Supplementary Data

Supplementary Methods

High-resolution ecological niche modeling of *Ixodes scapularis* ticks based on passive surveillance data at the northern frontier of Lyme disease emergence in North America.

Sampling Bias

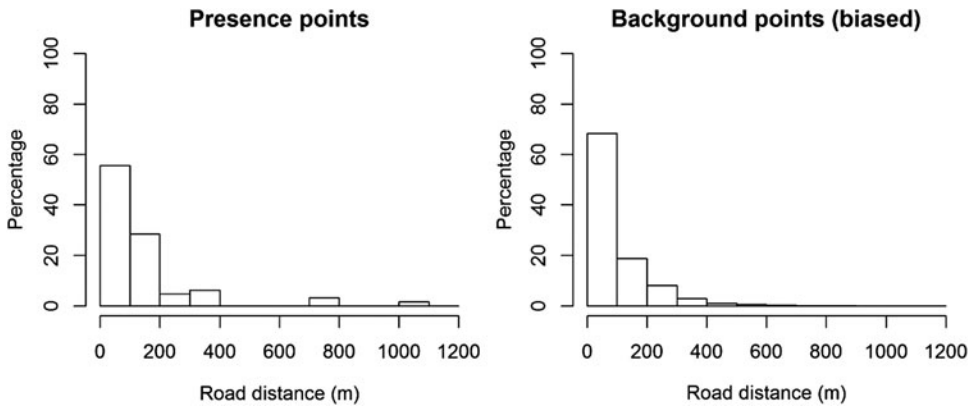
We built a MaxEnt model (with default settings) using nine potential indicators of spatial sampling bias, including variables like distance from roads and population density (Supplementary Table S2). Distance from roads was by far the most important variable in this model, as measured by permutation importance. Thus, we decided to use this variable as our proxy for likelihood of sampling.

All but one of the 63 presence points were found within 1 km of a road, whereas none were directly on a road (0 m). An exponential distribution was fit to the distribution of road distance in presence points ($\lambda \approx 0.00721$). Background points ($n = 10{,}000$) were then sampled from cells according to road distance: probability 0 from cells on roads or more than 1 km from them and with probability proportional to the fitted exponential distribution between 15–1000 m. The distribution of distances from roads of this background point dataset (Supplementary Fig. S1) much more closely reflects the fact that the vast majority of presence points occur within 400 m of a road than the naively sampled set of background points (Supplementary Fig. S2).

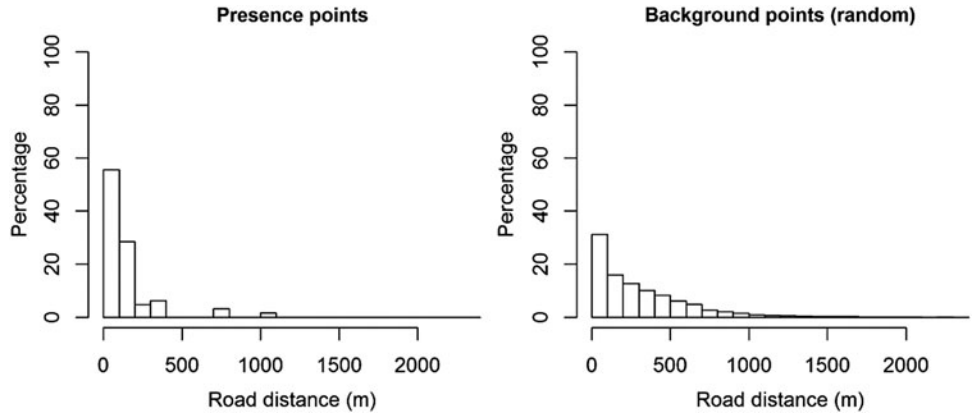
Model Selection

Beginning with the full variable set, we built models with all feature class combinations of linear, quadratic, and hinge features and regularization multiplier values between 1 and 10 in increments of 0.5. We used AICc to determine the optimal model, as described in Warren and Seifert (2011) with modifications by Wright et al. (2015) and implemented in the R package *enmSdm* (Smith 2017).

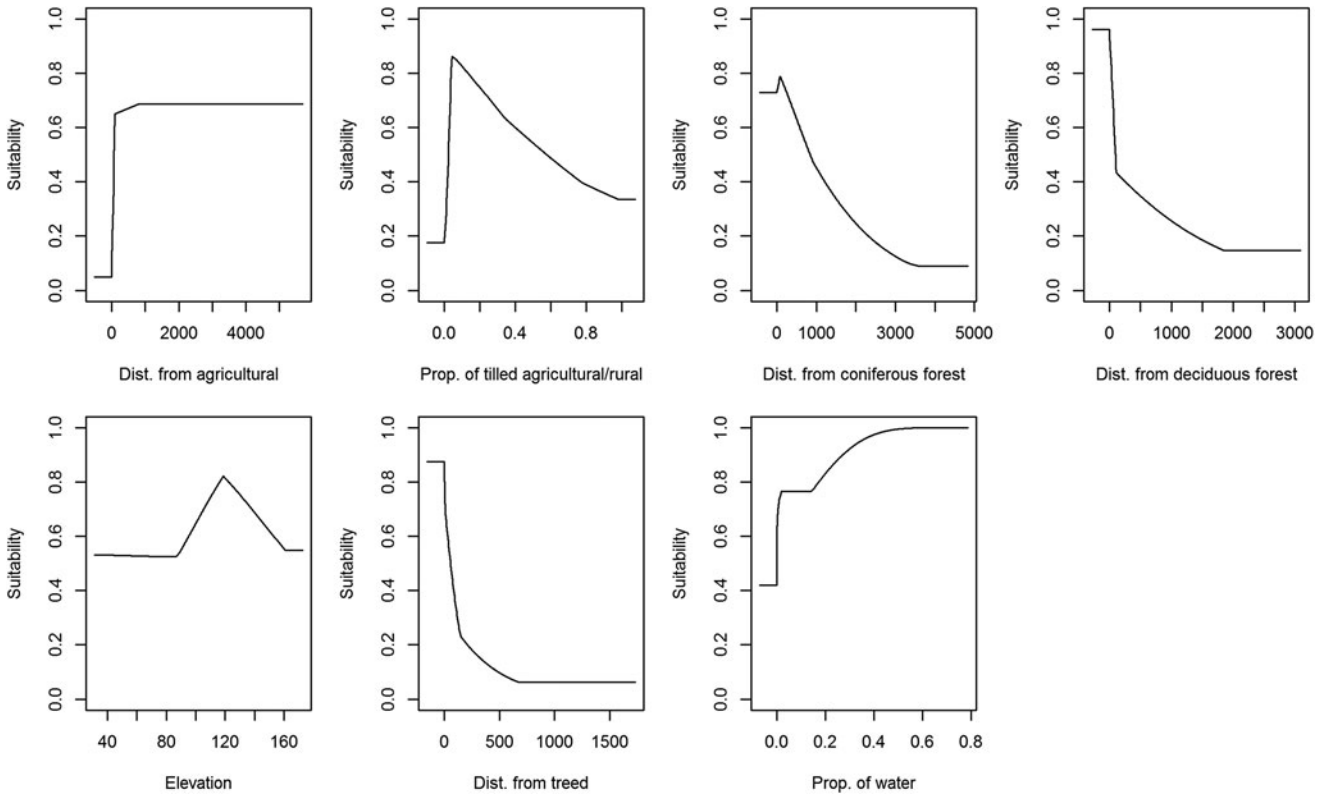
To improve the interpretability of the model, the variable set was reduced according to the method described in Yost et al. (2008). Briefly, at each step the lowest contributing variable (as measured by average loss in regularized training gain) is removed, and a 95% confidence interval of regularized training gain for the whole model is calculated (based on 10 random 75/25 training/testing partitions of the dataset). This procedure is continued until the calculated confidence interval no longer overlaps with the confidence interval of the full model with all variables, at which point removal of further variables is determined to produce an inferior model.



SUPPLEMENTARY FIG. S1. Distribution of distance from roads for presence points ($n = 63$) and background points sampled with spatial bias ($n = 10{,}000$).



SUPPLEMENTARY FIG. S2. Distribution of distance from roads for presence points ($n = 63$) and randomly sampled background points ($n = 10{,}000$).



SUPPLEMENTARY FIG. S3. Response curves for variables in the *Ixodes scapularis* model. The *x*-axis shows the value of predictor, and the *y*-axis shows the predicted suitability. Curves obtained by varying each variable in a univariate model.

Supplementary Table S1. Name, Processing, and Original Resolution of Environmental Layers Tested in the *Ixodes scapularis* Model

Layer name	*Processing*	*Resolution*
Elevation	Resampled elevation raster to 15 m	90 m
Proportion of tilled agricultural	Calculated proportion of tiled agricultural cells within a 1000 m buffer	15 m
Proportion of tilled agricultural/ undifferentiated rural	Calculated proportion of tilled agricultural/undifferentiated rural cells within a 1000 m buffer	15 m
Proportion of coniferous forest	Calculated proportion of coniferous forest cells within a 1000 m buffer	15 m
Proportion of deciduous forest	Calculated proportion of deciduous forest cells within a 1000 m buffer	15 m
Proportion of mixed forest	Calculated proportion of forest cells within a 1000 m buffer	15 m
Proportion of deciduous or mixed forest	Calculated proportion of deciduous or mixed forest cells within a 1000 m buffer	15 m
Proportion of forest (not including hedgerows)	Calculated proportion of forest cells within a 1000 m buffer	15 m
Proportion of hedgerow	Calculated proportion of hedgerow cells within a 1000 m buffer	15 m
Proportion of treed (including hedgerows)	Calculated proportion of treed cells within a 1000 m buffer	15 m
Proportion of water	Calculated proportion of water cells within a 1000 m buffer	
Distance from tilled agricultural	Calculated distance from each cell to a tilled agricultural cell	15 m
Distance from tilled agricultural/ undifferentiated rural	Calculated distance from each cell to a tilled agricultural/ undifferentiated rural cell	15 m
Distance from coniferous forest	Calculated distance from each cell to a coniferous forest cell	15 m
Distance from deciduous forest	Calculated distance from each cell to a deciduous forest cell	15 m
Distance from mixed forest	Calculated distance from each cell to a mixed forest cell	15 m
Distance from deciduous or mixed forest	Calculated distance from each cell to a deciduous or mixed forest cell	15 m
Distance from forest (not including hedgerows)	Calculated distance from each cell to a forest cell	15 m
Distance from hedgerow	Calculated distance from each cell to a hedgerow cell	15 m
Distance from treed (including hedgerows)	Calculated distance from each cell to a treed cell	15 m
Distance from water	Calculated distance from each cell to a water cell	15 m

Supplementary Table S2. Reclassification of 21 Land Cover Classes Into 13 Classes in the Southern Ontario Land Resource Information System Version 2.0

SOLRIS value	*SOLRIS class*	*New class*
51	Open alvar	Barren/other
53	Treed alvar	Grassland/shrubland
90	Forest	Mixed forest
91	Coniferous forest	Coniferous forest
92	Mixed forest	Mixed forest
93	Deciduous forest	Deciduous forest
131	Treed swamp	Wetland
135	Thicket swamp	Wetland
140	Fen	Wetland
150	Bog	Wetland
160	Marsh	Wetland
170	Open water	Water
191	Plantations—tree cultivated	Coniferous forest
192	Hedgerows	Hedgerows
193	Tilled	Tilled agricultural
201	Transportation	Transportation
202	Built-up area—pervious	Built-up pervious
203	Built-up area—impervious	Built-up impervious
204	Extraction—aggregate	Barren/other
205	Extraction—peat/topsoil	Barren/other
250	Undifferentiated	Rural/undifferentiated

Only land cover classes present in Ottawa, Canada are shown.

Supplementary Table S3. Name, Processing, and Original Resolution of Environmental Layers Tested for Spatially Biased Background Sampling in the *I. scapularis* Model

Layer name	*Processing*	*Resolution*
Population density	Rasterized population density (population/area) in Census Dissemination Areas of the Canada 2016 census (Statistics Canada 2017)	Vector
Proportion of built-up area—pervious	Calculated proportion of built-up area—pervious cells within a 1000 m buffer	15 m
Proportion of built-up area—impervious	Calculated proportion of built-up area—impervious cells within a 1000 m buffer	15 m
Proportion of roads	Calculated proportion of road cells within a 1000 m buffer	15 m
Proportion of roads and built-up area—impervious	Calculated proportion of road and built-up area—impervious cells within a 1000 m buffer	15 m
Distance from built-up area—pervious	Calculated distance from each cell to a built-up area—pervious cell	15 m
Distance from built-up area—impervious	Calculated distance from each cell to a built-up area—impervious cell	15 m
Distance from roads	Calculated distance from each cell to a road cell	15 m
Distance from roads and built-up area—impervious	Calculated distance from each cell to a road or built-up area—pervious cell	15 m

References

Smith AB. enmSdm: Tools for modeling species niches and distributions: R package version 0.1.0.0. 2017. Available at www.earthskysea.org

Statistics Canada. Canada, provinces, territories, census divisions, census subdivisions and dissemination areas. 2017. Available at http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page_dl-tc.cfm

Warren DL, Seifert SN. Ecological niche modeling in maxent: The importance of model complexity and the performance of model selection criteria. Ecol Appl 2011; 21:335–342.

Wright AN, Hijmans RJ, Schwartz MW, Shaffer HB. Multiple sources of uncertainty affect metrics for ranking conservation risk under climate change. Divers Distrib 2015; 21:111–122.

Yost AC, Petersen SL, Gregg M, Miller R. Predictive modeling and mapping sage grouse (*Centrocercus urophasianus*) nesting habitat using maximum entropy and a long-term dataset from Southern Oregon. Ecol Inform 2008; 3:375–386.