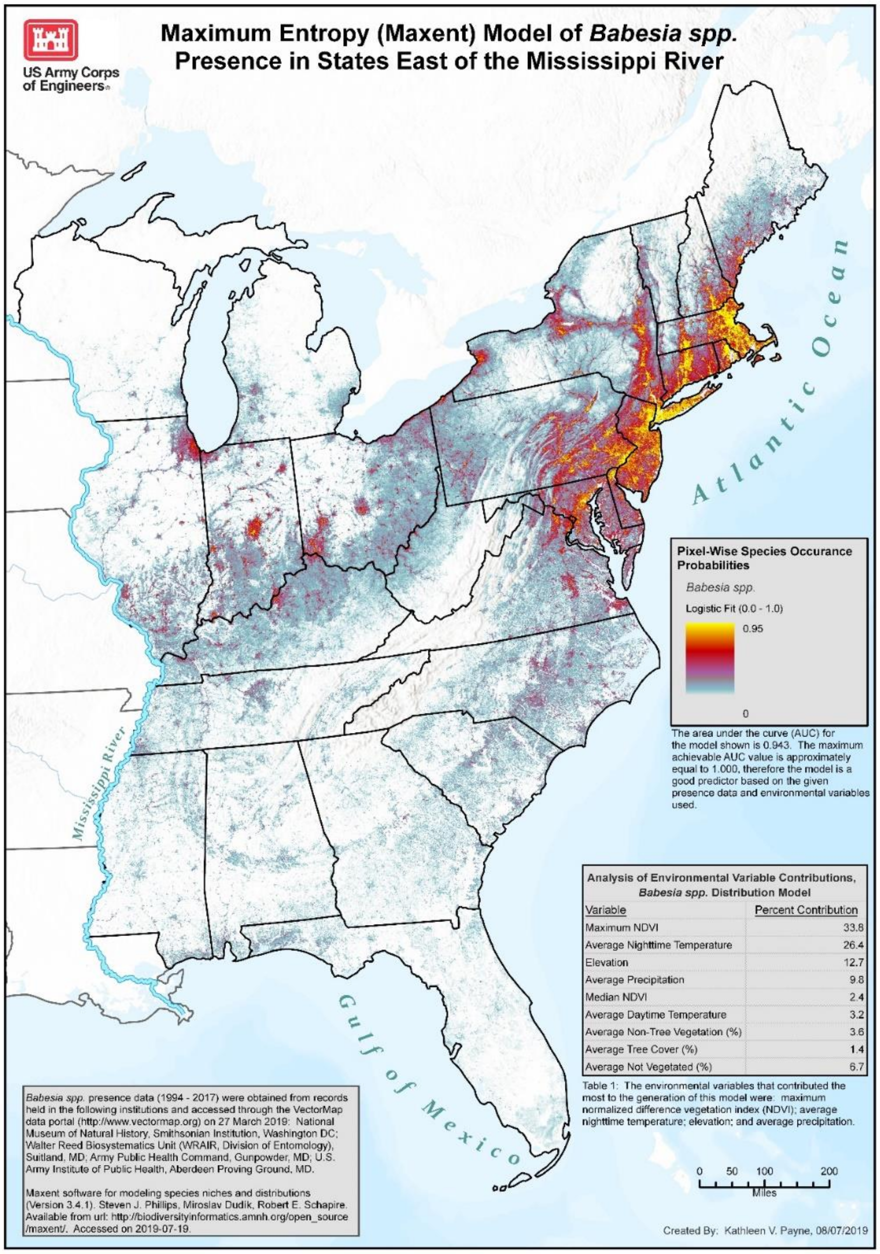
Table 1: The environmental variables that contributed the most to the generation of this model were: maximum normalized difference vegetation index (NDVI); average nighttime temperature; elevation; and average precipitation.

A. phagocytophilum presence data (1994 - 2017) were obtained from records held in the following institutions and accessed through the VectorMap data portal (http://www.vectormap.org) on 27 March 2019: National Museum of Natural History, Smithsonian Institution, Washington DC; Walter Reed Biosystematics Unit (WRAIR, Division of Entomology), Suitland, MD; Army Public Health Command, Gunpowder, MD; U.S. Army Institute of Public Health, Aberdeen Proving Ground, MD.

Maxent software for modeling species niches and distributions (Version 3.4.1). Steven J. Phillips, Miroslav Dudik, Robert E. Schapire. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2019-07-19.

Created By: Kathleen V. Payne, 08/07/2019

A.4 Pathogen Distribution Model for *Babesia spp.*



Maximum Entropy (Maxent) Model of *Babesia spp.* Presence in States East of the Mississippi River

The area under the curve (AUC) for the model shown is 0.943. The maximum achievable AUC value is approximately equal to 1.000, therefore the model is a good predictor based on the given presence data and environmental variables used.

Analysis of Environmental Variable Contributions, *Babesia spp.* Distribution Model

Variable	Percent Contribution
Maximum NDVI	33.8
Average Nighttime Temperature	26.4
Elevation	12.7
Average Precipitation	9.8
Median NDVI	2.4
Average Daytime Temperature	3.2
Average Non-Tree Vegetation (%)	3.6
Average Tree Cover (%)	1.4
Average Not Vegetated (%)	6.7

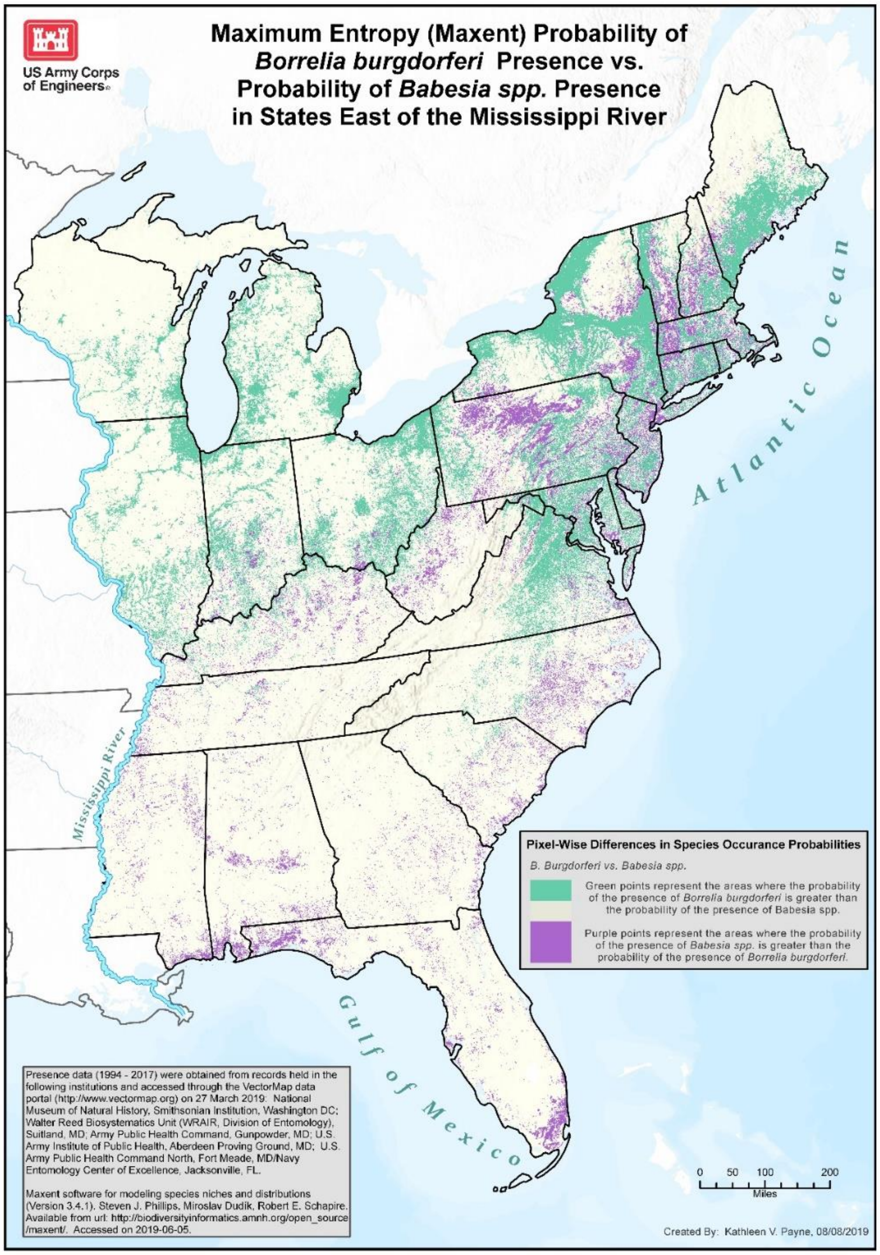
Table 1: The environmental variables that contributed the most to the generation of this model were: maximum normalized difference vegetation index (NDVI); average nighttime temperature; elevation; and average precipitation.

Babesia spp. presence data (1994 - 2017) were obtained from records held in the following institutions and accessed through the VectorMap data portal (http://www.vectormap.org) on 27 March 2019: National Museum of Natural History, Smithsonian Institution, Washington DC; Walter Reed Biosystematics Unit (WRAIR, Division of Entomology), Suitland, MD; Army Public Health Command, Gunpowder, MD; U.S. Army Institute of Public Health, Aberdeen Proving Ground, MD.

Maxent software for modeling species niches and distributions (Version 3.4.1). Steven J. Phillips, Miroslav Dudik, Robert E. Schapire. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2019-07-19.

Created By: Kathleen V. Payne, 08/07/2019

A.5 Probability of *Borrelia burgdorferi* Presence vs. Probability of *Babesia spp.* Presence



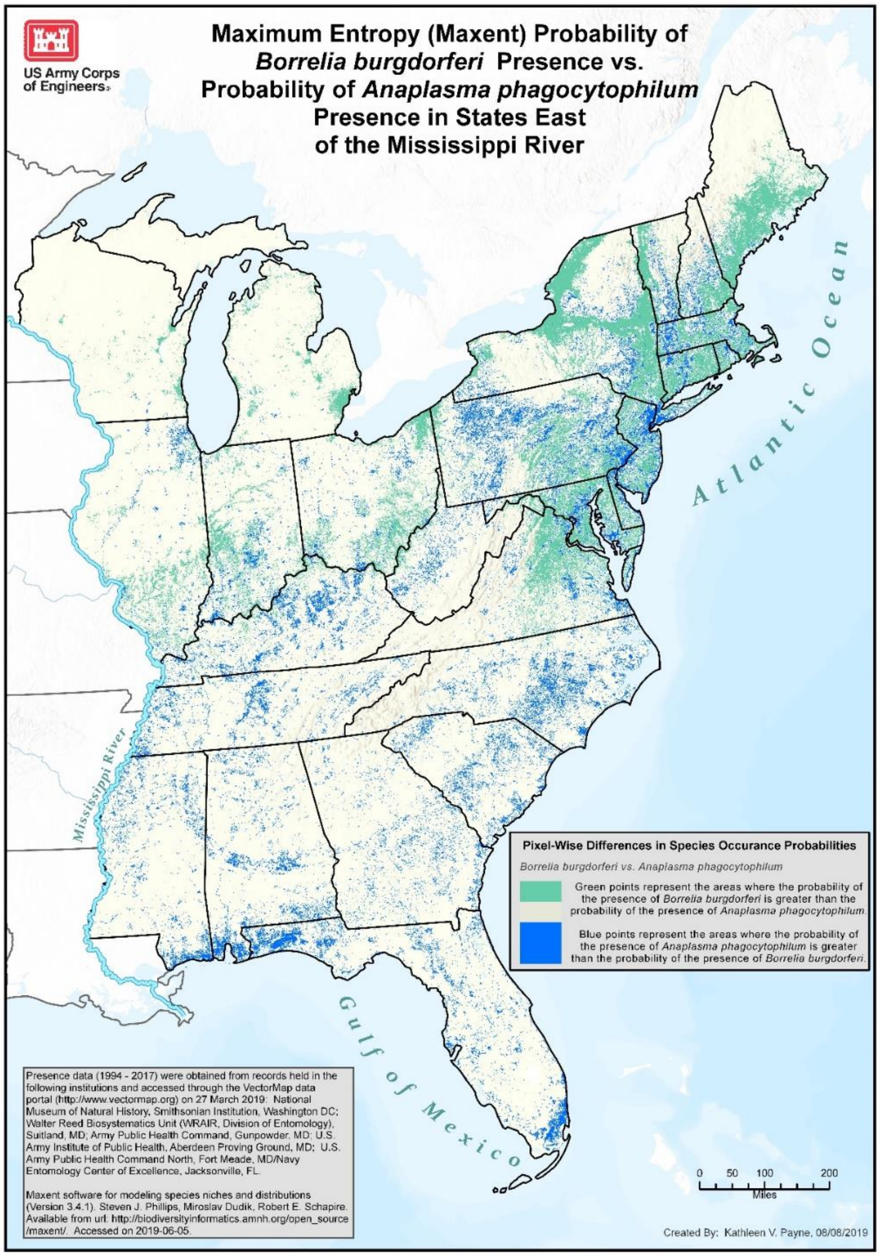
Maximum Entropy (Maxent) Probability of *Borrelia burgdorferi* Presence vs. Probability of *Babesia spp.* Presence in States East of the Mississippi River

Presence data (1994 - 2017) were obtained from records held in the following institutions and accessed through the VectorMap data portal (http://www.vectormap.org) on 27 March 2019: National Museum of Natural History, Smithsonian Institution, Washington DC; Walter Reed Biosystematics Unit (WRAIR, Division of Entomology), Suitland, MD; Army Public Health Command, Gunpowder, MD; U.S. Army Institute of Public Health, Aberdeen Proving Ground, MD; U.S. Army Public Health Command North, Fort Meade, MD/Navy Entomology Center of Excellence, Jacksonville, FL.

Maxent software for modeling species niches and distributions (Version 3.4.1). Steven J. Phillips, Miroslav Dudik, Robert E. Schapire. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2019-06-05.

Created By: Kathleen V. Payne, 08/08/2019

A.6 Probability of *Borrelia burgdorferi* Presence vs. Probability of *Anaplasma phagocytophilum* Presence



Maximum Entropy (Maxent) Probability of *Borrelia burgdorferi* Presence vs. Probability of *Anaplasma phagocytophilum* Presence in States East of the Mississippi River

US Army Corps of Engineers

Pixel-Wise Differences in Species Occurance Probabilities

Borrelia burgdorferi vs. Anaplasma phagocytophilum

🟩 Green points represent the areas where the probability of the presence of *Borrelia burgdorferi* is greater than the probability of the presence of *Anaplasma phagocytophilum*.

🟦 Blue points represent the areas where the probability of the presence of *Anaplasma phagocytophilum* is greater than the probability of the presence of *Borrelia burgdorferi*.

Presence data (1994 - 2017) were obtained from records held in the following institutions and accessed through the VectorMap data portal (http://www.vectormap.org) on 27 March 2019: National Museum of Natural History, Smithsonian Institution, Washington DC; Walter Reed Biosystematics Unit (WRAIR, Division of Entomology), Suitland, MD; Army Public Health Command, Gunpowder, MD; U.S. Army Institute of Public Health, Aberdeen Proving Ground, MD; U.S. Army Public Health Command North, Fort Meade, MD/Navy Entomology Center of Excellence, Jacksonville, FL.

Maxent software for modeling species niches and distributions (Version 3.4.1). Steven J. Phillips, Miroslav Dudik, Robert E. Schapire. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2019-06-05.

Created By: Kathleen V. Payne, 08/08/2019

Unit Conversion Factors

Multiply	By	To Obtain
acres	4,046.873	square meters
degrees Fahrenheit	(F-32)/1.8	degrees Celsius
feet	0.3048	meters
inches	0.0254	meters
microns	1.0 E-06	meters
miles (nautical)	1,852	meters
miles (U.S. statute)	1,609.347	meters
mils	0.0254	millimeters
square feet	0.09290304	square meters
square inches	6.4516 E-04	square meters
square miles	2.589998 E+06	square meters
square yards	0.8361274	square meters
yards	0.9144	meters

Acronyms

APHC	Army Public Health Center
AUC	area under the curve
BAB	babesiosis
CDC	Centers for Disease Control
DoD	Department of Defense
HTTKP	Human Tick Test Kit Program
LD	Lyme disease
NDVI	normalized difference vegetation index
PCR	polymerase chain reaction
ROC	receiver operating characteristic
ROI	region of interest
SOP	standard operating procedure
TBDL	Tick-Borne Disease Laboratory
WRBU	Walter Reed Biosystematics Unit

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
February 2020

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

Species Distribution Modeling of *Ixodes scapularis* and Associated Pathogens in States East of the Mississippi River

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62784

6. AUTHOR(S)

Kathleen V. Payne, Sean P. Griffin, Susan L. Lyon,
Robin E. Lopez, and Nicole M. Wayant

5d. PROJECT NUMBER
855

5e. TASK NUMBER
22

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Geospatial Research Laboratory
U.S. Army Engineer Research and Development Center
7701 Telegraph Road
Alexandria, VA 22315-3864

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDC/GRL TR-20-2

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Corps of Engineers
Washington, DC 20314-1000

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of this technical report is to present results of an investigation of the spatial distribution of the deer tick, *Ixodes scapularis,* and the three parasites it carries that cause serious diseases (Lyme disease, anaplasmosis, and babesiosis) in humans. The study used the maximum entropy (MaxEnt) species niche modeling technique to produce maps predicting the probability of the presence of *Ixodes scapularis* in the eastern United States. The model makes predictions based on tick and disease surveillance data from the Army Public Health Center, and environmental data collected from satellite remote sensing platforms. Geospatial analysis was also used to locate patterns between the disease-causing parasites. The resulting prediction maps of deer tick location can be used to inform vector interception planning, which attempts to lower the risk of disease-carrying ticks from infecting humans. The maps comparing the spatial distribution of the diseases related to deer ticks can be used as a launch point for further public health study into the drivers behind parasite spread, or to direct treatment resources.

15. SUBJECT TERMS

Ixodes scapularis – Geographical distribution – Mathematical models
Habitat (Ecology) – Mathematical models
Pathogens
Lyme disease
Military bases

16. SECURITY CLASSIFICATION OF:

a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
Unclassified	Unclassified	Unclassified	SAR	38	**19b. TELEPHONE NUMBER *(include area code)***

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18