

1 Exploring the niche of *Rickettsia montanensis* (Rickettsiales: Rickettsiaceae) infection of the  
2 American dog tick (Acari: Ixodidae), using multiple species distribution model approaches.

3

4 Catherine A. Lippi<sup>1,2</sup>, Holly D. Gaff<sup>3,4</sup>, Alexis L. White<sup>1,2</sup>, Heidi K. St. John<sup>5,6</sup>, Allen L.  
5 Richards<sup>5</sup>, Sadie J. Ryan<sup>\*1,2,7</sup>

- 6 1. Quantitative Disease Ecology and Conservation (QDEC) Lab Group, Department of  
7 Geography, University of Florida, Gainesville, FL, 32610 USA
- 8 2. Emerging Pathogens Institute, University of Florida, Gainesville, FL, 32610 USA
- 9 3. Department of Biological Sciences, Old Dominion University, Norfolk, 23529 VA, USA
- 10 4. School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal,  
11 Durban, 4041, South Africa
- 12 5. Viral and Rickettsial Disease Program (VRDD) Naval Medical Research Center, Silver  
13 Spring, MD
- 14 6. Henry M. Jackson Foundation for the Advancement of Military Medicine, 6720A  
15 Rockledge Dr, Bethesda, MD 20817 USA
- 16 7. School of Life Sciences, University of KwaZulu-Natal, Durban, 4041, South Africa

17

18 \*corresponding: [sjryan@ufl.edu](mailto:sjryan@ufl.edu)

19

## 20 **Abstract:**

21 The American dog tick, *Dermacentor variabilis* (Say), is a vector for several human disease  
22 causing pathogens such as tularemia, Rocky Mountain spotted fever, and the understudied  
23 spotted fever group rickettsiae (SFGR) infection caused by *Rickettsia montanensis*. It is  
24 important for public health planning and intervention to understand the distribution of this tick  
25 and pathogen encounter risk. Risk is often described in terms of vector distribution, but greatest  
26 risk may be concentrated where more vectors are positive for a given pathogen. When assessing  
27 species distributions, the choice of modeling framework and spatial layers used to make  
28 predictions are important. We first updated the modeled distribution of *D. variabilis* and *R.*  
29 *montanensis* using MaxEnt, refining bioclimatic data inputs, and including soils variables. We  
30 then compared geospatial predictions from five species distribution modeling (SDM)  
31 frameworks. In contrast to previous work, we additionally assessed whether the *R. montanensis*

32 positive *D. variabilis* distribution is nested within a larger overall *D. variabilis* distribution,  
33 representing a fitness cost hypothesis. We found that 1) adding soils layers improved the  
34 accuracy of the MaxEnt model; 2) the predicted ‘infected niche’ was smaller than the overall  
35 predicted niche across all models; and 3) each model predicted different sizes of suitable niche,  
36 at different levels of probability. Importantly, the models were not directly comparable in output  
37 style, which could create confusion in interpretation when developing planning tools. The  
38 random forest (RF) model had the best measured validity and fit, suggesting it may be most  
39 appropriate to these data.

40 **Keywords:** Species distribution models; ecological niche models; spotted fever group; boosted  
41 regression trees; MaxEnt; Random Forests

42

## 43 **Introduction**

44           The American dog tick (*Dermacentor variabilis*) is primarily distributed east of the  
45 Rocky Mountains and the Pacific coastal region in the USA, where it is a known vector of the  
46 pathogens that cause both tularemia, (*Francisella tularensis*) and Rocky Mountain spotted fever  
47 (RMSF) (*Rickettsia rickettsii*). Both diseases can be fatal if left untreated, and therefore  
48 understanding the risk of exposure to *D. variabilis* bites is an essential part of public health  
49 planning. In addition to these more well-known vector-borne diseases, *D. variabilis* can also  
50 transmit *Rickettsia montanensis*, a spotted fever group rickettsiae (SFGR). *Rickettsia*  
51 *montanensis* was previously thought to be nonpathogenic in humans (Baldrige et al. 2010) but  
52 has more recently been implicated as the agent in an afebrile rash illness (McQuiston et al. 2012).  
53 In addition to the potential for human pathogenicity, *R. montanensis* may play an interesting role  
54 in the manifestation of other SFGR dynamics by inhibiting tick coinfection of another *Rickettsia*  
55 spp., or conferring antigenic responses in humans exposed to *R. montanensis*, providing  
56 immunity (partial or complete) to other SFGR pathogens (Baldrige et al. 2010). This has  
57 potential implications for the transmission cycles of other SFGR, namely *R. rickettsii*, which  
58 occurs sympatrically with *R. montanensis*. *Rickettsia montanensis* infections may also affect  
59 SFGR disease surveillance and case detection. In 2010, the Centers for Disease Control (CDC)  
60 designated a new category for reporting rickettsial diseases to reflect diagnostic uncertainty in  
61 cases (CDC 2010, 2019a). The new reporting group, Spotted Fever Rickettsiosis (SFR), includes  
62 cases of RMSF, Pacific Coast tick fever, *Rickettsia parkeri* rickettsiosis (Tidewater spotted  
63 fever), and rickettsialpox. The number of SFR cases reported in the USA has shown a generally  
64 increasing trend since 2010, with more than 6,200 cases reported in 2017 (CDC 2019a). It is  
65 plausible that *R. montanensis* is the agent responsible for some of these SFR cases, as commonly

66 used serologic tests are not able to differentiate rickettsial pathogens because of immunological  
67 cross-reactivity (CDC 2019a, Nicholson and Paddock 2019). The public health implications of *R.*  
68 *montanensis* infections, both for human health and case surveillance, has fueled interest in  
69 investigating this pathogen.

70 The previous assumption of *R. montanensis* being nonpathogenic in humans, and non-  
71 specific point of care tests for SFGR has led to few publications on this pathogen and its known  
72 or potential distribution, as pointed out by Hardstone, Yoshimizu and Billeter (2018). Estimation  
73 of the geographic distribution for pathogens and their vectors is a crucial component in the  
74 development of public health agency policies and recommendations. In 2016, St. John and  
75 colleagues (St. John et al. 2016) conducted a study to describe the predicted distributions of *R.*  
76 *montanensis* positive and negative *D. variabilis* in the United States using the MaxEnt modeling  
77 environment to generate species distribution models (SDMs). Species distribution models have  
78 become increasingly prevalent in the disease ecology literature, with examples spanning a range  
79 of infectious disease systems, spatial scales, and geographic foci (Gurgel-Gonçalves et al. 2012,  
80 Blackburn et al. 2017, Lippi et al. 2019). This methodology has been embraced as an accessible  
81 means of quickly estimating the geographic range of a given pathogen or vector, which is used in  
82 many instances to infer risk of exposure.

83 Species distribution models are attractive from a logistical standpoint for mapping  
84 potential exposure to vectors as they provide a means of estimating suitable geographic ranges  
85 with presence-only data, which are often available through public surveillance networks. Briefly,  
86 SDMs are made by correlating location records of species occurrence with underlying  
87 environmental conditions in a geospatial modeling environment, and the model is then projected  
88 to unsampled portions of the landscape (Peterson and Soberón 2012). There is now a range of

89 modeling algorithms and freely available software packages that, when coupled with the  
90 availability of georeferenced occurrence records, have made SDMs commonly used across  
91 disciplines in recent years (Townsend Peterson et al. 2007, Elith et al. 2008, Phillips and Dudík  
92 2008, Elith and Leathwick 2009, Evans et al. 2011, Naimi and Araújo 2016). Nevertheless,  
93 caution must be exercised when modeled distributions are put into a health advisory context.  
94 Consensus in SDMs is notoriously difficult to achieve, as discrepancies in potential distributions  
95 may arise from choice of modeling method, user-specified parameters, selection of  
96 environmental predictors, and biases in data inputs (Carlson et al. 2018).

97 The objectives of this study were as follows: i) expand the methodology used to generate  
98 the predicted range map of *D. variabilis* presented in St. John et al. (St. John et al. 2016) by  
99 creating SDMs in MaxEnt with an updated and refined set of environmental predictors; ii)  
100 explore differences across four additional SDM algorithms; and iii) compare potential  
101 geographic distributions of *D. variabilis* and the subset of *D. variabilis* that tested positive for *R.*  
102 *montanensis* to assess if there were any appreciable differences in the ‘infected’ niche.

103

## 104 **Methods**

### 105 ***Presence data***

106 Locations of *D. variabilis* in the US from 2002-2012, which tested both positive and  
107 negative for *R. montanensis*, are described in St John et al. (2016). Data were openly available  
108 through VectorMap (<http://vectormap.si.edu/dataportal/>), a project of the Walter Reed  
109 Bioinformatics Unit, housed at the Smithsonian Institution Washington DC (St. John et al. 2016).  
110 These data were collected primarily through reports to United States military installations and as

111 part of passive vector surveillance studies. Prior to implementing modeling procedures, we  
112 conducted data thinning on species occurrence points via the spThin package in R (ver. 3.6.1) (R  
113 Core Team 2019), which uses a spatial thinning algorithm to randomize the removal of  
114 occurrence locations within a specified distance threshold (Aiello-Lammens et al. 2015). The  
115 resulting dataset retained spatially unique records of species presence within 10km, the spatial  
116 resolution of the study. Locations of tick occurrences in these data were provided as either the  
117 location of collection or the associated medical treatment facility. Correlative SDMs are  
118 susceptible to the effects of geographic sampling biases, where overrepresented locations may  
119 erroneously drive associations between environmental conditions at oversampled locations and  
120 species occurrence (Aiello-Lammens et al. 2015). Spatial thinning of occurrences at the chosen  
121 spatial resolution minimizes the potential effects of sampling bias in this dataset, where locations  
122 near reporting medical facilities may be overrepresented.

### 123 *Environmental data layers*

124 Species distribution models, of the type we present in this study, require gridded  
125 environmental data layers as input for building models and making spatial predictions. For  
126 comparability with the previous study (St. John et al. 2016), and to maintain consistency across  
127 algorithms in this paper, we used interpolated bioclimatic (BIOCLIM) layers from  
128 WorldClim.org at a 10km resolution selected to match the spatial resolution of tick occurrence  
129 data (Fick and Hijmans 2017). The 19 BIOCLIM variables consist of long-term averages of  
130 temperature, precipitation, and associated measures of extremes and seasonality.

131 In addition to the original layer set from the previous study, we chose to add soils layers  
132 to our candidate environmental variables, as ticks are frequently found in the leaf-litter or debris  
133 and considered largely soil-dwelling organisms (Burtis et al. 2019). The International Soil

134 Reference Information Centre (ISRIC) SoilGrids product provides a global suite of 195 standard  
135 numeric and taxonomic soil descriptors at seven standard depths (Hengl et al. 2017). As ticks are  
136 sensitive to abiotic soil attributes such as soil moisture (Burtis et al. 2019), we selected two  
137 layers of soil data for inclusion in the candidate variable set for model building; soil organic  
138 carbon density and available soil water capacity until wilting point to describe the potential water  
139 capacity and retention of soils. A standard depth of 0cm was chosen to approximate the surface  
140 and leaf litter conditions that ticks may encounter in the environment. Gridded soil layer  
141 products, aggregated to a spatial resolution of 10km with the GDAL software package, were  
142 used to match the resolution of the study (GDAL/OGR contributors 2020).

143 Collinearity in environmental predictor variables is a well-described issue affecting SDM  
144 output, potentially increasing model instability and uncertainty in predictions (De Marco and  
145 Nóbrega 2018). We reduced collinearity in environmental variable inputs via variance inflation  
146 factor (VIF), wherein only those layers with values below a specified threshold ( $th=10$ ) were  
147 used in model building (Chatterjee and Hadi 2006).

#### 148 ***Model implementation***

149 We used the ‘sdm’ package in R (ver. 3.6.2) to fit and spatially project SDMs for positive  
150 ticks and the combined dataset of positive and negative ticks (Naimi and Araújo 2016). The sdm  
151 package provides a flexible modeling platform for building SDMs, assessing model accuracy,  
152 and projecting output. Choice of modeling method can result in drastically different predictions  
153 of species ranges. To assess variation in potential distributions as an artifact of methodology, we  
154 used five commonly implemented modeling algorithms for estimating species ranges. These  
155 included two regression methods, generalized linear model (GLM) and generalized additive  
156 model (GAM), and three machine learning methods: maximum entropy (MaxEnt), random

157 forests (RF), and boosted regression trees (BRT) (McCullagh and Nelder 1998, Breiman 2001,  
158 Wood 2006, Elith et al. 2008, Phillips and Dudík 2008). We additionally created an SDM  
159 modeling ensemble, where predictions were weighted by accuracy metrics and averaged across  
160 methods (Araujo and New 2007). Model ensembles have been criticized due to performance  
161 issues and poor reporting practices (Hao et al. 2019) but are widely used. In spite of known  
162 issues, we included the ensemble method for the purposes of comparison, as SDM ensembles are  
163 prolific in the literature and provide a useful tool for combining the results from the various  
164 model approaches.

165         Model parameterization also heavily influences resulting predictions of species'  
166 distributions. While model settings vary depending on method, they are generally chosen based  
167 on known attributes of the target species (e.g. physiological thermal limits, nonlinear responses  
168 to environmental drivers, etc.), intended application of results (e.g. hypothesis testing, biological  
169 interpretation of niche, designing interventions, etc.), or to address issues with bias and small  
170 sample size (Merow et al. 2013, Morales et al. 2017). *Dermacentor variabilis* are habitat  
171 generalists, and in contrast with many arthropod systems, little is known for acarids particularly  
172 regarding quantitative relationships with environmental conditions. We therefore used default  
173 modeling parameters for each method as defined in the sdm modeling platform. Five hundred  
174 model replications were run for each method, using a random subsampling of occurrence records  
175 (80%) for each model. Because we used presence-only data of species occurrences, pseudo-  
176 absences (n=1,000) were randomly generated throughout the study region within the sdm  
177 modeling procedure for each model run. Accuracy metrics were derived via a random  
178 subsampling (20%) of testing data, withheld from the model building process. Four measures of  
179 model accuracy were used to assess model output. These included the receiver operator



180 characteristic (ROC) curve with area under the curve (AUC), true skill statistic (TSS), model  
181 deviance, and mean omission (i.e. false negatives).

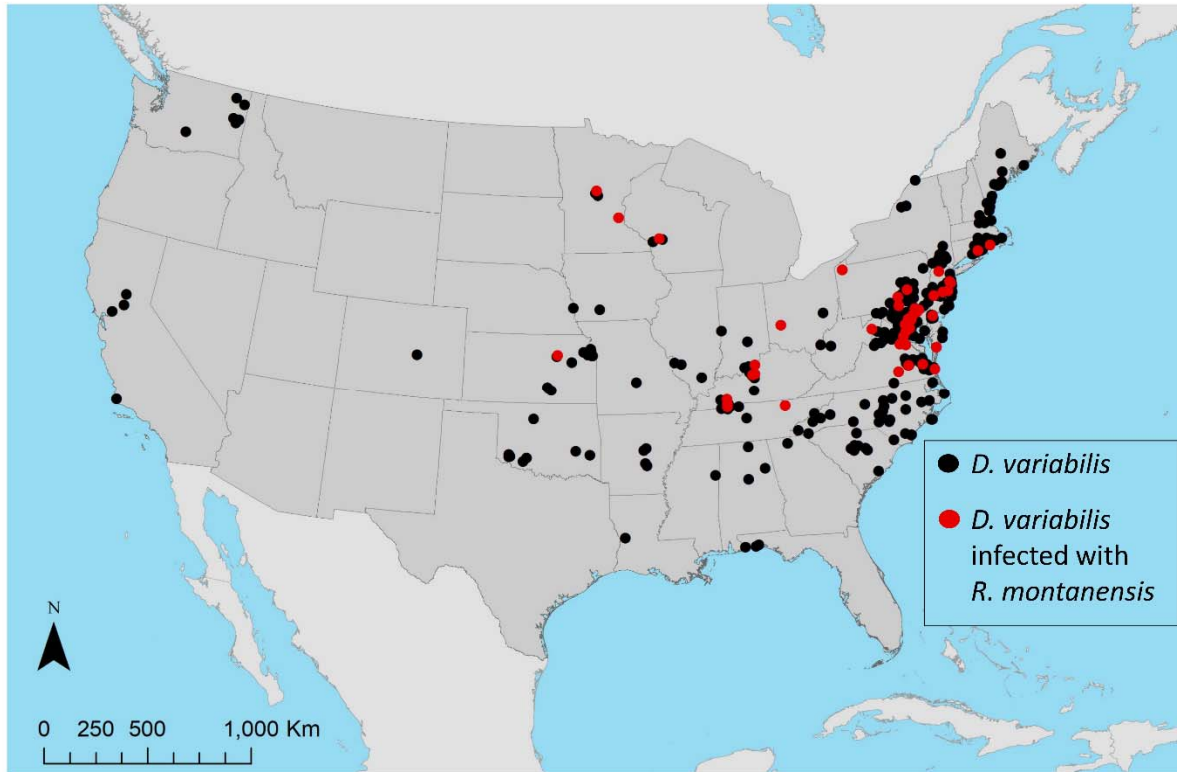
182 Spatial predictions were made by projecting the mean of all 500 models for a given  
183 method onto the study area. Overlap in geographic predictions between ticks positive for *R.*  
184 *montanensis* and the full dataset of tick occurrences (i.e. *R. montanensis* positive and negative  
185 ticks) was assessed by reclassifying modeled probabilities as binary geographic distributions (i.e.  
186 presence and absence) in ArcMap (ver.10.4), where raster cells with predicted occurrence  $\geq 50\%$   
187 were considered present (ESRI 2016). Reclassified distributions were combined using the  
188 ‘Raster Calculator’ tool in the Spatial Analyst extension of the program ArcMap, allowing for  
189 the visualization of range overlap between datasets.

## 190 **Results**

191 The full dataset of georeferenced *D. variabilis* occurrences downloaded from VectorMap  
192 was comprised of 3,771 records, 135 of which had tested positive for *R. montanensis*. Spatial  
193 thinning of occurrence records resulted in 432 unique locations of *D. variabilis*, where a subset  
194 of 44 records was positive for *R. montanensis* (Fig. 1). Models for both the full set of *D.*  
195 *variabilis* occurrence records, and the pathogen positive subset, were built with a reduced set of  
196 VIF-selected environmental variables which included annual mean temperature, mean diurnal  
197 temperature range, temperature seasonality, mean temperature of the wettest quarter, mean  
198 temperature of the driest quarter, precipitation seasonality, precipitation of the warmest quarter,  
199 precipitation of the coldest quarter, soil organic carbon density, and available soil water capacity  
200 (Table 1).

201

202 **Fig. 1.** Occurrence records for *D. variabilis*, and *D. variabilis* infected with *R. montanensis*, used  
203 in building species distribution models (SDMs).



204

205 Table 1. Environmental input datasets used in model building selected via variable inflation  
206 factor (VIF).

Environmental Variable (unit)	Coded Variable Name	Data Source
Annual Mean Temperature (°C)	Bio 1	Bioclim
Mean Diurnal Range (°C)	Bio 2	Bioclim
Temperature Seasonality	Bio 4	Bioclim
Mean Temp of Wettest Quarter (°C)	Bio 8	Bioclim
Mean Temp of Driest Quarter (°C)	Bio 9	Bioclim
Precipitation Seasonality	Bio 15	Bioclim
Precip of Warmest Quarter (mm)	Bio 18	Bioclim
Precip of Coldest Quarter (mm)	Bio 19	Bioclim
Soil Organic Carbon Density	OC Dens	ISRIC
Available Soil Water Capacity Until Wilting	WWP	ISRIC

207

208

209 Accuracy metrics for averaged SDMs produced with each modeling method are  
210 presented in Table 2. The averaged model built with RF had the highest predictive power,  
211 relative to low deviation and omission error, for the full dataset (AUC=0.96, TSS=0.79,  
212 deviance=0.47, mean omission=0.11) and subset of positive ticks (AUC=0.93, TSS=0.81,  
213 deviance=0.20, mean omission=0.10). MaxEnt models also performed well, albeit with higher  
214 mean omission error for the full dataset (AUC=0.95, TSS=0.83, deviance=0.23, mean  
215 omission=0.12) and subset of positive ticks (AUC=0.95, TSS=0.77, deviance=0.66, mean  
216 omission=0.12). Averaged models produced with GLM for full dataset (AUC=0.90, TSS=0.70,  
217 deviance=0.80, mean omission=0.15) and positive subset (AUC=0.92, TSS=0.76,  
218 deviance=0.23, mean omission=0.15), and BRT for the full dataset (AUC=0.90, TSS=0.69,  
219 deviance=0.89, mean omission=0.15) and positive subset (AUC=0.91, TSS=0.77,  
220 deviance=0.27, mean omission=0.13) had relatively lower performance, with lower accuracy  
221 metrics and higher error compared to other methods.

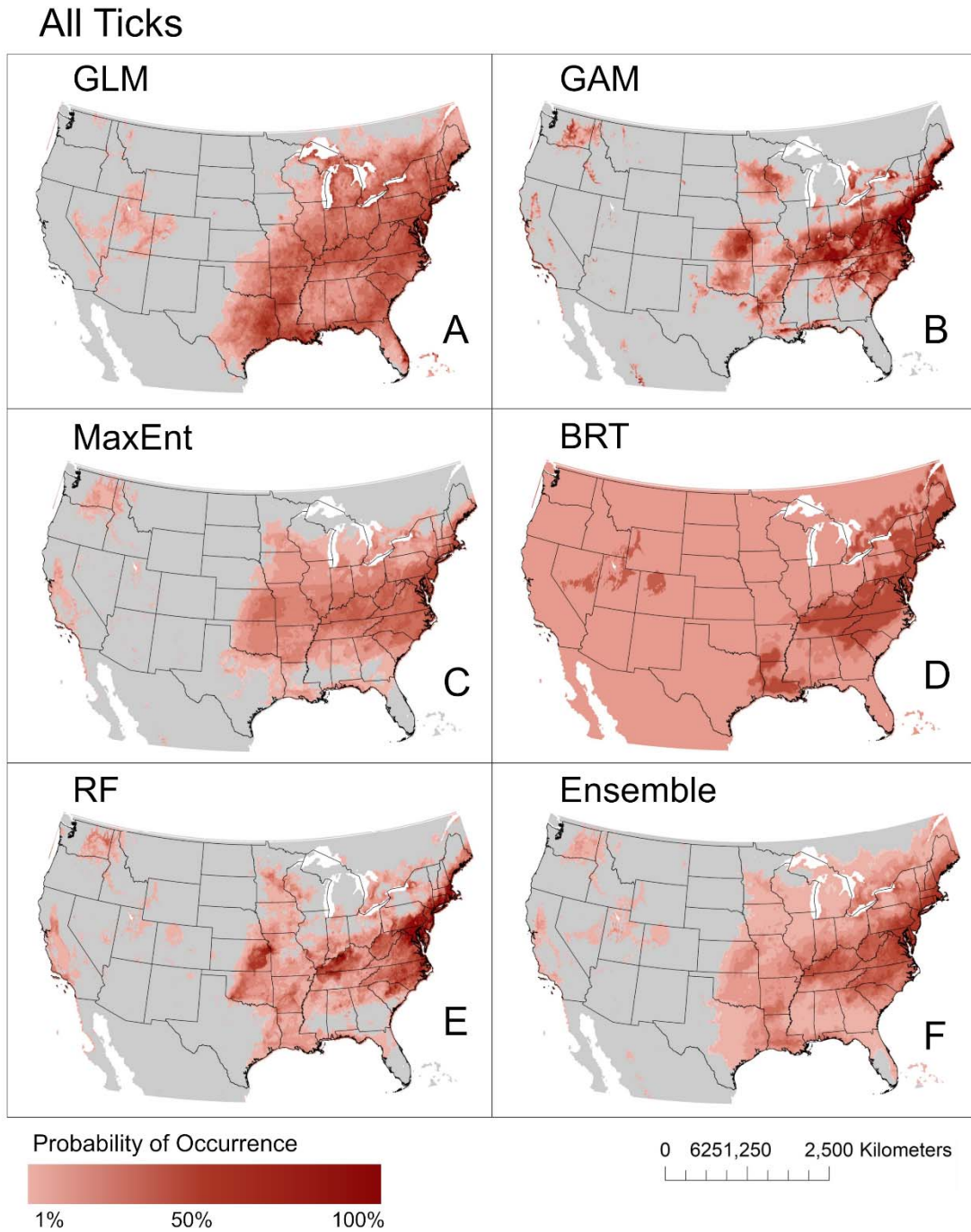
222 Table 2. Accuracy metrics for species distribution models of *Dermacentor variabilis* ticks built  
223 with five modeling methods including generalized linear model (GLM), maximum entropy  
224 (MaxEnt), generalized additive model (GAM), random forests (RF), and boosted regression trees  
225 (BRT).

<b>Method</b>	<b>Dataset</b>	<b>AUC</b>	<b>TSS</b>	<b>Deviance</b>	<b>Mean Omission</b>
GLM	Positive	0.92	0.76	0.23	0.15
	All	0.90	0.70	0.80	0.15
GAM	Positive	0.92	0.79	0.70	0.11
	All	0.95	0.79	0.55	0.12
MaxEnt	Positive	0.95	0.83	0.23	0.12
	All	0.95	0.77	0.66	0.12
BRT	Positive	0.91	0.77	0.27	0.13
	All	0.90	0.69	0.89	0.15
RF	Positive	0.93	0.81	0.20	0.10
	All	0.96	0.79	0.47	0.11

227           The overall pattern of estimated geographic distributions resulting from models built with  
228 all *D. variabilis* records was similar across methods, where the highest probabilities of suitable  
229 habitat were predicted in eastern United States, with some area in the West also identified as  
230 potentially suitable (Fig. 2). However, estimated probabilities of occurrence varied greatly across  
231 methods, with BRT (maximum 66.68%) and MaxEnt (maximum 67.08%) yielding generally low  
232 probabilities, and GLM (maximum 97.81%), GAM (maximum 99.60%), and RF (maximum  
233 99.98%) producing models with generally higher probabilities of occurrence. The averaged  
234 ensemble of model predictions yielded intermediate probabilities (maximum 81.38%). *D.*  
235 *variabilis* was predicted across models to occur with relatively high probability in the  
236 northeastern United States, including areas in the states of Delaware, Pennsylvania, New Jersey,  
237 Connecticut, Maryland, Massachusetts, and Maine, and in the southern states of Virginia, West  
238 Virginia, Kentucky, Tennessee, North Carolina, and South Carolina. Lower probabilities of  
239 occurrence (i.e. < 50%), spanning much of the Midwest and limited areas in the West, were also  
240 consistent across models. The BRT model estimates the potential range of *D. variabilis* to extend  
241 across the entire North American continent, albeit with very low probability ranging 20-30%.

242

243 **Fig. 2.** Predicted geographic distributions of *D. variabilis* ticks. Distributions were estimated  
244 using five common modeling methods including generalized linear model (GLM, A),  
245 generalized additive model (GAM, B), maximum entropy (MaxEnt, C), boosted regression trees  
246 (BRT, D), random forests (RF, E), and a weighted ensemble of these five methods (F).

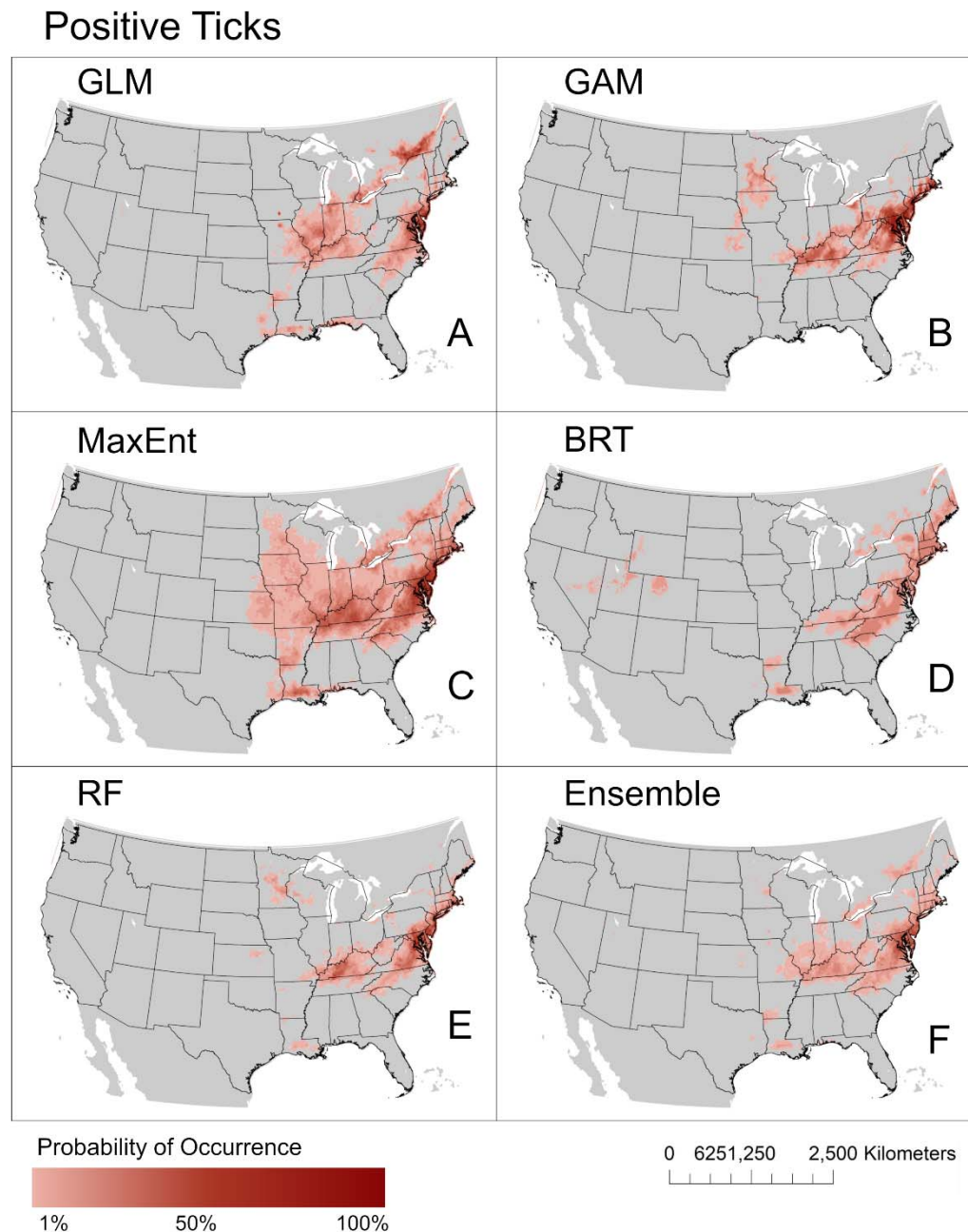


247

248

249 The predicted distribution of *D. variabilis* which tested positive for *R. montanensis* was  
250 geographically constrained compared to the distributions estimated with the full dataset (Fig. 3).

251 **Fig. 3.** Predicted geographic distributions of *D. variabilis* ticks infected with *R. montanensis*.  
252 Distributions were estimated using five common modeling methods including generalized linear  
253 model (GLM, A), generalized additive model (GAM, B), maximum entropy (MaxEnt, C),  
254 boosted regression trees (BRT, D), random forests (RF, E), and a weighted ensemble of these  
255 five methods (F).



257 This pattern of restricted distribution was observed for all projected models, regardless of  
258 methodology. Four models (GLM, GAM, MaxEnt, and RF) project the highest probability for  
259 the occurrence of positive ticks in eastern states. The GAM, MaxEnt, and RF models predicted  
260 larger geographic extents for positive ticks, where high probabilities of occurrence (> 50%)  
261 spanned portions of states including Virginia, Maryland, Delaware, Pennsylvania, Connecticut,  
262 Rhode Island, and Massachusetts. The potential range of positive ticks in the East predicted via  
263 GLM were further restricted, where high probabilities for occurrence were observed in Virginia,  
264 Maryland, and New Jersey. Portions of Kentucky were also predicted to be suitable for positive  
265 ticks with high probability with three models (GAM, MaxEnt, and RF). The estimated range of  
266 positive ticks generated with BRT also included areas spanning states in the South, the  
267 Northeast, and some locations in the West. However, maximum probabilities of occurrence for  
268 positive ticks predicted with the BRT model are very low, not exceeding 40%.

269 Environmental factors that contributed the most to model averages for *D. variabilis*  
270 presence varied across methods. The top contributing environmental variable was precipitation  
271 seasonality for the GLM (81%), BRT (84%), and RF (26%) models; mean diurnal temperature  
272 range (30%) and mean annual temperature (25%) contributed the most to GAM models; mean  
273 annual temperature (28%) was most important for MaxEnt models (STable 1). Variable response  
274 curves varied greatly between models, but generally indicate that suitability is predicted when  
275 the precipitation in the warmest quarter is high and variation in precipitation seasonality is low  
276 (SFig. 1). Precipitation seasonality was the top contributing environmental factor for predictions  
277 of ticks testing positive for *R. montanensis* across all methods and was the single most important  
278 variable for MaxEnt (72%), BRT (98%), and RF (25%) models. Precipitation seasonality (99%),  
279 mean annual temperature (63%), temperature seasonality (60%), and precipitation of the

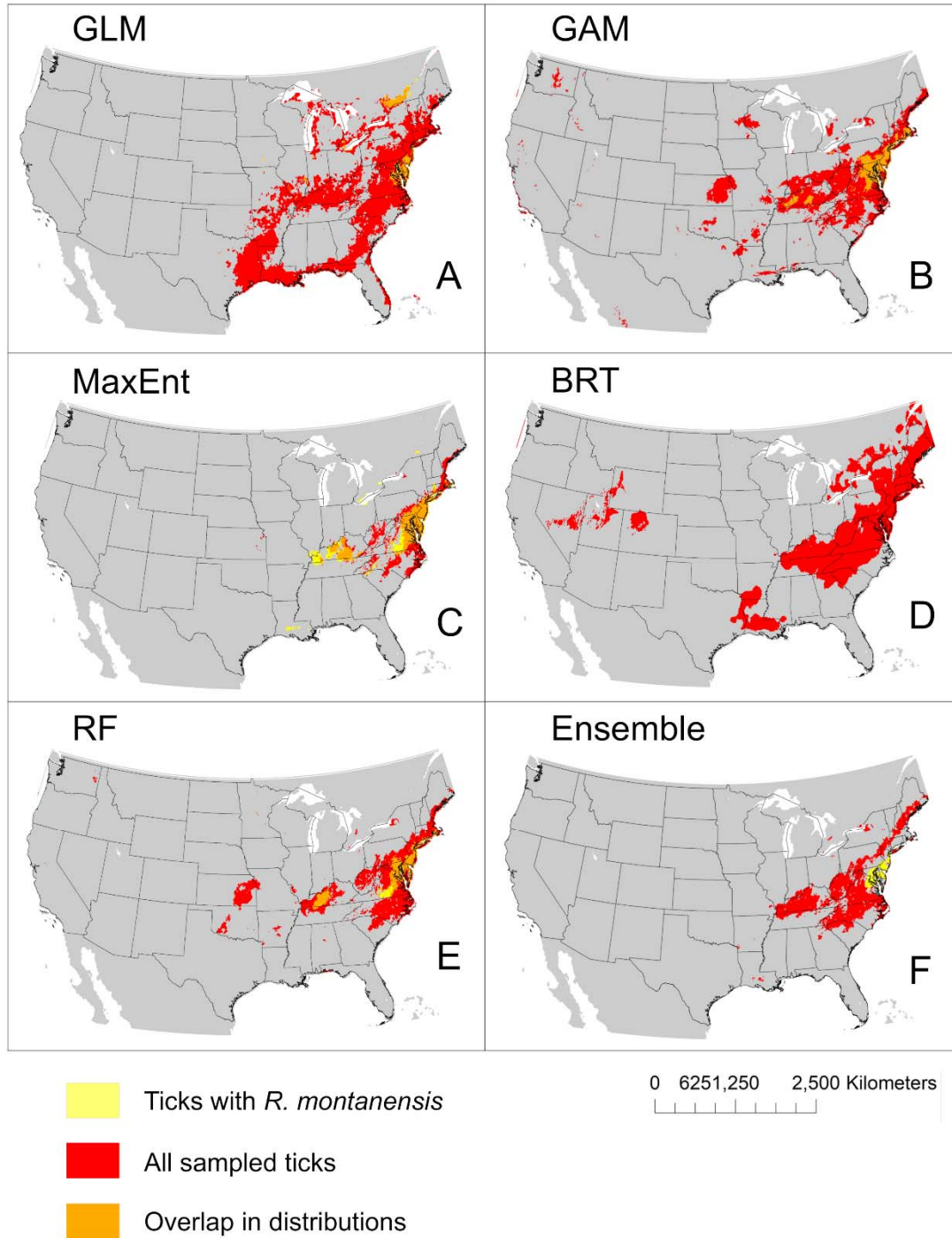
280 warmest quarter (51%) were important variables in GLM models; precipitation seasonality  
281 (68%) and precipitation of the warmest quarter (50%) were important variables in GAM models  
282 (STable 2). Low variation in precipitation seasonality was the only common variable response  
283 across models of ticks positive for *R. montanensis* (SFig. 2).

284         Applying a conservative threshold for presence (i.e. where probabilities of occurrence  
285 less than 50% were considered unsuitable habitat) highlights where potential *D. variabilis* habitat  
286 occurs with the greatest probability (Fig. 4). High model predictions for presence of *D. variabilis*  
287 are predominantly concentrated in the eastern US across methods. Ticks positive for *R.*  
288 *montanensis* were predicted in geographically limited areas within the broader range of *D.*  
289 *variabilis* in all projected models, with the exception of the BRT model, which did not produce  
290 any probabilities for positive tick occurrences that met the threshold for presence (Fig. 4D).  
291 Restricted geographic distribution of positive ticks within the broader range of *D. variabilis* was  
292 also predicted in the model ensemble, although these were distinct potential ranges that did not  
293 overlap in areas with high probabilities (Fig. 4F).

294



295 **Fig. 4.** Overlap in the predicted geographic distributions (probability of occurrence > 50%) for  
296 *D. variabilis* ticks and *D. variabilis* ticks positive for *R. montanensis* infections. Distributions  
297 were estimated using five common modeling methods including generalized linear model (GLM,  
298 A), generalized additive model (GAM, B), maximum entropy (MaxEnt, C), boosted regression  
299 trees (BRT, D), and random forests (RF, E), and a weighted ensemble of these five methods (F).



301

## 302 **Discussion**

303         The potential geographic ranges of *D. variabilis* produced in this study are largely in  
304 concurrence with maps of potential tick exposure issued by health authorities, where the ticks are  
305 reported to be widely distributed east of the Rocky Mountains, with limited distribution along the  
306 Pacific Coast (CDC 2019b). However, every modeling method, except for BRT, yielded  
307 geographic predictions for occurrence that were limited in area compared to general range maps  
308 for *D. variabilis*. These restrictions are more pronounced when locations with low probabilities  
309 of occurrence are omitted from mapped results, excluding known areas of occurrence in  
310 southern, midwestern, and Pacific coastal states (Fig. 4). *Dermacentor variabilis* with confirmed  
311 *R. montanensis* infections had constrained potential ranges within the broader geographic extent  
312 of *D. variabilis*. The potential range of positive *D. variabilis* modeled with a reduced variable set  
313 via MaxEnt in this study mirrors the results presented in St. John et al. (2016), where the highest  
314 probabilities of occurrence for positive ticks were focused in the Northeast and the Midwest,  
315 within the predicted range of *D. variabilis* that tested negative for *R. montanensis*. This restricted  
316 spatial pattern of predicted positive tick occurrences was also observed to varying degrees in  
317 models produced with three other methods (GLM, GAM, and RF).

318         *Dermacentor variabilis* is a habitat generalist, capable of exploiting a wide range of  
319 environmental conditions and host species (Sonenshine 1993). The bioclimatic variables that  
320 drove suitability predictions varied greatly between models, possibly reflecting the generalist life  
321 history of *D. variabilis*, and its ability to withstand conditions that would significantly limit other  
322 arthropod vectors. Nevertheless, indicators of seasonality, both in temperature and precipitation,  
323 were typically represented in averaged models, possibly reflecting the phenology of tick

324 reproductive cycles. The life history of this vector may also elucidate the observed discrepancies  
325 in overall predicted range for *D. variabilis* presence versus the subset of pathogen positive tick  
326 occurrences. The restricted range of ticks positive for *R. montanensis*, within the larger area of  
327 suitable *D. variabilis* habitat is possibly indicative of an underlying range restriction in host  
328 availability. The hosts involved in zoonotic transmission cycles for *R. montanensis* are unknown,  
329 as they are with many tick-borne rickettsiae (Parola et al. 2005). Thus, geographic range  
330 constraint of a competent reservoir host, determined by the host's ecological niche, is a potential  
331 driver of the constrained spatial pattern of pathogen positive ticks observed in this study.  
332 Furthermore, environmental conditions may mediate bacterial replication and transmission  
333 cycles, limiting where arthropod vectors can successfully acquire or transmit new infections  
334 (Galletti et al. 2013). Of note, while we estimated an 'infected niche', these ticks tested positive  
335 for the pathogen, but even with this nuanced information available, we do not precisely know if  
336 they were capable of onward transmission. In the Rocky Mountain wood tick, *Dermacentor*  
337 *andersoni*, there is evidence that ticks infected with *R. rickettsii* have reduced fitness (Niebylski  
338 et al. 1999). However, there are other tick-pathogen relationships that may increase fitness, such  
339 as the blacklegged tick, *Ixodes scapularis*, and the bacteria *Anaplasma phagocytophilum*, that  
340 causes the tick to express anti-freeze-like protein to enhance its survival in colder climates  
341 (Neelakanta et al. 2010). If there is an interaction between pathogens and tick fitness, this may  
342 impact the predicted suitable habitat for the infected vector.

343         Discrepancies in predicted geographic ranges are expected to arise as a result of the  
344 inherent differences in SDM algorithms. While there is considerable overlap in the projected  
345 output across methods in this study, low probabilities of occurrence are abundant throughout  
346 much of the projected ranges. These finding contrast with other published distributions of *D.*

347 *variabilis*, where MaxEnt models have projected high bioclimatic suitability throughout the  
348 range, particularly in the central and southern United States (James et al. 2015, Minigan et al.  
349 2018, Boorgula et al. 2020). Model results are also subject to differences in data inputs and  
350 limitations of data sampling, such as when the geographic extent of collections is subject to bias.  
351 This study was limited to data associated with a military installation reporting, and thus  
352 contained repeated observations at some locations – a common finding in surveillance  
353 geolocation data. Spatial thinning of occurrence records addresses potential oversampling of  
354 localities in these data, as many reported bites occurred in the vicinity of reporting military  
355 installations and medical treatment facilities. The thinned data used in model building is an  
356 appropriate representation of the full dataset of tick occurrences collected at clinics across the  
357 study area, where sufficient locations were represented while controlling for repeated  
358 observations (SFig. 3). Nevertheless, this does not address issues of potential under sampling  
359 across the full range of suitable habitats, or across life-stages, which could be underrepresented  
360 in this particular surveillance system. The majority of *D. variabilis* surveillance data are adult  
361 ticks collected through field sampling or found biting humans. Juvenile life stages of *D.*  
362 *variabilis* are rarely collected through standard field sampling methods of flagging or dragging  
363 because they live off-host, in areas that are not sampled, such as rodent burrows (Sonenshine  
364 1993). The bias of using data from sampling methods which collect a majority of adult *D.*  
365 *variabilis*, may confound our understanding of predicted *D. variabilis* ranges and the dynamics  
366 of *R. montanensis*. More data are needed from all life stages to create better understanding of  
367 tick-rickettsia range distribution.

368 Interpolated bioclimatic variables are the primary input environmental predictors used  
369 frequently throughout the SDM literature, but they may not fully describe geographically

370 limiting factors when modeling generalist tick species. Due to the prevalence of SDM studies  
371 that used BioClim data, we included bioclimatic layers in this study for comparison with other  
372 published models. However, few of these environmental variables independently contributed to  
373 predictions of *D. variabilis* presence, and those that were identified as underlying drivers for  
374 averaged model output were related to seasonality. When building SDMs for ticks, identification  
375 of better environmental predictors may be necessary, such as those related to soil moisture.  
376 While we included some proxies of soil conditions, they did not appreciably contribute to  
377 estimates for *D. variabilis* presence. The Normalized Difference Vegetation Index (NDVI), a  
378 quantification of surface vegetation using remote sensing measurements, has been used in other  
379 studies of tick distributions to approximate conditions that may be limiting to questing ticks.  
380 However, the importance of the contribution of NDVI to models in the literature has been mixed.  
381 We therefore suggest that future studies include further exploration of additional geospatial data  
382 layers and tick-host interactions, to better capture the environmental limits to tick and pathogen  
383 positive tick distributions, in order to use SDM approaches effectively.

384 This study leveraged a unique surveillance dataset to assess differences in geographic  
385 distributions between *D. variabilis* ticks, and those infected with *R. montanensis*. We have  
386 consistently demonstrated that presence of a potential vector does not inherently imply presence  
387 of the pathogen, across a range of modeling methods. This has important implications for public  
388 health agencies, which may use SDMs of vectors to infer risk and make management decisions.  
389 Moving forward, future research that uses expanded georeferenced tick surveillance data, with  
390 accompanying seroprevalence screening, will help reconcile our distribution maps with the full  
391 range of *D. variabilis* in the United States. Additional data points will allow the opportunity to  
392 better assess model performance with independent validation data that control for spatial

393 autocorrelation, as the use of holdout data can lead to inflated model accuracy metrics (Bahn and  
394 McGill 2013). Although RF was the best performing modeling algorithm in this study,  
395 methodological choices should be made with specific goals in mind, and we caution against  
396 extrapolating the model performance shown here to other datasets or geographic foci.

## 397 **Conclusion**

398 There is considerable overlap in the estimated geographic range of *D. variabilis* across modeling  
399 methods used in this study. Nevertheless, by conserving input data layers across modeling  
400 approaches, we demonstrated that differences in these predictions can arise as an artefact of  
401 methodology. These discrepancies in predicted range may be quite profound in impact and  
402 interpretation, depending on the intended application of results:for example, if these mapped  
403 model outcomes are used for communicating either tick encounter or disease risk at sub-regional  
404 scales. Further, we find that the predicted “infected niche” is smaller than the overall predicted  
405 tick niche, and thus predicted vector distributions may not best reflect human risk of acquiring a  
406 vector-borne disease. We therefore recommend caution in relying on single method SDMs, or  
407 those that imply disease risk from vector niches, to inform public health operations.

## 408 **Acknowledgments**

409 A.L.R. and H.K.S. are federal/contracted employee of the United States government. This work  
410 was prepared as part of their official duties. Title 17 U.S.C. 105 provides that ‘copyright  
411 protection under this title is not available for any work of the United States Government.’ Title  
412 17 U.S.C. 101 defines a U.S. Government work as work prepared by a military service member  
413 or employee of the U.S. Government as part of that person's official duties.

414 **Funding:** CAL, HDG, and SJR were funded by NIH 1R01AI136035-01. ALW and SJR were  
415 additionally funded by CDC grant 1U01CK000510-01: Southeastern Regional Center of  
416 Excellence in Vector-Borne Diseases: The Gateway Program. This project was also funded by  
417 the Department of Defense Global Emerging Infections System (GEIS), work unit  
418 000188M.0931.001.A0074. This publication was supported by the Cooperative Agreement  
419 Number above from the Centers for Disease Control and Prevention. Its contents are solely the  
420 responsibility of the authors and do not necessarily represent the official views of the Centers for

421 Disease Control and Prevention. The views expressed in this article reflect the results of research  
422 conducted by the author and do not necessarily reflect the official policy or position of the  
423 Department of the Navy, Department of Defense, nor the United States Government.  
424

## 425 References

- 426 **Aiello-Lammens, M. E., R. A. Boria, A. Radosavljevic, B. Vilela, and R. P. Anderson. 2015.**  
427 spThin: an R package for spatial thinning of species occurrence records for use in  
428 ecological niche models. *Ecography*. 38: 541–545.
- 429 **Araujo, M., and M. New. 2007.** Ensemble forecasting of species distributions. *Trends in*  
430 *Ecology & Evolution*. 22: 42–47.
- 431 **Bahn, V., and B. J. McGill. 2013.** Testing the predictive performance of distribution models.  
432 *Oikos*. 122: 321–331.
- 433 **Baldrige, G. D., N. Y. Burkhardt, A. S. Oliva, T. J. Kurtti, and U. G. Munderloh. 2010.**  
434 Rickettsial ompB promoter regulated expression of GFPuv in transformed *Rickettsia*  
435 *montanensis*. *PloS one*. 5: e8965.
- 436 **Blackburn, J. K., S. Matarimov, S. Kozhokeeva, Z. Tagaeva, L. K. Bell, I. T. Kracalik,**  
437 **and A. Zhunushov. 2017.** Modeling the ecological niche of *Bacillus anthracis* to map  
438 anthrax risk in Kyrgyzstan. *Am. J. Trop. Med. Hyg.* 96: 550–556.
- 439 **Boorgula, G. D. Y., A. T. Peterson, D. H. Foley, R. R. Ganta, and R. K. Raghavan. 2020.**  
440 Assessing the current and future potential geographic distribution of the American dog  
441 tick, *Dermacentor variabilis* (Say) (Acari: Ixodidae) in North America. *PLOS ONE*. 15:  
442 e0237191.
- 443 **Breiman, L. 2001.** Random Forests. *Machine Learning*. 45: 5–32.
- 444 **Burtis, J. C., J. B. Yavitt, T. J. Fahey, and R. S. Ostfeld. 2019.** Ticks as soil-dwelling  
445 arthropods: an intersection between disease and soil ecology. *Journal of Medical*  
446 *Entomology*. 56: 1555–1564.
- 447 **Carlson, C. J., E. Dougherty, M. Boots, W. Getz, and S. J. Ryan. 2018.** Consensus and  
448 conflict among ecological forecasts of Zika virus outbreaks in the United States.  
449 *Scientific Reports*. 8.
- 450 **CDC. 2010.** Spotted Fever Rickettsiosis (*Rickettsia* spp.) 2010 Case Definition (No. CSTE  
451 Position Statement: 09-ID-16). Centers for Disease Control and Prevention.
- 452 **CDC. 2019a.** Rocky Mountain Spotted Fever (RMSF) Epidemiology and Statistics.
- 453 **CDC. 2019b.** Regions where ticks live.
- 454 **Chatterjee, S., and A. S. Hadi. 2006.** Analysis of Collinear Data, pp. 221–258. *In* *Regression*  
455 *Analysis by Example*. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- 456 **De Marco, P., and C. C. Nóbrega. 2018.** Evaluating collinearity effects on species distribution  
457 models: An approach based on virtual species simulation. *PLOS ONE*. 13: e0202403.

- 458 **Elith, J., and J. R. Leathwick. 2009.** Species Distribution Models: Ecological Explanation and  
459 Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and*  
460 *Systematics*. 40: 677–697.
- 461 **Elith, J., J. R. Leathwick, and T. Hastie. 2008.** A working guide to boosted regression trees.  
462 *Journal of Animal Ecology*. 77: 802–813.
- 463 **ESRI. 2016.** ArcGIS 10.4. Environmental Systems Research Institute (ESRI), Redlands, CA.
- 464 **Evans, J. S., M. A. Murphy, Z. A. Holden, and S. A. Cushman. 2011.** Modeling Species  
465 Distribution and Change Using Random Forest, pp. 139–159. *In* Drew, C.A., Wiersma,  
466 Y.F., Huettmann, F. (eds.), *Predictive Species and Habitat Modeling in Landscape*  
467 *Ecology*. Springer New York, New York, NY.
- 468 **Fick, S. E., and R. J. Hijmans. 2017.** WorldClim 2: new 1 km spatial resolution climate  
469 surfaces for global land areas. *International Journal of Climatology*. 37: 4302–4315.
- 470 **Galletti, M. F. B. M., A. Fujita, M. Y. Nishiyama Jr, C. D. Malossi, A. Pinter, J. F. Soares,**  
471 **S. Daffre, M. B. Labruna, and A. C. Fogaça. 2013.** Natural Blood Feeding and  
472 Temperature Shift Modulate the Global Transcriptional Profile of *Rickettsia rickettsii*  
473 Infecting Its Tick Vector. *PLoS ONE*. 8: e77388.
- 474 **GDAL/OGR contributors. 2020.** GDAL/OGR Geospatial Data Abstraction software Library.  
475 Open Source Geospatial Foundation.
- 476 **Gurgel-Gonçalves, R., C. Galvão, J. Costa, and A. T. Peterson. 2012.** Geographic distribution  
477 of Chagas disease vectors in Brazil based on ecological niche modeling. *Journal of*  
478 *Tropical Medicine*. 2012: 1–15.
- 479 **Hao, T., J. Elith, G. Guillera-Aroita, and J. J. Lahoz-Monfort. 2019.** A review of  
480 evidence about use and performance of species distribution modelling ensembles like  
481 BIOMOD. *Diversity and Distributions*. 25: 839–852.
- 482 **Hardstone Yoshimizu, M., and S. A. Billeter. 2018.** Suspected and confirmed vector-borne  
483 rickettsioses of North America associated with human diseases. *Tropical Medicine and*  
484 *Infectious Disease*. 3: 2.
- 485 **Hengl, T., J. Mendes de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A.**  
486 **Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A.**  
487 **Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I.**  
488 **Wheeler, S. Mantel, and B. Kempen. 2017.** SoilGrids250m: Global gridded soil  
489 information based on machine learning. *PLOS ONE*. 12: e0169748.
- 490 **James, A. M., C. Burdett, M. J. Mccool, A. Fox, and P. Riggs. 2015.** The geographic  
491 distribution and ecological preferences of the American dog tick, *Dermacentor variabilis*  
492 (Say), in the U.S.A. *Medical and Veterinary Entomology*. 29: 178–188.
- 493 **Lippi, C. A., A. M. Stewart-Ibarra, M. E. F. B. Loo, J. E. D. Zambrano, N. A. E. Lopez, J.**  
494 **K. Blackburn, and S. J. Ryan. 2019.** Geographic shifts in *Aedes aegypti* habitat  
495 suitability in Ecuador using larval surveillance data and ecological niche modeling:  
496 Implications of climate change for public health vector control. *PLOS Neglected Tropical*  
497 *Diseases*. 13: e0007322.



- 498 **McCullagh, P., and J. A. Nelder. 1998.** Generalized linear models, 2nd ed. ed, Monographs on  
499 statistics and applied probability. Chapman & Hall/CRC, Boca Raton.
- 500 **McQuiston, J. H., G. Zemtsova, J. Perniciaro, M. Hutson, J. Singleton, W. L. Nicholson,**  
501 **and M. L. Levin. 2012.** Afebrile spotted fever group Rickettsia infection after a bite  
502 from a *Dermacentor variabilis* tick infected with *Rickettsia montanensis*. *Vector-Borne*  
503 *and Zoonotic Diseases*. 12: 1059–1061.
- 504 **Merow, C., M. J. Smith, and J. A. Silander. 2013.** A practical guide to MaxEnt for modeling  
505 species' distributions: what it does, and why inputs and settings matter. *Ecography*. 36:  
506 1058–1069.
- 507 **Minigan, J. N., H. A. Hager, A. S. Peregrine, and J. A. Newman. 2018.** Current and potential  
508 future distribution of the American dog tick (*Dermacentor variabilis*, Say) in North  
509 America. *Ticks and Tick-borne Diseases*. 9: 354–362.
- 510 **Morales, N. S., I. C. Fernández, and V. Baca-González. 2017.** MaxEnt's parameter  
511 configuration and small samples: are we paying attention to recommendations? A  
512 systematic review. *PeerJ*. 5: e3093.
- 513 **Naimi, B., and M. B. Araújo. 2016.** sdm: a reproducible and extensible R platform for species  
514 distribution modelling. *Ecography*. 39: 368–375.
- 515 **Neelakanta, G., H. Sultana, D. Fish, J. F. Anderson, and E. Fikrig. 2010.** *Anaplasma*  
516 *phagocytophilum* induces *Ixodes scapularis* ticks to express an antifreeze glycoprotein  
517 gene that enhances their survival in the cold. *The Journal of Clinical Investigation*. 120:  
518 3179–3190.
- 519 **Nicholson, W. L., and C. D. Paddock. 2019.** Rickettsial Diseases (Including Spotted Fever &  
520 Typhus Fever Rickettsioses, Scrub Typhus, Anaplasmosis, and Ehrlichioses). *In* *CDC*  
521 *Yellow Book*. Centers for Disease Control and Prevention.
- 522 **Niebylski, M. L., M. G. Peacock, and T. G. Schwan. 1999.** Lethal effect of *Rickettsia rickettsii*  
523 on its tick vector (*Dermacentor andersoni*). *Appl. Environ. Microbiol.* 65: 773–778.
- 524 **Parola, P., C. D. Paddock, and D. Raoult. 2005.** Tick-borne rickettsioses around the world:  
525 emerging diseases challenging old concepts. *Clinical Microbiology Reviews*. 18: 719–  
526 756.
- 527 **Peterson, A. T., and J. Soberón. 2012.** Species Distribution Modeling and Ecological Niche  
528 Modeling: Getting the Concepts Right. *Natureza & Conservação*. 10: 102–107.
- 529 **Phillips, S. J., and M. Dudík. 2008.** Modeling of species distributions with Maxent: new  
530 extensions and a comprehensive evaluation. *Ecography*. 31: 161–175.
- 531 **R Core Team. 2019.** R: A language and environment for statistical computing. R Foundation for  
532 Statistical Computing, Vienna, Austria.
- 533 **Sonenshine, D. E. 1993.** *Biology of Ticks*. Oxford University Press, New York.
- 534 **St. John, H. K., M. L. Adams, P. M. Masuoka, J. G. Flyer-Adams, J. Jiang, P. J. Rozmajzl,**  
535 **E. Y. Stromdahl, and A. L. Richards. 2016.** Prevalence, distribution, and development  
536 of an ecological niche model of *Dermacentor variabilis* ticks positive for *Rickettsia*  
537 *montanensis*. *Vector-Borne and Zoonotic Diseases*. 16: 253–263.

538 **Townsend Peterson, A., M. Papeş, and M. Eaton. 2007.** Transferability and model evaluation  
539 in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*. 30: 550–  
540 560.

541 **Wood, S. N. 2006.** Generalized additive models: an introduction with R, Texts in statistical  
542 science. Chapman & Hall/CRC, Boca Raton, FL.

543

544

545

546

547

548 Supplemental Materials

549 **Table S1.** Model variable importance with confidence interval limits for models built with all  
 550 *Dermacentor variabilis* ticks.

Method	Variable	Model Importance	Lower CI Limit	Upper CI Limit
GLM	Bio 1	0.29	0.26	0.32
	Bio 2	0.04	0.03	0.05
	Bio 4	0.08	0.06	0.10
	Bio 8	0.02	0.01	0.02
	Bio 9	0.01	0.00	0.01
	Bio 15	0.81	0.77	0.84
	Bio 18	0.02	-0.00	0.04
	Bio 19	0.07	0.05	0.09
	OC Dens	0.18	0.16	0.20
	WWP	0.04	0.03	0.05
GAM	Bio 1	0.25	0.22	0.29
	Bio 2	0.30	0.27	0.32
	Bio 4	0.22	0.18	0.26
	Bio 8	0.14	0.12	0.15
	Bio 9	0.11	0.09	0.13
	Bio 15	0.17	0.14	0.19
	Bio 18	0.13	0.11	0.16
	Bio 19	0.03	-0.00	0.06
	OC Dens	0.00	-0.00	0.01
	WWP	0.02	0.01	0.04
MaxEnt	Bio 1	0.28	0.26	0.30
	Bio 2	0.14	0.12	0.15
	Bio 4	0.06	0.05	0.07
	Bio 8	0.03	0.03	0.03
	Bio 9	0.06	0.05	0.08
	Bio 15	0.07	0.06	0.08
	Bio 18	0.06	0.05	0.07
	Bio 19	0.08	0.06	0.09
	OC Dens	0.01	0.00	0.01
	WWP	0.00	0.00	0.00
BRT	Bio 1	0.00	0.00	0.00
	Bio 2	0.00	0.00	0.00
	Bio 4	0.00	0.00	0.00
	Bio 8	0.00	0.00	0.00
	Bio 9	0.00	0.00	0.00
	Bio 15	0.84	0.81	0.87
	Bio 18	0.01	0.00	0.02
	Bio 19	0.01	0.00	0.01
	OC Dens	0.00	0.00	0.00

It is made available under a [CC-BY-NC-ND 4.0 International license](#) .

	WWP	0.00	0.00	0.00
RF	Bio 1	0.05	0.05	0.05
	Bio 2	0.05	0.04	0.05
	Bio 4	0.04	0.04	0.05
	Bio 8	0.02	0.02	0.02
	Bio 9	0.04	0.03	0.04
	Bio 15	0.26	0.24	0.27
	Bio 18	0.03	0.03	0.04
	Bio 19	0.04	0.03	0.04
	OC Dens	0.01	0.00	0.01
	WWP	0.01	0.01	0.01

551

552

553 **Table S2.** Model variable importance with confidence interval limits for models built with the  
 554 subset of ticks infected with *R. montanensis*.

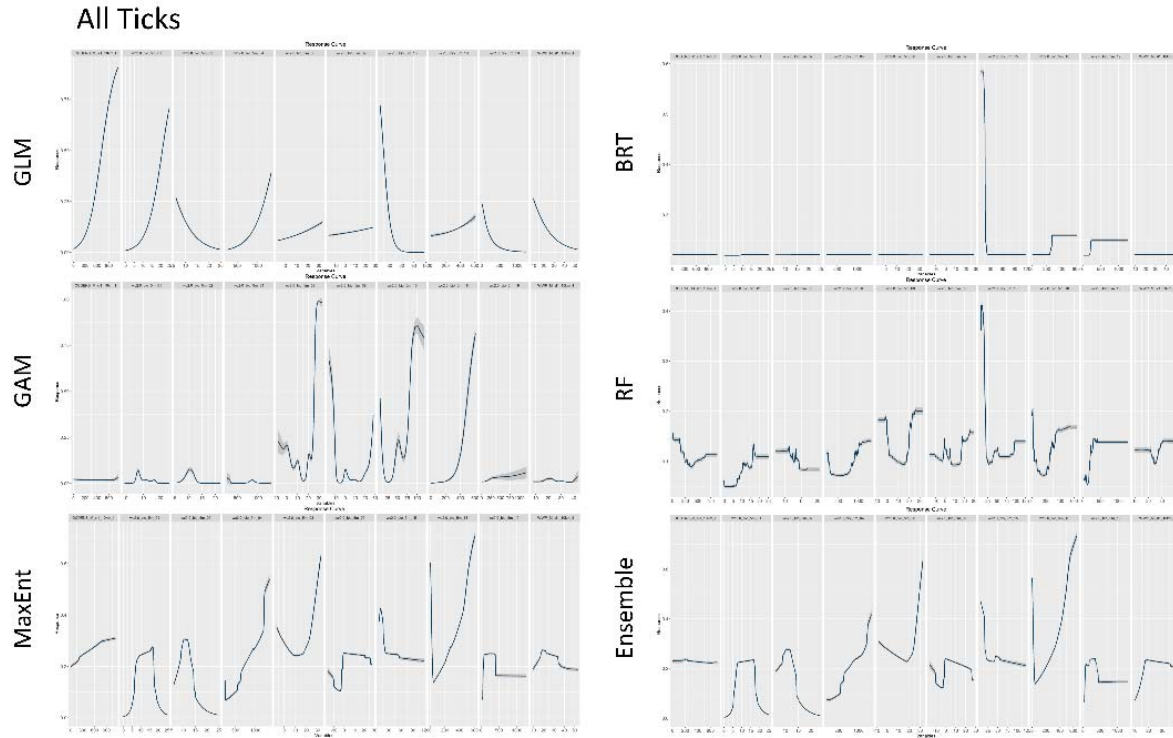
Method	Variable	Model Importance	Lower CI Limit	Upper CI Limit
GLM	Bio 1	0.63	0.60	0.66
	Bio 2	0.24	0.20	0.27
	Bio 4	0.60	0.57	0.62
	Bio 8	0.07	0.04	0.10
	Bio 9	0.15	0.10	0.20
	Bio 15	0.99	0.98	0.99
	Bio 18	0.51	0.46	0.56
	Bio 19	0.40	0.35	0.44
	OC Dens	0.05	0.02	0.07
	WWP	0.26	0.21	0.31
GAM	Bio 1	0.41	0.25	0.58
	Bio 2	0.28	0.20	0.36
	Bio 4	0.46	0.35	0.57
	Bio 8	0.13	0.02	0.24
	Bio 9	0.22	0.06	0.38
	Bio 15	0.68	0.58	0.77
	Bio 18	0.50	0.39	0.61
	Bio 19	0.26	0.13	0.38
	OC Dens	0.10	0.01	0.18
	WWP	0.17	0.07	0.26
MaxEnt	Bio 1	0.37	0.31	0.43
	Bio 2	0.18	0.15	0.21
	Bio 4	0.33	0.29	0.37
	Bio 8	0.02	0.00	0.03
	Bio 9	0.01	0.00	0.02
	Bio 15	0.72	0.69	0.75
	Bio 18	0.27	0.22	0.32
	Bio 19	0.09	0.06	0.11
	OC Dens	0.01	0.00	0.01
	WWP	0.07	0.05	0.09
BRT	Bio 1	0.00	0.00	0.00
	Bio 2	0.00	0.00	0.00
	Bio 4	0.00	0.00	0.00
	Bio 8	0.00	0.00	0.00
	Bio 9	0.00	0.00	0.00
	Bio 15	0.98	0.97	0.99
	Bio 18	0.00	0.00	0.00
	Bio 19	0.00	0.00	0.00
	OC Dens	0.00	0.00	0.00
	WWP	0.00	0.00	0.00

RF	Bio 1	0.04	0.03	0.04
	Bio 2	0.05	0.04	0.05
	Bio 4	0.05	0.04	0.05
	Bio 8	0.04	0.04	0.05
	Bio 9	0.02	0.02	0.02
	Bio 15	0.25	0.23	0.27
	Bio 18	0.04	0.03	0.04
	Bio 19	0.04	0.04	0.05
	OC Dens	0.01	0.00	0.01
	WWP	0.03	0.02	0.04

555

556

557 **Fig. S1.** Averaged variable response curves, with 95% confidence intervals, for environmental  
558 variables used to build species distribution models (SDMs) with the full dataset of *D. variabilis*  
559 occurrences. The variables that influenced the predicted presence of ticks varied between the  
560 modeling methods used, including generalized linear model (GLM), generalized additive model  
561 (GAM), maximum entropy (MaxEnt), boosted regression trees (BRT), and random forests (RF),  
562 and a weighted ensemble of these five methods.

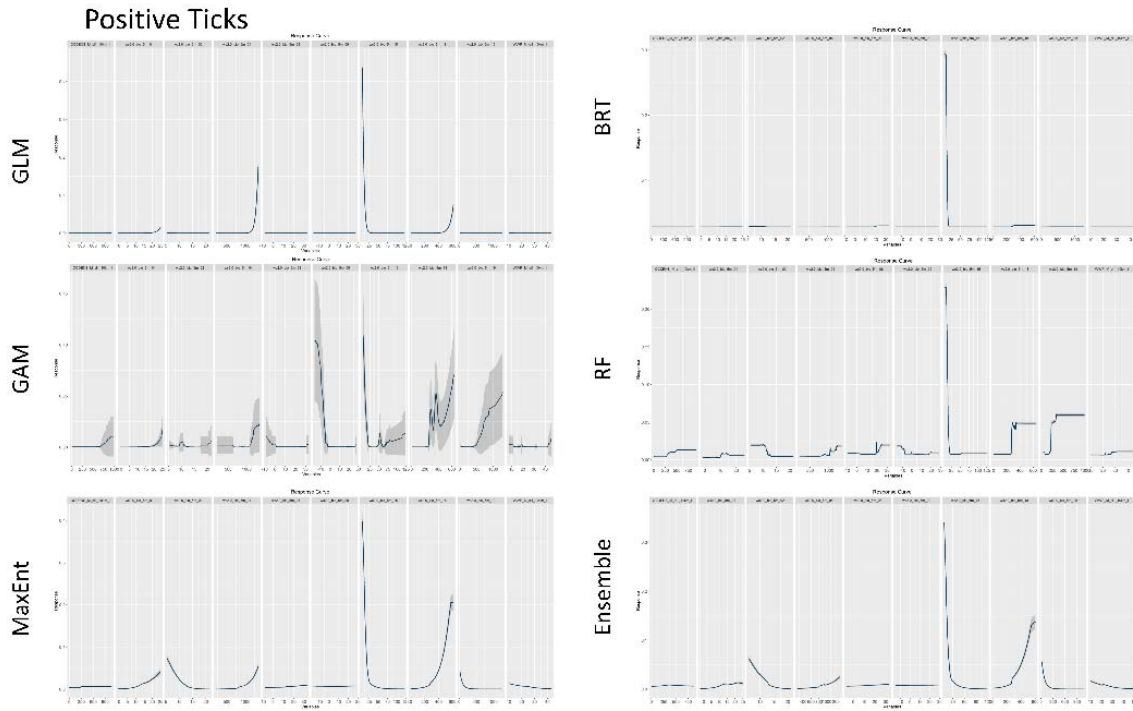


563

564

565

566 **Fig. S2.** Averaged variable response curves, with 95% confidence intervals, for environmental  
567 variables used to build species distribution models (SDMs) with the subset of *D. variabilis* ticks  
568 infected with *R. montanensis*. The variables that influenced the predicted presence of infected  
569 ticks varied between the modeling methods used, including generalized linear model (GLM),  
570 generalized additive model (GAM), maximum entropy (MaxEnt), boosted regression trees  
571 (BRT), and random forests (RF), and a weighted ensemble of these five methods.



572

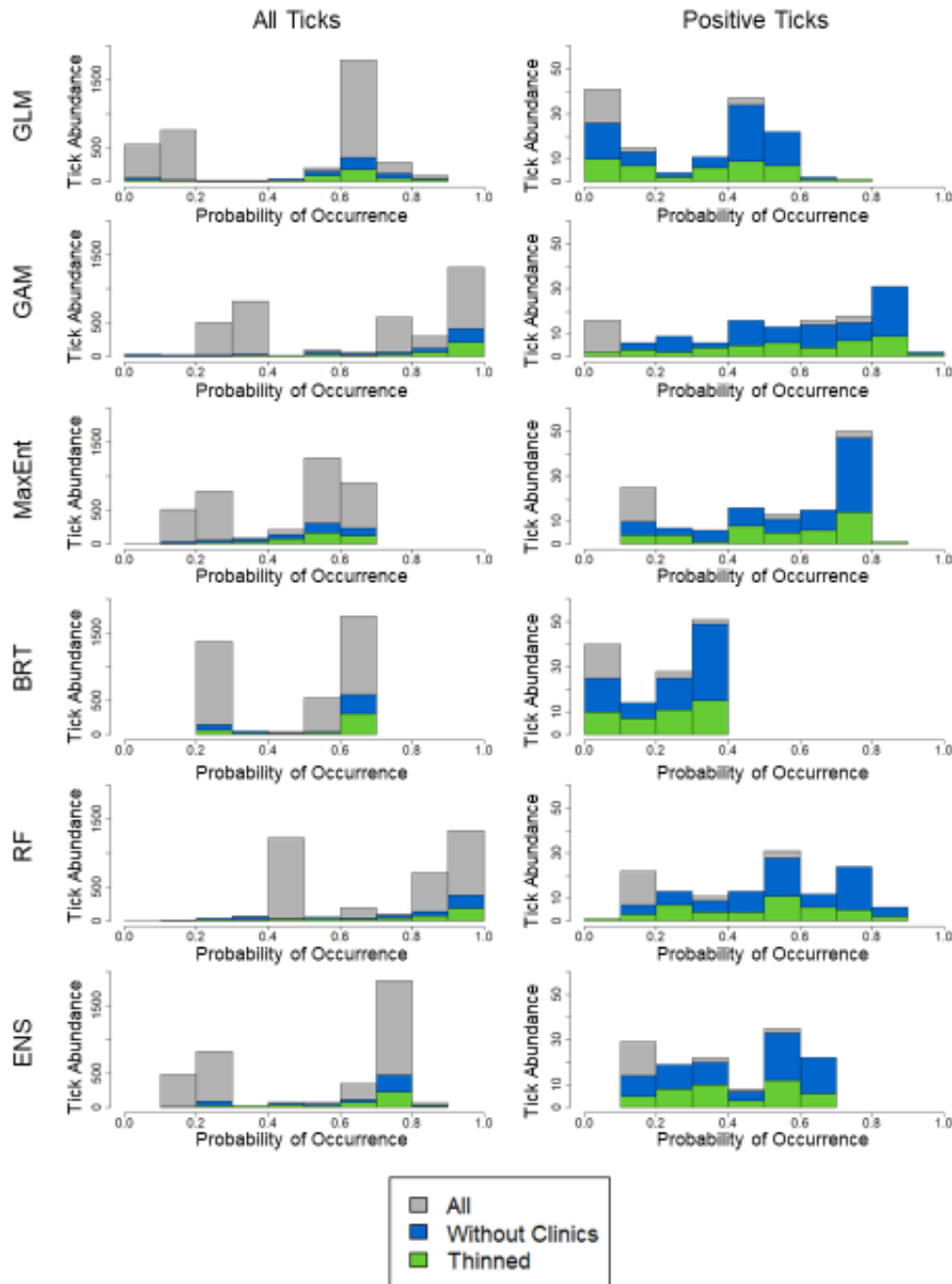
573

574

575



576 **Fig. S3.** Abundance of tick occurrence points shown against predicted probability of occurrence  
577 across modeling methods for the full dataset (grey), data with repeated observations at reporting  
578 clinic locations removed (blue), and spatially thinned data used in SDM building (green). All  
579 ticks include all *Dermacentor variabilis* ticks, and positive ticks are *D. variabilis* ticks that tested  
580 positive for *Rickettsia montanensis* infections.



581

582