



Published in final edited form as:

*Ticks Tick Borne Dis.* 2025 March ; 16(2): 102446. doi:10.1016/j.ttbdis.2025.102446.

## Estimating the density of questing *Ixodes scapularis* nymphs in the eastern United States using climate and land cover data

Karen M Holcomb\*, Erik Foster,  
Rebecca J Eisen

Division of Vector Borne Diseases, Centers for Disease Control and Prevention, 3156 Rampart Road, Fort Collins, CO 80521 USA

### Abstract

Tick-borne diseases pose a persistent and increasing threat to public health. In the United States, the majority of human infections are caused by pathogens spread by the blacklegged tick, *Ixodes scapularis*. Most infections are reported during the summer months, when nymphal ticks are active in states in the Northeast and Upper Midwest. The density of questing *I. scapularis* nymphs (DON) provides an estimate for the risk of human encounters with nymphs, but it is a resource intensive metric to obtain from field sampling. Thus, DON estimates are limited in the US national tick surveillance database, the ArboNET Tick Module. We estimated DON across all counties in the eastern US using a zero-inflated negative binomial model utilizing tick surveillance data reported to ArboNET (2004–2023) as well as climate and land cover data. The model estimated generally low DON across the southeastern US and Great Plains states with higher estimates in the Upper Midwest and Northeast regions. We assigned counties to relative acarological encounter risk categories based on estimated DON: zero or lower quartile DON estimates were scored as low risk, whereas inter- and upper-quartile DON estimates were scored as moderate-high risk. Counties with moderate-high DON reported from field sampling were accurately categorized by the model as moderate-high encounter risk (99 % sensitivity). However, 80 % of sampled counties reporting low DON were classified as moderate-high risk (20 % specificity). These misclassified counties were typically situated in recently colonized areas in the Northeast and Upper Midwest and likely indicated areas potentially suitable for tick population expansion. Our model yielded a very high negative predictive value (96 %) indicating the model did very well estimating low relative encounter risk in counties where no or few nymphs were collected, and a fair positive predictive value (60 %) indicated that densities may not have reached an expected peak in some

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

\*Corresponding author. [kholcomb@cdc.gov](mailto:kholcomb@cdc.gov) (K.M. Holcomb).

**Disclaimer:** The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

CRedit authorship contribution statement

**Karen M Holcomb:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis. **Erik Foster:** Writing – review & editing, Data curation. **Rebecca J Eisen:** Writing – review & editing, Supervision, Conceptualization.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ttbdis.2025.102446.

locations, particularly in the Northeast, Upper Midwest, and northern states in the Southeast. Further tick surveillance is needed to evaluate and to refine these predictions. The resulting maps are useful for estimating relative risk of nymphal encounters across the eastern US where field data are sparse and may aid in efforts aimed at promoting the use of personal protective measures in communities that are at risk for nymphal tick encounters.

## Keywords

Blacklegged tick; Deer tick; Host-seeking; Encounter risk; Surveillance; Zero-inflated negative binomial model

## 1. Introduction

Blacklegged ticks (*Ixodes scapularis*) pose a significant risk to human health. They are becoming increasingly common in the eastern United States (Eisen and Eisen, 2023a) where they serve as bridge vectors of the causative agents of six known human diseases (Eisen and Eisen, 2018) with the number of cases increasing recently (Rosenberg et al., 2018). For example, Lyme disease (caused primarily by *Borrelia burgdorferi* sensu stricto) accounts for an annual average of 82 % of all domestically acquired vector-borne diseases cases in the United States (US); the number and spatial distribution of cases have dramatically increased during the previous two decades (Kugeler et al., 2015; Schwartz et al., 2017). The majority of Lyme disease cases in the US are reported from the states in Northeast and Upper Midwest regions where disease onset coincides with nymphal tick activity in late spring and early summer (May-July; Mead, 2015).

In the US, active tick surveillance is conducted nationally to monitor changes in the distribution and abundance of *Ixodes* spp. ticks. Tick surveillance provides potentially useful information on the risk of human encounters with ticks and their pathogens (Eisen and Paddock, 2021). *Ixodes scapularis* of northern vs. southern origin differ in nymphal questing behaviors and thus potential risk of encountering humans. Nymphs of northern origin openly quest while those of southern origin are known to not quest above leaf litter (Arsnoe et al., 2015; Diuk-Wasser et al., 2006; Goddard and Piesman, 2006), making them extremely difficult to collect via dragging or flagging. While the distribution of *I. scapularis* displaying nymphal questing behavior of ticks of northern origin has expanded, the precise distribution is poorly defined (Beati et al., 2022) and no genetic marker responsible for variation in nymphal questing behavior has been defined (Arsnoe et al., 2015).

One of the standard acarological metrics derived from surveillance, the density of questing nymphs (DON), provides an indication of the likelihood of human encounters with ticks, assuming similar human behaviors between locations that would put people in contact with ticks (Falco et al., 1998; Pepin et al., 2012). However, collecting data to estimate DON is costly and labor-intensive. Therefore, estimates across the eastern US are spatially and temporally sparse (Foster et al., 2024). Statistical models can be useful for providing DON estimates in unsampled locations. A previous study used data from surveillance performed at evenly gridded surveillance sites (2004–2006) to fit a statistical model to estimate *I. scapularis* DON across the eastern US based on altitude, monthly mean vapor pressure

deficit, and spatial autocorrelation in DON estimates among sampled sites (Diuk-Wasser et al., 2010). This well-performing model was later expanded to estimate the density of *B. burgdorferi* s.s.-infected *I. scapularis* nymphs (Diuk-Wasser et al., 2012) with modeled estimates statistically significantly correlated with the incidence of Lyme disease cases in the eastern US (Pepin et al., 2012). However, dramatic changes have been reported in the tick's distribution in the nearly two decades since this effort was undertaken (Eisen and Eisen, 2023a), underscoring the need for an update.

Following the abovementioned systematic sampling effort from 2004 to 2006, density sampling has been conducted sporadically by public health agencies and their academic and organizational partners (Foster et al., 2024). Starting in 2018, reported tick surveillance data were standardized and collated into the Tick Module of the national arthropod-borne disease surveillance database (ArboNET) administered by the US Centers for Disease Control and Prevention (CDC). To provide updated county-level estimates of DON summarized over space and time, we developed a zero-inflated negative binomial model using reported DON, climatic, and land cover data. We estimated the county-level probability of collecting questing nymphs and DON across the eastern US, ultimately providing an estimate of the relative risk of human-nymph encounters by county throughout the suitable range of *I. scapularis*.

## 2. Materials and methods

### 2.1. Reported nymphal density sampling data

We used collection site-level *I. scapularis* nymphal densities reported to the CDC ArboNET Tick Module (2004–2023) to estimate average expected county-level densities during the peak of nymphal questing activity (May–July). Briefly, each site-level record included the location and date of the tick surveillance activity, the collection method used (e. g., flagging, CO<sub>2</sub>-trap), the number of ticks collected by species and life stage, and the linear distance (m) over which collections were made if the goal of collection was density sampling. We chose county as our unit of analysis primarily because this is the spatial scale used by the national tick and tick-borne disease surveillance program and secondarily, due to lack or uncertainty in the reliability of site-level geolocations (precision in reporting sub-county level data is not required by the surveillance program). Approximately 13 % of the individual collection reports did not contain coordinates and when provided, surveillance personnel may not have provided precise coordinates (e.g., coordinate correspond to county centroid, closest intersection or city, or other jittered location to protect site privacy) that could have reduced the precision of estimates on a sub-county scale. The quality of coordinates could have also varied by agency and potentially across years; we were confident though that surveillance activities were correctly attributed to the county.

To obtain a county-level estimate, we followed a similar procedure as previously described (Foster et al., 2024). Briefly, we only included data with at least 750 m<sup>2</sup> dragged or flagged per collection site visit during the expected peak of *I. scapularis* nymphal questing activity (defined here as 1 May–31 July). The sampling distance inclusion criterion was based on CDC's recommended minimum distance for density sampling (Centers for Disease Control and Prevention, 2022). Foster et al. (2024) summarized densities over the duration of the

expected season of nymphal activity (1 March–30 September), but we chose to focus on the three-month (May–July) peak of questing activity to increase the likelihood that surveillance data were representative of local maximum densities. For each unique collection site and year, we selected the maximum estimated DON per 1000 m<sup>2</sup> (“site peak DON”). When multiple sites were visited in a county in a single year, we averaged the site-level peak DONs for these sites to represent densities across variable conditions and to reduce the impact of outliers. We then selected the maximum county-year peak DON per county across the whole study timeframe (2004–2023). This value, henceforth referred to as “county peak DON”, represents the highest observed average risk of encountering nymphal *I. scapularis* across the county and years of surveillance. For model fitting, we rounded the county peak DON to the nearest integer to represent the number of questing nymphs collected per 1000 m<sup>2</sup> dragged or flagged.

## 2.2. Environmental covariates

We used climate, land cover, and elevation data at the county-scale to identify associations with peak DON across the eastern United States. Here we use “eastern US” to denote all states in the contiguous US that are split by or completely east of the 100th meridian (Fig. 1A). This area encompasses the estimated environmentally suitable range of *I. scapularis* (Fig. 1B).

Using rasters of bioclimatic data (1000-m resolution) from WorldClim (Fick and Hijmans, 2017; WorldClim, 2020), we calculated the county average for each of the 19 bioclimatic variables. Bioclimatic data captures average annual and seasonality trends in temperature and precipitation as well as biologically limiting conditions related to combinations of temperature and precipitation. Bioclimatic variables were derived based on weather observations from 1970 to 2000.

We used the 30-m resolution land cover raster from the 2019 National Landcover Database (Dewitz and U.S. Geological Society, 2021) to derive county-level coverage of land cover categories. For each county, we calculated the percent of forest cover (“deciduous”, “evergreen”, and “mixed forest” categories), low urbanization (“developed, open space” and “developed, low intensity” categories), moderate/high urbanization (“developed, moderate intensity” and “developed, high intensity” categories), and croplands (“hay/pasture” and “cultivated crops” categories). Urbanization categories capture the coverage by impervious surfaces. Impervious surfaces are hard surfaces including rooftops, parking lots, streets, and other surfaces impermeable to infiltration of rainfall to underlying soils or groundwater. The low urbanization category contained <50 % impervious surfaces often mixed with vegetation like lawns, golf courses, and parks which potentially provide habitat for ticks and their hosts in proximity to humans. The moderate/high urbanization category (≥50 % impervious surfaces) captures the opposite, areas likely to be unsuitable for tick populations due to lack of vegetated harborages, conducive microclimates, and hosts.

We calculated mean elevation per county using the global digital elevation model (1000-m resolution) from the U.S. Geological Survey (Danielson and Gesch, 2011; U.S. Geological Survey and National Geospatial-Intelligence Agency, 2010).

To reduce multi-collinearity and improve interpretability of model results, we used hierarchical clustering to select a subset of rather un-correlated variables (pairwise correlation coefficients  $< |0.65|$ ) for model building. To do this, we first grouped variables from counties with included tick surveillance data into clusters such that variables within a cluster had similar patterns of average magnitudes of Pearson correlations to other covariates, using the *klaR* R package (Weihs et al., 2005). Five clusters provided defined groupings of variables with relatively high within-cluster and relatively low between-cluster correlations (Table A.1). Within each cluster, we selected the covariate with the largest average within-cluster correlation (i.e., capturing most of the information in a cluster) as well as any other covariates with an average within-cluster correlation magnitude  $< 0.5$ . Among the covariates selected in this manner, we removed any with a pairwise correlation coefficient  $\geq |0.65|$  to other selected covariates (Table 1) and added percent forest, a covariate often included in previous *I. scapularis* distribution models (Burtis et al., 2022; Hahn et al., 2016). Percent forest had an average within-cluster correlation of 0.60 (Table A.1) so was not initially selected with the above 0.5 threshold.

Due to the large variation in the scale across candidate covariates (i. e., percentages, °C, mm of precipitation, and meters above sea level), we normalized all covariates to have a mean of 0 and standard deviation of 1 before model fitting. For interpretability, we back-transformed covariates to their original scale for plotting and discussion.

### 2.3. Model fitting

Using counties with reported surveillance data meeting our inclusion criteria, we built a model of county-level peak DON using a zero-inflated negative binomial (ZINB) modeling framework to account for the overdispersion and excess zeros in the observed DON data. We chose a negative binomial distribution because we identified overdispersion relative to a Poisson distribution using the *AER* R package (Kleiber and Zeileis, 2008) (*dispersiontest* function:  $P < 0.0001$ ). There are two possible reasons for there to be collections of zero nymphs: 1) nymphs were not present (environmentally unsuitable), or 2), nymphs could be present, but dragging or flagging failed to collect any. The ZINB model, a mixture model, accounts for these two possibilities by combining a binary distribution describing the presence or absence of questing nymphs (i.e., structural zeros) with a negative binomial distribution describing the expected number of questing nymphs collected per 1000 m<sup>2</sup> (i.e., random zeros). Each of these two pieces can include any of the covariates listed in Table 1. Together, these pieces estimate DON (per 1000 m<sup>2</sup>). Since we calculated county peak DON per 1000 m<sup>2</sup> in all sampled counties, no model offset was needed to account for variation in collection area. See Text A.1 for mathematical equations of ZINB models.

We used backward stepwise selection to select the environmental variables to include in our final model. We first fitted a full ZINB model using the *psc* R package (Jackman, 2020; Zeileis et al., 2008) with all candidate environmental variables (Table 1) included in both portions of the ZINB model. We then iteratively removed non-significant variables ( $P > 0.05$  and  $\Delta AIC > 2$ ) to reach the final reduced model. We used likelihood ratio tests (*lmtest* R package; Zeileis and Hothorn, 2002) to confirm dropping each variable did not degrade the model fit.

Considering the covariates in the final fitted ZINB model, we calculated two extrapolation metrics (Shape and MESS) to compare the similarity of environmental conditions in counties included in training the model (training dataset) with those in the rest of the eastern US (projected dataset). These metrics are often used to identify non-analogous conditions where species distribution and ecological niche models apply, or “project” fitted relationships to wider geographical areas. Identifying non-analogous conditions indicates the robustness of out-of-sample predictions because predictions in areas with environmental conditions largely dissimilar to those in the training set may be statistically unsupported and biologically meaningless. Shape is a metric representing the multivariate distance between each projected point and the closest training point, accounting for the dispersion in conditions observed in the training data (Velazco et al., 2023). The multivariate environmental suitability surface (MESS; Elith et al., 2010) is a metric that captures the single most dissimilar condition in the projected data, using the Euclidean distance between the center of the training data distribution and the projected data for each covariate separately. For these calculations, we used the *flexsdm* (Shape; Velazco et al., 2022) and *modEvA* (MESS; Barbosa et al., 2013) R packages, respectively.

We used R (version 4.3.1) (R Core Team, 2023) for all data processing, modeling, and visualization. For this, in addition to those listed elsewhere, we used the *cowplot* (Wikle, 2023), *data.table* (Barrett et al., 2024), *dplyr* (Wickham et al., 2023a), *ggplot2* (Wickham, 2016), *sf* (Pebesma, 2018; Pebesma and Bivand, 2023), and *tidyr* (Wickham et al., 2023b) R packages.

## 2.4. Estimated quantities from the final ZINB model

Using the full final ZINB model, we estimated two quantities, the probability of collecting  $\geq 1$  questing *I. scapularis* nymphs (collection probability) and the density of questing nymphs collected (DON), for each county in the eastern US ( $N = 2694$ ). Both quantities incorporated both portions of the ZINB model and give complementary views of the model results. Collection probability estimates provide gradations among counties in the likelihood that questing nymphs could be encountered while DON estimates provide gradations in the potential magnitude of densities encountered. In the following sub-sections, we describe how we estimated each quantity and its corresponding 95 % confidence interval (CI), as well as model evaluations and validations relevant to each quantity.

### 2.4.1. Estimating the probability of collecting at least one questing *I. scapularis* nymph (collection probability) and evaluating model performance

—Combining the probabilities from both the zero-inflated (logistic piece; structural zeros) and negative binomial (random zeros) portions of the final ZINB model, we calculated the probability that no questing nymphs would be collected. For presentation and discussion, we inverted this probability to the probability that at least one questing nymph would be collected (“collection probability”).

To obtain a 95 % CI for collection probability, we resampled the multivariate normal distribution of fitted covariate estimates 1000 times, using each set of resampled coefficients to obtain an estimate per county. From the resulting estimates, we used the 2.5 % and 97.5

% quantiles to represent the lower and upper 95 % CI bounds. Accounting for the 95 % CI, we categorized the collection probability for each county as unsuitable (<50 % probability of collection), marginal (95 % CI contains 50 % probability), or suitable (>50 % probability), using the 50 % threshold as a natural break in probabilities. We performed a ROC curve analysis to confirm that the collection probability threshold for dichotomizing presence and absence that maximized sensitivity and specificity was not largely different than a 50 % cutoff. For this, we used the *pROC* R package (Robin et al., 2011).

We compared these predicted categories to the field-derived county peak DON using contingency table analysis. For this, we dichotomized county peak DON as absent (DON = 0) or present (DON ≥ 1). To calculate diagnostic metrics, we only used counties with an estimated collection probability categorized as suitable or unsuitable (i.e., not categorized as marginal because we could not confidently assign these counties to one of the categories due to model uncertainty). We calculated sensitivity as the proportion of counties with collections of questing nymphs (observed presence) that were estimated as suitable and specificity as the proportion of counties with observed absence that were estimated as unsuitable. We calculated positive and negative predictive values as the proportion of counties predicted to be suitable or unsuitable with observed presence or absence of questing nymphs, respectively.

As an external validation, we also compared these collection probability categories with the reported collections of nymphs in the full dataset of reported tick surveillance data in the ArboNET Tick Module (i. e., without distance or time-of-year exclusion criteria applied). This provided a more extensive set of counties for comparison of presence or absence of collected questing nymphs. We note that distance was not reported for all collections since density sampling was not always the goal of sampling. Like above, if at least one questing nymph was reported as collected by drag/flag sampling, we categorized that county as present. If no nymphs were collected in sampling, we classified that county as absent. We calculated diagnostic metrics using contingency table analysis as described above.

**2.4.2. Estimating the density of questing *I. scapularis* nymphs collected (DON) and evaluating model performance**—Using the final fitted ZINB model, we also estimated DON per 1000 m<sup>2</sup>. This incorporated the expected number of questing nymphs collected per 1000 m<sup>2</sup> (negative binomial portion) with the estimated presence or absence of questing nymphs (zero-inflated portion). For plotting, we binned predictions as zero (i.e., <1 nymph collected/1000 m<sup>2</sup>) or quartiles of non-zero (i.e., ≥1 nymph collected/1000 m<sup>2</sup>) predictions. Using the same 1000 draws from the multivariate normal distribution of fitted coefficients for the collection probability calculation above, we obtain a 95 % CI for DON for each county. We calculated the standard error of prediction using these 1000 refit estimates as a measure of model uncertainty.

We evaluated model fit and performance by performing cross-validation (CV) and examining scaled Pearson's residuals. We repeated 10-fold CV 10 times and calculated the mean absolute prediction error (average absolute difference between observed and predicted DON) and 95 % CI across repeated CV iterations. Examination of the Pearson residuals enabled us to identify outliers and further evaluate model performance. Residuals with a

magnitude greater than two (“high” residual) indicate the model did not fit well to these data; magnitudes greater than four are considered “extreme” residuals.

## 2.5. Categorization of relative human-nymph encounter risk

Using estimated DON, we then assigned counties to relative risk categories for human-nymph encounters based on their associated 95 % CIs. The relative risk categories assume similar human activity patterns between counties. The low risk category contained values up through the lower quartile outlined above (i.e., zero and the lower quartile of non-zero predictions). The moderate-high risk category contained values in the remaining three quartiles (inter- and upper quartile of non-zero predictions). In other words, we used the 25th percentile of non-zero DON predictions as the breakpoint between the low and moderate-high risk categories. Counties for which the 95 % CI were wholly contained within either risk category were assigned to that category. When 95 % CIs contained values from both categories, we labeled these counties as equivocal risk; due to model uncertainty, we could not confidently assign these counties to one of our risk categories. We chose quartiles as the basis for relative risk categorization to align with a previous summarization of DON estimates (Foster et al., 2024). We binned county peak DON as low or moderate-high risk and compared the estimated and observed risk categories using contingency table analysis to evaluate predictive performance.

To calculate diagnostic metrics, we only used counties with an estimated collection probability categorized as low or moderate-high relative risk (i.e., not categorized as equivocal risk). We calculated sensitivity as the proportion of counties with observed moderate-high county peak DON that were estimated as moderate-high risk and specificity as the proportion of counties with observed low county peak DON that were estimated as low risk. We calculated positive predictive value as the proportion of counties predicted to be moderate-high relative risk that had an observed moderate-high DON and negative predictive value as the proportion of counties predicted to be low relative risk that had an observed low DON.

## 3. Results

### 3.1. Characteristics of included surveillance data

As of 4 March 2024, 1075 counties (39.9 %) from 37 of the 38 states in the eastern US had *I. scapularis* density data from drag/flag sampling reported to the ArboNET Tick Module. A state-level summary of collections is presented in Table A.2. Of all the reported collection data, 668 counties from 35 states had reports that met our inclusion criteria; data were not included from Delaware (no reports), Louisiana (no reports met inclusion criteria), or Florida (no reports met inclusion criteria). County peak DON in counties that met our inclusion criteria ranged from 0.0 – 488.5 *I. scapularis* nymphs per 1000 m<sup>2</sup>.

Sampling effort and consistency (referred to also as surveillance intensity) as measured by median and total number of visits to unique sites, number of sites, and number of years with surveillance varied across included counties (Fig. A.1). Generally, higher amounts of surveillance intensity occurred in states with historically high incidence of Lyme disease

(northern states in the Upper Midwest and Northeast regions) as well as neighboring states (Indiana, Michigan, and West Virginia) to these high incidence states (Schwartz et al., 2017). Effort often varied across counties within a state. States with smaller geographical areas tended to have more uniform surveillance intensity spatially and temporally with surveillance occurring in all counties (e.g., Connecticut and Vermont). Across all included counties, each had a median of one surveillance site (range: 1–19) per year (Fig. A.1B). Per county, each site was visited for sampling a median of two times (range: 1–14) per year (Fig. A.1A). Across years, counties had a median of two (range: 1–39) total sites visited (Fig. A.1E) and a median of four (range: 1–93) total visits across all sites (Fig. A.1D). Compared to all included counties, counties in Minnesota, Michigan, Connecticut, Vermont, and Maine generally had higher numbers of sites and visits per year and more years sampled.

Surveillance occurred from 2004 to 2023 with each county having a median of one year (range: 1–14) of collection data (Fig. A.1C). The most recent year of included surveillance data was during 2018–2023 for 451 (67.5 %) counties (Fig. A.1F); two-thirds of included counties were sampled at least once since the initiation of national tick surveillance efforts in 2018. Of the 217 counties for which the most recent year of surveillance data was prior to 2018, surveillance in 155 (71.4 %) counties dated back to the systematic sampling used to inform the previous DON model (Diuk-Wasser et al., 2010). Approximately half of these counties (N = 77) were in the Upper Midwest while the others were somewhat evenly distributed across the Northeast (N = 26), Southeast (N = 32), and Great Plains (N = 20) regions. Across the Southeast and southern states of the Great Plains, most counties only had a single year (2004) of data.

### 3.2. Final fitted zinb model

The zero-inflated piece of our final ZINB model (predicting absence of questing nymphs; structural zeros) showed support for a positive association with absence of questing nymphs and increasing mean diurnal temperature range (Bio2) as well as negative associations with increasing precipitation of the driest quarter (Bio17) and precipitation of the warmest quarter (Bio18; Table 2). In other words, the probability of the presence of questing nymphs in a county decreased with increasing diurnal temperature ranges. As diurnal temperature ranges are generally larger in arid as compared to humid climates, aridity appears to have a negative impact of on the presence of questing nymphs. Additionally, the probability of the presence of questing nymphs increased with increasing precipitation during the driest quarter of the year as well as during the warmest quarter (summer). Larger amounts of precipitation, especially during the summer when nymphs could experience heat stress, appears conducive for the presence of questing nymphs. Precipitation of the driest quarter was highly correlated with annual temperature and precipitation patterns as well as extreme conditions like mean temperature of the coldest quarter and minimum temperature of the warmest month (Table 1). Geographically, the majority (501/535) of estimated structural zeros (predicted absence of questing nymphs) occurred in the drier, more arid counties of the Great Plains region (Figs. A.2, A.3).

The negative binomial piece of our final ZINB model (predicting number of questing nymphs collected per 1000 m<sup>2</sup>) showed support for a positive association of increasing

numbers of nymphs collected with increasing forest and increasing low urbanization land cover in a county (Table 2). These land covers could provide harborage and hosts for ticks, thus facilitating greater local abundance. Also, these land covers could provide adequate microclimates (shade) to minimize desiccation and promote host-seeking behavior, increasing the number of questing nymphs collected. We also found a positive association with the number of questing nymphs collected and mean temperature of the wettest month (Bio8). Warmer temperatures during the wettest month means precipitation likely falls as rain instead of snow, potentially providing more moisture during the time of year when nymphs would be actively host-seeking. We found a negative association with the number of collected nymphs and maximum temperature of the warmest month (Bio5). Maximum temperature of the warmest month was highly correlated with temperature extremes like mean temperature of the warmest and coldest quarters (Table 1). Overall, the density of questing nymphs collected appears to be modulated by the presence of land cover types providing suitable tick habitats as well as cooler summer temperatures and the interplay of temperature and precipitation coinciding with periods of activity. Geographically, these fitted relationships resulted in counties assigned as random zeros ( $N = 784$ ) predominantly across non-coastal areas of the southern states in the eastern US, over-lapping with counties assigned as structural zeros ( $N = 322$ ) in the southern states of the Great Plains region (Fig. A.3).

Counties in the western part of the Great Plains and along the Gulf Coast had more dissimilar environmental conditions than those in counties used to fit the ZINB model (Fig. 2A). The amount of dissimilarity, as measured by the Shape metric, increased with increasing geographic distance from counties in the training set, as expected based on Tobler's first law of geography (Tobler, 1970); closer things are more similar than more distantly situated things. The dissimilar conditions in the Great Plains and Gulf Coast were largely novel combinations of environmental data while conditions throughout the rest of the eastern US were largely out-of-range of the training data for individual covariates (Fig 2B). Mean diurnal temperature range (Bio2) and precipitation of the warmest quarter (Bio18) were the most dissimilar univariate conditions (Fig. A.4).

### 3.3. Estimated probability of collecting at least one nymph

Combining both the zero-inflated and negative binomial pieces of the final ZINB model, we estimated the probability that at least one questing nymph would be collected per 1000 m<sup>2</sup> dragged or flagged. The response curves for this probability (Fig. 3A) identified two covariates with a threshold: mean diurnal temperature range (Bio2) and precipitation of the warmest quarter (Bio18). Counties where mean diurnal temperatures ranged by >13 °C and total precipitation in the warmest quarter was <250 mm resulted in dramatically lower probabilities of collecting questing nymphs than below or above this range, respectively. Maximum temperature of the warmest month (Bio5) showed a somewhat sigmoidal, negative relationship with <50 % probability of collecting at least one nymph at temperatures above 30–32 °C. The remaining covariates had positive relationships with collection probability. Increasing precipitation of the driest quarter (Bio17) resulted in only slight increases in collection probability.

The spatial distribution of predicted probability that at least one questing nymph would be collected per 1000 m<sup>2</sup> (Fig. 3B–C) illustrated three distinct regions in the eastern US. The highest probabilities were predicted in the Upper Midwest and Northeast. The Southeast region had relatively low predicted probabilities while the Great Plains region had a markedly low probability (<5 %). Maximum temperature of the warmest month (Bio5; Fig. A.2B) and mean temperature of the wettest month (Bio8; Fig. A.2C) largely described the north-south division in collection probability. Lower collection probability in interior counties in Michigan were largely due to mean diurnal temperature ranges >13 °C. High mean diurnal ranges (Bio2; Fig. A.2A) and low amounts of precipitation during the warmest quarter (Bio18; Fig. A.2E) largely account for these very low probabilities along the Great Plains.

The largest amount of uncertainty, as measured by the width of the 95 % CI (Fig. 3D), occurred in counties along the eastern edge of the Great Plains region, northern parts of the Upper Midwest, and the largely forested areas (Fig. A.2G) of Missouri, Arkansas, and Virginia. The Southeastern region had larger uncertainty than the Northeast and southern Upper Midwest regions (Fig. 3D). The western part of the Great Plains region had the lowest uncertainty (i.e., highest certainty in predicted collection probability).

Considering the 95 % CI of predicted collection probability and dichotomizing at 50 %, we categorized counties as unsuitable, marginal, or suitable (Fig. 3E). The ROC curve analysis identified 48.9 % as the optimal collection probability cutoff, a threshold that resulted in nearly identical spatial distribution of categories and model performance (Fig. A.5) as the 50 % threshold. The AUC of the model was 0.71, indicating acceptable discriminatory performance (Hosmer et al., 2013). With the 50 % threshold, most of the Northeast and Upper Midwest regions were categorized as suitable with a band of marginal counties along the south and east of this area. Northcentral counties in Michigan's lower peninsula were also categorized as marginal with four of these counties categorized as unsuitable. The Great Plains region and most of the Southeast were also categorized as unsuitable with scattered counties along the Gulf Coast and in Florida categorized as suitable or marginal.

A total of 551 counties were categorized as either suitable (N = 485) or unsuitable (N = 66) while 117 were categorized as marginal (i.e., 95 % CI for collection probability contained 50 %; Table 3). Of these 117 counties categorized as marginal for collection of questing nymphs, 72 had a county peak DON ≥ 1 (observed presence) and 45 had a county peak DON of zero (observed absence). These observed present counties were primarily in western Minnesota, southern Indiana, and northern lower peninsula Michigan. The observed absent counties were also in northern parts of Michigan's lower peninsula and generally more spatially distributed throughout the band of marginal counties.

Considering only counties with suitable or unsuitable predictions of collection probability (N = 551, Table 3), the final model correctly classified 372 counties where questing nymphs were collected as suitable and 59 counties where questing nymphs were not collected as unsuitable (i.e., 78 % accuracy). The model also correctly identified 98 % (372/379) of counties with observed nymph collections as suitable (i.e., sensitivity); only 34 % (59/172) of counties with observed absence were predicted to be unsuitable (i.e., specificity). The

majority of these misclassified counties occurred in the Northeast and Upper Midwest where no nymphs were collected, but the model predicted suitability. Of the 485 counties predicted to be suitable, 372 had at least one questing nymph collected (i.e., 77 % positive predictive value) and of the 66 counties predicted to be unsuitable, 59 had no questing nymphs collected (i.e., 89 % negative predictive value). The high negative predictive power suggests that the model was good at predicting counties where questing nymphs would not be collected. The fair positive predictive value indicates that the model predicted a wider area in which nymphs could be collected than was observed, potentially indicating an envelope of environmentally suitable area for nymphal questing activity. Most of such mismatches (predicted suitable, but none collected) occurred in the Upper Midwest and Northeast regions (Fig. 3E).

Examination of reported nymphal collections to the ArboNET Tick Module without our inclusion criteria applied (e.g., we removed the requirements of drag/flag sampling 750 m<sup>2</sup> and did not restrict collections to peak questing periods) identified similar patterns of presence/absence of collected nymphs as our included data. Inclusion of these extra questing nymph observations provided a wider spatial coverage of the eastern US for model evaluation (Fig. A.6). Collections of zero nymphs largely occurred in the Great Plains and Southeastern regions (Fig. A.6A) while collections of at least one nymph occurred in the Northeast and Upper Midwest regions (Fig. A.6B). Of note, all counties in Pennsylvania reported collecting at least one questing nymph while the neighboring state of Ohio had mixed reports of collection success. Of the 55 counties with reported surveillance in Mississippi, a questing nymph was collected in a single county. Similarly, collections in Florida tended not to find questing nymphs (16 of 18 counties reported no nymphs collected). Comparing these observations to the categorized collection probabilities used above resulted in very similar performance metrics; 79 % accuracy, 96 % sensitivity, 50 % specificity, 77 % positive predictive value, and 87 % negative predictive value (Fig. A.6C–D).

### 3.4. Estimated density of questing nymphs (DON)

The response curves for DON from the final ZINB model (Fig. 4A) exhibited similar associations as for collection probability. Mean diurnal temperature range (Bio2) and precipitation of the warmest quarter (Bio18) exhibited a threshold effect with 13 °C (Bio2) and 250 mm (Bio18) as change points. The largest magnitude of variation in DON due to variation in the maximum temperature of the warmest month (Bio5; reduction in DON with increasing temperature) and percent of a county with open space and low intensity urban landcover (increase in DON with increasing cover). Increased mean temperature of the wettest month (Bio8) and percent forest cover were associated with increased DON. Variation in precipitation of the driest quarter (Bio17) had minimal impact on DON with a slightly positive relationship.

Estimated DON illustrated two distinct areas in the eastern US. DON was generally low (< 2.2 nymphs per 1000 m<sup>2</sup>) across the Southeast and Great Plains region and higher in the Upper Midwest and Northeast (Fig. 4B). Nearly all counties throughout the Great Plains region were predicted to have low DON (<1 nymph collected per 1000 m<sup>2</sup>). Scattered

counties along the Gulf Coast were predicted to have 1.0–14.9 nymphs collected per 1000 m<sup>2</sup>. Maximum temperatures <25 °C in the warmest quarter (Bio5; Fig. A.2B) resulted in high DON estimates across the northern edge of the Upper Midwest and Northeast as well as along the Appalachian Mountains through Virginia and North Carolina. Low urbanization land cover was highly associated (correlation coefficient > 0.65, Table 1) with moderate/high urbanization land cover within a county, as these counties usually contained cities and surrounding suburbs. Thus, counties that contained large cities (e.g., Kansas City, Kansas; St. Louis, Missouri; Indianapolis, Indiana; Minneapolis, Minnesota; and Atlanta, Georgia; Fig. A.2H) and were thus largely urbanized were often predicted to have higher DON than neighboring, less urbanized counties. Large amounts of precipitation (>450 mm) during the warmest quarter (Bio18; Fig. A.2E) as well as low diurnal temperature ranges (<13 °C; Bio2; Fig. A.2A) and warmer mean temperatures (>26 °C) during the wettest month (Bio8; Fig. A.2C) along the Gulf Coast, especially throughout Florida, resulted in the higher DON estimates in this region.

As expected for a negative binomial distribution, the highest standard errors for predicted DON were estimated in areas with high predicted DON, largely occurring in the northern parts of the Upper Midwest and Northeast (Fig. 4C). Similarly, the highly urbanized counties mentioned above generally had higher standard error.

We used cross-validation and Pearson residuals to assess model fit and performance for predicting DON. The average magnitude of error across cross-validation iterations was 19.6 nymphs per 1000 m<sup>2</sup> (95 % CI: 19.4–19.7) indicating that, on average across all counties in the training set, the model predictions were 19.6 nymphs per 1000 m<sup>2</sup> different than observed county peak DON. Higher magnitudes of error occurred in counties with large magnitudes of predicted DON. The tight confidence interval indicated that random subsets of counties were not having a large impact on model performance (i.e., robust model performance). The average absolute error of the ZINB model fitted with all data (19.3) fell slightly below this confidence interval. Of the 668 Pearson's  $\chi^2$  standardized residuals from the ZINB model, 35 had an absolute value >2 (i.e., "high") and 7 of which had an absolute value >4 (i.e., "extreme"). All high or extreme residuals were positive, indicating that the model tended to underestimate observed densities. Extreme residuals occurred in Minnesota (Morrison County), Wisconsin (Vernon and Iowa counties), Virginia (Montgomery, Pulaski, and Carrol counties), and Washington DC, all counties with an observed county peak DON >100 per 1000 m<sup>2</sup>. Counties with high or extreme residual values tended to be for counties with lower relative sampling effort in our data (e.g., fewer sites per year, fewer visits per year, few years of surveillance), but not uniquely so (Fig. A.7).

### 3.5. Relative encounter risk

Collapsing predicted DON into two broad categories of low and moderate-high risk of nymphal encounters based on the lower non-zero quartile of predicted DON illustrated a contiguous area of moderate-high risk encompassing the majority of the Upper Midwest and Northeast regions (Fig. 5), reflecting major differences in the magnitude of predicted DON (Fig. 4B). A zone of equivocal risk (i.e., 95 % CI contains estimates in both the low and moderate-high risk groups) separated this area from a broad contiguous area of low

risk spanning the Great Plains region and across much of the Southeastern US. Scattered counties along the central Gulf Coast and in Florida were also predicted as equivocal or moderate-high risk. Using one questing nymph as the cut-point for categorizing relative risk resulted in a wider geographic region of elevated risk including much of the Southeast with the Northeast and Upper Midwest (Fig. A.8) and did not distinguish between difference in magnitude of DON between these regions (Fig. 4B).

Of the 668 counties included in the analysis, the 95 % CI of 596 were wholly included in a single risk category (Table 4). The other 72 were equivocal with the 95 % CI containing estimates in both the low and moderate-high risk categories. Of these, 58 had observed county peak DON in the low risk category while 14 had observations in the moderate-high risk category. These moderate-high observed counties mainly occurred in Illinois and Indiana while the observed low counties were distributed more widely throughout the band of equivocal risk (Fig. 5).

Considering only counties with unequivocal predictions of risk category (Table 4), the model correctly identified 99 % (323/325) of counties with observed moderate-high risk (i.e., sensitivity); only 20 % (55/271) of counties with observed low densities were predicted as low risk (i.e., specificity). Much of this error was attributable to observing low county peak DON in areas of the Northeast and Upper Midwest that were predicted to be highly suitable for elevated DON. Of the 539 counties predicted to be moderate-high risk, 323 were observed as moderate-high (i.e., 60 % positive predictive value) and of the 57 counties predicted to be low risk, 55 were observed as low (i.e., 96 % negative predictive value). In short, the very high negative predictive value illustrates that the model did very well predicting low encounter risk. In contrast, the fair positive predictive value suggests that perhaps densities have not yet reached an expected peak in some locations, especially in the Upper Midwest and Northeast regions; alternatively, the model could be overestimating risk in these areas. Looking at the distribution of error in sensitivity and specificity, our model erred on the side of overprediction of the relative risk of encounters with questing nymphs.

We observed variation in the effort and consistency of sampling (i.e., surveillance intensity) in counties across our predicted relative risk categories (Fig. A.9). Counties predicted as low or equivocal relative risk had a narrower range and lower degree of intensity (e.g., fewer sites, visits per site, and years of data) than counties predicted as moderate-high risk. This could be because lower density sites were visited less often than higher density sites due to their perceived low risk. Counties misclassified as moderate-high risk generally had low surveillance intensity, but not uniquely so, indicating that misclassification was not solely due to low sampling effort and consistency. We observed higher Pearson standardized  $\chi^2$  residuals in counties with observed moderate-high relative risk (Fig. A.10), with most (33/35) of the residuals > 2 occurring in counties predicted to be moderate-high risk. Thus, the majority of ill-fitting points were contained in the higher relative risk category. The other two high residuals occurred in southern Illinois counties predicted to have low relative risk, adjacent to the band of predicted equivocal relative risk.

## 4. Discussion

We developed a predictive model for the county-level average peak density of questing *I. scapularis* nymphs in the eastern United States that integrated reported collections of questing nymphs over two decades from the ArboNET Tick Module with environmental data. Our model results provide an estimate of the average county-level likelihood of encountering at least one questing nymph (collection probability) as well as the expected number of questing nymphs encountered per 1000 m<sup>2</sup> sampled (DON). To overcome a lack of precision in DON estimates that are inherent to opportunistic sampling, reduce underestimation of nymphal encounters based on DON estimates, and to improve interpretability of resulting risk maps, we categorized counties into low or moderate-high risk of encounters with questing nymphs. This categorization accurately identified moderate-high encounter risk in the Northeast and Upper Midwest compared with the Southeast and Great Plains regions where we predicted low risk (high sensitivity). Moreover, compared with the only previous DON model for the eastern US (Diuk-Wasser et al., 2010), our model showed higher nymphal densities extending further south. Several low DON observations from areas predicted to have elevated risk contributed to diminished specificity. Such areas of discordance between observed and predicted densities point to counties where additional surveillance is needed to improve confidence in observations reported to ArboNET. They also highlight avenues for future investigation into additional factors modulating densities or heralding the potential for increased densities as the tick reclaims areas of its historic range (Eisen and Eisen, 2023a). The model was particularly useful in providing estimates of nymphal encounters across the broad geographic areas where tick surveillance records were lacking or incomplete. Because of the resource-intensive nature of tick surveillance, obtaining a uniform distribution of surveillance data would be infeasible. Models can inexpensively provide estimates in unsampled locations, be used to target limited resources to areas of high risk or large amounts of uncertainty (Case et al., 2024), and be useful for communicating risk.

### 4.1. Utility of relative risk categorization

While our ZINB modeling framework allowed the estimation of both collection probability and DON, our relative risk categorization likely provides a more robust and practically meaningful estimate for public health compared with either of these metrics. Variation in the amount of surveillance data used to calculate the county peak DON means that the precision of estimates likely varied between counties, potentially introducing uncertainty into our estimates that our use of our wide relative risk categories helped to mitigate. A previous simulation study based on field data indicated that sampling a site only once during the two months of assumed peak activity introduced profound errors in estimated density (Dobson, 2013); variation in weather or collector technique can result in wide variation in collections. Sampling twice improved precision while sampling every three weeks resulted in the least biased estimate. Of the 2448 unique site-year records included in our dataset, 1112 (45.5 %) sites were only visited once, 611 (25.0 %) were visited twice, and 725 (29.6 %) were visited three or more times during May-July (assumed peak of nymphal activity). Across the 668 counties with included data in our analysis, 101 (15.1 %) had county peak DON estimates based on a single visit to a single collection site. Therefore, while a relatively high

proportion of site-level estimated peak DONs could have been biased due to a single visit to that site, the rolled-up estimate across years at the county level (county peak DON) were likely less biased due to including multiple sampling events across space and/or time. High or extreme residuals often occurred in counties with lower intensities of sampling (e.g., few visits per site or few years of data), but not uniquely so. Thus, this potential variation in accuracy did not appear to dramatically impact the resulting model fit. Grouping DON into relative risk categories further reduced the impact of underlying uncertainty and variation of DON within a season and between sites. While the ZINB model tended to underestimate the magnitude of DON, as highlighted by the sign of the Pearson residuals, most error in this classification scheme was due to observed low relative risk counties predicted to be moderate-high relative risk (overestimation). Thus, our relative risk categories mitigated the impact of imprecise DON estimates due to underestimation of local DON in sampling.

While the relationship between DON and human incidence of tickborne diseases is overall positive (Pepin et al., 2012; Stafford et al., 1998), the exact shape of the relationship is not well characterized (Ginsberg et al., 2021) with the exact density collected of little predictive value except in comparison to other densities. Our relative risk categorizations attempted to provide this contextualization to DON, making them more intuitive and useful than interpreting pure DON values. Additionally, the incorporation of prediction variability into defining the categories highlighted counties for which the variability in the magnitude of estimated DON resulted in largely different groupings vs. overall similar levels of DON (i.e., low, equivocal, or moderate-high relative risk). Our choice of breakpoint between low and moderate-high relative risk categories was rather arbitrary, simply based on quartiles of non-zero predictions, and may not necessarily reflect a biological threshold for risk. Thus, a direct correspondence between the estimated relative encounter risk and human incidence was not implied. Further work is needed to identify the relationship between surveillance metrics and human incidence of disease to develop meaningful risk thresholds. Incorporation of pathogen presence and prevalence data with density have been shown to provide a more meaningful metric of infection risk compared with tick densities alone (Mather et al., 1996; Pepin et al., 2012), but further work is needed in this arena as well to identify predictive thresholds and relationships.

Interestingly, our categorization of collection probability produced a very similar risk map as categorizing DON as relative risk levels. Areas with high likelihoods of collecting at least one questing nymph largely corresponded with areas with moderate-high relative risk; relative risk categorization resulted in a slightly wider spatial distribution, incorporated more counties on the eastern edge of the Upper Midwest region and in northern Michigan. Both categorizations smoothed over gradation of predictions, producing largely contiguous areas of each categorization and scattered counties categorized as higher risk across the more uniformly low risk states of the Gulf Coast. The relative risk categorization had slightly lower specificity (20 % vs. 34 %) and positive predictive value (60 % vs. 77 %), but a very similar sensitivity (99 % vs. 98 %) and slightly higher negative predictive value (96 % vs. 89 %) than the categorization of collection probability; overall very similar diagnostic performance for these categorization methods. Similarity in performance could imply that once northern origin ticks become established in a county, densities may stabilize within the moderate-high relative risk level. If so, identification of suitability for questing nymphs (i.e.,

collection probabilities > 50 %) would be sufficient to define areas of elevated risk. Great variation in density values would be expected within these areas (Wilson 1998). Further work is needed to identify if a meaningful risk threshold is practically different than at least one nymph collected or if further categorization of densities provides further insight into risk.

#### 4.2. Spatial patterns of predicted DON

Our results provide an updated prediction of the spatial distribution and average peak density of questing *I. scapularis* nymphs across counties in the eastern United States. A previous predictive model (Diuk-Wasser et al., 2010) used surveillance from uniformly distributed sampling sites east of the 100th meridian from 2004 to 2006 to estimate nymphal densities. Notably, the DON data from that study were included in ArboNET and used to build the model presented here, but our dataset also included more recent estimates that reflect expansion in the tick's range and increases in DON in some areas (Eisen and Eisen, 2023a). While we employed a different density calculation method (Diuk-Wasser et al. (2010) calculated daily average density), thus limiting our ability to directly compare densities, comparison of spatial patterns of crude groupings (high or low densities) provide insight into broadscale changes in populations of *I. scapularis* ticks of northern origin that display open questing behavior (Arsnoe et al., 2015). Diuk-Wasser et al. (2010) predicted two foci of high DON in the coastal Northeast and most of Wisconsin in the Upper Midwest. Our model expands this area of elevated risk to encompass the majority of the Upper Midwest and Northeast regions with the inclusion of more recent surveillance data. Specifically, we predicted higher densities in eastern Pennsylvania and New York; throughout Illinois, Indiana, Ohio, and in more coastal lower peninsula Michigan; as well as along the Appalachian Mountains into West Virginia and Virginia. These areas align with the described pattern of expansion of *I. scapularis* from these originally described foci over the previous several decades (Eisen and Eisen, 2023a). These changes also align with dramatic increases in states (West Virginia, Virginia, northern North Carolina) reporting tremendous increases in Lyme disease cases (Schwartz et al., 2017). Our model also predicted lower DON in northcentral lower peninsula Michigan. These predictions align with field observations where lower densities were collected at inland versus coastal sites in Michigan (Hamer et al., 2010).

Given that tick surveillance was performed for public health goals and not research aims, our dataset of included surveillance data contained a patchwork of collection locations across the eastern US. Most sampling occurred in the northern part of the region where sampled counties were often contiguous while sampled counties across the Southeast were sparsely distributed. Due to the large variation in distances between the centroids of sampled counties (mean: 42.9 km; range: 0.9–231.6 km), we were unable to incorporate an autocorrelation term into our model. In the previous model (Diuk-Wasser et al., 2010), variation in abundance was primarily driven by an autocorrelation term, indicating that neighboring populations had larger impacts on local abundance than environmental or landscape features. Although not accounted for in our model, we believe this biological phenomenon remains relevant and suggest viewing with caution results predicting elevated DON in discontinuous counties (e.g. those in the Gulf Coast region). In our model, spatially

smoothed environmental conditions likely incorporated some of this autocorrelation among neighboring populations. However, the observed county peak DON was more variable county-to-county than predicted DON indicating that environmental covariates did not explain all variation. Further work could be done to fully elucidate drivers of variation in DON because our summarization of data spatially and temporally smoothed over sometimes substantial year-to-year and site-to-site heterogeneity at subcounty scales (Foster et al., 2024; Schulze and Jordan, 2022).

Our model predicted that the probability of collecting questing nymphs was low across the majority of the Southeast with the exception of some scattered counties along the Gulf Coast and in Florida. This regional difference was expected and encouraging given we did not explicitly account for variation in nymphal questing behavior of northern and southern origin ticks (Arsnoe et al., 2015; Diuk-Wasser et al., 2006; Goddard and Piesman, 2006). The difference in questing was largely captured by fitted relationship with the maximum temperature of the warmest month (Bio5) and the mean temperature of the wettest month (Bio8) in the negative binomial piece predicting no to low densities in this region (i.e., assigned random zeros here). The zero-inflated piece captured suitability throughout much of the eastern US except the Great Plains region (i.e., fitted structural zeroes here), largely aligning with the predicted potential range of *I. scapularis* ticks (Burtis et al., 2022; Hahn et al., 2016, 2017). Broadly, maximum temperature defined a north-south high risk-low risk division while mean diurnal temperature separated low risk in arid areas that experience larger temperature differences between day and night from elevated risk in the more humid areas further east. The sporadic pockets of higher collection probability and DON in the Southeast were surprising. As all elevated predictions in the Southeast occurred in states for which no data were included in the model (Florida and Louisiana), sparse sampling in southern states could have contributed to these unexpected predictions; the model used associations primarily informed by northern counties with robust surveillance to predict in these sparsely surveilled areas. Additionally, given the dissimilarity of environmental conditions relative to the training data, the discordance between model predictions and reported collections of questing nymphs in the full ArboNET Tick Module, and the geographic isolation of moderate-high risk counties in Florida and Louisiana from all other moderate-high risk counties, we urge caution in attributing elevated risk to these counties. These results could potentially indicate that if *I. scapularis* ticks of northern origin were introduced here, populations could establish and reach moderately high densities, putting these populations at higher risk for tick-borne diseases. Alternatively, this could simply reflect model error based on limited training data available from the Southeast.

Florida reports an annual average of 150 Lyme disease cases (0.7 cases/100,000, 2008–2022; Centers for Disease Control and Prevention, 2024), the highest number in the southeastern US. Most cases (>80 %) of Lyme disease reported in low-incidence states like Florida reported recent travel to high-incidence states where exposure likely occurred; misdiagnosis of patients without travel history may occur frequently (Forrester et al., 2015). At face value, our results seem to indicate there is moderate-high risk for human encounters with questing nymphs scattered throughout Florida. However, several observations yield low confidence that the actual risk is as high as predicted. Given the geographical isolation of Florida from the large contiguous areas with consistent documentation of collections

of questing nymphs with drags/flags in the Northeast and Upper Midwest (Foster et al., 2024), it is unlikely that local populations of northern origin ticks have established in these counties in Florida, especially since the change in population distribution over the previous decades appears to have been primarily via contagious spread (Eisen and Eisen, 2023a). Established populations of *I. scapularis* in Florida are genetically grouped with other ticks of southern origin (Beati et al., 2022; Bhosale et al., 2023; Kessler et al., 2019; Xu et al., 2020). Our external validation dataset did identify collections in 2 of the 18 Florida counties with collection data, indicating that collection of questing nymphs by dragging/flagging is possible, but not frequent. Also, as mentioned previously, no data from Florida met our inclusion criteria so predictions here relied on fitted associations that were not necessarily causal. Further surveillance is needed to investigate the current encounter risk and if these counties develop moderate-high DON if northern origin *I. scapularis* ticks are introduced.

#### 4.3. Biological plausibility of fitted associations with environmental covariates

Although we are reluctant to attribute causality to the observed correlations, we found that the presence of questing nymphs was associated primarily with measures of humidity and precipitation, whereas the number of nymphs collected per area was predicted mainly by land use categories (forest and low urbanization) and temperature. We note these associations are not representative or predictive of any single collection event, but rather they segregate areas of expected low from elevated risks of encountering questing nymphs.

The fitted model found that the presence of questing nymphs was associated with environmental factors related to the balance of humidity and moisture. Overall, *I. scapularis* ticks have low tolerance for desiccation (Diuk-Wasser et al., 2021) so these associations align with the known biology of the tick. Mean diurnal temperature ranges (the difference between day and night temperatures) are greater in arid climates compared with more humid or tropical climate zones. Thus, the inverse relationship of collection probability and mean diurnal temperature range (Bio2), a metric related to humidity, suggests that, consistent with previous studies, more arid conditions impede survival or limit nymphal questing. For example, active ticks experiencing high vapor pressure deficits (i.e., less humid conditions) desiccate rapidly (Needham and Teel, 1991) and thus reduce the height and time of questing above leaf litter, return to refugia more often to rehydrate, or simply fail to survive in too arid of conditions (Burtis et al., 2016; Eisen et al., 2016; Schulze and Jordan, 2003). Precipitation during the driest (Bio17) and warmest (Bio18) quarters captured extreme condition in terms of water and temperature balance. Increases in precipitation during the driest quarter is highly correlated with increases in the total annual amount of precipitation (Pearson correlation = 0.70), but variation in the timing of the driest quarter influences the kind of precipitation. For counties throughout the Southeast and coastal Northeast, the driest quarter occurs during warmer times of year (mean temperature 10.0 – 28.2 °C; WorldClim, 2020) when precipitation would likely be in the form of rain, thus providing moisture when ticks are more active. In contrast, the driest quarter for the northern parts of the eastern US occurs during colder months (mean temperature -14.3 – 0 °C; WorldClim, 2020) when precipitation would likely be in the form of snow that could provide insulation for overwintering ticks and promote survival (Eisen et al., 2016). Precipitation during the warmest quarter relates to the amount of moisture present during the summer months when

questing nymphs would experience heat stress, thus giving a measure of dehydration risk (Eisen and Eisen, 2023b; Eisen et al., 2016). DON has been observed to decrease with hot, dry weather during nymphal questing periods (Burtis et al., 2016), likely due to nymphs questing at lower heights, retuning to refugia more often to rehydrate, or reducing questing activities overall (Vail and Smith, 1998, 2002). We observed larger magnitudes of variation in the associations of collection probability and DON with precipitation during the warmest quarter vs. the driest quarter, potentially because precipitation during the warmest quarter temporally aligns with nymphal questing activity thus resulting in larger modulation in surveillance-based estimates.

The model predicted that the density of collected nymphs was related to land cover and temperature. These predictions generally aligned with previously described associations. Higher densities of nymphs have been collected in forests as compared to open meadows or croplands (Ginsberg et al., 2020; Ostfeld et al., 1995, 1996). Forests provide habitat for hosts (rodents and deer) as well as microclimate refugia (Eisen and Eisen, 2023b; Killilea et al., 2008), thus supporting larger local abundances. DON has been found to decrease with increasing hot, dry summer conditions (Burtis et al., 2016; Crandall et al., 2024), likely due to nymphs questing closer to leaf litter to mitigate desiccation (Vail and Smith, 2002). The number of questing nymphs collected was also predicted to increase with increasing low intensity urban landcover (i.e., open space and low intensity urbanization) in a county, an association not previously identified. This relationship could have been driven by underlying surveillance methodologies instead of biological relationships. In order to estimate the risk of human-tick encounters in areas where people frequent, tick surveillance is often performed in proximity to human populations (e.g., heavily used parks in or near urban/suburban areas) (Centers for Disease Control and Prevention, 2022). Thus, the model identified an association between open space and low intensity urbanization and increasing number of nymphs collected, which may simply reflect a bias in where sampling was conducted.

Previous models, like that of Diuk-Wasser et al. (2010), have found elevation as a significant predictor of presence or density of *Ixodes* ticks (Burtis et al., 2022; Hahn et al., 2016; Johnson et al., 2018; Sharma et al., 2024; Tran et al., 2021). In contrast, we did not find elevation to be significant in our model. Elevation has been proposed to primarily delineate suitable tick habitats with an unclear direction of relation within the suitable range (Rand et al., 2003). In our model, other environmental covariates (i.e., mean diurnal temperature range, precipitation of the driest and warmest quarters) weakly associated with mean elevation (Pearson correlation  $< |0.37|$ ) were defining suitability for questing nymphs (zero-inflated piece of the model). This difference is potentially reflective of the distribution of surveillance data and environmental conditions included in this and previous models and highlights that not all relationships are necessarily causal.

#### 4.4. Limitations and future directions

Given our choice to use county as our spatial unit of analysis, averaged environmental data were not reflective of weather conditions or habitat present at the location and time of dragging or flagging or changes in densities over time. Thus, our county peak estimates,

while providing a more global picture of density, smoothed over sub-county heterogeneity in density and associated environmental factors which themselves were long-term climate averages. Our DON estimates therefore reflect expected densities given average conditions and not on-the-ground predictions of DON at any particular collection event. Future work could investigate environmental drivers of inter-annual and site-to-site variation in densities.

Also of importance is that our DON and collection probability estimates do not indicate the presence or absence of *I. scapularis* populations and should not be taken as estimating the spatial extent or proportion of suitable habitat occupied by this species in the eastern US. As we did not include any data on adult or larval collections or counties with known established populations, our scope aimed specifically at quantifying only human encounters with questing nymphs (i.e., nymphal *I. scapularis* that quest in a manner such that they are more likely to encounter a human).

Counties with discordant predictions and observations as well as those categorized as equivocal risk are potential targets for future surveillance, maximizing the usefulness of limited resources for surveillance. Our ZINB model generally underestimated the observed county peak DON, especially in the Upper Midwest and Northeast. Diuk-Wasser et al. (2010)'s results had similar instances of predicted high densities, but no successful collections of questing nymphs (e.g., eastern Wisconsin). These areas have been reported to have higher densities, but albeit in patchier populations (Eisen and Eisen, 2023a). Sampling in these areas of discordance could identify uncounted factors that are limiting densities in these areas or to monitor for increasing densities. Throughout counties in Pennsylvania, New Hampshire, Ohio, Vermont, Maine, and the upper peninsula of Michigan, no nymphs were collected while the model estimated >15 nymphs per 1000 m<sup>2</sup>. These areas represent recently colonized areas (Eisen and Eisen, 2023a; Elias et al., 2022; Lantos et al., 2017; Wang et al., 2014) where tick populations may still be increasing. Also, areas of equivocal risk could indicate areas of emergence where tick populations are currently or may shortly establish in new counties. Spread of *I. scapularis* ticks across the eastern US has largely occurred through contagious spread of tick populations from neighboring counties (Eisen and Eisen, 2023a). Thus, those counties bordering moderate-high risk counties may be at an increased risk for establishing populations, dependent on the presence of suitable habitats and hosts.

## 5. Conclusions

Our ZINB incorporated reports obtained from national tick surveillance to provide an updated estimate of the distribution, density, and relative risk of encountering questing *I. scapularis* nymphs in the eastern US. The ZINB model predicted generally low DON and encounter risk across the southeastern US and Great Plains states with higher estimates in the Upper Midwest and Northeast regions. Our relative encounter risk categorization highlighted counties of potential suitability for tick population expansion in recently colonized areas of the Northeast and Upper Midwest and did very well at predicting areas of low encounter risk. Categorizing DON as low or moderate-high risk also reduced the potential impact of non-systematic surveillance across space and time on the exact magnitude of DON and provided a more intuitive metric of the relative

encounter risk. The utilization of a zero-inflated model enabled us to identify environmental conditions statistically associated with presence/absence of questing nymphs separately from those conditions associated with increasing densities of questing nymphs. While fitted relationships with these bioclimatic variables may not be causally predictive of nymphal abundance, they define an environmentally suitable envelope that can be used to raise awareness of changing risks for encounters with medically important ticks in the US. The collection probability and DON estimates provide county-level gradations in the likelihood of encountering any questing nymphs and the number of encountered nymphs, respectively. Our risk map provides a step towards a large-scale synthesis of human-tick encounter risk. To reach this, further work is needed to identify and refine the relationship of surveillance estimates of tick density with human encounters with ticks and their pathogens.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements:

We thank all our public health and academic partners who participate in proposing, designing, funding, conducting, and reporting tick and tickborne disease surveillance activities in the United States. We also thank Mark Delorey and James Burtis (Centers for Disease Control and Prevention) for their input on fitting, assessing, and presenting zero-inflated negative binomial models.

## Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Data availability

Data sharing agreements in ArboNET preclude sharing data below the county spatial scale. Aggregated and summarized county data are presented in this publication.

## References

- Arsnoe IM, Hickling GJ, Ginsberg HS, McElreath R, Tsao JI, 2015. Different populations of blacklegged tick nymphs exhibit differences in questing behavior that have implications for human Lyme disease risk. *PLoS. One* 10, e0127450. 10.1371/journal.pone.0127450. [PubMed: 25996603]
- Barbosa AM, Real R, Muñoz AR, Brown JA, Robertson M, 2013. New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Divers. Distrib* 19, 1333–1338. 10.1111/ddi.12100.
- Barrett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, Hocking T, 2024. data. table: Extension of `data.frame` (R package version 1.15.2). [software]. <https://CRAN.R-project.org/package=data.table>.
- Beati L, Ludwig JM, Childs CT, Patel J, Bhagatwala J, Arsnoe IM, Burke RL, Ginsberg HS, Hamer SA, Hickling GJ, LeBrun RA, Ogden NH, Oliver JD, Tsao JI, Meeûs TD, 2022. Population genetics of *Ixodes scapularis* Say, 1821 in the United States using microsatellite loci: more than a North-South phenomenon. *Zenodo*. 10.5281/zenodo.7265193.
- Bhosale CR, Wilson KN, Ledger KJ, White ZS, Dorleans R, De Jesus CE, Wisely SM, 2023. Ticks and tick-borne pathogens in recreational greenspaces in North Central Florida, USA. *Microorganisms*. 11, 756. 10.3390/microorganisms11030756. [PubMed: 36985329]
- Burtis JC, Foster E, Schwartz AM, Kugeler KJ, Maes SE, Fleshman AC, Eisen RJ, 2022. Predicting distributions of blacklegged ticks (*Ixodes scapularis*), Lyme disease spirochetes (*Borrelia*

- burgdorferi sensu stricto*) and human Lyme disease cases in the eastern United States. Ticks. Tick. Borne Dis 13, 102000. 10.1016/j.ttbdis.2022.102000. [PubMed: 35785605]
- Burtis JC, Sullivan P, Levi T, Oggenfuss K, Fahey TJ, Ostfeld RS, 2016. The impact of temperature and precipitation on blacklegged tick activity and Lyme disease incidence in endemic and emerging regions. Parasit. Vectors 9, 606. 10.1186/s13071-016-1894-6. [PubMed: 27887625]
- Case BKM, Dye-Braumuller KC, Evans C, Li H, Rustin L, Nolan MS, 2024. Adapting vector surveillance using Bayesian experimental design: an application to an ongoing tick monitoring program in the southeastern United States. Ticks. Tick. Borne Dis 15, 102329. 10.1016/j.ttbdis.2024.102329. [PubMed: 38484538]
- Centers for Disease Control and Prevention, 2022. Surveillance For *Ixodes scapularis* and Pathogens Found in This Tick Species in the United States, in: Eisen RJ, Eisen L, Graham CB, Foster E, Hojgaard A, Mead PS, Kersh G, Karpathy SE, Paddock CD, Savage H, Herwaldt BL, Bradbury R, Backenson B, Prusinski MA, Neitzel DF, Bjork J (Eds.). [https://www.cdc.gov/ticks/resources/TickSurveillance\\_Iscapularis-P.pdf](https://www.cdc.gov/ticks/resources/TickSurveillance_Iscapularis-P.pdf).
- Centers for Disease Control and Prevention, 2024. Lyme Disease Cases By State Or Locality. [https://www.cdc.gov/lyme/datasurveillance/case-counts/Lyme\\_Disease\\_Cases\\_by\\_State\\_or\\_Locality.csv](https://www.cdc.gov/lyme/datasurveillance/case-counts/Lyme_Disease_Cases_by_State_or_Locality.csv) (accessed 18 July 2024).
- Crandall KE, Millien V, Kerr JT, 2024. High-resolution environmental and host-related factors impacting questing *Ixodes scapularis* at their northern range edge. Ecol. Evol 14, e10855. 10.1002/ece3.10855. [PubMed: 38384829]
- Danielson JJ, Gesch DB. Global multi-resolution terrain elevation data 2010 (GMTED2010). U.S. Geological Survey, Open-File Report 2011–1073. 10.3133/ofr20111073.
- Dewitz J, U.S. Geological Survey, 2021. National land cover database (NLCD) 2019 land cover conterminous United States. U.S. Geological Survey data release, v 2.0. doi:10.5066/P9KZCM54.
- Diuk-Wasser MA, Gatewood AG, Cortinas MR, Yaremych-Hamer S, Tsao J, Kitron U, Hickling G, Brownstein JS, Walker E, Piesman J, Fish D, 2006. Spatiotemporal patterns of host-seeking *Ixodes scapularis* nymphs (Acari: Ixodidae) in the United States. J. Med. Entomol 43, 166–176. 10.1603/0022-2585(2006)043[0166:spohis]2.0.co;2. [PubMed: 16619595]
- Diuk-Wasser MA, Hoen AG, Cislo P, Brinkerhoff R, Hamer SA, Rowland M, Cortinas R, Vourc'h G, Melton F, Hickling GJ, Tsao JI, Bunikis J, Barbour AG, Kitron U, Piesman J, Fish D, 2012. Human risk of infection with *Borrelia burgdorferi*, the Lyme disease agent, in eastern United States. Am. J. Trop. Med. Hyg 86, 320–327. 10.4269/ajtmh.2012.11-0395. [PubMed: 22302869]
- Diuk-Wasser MA, VanAcker MC, Fernandez MP, 2021. Impact of land use changes and habitat fragmentation on the eco-epidemiology of tick-borne diseases. J. Med. Entomol 58, 1546–1564. 10.1093/jme/tjaa209. [PubMed: 33095859]
- Diuk-Wasser MA, Vourc'h G, Cislo P, Hoen AG, Melton F, Hamer SA, Rowland M, Cortinas R, Hickling GJ, Tsao JI, Barbour AG, Kitron U, Piesman J, Fish D, 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 Ecol Biogeogr 19, 504–514.
- Dobson AD, 2013. Ticks in the wrong boxes: assessing error in blanket-drag studies due to occasional sampling. Parasit. Vectors 6, 344. 10.1186/1756-3305-6-344. [PubMed: 24321224]
- Eisen L, Eisen RJ, 2023a. Changes in the geographic distribution of the blacklegged tick, *Ixodes scapularis*, in the United States. Ticks. Tick. Borne Dis 14, 102233. 10.1016/j.ttbdis.2023.102233. [PubMed: 37494882]
- Eisen RJ, Eisen L, 2018. The blacklegged tick, *Ixodes scapularis*: an increasing public health concern. Trends. Parasitol 34, 295–309. 10.1016/j.pt.2017.12.006. [PubMed: 29336985]
- Eisen RJ, Eisen L, 2023b. Evaluation of the association between climate warming and the spread and proliferation of *Ixodes scapularis* in northern states in the Eastern United States. Ticks. Tick. Borne Dis 15, 102286. 10.1016/j.ttbdis.2023.102286. [PubMed: 38016209]
- Eisen RJ, Eisen L, Ogden NH, Beard CB, 2016. 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 53, 250–261. 10.1093/jme/tjv199. [PubMed: 26681789]

- Eisen RJ, Paddock CD, 2021. Tick and tickborne pathogen surveillance as a public health tool in the United States. *J. Med. Entomol* 58, 1490–1502. 10.1093/jme/tjaa087. [PubMed: 32440679]
- Elias SP, Witham JW, Schneider EF, Rand PW, Hunter ML, Lubelczyk C, Smith RP, Hamer S, 2022. Emergence of *Ixodes scapularis* (Acari: Ixodidae) in a small mammal population in a coastal oak-pine forest, Maine, USA. *J. Med. Entomol* 59, 725–740. 10.1093/jme/tjab209. [PubMed: 34958101]
- Elith J, Kearney M, Phillips S, 2010. The art of modelling range-shifting species. *Methods Ecol. Evol* 1, 330–342. 10.1111/j.2041-210X.2010.00036.x.
- Falco RC, McKenna DF, Daniels TJ, Nadelman RB, Nowakowski J, Fish D, Wormser GP, 1998. Temporal relation between *Ixodes scapularis* abundance and risk for Lyme disease associated with erythema migrans. *Am. J. Epidemiol* 149, 771–776. 10.1093/oxfordjournals.aje.a009886.
- Fick S, Hijmans R, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol* 37, 4302–4315. 10.1002/joc.5086.
- Forrester JD, Brett M, Matthias J, Stanek D, Springs CB, Marsden-Haug N, Oltean H, Baker JS, Kugeler KJ, Mead PS, Hinckley A, 2015. Epidemiology of Lyme disease in low-incidence states. *Ticks. Tick. Borne Dis* 6, 721–723. 10.1016/j.ttbdis.2015.06.005. [PubMed: 26103924]
- Foster E, Holcomb KM, Eisen RJ, 2024. Density of host-seeking *Ixodes scapularis* nymphs by region, state, and county in the contiguous United States generated through national tick surveillance. *Ticks. Tick. Borne Dis* 15, 102316. 10.1016/j.ttbdis.2024.102316. [PubMed: 38325243]
- Ginsberg HS, Hickling GJ, Burke RL, Ogden NH, Beati L, LeBrun RA, Arsnoe IM, Gerhold R, Han S, Jackson K, Maestas L, Moody T, Pang G, Ross B, Rulison EL, Tsao JI, 2021. Why Lyme disease is common in the northern US, but rare in the south: the roles of host choice, host-seeking behavior, and tick density. *PLoS. Biol* 19, e3001066. 10.1371/journal.pbio.3001066. [PubMed: 33507921]
- Ginsberg HS, Rulison EL, Miller JL, Pang G, Arsnoe IM, Hickling GJ, Ogden NH, LeBrun RA, Tsao JI, 2020. Local abundance of *Ixodes scapularis* in forests: effects of environmental moisture, vegetation characteristics, and host abundance. *Ticks. Tick. Borne Dis* 11, 101271. 10.1016/j.ttbdis.2019.101271. [PubMed: 31677969]
- Goddard J, Piesman J, 2006. New records of immature *Ixodes scapularis* from Mississippi. *J. Vector. Ecol* 31, 421–422. 10.3376/1081-1710(2006)31[421:Nrois]2.0.Co;2. [PubMed: 17249363]
- Hahn MB, Jarnevich CS, Monaghan AJ, Eisen RJ, 2016. Modeling the geographic distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the contiguous United States. *J. Med. Entomol* 53, 1176–1191. 10.1093/jme/tjw076. [PubMed: 27282813]
- Hahn MB, Jarnevich CS, Monaghan AJ, Eisen RJ, 2017. Response: the geographic distribution of *Ixodes scapularis* (Acari: Ixodidae) revisited: the importance of assumptions about error balance. *J. Med. Entomol* 54, 1104–1106. 10.1093/jme/tjx096. [PubMed: 28874013]
- Hamer SA, Tsao JI, Walker ED, Hickling GJ, 2010. Invasion of the Lyme disease vector *Ixodes scapularis*: implications for *Borrelia burgdorferi* endemicity. *EcoHealth* 7, 47–63. 10.1007/s10393-010-0287-0. [PubMed: 20229127]
- Hosmer DWJ, Lemeshow S, Sturdivant RX, 2013. Chapter 5: Assessing the fit of the model, Applied logistic regression. John Wiley & Sons, New York.
- Jackman S, 2020. pscl: Classes and Methods For R developed in the Political Science Computational Laboratory (R Package Version 1.5.5.1). [software]. United States Studies Centre University of Sydney, Australia. <https://github.com/atahk/pscl/>.
- Johnson TL, Boegler KA, Clark RJ, Delorey MJ, Bjork JKH, Dorr FM, Schiffman EK, Neitzel DF, Monaghan AJ, Eisen RJ, 2018. An acarological risk model predicting the density and distribution of host-seeking *Ixodes scapularis* nymphs in Minnesota. *Am. J. Trop. Med. Hyg* 98, 1671–1682. 10.4269/ajtmh.17-0539. [PubMed: 29637876]
- Kessler WH, Ganser C, Glass GE, 2019. Modeling the distribution of medically important tick species in Florida. *Insects*. 10, 190. 10.3390/insects10070190. [PubMed: 31261713]
- Killilea ME, Swei A, Lane RS, Briggs CJ, Ostfeld RS, 2008. Spatial dynamics of Lyme disease: a review. *EcoHealth* 5, 167–195. 10.1007/s10393-008-0171-3. [PubMed: 18787920]
- Kleiber C, Zeileis A, 2008. Applied Econometrics With R. Springer-Verlag, New York.

- Kugeler KJ, Farley GM, Forrester JD, Mead PS, 2015. Geographic distribution and expansion of human Lyme disease, United States. *Emerg. Infect. Dis* 21, 1455–1457. 10.3201/eid2108.141878. [PubMed: 26196670]
- Lantos PM, Tsao J, Nigrovic LE, Auwaerter PG, Fowler VG, Ruffin F, Foster E, Hickling G, 2017. Geographic expansion of Lyme disease in Michigan, 2000–2014. *Open. Forum. Infect. Dis* 4. 10.1093/ofid/ofw269ofw269.
- Mather TN, Nicholson MC, Donnelly EF, Matyas BT, 1996. Entomologic index for human risk of Lyme disease. *Am. J. Epidemiol* 144, 1066–1069. 10.1093/oxfordjournals.aje.a008879. [PubMed: 8942438]
- Mead PS, 2015. Epidemiology of Lyme disease. *Infect. Dis. Clin. North Am* 29, 187–210. 10.1016/j.idc.2015.02.010. [PubMed: 25999219]
- Needham GR, Teel PD, 1991. Off-host physiological ecology of ixodid ticks. *Annu. Rev. Entomol* 36, 659–681. 10.1146/annurev.en.36.010191.003303. [PubMed: 2006871]
- O'Donnell MS, Ignizio DA, 2012. Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey, Data Series 691 10. <http://pubs.usgs.gov/ds/691>.
- Ostfeld RS, Cepeda OM, Hazler KR, Miller MC, 1995. Ecology of Lyme disease: habitat associations of ticks (*Ixodes scapularis*) in a rural landscape. *Ecol. Appl* 5, 353–361. 10.2307/1942027.
- Ostfeld RS, Hazler KM, Cepeda OM, 1996. Temporal and spatial dynamics of *Ixodes scapularis* (Acari: Ixodidae) in a rural landscape. *J. Med. Entomol* 33, 90–95. 10.1093/jmedent/33.1.90. [PubMed: 8906910]
- Pebesma E, 2018. Simple features for R: standardized support for spatial vector data. *R. J* 10, 439–446. 10.32614/RJ-2018-009.
- Pebesma E, Bivand R, 2023. *Spatial Data Science: With applications in R*. Chapman and Hall/CRC, New York. 10.1201/9780429459016.
- Pepin KM, Eisen RJ, Mead PS, Piesman J, Fish D, Hoen AG, Barbour AG, Hamer S, Diuk-Wasser MA, 2012. 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* 86, 1062–1071. 10.4269/ajtmh.2012.11-0630. [PubMed: 22665620]
- R Core Team, 2023. R: a language and environment for statistical computing (Version 4.3.1). R foundation for statistical computing. [software]. <https://www.R-project.org/>.
- Rand PW, Lubelczyk C, Lavigne GR, Elias S, Holman MS, Lacombe EH, Smith RP, 2003. Deer density and the abundance of *Ixodes scapularis* (Acari: Ixodidae). *J. Med. Entomol* 40, 179–184. 10.1603/0022-2585-40.2.179. [PubMed: 12693846]
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M, 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC. Bioinformatics* 12, 77. 10.1186/1471-2105-12-77. [PubMed: 21414208]
- Rosenberg R, Lindsey NP, Fischer M, Gregory CJ, Hinckley AF, Mead PS, Paz-Bailey G, Watermann SH, Drexler N, Kersh GJ, Hooks H, Partridge SK, Visser SN, Beard CB, Petersen LR, 2018. Vital signs: trends in reported vectorborne disease cases - United States and territories, 2004–2016. *Morb Mortal W* 67, 496–501. 10.15585/mmwr.mm6717e1.
- Schulze TL, Jordan RA, 2003. Meteorologically mediated diurnal questing of *Ixodes scapularis* and *Amblyomma americanum* (Acari: ixodidae) nymphs. *J. Med. Entomol* 40, 395–402. 10.1603/0022-2585-40.4.395. [PubMed: 14680102]
- Schulze TL, Jordan RA, 2022. Daily variation in sampled densities of *Ixodes scapularis* and *Amblyomma americanum* (Acari: Ixodidae) nymphs at a single site-implications for assessing acarological risk. *J. Med. Entomol* 59, 741–751. 10.1093/jme/tjab213. [PubMed: 34994380]
- Schwartz AM, Hinckley AF, Mead PS, Hook SA, Kugeler KJ, 2017. Surveillance for Lyme disease - United States, 2008–2015. *MMWR Surveil Summ* 66, 1–12. 10.15585/mmwr.ss6622a1.
- Sharma Y, Laison EKE, Philippsen T, Ma J, Kong J, Ghaemi S, Liu J, Hu F, Nasri B, 2024. Models and data used to predict the abundance and distribution of *Ixodes scapularis* (blacklegged tick) in North America: a scoping review. *Lancet Reg. Health Am* 32, 100706. 10.1016/j.lana.2024.100706. [PubMed: 38495312]

- Stafford KCI, Carter ML, Magnarelli LA, Ertel S-H, Mshar P, 1998. Temporal correlations between tick abundance and prevalence of ticks infected with *Borrelia burgdorferi* and increasing incidence of Lyme disease. *J. Clin. Microbiol* 36, 1240–1244. 10.1128/jcm.36.5.1240-1244.1998. [PubMed: 9574684]
- Tobler WR, 1970. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr* 46, 234–240. 10.2307/143141.
- Tran T, Prusinski MA, White JL, Falco RC, Vinci V, Gall WK, Tober K, Oliver J, Sporn LA, Meehan L, Banker E, Backenson PB, Jensen ST, Brisson D, 2021. Spatio-temporal variation in environmental features predicts the distribution and abundance of *Ixodes scapularis*. *Int. J. Parasitol* 51, 311–320. 10.1016/j.ijpara.2020.10.002. [PubMed: 33359203]
- U.S. Census Bureau, 2021. Regions and divisions. [https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us\\_regdiv.pdf](https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf). (accessed 14 June 2024).
- U.S. Geological Survey, National Geospatial-Intelligence Agency, 2010. Global multi-resolution Terrain Elevation Data 2010 (GMTED2010). USGS EROS Archive, 10.5066/F7J38R2.
- Vail SC, Smith G, 1998. Air temperature and relative humidity effects on behavioral activity of blacklegged tick (Acari: Ixodidae) nymphs in New Jersey. *J. Med. Entomol* 35, 1025–1028. 10.1093/jmedent/35.6.1025. [PubMed: 9835697]
- Vail SC, Smith G, 2002. Vertical movement and posture of blacklegged tick (Acari: Ixodidae) nymphs as a function of temperature and relative humidity in laboratory experiments. *J. Med. Entomol* 39, 842–846. 10.1603/0022-2585-39.6.842. [PubMed: 12495181]
- Velazco SJE, Rose B, Andrade AFA, Minoli I, Franklin J, 2022. flexsdm: an R package for supporting a comprehensive and flexible species distribution modelling workflow. *Methods Ecol. Evol* 13, 1661–1669. 10.1111/2041-210X.13874.
- Velazco SJE, Rose MB, De Marco P, Regan HM, Franklin J, 2023. How far can I extrapolate my species distribution model? Exploring shape, a novel method. *Ecography*. e06992. 10.1111/ecog.06992, 2024.
- Wang P, Glowacki MN, Hoet AE, Needham GR, Smith KA, Gary RE, Li X, 2014. Emergence of *Ixodes scapularis* and *Borrelia burgdorferi*, the Lyme disease vector and agent. *Ohio. Front Cell Inf Mi* 4, 70. 10.3389/fcimb.2014.00070.
- Weihls C, Ligges U, Luebke K, Raabe N, 2005. klaR Analyzing German Business Cycles, in: Baier D, Decker R, Schmidt-Thieme L (Eds.). Springer-Verlag, Berlin, pp. 335–343.
- Wickham H, 2016. ggplot2: Elegant Graphics For Data Analysis. Springer-Verlag, New York.
- Wickham H, François R, Henry L, Müller K, Vaughan D, 2023a. dplyr: a grammar of data manipulation (R package version 1.1.4). [software]. <https://CRAN.R-project.org/package=dplyr>.
- Wickham H, Vaughan D, Girlich M, 2023b. tidyr: tidy messy data (R package version 1.3.0). [software]. <https://CRAN.R-project.org/package=tidyr>.
- Wilson ML, 1998. Distribution and abundance of *Ixodes scapularis* (Acari: Ixodidae) in North America: ecological processes and spatial analysis. *J. Med. Entomol* 35, 446–457. 10.1093/jmedent/35.4.446. [PubMed: 9701926]
- Wikle CO, 2023. cowplot: streamlined plot theme and plot annotations for ‘ggplot2’ (R package version 1.1.2). [software]. <https://CRAN.R-project.org/package=cowplot>.
- WorldClim, 2020. Bioclimatic variables. v2.1. <https://www.worldclim.org/data/worldclim21.html>.
- Xu G, Wielstra B, Rich SM, 2020. Northern and southern blacklegged (deer) ticks are genetically distinct with different histories and Lyme spirochete infection rates. *Sci. Rep* 10, 10289. 10.1038/s41598-020-67259-0. [PubMed: 32581236]
- Zeileis A, Hothorn T, 2002. Diagnostic checking in regression relationships. *R News* 2, 7–10.
- Zeileis A, Kleiber C, Jackman S, 2008. Regression models for count data in R. *J. Stat. Softw* 27, 1–25. 10.18637/jss.v027.i08.

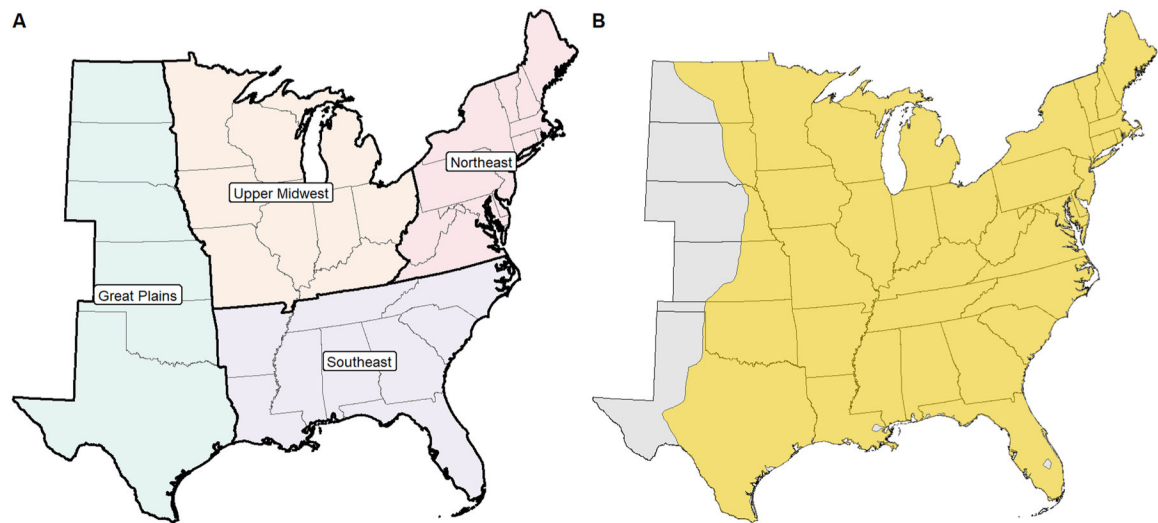


**Fig. 1. Geographic regions and estimated suitable distribution for *I. scapularis* in the eastern United States.**

A) Geographic area of the contiguous United States defined here as the eastern United States. Geographic-based regional naming scheme modified from the U.S. Census Bureau (U.S. Census Bureau, 2021). Thin lines denote state boundaries. B) Estimated distribution for *I. scapularis* ticks (yellow shading) generalized from habitat suitability modeling (Burtis et al., 2022; Hahn et al., 2016, 2017).

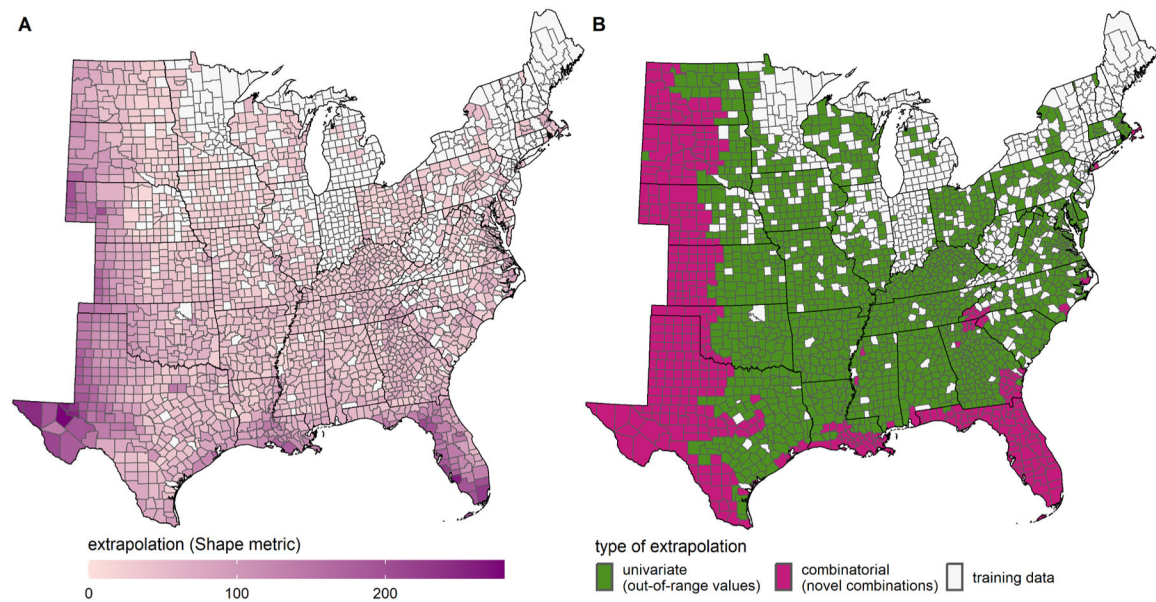


**Fig. 2. Extrapolation between environmental conditions in counties used to fit the zero-inflated negative binomial model (training data) and the other counties in the eastern US to which the model was projected.**

Shape used as a metric of the A) magnitude and B) type of extrapolation (dissimilarity). Larger values of Shape indicate more dissimilar environmental conditions between the training and projected datasets. See Fig. A.4 for univariate similarity for each covariate.

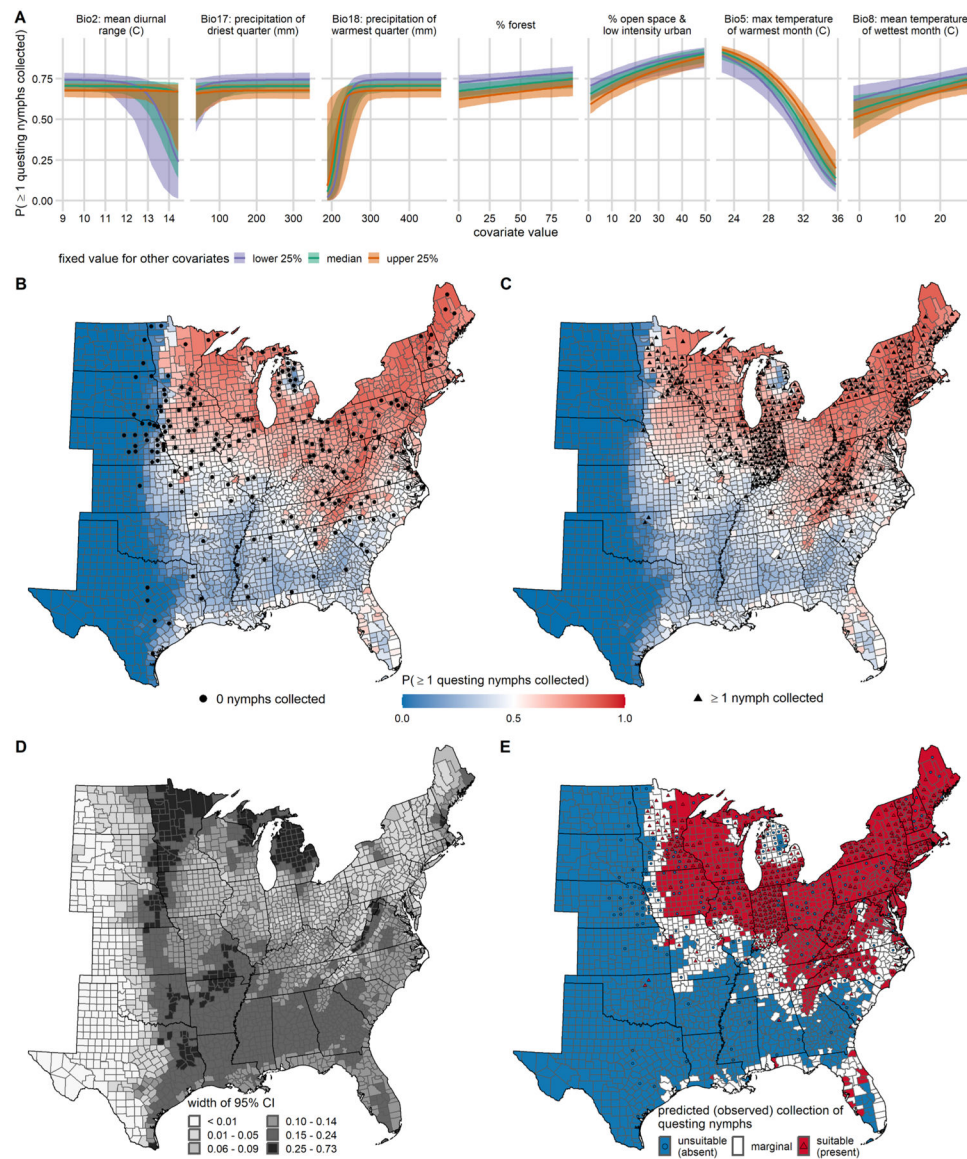


**Fig. 3. Predicted probability of collecting at least one questing *I. scapularis* nymph per 1000 m<sup>2</sup> (collection probability).**

A) Response curves for collection probability for covariates included in the final zero-inflated negative binomial (ZINB) model. The order of panels matches the order of covariates presented in Table 2. To visualize the relationship for each covariate, all other covariates were held at their respective lower 25 %, median, or upper 25 % value while the covariate of interest was varied across its observed range in included counties. County-level predicted collection probability across the eastern US with observed county peak DON of B) zero (dots) or C) at least one (triangles) indicated. D) Width of 95 % confidence interval (CI) for predicted collection probabilities. E) Categorization of collection probability, accounting for the 95 % CI, as unsuitable (<50 % probability), marginal (95 % CI contains 50 % probability), or suitable (>50 % probability), compared to observed DON (present or absent).

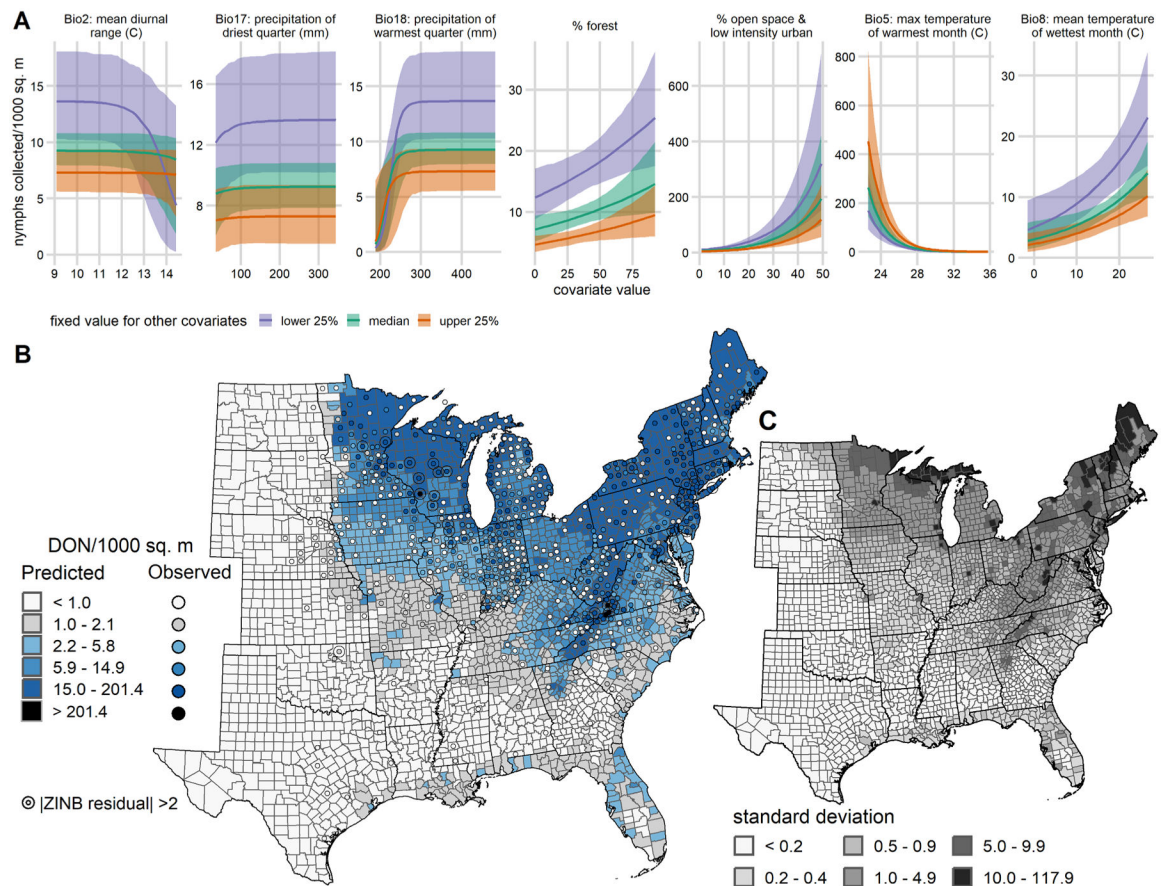


**Fig. 4. Predicted density of *I. scapularis* nymphs (DON) per 1000 m<sup>2</sup>.**

A) Response curves for DON for covariates included in the final zero-inflated negative binomial (ZINB) model. The order of panels matches the order of covariates in Table 2. To visualize the relationship for each covariate, all other covariates were held at their respective lower 25 %, median, or upper 25 % value while the covariate of interest was varied across its observed range in included counties. B) County-level predicted DON compared to observed county peak DON derived from tick surveillance. C) Standard deviation of predicted DON (nymphs per 1000 m<sup>2</sup>).

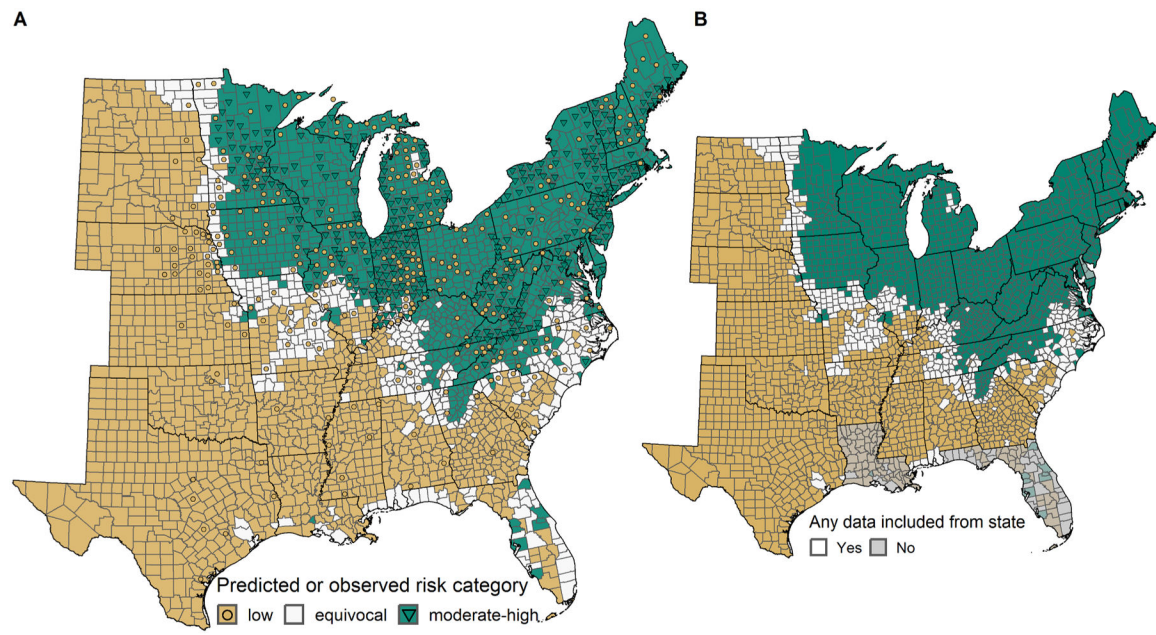


**Fig. 5. Predicted and observed human encounter risk with questing *I. scapularis* nymphs.**

A) Predicted (shading) and observed (shape) encounter risk and B) attenuated predicted risk. Attenuation of risk (grey shading) based on states without any observed data included in the model. Florida and Louisiana are also geographically isolated from contiguous locations of observed collections of questing *I. scapularis* nymphs (Fig. A.6C). Relative risk categories based on quartiles of predicted DON and associated 95 % CI (Fig. 4). The low risk category encompassed predictions for which the upper 95 % CI bound < 2.2 nymphs/1000 m<sup>2</sup>. The moderate-high risk category encompassed predictions for which the lower 95 % CI bound ≥ 2.2 nymphs/1000 m<sup>2</sup>. The equivocal risk category contained predictions for which the 95 % CI contained values in both the low and moderate-high categories.

**Table 1**

Climate and land cover variables included for initial consideration in the zero-inflated negative binomial (ZINB) model.

| Candidate variables for ZINB model                                                | Correlated variable(s) <sup>+</sup> not included in model                                |
|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| Percent of county forest land cover                                               | Percent of county cropland cover                                                         |
| Percent of county low urbanization (open space or low-intensity urban land cover) | Percent of county moderate/high urbanization (medium or high intensity urban land cover) |
| Mean diurnal temperature range <sup>=</sup> (Bio2, °C)                            | –                                                                                        |
| Max temperature of warmest month (Bio5, °C)                                       | Annual mean temperature (Bio1, °C)                                                       |
|                                                                                   | Min temperature of warmest month (Bio6, °C)                                              |
|                                                                                   | Mean temperature of warmest quarter (Bio10, °C)                                          |
|                                                                                   | Mean temperature of coldest quarter (Bio11, °C)                                          |
| Mean temperature of wettest quarter (Bio8, °C)                                    | –                                                                                        |
| Total precipitation of driest quarter (Bio17, mm)                                 | Isothermality <sup>‡</sup> (Bio3, %)                                                     |
|                                                                                   | Temperature seasonality <sup>#</sup> (Bio4, °C)                                          |
|                                                                                   | Min temperature of warmest month (Bio6, °C)                                              |
|                                                                                   | Annual temperature range (Bio7, °C)                                                      |
|                                                                                   | Mean temperature of driest quarter (Bio9, °C)                                            |
|                                                                                   | Mean temperature of coldest quarter (Bio11, °C)                                          |
|                                                                                   | Total annual precipitation (Bio12, mm)                                                   |
|                                                                                   | Total precipitation of driest month (Bio14, mm)                                          |
|                                                                                   | Precipitation seasonality <sup>^</sup> (Bio15, mm)                                       |
|                                                                                   | Total precipitation of coldest quarter (Bio19, mm)                                       |
| Total precipitation of warmest quarter (Bio18, mm)                                | Total precipitation of wettest month (Bio13, mm)                                         |
|                                                                                   | Total precipitation of wettest quarter (Bio16, mm)                                       |
| Mean elevation (m)                                                                | –                                                                                        |

Correlations assessed for each pairwise combination using values from counties included in the model. A quarter is a three-month period. Bioclimatic variables were downloaded from WorldClim (WorldClim, 2020), land cover data from the National Landcover Database (Dewitz and U.S. Geological Society, 2021), and elevation from the U.S. Geological Survey (U.S. Geological Survey and National Geospatial-Intelligence Agency, 2010). For additional details on calculation of bioclimatic variables, see O'Donnell and Ignizio (2012).

<sup>+</sup> Pairwise Pearson correlation coefficient  $\geq |0.65|$ .

<sup>=</sup> Mean of monthly temperature range (i.e., monthly max temperature - monthly min temperature).

<sup>‡</sup> Mean diurnal temperature range (Bio2) / annual temperature range (Bio7)  $\times 100$ .

<sup>#</sup> Standard deviation  $\times 100$ .

<sup>^</sup> Coefficient of variation.

– All pairwise correlation coefficients less than our cutoff ( $\geq |0.65|$ ).

**Table 2**  
Fitted parameter estimates for variables included in the zero-inflated negative binomial (ZINB) model.

| Model piece                    | Covariate                                      | Estimate <sup>#</sup> | Standard error | z-value | P (> z ) |
|--------------------------------|------------------------------------------------|-----------------------|----------------|---------|----------|
| Zero-inflated                  | Intercept                                      | -7.74                 | 2.64           | -2.94   | 0.003    |
|                                | Mean diurnal temperature range (Bio2)          | 1.67                  | 0.63           | 2.65    | 0.008    |
|                                | Total precipitation of driest quarter (Bio17)  | -2.35                 | 1.08           | -2.18   | 0.029    |
|                                | Total precipitation of warmest quarter (Bio18) | -3.48                 | 1.20           | -2.91   | 0.004    |
| Negative binomial <sup>+</sup> | Intercept                                      | 2.36                  | 0.07           | 34.44   | < 0.001  |
|                                | % forest land cover                            | 0.20                  | 0.09           | 2.35    | 0.019    |
|                                | % low urbanization land cover <sup>#</sup>     | 0.57                  | 0.07           | 7.96    | < 0.001  |
|                                | Maximum temperature of warmest month (Bio5)    | -1.16                 | 0.10           | -11.65  | < 0.001  |
|                                | Mean temperature of wettest month (Bio8)       | 0.25                  | 0.07           | 3.40    | < 0.001  |

The zero-inflated piece of the model predicts the probability of absence of questing nymphs using a logistic model and the negative binomial piece predicts the number of questing nymphs collected. See Text A.1 for further mathematical details of the ZINB model.

<sup>#</sup>Covariate data normalized (i.e., have a mean of 0 and standard deviation of 1) prior to model fitting so the magnitudes of parameter estimates are in context of standard deviations.

<sup>+</sup> Estimated shape parameter (θ): 0.38 (standard error: 1.07).

<sup>#</sup> Percent of a county with open space and low intensity urban land cover.

**Table 3**

Contingency table of observed and predicted collection of questing nymphs per 1000 m<sup>2</sup>.

|                                           |                         | Observed collection result (questing nymphs/1000 m <sup>2</sup> ) |                     |                                       |
|-------------------------------------------|-------------------------|-------------------------------------------------------------------|---------------------|---------------------------------------|
|                                           |                         | Absent (0)                                                        | Present (≥1)        | Total                                 |
| Predicted collection probability category | Unsuitable <sup>†</sup> | 59                                                                | 7                   | 66                                    |
|                                           | Marginal <sup>#</sup>   | 45                                                                | 72                  | 117                                   |
|                                           | Suitable <sup>‡</sup>   | 113                                                               | 372                 | 485                                   |
|                                           | Total                   | 217                                                               | 451                 | 668                                   |
| Acc: (59+372)/551 = 78 %;                 |                         | Sens: 372/379 = 98 %                                              | Spec: 59/172 = 34 % | PPV: 372/485 = 77 % NPV: 59/66 = 89 % |

Predicted collection probability (95 % CI) dichotomized using a threshold of 0.5. Observed presence or absence of questing nymphs based on county peak DON calculated from reported tick surveillance data. Last row presents diagnostic metrics calculated using counties predicted as either unsuitable or suitable, not marginal. Acc: accuracy; Sens: sensitivity; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value.

<sup>†</sup>Unsuitable: upper 95 % CI bound for collection probability <50 %.

<sup>#</sup>Marginal: 95 % CI for collection probability contains 50 %.

<sup>‡</sup>Suitable: lower 95 % CI bound for collection probability ≥50 %.

**Table 4**

Contingency table of observed and predicted relative risk categories for human-nymph encounters.

|                                  |                            | Observed relative risk category |                            |                                          |
|----------------------------------|----------------------------|---------------------------------|----------------------------|------------------------------------------|
|                                  |                            | Low <sup>†</sup>                | Moderate-high <sup>‡</sup> | Total                                    |
| Predicted relative risk category | Low <sup>†</sup>           | 55                              | 2                          | 57                                       |
|                                  | Equivocal <sup>#</sup>     | 58                              | 14                         | 72                                       |
|                                  | Moderate-high <sup>‡</sup> | 216                             | 323                        | 539                                      |
|                                  | Total                      | 329                             | 339                        | 668                                      |
| Acc: (55+323)/596 = 78 %         |                            | Sens: 323/325 = 99 %            | Spec: 55/271 = 20%         | PPV: 323/539 = 60 %<br>NPV: 55/57 = 96 % |

Observed risk category based on county peak DON calculated from tick surveillance data and predicted risk based on 95 % CI from the ZINB model. Last row presents diagnostic metrics calculated using counties predicted as either low or moderate-high relative risk, not equivocal. Acc: accuracy; Sens: sensitivity; Spec: specificity; PPV: positive predictive value; NPV: negative predictive value.

<sup>†</sup>Low relative risk: upper 95 % CI bound < 2.2 questing nymphs collected/1000 m<sup>2</sup>.

<sup>#</sup>Equivocal relative risk: 95 % CI contains values above and below 2.2 nymphs/1000 m<sup>2</sup>.

<sup>‡</sup>Moderate-high relative risk: lower 95 % CI bound ≥ 2.2 questing nymphs collected/1000 m<sup>2</sup>.