Ticks and Tick-borne Diseases 14 (2023) 102126

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

journal homepage: www.elsevier.com/locate/ttbdis

Original article

Microhabitat modeling of the invasive Asian longhorned tick (*Haemaphysalis longicornis*) in New Jersey, USA

Ilia Rochlin [a,b,*], Andrea Egizi [a,c], Zoe Narvaez [a], Denise L. Bonilla [d], Mike Gallagher [e], Gregory M. Williams [f], Tadhgh Rainey [g], Dana C. Price [a], Dina M. Fonseca [a,**]

[a] Center for Vector Biology, Rutgers University, New Brunswick, NJ 08901, USA
[b] Department of Microbiology and Immunology, Center for Infectious Diseases, Stony Brook University, Stony Brook, NY 11794, USA
[c] Monmouth County Mosquito Control Division, Tick-borne Disease Program, Tinton Falls, NJ 07724, USA
[d] USDA/APHIS/Veterinary Services, Strategy and Policy, National Cattle Fever Tick Eradication Program, Fort Collins, CO 80526, USA
[e] USDA Forest Service Northern Research Station, New Lisbon, NJ 08064, USA
[f] Tarsal Claw Industries LLC, Milltown NJ 08850, USA
[g] Public Health Entomologists LLC, Milford, NJ 08848, USA

A B S T R A C T

The Asian longhorned tick (*Haemaphysalis longicornis*) is a vector of multiple arboviral and bacterial pathogens in its native East Asia and expanded distribution in Australasia. This species has both bisexual and parthenogenetic populations that can reach high population densities under favorable conditions. Established populations of parthenogenetic *H. longicornis* were detected in the eastern United States in 2017 and the possible range of this species at the continental level (North America) based on climatic conditions has been modeled. However, little is known about factors influencing the distribution of *H. longicornis* at geographic scales relevant to local surveillance and control. To examine the importance of local physiogeographic conditions such as geology, soil characteristics, and land cover on the distribution of *H. longicornis* we employed ecological niche modeling using three machine learning algorithms - Maxent, Random Forest (RF), and Generalized Boosting Method (GBM) to estimate probability of finding *H. longicornis* in a particular location in New Jersey (USA), based on environmental predictors. The presence of *H. longicornis* in New Jersey was positively associated with Piedmont physiogeographic province and two soil types - Alfisols and Inceptisols. Soil hydraulic conductivity was the most important predictor explaining *H. longicornis* habitat suitability, with more permeable sandy soils with higher hydraulic conductivity being less suitable than clay or loam soils. The models were projected over the state of New Jersey creating a probabilistic map of *H. longicornis* habitat suitability at a high spatial resolution of 90×90 meters. The model's sensitivity was 87% for locations sampled in 2017-2019 adding to the growing evidence of the importance of soil characteristics to the survival of ticks. For the 2020-2022 dataset the model fit was 57%, suggestive of spillover to less optimal habitats or, alternatively, heterogeneity in soil characteristics at the edges of broad physiographic zones. Further modeling should incorporate abundance and life-stage information as well as detailed characterization of the soil at collection sites. Once critical parameters that drive the survival and abundance of *H. longicornis* are identified they can be used to guide surveillance and control strategies for this invasive species.

1. Introduction

The Asian longhorned tick (*Haemaphysalis longicornis* Neumann) is a species of medical and veterinary importance in East Asia, Australia, and New Zealand/Oceania, recently established in the United States (Bickerton and Toledo, 2020; Hammer et al., 2015; Heath, 2016; Luo et al., 2015; Rainey et al., 2018; Thompson et al., 2020). In its Australasian range, *H. longicornis* represents a major threat to domestic livestock, heavily parasitizing large ruminants and impeding production of animal products (Heath 2016). It is a vector of *Theileria orientalis* group bovine pathogens, and has recently been implicated in the transmission of the virulent *T. orientalis* Ikeda genotype to US cattle (Dinkel et al., 2021; Thompson et al., 2020). Though humans are not favored hosts of *H. longicornis*, opportunistic feeding is well documented across native and invasive ranges (Bickerton and Toledo, 2020; Wormser et al., 2020). In East Asia, *H. longicornis* vectors severe fever with thrombocytopenia syndrome virus (SFTSV), an emerging human tick-borne disease recently reclassified as Dabie bandavirus (Li et al., 2021; Liu et al., 2015; Luo et al., 2015). Under laboratory conditions, *H. longicornis* has demonstrated vector competency for the closely related Heartland virus, emergent in parts of the US (Raney et al., 2022). Endemic bacterial pathogens, including *Borrelia burgdorferi* sensu stricto, the human

* Corresponding author at: 238 Centers for Molecular Medicine, Stony Brook University, Stony Brook, NY 11794, USA.
** Corresponding author.
E-mail addresses: ilia.rochlin@stonybrook.edu (I. Rochlin), dina.fonseca@rutgers.edu (D.M. Fonseca).

https://doi.org/10.1016/j.ttbdis.2023.102126
Received 2 October 2022; Received in revised form 13 January 2023; Accepted 13 January 2023
Available online 19 January 2023

variant of *Anaplasma phagocytophilum*, and Bourbon virus have been detected sporadically in field collected specimens in the US, although the epidemiological significance of these findings is not yet known (Breuner et al., 2020; Cumbie et al., 2022; Levin et al., 2021; Price et al., 2022, 2021). While *H. longicornis* is currently not perceived as a major public health threat in the US, this status may change given the enormous tick densities reached in favorable habitats (Bickerton et al., 2021; Rochlin et al., 2022). Within its native range, the tick is most abundant in tall grassy transitional areas such as hillside gravesites and farmland bordering forests, which are places often frequented by humans, increasing the risk of tick bites (Coburn et al., 2016; Johnson et al., 2017). High *H. longicornis* abundance was also observed in similar ecotone and open field habitat in the US (Rochlin et al., 2022).

In 2017, the first established *H. longicornis* population was discovered in New Jersey, USA (Rainey et al., 2018). As of April 2022, the US range of *H. longicornis* covers 17 states and 297 counties (USDA APHIS situation report 2022). This tick species appears to be expanding its range southward, westward, and northward to occupy extensive suitable habitat found in eastern North America (Rochlin, 2019) although some of these new detections are likely the result of increased surveillance. The potential continental level distribution of *H. longicornis* has been characterized (Raghavan et al., 2019; Rochlin, 2019), but modeling its localized distribution at scales relevant for both surveillance and control of this species is a more difficult proposition. Climate data without constraints afforded by the use of habitat environmental data provide a ''broad-brush'' picture of suitability for a given species (Sohl, 2014) whereas local environmental data can provide insight into site-level habitat suitability. We hypothesized that *H. longicornis* fine-scale distribution is largely determined by the ecological patterns of natural communities including soil, vegetation, and host species and physiogeographic conditions. Specifically, we observed the appearance of an association between this tick species' distribution in New Jersey and the Piedmont ecological region (Dalton et al., 2014) (Fig. 1). Ecological regions are characterized by geology and topography, natural communities, soils, and hydrologic conditions (Cleland et al., 1997). New Jersey is home to several geological provinces with three major types - Coastal Plain, New England, and Piedmont - occupying most of the state (Cleland et al., 2007, 1997; McNab et al., 2007).

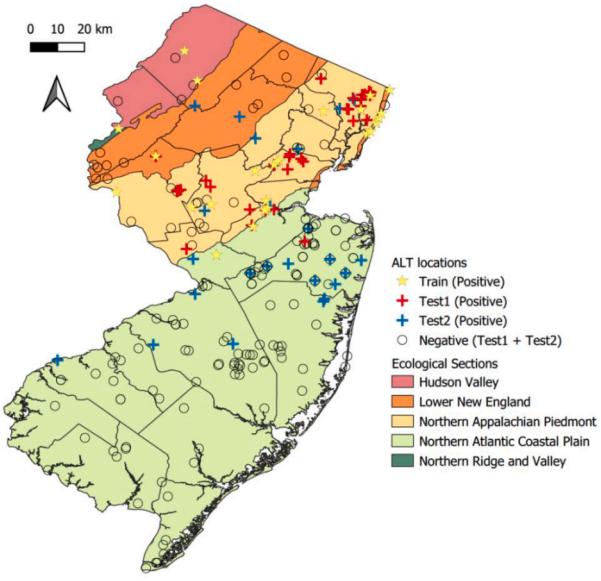

Fig. 1. Ecological sections and tick collections. Ecological sections in New Jersey and surrounding states are shown. New Jersey state and county boundaries are indicated. *Haemaphysalis longicornis* (ALT) locations for training (Train) and both testing (Test1 and Test2) datasets are shown. All negative locations from Test1 and Test2 datasets were combined.

The aim of this study was to identify the environmental parameters and conditions that define known *H. longicornis* habitat in New Jersey on a local scale appropriate for further investigations into this tick species' distribution and for guiding management decisions. We employed ecological niche modeling using three common machine learning algorithms - Maxent, Random Forest (RF), and Generalized Boosting Method (GBM) for modeling species distribution estimated from presence data-only and from environmental variables (Elith et al., 2011; Marmion et al., 2009; Phillips et al., 2006). The individual models were projected over the state of New Jersey creating a probabilistic map of *H. longicornis* habitat suitability at a high spatial resolution of 90×90 m.

2. Materials and methods

2.1. Data sources

We obtained data on *H. longicornis* surveillance in the state of New Jersey (NJ) from multiple independent sources including county, state and federal agencies and several Rutgers University research projects (sources detailed below and in Table 1). While we only used detections of *H. longicornis* from the environment, as opposed to on-host detections of uncertain geographic provenance, we were unable to compare tick density and abundance across datasets due to variable sampling methodology. Therefore, we used only presence/absence data in our analyses. In total, information on 283 sites in all 21 NJ counties was obtained resulting in 275 sites used in the analyses (see Results section for more details). All sites were geocoded to the exact geographic coordinates.

2.1.1. Tick data collections

During 2017-2018 after the Asian longhorned tick was first detected in NJ (Rainey et al. 2018), the US Department of Agriculture (USDA) conducted environmental surveys at multiple sites across NJ using a variety of tick sweeps made of white crib flannel or corduroy cloth pulled beside or behind the operator over the ground and through low vegetation ('USDA dataset'). The cloth was inspected every 1-2 m and attached ticks were removed with forceps and stored in 80-95% ethanol in labeled 1.5-2 ml plastic tubes. Sampling was focused on premises where the Asian longhorned tick had been detected on livestock (USDA APHIS situation report 2022)). Ticks were sent to the National Veterinary Services Laboratories (NVSL, Ames, IA, USA) where experts performed the taxonomic identifications.

In May 2018, faculty and students at the Rutgers Center for Vector Biology developed weekly surveys for *H. longicornis* at 12 sites across three Rutgers New Brunswick campuses (Cook, Douglas, and Livingston, Middlesex county, NJ) to assess the distribution and habitat relationships of *H. longicornis* in urban environments where contact with people is more likely ('RU1 dataset', see Table 1 for tick data). Based on existing information from its native range (Coburn et al., 2016; Johnson et al., 2017) we sampled forest edges and it became immediately evident that *H. longicornis* was especially abundant in bushes and tall grasses on the sides of trails and paths (data not shown). Because it is often difficult to walk through those areas, we chose to sample with a tick sweep, a system where operators walk transects holding the long side of a L-shaped PVC pipe and brush the vegetation and ground directly to the side of the transect with a cloth attached to the shorter side (Egizi et al., 2019a). Three 30.5 m long transects were sampled at each site and ticks were checked every meter. Opportunistic surveys for *H. longicornis* were also developed in August 2018 and then again in June, August and September 2019 at several locations across multiple counties including Atlantic (May's Landing), Salem (Mad Horse Fish and Wildlife, Stony Point), Cumberland (Tindall Island, Sea Breeze, New Sweden Wildlife Management Area) and Warren (Musconetcong Gorge Nature Preserve). During each survey, multiple sites and a minimum of three transects 100 m long were sampled. Ticks were stored in vials with 100% EtOH and identified using established morphological keys (Egizi et al., 2019b;

Table 1
Haemaphysalis longicornis datasets from New Jersey, USA, used in the study.

Dataset Name	Source[1]	Site Type[2]	Years collected	#Records	#Records in analysis[3]	Used for
Train	RU1, NJ, MCMCD	+	2018-2019	29	28	Training
Test1	RU1, NJ, MCMCD	-	2018-2019	110	106	Testing
Test1	USDA	+	2017-2018	32	31	Testing
Test2	RU2, USFS, MCMCD	-	2019-2022	82	80	Testing
Test2	RU2, USFS, MCMCD	+	2019-2022	30	30	Testing

[1] RU1 – Rutgers University, project 1 in New Brunswick, NJ – State of New Jersey, NJ – DOH/DEP sponsored statewide survey, MCMCD – Monmouth County Mosquito Control Division, USDA– US Department of Agriculture, USFS – US Forest Service, RU2- Rutgers University, project 2 statewide survey,
[2] Site type: + positive, *H. longicornis* found, - negative, *H. longicornis* not found
[3] Some records were located outside of environmental predictor coverage and thus were omitted from the analysis

Keirans and Litwak, 1989); specimens identified as *H. longicornis* were barcoded at the cytochrome oxidase locus and integrated in a broader analysis (Egizi et al., 2020).

In August 2018, after a Monmouth County resident submitted a female *H. longicornis* to the county's passive tick surveillance program (USDA APHIS situation report, 2022)), the Monmouth County Mosquito Control Division developed active surveillance to confirm the extent of the species' presence in the county ('MCMCD dataset'). In 2018-2019 surveillance focused on ecotone and/or old field habitat in public parks and recreation areas both near the resident's home and throughout western Monmouth County (because neighboring Middlesex County had known established populations). Sampling was performed only on rain-free days when vegetation was dry, and winds did not exceed 10 km/h between the hours of 0800 and 1300. For this exploratory sampling, three 30.5 m long transects were sampled at each site. In following years (2019-2022) detections of *H. longicornis* were a byproduct of routine active surveillance targeting *Ixodes scapularis* Say and *Amblyomma americanum* (L.), in Monmouth County (Jordan et al., 2022). Sampling was conducted in forest interior sites in public parks and open spaces. Five 100-m transects at each site were sampled on three different days during May-June each year, by working a 1-m^2 cotton duck cloth drag held behind the operator in a straight line and inspecting the drag at 10 m intervals (Jordan et al., 2022). Ticks were removed from drags with forceps and placed in vials with 90% EtOH, identified to species as described.

In 2019 and 2020, the US Forest Service in collaboration with Pennsylvania State University ('USFS dataset') collected data from plots in 27 forest stands of the Pinelands National Reserve in southern New Jersey for 3 days in early June, late July/ early August, and early October using 1 m^2 cotton drags and 1 m^2 sweeps. Plots were circular with a radius of 7.3 m and were located at least 30 m from a road to minimize edge effects. All sampling radiated from plot center and 3 drags and 3 sweeps were conducted along different 7.3 m transects (total of 65.7 m sampled each time at each plot). Ticks were removed from the sampling cloth using a lint roller sheet which was then placed in a Ziploc bag. Ticks were later removed from the lint roller sheets, identified according to species and life stage using a stereo microscope (Leica S8AP0, Leica Microsystems, Wetzlar, Germany) and morphological keys, and stored in vials of 90% ethanol.

From March through November 2019 a NJ Department of Health and NJ Department of Environmental Protection jointly-sponsored survey was conducted at 64 sites in 20 counties across New Jersey to specifically target *H. longicornis* ('NJ dataset'). Potential sites were chosen by the collectors based on their experience in other *H. longicornis* habitat and through local accounts given by county mosquito control personnel. Sites were sampled using tick sweeps constructed with 0.25 m^2 (0.5×0.5 m) crib flannel and a PVC pipe handle (Egizi et al., 2019a). Sampling distance was assessed by measuring the collector's stride length and walking a fixed number of strides prior to checking the sweep. Tick sweeps were inspected every 10 m based on observations that *H. longicornis* does not remain on drags as long as other tick species (Sherpa et al., 2021). A minimum of 750 m^2 was sampled at each location. Any ticks collected were transferred directly from the sweep to sterile vials. Ticks were stored at 0 °C and then identified using the morphological keys mentioned above.

A final dataset originated from a second Rutgers University project ('RU2 dataset') conducting tick surveillance in 30 forested public parklands in north, central and southern New Jersey between May and July 2022, primarily focusing on forested areas but often also including ecotones. Each sampling event involved two hours of dragging with a 1 m^2 crib flannel cloth attached to a 1 m long PVC pipe. Both sides of the drag were checked for ticks every 10 m and ticks were immediately transferred to sterile vials of 90% EtOH, transported back to the lab and identified using the morphological keys mentioned above. Whenever specimen condition was poor, species ID was confirmed using DNA barcoding (Egizi et al., 2020).

2.1.2. Environmental predictors

Ecological, soil, and landscape variables used in this study are listed in Table 2. All data layers are publicly available from governmental or

Table 2
Environmental predictors used in the analysis. All predictors were entered in random forest (RF) and generalized boosted regression (GBM) analyses. Maxent predictors are indicated with a superscript[m]. Abbreviations for the SSURGO soil parameters are identical to those used in the SSURGO database and documentation. For data sources see Supplemental Table S1

Predictors	Description	Abbreviations
Geology[m]	National scale geologic map	geo_nj_90m
Ecological regions[m]	US ecological and physiogeographic sections	ecomap_nj_90m
Tree cover[m]	% Tree cover (all forest types)	treecov_nj_90m
Soil order[m] (STATSGO)	Soil taxonomic order	taxorder_nj_90m
Available water capacity / storage[m] (STATSGO)	The volume of water the soil can store that is available to plants to 25 cm depth	moist_nj_90m
Slope (STATSGO)	Slope grade	slope_nj_90m
Soil group[m] (SSURGO)	Soil taxonomic group	taxgrtgroup
Soil hydric group (SSURGO)	A group of soils having similar runoff potential under similar storm and cover conditions	hydgrp
Saturated Hydraulic Conductivity[m] (SSURGO)	The amount of water that would move vertically through saturated soil under hydraulic gradient	ksat_r
Satiated soil (SSURGO)	Volumetric soil water content at or near zero bar tension, % whole soil	wsatiated_r
Liquid limit (SSURGO)	The water content of the soil at the change between the liquid and plastic states	ll_r
Coarse sand (SSURGO)	Mineral particles 0.5 mm to 1.0 mm as a weight % of < 2 mm fraction	sandco_r
Very fine sand (SSURGO)	Mineral particles 0.05 mm to 0.1 mm as a weight % of < 2 mm fraction	sandvf_r
Sieve #4 soil fraction (SSURGO)	Soil fraction passing a number 4 sieve (4.7 mm square opening)	sieveno4_r
Rock fragments 3 to 10 inch - gravel (SSURGO)	The percent by weight of the horizon occupied by rock fragments 7.6-25.4 cm (3-10 inches)	frag3to10_r

academic sources (see Supplemental Table S1 for sources). Elevation was considered but not used in the analysis due to strong correlation with soil and ecological region variables. All shapefiles were converted to raster files using National Land Cover Dataset 30×30 m resolution and extent and cropped to the state of New Jersey boundaries. The resulting raster files were then resampled to 90×90 m resolution to upscale to coarser resolution layers. All Geographic Information Systems (GIS) processing and analyses were conducted in R statistical software v. R-3.6.1 (R Development Core Team, 2019) and QGIS v.3.4.13 (http://www.qgis.org).

2.2. Environmental predictor selection

Species distribution model performance is sensitive to the choice of modeling parameters with model over-fitting, multicollinearity, and data-dredging (i.e. using large number of environmental layers) negatively affecting the prediction's accuracy (Kuhn, 2008; Merow et al., 2013; Rodda et al., 2011). Therefore, a step-wise variable selection process to address these issues was carried out as follows. The initial dataset included predictors identified by the investigators and partially based on previous studies, as well as fine scale soil variables derived from the USDA SSURGO database using tables muaggatt, chorizon, component, and valu1 (Soil Survey, 2020). Categorical predictors with more than 50 levels, invariable numerical predictors, or those with sparse coverage were excluded from the analysis. Highly correlated numerical variables were identified by Pearson's pairwise correlation coefficient $r > |0.8|$ and removed using function findCorrelation in the package 'caret'.

Further reduction was made by using the variable selection procedure within the package 'Boruta' 6.0.0 (Kursa and Rudnicki, 2010), which establishes variable importance based on their permutations (i.e. "shadow") using a Random Forest algorithm. The 'Boruta' approach was identified as the most powerful of the conventional variable selection methods (Degenhardt et al., 2019; Zhou et al., 2019). The resulting set of predictors (Table 2) was used for preliminary analysis with all three algorithms. Maxent and GBM have well developed variable importance selection that was utilized to further reduce the number of predictors (Ashrafzadeh et al., 2019; Kuhn, 2008). Specifically, only a subset of seven predictors (Table 2) was used by Maxent, with the remainder having null variable and permutation importance. GBM variable thinning was conducted using package 'dismo' (Elith et al., 2008).

2.3. Statistical modeling

Three algorithms were selected for modeling *H. longicornis* habitat suitability in New Jersey. Maxent is a machine learning algorithm for modeling species distribution estimated from presence-only data and environmental variables (Elith et al., 2011; Phillips et al., 2006; Phillips and Dudík, 2008). Compared to other available algorithms, Maxent performance is consistently ranked among the best and it is widely used to model invasive species making it the benchmark for accuracy and performance (Elith et al., 2006; Merow et al., 2013). In particular, Maxent has been able to accurately predict ranges of rare species with only a few presence records (Anderson and Gonzalez, 2011; Pearson et al., 2007) including invasive species with limited collection records (West et al., 2016, p. 201). Other machine learning algorithms, in particular random forest (RF) and generalized boosted regression (GBM) have been used extensively for species distribution modeling (Bucklin et al., 2015; Elith et al., 2008; Gallien et al., 2012; Heikkinen et al., 2012; Marmion et al., 2009). Inclusion of multiple modeling approaches greatly improves the predictive accuracy of species distribution models (Araújo and New, 2007).

Statistical modeling was conducted in R statistical software using package 'dismo' 1.1-4 with Maxent 3.4.1 (Hijmans et al., 2017), and 'caret' 6.0-84 (Kuhn, 2008) for RF and GBM. Utilities for model evaluation were provided in the packages 'dismo', 'ENMeval' 0.3.0 (Muscarella et al., 2014), and 'PresenceAbsence' 1.1.9 (Freeman and Moisen, 2008).

2.4. Modeling process and optimization

Datasets with biased and few localities are especially prone to overfitting because of a high "noise" to "signal" ratio (Anderson and Gonzalez, 2011). Thus, the goal of model tuning is to achieve a balance between a model's complexity and predictive accuracy. An optimal model provides the maximal explanation of calibration data while still maintaining generality – the ability to predict independent data. Therefore, the approach to reduce overfitting used for all three modeling algorithms was to maximize testing AUC (area under the receiver operating curve) while minimizing the difference between training and testing AUC (ΔAUC) (Merow et al., 2013; Radosavljevic and Anderson, 2014; Warren and Seifert, 2011). The model performance using training data and the unseen (testing) data is expected to be similar if no overfitting (i.e., too much noise has been modeled in the training data) or underfitting (i.e., lack of ability to model old and/or new data) has occurred.

This study's goal was to model suitable habitat; as a result, the background for the model can be chosen from locations that the user is interested in contrasting against presences to characterize the discriminating environmental conditions (Merow et al., 2013). The absence points were included based on the presence of other tick species but not *H. longicornis* in the field collected samples. Having both presence and absence data is always preferable, particularly when there is sample selection bias (Piesman and Gray, 1994; Yackulic et al., 2013). The tick location data were split into one training and four testing sets, two with sites where *H. longicornis* was present, and two where this species was absent (Table 1). The purpose of the training set was to create or "train" the model, whereas the independent testing datasets were used to assess model's accuracy.

For Maxent, the model tuning goal is to identify the best combination of features and regularization parameters that protect the model against overfitting (Anderson and Gonzalez, 2011; Merow et al., 2013). In Maxent, linear features are automatically selected for sample size < 10, quadratic features are added for sample size > 10, hinge features are added for samples > 15, and threshold and product features are added when the sample size > 80 (Elith et al., 2011). Hinge features are a special case of linear and threshold features thus making the threshold features redundant in the model (Elith et al., 2011; Merow et al., 2013).

Following these recommendations, package ENMeval was used to compare regularization values (0.5, 1, 2, 4, 5, 10) with select combinations of linear (L), quadratic (Q), product (P), hinge (H), and threshold (T) features ("LQ", "LQP", "H", "QH", "QHP", "LQHPT") on a full set of environmental variables (Table 2). Other settings included using spatial blocks with 5 k-fold random data partitioning. ENMeval quantifies several evaluation metrics including the difference between training and testing AUC. The best feature and regularization combination (i.e. difference between training and testing AUC) were then used to replace default Maxent settings for the final model creation. The Maxent model was run 10 times using cross-validation to estimate variable importance. Unimportant variables with zero percent permutation significance were removed from the set and the best combination of features and regularization for the reduced model were determined as described above. The reduced model was then projected into New Jersey state boundaries to identify areas suitable for *H. longicornis* with a complementary log-log (cloglog) transform to produce an estimate of occurrence probability output providing an estimate between 0 and 1 of probability of presence.

For Random Forest and generalized boosted regression, training control parameters in the package caret were set at 10 cross-validations with 5 repeats each, for tuning parameters the optimization length (other than described below for each method) was set at 15. For RF, the terminal node size determines the tree complexity and thus influences overfitting. Optimal node size (between 1 and 15) was selected based on

maximizing AUC_{test} and minimizing ΔAUC using 0.7 random data partitioning. The best node size and 1000 trees were used to create 10 RF models to capture some of the stochastic variability, and then averaged for the final model.

GBM optimization followed the established protocol using the package dismo (Elith et al., 2008; Hijmans et al., 2017). First, the optimal number of boosting trees was determined using a low learning rate (0.001) and low tree complexity (2) that allowed two-way interactions. The resulting cross-validated model was then used to remove the lowest contributing predictors to simplify the model. The resulting model was used to determine the tree complexity based on the minimum number of observations in the terminal nodes (between 1 and 15) as described for RF. The interaction depth was set to 2, the number of trees was set to 5000 to approximate the optimal number identified in the first step, and the shrinkage (i.e. learning rate) parameter was set at 0.001.

The resulting probabilistic outputs generated by each algorithm were then averaged to produce a single probabilistic map for Maxent, RF, and GBM. A threshold value is applied to convert the probabilistic map to a binary presence-absence classification (Freeman and Moisen, 2008), for example for survey design (Pearson et al., 2007). While there are strong arguments against applying such largely arbitrary thresholding (Merow et al., 2013), it is often required for policy perspective or practical considerations (Liu et al., 2013). Several different thresholds have been identified in literature (Freeman and Moisen, 2008; Liu et al., 2013; Yackulic et al., 2013). Sensitivity represents the percentage of locations where the tick was collected where the model predicted suitability, and specificity is the percentage of sites where the tick was not collected that were classified as unsuitable. Maximizing the sum of sensitivity and specificity, also called True Skill Statistics (TSS) and predicted prevalence equal to the observed prevalence thresholds produced higher sensitivity and kappa, i.e. a measure of how the classification results compare to values assigned by chance, in several studies (Freeman and Moisen, 2008; Liu et al., 2013).

3. Results

3.1. Tick locations

The statewide dataset assembled in 2018-2019 contained 139 tick sampling locations (Table 1). Out of the 139, 110 were negative, i.e., no *H. longicornis* were detected, and 29 locations were positive for *H. longicornis*. Due to gaps in the predictor spatial coverage, four negative locations and one positive were excluded from the analysis resulting in a final tally of 106 negative and 28 positive locations (Test1 negative and Train datasets, respectively, hereafter). Therefore, the Train dataset with 28 New Jersey positive locations was used for the model development and training (species distribution modeling relies on presence only data). The USDA dataset from 2017-2018 contained 32 sites positive for *H. longicornis* with 31 valid locations (Test1 positive dataset hereafter). Therefore, Test1 was a combined dataset containing both positive and negative locations (Table 1) used for model assessment. More recent *H. longicornis* surveys in 2019-2022 were used for additional model evaluation. Out of 112 sites, two locations were excluded from the analysis because of the gap in the predictors resulting in 110 valid locations, 30 of which were positive and 80 negative for *H. longicornis* ticks (Test2 dataset hereafter). All three datasets are provided as Supplemental Table S2.

3.2. Environmental predictor selection

Out of 58 initial environmental variables (see Supplemental Table S1), 11 predictors with too many levels (categorical) or invariable (numeric) were removed during the first selection step. An additional 13 variables were excluded as highly correlated during the second selection step. Of the remaining 35 variables, 20 were filtered out by the final 'Boruta' selection leaving a final set of 15 variables (Fig. 2, Table 2). Out of 15 predictors in the final set, 8 had zero percent contribution or permutation importance in Maxent based on 10 cross-validations and were removed from the model resulting in the set of 7 Maxent predictors (Table 2). GBM selection process did not identify non-contributing predictors, and the full 15 predictor dataset was used to fit both GBM and RF models.

3.3. Environmental predictor importance and responses

Modeling revealed that the environmental predictor importance varied between Maxent and the tree-based algorithms. For Maxent, the permutation importance is measured by randomly permuting the values of each variable and recording the drop in training AUC normalized to percentages. The two most important variables determined by Maxent were saturated hydraulic conductivity (ksat_r: 46%) and soil taxonomic order (taxorder_nj_90m: 31%). The remaining 5 variables (Table 2) contributed less than 10% each. For RF and GBM, the prediction accuracy on the out-of-bag portion of the data is recorded. Then the same is done after permuting each predictor variable. The differences between the two accuracies are then averaged over all trees, normalized by the

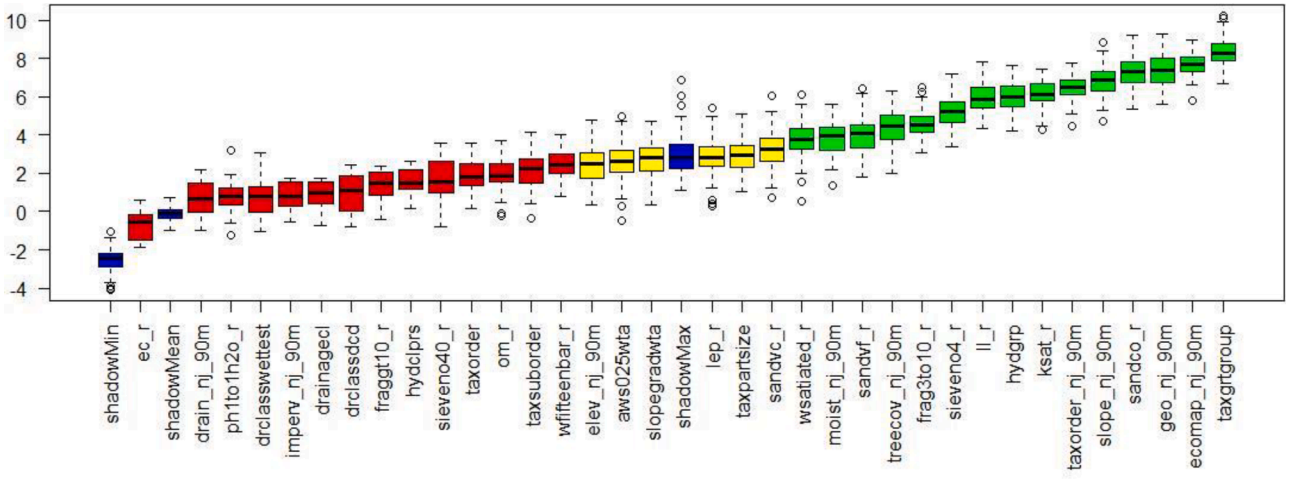

Fig. 2. Environmental predictor selection using package 'Boruta'. Green color designates important predictors, yellow or red color indicates weakly correlated or unimportant predictors excluded from the final set, and blue color indicates shadow attributes created by shuffling original ones. Attributes that have significantly worst importance than shadow ones are being consecutively dropped, while those significantly better than shadows are considered important. Shadows are re-created in each iteration (n = 10,000). For predictor abbreviations and data sources see Table 2 (significant or "green" predictors) and Supplemental Table S1.

standard error, and scaled to have a maximum score of 100. The three most important RF and GBM variables were slope grade (slope_nj_90m: score = 100), hydraulic conductivity (ksat_r: score = 98), and ecological region (ecomap_nj_90m: score = 91). Two predictors with the lowest importance were liquid limit (ll_r; score = 0) and satiated soil (wsatiated_r: score = 11).

The partial dependence or marginal effect curves for numeric predictors show the effect of changing a particular predictor keeping all other variables at their average sample values (Fig. 3). The curves were produced by Maxent and by the package 'pdp' in R for RF and GBM (Greenwell, 2017). The predictor responses exhibited high degree of similarity in different models. For all models, the probability of *H. longicornis* presence declined with saturated hydraulic conductivity (ksat_r), available water capacity (moist_nj_90m), and the proportion of coarse sand (sandco_r), whereas it increased with % tree cover (treecov_nj_90m) (Fig. 3). For RF and GBM models, the probability of *H. longicornis* presence also declined with increased sieve #4 soil fraction (sieve4_r) and it increased with the proportion of very fine sand (sandvf_r) and steeper slope grade (slope_nj_90m). The relationship with less important predictors - rock fragments 3 to 10 inch or 7.6-25.4 cm (frag3to10_r) and satiated soils (wsatiated_r) were not as uniform, but generally showed a decline of the tick's presence probability.

A consistent agreement was also exhibited among the three models regarding the contributions of categorical environmental predictors (Fig. 4). For ecological regions (ecomap_nj_90m), Northern Appalachian Piedmont (17 negative / 16 positive) had the largest positive effect on the probability of presence containing 55% of positive locations, whereas Northern Atlantic Coastal Plain (76 negative / 2 positive) had the largest negative effect. For geology (geo_nj_90m), those categories that had a positive effect on the tick presence were identified as more important in the models, specifically, basalt (1 negative / 4 positive), diabase (0 negative / 7 positive), sandstone (2 negative / 4 positive), and siltstone (11 negative / 7 positive) containing 76% of the positive locations altogether. For soil hydric group, group C (soils with moderately high runoff potential that are 20-40% clay) (18 negative / 16 positive) had the largest positive effect on the probability of presence, followed by group D (soils with high runoff potential containing > 40 % clay) (6 negative / 4 positive).

Soil taxonomy was selected by all models as highly important. For soil taxonomic orders (taxoder_nj_90m), Alfisols (14 negative / 17 positive) and Inceptisols (0 negative / 5 positive) had the largest positive effect on the probability of *H. longicornis* presence followed by rock outcrop (2 negative / 2 positive) and urban land (3 negative / 2 positive). Conversely, Entisols (37 negative / 2 positive) and Ultisols (44

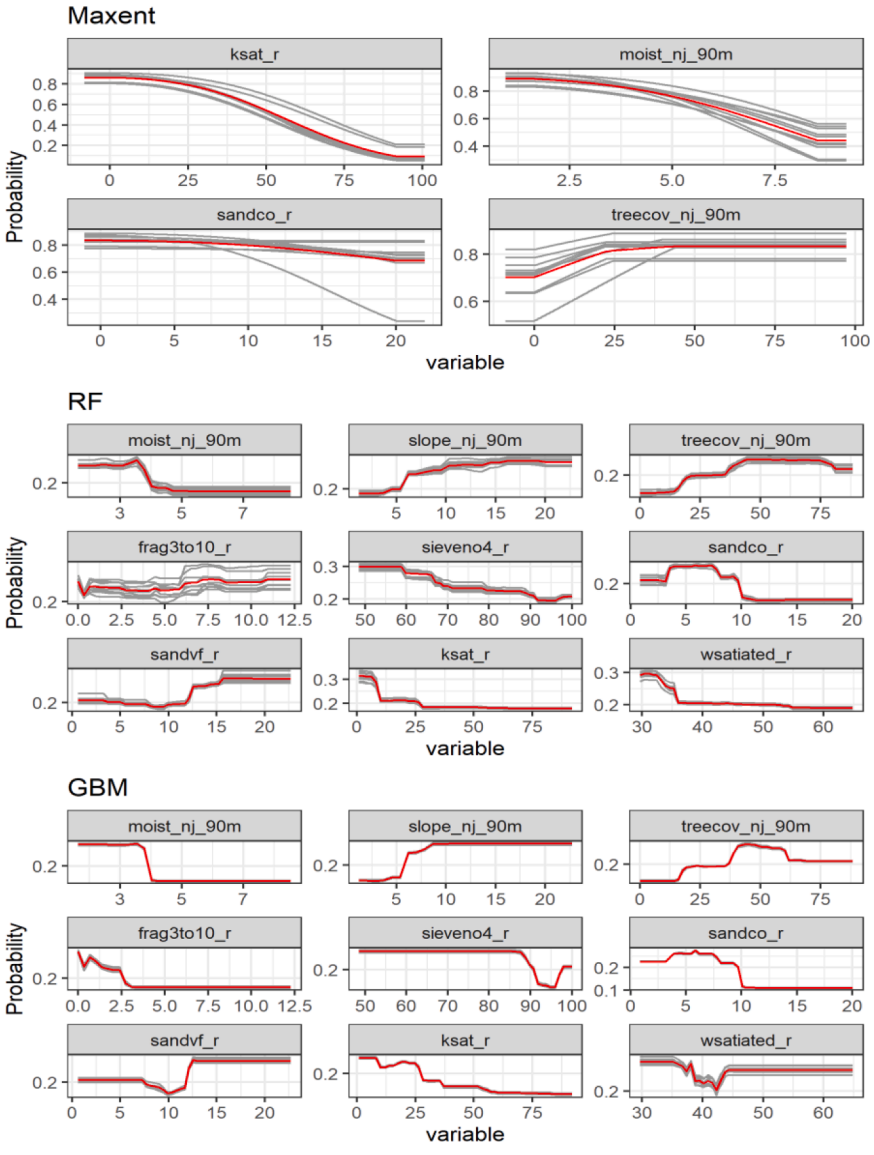

Fig. 3. Averaged response curves of numerical environmental predictors used to model the potential habitat of the Asian longhorned tick (*Haemaphysalis longicornis*) in New Jersey, USA. Grey curves correspond to each of the 10 cross-validated models and indicate the variability, whereas red curves indicate the average response. Response curves do not account for interactions between the variables. Response curves for 'll_r' are not shown because of low importance in the models. For predictor abbreviations see Table 2. Maxent, RF (Random Forest), and GBM (Gradient Boosting Machines) indicate three different algorithms used in the study.

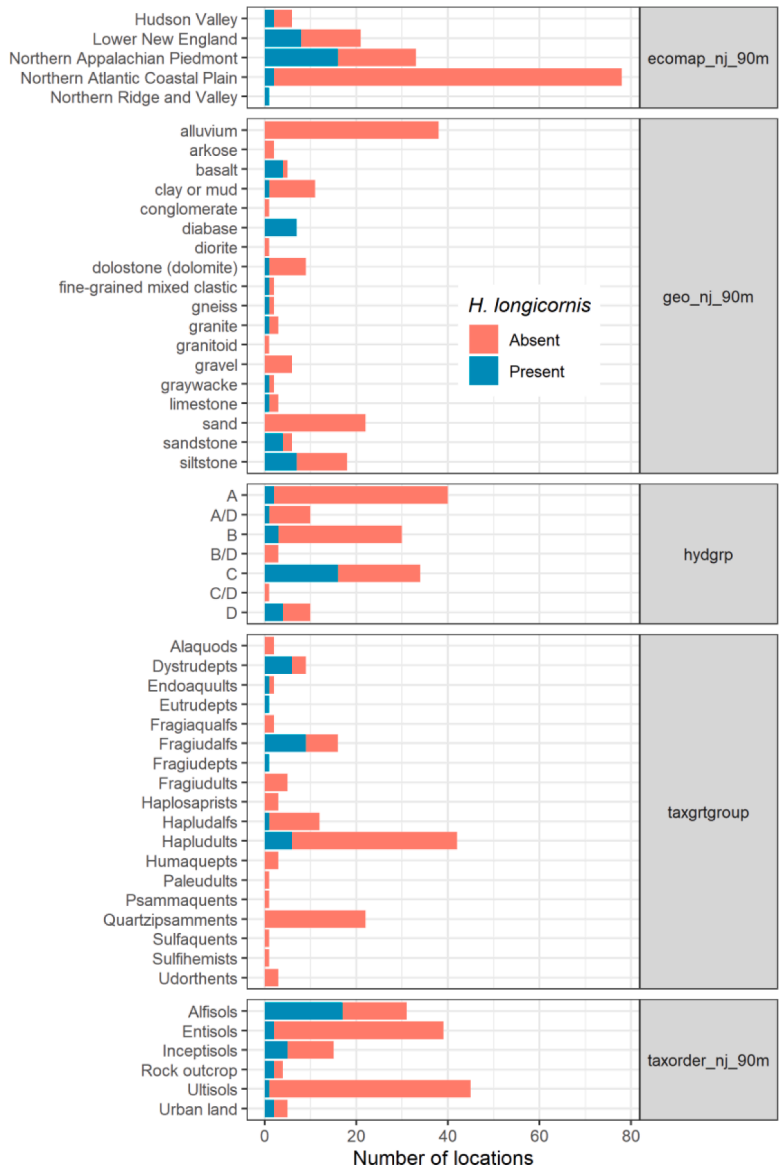

Fig. 4. Distribution of *Haemaphysalis longicornis* positive and negative locations (Train and Test1 datasets) in New Jersey, USA, in relation to categorical environmental predictors used in modeling. For predictor abbreviations see Table 2.

negative / 1 positive) had a negative effect on the probability of the tick's presence. For soil taxonomic groups (taxgrtgroup), positive effects on the probability of presence were identified as more important, with Fragiudalfs (7 negative / 9 pos.) and Dystrudepts (3 negative / 6 positive) having the largest effect followed by Hapludults (36 negative / 6 positive) to a lesser extent.

3.4. Model selection and output

Maxent model selection was carried out using recommended protocols to avoid data dredging and collinearity (Merow et al., 2013; Pearson et al., 2007; Warren and Seifert, 2011). The model selection process sought to improve the models by removing the least important variables (Table 2), changing the default features, and increasing the regularization parameter. The best fitting model included hinge, quadratic, and product features combined with a higher regularization parameter of 4.0, which was in line with previous findings (Anderson and Gonzalez, 2011). The final model had similar $AUC_{train} = 0.844 \pm 0.007$ and $AUC_{test} = 0.831 \pm 0.077$ (mean±SD based on 10 replicates) indicating a lack of overfitting (Merow et al., 2013). The omission rate (proportion of test points not predicted) for the lowest training presence was 0% and for the 10th percentile training presence it was 8.2% ± 0.001 as expected in a model with a moderate to good fit to the data (Radosavljevic and Anderson, 2014), and which were significantly better than random predictions at $P < 0.001$ by binomial test.

Both tree-based models, RF and GBM, contained all 15 predictors (Table 2) as GBM variable selection (Elith et al., 2008; Hijmans et al., 2017) failed to identify non-important predictors. For RF, the selected terminal node size was set to 10, which resulted in the model with $AUC_{train} = 0.992 \pm 0.006$ and $AUC_{test} = 0.907 \pm 0.052$ (mean±SD based on 10 replicates). For GBM, the number of minimal observations was set at 14 resulting in the model with $AUC_{train} = 0.974 \pm 0.015$ and $AUC_{test} = 0.904 \pm 0.059$ (mean±SD based on 10 replicates).

The models were projected into New Jersey's environmental conditions (Fig. 5). Continuous maps produced by the models were converted to binary distribution maps, using two threshold-dependent measures. Maximum specificity and sensitivity, also known as true skill statistic, TSS (Freeman and Moisen, 2008) was a more permissible threshold with the following cutoff probability values, Maxent = 0.410, RF = 0.375, and GBM = 0.190. The cutoff values of TSS/Maximum sensitivity and

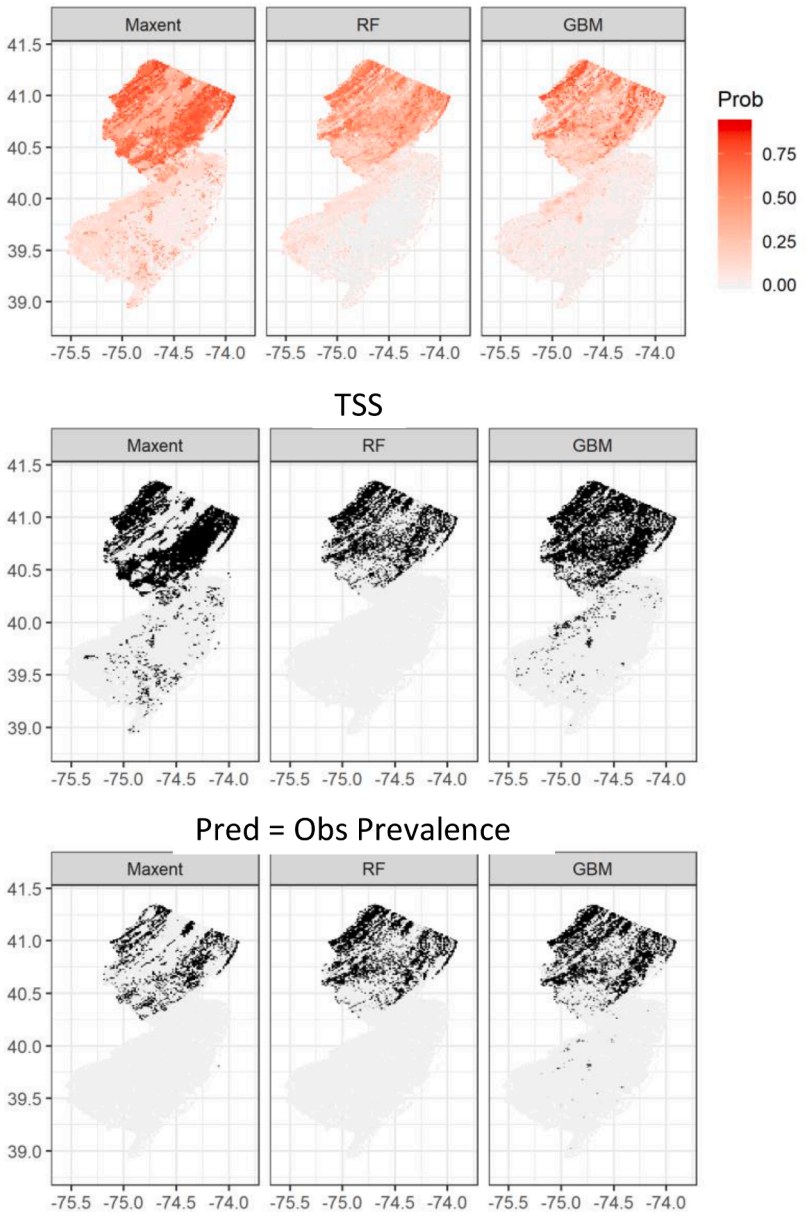

Fig. 5. *Haemaphysalis longicornis* projected habitat suitability in New Jersey, USA. The output for all three models, Maxent, Random Forest (RF), and generalized boosted regression (GBM) is shown. Top row: probabilistic maps with a range from 0 (not suitable) to 1 (highly suitable). Middle row: binary maps indicating presence (black) and absence (grey) produced by maximizing specificity and sensitivity, also known as true skill statistic, TSS (Freeman and Moisen, 2008). Bottom row: binary maps indicating presence (black) and absence (grey) produced by applying the threshold with the predicted prevalence equal to the observed prevalence.

specificity were very similar to the 10th percentile presence threshold (Pearson et al., 2007; Radosavljevic and Anderson, 2014). A stricter presence threshold of the predicted prevalence equal to the observed prevalence (Freeman and Moisen, 2008) was also applied with the following cutoff values: Maxent = 0.710, RF = 0.425, and GBM = 0.345.

3.5. Model comparison and assessment

All three models identified northern New Jersey as the most suitable habitat for *H. longicornis* (Figs. 5 and 6). To compare the model outputs, package 'dismo' niche similarity metric *I* ranging from no similarity = 0 to identical = 1 was used (Warren et al., 2008). The environmental niches produced by the three models largely overlapped with high degree of similarity: $I = 0.927$ for Maxent vs. RF, $I = 0.931$ for Maxent vs. GBM, and $I = 0.968$ for RF vs. GBM. Despite this general similarity, several important differences were noted among the model outputs. A subset of 7 out of 15 predictors in Maxent (Table 2) generated a smoother probability surface with two distinct bands approximately corresponding to Northern Appalachian Piedmont and Hudson Valley ecological regions, with an area of lower suitability dividing the two. Maxent also identified scattered areas of suitable habitat in western and southern New Jersey, although the probability was generally lower (Fig. 5 comparing TSS versus prevalence). Both tree based algorithms, RF and GBM, used a full set of 15 predictors generating maps that predicted suitable habitat more uniformly across the northern New Jersey, but with much more granularity or patchiness at the local scales. GBM was also more like Maxent in identifying suitable areas in southern and, in particular, western New Jersey. In contrast, RF classified the entire southern part of New Jersey as unsuitable.

The consensus probabilistic map indicating suitable habitats for *H. longicornis* was created by averaging the projections made by the three algorithms (Marmion et al., 2009). The average probability map combining the three models and the overlapping binary outputs is shown in Fig. 6. The same figure also shows the Train and Test1 site locations (Table 1).

Model accuracy was assessed using two testing datasets, Test1 and Test2 (see Table 1 for description). The probability scores for the locations where Asian longhorned ticks were detected declined from 0.651 $\pm$ 0.155 for the train set to 0.39 $\pm$ 0.178 for Test1 set, to 0.2 $\pm$ 0.177 for Test2 set (all pairwise comparisons $P < 0.001$, Mann Whitney U test).

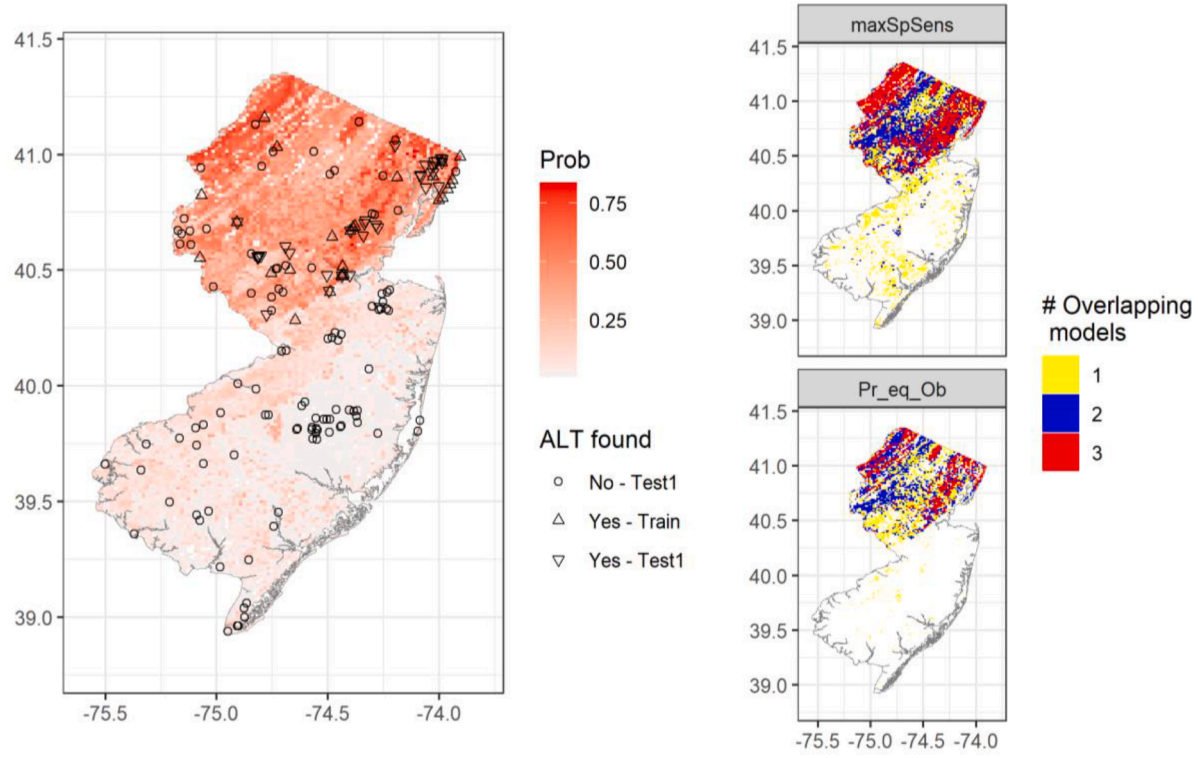

Fig. 6. Combined model projections for *Haemaphysalis longicornis* suitable habitat in New Jersey, USA. Left panel: averaged probabilistic model. Asian longhorned tick (ALT) locations used for developing or training the model (No – other tick species but not ALT were found – used as Test1, Yes – ALT found, used as Train) and testing ALT positive locations (Yes- ALT found, used as Test1) from an independent dataset are shown. Right panel: combined binary outputs (see Fig. 5) indicating ALT presence agreement among the three models. Two thresholds were used: maximum specificity and sensitivity, also known as true skill statistic, TSS (maxSpSens) and a stricter predicted prevalence equal to the observed prevalence (Pr_eq_Ob).

Whereas all of Train set locations were accurately predicted as "presence" under the maximum specificity and sensitivity threshold, i.e. 100% sensitivity (Fig. 7), 27 out of 31 (87% sensitivity) and 17 out of 30 (57% sensitivity) were correctly classified in Test1 and Test2 datasets, respectively. A more stringent equal predicted and observed prevalence threshold resulted in 93% sensitivity for the Train set (26 out of 28), 48% sensitivity (15 out of 31) for Test1 set, and 27% sensitivity (8 out of 30) for Test2 set. The model's specificity, i.e. the ability to predict negative locations from which the Asian longhorned ticks were absent was 69% (73 out of 106) for Test1 set, and 82% (41 out of 50) for Test2 set under maximum specificity and sensitivity threshold, and 85% (90 out of 106) for the Test1 set, and 100% (50 out of 50) for the Test2 set under equal predicted and observed prevalence threshold (Fig. 7). For TSS threshold, the positive predictive value (PPV) was 0.93 (Test1) vs. 0.57 (Test2) and the negative predictive value (NPV) was 0.68 (Test1) vs. 0.82 (Test2 set). A more stringent equal predicted and observed

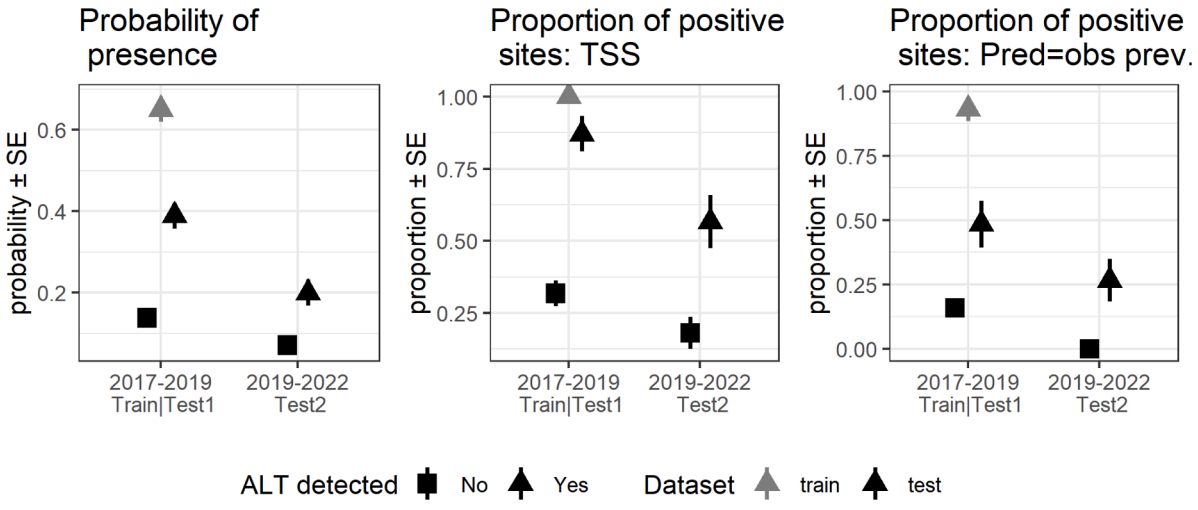

Fig. 7. Model assessment. Train dataset (presence only) was used to develop the model, whereas two test datasets, one from 2017-2019 (Test1) and the second one from 2019-2022 (Test2) were used to evaluate the model. All three models were combined. Probability of presence (continuous scale) determined for points where *Haemaphysalis longicornis* (ALT) was present or absent in New Jersey, USA. Maximum specificity and sensitivity or equal predicted and observed prevalence are on categorical scale (yes or no) and were evaluated based on the proportion of points falling into the "presence" category.

prevalence threshold resulted in PPV of 0.7 (Test1) vs. 0.26 (Test2) and NPV of 0.84 (Test1) vs. 1.0 (Test2 set).

4. Discussion

Ecological zones exhibit similar patterns in: (1) geology and topography, (2) biotic communities, (3) soils, (4) hydrologic function, (5) climate, and (6) natural processes such as nutrient cycling, productivity, succession, and disturbance regimes (Cleland et al., 1997). All three models identified northern New Jersey as the area with the highest habitat suitability for *H. longicornis* with 26 out of 28 total positive locations (Fig. 1). Of note, the first detection of an established population *H. longicornis* in the US was within this region, (Rainey et al 2018), which is broadly included in the Eastern broadleaf forest province subject to more continental climate with vegetation characterized by tall deciduous broadleaf forest often dominated by oak-hickory cover (McNab et al., 2007). *Haemaphysalis longicornis* higher abundance and association with certain ecological communities, i.e. forested to shrubby or mixed habitats, roughly similar to eastern broadleaf forest habitat in the US, have been observed in many previous studies (Heath, 2016; Li et al., 2016; Neilson, 1980; Oh et al., 2009; Tsunoda and Tatsuzawa, 2004).

Eastern broadleaf forest ecological province is further divided into four sections, Northern Appalachian Piedmont, Lower New England, Hudson Valley, and Northern Ridge and Valley (Cleland et al., 2007, 1997; McNab et al., 2007). Among those, Piedmont clearly stood out having over one-half of all *H. longicornis* positive locations (Fig. 1). The terrain forms a Piedmont peneplain, i.e. gently rolling hills with erosional features. These land cover and topographic features are reminiscent of China's more mountainous parts of Henan, Hubei, and Anhui provinces, which are considered prime *H. longicornis* habitat in the epicenter of the Dabie bandavirus outbreak vectored by this tick species (Liu et al., 2015). This affinity of *H. longicornis* for wooded hilly landscape was captured on the models by positive association of slope grade and tree cover with *H. longicornis* habitat suitability (Fig. 3).

Piedmont's geology is complex, consisting of bedrock, shales, siltstones, and sandstones that are also present in other ecological sections in northern New Jersey (Dalton et al., 2014; McNab et al., 2007). Not surprisingly, geology was also a strong contributor to *H. longicornis* presence with basalt, diabase (especially prominent in the diabase sill of New Jersey Palisades), sandstone, and siltstone identified as most suitable habitat containing 22 or 76% of the *H. longicornis* collections in the training dataset and 30 or 93% of the positive sites in the Test1 dataset. While only 7 or 22% of the *H. longicornis* positive sites in the Test2 dataset (2019-2022) were within Piedmont, most positive locations were in close proximity to this ecological region (Fig. 1). Geology may also serve as a proxy for other microhabitat factors not included in the current model.

The southern part of New Jersey is included within the Coastal Plain Mixed Forest province with different geology, soil, and land cover. Its single ecological section, Northern Atlantic Coastal Plain, contained the majority of *H. longicornis* negative locations (76 out of 110, 70%) where other tick species, but not *H. longicornis*, were found prior to 2020. This ecological region is subject to humid, maritime climate with forest vegetation dominated by conifers (McNab et al., 2007). The terrain is a flat alluvial plain with soils formed by recent marine deposits of sands and shales. Coastal plain ranked as low suitability *H. longicornis* habitat: its dominant geological compositions alluvium, sand, and gravel (Dalton et al., 2014; McNab et al., 2007) were devoid of *H. longicornis* positive locations, despite comprising 48% of collection sites in the Test1 dataset (66 negative sites/137 total sites) (Fig. 4). These broadly defined distribution patterns were similar to some areas of northern China, where *H. longicornis* ticks were mostly found in shrubland and broadleaf forested areas, but not in coniferous forests (Zheng et al., 2012). More recently (since 2020) *H. longicornis* were found in the northern, transitional part of Northern Atlantic Coastal Plain suggesting that the range of this species is expanding into new habitat types from its original establishment in northern New Jersey (Fig. 1). Similar *H. longicornis* distribution was observed on nearby Long Island, New York, where this species was much more abundant in the Appalachian/Piedmont ecological zone in comparison to Coastal Plain/Pine Barrens, however incursions of *H. longicornis* into less favorable Pine Barrens were also observed (Rochlin et al., 2022).

Not much information is available on *H. longicornis* associations with different soils. Types of soils are classified based on their physical and chemical characteristics into soil orders, suborders, groups and other taxonomic units (Soil Survey Staff, 1999). In this study, presence of *H. longicornis* was strongly associated with two soil taxonomic orders, Alfisols and Inceptisols. Alfisols are formed under forest or grass vegetation, have distinct clay accumulation in the subsoil, and are rich in nutrients, fertile, and productive. In the US, Alfisols are common in the Ohio and Mississippi river valleys ("NRCS Soils," 2023). One particular group of Alfisols, Fragiudalfs, contained about 30% of all positive *H. longicornis* locations. Fragiudalfs are found in gently sloping terrain under forest vegetation, and are characterized by a fragipan that is a dense subsurface soil layer severely restricting water flow thus reducing drainage and increasing soil moisture (Soil Survey Staff, 1999). Inceptisols, the second soil taxonomic order closely associated with *H. longicornis* presence are those soils that are in the beginning stages of soil profile development commonly found in mountainous terrain, especially in southern New England, the Appalachian Mountains, and the Allegheny Plateau ("NRCS Soils," 2023). One group within this order, Dystrudepts, contained over 20% of all positive *H. longicornis* locations. These are acidic soils with lower pH that are freely drained but remain moist throughout the year due to abundant rainfall supporting either forested or agricultural land uses (Soil Survey Staff, 1999).

The dry and sandy soils of southern New Jersey appeared less suitable as *H. longicornis* habitat. Entisols are the soils of unstable environments, such as floodplains and sand dunes commonly found at the site of recently deposited materials (e.g., alluvium), or in parent materials resistant to weathering (e.g. sand) (Soil Survey Staff, 1999). The sandy Entisols group known as Psamments were especially unsuitable for *H. longicornis*. Quartzipsamments (quartz sand Psamments) contained about 15% of all tick collections but not a single positive *H. longicornis* location. The low suitability of sandy and gravelly soils for *H. longicornis* is also corroborated by the negative influence of the coarse sand and gravel sized rocks as opposed to more loamy or clay soils with a higher proportion of finer mineral particles (very fine sand) that are more plastic and not easily separated in a sieve (Fig. 3).

Another soil taxonomic order occupying most of the southeastern US ("NRCS Soils," 2023), Ultisols, presented a more complex picture. Ultisols are acidic weathered low-fertility soils of warm and humid climates. While when considered at larger aggregated scales (STATGO, Fig. 4, taxorder_nj_90m), Ultisols appeared unsuitable for *H. longicornis*, the analysis at finer smaller spatial scales and finer resolution (SSURGO, Fig. 4, taxgrtgroup) revealed that Hapludults, a common Ultisols group of southeastern US, contained about 20% of positive *H. longicornis* locations. However, the locations where *H. longicornis* was found represented a much smaller fraction of the total within this soil group (Fig. 4) compared to the more favorable Alfisols and Inceptisols suggesting generally lower or patchier habitat suitability.

Although Hapludults are well drained, these soils usually receive abundant rainfall throughout the year (Soil Survey Staff, 1999). Humidity and dehydration have been identified as major factors affecting *H. longicornis* survival (Heath, 2016). The consistency of rainfall in supporting water content as well as the hydrological soil features are likely among the most critical factors in defining the environmental conditions for *H. longicornis* (Rochlin, 2019). Approximately 70% of all positive *H. longicornis* locations were from the soil hydrologic groups C (55%) and D (14%). These soils have moderate to high runoff potential because the movement of water through the soil is restricted as their composition with 20 to >50% clay tends to retain water (NRCS, 2007).

The critical role of soil hydrologic characteristics is underscored by the importance of saturated hydraulic conductivity in all models (Fig. 3). Soil hydraulic conductivity describes the ease with which pores in a saturated soil transmit water. Clay soils that support *H. longicornis* habitat have low saturated hydraulic conductivity, which is much higher in sandy soils. Other soil parameters related to hydrologic properties of the soil such as available water capacity, satiated soil, and liquid limit generally followed the same pattern.

Applying the true skill statistics (TSS) as a threshold (Freeman and Moisen, 2008), the model sensitivity was initially as high as 87% when tested with an independent dataset collected within the same period (2017-2019). However, it declined to 57% for the more recent dataset collected in 2020-2022. This decline in accuracy predicting positive (i.e. presence) locations can have several explanations including the possibility that *H. longicornis* is still in the process of expanding its range and occupying the entirety of suitable habitat. Therefore, the model developed at the initial stages of the invasion may not have captured all relevant microhabitat characteristics (Srivastava et al., 2019). Another possible explanation for poorer model sensitivity using test data from recent years is potential impacts of sampling bias on model performance. Some environments or locations were likely oversampled due to accessibility or other factors, whereas other may have been undersampled. As the later datasets were largely generated by different agencies/-personnel, they could have been biased in different ways than the earlier datasets yielding reduced sensitivity.

The distribution of the ALT appears to be expanding in NJ, northwards into suitable habitat as well as southwards into less suitable habitat (Fig. 1). However, the populations of *H. longicornis* are still largest at the original sites (data not shown), a feature not exploited by the current presence/absence modeling approach. Overall, even with reduced sensitivity, the model predicts *H. longicornis* presence significantly better than random ($X^2 = 11.1$, $P < 0.001$) and retained high specificity, i.e., the ability to differentiate locations not suitable for *H. longicornis* (82%) in 2020-2022. Therefore, in its current format the model remains useful for surveillance, indicating that areas with xeric soils and coniferous vegetation should be de-prioritized. While the generated maps might be appropriate for some management decisions (for example to protect livestock in high-risk areas), better accuracy is needed to inform preventative interventions such as treating recreational or residential areas. As with all modeling efforts, they are meant to be used as guidance but assessing the situation on the ground is still recommended before investment in these types of interventions.

It is possible that the model did not capture all factors contributing to *H. longicornis* habitat suitability. One such factor might be anthropogenic influence on this species' occurrence and detection in new areas (Egizi et al., 2020). The "urban corridor" of New Jersey, an area with the greatest human population density and connected by major highways, aligns closely with the Piedmont region. This fact raises sampling biases due to higher propagule pressure as well as higher likelihood of detection (Pearson, 2007; Phillips and Dudík, 2008; Yackulic et al., 2013). As a result of its high population density, the Piedmont region is also where risk of human, pet, and/or livestock exposure to *H. longicornis* bites and potential pathogen transmission is expected to be among the highest in the state.

In conclusion, at present there are still considerable knowledge gaps with regard to environmental predictors that define the microhabitat distribution of *H. longicornis*. As mentioned, presence/absence models also do not reflect species abundance, which can be much higher in optimal microhabitats resulting in a higher rate of tick bites and risk of pathogen transmission. This study also highlights the utility and need for geographically representative sampling of tick populations (i.e., sampling outside areas already known to be suitable) to improve the ability to accurately predict species distributions in future studies. More accurate range and population density predictions will provide better guidance for local public health and vector control agencies to mitigate the impact of this species on veterinary and public health.

Declaration of Competing Interest

The authors declare no competing interests

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported in part through Cooperative Agreement Number U01CK000509 between the Centers for Disease Control and Prevention (CDC) and the Northeast Regional Center for Excellence in Vector Borne Diseases subcontracted to Rutgers University (D.M.F.). The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Department of Health and Human Services. The 2018-2019 statewide surveillance performed by T.R and G.M.W was funded by a CDC Epidemiology and Laboratory Capacity (ELC) Cooperative agreement awarded to New Jersey Department of Health (NJDOH) in partnership with New Jersey Department of Environmental Protection (NJDEP). Collections by Z.N. were supported by a USDA-NIFA McIntire-Stennis capacity grant (NJ08330) to D.C.P. We would like to acknowledge Robert A. Jordan (MCMCD) for the sampling effort in Monmouth County and the support of the Monmouth County Board of County Commissioners. We thank the numerous USDA, New Jersey Department of Agriculture, and other agency personnel that were instrumental in the surveillance, detection, and records management for *H. longicornis* in New Jersey.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ttbdis.2023.102126.

References

Anderson, R.P., Gonzalez, I., 2011. Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. Ecol. Model. 222, 2796–2811. https://doi.org/10.1016/j.ecolmodel.2011.04.011.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47. https://doi.org/10.1016/j.tree.2006.09.010.

Ashrafzadeh, M.R., Naghipour, A.A., Haidarian, M., Kusza, S., Pilliod, D.S., 2019. Effects of climate change on habitat and connectivity for populations of a vulnerable, endemic salamander in Iran. Glob. Ecol. Conserv. 19, e00637. https://doi.org/10.1016/j.gecco.2019.e00637.

Bickerton, M., McSorley, K., Toledo, A., 2021. A life stage-targeted acaricide application approach for the control of *Haemaphysalis longicornis*. Ticks Tick Borne Dis. 12, 101–581. https://doi.org/10.1016/j.ttbdis.2020.101581.

Bickerton, M., Toledo, A., 2020. Multiple pruritic tick bites by Asian Longhorned tick larvae (*Haemaphysalis longicornis*). Int. J. Acarol. 46, 373–376. https://doi.org/10.1080/01647954.2020.1805004.

Breuner, N.E., Ford, S.L., Hojgaard, A., Osikowicz, L.M., Parise, C.M., Rosales Rizzo, M. F., Bai, Y., Levin, M.L., Eisen, R.J., Eisen, L., 2020. Failure of the Asian longhorned tick, *Haemaphysalis longicornis*, to serve as an experimental vector of the Lyme disease spirochete, *Borrelia burgdorferi* sensu stricto. Ticks Tick Borne Dis. 11, 101–311. https://doi.org/10.1016/j.ttbdis.2019.101311.

Bucklin, D.N., Basille, M., Benscoter, A.M., Brandt, L.A., Mazzotti, F.J., Romañach, S.S., Speroterra, C., Watling, J.I., 2015. Comparing species distribution models constructed with different subsets of environmental predictors. Divers. Distrib. 21, 23–35. https://doi.org/10.1111/ddi.12247.

Cleland, D.T., Avers, P.E., McNab, W.H., Jensen, M.E., Bailey, R.G., King, T., Russell, W. E., 1997. National hierarchical framework of ecological units. Ecosystem Management Applications for Sustainable Forest and Wildlife Resources. Yale University Press, New Haven & London, pp. 181–200.

Cleland, D.T., Freeouf, J.A., Keys, J.E., Nowacki, G.J., Carpenter, C.A., McNab, W.H., 2007. Ecological subregions: sections and subsections for the conterminous United States. Gen. Tech. Rep. WO-76D 76.

Coburn, J.M., Chong, S.T., Kim, H.C., Chang, N.W., Calix, L.C., Resto, K., Lee, D.J., Johnson, J.L., Robbins, R.G., Klein, T.A., 2016. Tick surveillance in four southwestern provinces of the Republic of Korea during 2013. Syst. Appl. Acarol. 21, 147–165. https://doi.org/10.11158/saa.21.2.1.

Cumbie, A.N., Trimble, R.N., Eastwood, G., 2022. Pathogen spillover to an invasive tick species: first detection of Bourbon virus in *Haemaphysalis longicornis* in the United States. Pathogens 11, 454. https://doi.org/10.3390/pathogens11040454.

Dalton, R., Monteverde, R., Sugarman, P., Volkert, R., 2014. Bedrock geologic map of New Jersey.
Degenhardt, F., Seifert, S., Szymczak, S., 2019. Evaluation of variable selection methods for random forests and omics data sets. Brief. Bioinform. 20, 492–503.
Dinkel, K.D., Herndon, D.R., Noh, S.M., Lahmers, K.K., Todd, S.M., Ueti, M.W., Scoles, G. A., Mason, K.L., Fry, L.M., 2021. A U.S. isolate of *Theileria orientalis*, Ikeda genotype, is transmitted to cattle by the invasive Asian longhorned tick, *Haemaphysalis longicornis*. Parasites Vectors 14, 157. https://doi.org/10.1186/s13071-021-04659-9.
Egizi, A., Bulaga-Seraphin, L., Alt, E., Bajwa, W.I., Bernick, J., Bickerton, M., Campbell, S. R., Connally, N., Doi, K., Falco, R.C., Gaines, D.N., Greay, T.L., Harper, V.L., Heath, A.C.G., Jiang, J., Klein, T.A., Maestas, L., Mather, T.N., Occi, J.L., Oskam, C. L., Pendleton, J., Teator, M., Thompson, A.T., Tufts, D.M., Umemiya-Shirafuji, R., VanAcker, M.C., Yabsley, M.J., Fonseca, D.M., 2020. First glimpse into the origin and spread of the Asian longhorned tick, *Haemaphysalis longicornis*, in the United States. Zoonoses Public Health. https://doi.org/10.1111/zph.12743.
Egizi, A.M, Occi, J.L., Price, D.C., Fonseca, D.M., 2019a. Leveraging the expertise of the New Jersey mosquito control community to jump start standardized tick surveillance. Insects 10, 219. https://doi.org/10.3390/insects10080219.
Egizi, A.M., Robbins, R.G., Beati, L., Nava, S., vans, C.R., Occi, J.L., Fonseca, D.M., 2019b. A pictorial key to differentiate the recently detected exotic *Haemaphysalis longicornis* Neumann, 1901 (Acari, Ixodidae) from native congeners in North America. ZooKeys 117–128. https://doi.org/10.3897/zookeys.818.30448.
Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., 2006. Novel methods improve prediction of species distributions from occurrence data. Ecography 29, 129–151.
Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813.
Elith, J., Phillips, S.J., Hastie, T., Dudik, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.
Freeman, E.A., Moisen, G.G., 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol. Model. 217, 48–58. https://doi.org/10.1016/j.ecolmodel.2008.05.015.
Gallien, L., Douzet, R., Pratte, S., Zimmermann, N.E., Thuiller, W., 2012. Invasive species distribution models – how violating the equilibrium assumption can create new insights. Glob. Ecol. Biogeogr. 21, 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x.
Greenwell, B.M., 2017. pdp: An R package for constructing partial dependence plots. R J 9, 421–436.
Hammer, J.F., Emery, D., Bogema, D.R., Jenkins, C., 2015. Detection of *Theileria orientalis* genotypes in *Haemaphysalis longicornis* ticks from southern Australia. Parasites Vectors 8, 229.
Heath, A.C.G., 2016. Biology, ecology and distribution of the tick, *Haemaphysalis longicornis* Neumann (Acari: Ixodidae) in New Zealand. N. Z. Vet. J. 64, 10–20.
Heikkinen, R.K., Marmion, M., Luoto, M., 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? Ecography 35, 276–288. https://doi.org/10.1111/j.1600-0587.2011.06999.x.
Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2017. dismo: Species distribution modeling. R Package Version 11-4.
Johnson, J.L., Kim, H., Coburn, J.M., Chong, S., Chang, N.W., Robbins, R.G., Klein, T.A., 2017. Tick surveillance in two southeastern provinces, including three metropolitan areas, of the Republic of Korea during 2014. Syst. Appl. Acarol.
Jordan, R.A., Gable, S., Egizi, A., 2022. Relevance of spatial and temporal trends in nymphal tick density and infection prevalence for public health and surveillance practice in long-term endemic areas: a case study in Monmouth County, NJ. J. Med. Entomol. 59, 1451–1466. https://doi.org/10.1093/jme/tjac073.
Keirans, J.E., Litwak, T.R., 1989. Pictorial key to the adults of hard ticks, family Ixodidae (Ixodida: Ixodoidea), east of the Mississippi river. J. Med. Entomol. 26, 435–448. https://doi.org/10.1093/jmedent/26.5.435.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26.
Kursa, M.B., Rudnicki, W.R., 2010. Feature Selection with the boruta package. J. Stat. Softw. 36 https://doi.org/10.18637/jss.v036.i11.
Levin, M.L., Stanley, H.M., Hartzer, K., Snellgrove, A.N., 2021. Incompetence of the Asian longhorned tick (Acari: Ixodidae) in transmitting the agent of human granulocytic anaplasmosis in the United States. J. Med. Entomol. 58, 1419–1423. https://doi.org/10.1093/jme/tjab015.
Li, A., Liu, L., Wu, W., Liu, Y., Huang, X., Li, C., Liu, D., Li, J., Wang, S., Li, D., Liang, M., 2021. Molecular evolution and genetic diversity analysis of SFTS virus based on next-generation sequencing. Biosaf. Health 03, 105–115. https://doi.org/10.1016/j.bsheal.2021.02.002.
Li, Z., Bao, C., Hu, J., Liu, W., Wang, X., Zhang, L., Ji, Z., Feng, Z., Li, L., Shen, A., 2016. Ecology of the tick-borne phlebovirus causing severe fever with thrombocytopenia syndrome in an endemic area of China. PLoS Negl. Trop. Dis. 10, e0004574.
Liu, C., White, M., Newell, G., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. J. Biogeogr. 40, 778–789. https://doi.org/10.1111/jbi.12058.
Liu, K., Zhou, H., Sun, R.X., Yao, H.W., Li, Y., Wang, L.P., Mu, D., Li, X.L., Yang, Y., Gray, G.C., 2015. A national assessment of the epidemiology of severe fever with thrombocytopenia syndrome, China. Sci. Rep. 5, 9679.
Luo, L.M., Zhao, L., Wen, H.L., Zhang, Z.T., Liu, J.W., Fang, L.Z., Xue, Z.F., Ma, D.Q., Zhang, X.S., Ding, S.J., 2015. *Haemaphysalis longicornis* ticks as reservoir and vector of severe fever with thrombocytopenia syndrome virus in China. Emerg. Infect. Dis. 21, 1770.
Marmion, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecol. Model. 220, 3512–3520. https://doi.org/10.1016/j.ecolmodel.2008.10.019. Selected Papers on Spatially Explicit Landscape Modelling: Current practices and challenges.
McNab, W.H., Cleland, D.T., Freeouf, J.A., Keys, J.E., Nowacki, G.J., Carpenter, C.A., 2007. Description of ecological subregions: sections of the conterminous United States. Gen. Tech. Rep. WO-76B 76B, 1–82. 10.2737/WO-GTR-76B.
Merow, C., Smith, M.J., Silander Jr, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36, 1058–1069.
Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., Anderson, R.P., 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods Ecol. Evol. 5, 1198–1205. https://doi.org/10.1111/2041-210X.12261.
Neilson, F.J.A., 1980. An investigation into the ecology, biology, distribution and control of *Haemaphysalis longicornis* Neumann, 1901: a thesis presented in partial fulfilment of the requirements for the degree of Master of Veterinary Science at Massey University.
NRCS Soils [WWW Document], 2023 https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/class/maps/(accessed 3.12.20).
NRCS, U., 2007. National engineering handbook: Part 630—Hydrology. USDA Soil Conserv. Serv. Wash. DC USA.
Oh, J.Y., Moon, B.-C., Bae, B.K., Shin, E., Ko, Y.H., Kim, Y.-J., Park, Y.H., Chae, J.-S., 2009. Genetic identification and phylogenetic analysis of *Anaplasma* and *Ehrlichia* species in *Haemaphysalis longicornis* collected from Jeju Island, Korea. J. Bacteriol. Virol. 39, 257–267.
Pearson, R.G., 2007. Species' distribution modeling for conservation educators and practitioners. Synth. Am. Mus. Nat. Hist. 50.
Pearson, R.G., Raxworthy, C.J., Nakamura, M., Townsend Peterson, A., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34, 102–117.
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.
Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.
Piesman, J., Gray, J.S., 1994. Lyme disease/Lyme borreliosis. Ecol. Dyn. Tick-Borne Zoonoses Sonenshine TN Mather Eds Oxf. Univ. Press Inc N. Y. 327–350.
Price, K.J., Ayres, B.N., Maes, S.E., Witmier, B.J., Chapman, H.A., Coder, B.L., Boyer, C. N., Eisen, R.J., Nicholson, W.L., 2022. First detection of human pathogenic variant of *Anaplasma phagocytophilum* in field-collected *Haemaphysalis longicornis*, Pennsylvania, USA. Zoonoses Public Health 69, 143–148. https://doi.org/10.1111/zph.12901.
Price, K.J., Graham, C.B., Witmier, B.J., Chapman, H.A., Coder, B.L., Boyer, C.N., Foster, E., Maes, S.E., Bai, Y., Eisen, R.J., Kyle, A.D., 2021. *Borrelia burgdorferi* sensu stricto DNA in field-collected *Haemaphysalis longicornis* ticks, Pennsylvania, United States. Emerg. Infect. Dis. 27, 608–611. https://doi.org/10.3201/eid2702.201552.
R Development Core Team, 2019. R: a language and environment for statistical computing.
Radosavljevic, A., Anderson, R.P., 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. J. Biogeogr. 41, 629–643.
Raghavan, R.K., Peterson, A.T., Cobos, M.E., Ganta, R., Foley, D., 2019. Current and future distribution of the lone star tick, *Amblyomma americanum* (L.) (Acari: Ixodidae) in North America. PLoS One 14, e0209082. https://doi.org/10.1371/journal.pone.0209082.
Rainey, T., Occi, J.L., Robbins, R.G., Egizi, A., 2018. Discovery of *Haemaphysalis longicornis* (Ixodida: Ixodidae) parasitizing a sheep in New Jersey, United States. J. Med. Entomol. 55, 757–759.
Raney, W.R., Perry, J.B., Hermance, M.E., 2022. Transovarial transmission of Heartland virus by invasive Asian longhorned ticks under laboratory conditions. Emerg. Infect. Dis. 28, 726–729. https://doi.org/10.3201/eid2803.210973.
Rochlin, I., 2019. Modeling the Asian longhorned tick (Acari: Ixodidae) suitable habitat in North America. J. Med. Entomol. 56, 384–391. https://doi.org/10.1093/jme/tjy210.
Rochlin, I., Benach, J.L., Furie, M.B., Thanassi, D.G., Kim, H.K., 2022. Rapid invasion and expansion of the Asian longhorned tick (*Haemaphysalis longicornis*) into a new area on Long Island, New York, USA. Ticks Tick Borne Dis. 102088 https://doi.org/10.1016/j.ttbdis.2022.102088.
Rodda, G.H., Jarnevich, C.S., Reed, R.N., 2011. Challenges in identifying sites climatically matched to the native ranges of animal invaders. PLoS ONE 6, e14670.
Sherpa, P., Harrington, L.C., Piedmonte, N.P., Wunderlin, K., Falco, R.C., 2021. Optimal collection methods for Asian longhorned ticks (Ixodida: Ixodidae) in the Northeast United States. J. Med. Entomol. 58, 2255–2263. https://doi.org/10.1093/jme/tjab083.
Sohl, T.L., 2014. The relative impacts of climate and land-use change on conterminous United States bird species from 2001 to 2075. PLoS ONE 9, e112251. https://doi.org/10.1371/journal.pone.0112251.
Soil Survey, 2020. Web soil survey.
Soil Survey Staff, 1999. Soil taxonomy—a basic system of soil classification for making and interpreting soil surveys.
Srivastava, V., Lafond, V., Griess, V.C., 2019. Species distribution models (SDM): applications, benefits and challenges in invasive species management. CABI Rev. 1–13.
Thompson, A.T., White, S., Shaw, D., Egizi, A., Lahmers, K., Ruder, M.G., Yabsley, M.J., 2020. *Theileria orientalis* Ikeda in host-seeking *Haemaphysalis longicornis* in Virginia, U.S.A. Ticks Tick Borne Dis. 11, 101450 https://doi.org/10.1016/j.ttbdis.2020.101450.

Tsunoda, T., Tatsuzawa, S., 2004. Questing height of nymphs of the bush tick, *Haemaphysalis longicornis*, and its closely related species, *H. mageshimaensis*: correlation with body size of the host. Parasitology 128, 503–509.

USDA APHIS situation report, 2022. National *Haemaphysalis longicornis* (Asian longhorned tick) situation report - August 23, 2022 https://www.aphis.usda.gov/animal_health/animal_diseases/tick/downloads/longhorned-tick-sitrep.pdf. US Department of Agriculture.

Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution 62, 2868–2883.

Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol.Appl. 21, 335–342.

West, A.M., Kumar, S., Brown, C.S., Stohlgren, T.J., Bromberg, J., 2016. Field validation of an invasive species Maxent model. Ecol. Inform. 36, 126–134. https://doi.org/10.1016/j.ecoinf.2016.11.001.

Wormser, G.P., McKenna, D., Piedmonte, N., Vinci, V., Egizi, A.M., Backenson, B., Falco, R.C., 2020. First recognized human bite in the United States by the Asian longhorned tick, *Haemaphysalis longicornis*. Clin. Infect. Dis. 70, 314–316.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., Veran, S., 2013. Presence-only modelling using MAXENT: when can we trust the inferences? Methods Ecol. Evol. 4, 236–243.

Zheng, H., Yu, Z., Zhou, L., Yang, X., Liu, J., 2012. Seasonal abundance and activity of the hard tick *Haemaphysalis longicornis* (Acari: Ixodidae) in North China. Exp. Appl. Acarol. 56, 133–141.

Zhou, L., Liang, T., Shi, L., 2019. Amphibian and reptilian chorotypes in the arid land of central Asia and their determinants. Sci. Rep. 9, 1–13. https://doi.org/10.1038/s41598-019-45912-7.