Modeling the Geographic Distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the Contiguous United States

Micah B. Hahn[1,2], **Catherine S. Jarnevich**[3], **Andrew J. Monaghan**[4], and **Rebecca J. Eisen**[1]

[1]Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, 3156 Rampart Rd., Fort Collins, CO 80521

[3]U.S. Geological Survey, 2150 Centre Avenue, Bldg C, Fort Collins, CO 80526

[4]National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307

Abstract

In addition to serving as vectors of several other human pathogens, the black-legged tick, Ixodes scapularis Say, and western black-legged tick, *Ixodes pacificus* Cooley and Kohls, are the primary vectors of the spirochete (*Borrelia burgdorferi*) that causes Lyme disease, the most common vector-borne disease in the United States. Over the past two decades, the geographic range of *I. pacificus* has changed modestly while, in contrast, the *I. scapularis* range has expanded substantially, which likely contributes to the concurrent expansion in the distribution of human Lyme disease cases in the Northeastern, North-Central and Mid-Atlantic states. Identifying counties that contain suitable habitat for these ticks that have not yet reported established vector populations can aid in targeting limited vector surveillance resources to areas where tick invasion and potential human risk are likely to occur. We used county-level vector distribution information and ensemble modeling to map the potential distribution of *I. scapularis* and *I. pacificus* in the contiguous United States as a function of climate, elevation, and forest cover. Results show that *I. pacificus* is currently present within much of the range classified by our model as suitable for establishment. In contrast, environmental conditions are suitable for *I. scapularis* to continue expanding its range into northwestern Minnesota, central and northern Michigan, within the Ohio River Valley, and inland from the southeastern and Gulf coasts. Overall, our ensemble models show suitable habitat for *I. scapularis* in 441 eastern counties and for *I. pacificus* in 11 western counties where surveillance records have not yet supported classification of the counties as established.

The black-legged and western black-legged ticks, Ixodes scapularis Say and *Ixodes pacificus* Cooley and Kohls, respectively (herein referred to as *Ixodes* spp.), are the primary vectors to

[2]Corresponding author, mhahn@cdc.gov.

humans of the bacterial causative agents of Lyme disease (*Borrelia burdorferi* sensu stricto), as well as of pathogens that cause other human diseases including Anaplasmosis, Babesiosis, and Powassan virus disease (Piesman and Gern 2004, Brown and Lane 2005, Ebel 2010). Lyme disease is the most commonly reported vector-borne disease in the United States with over 30,000 cases reported annually in recent years (Mead 2015, Nelson et al. 2015). Cases in the United States are focused in 14 high-incidence states in the Northeast and North-Central regions and in small numbers in the western United States. Since the late 1990s, the number of reported cases of Lyme disease in the United States has tripled (Mead 2015). Furthermore, within the North-Central and Northeastern foci, the geographic range of reported Lyme disease cases has expanded. For example, in the Northeast, the number of counties considered high risk for Lyme disease has increased by more than 320 percent since the mid-1990s (Kugeler et al. 2015). In Minnesota, the number of *I. scapularis*-borne disease cases expanded in distribution across the state and increased by 742 percent from 1996 through 2011 (Robinson et al. 2015).

Coinciding with the increasing geographic range over which Lyme disease cases have been reported during the previous two decades, the number of counties in which *I. scapularis* is considered to be established has increased in the Northeastern, Mid-Atlantic, and North Central United States, while the range of *I. pacificus* has remained relatively stable (Eisen et al. 2016). This suggests that the realized niche of *I. pacificus*, or the areas that the tick is currently found, has nearly reached the extent of its fundamental niche, or the regions of the country where the tick can hypothetically survive given local environmental and climatic conditions. Alternatively, the fundamental niche may be much larger than the realized niche, but substantial barriers to migration (e.g., mountain ranges and vast deserts) or biotic factors such as lack of hosts have slowed expansion of the tick's range. Similarly, competition with established *I. scapu-laris* populations in the eastern United States may have prevented *I. pacificus* from becoming established outside of the West. The continued range expansion of *I. scapularis* suggests areas likely exist in the United States where this species can survive and reproduce but where established populations have not been reported. The goal of this study is to explore the degree to which the realized niches of *I. scapularis* and *I. pacificus* overlap with their modeled suitable habitat. This information would allow identification of counties where enhanced vector surveillance might be needed, for example in counties classified by our model as suitable for establishment but where vector populations have not yet been documented. Such areas may represent the leading edge of range expansion or where the tick is already established but surveillance activities are lacking. Habitat suitability models can aid in identifying whether and to what extent these medically important ticks are likely to continue expanding their ranges.

Others have developed habitat suitability models for *I. scapularis* (Estrada-Peña 2002, Brownstein et al. 2003) using similar county-level data on the distribution of *I. scapularis* compiled nearly two decades ago (Dennis et al. 1998). Each of these modeling efforts predicted the potential for *I. scapularis* range expansion to some degree. An updated survey of the tick's distribution (Eisen et al. 2016) supports the earlier model predictions in some areas, but also reveals establishment in areas not predicted to be suitable by the models. Habitat suitability models based on the updated distribution records will likely differ substantially from previous models because the current distribution of these ticks indicates

that they can survive under a broader range of climatic conditions than was captured using the geographical distributions of nearly two decades ago. Here we utilize updated *Ixodes* spp. distribution data (Eisen et al. 2016) and statistical ensemble modeling (Araujo and New 2007) to map the potential distribution of the tick vectors of Lyme disease spirochetes and other human pathogens in the United States. An ensemble modeling approach is used to address uncertainty in individual modeling algorithms (Buisson et al. 2010, Springer et al. 2015).

Materials and Methods

Tick Distribution Data

We used published data on the reported distribution of *I. scapularis* and *I. pacificus* by county in the United States as the basis of our modeling. Eisen et al. (2016) recently updated these data using literature searches, state health department data, and personal communications with tick and Lyme disease researchers throughout the United States. The county status in the database is defined using the definitions presented by Dennis et al. (1998). A county was classified as "established" if at least six ticks, or two or more life stages, were collected in a single year within the county. Counties were classified as "reported" if the specimen collections did not reach these thresholds, or if the number of ticks collected was not specified. All other counties lacked collection records for these tick species and were classified as "no records."

Climate, Elevation, and Land Cover Data Sources and Variable Selection

We selected a variety of environmental predictor candidates for our distribution models based on previous research on the biology and ecological requirements of *Ixodes* spp. Derivation methods and data sources for the candidate variables have been described previously in Springer et al. (2015) and are briefly summarized below and in the Supp. File (online only).

We used the 19 bioclimatic variables (Nix 1986) from WorldClim (version 1.4) at 2.5 arc minute resolution (roughly 5 km). WorldClim is a set of global climate layers that represent average conditions between 1950 and 2000 (Hijmans et al. 2005). We also used two estimates of growing degree days (GDDs), which is a measure of cumulative heat over a baseline temperature. We included the mean number of GDDs > 10°C for each month as well as cumulative GDDs > 10°C from the start of the year (Eisen et al. 2006, Moore et al. 2014). Because GDDs are not a direct output of WorldClim, they were estimated using WorldClim temperature data and calibrated using daily data from Daymet (Thorton et al. 2012), as described in Springer et al. (2015). We also tested monthly average values of vapor pressure (a measure of humidity) and the average number of days per month with snow cover (based on values of snow water equivalent >0mm) using data for the period 1980-2000 from Daymet. We obtained elevation data from the U.S. Geological Survey GTOPO30 digital elevation model (1996). Finally, land cover data were obtained from the USGS 2011 National Land Cover Database at 30-m resolution (Homer et al. 2015). The percent forest cover in a county was calculated by summing the area of the pixels in the deciduous, evergreen, and mixed forest classes and dividing by the county area.

Because the tick surveillance data were at the county level, we used the Zonal Statistics tool in ArcGIS version 10.2 (Environmental Systems Resource Institute; ESRI, Redlands, CA) to calculate a county-level mean value for each of the climate and elevation variables. Means for each county were calculated using data from grid cells whose centroid fell inside the county boundary.

We evaluated collinearity between predictors by generating a matrix that listed the largest value from among three correlation coefficients (Pearson, Spearman, and Kendall) calculated for each pair of variables. Then we ranked the variables in descending order by their values of deviance explained, or the amount of variation explained in a univariate model for a given predictor. We used three methods to choose among correlated variables and limited the analysis to variables with pairwise correlations <0.80. For the first method, we used expert knowledge on tick ecological requirements and deviance explained to narrow the predictor list. For the second method, we dropped variables that explained <5% deviance and then, from the ordered list of remaining variables, we retained the one with the highest deviance explained from among correlated variables. Then we continued down the list, adding each successive variable to the candidate pool if it was not correlated with any of the predictors already selected. For the third method, we dropped all variables that explained <1% deviance and selected variables based on their deviance explained as described in the second method; however, we also categorized each variable as a temperature, precipitation, or humidity variable. If, when moving down the ranked predictor list, we encountered a variable in a climate category that was not represented in the candidate pool, we included it and dropped any correlated predictors that had already been selected if there were other selected predictors in the same climate category. These variables are later referenced as *I. scapularis* or *I. pacificus* predictor sets 1, 2, or 3, respectively. Unless otherwise noted, all statistical analyses were conducted using VisTrails Software for Assisted Habitat Modeling (SAHM; version 2.0) (Morisette et al. 2013).

Modeling *Ixodes* Species Distributions

We developed the habitat suitability models for *I. scapularis* and *I. pacificus* separately because their distributions do not overlap (Eisen et al. 2016). We limited the study extent for *I. scapularis* to the Midwestern and Eastern United States, using the western borders of North and South Dakota, Nebraska, Kansas, Oklahoma, and Texas as the boundary. We modeled *I. pacificus* in Washington, Oregon, California, Idaho, Nevada, Utah, and Arizona.

We modeled the distribution of suitable habitat for these tick species using five algorithms: 1) boosted regression tree (BRT), 2) generalized linear model (GLM), 3) multivariate adaptive regression spline (MARS), 4) maximum entropy (Maxent), and 5) random forest (RF) (Talbert and Talbert 2001). Consideration of multiple algorithms allowed us to evaluate potential biases of individual approaches and optimize model parameters for subsequent analyses. We ran this set of algorithms for the three predictor sets described above for a total of 15 models for each tick species (i.e., 5 algorithms for each of 3 predictor sets). Inputs to the models included county tick status as "present" (counties that have established *I. scapluaris* or *I. pacificus* populations) or "absent" (counties with reported tick populations or that have no reported tick surveillance data) as well as county-level values of each of the

climate, elevation, and land cover predictors. We chose not to include counties with "reported" tick populations in our presence points because it is possible that the few ticks collected in that county represent anomalous, imported ticks that will not survive to reproduce.

In addition to running the models using all of the training data, we used a 10-fold cross-validation method to generate performance statistics for our models. This process included dividing the data into 10 equal subsets and running the models 10 times, leaving out one subset each time (henceforth referred to as the "test" dataset). Each model run produced a continuous relative probability surface of suitable habitat within the study extent. For each run, the probability threshold that maximized the sensitivity and specificity of the results based on the true presence or absence of a tick species was used to convert the continuous habitat suitability score into a binary score that classified each county as either suitable (score = 1) or unsuitable (score=0) (Fielding and Bell 1997, Guisan et al. 2007). These binary results from the cross-validation model runs were then aggregated into several metrics that we used to assess model over-fitting and to optimize model parameters.

First we assessed the Receiver Operating Characteristic (ROC) curve and associated values of area under the curve (AUC) for training and testing runs. The ROC curve is a plot of the true positive rate against the false positive rate for different cutpoints of the continuous habitat suitability score. 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 (Fielding and Bell 1997). We assessed over-fitting by looking to see if the difference between the training AUC and mean of the testing AUC values exceeded 0.05 or if there were large differences between the training and testing sets in percent correctly classified, percent deviance explained, and sensitivity and specificity of training and testing models for each algorithm. Based on these assessments, we optimized the BRT model using a tree complexity of two, a learning rate of 0.005, and 5,000 trees (Elith et al. 2008, Springer et al. 2015). All other models used the default SAHM parameters.

After the models were optimized, we compared the performance metrics produced by the three predictor sets using each of the five algorithms to select one model for each algorithm. For example, we compared the AUC, sensitivity, specificity, and AIC for the three GLM models and selected the model with the best performance based on these statistics. We repeated this for each model algorithm to select the top five models for each tick species.

Visualization and Evaluation of Modeling Results

After selecting the top five models, we calculated the relative contribution of each climate or land use predictor to each of the models. The variables selected by each algorithm varied, so in order to compare predictors across models, we normalized the contribution values by converting them to percentages relative to all variables in an individual model. We also examined response curves of each selected predictor across the five models. These curves show the relationship between an environmental predictor and tick habitat suitability.

We completed model selection by identifying, among the top five optimized models, those that had average testing AUC >0.7, biologically realistic response curves based on expert knowledge, and sensitivity and specificity values >0.70. If any model did not meet these criteria, it was not included in the final model. We created an ensemble model from the models that met these criteria by summing their binary habitat suitability maps. Thus, each county in the ensemble model had a habitat suitability score indicating the number of individual habitat suitability models that classified it as suitable tick habitat. To evaluate the ensemble model predictions and identify areas of likely tick invasion, we compared the ensemble map to the tick surveillance records used to train the model (Eisen et al. 2016). We also created binary distribution maps from each of the ensemble members to evaluate differences in the geographic range predictions from each modeling algorithm. Finally, we identified counties where surveillance records indicate that *Ixodes* ticks are not established but at least two ensemble members predicted suitable habitat.

Results

Ixodes scapularis Variable Selection and Model Performance

Using the three variable selection methods described above, we reduced the original set of 68 environmental predictors to 11 (those listed in Table 1 and cumulative growing degree days in December, which subsequently was not retained by any model). Although we consistently selected many of the same predictors using the three selection methods, there were slight differences in the three groups of predictors presented to the model algorithms and at least one modeling algorithm performed best with each of the predictor sets (Table 2). The best performing BRT and GLM models used variables selected using expert opinion (predictor set 1). The best MARS model used the candidate variables selected based only on percent deviance explained (predictor set 2), and the Maxent model used variables selected with a combination of percent deviance explained and climate category (predictor set 3). We dropped the RF model from the results because there was evidence of substantial over-fitting of the training data using this algorithm. In particular, the training and testing sensitivity varied by almost 15 percentage points, while the difference in the other algorithms was 0 to 3 percentage points. When interpreting the results of the modeling algorithms, it is important to note that both the Maxent and RF models do not have an internal variable selection process and therefore retain all predictors presented to them as candidates. Predictor variables that do not explain a significant amount of variation in these models will have low normalized contribution values in these models.

Several of the predictors were retained by at least three of the four modeling algorithms (Table 1). Maximum temperature of the warmest month (Bio5) was retained by all of the models that used predictor sets 1 or 2 which included Bio5 as a candidate, and its normalized contribution was between 11 and 55 percent. Precipitation in the warmest quarter (Bio18) was also in three of four final models and had normalized contributions between 8 and 13 percent. Precipitation of the coldest quarter (Bio19) and precipitation of the driest quarter (Bio17) were highly correlated so were not included in the same model, but three of four models retained one of these precipitation variables and normalized contributions were between 5 and 16 percent. In addition to climate variables, percent forest

cover in a county was a strong predictor of suitable *I. scapularis* habitat. Forest cover was retained by three of four modeling algorithms and had normalized contributions between 14 and 26 percent. Elevation was the only predictor retained by all four optimized models and had normalized contributions between 21 and 44 percent. Among the four optimized models, nonclimatic predictors (percent forest cover and elevation) explained an average of 46 percent (range: 40 to 55 percent) of the variation in *I. scapularis* habitat suitability, and climate predictors explained an average of 54 percent of the variation (range: 45 to 59 percent).

All models had high accuracy with average test AUC values between 0.85 and 0.86 (Table 2). Across all models, ~ 77 percent of counties were correctly classified in the testing datasets. Sensitivity, or the percent of counties with known suitable habitat that were classified as such by the models, was 75 to 77 percent in the testing data. Most of the sensitivity loss was in the inland counties in the Southeast where 106 counties with established vector populations were predicted as unsuitable by three or four ensemble members (Fig. 1) . The specificity range, or the percent of counties without reported established tick populations that were classified as such by the models, was 77 to 78 percent. Performance metrics were similar for training and testing data across all models indicating that the models were not overfit to the training data and performed well on testing datasets.

Ixodes scapularis Response Curves

Response curves of the climate predictors show a consistent relationship between maximum temperature in the warmest month (Bio5) and relative probability of *I. scapularis* occurrence with the least suitable habitat in areas experiencing summer temperatures between 28 and 35°C (Fig. 2). Curves of precipitation of the warmest quarter (Bio18) show a nearly linear relationship between increasing suitability and increasing precipitation. Precipitation of the coldest quarter (Bio19) and precipitation of the driest quarter (Bio17) show similar relationships and indicate the most suitable habitat occurs in areas that receive around 200 mm of precipitation during these time periods. Although only selected by one model, the curve of the mean temperature of the coldest quarter (Bio11) shows high habitat suitability at temperatures less than 0°C. There was a strong positive linear association between vapor pressure in February and habitat suitability, while the association between average days per month with snow cover and habitat suitability was weakly negative. Mean diurnal temperature range (Bio2) was the only predictor with conflicting response curves. Both the GLM and Maxent models showed decreasing suitability with increasing mean diurnal range, while the MARS model showed a threshold effect where habitat suitability increased in areas with high variability in mean diurnal range. The normalized contribution of this variable was comparatively small (0.6–3.7%). The curves of percent forest cover showed increasing habitat suitability from zero to ~75 percent forest cover in a county and then decreasing suitability in areas of higher forest cover. Finally, the elevation curves show the most suitable habitat at elevations below 150–200 m, and low suitability at elevations higher than 500 m.

Ixodes scapularis Ensemble Model

The area predicted by the ensemble distribution model as suitable *I. scapularis* habitat largely followed the known distribution of the tick represented by the county surveillance records compiled by Eisen, Eisen, and Beard (2016); however, there were notable areas of discordance between the distribution model and known *I. scapularis* occurrence (Fig. 3). The ensemble model predicted suitable habitat restricted to a coastal ring around the southeastern United States. As a result, it appears to underpredict suitability along the Appalachian Mountains extending through Tennessee and northern Alabama where surveillance data show *I. scapularis* to be either reported or established (Fig. 1). Additionally, only one or two of the distribution models predicted suitable tick habitat throughout Arkansas and southern Missouri where the tick surveillance data demonstrated *I. scapularis* occurrence. The ensemble model also predicted several areas of suitable tick habitat where *I. scapularis* has not yet been identified (Fig. 3). For example, the BRT model predicted suitable *I. scapularis* habitat extending into North Dakota where current surveillance information shows patchy occurrence. Similarly, *I. scapularis* has currently only been reported on the western edges of Michigan's lower and upper peninsulas, while the distribution models show strong agreement that suitable habitat is found throughout most of the state. Finally, the ensemble model shows contiguous suitable *I. scapularis* habitat within the Ohio River Valley, another area where the tick has only been found sporadically.

Ixodes pacificus Variable Selection and Model Performance

The second and third variable selection methods produced the same predictor sets for the *I. pacificus* models, so we only present modeling results using expert opinion and percent deviance explained (predictor sets 1 and 2, respectively). Using these two variable selection methods described above, we reduced the original set of 68 environmental predictors to 17 (those listed in Table 3 and temperature seasonality (Bio4), precipitation of driest quarter (Bio17), vapor pressure in October, and cumulative growing degree days in February, which were not retained by any model). The best performing GLM, Maxent, and RF models used variables selected using expert opinion (predictor set 1, Table 4). The best MARS model used the candidate variables selected based only on percent deviance explained (predictor set 2). We dropped the BRT model from the results because there was evidence of substantial overfitting of the training data using this algorithm. In particular, the training and testing sensitivity varied by 10 percentage points while the difference in the remaining algorithms was 0 to 5 percentage points.

Several of the predictors were retained by three or four modeling algorithms (Table 3). Precipitation of the coldest quarter (Bio19) was retained by all four algorithms, and its normalized contribution was between 3 and 80 percent. Percent forest cover was also retained by all four modeling algorithms and had normalized contributions between 0.1 and 31 percent. Precipitation of the warmest quarter (Bio18) and mean temperature of the wettest quarter (Bio8) were retained by three of four modeling algorithms and had normalized contributions between 6 and 18 percent and 8 and 28 percent, respectively. The cumulative number of growing degree days in December was also retained by three modeling algorithms, but the normalized contribution values were less than 8 percent. Minimum temperature of the coldest month (Bio6) was retained by the Maxent and RF

models, and the normalized contribution values were 26 and 8 percent, respectively. Isothermality (Bio3), maximum temperature of the warmest month (Bio5), and vapor pressure in July were retained by the Maxent and RF models as well, but their normalized contribution values were less than 4 percent. Precipitation seasonality (Bio15) and annual mean temperature (Bio1) had high values of deviance explained in univariate models, but they were dropped as predictor candidates in the expert opinion model (predictor set 1) because they were highly correlated with other predictors selected to represent temperature extremes such as the minimum temperature of the coldest month (Bio6). The MARS algorithm performed best with the deviance explained predictor set (predictor set 2), which included Bio15 and Bio1. These variables were retained in the final model and had normalized contribution values of 47 and 8 percent, respectively. Among the four optimized models, land cover predictors (percent forest cover) explained 12 percent (range: 0.1 to 31 percent) of the variation in *I. pacificus* habitat suitability on average, and climate predictors explained ~88 percent of the variation (range: *69* to 100 percent).

The *I. pacificus* models had higher accuracy than the *I. scapula-ris* models with average test AUC values between 94 and 95 percent (Table 4). Across all models, 88 to 93 percent of counties were correctly classified in the testing datasets. Sensitivity was 89 to 91 percent. Specificity was 89 to 95 percent. Performance metrics were similar for training and testing data across all models indicating that the models were not over-fit to the training data and performed well on testing datasets.

Ixodes pacificus Response Curves

The response curve of precipitation seasonality (Bio15) shows that *I. pacificus* can survive in areas with substantial wet and dry seasons, although this is likely a proxy for other highly correlated cli-matological variables (Fig. 4). High precipitation in the warmest (Bio18) or coldest (Bio19) quarter can reduce *I. pacificus* habitat suitability, but between 300 and 600mm of precipitation in the coldest quarter may be conducive for the ticks. Curves of minimum temperature of the coldest month (Bio6) show decreasing suitability in areas that experience extreme cold temperatures below ~−5°C. Other temperature predictors such as annual mean temperature (Bio1), cumulative growing degree days in December, and maximum temperature of the warmest month (Bio5), show similar relationships and indicate that warmer temperatures provide more suitable *I. pacificus* habitat. However, response curves for the mean temperature of the wettest (Bio8) and driest (Bio9) quarters show that average temperatures greater than 20°C tend to reduce habitat suitability in either season. This maximum temperature threshold is ~8°C lower than the maximum temperature for *I. scapularis* in the eastern United States, perhaps due to the overall drier climate in the West. The two response curves for isothermality (Bio3) show conflicting results with the RF model showing increasing suitability with increased variation in the daily temperatures relative to the annual temperature fluctuation while the Maxent models shows a negative relationship; however, the normalized contribution of Bio3 in both models is <1%. The response curve for the annual diurnal temperature range (Bio2), or the annual mean of the monthly temperature range, shows a negative relationship with *I. pacificus* habitat suitability indicating that the ticks are more likely found in areas with smaller range in monthly temperature extremes. Finally, the response curves for percent forest cover show increased suitability in areas with

more forest cover, and vapor pressure in July (a measure of humidity) also shows a positive relationship with *I. pacificus* habitat suitability.

Ixodes pacificus Ensemble Model

The *I. pacificus* ensemble members show strong agreement in predicting suitable habitat across California and the west coasts of Oregon and Washington (Fig. 5). This largely follows the distribution of known *I. pacificus* populations from surveillance data (Eisen et al. 2016), although only the RF model shows suitable habitat in northwestern Utah where established *I. pacificus* populations have been reported. The ensemble model also shows suitable habitat in western Washington as well as potential habitat expanding along the Oregon-Washington border. Surveillance data show reported, but not established, populations in western Washington and no recorded occurrences along the eastern half of the Oregon-Washington border.

Expansion of *Ixodes* spp. Distribution

Taken together, the *I. scapularis* and *I. pacificus* ensemble models show that there are considerable areas of the United States that could provide suitable habitat for *Ixodes* spp. ticks but where surveillance records indicate that these species are not yet established (Fig. 1). The ensemble model in the eastern United States shows suitable *I. scapularis* habitat in 441 additional counties where the tick is not yet classified as established, representing a potential 52% increase in the number of counties with established populations. If realized, this increase could be attributed to the initiation of enhanced surveillance that recognizes currently established populations or further expansion of the tick's range. In the North-Central states, the models predict *I. scapularis* expansion into northwestern Minnesota, complete coverage of Wisconsin, and substantial spread into central and northern Michigan as well as the Upper Peninsula. The models also show potential movement of *I. scapularis* within the Ohio River Valley, connecting the previously distinct North-Central and Northeastern populations. In the Southeast, the models show suitable habitat along the coasts and inland from the coasts, filling in gaps in regions where surveillance data show only sporadic records of counties where *I. scapularis* is established. In the western United States, *I. pacificus* has been found in almost every county along the Pacific coast. The ensemble model shows suitable *I. pacif-icus* habitat in 11 additional counties where the tick is not yet established, representing a potential 12% increase in the number of counties with established populations. Most of this suitable habitat lacking establishment records is in the Pacific Northwest (Fig. 1).

Discussion

We used an ensemble modeling approach to predict areas of suitable habitat for *I. scapularis* and *I. pacificus* in the contiguous United States. Each of the four statistical models comprising the ensembles retained different predictors, limiting our ability to identify the key factors limiting the distributions of the two ticks. Nonetheless, the *I. scapularis* ensemble members consistently predicted suitable habitat in the North-Central and Northeastern United States as well as coastal counties along the Atlantic and Gulf coasts. They also consistently predicted unsuitable habitat in Tennessee, northern Georgia, and northern

Alabama. Similarly, the *I. pacificus* models agreed in predicting suitable habitat along much of the Pacific coast and unsuitable habitat in alpine and desert counties. Overall, our ensemble models showed suitable habitat for *I. scapularis* in 441 eastern counties and for *I. pacificus* in 11 western counties where surveillance records have not yet supported classification of the counties as established. Based on these findings, continued range expansion is expected for *I. scapularis*, particularly in the North-Central states.

Predicted Distribution of Suitable Habitat for *Ixodes scapularis*

There have been two previous efforts to model the nation-wide distribution of *I. scapularis*, both of which used climate and land use predictors and modeling algorithms similar to those used in the present study (Estrada-Peña 2002, Brownstein et al. 2003). Similar to the results of our ensemble model, both previous studies characterized the two primary *I. scapularis* foci in the Northeastern states and to a lesser extent in the North-Central states as suitable habitat. Nonetheless, there were notable differences in the predicted distributions when compared with each other and with the predicted distribution from this study. For example, using county-level records of the distribution of *I. scapularis* (Dennis et al. 1998) similar to the data used in our models, Brownstein et al. (2003) used a logistic regression modeling framework to predict the spatial distribution of the tick in the contiguous United States. Model predictions were strongly driven by several temperature variables as well as vapor pressure. The model correctly predicted the eventual expansion of established *I. scapularis* populations throughout the Northeast and in eastern North Carolina in areas where the tick was only reported at the time. The model predicted several regions of the eastern United States, including northern Minnesota, western Michigan, and the Ohio River Valley, to be unsuitable or very marginally suitable for *I. scapularis*. Updated tick distribution records indicate establishment throughout these regions. Consistent with our model, Brownstein et al. (2003) showed unsuitability or low suitability for inland counties in the Southeast. Using the same county-level records (Dennis et al. 1998) to train a discriminant analysis model of *I. scapularis* in the eastern United States, Estrada-Peña (2002) predicted a near complete expansion of *I. scapularis* in the eastern United States with gaps in the distribution in western Minnesota, Michigan, and much of Indiana, Illinois, and Ohio. The primary variables defining suitability included winter temperature and vegetation vitality.

Our current study provides an update to these modeling efforts using recent tick surveillance data that document a much expanded distribution and broader climatic envelope of *Ixodes* spp., particularly in the eastern United States. For example, expansion of the tick into Minnesota and the Northeast shows that *I. scapularis* can survive in a broader range of climatic conditions than would have been thought based on the distribution from Dennis et al. (1998). The average temperature of the coldest quarter (Bio11) in counties that had established *I. scapularis* populations in 1998 was 2°C (Dennis et al. 1998). Two decades later, updated surveillance records show that the average temperature during the coldest quarter in counties with established tick populations is lower, 0.25°C (Eisen et al. 2016). This change in habitat range likely accounts for some of the discrepancy between the habitat distribution results from Brownstein et al. (2003) and Estrada-Pena (2002) and this study.

In addition, the ensemble modeling approach used in this study can provide more accurate predictions than a single model when there is a high degree of uncertainty in the system being modeled (Buisson et al. 2010). Several modeling frameworks have been used to model species distributions with varying degrees of accuracy (Elith 2002, Segurado and Araujo 2004, Elith et al. 2006) and model performance in part depends on if the method used by the algorithm to identify relationships between the species locations and environmental gradients corresponds to the species' empirical presence or absence response to environmental variables (Segurado and Arauijo 2004). For example, the distribution of a species that responds strongly to an environmental gradient, perhaps that requires a minimum temperature threshold for survival, could likely be captured using a regression technique such as GLM or MARS (Segurado and Araujo 2004). In contrast, a species with a complex distribution pattern that does not respond to clear environmental gradients might be more accurately modeled using machine learning techniques such as BRTs, RF, or Maxent (Segurado and Araujo 2004). A third class of models that relies only on presence locations such as the DOMAIN algorithm used in Estrada-Peña (2002) tends to overestimate species presence and is often out-performed by modeling algorithms that characterize the background environment during model training either with absence or pseudo-absence data (Engler et al. 2004, Elith et al. 2006). There are known temperature and humidity thresholds for *Ixodes* spp. survival that are discussed below. But also, distribution records of *I. scapularis* and *I. pacificus* (Eisen et al. 2016) have shown that these ticks are able to survive in a wide variety of climates, perhaps due to their ability to find suitable microclimates in otherwise inhospitable areas (Bertrand and Wilson 1996, Vail and Smith 1998). As a result, their biological response to environmental gradients may sometimes be captured with a simple function, while in other cases, their response to broad-scale climatic predictors may require a more flexible model. Our ensemble modeling approach incorporated both classes of modeling algorithms and likely captured more of this variation in tick biology than a single model approach.

The *I. scapularis* ensemble in the current study predicted no suitable habitat in Tennessee, or northern portions of neighboring states (Mississippi, Georgia, and Alabama) despite reports of established tick populations in these areas (Rosen et al. 2012, Goltz and Goddard 2013, Goltz et al. 2013). Estrada-Pena (2002) and Brownstein et al. (2003) showed low to moderate habitat suitability for *I. scapularis* in this region, although the predicted distribution from these studies may have been less extensive if the continuous probability values were dichotomized following the methods used in the present study. Notably, several of the climate variables used in our habitat suitability models distinguish this area from the coastal counties where the model predicts suitable habitat (Fig. 2 and Supp. Fig. 1 [online only]). The models showed that suitability is lowest when the maximum temperature of the warmest month (Bio5) is between 30-33°C (Ogden et al. 2004, Eisen et al. 2015). Such a temperature range is typical for inland counties in the Southeast during the warmest month of the year. Similarly, the model that employed temperatures in the coldest quarter as a predictor showed decreased suitability between temperatures 5–12°C, also characteristic for the region predicted as unsuitable. Several other climatic factors associated with low suitability in the models characterize the inland counties in the Southeast that were classified as unsuitable. These included low February vapor pressure, between 250–350mm of

precipitation in the warmest quarter (Bio18), and between 350–450 mm of precipitation during the coldest quarter (Bio19). In addition, high elevation along the southern portion of the Appalachian mountain range along the Tennessee-North Carolina border into northern Georgia and the low forest cover along the Mississippi River Valley may account for the lack of suitable habitat in these areas (Supp. Fig. 1 [online only]).

These climatic conditions and land cover characteristics that separate inland counties in the Southeast from other counties within the *I. scapularis* range likely explain why the area was classified as unsuitable. However, contradictory to model predictions, surveillance records indicate that *I. scapularis* is established in some of these counties (Rosen et al. 2012, Goltz and Goddard 2013, Goltz et al. 2013, Eisen et al. 2016), demonstrating that the tick can survive and reproduce under the climatic conditions representative of much of this region and highlight an area where model sensitivity is relatively low (a high false negative rate). It is likely that because of the relatively low number of records from inland counties in the Southeast, the climatic conditions of that region were underrepresented in the model of the eastern United States. We attempted to test the impact of the lack of surveillance records in this area by training a regional model that included Tennessee, Kentucky, Illinois, Missouri, Arkansas, northern Mississippi, Alabama, and Georgia, western North and South Carolina, and southwestern Virginia using the limited presence locations from counties in this region. Attempts to train the regional model yielded low accuracy, likely due to the paucity of presence records. Overall, this is a region where additional surveillance and fine-scale modeling studies are needed to elucidate the true distribution of the tick and the variables that define the habitat-climate suitability envelope.

A range of variables capturing warm- and cold-season temperatures and moisture was retained in the *I. scapularis* models. Across models, the most suitable habitat was found in forested areas below 500m where summer temperatures are generally below 25–32°C and moisture throughout the year is sufficient to reduce the likelihood of desiccation-induced mortality. Two of three models that retained maximum temperature of the warmest month (Bio5) showed declining suitability as temperature increased from ~25–32°C. These findings are consistent with laboratory studies that showed that host-seeking activity of *I. scapularis* nymphs peaked at 25°C and fell after 30°C (Vail and Smith 2002). Typically, at very high temperatures, ticks are vulnerable to desiccation-induced mortality (Eisen et al. 2015). There is evidence for this to occur at temperatures as low as 32°C for *I. ricinus* but generally occurs above 40°C for other ixodid ticks (Balashov 1971, Sonenshine and Roe 2013). Furthermore, ticks are less likely to seek hosts when temperatures are high (Vail and Smith 1998), which may increase tick mortality rates by reducing host-finding success (Randolph and Storey 1999). Above 32°C, one ensemble member showed a continued decline in habitat suitability while the other two models showed an increase in suitability. The models showing increasing suitability at high temperatures are influenced by the presence of ticks in counties along the Gulf coast where precipitation in the warmest quarter helps to overcome the risk of heat-induced desiccation (Supp. Fig. 1 [online only]).

There was a consistent positive relationship between precipitation in the warmest quarter and habitat suitability (Bio18). Precipitation is generally positively correlated with humidity, which in turn is important for tick survival (Needham and Teel 1991, Stafford 1994, Vail and

Smith 1998, Eisen et al. 2003). Ticks are highly sensitive to desiccation and need humidity when they are questing on the upper part of vegetation or a moist refuge near the soil (Lees and Milne 1951, Randolph and Storey 1999). Experiments have shown that low humidity can force questing ticks to return more frequently to the leaf litter to rehydrate, depleting their energy and decreasing their ability to find a host (Lees and Milne 1951, Eisen et al. 2015).

All three *I. scapularis* ensemble members that retained percent forest cover as a predictor showed increasing habitat suitability as the percent of a county classified as forest increased, up to about 60–75 percent coverage. A review of studies on the spatial distribution of Lyme disease and the vector ticks found the only environmental variable consistently associated with increased Lyme disease risk was presence of forests (Killilea et al. 2008). In addition to providing leaf litter, shade, and humidity for ticks (Stafford and Magnarelli 1993), wooded areas may support deer populations or the ecotone areas that separate the woods and lawns or public green space may be important habitat for small mammal hosts, both of which could support higher tick populations (Wilson et al. 1985, 1990; Ostfeld et al. 1995). The perhaps counterintuitive finding that habitat suitability declines when forest cover is greater than 75 percent may be explained by the observation that counties with >75 percent forest cover in our study were located predominately in the Appalachian Mountain range at elevations over 500 m where *I. scapularis* populations have not been reported.

Finally, our models consistently showed that the most suitable *I. scapularis* habitat was found at sea level, and then suitability declined to zero at an elevation of around 800 m in the Maxent model and at ~500m in the other three models. Other studies have found decreasing tick density with increasing elevation in Europe and North America (Jouda et al. 2004, Diuk-Wasser et al. 2010). Specifically, Diuk-Wasser et al. (2010) found no *I. scapularis* nymphs above 510 m in their field study across the eastern United States, perhaps due to correlation with unmeasured variables associated with an altitudinal gradient such as minimum temperature.

Predicted Distribution of Suitable Habitat for *Ixodes pacificus*

Habitat suitability models for *I. pacificus* across the western United States are lacking. Overall, our ensemble members showed good agreement in the expected distribution of suitable habitat along the Pacific coast. Although other variables, including percent forest cover and temperature during key periods contributed to describing suitable habitat for *I. pacificus,* seasonal variability in precipitation contributed strongly and fairly consistently across models, albeit captured by different variables in the various ensemble members. For example, precipitation seasonality (Bio15) was only retained by one ensemble member (MARS) where it explained almost half of the variation in *I. pacificus* habitat suitability. Precipitation seasonality is a measure of month-to-month precipitation variability. Areas that have distinct wet and dry seasons, like much of the western coast of the United States, have high precipitation seasonality. Our response curves indicate that areas with a coefficient of variation >40 which includes all of California, southern Arizona, and the western portions of Washington and Oregon (Supp. Fig. 2 [online only]) are highly suitable for *I. pacificus.* The other three remaining models captured the variability in precipitation using a combination of

precipitation during the coldest (Bio19) and warmest (Bio18) quarters. Across models, these variables combined explained between 41– 86% of variation in suitability. In the expert opinion predictor set, precipitation seasonality (Bio15) was not selected as a predictor because the timing of precipitation (Bio18 and Bio19) was more directly linked to the life cycle of the tick than month to month variability. Although there was disagreement among models about the impact of more than 500 mm of precipitation during the coldest quarter (Bio19) on *I. pacificus* habitat suitability, all models agreed that suitability increases as precipitation increases from 200 to 500 mm, which corresponds to much of central Washington, Oregon, and California (Supp. Fig. 2 [online only]). This finding may be related to larval or nymphal survival over winter, or winter precipitation may be important for *Ixodes* adults questing during winter months (Carroll and Kramer 2003, Dautel et al. 2008). At higher elevation sites that typically experience very low minimum temperature and consistent snow pack over the winter months, precipitation in the form of snow may provide additional insulation to enhance survival during the extreme cold (Templer et al. 2012). This insulating snow may be particularly important in areas above tree line where the leaf litter does not maintain a viable soil temperature (Burtis et al. 2016). Areas that are dry during the warmest quarter (Bio18) were more suitable for *I. pacificus* than areas that receive more than 100 mm of precipitation during these months. Most of the western United States falls into this drier category except for the western coast of Washington, the northwest coast of Oregon, northern Idaho, and most of Arizona (Supp. Fig. 2 [online only]). Field and laboratory work to elucidate the life cycle of *I. pacificus* living in northern California has shown that the tick has adapted to the dry summer climate in the western United States (Padgett and Lane 2001). In particular, larvae and nymphs in this area quest in the spring to avoid hot and dry conditions during the summer (Padgett and Lane 2001). Field sampling at sites in central and southern California, an area that is hotter and drier than the rest of the state, showed that all *I. pacificus* life stages have a relatively truncated questing season compared to ticks collected in northern California (MacDonald and Briggs 2015).

The relationship between increasing percent forest and habitat suitability that was found for *I. scapularis* in the eastern United States was found for *I. pacificus* in the West. Unlike the *I. scapularis* model, there was no decline in habitat suitability at high percent forest cover for *I. pacificus,* but most counties in the western United States had less than 75 percent forest cover. As mentioned above, the leaf litter may support overwintering *I. pacificus* by shielding the ticks from prolonged cold temperatures (Burtis et al. 2016), may provide a refuge of cooler temperatures and increased humidity during hot summer months (Schulze et al. 1995, Vail and Smith 2002), or could provide habitat for vertebrate hosts (Eisen et al. 2006).

Three of four models showed that an average temperature around 10°C during the wettest quarter (Bio8) is ideal for *I. pacificus* habitat but that suitable habitat is found in areas with temperature ranges between 0 and 20°C during this time period. This climate predictor therefore excludes western Washington and Oregon, most of Idaho, and southern Arizona as suitable *I. pacificus* habitat (Supp. Fig. 2 [online only]). For most of the West coast, the wettest quarter of the year falls between October and January, and having mild temperatures during the cold season may improve overwintering success of immature life stages and could

facilitate survival of host-seeing adults that are active mainly from November to May (Lane 1990, Lane et al. 1991).

Limitations and Future Directions

The *I. pacificus* models had higher accuracy, sensitivity, and specificity than the *I. scapularis* models, likely in part because *I. pacificus* occupies nearly its full fundamental niche. The geographic distribution of *I. pacificus* has remained stable over the last two decades (Dennis et al. 1998), and the habitat model closely follows this known distribution, suggesting that there are few counties where we might expect to see expansion of this tick species. *I. pacificus* would have to overcome significant physical and climatological barriers to continue an eastward expansion over the Sierra Mountains and into the deserts of Nevada and Arizona.

In contrast to *I. pacificus*, the distribution of *I. scapularis* has expanded substantially over the same time period (Eisen et al. 2016), and will likely continue to expand as predicted by our ensemble model. Specifically, we expect expansion into northwestern Minnesota, central and northern Michigan, within the Ohio River Valley, and inland from the southeastern and Gulf coasts. As a result, the fundamental niche of *I. scapularis* is only partially occupied, with an estimated 441 counties considered suitable but currently lacking records of established populations. As a result, sensitivity of the model is compromised (i.e., the tick is absent from counties where climatic conditions and land cover are expected to be suitable). Another reason for low sensitivity in the *I. scapularis* model, particularly in the inland counties in the Southeast, is that the lack of routine and systematic tick surveillance throughout the United States means that some areas where *I. scapularis* ticks are already established may not have been included in our models as presence points. For example, tick surveillance records from Tennessee were collected from a convenience sample of hunter-killed deer which likely do not capture the extent of the tick distribution in that region (Rosen et al. 2012, Eisen et al. 2016). Additionally, in the South, *I. scapularis* larvae and nymphs are rarely collected by drag cloths so established tick populations may be missed even in areas with systematic drag surveillance (Goddard and Piesman 2006). If any of these missing counties has climatic features that are unique compared to the presence points included in our model, then our predictions would leave out potentially suitable habitat within the same climate regime.

We used average climate, elevation, and land cover values for each county as predictors in our habitat models. Although averaging the pixel values across a county provides a more representative estimate of a variable than using a single value for each county, some counties may cover a substantial gradient with regard to these environmental variables. For example, there are counties along the Appalachian Mountain range that include areas with elevations less than 500m to nearly 2,000 m. If a county contains unsuitable and suitable habitat, models based on the average climatic conditions may identify the county as unsuitable despite the availability of tick habitat at lower elevations. Subcounty models for these ecologically diverse regions may improve the accuracy of habitat suitability predictions. Additionally, the baseline climate data we used may have impacted model calibration as the locations records extended beyond the year 2000.

Although our models have defined the potential distribution of *Ixodes* spp. with reasonable accuracy based on current vector surveillance data, the expansion of tick populations into new areas depends on either dispersal of a gravid female, male and female adults, or several immature ticks that survive until reproduction (Springer et al. 2015). Ticks are generally dispersed by the movement of hosts (Madhav et al. 2004, Khatchikian and Prusinski 2012) and therefore, *Ixodes* spp. range expansion will also be affected by available host habitat or host movement, which was not explicitly included in our models. Although the expansion in the distribution of *I. scapula-ris* ticks is driven in large part by the movement of white-tailed deer (Rand et al. 2003), the primary host of the adult life stage of the tick, others have pointed to the potential role of migratory birds in *B. burgdorferi* transmission dynamics through the introduction of infected larvae and nymphs to areas without established *I. scapularis* populations or with only pathogenically naïve tick populations (Scott et al. 2001, Brinkerhoff et al. 2011, Newman et al. 2015). Future tick surveillance and modeling efforts could target riparian corridors or known migration routes that could be a factor in the long-range dispersal of ticks (Weisbrod and Johnson 1989).

Vector surveillance coupled with habitat modeling can provide a useful public health tool for detecting new areas of tick invasion and potential human risk (Koffi et al. 2012). This study identified several areas that could provide suitable habitat for medically important ticks but where surveillance records indicate that these tick species are not yet established. These areas in northwestern Minnesota, central and northern Michigan, within the Ohio River Valley, along the southeastern and Gulf coasts, and in eastern Washington are regions where enhanced tick surveillance could verify the presence of *Ixodes* spp. which could serve as an early indicator of the potential for Lyme disease risk (Mather et al. 1996, Pepin et al. 2012). However, follow up studies would then be needed to assess the density of host-seeking infected ticks in these areas. We also showed the limitations of this approach when there is no routine or systematic vector surveillance, in particular, this type of modeling is likely to have low sensitivity in areas with limited surveillance records. Additionally, because we utilized surveillance data that included ticks from all clades and life stages and that did not incorporate information on the presence of *B. burgdorferi* infection, the distribution map from the present study is not predictive of human risk. For example, others have noted the expansion of human Lyme disease cases in southwestern Virginia, an area where our *I. scapularis* habitat models show disagreement on the suitability of the region for establishment of tick populations (Brinkerhoff et al. 2014, Lantos et al. 2015). Similarly, in the Southeast where *I. scapularis* is established, Lyme disease cases are rare, which may be attributable to differences in host-seeking behavior between northern and southern clades of the tick, and/or low infection rates in ticks (Stromdahl and Hickling 2012, Arsnoe et al. 2015). To more closely approximate human risk for exposure to ticks infected with human pathogens, we recognize a need for modeling the density of host-seeking infected nymphs, similar to work by Diuk-Wasser et al. (2012), rather than simply assessing the establishment of tick populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Araujo M, New M. Ensemble forecasting of species distributions. Trends Ecol Evol. 2007; 22:42–47. [PubMed: 17011070]

Arsnoe IM, Hickling GJ, Ginsberg HS, McElreath R, Tsao JI. Different populations of blacklegged tick nymphs exhibit differences in questing behavior that have implications for human Lyme disease risk. PLoS ONE. 2015; 10:e0127450. [PubMed: 25996603]

Balashov, Y. Blood sucking ticks (*Ixodoidea*): Vectors of diseases of man and animals. University of Michigan, MI: 1971.

Bertrand MR, Wilson ML. Microclimate-dependent survival of unfed adult *Ixodes scapularis* (Acari: Ixodidae) in nature: Life cycle and study design implications. Entomol Soc Am. 1996; 33:619–627.

Brinkerhoff RJ, Folsom-O'Keefe CM, Tsao K, Diuk-Wasser MA. Do birds affect Lyme disease risk? Range expansion of the vector-borne pathogen *Borrelia burgdorferi.* Front Ecol Environ. 2011; 9:103–110.

Brinkerhoff RJ, Gilliam WF, Gaines D. Lyme disease, Virginia, USA, 2000–2011. Emerg Infect Dis. 2014; 20:19–22.

Brown, R., Lane, R. Geographic distribution of tick-borne disease and their vectors, pp. 363–391. In: Goodman, J.Dennis, D., Sonenshine, D., editors. Tick-Borne Dis Humans. ASM Press; Washington, DC: 2005.

Brownstein JS, Holford TR, Fish D. A climate-based model predicts the spatial distribution of the Lyme disease vector Ixodes scapularis in the United States. Environ Health Perspect. 2003; 111:1152–1157. [PubMed: 12842766]

Buisson L, Thuiller W, Casajus N, Lek S, Grenouillet G. Uncertainty in ensemble forecasting of species distribution. Glob Chang Biol. 2010; 16:1145–1157.

Burtis JC, Ostfeld RS, Yavitt JB, Fahey TJ. The relationship between soil arthropods and the overwinter survival of *Ixodes scapularis* (Acari: Ixodidae) under manipulated snow cover. J Med Entomol. 2016; 53:225–229. [PubMed: 26487243]

Carroll JF, Kramer M. Winter activity of *Ixodes scapularis* (Acari: Ixodidae) and the operation of deer-targeted tick control devices in Maryland. J Med Entomol. 2003; 40:238–244. [PubMed: 12693854]

Dautel H, Dippel C, Ka D, Werkhausen A, Kahl O. Winter activity of *Ixodes ricinus* in a Berlin forest. J Med Microbiol. 2008; 298:50–54.

Dennis DT, Nekomoto TS, Victor JC, Paul WS, Piesman J. Reported distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the United States. J Med Entomol. 1998; 35:629–638. [PubMed: 9775584]

Diuk-Wasser MA, Vourc'h G, Cislo P, Hoen AG, Melton F, Hamer SA, Rowland M, Cortinas R, Hickling GJ, Tsao JI, et al. 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. Glob Ecol Biogeogr. 2010; 19:504–514.

Ebel G. Update on Powassan virus: emergence of a North American tick-borne flavivirus. Annu Rev Entomol. 2010; 55:95–110. [PubMed: 19961325]

Eisen RJ, Eisen L, Castro MB, Lane RS. Environmentally related variability in risk of exposure to Lyme disease spirochetes in Northern California: Effect of climatic conditions and habitat type. Environ Entomol. 2003; 32:1010–1018.

Eisen RJ, Eisen L, Lane RS. Predicting density of *Ixodes pacificus* nymphs in dense woodlands in Mendocino County, California, based on geographic information systems and remote sensing versus field-derived data. Am J Trop Med Hyg. 2006; 74:632–640. [PubMed: 16606998]

Eisen RJ, Eisen L, Ogden N, Beard C. Linkages of weather and climate with *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae), enzootic transmission of *Borrelia burgdorferi,* and Lyme disease in North America. J Med Entomol. 2015; 53:250–261.

Eisen RJ, Eisen L, Beard C. County-scale distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the continental United States. J Med Entomol. 2016; 53:349–386. [PubMed: 26783367]

Elith, J. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Ferson, S., Burgman, M., editors. Quant Methods Conserv Biol. Springer-Verlag; New York, NY: 2002. p. 39-58.

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans R, Huettmann F, Leathwick JR, Lehmann A, et al. Novel methods improve prediction of species' distributions from occurrence data. Ecography (Cop). 2006; 29:129–151.

Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008; 77:802–813. [PubMed: 18397250]

Engler R, Guisan A, Rechsteiner L. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J Appl Ecol. 2004; 41:263–274.

Estrada-Peña A. Increasing habitat suitability in the United States for the tick that transmits Lyme disease: A remote sensing approach. Environ. Health Perspect. 2002; 110:635–640.

Fielding A, Bell J. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv. 1997; 24:38–49.

Goddard J, Piesman J. New records of immature *Ixodes scapula-ris* from Mississippi. J Vector Ecol. 2006; 31:421–422. [PubMed: 17249363]

Goltz L, Goddard J. Observations on the seasonality of *Ixodes scapularis* Say in Mississippi, U.S.A. Syst. Appl Acarol. 2013; 18:212–217.

Goltz L, Varela-Stokes A, Goddard J. Survey of adult *Ixodes scapularis* Say for disease agents in Mississippi. J Vector Ecol. 2013; 38:401–403. [PubMed: 24581371]

Guisan A, Graham CH, Elith J, Huettmann F. Sensitivity of predictive species distribution models to change in grain size. Divers Distrib. 2007; 13:332–340.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005; 25:1965–1978. [9 May 2016] http://www.worldclim.org/.

Homer C, Dewitz J, Yang L, Jin S, Danielson P, Xian G, Coulston J, Herold N, Wickham J, Megown K. Completion of the 2011 National land cover database for the conterminous United States-Representing a decade of land cover change information. Photogramm Eng Remote Sensing. 2015; 8:345–354.

Jouda F, Perret J, Gern L. Ixodes ricinus density, distribution, and prevalence of *Borrelia burgdorferi* Sensu Lato infection along an alti-tudinal gradient. J Med Entomol. 2004; 41:162–169. [PubMed: 15061274]

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

Killilea ME, Swei A, Lane RS, Briggs CJ, Ostfeld RS. Spatial dynamics of Lyme disease: a review. Ecohealth. 2008; 5:167–195. [PubMed: 18787920]

Koffi JK, Leighton PA, Pelcat Y, Trudel L, Lindsay LR, Milord F, Ogden NH. Passive surveillance for *I. scapularis* ticks: Enhanced analysis for early detection of emerging Lyme disease risk. J Med Entomol. 2012; 49:400–409. [PubMed: 22493860]

Kugeler KJ, Farley GM, Forrester JD, Mead PS. Geographic distribution and expansion of human Lyme disease, United States. Emerg Infect Dis. 2015; 21:1455–1457. [PubMed: 26196670]

Lane R. Seasonal activity of two human-biting ticks. Calif Agric. 1990; 44:23–25.

Lane R, Piesman J, Burgdorfer W. Lyme Borreliosis: Relation of its causative agent to its vectors and hosts in North America and Europe. Annu Rev Entomol. 1991; 35:587–609.

Lantos PM, Nigrovic L, Auwaerter P, Fowler V, Ruffin F, Brinkerhoff R, Reber J, Williams C, Broyhill J, Pan W, Gaines D. Geographic expansion of Lyme disease in the southeastern United States, 2000-2014. Open Forum Infect Dis. 2015; 2:1–8.

Lees A, Milne A. The seasonal and diurnal activities of individual sheep ticks (*Ixodes ricinus L.*). Parasitology. 1951; 41:189–208. [PubMed: 14911213]

MacDonald AJ, Briggs CJ. Truncated seasonal activity patterns of the western blacklegged tick (*Ixodes pacificus*) in central and southern California. Ticks Tick Borne Dis. 2015; 7:234–242. [PubMed: 26564403]

Madhav NK, Brownstein JS, Tsao JI, Fish D. A dispersal model for the range expansion of blacklegged tick (Acari: Ixodidae). J Med Entomol. 2004; 41:842–852. [PubMed: 15535611]

Mather TN, Nicholson MC, Donnelly EF, Matyas BT. Entomologic index for human risk of Lyme disease. Am J Epidemiol. 1996; 144:1066–1069. [PubMed: 8942438]

Mead PS. Epidemiology of Lyme disease. Infect Dis Clin North Am. 2015; 29:187–210. [PubMed: 25999219]

Moore SM, Eisen RJ, Monaghan A, Mead P. Meteorological influences on the seasonality of Lyme disease in the United States. Am J Trop Med Hyg. 2014; 90:486–496. [PubMed: 24470565]

Morisette J, Jarnevich C, Holcombe T, Talbert C, Ignizio D, Talbert M, Silva C, Koop D, Swanson A, Young N. VisTrails SAHM: visualization and workflow management for species habitat modeling. Ecography (Cop). 2013; 36:129–135.

Needham GR, Teel PD. Off-host physiological ecology of Ixodid ticks. Annu Rev Entomol. 1991; 36:659–681. [PubMed: 2006871]

Nelson CA, Saha S, Kugeler KJ, Delorey MJ, Shankar MB, Hinckley AF, Mead PS. Incidence of clinician-diagnosed Lyme disease, United States, 2005–2010. Emerg Infect Dis. 2015; 21:1625–1631. [PubMed: 26291194]

Newman EA, Eisen L, Eisen RJ, Fedorova N, Hasty J, Vaughn C, Lane RS. *Borrelia burgdorferi* Sensu Lato spirochetes in wild birds in Northwestern California: Associations with ecological factors, bird behavior, and tick infestation. PLoS ONE. 2015; 10:e0118146. [PubMed: 25714376]

Nix, H. A biogeographic analysis of Australian elapid snakes, pp. 4–15. In: Longmore, R., editor. Atlas Elapid Snakes Aust. Australlian Government Publishing Service; Canberra: 1986.

Ogden NH, Lindsay LR, Beauchamp G, Charron D, Maarouf A, O'Callaghan CJ, Waltner-Toews D, Barker IK. Investigation of relationships between temperature and developmental rates of tick Ixodes scapularis (Acari: Ixodidae) in the laboratory and field. J Med Entomol. 2004; 41:622–633. [PubMed: 15311453]

Ostfeld RS, Cepeda O, Hazler K, Miller M. Ecology of Lyme disease: Habitat associations of ticks (*Ixodes scapularis*) in a rural landscape. Ecol Appl. 1995; 5:353–361.

Padgett K, Lane R. Life cycle of *Ixodes pacificus* (Acari: Ixodidae): timing of developmental processes under field and laboratory conditions. J Med Entomol. 2001; 38:684–693. [PubMed: 11580041]

Pepin KM, Eisen RJ, Mead PS, Piesman J, Fish D, Hoen AG, Barbour AG, Hamer S, Diuk-Wasser Ma. Geographic variation in the relationship between human Lyme disease incidence and density of infected host-seeking *Ixodes scapularis* nymphs in the Eastern United States. Am J Trop Med Hyg. 2012; 86:1062–1071. [PubMed: 22665620]

Piesman J, Gern L. Lyme borreliosis in Europe and North America. Parasitology. 2004; 129:S191–S220. [PubMed: 15938512]

Rand PW, Lubelczyk C, Lavigne GR, Elias S, Holman MS, Lacombe EH, Smith RP. Deer density and the abundance of *Ixodes scapularis* (Acari: Ixodidae). J Med Entomol. 2003; 40:179–184. [PubMed: 12693846]

Randolph SE, Storey K. Impact of microclimate on immature tick-rodent host interactions (Acari: Ixodidae): Implications for parasite transmission. J Med Entomol. 1999; 36:741–748. [PubMed: 10593075]

Robinson SJ, Neitzel DF, Moen Ra, Craft ME, Hamilton KE, Johnson LB, Mulla DJ, Munderloh UG, Redig PT, Smith KE, et al. Disease risk in a dynamic environment: The spread of tick-borne pathogens in Minnesota, USA. Ecohealth. 2015; 12:152–163. [PubMed: 25281302]

Rosen M, Hamer S, Gerhardt R, Jones C, Muller L, Scott M, Hickling G. Borrelia burgdorferi not detected in widespread *Ixodes scapularis* (Acari: Ixodidae) collected from white-tailed deer in Tennessee. J Med Entomol. 2012; 49:1473–14810. [PubMed: 23270178]

Schulze TL, Jordan Ra, Hung RW. Suppression of subadult *Ixodes scapularis* (Acari: Ixodidae) following removal of leaf litter. J Med Entomol. 1995; 32:730–733. [PubMed: 7473629]

Scott JD, Fernando K, Banerjee SN, Durden LA, Byrne SK, Banerjee M, Mann RB, Morshed MG. Birds disperse Ixodid (Acari: Ixodidae) and *Borrelia burgdorferi-infected* ticks in Canada. J Med Entomol. 2001; 38:493–500. [PubMed: 11476328]

Segurado P, Araujo MB. An evaluation of methods for modelling species distributions. J Biogeogr. 2004; 31:1555–1568.

Sonenshine, D., Roe, RM. Biology of Ticks. 2. Oxford University Press; United Kingdom: 2013.

Springer YP, Jarnevich CS, Barnett DT, Monaghan AJ, Eisen RJ. Modeling the present and future geographic distribution of the Lone Star Tick, *Amblyomma americanum* (Ixodida: Ixodidae), in the continental United States. Am J Trop Med Hyg. 2015; 93:875–890. [PubMed: 26217042]

Stafford KC. Survival of immature *Ixodes scapularis* (Acari: Ixodidae) at different relative humidities. J Med Entomol. 1994; 31:310–314. [PubMed: 8189424]

Stafford KC, Magnarelli La. Spatial and temporal patterns of *Ixodes scapularis* (Acari: Ixodidae) in Southeastern Connecticut. J Med Entomol. 1993; 30:762–771. [PubMed: 8360900]

Stromdahl E, Hickling G. Beyond Lyme: Aetiology of tick-borne human diseases with emphasis on the South-eastern United States. Zoonoses Public Health. 2012; 59:48–64. [PubMed: 22958250]

Talbert, C., Talbert, M. User documentation for the Software for Assisted Habitat Modeling (SAHM) package in VisTrails. U.S. Geoglogical Survey; Fort Collins, CO: 2001.

Templer PH, Schiller AF, Fuller NW, Socci AM, Campbell JL, Drake JE, Kunz TH. Impact of a reduced winter snowpack on litter arthropod abundance and diversity in a northern hardwood forest ecosystem. Biol Fertil Soils. 2012; 48:413–424.

Thorton, P., Thornton, M., Mayer, B., Wilhelmi, N., Wei, Y., Devarakonda, R., Cook, R. Daymet: Daily surface weather data on a 1-km grid for North America, 1980–2008. Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC); Oak Ridge, TN: 2012.

U.S. Geoglogical Survey. Global 30 Arc-Second Elevation (GTOPO30). Land Processes Distributed Active Archive Center (LP DAAC); Sioux Falls, SD: 1996.

Vail SG, Smith G. Air temperature and relative humidity effects on behavioral activity of blacklegged tick (Acari: Ixodidae) nymphs in New Jersey. J Med Entomol. 1998; 35:1025–1028. [PubMed: 9835697]

Vail SG, Smith G. Vertical Movement and posture of black-legged tick (Acari: Ixodidae) nymphs as a function of temperature and relative humidity in laboratory experiments. J Med Entomol. 2002; 39:842–846. [PubMed: 12495181]

Weisbrod A, Johnson R. Lyme disease and migrating birds in the St. Croix River Valley. Appl Environ Microbiol. 1989; 55:1921–1924. [PubMed: 2782872]

Wilson M, Adler G, Spielman A. Correlation between abundance of deer and that of the deer tick, *Ixodes dammini* (Acari: Ixodidae). Ann Entomol Soc. 1985; 78:172–176.

Wilson ML, Ducey AM, Litwin TS, Gavin Ta, Spielman A. Microgeographic distribution of immature *Ixodes dammini* ticks correlated with that of deer. Med Vet Entomol. 1990; 4:151–159. [PubMed: 2132979]
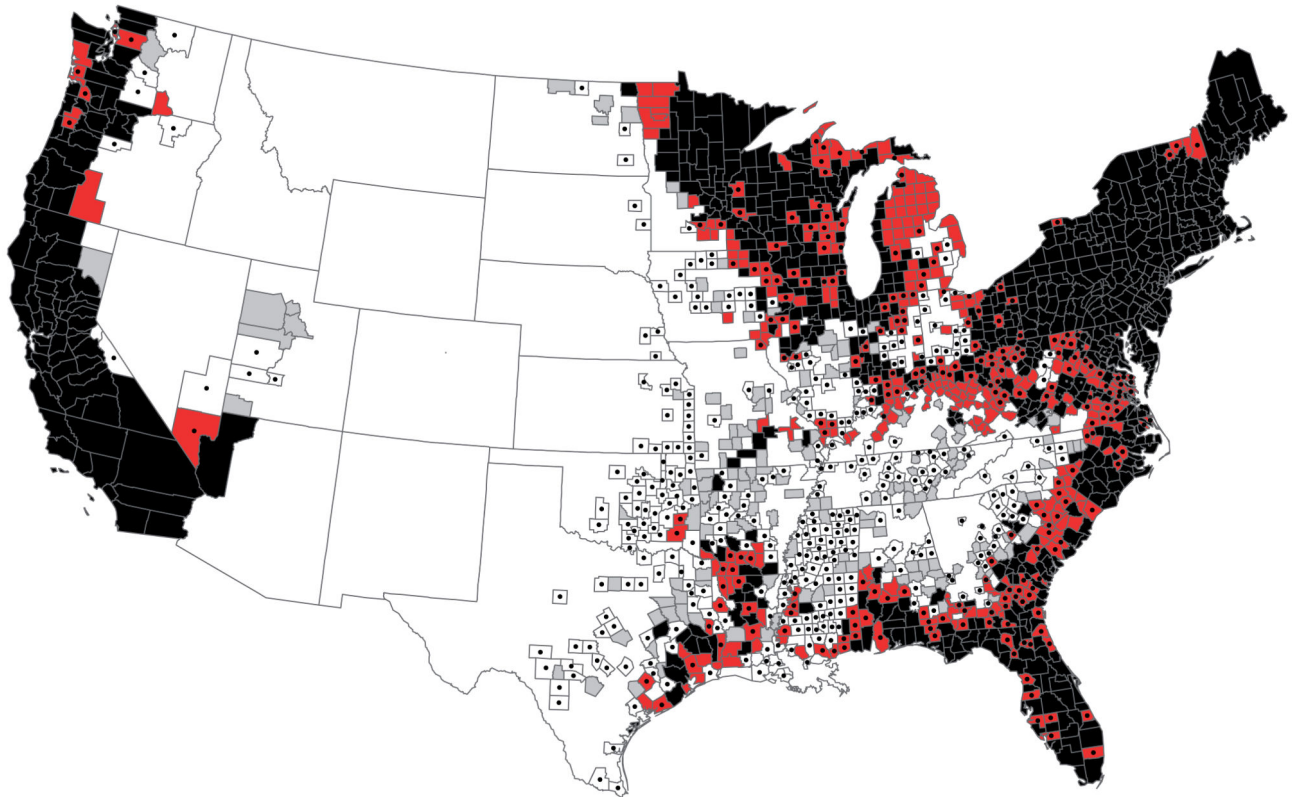
- ·   Reported, but not established, Ixodes spp. populations
- 🟥   No established Ixodes spp. population, but two or more ensemble members predicted suitable habitat
- ⬜   Established Ixodes spp. population, but were predicted as unsuitable by three or four ensemble members
- ⬛   Established Ixodes spp. populations

Fig. 1.

Map depicting counties on the leading edge of *Ixodes* spp. expansion and counties where the ticks are established but not predicted as suitable by the ensemble model. Counties in black are those that have established *Ixodes* spp. populations (Eisen et al. 2016). Counties in red are those that do not have an established *Ixodes* spp. population but are predicted to have suitable habitat by two or more ensemble model members. Counties in grey are those that have established *Ixodes* spp. populations but were predicted as unsuitable by three or four ensemble model members. Counties with black dots have reported (but not established) *Ixodes* spp. populations.
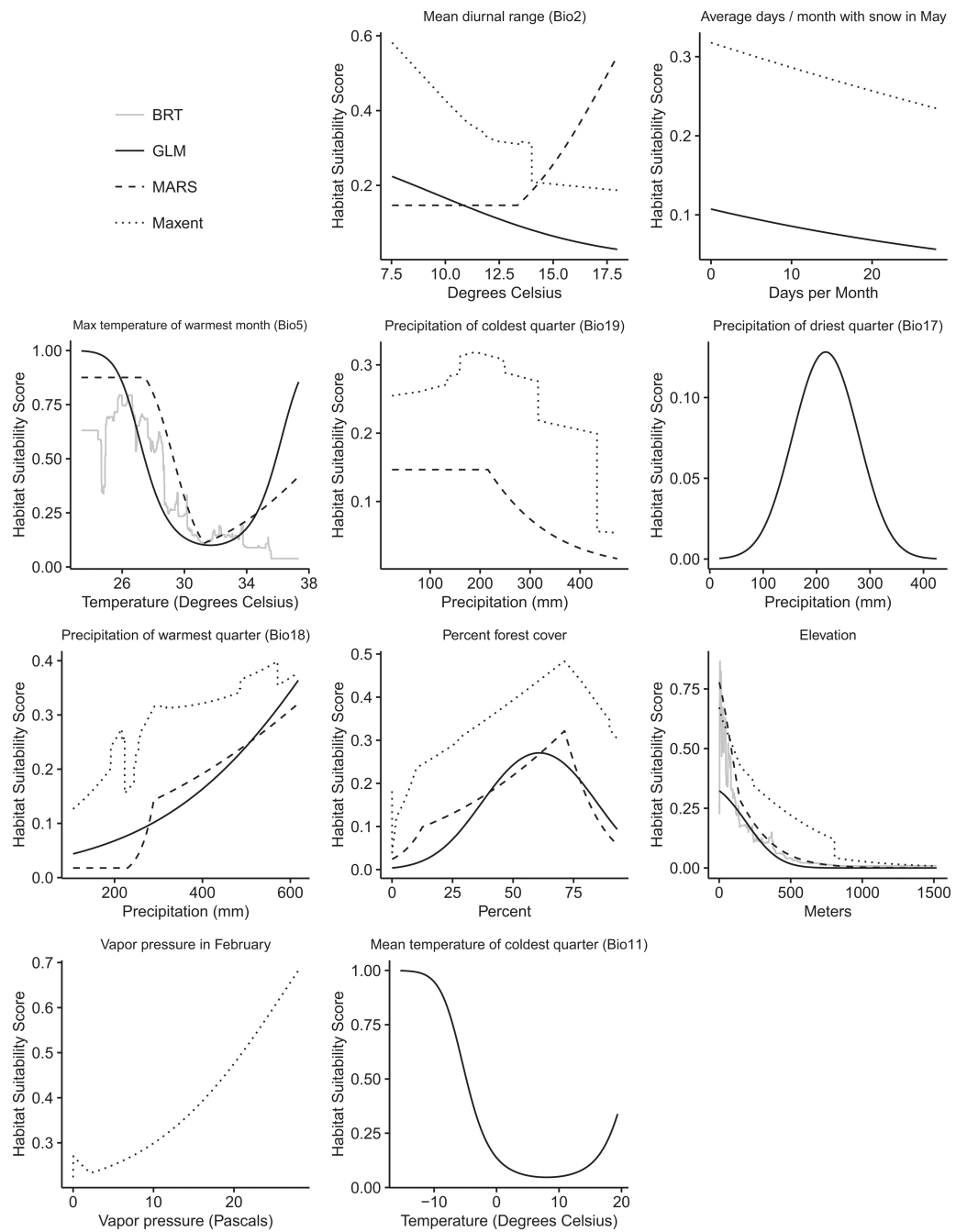
Fig. 2.

Response curves for the climate, elevation, or land cover predictor variables selected by the optimized *Ixodes scapularis* models. The x-axis represents the range of each predictor in the training dataset, and the y-axis represents the associated probability of suitable habitat (0= not suitable, 1 =maximum suitability).
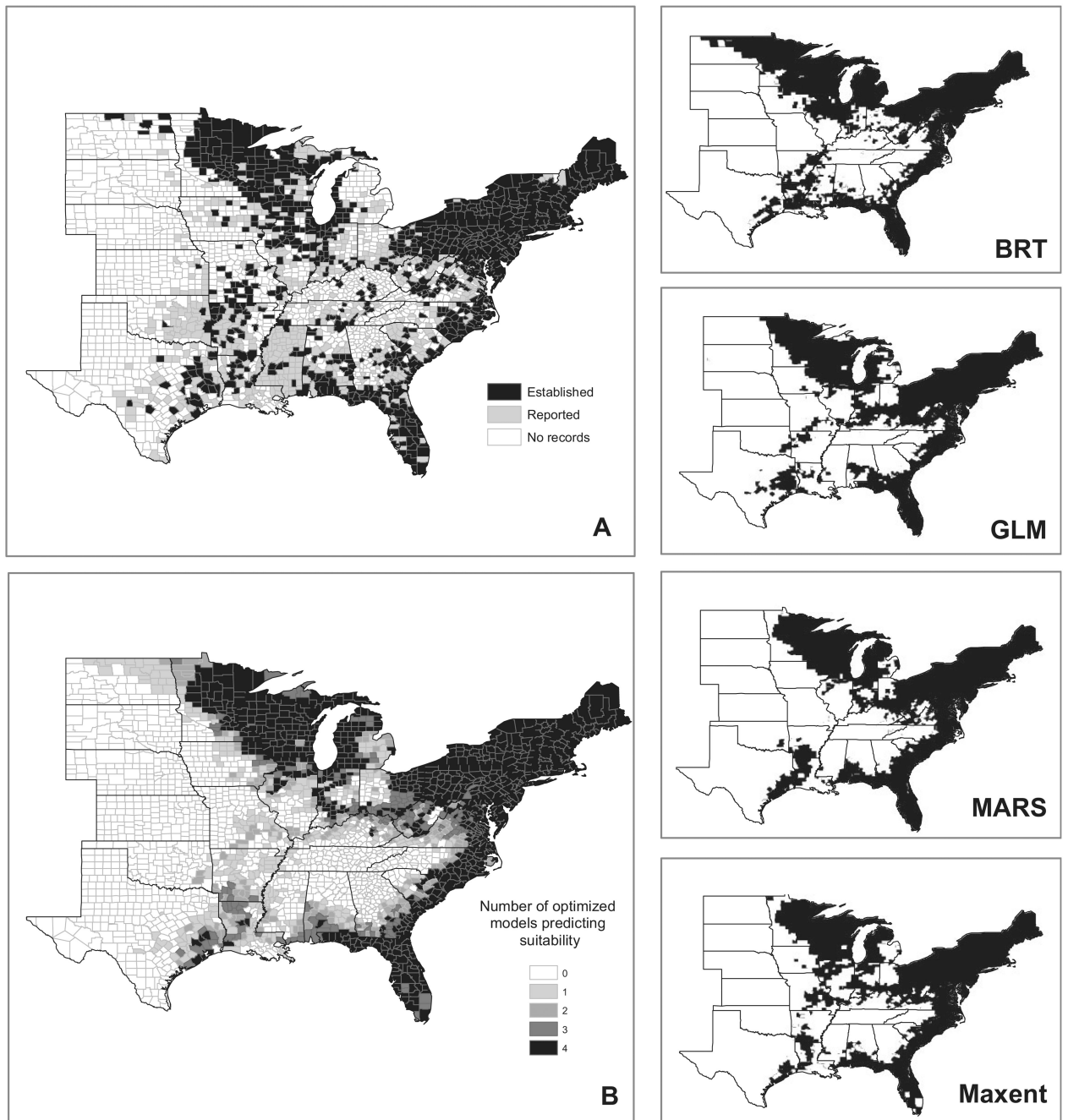
Fig. 3.

Maps depicting *Ixodes scapularis* surveillance records and results of the habitat suitability modeling. (A) County-level classification of *I. scapularis* surveillance records based on (Eisen et al. 2016). (**B**) Map of ensemble model consensus habitat suitability scores. Scores indicate the number of the four optimized models that classified a given county as having climate, elevation, and land cover conditions suitable for the establishment of *I. scapularis*. (**Right column**) Maps depicting the predicted distribution of *I. scapularis* by each of the individual optimized models: BRT, GLM, MARS, and Maxent, respectively. BRT=boosted

regression tree; GLM=generalized linear model; MARS=multivariate adaptive regression spline; Maxent=maximum entropy.
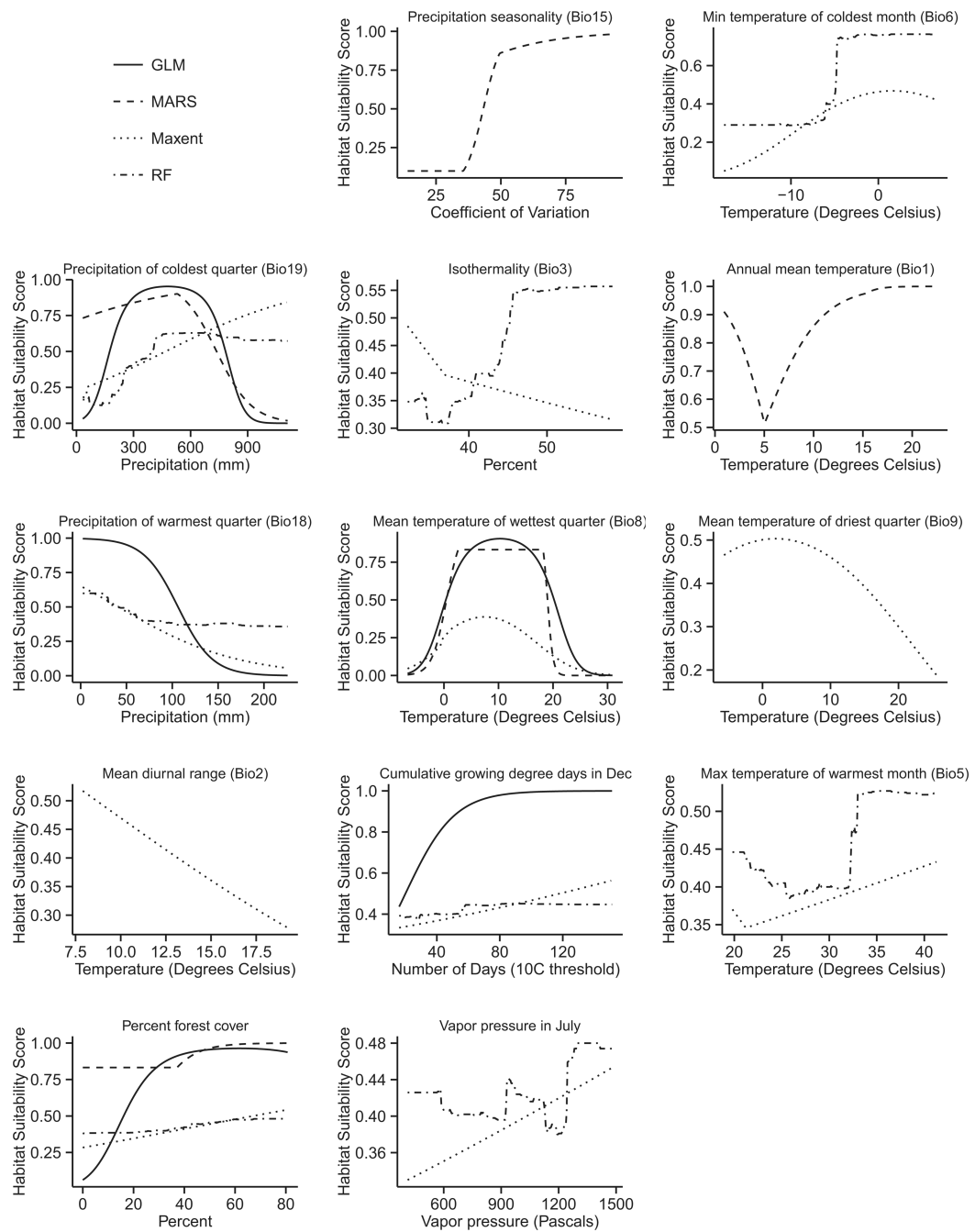
Fig. 4.

Response curves for the climate, elevation, or land cover predictor variables selected by the optimized *Ixodes pacificus* models. The x-axis represents the range of each predictor in the training dataset, and the y-axis represents the associated probability of suitable habitat (0= not suitable, 1=maximum suitability).
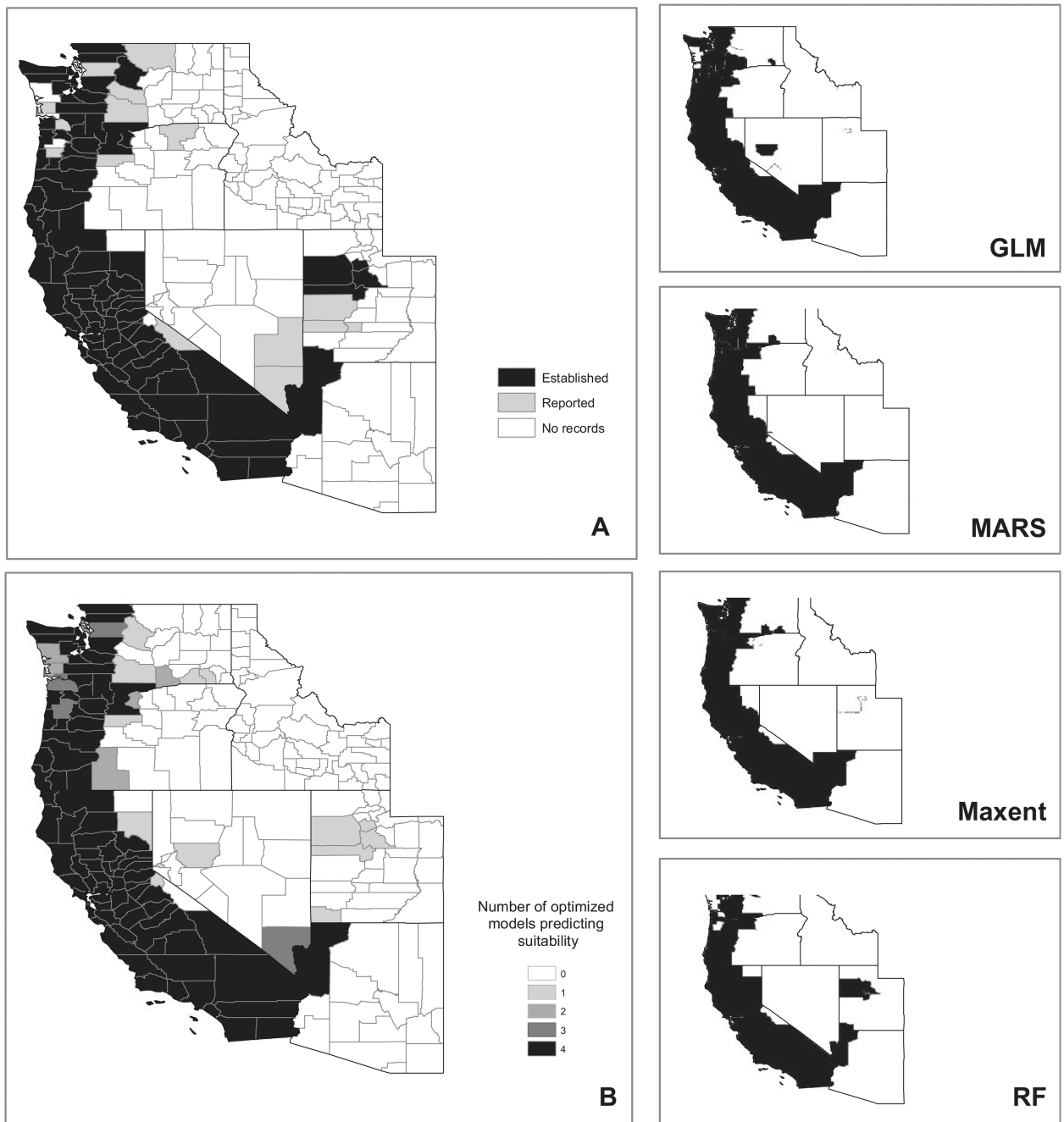
Fig. 5.

Maps depicting *Ixodes* pacificus surveillance records and results of the habitat suitability modeling. (**A**) County-level classification of *I. pacificus* surveillance records based on (Eisen et al. 2016). (**B**) Map of ensemble model consensus habitat suitability scores. Scores indicate the number of the four optimized models that classified a given county as having climate, elevation, and land cover conditions suitable for the establishment of I. pacificus. (**Right column**) Maps depicting the predicted distribution of *I. pacificus* by each of the individual optimized models: GLM, MARS, Maxent, and RF, respectively.

GLM=generalized linear model; MARS=multivariate adaptive regression spline; Maxent=maximum entropy; RF=random forest.

Table 1

Relative contributions of the climate, elevation, and land cover predictors selected by each of the distribution modeling algorithms for the *Ixodes scapularis* models

Climate, elevation, and land cover predictors	Percent deviance explained (rank)	Normalized contribution values (%)			
		BRT	GLM	MARS	Maxent
Mean diurnal temperature range (Bio2)	10.9 (1)	–	0.6	0.6	3.7
Average days with snow cover in May (Swe5)	10.8 (2)	–	1.4	–	0.3
Max temperature of warmest month (Bio5)	9.1 (3)	55.3	11.8	34.1	–
Precipitation of coldest quarter (Bio19)	8.4 (4)	–	–	7.1	5.3
Precipitation of driest quarter (Bio17)	7.4 (6)	–	16.3	–	–
Precipitation of warmest quarter (Bio18)	6.8 (7)	–	10.4	13.0	8.8
Percent forest cover (PercForest)	6.8 (8)	–	19.6	14.6	26.8
Elevation (Elev)	5.8 (10)	44.7	21.3	30.6	28.1
Vapor pressure in February (Vp2)	5.4 (14)	–	–	–	27.0
Mean temperature of coldest quarter (Bio11)	1.8 (37)	–	18.6	–	–

For each variable, the value of deviance explained is parenthetically accompanied by the rank of this value in the list of values for all of the 68 originally considered predictor variables, arranged in descending order (i.e., a rank of 1 indicates the highest value). Also provided are the normalized contribution values that quantify the amount of variation explained by each variable in each of the four optimized models. A blank cell indicates that this variable was not included in the corresponding optimized model. Variables obtained from WorldClim have their associated Bioclim labels indicated parenthetically. BRT, boosted regression tree; GLM, generalized linear model; MARS, multivariate adaptive regression spline; Maxent, maximum entropy.

Table 2

Model selection criteria and performance metrics for the models selected for each modeling algorithm used in the ensemble model of *Ixodes scapularis* distribution

Model selection	BRT[a]		GLM[a]		MARS[b]		Maxent[c]	
Performance metric	Test split	Train	Test split	Train	Test split	Train	Test split	Train
AUC[d]	0.85	0.87	0.86	0.87	0.85	0.86	0.86	0.88
Percent correctly classified	77.1	78.1	77.1	77.5	76.9	78.0	77.4	78.6
Percent deviance explained	30.26	33.38	34.02	36.81	30.24	32.57	26.76	28.64
Sensitivity	0.75	0.78	0.77	0.77	0.76	0.78	0.77	0.79
Specificity	0.78	0.78	0.77	0.78	0.77	0.78	0.78	0.78
Mean threshold	0.31	0.28	0.30	0.30	0.32	0.33	0.40	0.40

[a] Candidate variables (predictor set 1): Bio2, Swe5, Bio5, Bio11, Bio17, Bio18, PercForest, Elev, Cumulative growing degree days in December.

[b] Candidate variables (predictor set 2): Bio2, Swe5, Bio5, Bio18, Bio19, PercForest, Elev.

[c] Candidate variables (predictor set 3): Bio2, Swe5, Bio18, Bio19, PercForest, Elev, Vp2.

[d] Area under the (ROC) curve.

BRT, boosted regression tree; GLM, generalized linear model; MARS, multivariate adaptive regression spline; Maxent, maximum entropy.

Table 3

Relative contributions of the climate and land cover predictors selected by each of the distribution modeling algorithms for the *Ixodes pacificus* models

Climate and land cover predictors	Percent deviance explained (rank)	Normalized contribution values (%)			
		GLM	MARS	Maxent	RF
Precipitation seasonality (Bio15)	51.1 (1)	–	47.4	–	–
Minimum temperature of coldest month (Bio6)	42.1 (5)	–	–	26.6	8.5
Precipitation of coldest quarter (Bio19)	32.7 (10)	38.8	3.3	22.9	80.1
Isothermality (Bio3)	23.3 (20)	–	–	0.8	0.8
Annual mean temperature (Bio1)	22.8 (21)	–	8.8	–	–
Precipitation of warmest quarter (Bio18)	19.1 (28)	14.4	–	18.1	6.0
Mean temperature of wettest quarter (Bio8)	15.8 (31)	8.2	28.8	22.7	–
Mean temperature of driest quarter (Bio9)	12.1 (38)	–	–	1.8	–
Mean diurnal temperature range (Bio2)	11.6 (40)	–	–	1.0	–
Cumulative growing degree days in December (GDD12Cum)	9.2 (44)	7.2	–	1.5	0.1
Max temperature of warmest month (Bio5)	5.6 (53)	–	–	1.4	0.8
Percent forest cover (PercForest)	5.0 (54)	31.3	11.8	3.0	0.1
Vapor pressure in July (Vp7)	2.3 (61)	–	–	0.3	3.6

For each variable, the value of deviance explained is parenthetically accompanied by the rank of this value in the list of values for all of the 68 originally considered predictor variables, arranged in descending order (i.e., a rank of 1 indicates the highest value). Also provided are the normalized contribution values that quantify the amount of variation explained by each variable in each of the four optimized models. A blank cell indicates that this variable was not included in the corresponding optimized model. Variables obtained from WorldClim have their associated Bioclim labels indicated parenthetically. GLM, generalized linear model; MARS, multivariate adaptive regression spline; Maxent, maximum entropy; RF, random forest.

Table 4

Model selection criteria and performance metrics for the models selected for each modeling algorithm used in the ensemble model of *Ixodes pacificus* distribution

Model selection									
	GLM[a]		MARS[b]		Maxent[a]		RF[a]	
Performance metric	Test split	Train	Test split	Train	Test split	Train	Test split	Train
AUC[c]	0.95	0.98	0.94	0.97	0.94	0.97	0.95	0.93
Percent correctly classified	89.9	90.8	89.2	92.4	89.1	92.0	92.5	89.5
Sensitivity	0.91	0.91	0.88	0.93	0.90	0.92	0.89	0.89
Specificity	0.89	0.91	0.90	0.92	0.89	0.92	0.95	0.90
Mean threshold	0.35	0.26	0.33	0.31	0.30	0.27	0.49	0.42

[a]Candidate variables (predictor set 1): Bio6, Bio19, Bio3, Bio18, Bio8, Bio9, Bio2, GDD12Cum, Bio5, PercForest, Vp7.

[b]Candidate variables (predictor set 2): Bio15, Bio19, Bio1, Bio17, Bio8, Vp10, GDD2Cum, PercForest.

[c]Area under the (ROC) curve.

GLM, generalized linear model; MARS, multivariate adaptive regression spline; Maxent, maximum entropy; RF, random forest.