MAXENT vs. MAXLIKE: Empirical Comparisons with Ant Species Distributions

Matthew C. Fitzpatrick[1,†], Nicholas J. Gotelli[2], Aaron M. Ellison[3]

[1] *Appalachian Lab, University of Maryland Center for Environmental Science, Frostburg, Maryland, 21502 USA*

[2]*Department of Biology, University of Vermont, Burlington, Vermont, 05405 USA*

[3]*Harvard Forest, Harvard University, Petersham, Massachusetts, 01366 USA*

[†]E-mail: mfitzpatrick@umces.edu

ABSTRACT

MAXENT is one of the most widely used tools in ecology, biogeography, and evolution for modeling and mapping species distributions using presence-only occurrence records and associated environmental covariates. Despite its popularity, the exponential model implemented by MAXENT does not directly estimate occurrence probability, the natural quantity of interest when modeling species distributions. Instead, MAXENT generates an index of relative habitat suitability. MAXLIKE, a newly introduced maximum-likelihood technique, has been shown to overcome the problem of directly estimating the probability of occurrence using presence-only data. However, the performance and relative merits of MAXENT and MAXLIKE remain largely untested, especially when modeling species with relatively few occurrence data that encompass only a portion of the geographic range of the species. Using geo-referenced occurrence records for six species of ants in New England, we provide comparisons of MAXENT and MAXLIKE. We show that by most quantitative metrics, the performance of MAXLIKE exceeds that of MAXENT, regardless of whether MAXENT models account for sampling bias and include nonlinear relationships – neither of which was considered in our MAXLIKE models. More importantly, for most species, the relative suitability index estimated by MAXENT was poorly correlated with the probability of occurrence estimated by MAXLIKE, suggesting that the two methods are estimating different quantities. For species distribution modeling, MAXLIKE, and similar models that are based on an explicit sampling process and that directly estimate probability of occurrence, should be considered as important alternatives to the widely-used MAXENT framework.

KEYWORDS

Ecological niche modeling; myrmecology; New England; occurrence probability; presence-only data, species distribution modeling.

INTRODUCTION

The fitting of species distribution models (SDMs) to geo-referenced species occurrence records and environmental variables is a major research activity in biogeography and ecology (Elith and Leathwick 2009, Franklin 2009). When fit with presence-only data (i.e., using only species occurrence records, not species absence records), these models generate indices proportional to habitat suitability (Phillips et al. 2006) or probability of habitat use (Boyce et al. 2002) that can be mapped in geographic space. These distribution maps have figured prominently in modeling the distributions of invasive species (Ficetola et al. 2007, Fitzpatrick et al. 2007, Ward 2007), forecasting geographic range shifts caused by climatic change (Thuiller et al. 2005, Fitzpatrick et al. 2008, Lawler et al. 2009), and in describing or estimating macroecological patterns such as species richness (Svenning et al. 2010, Mateo et al. 2012, Pottier et al. 2012). The indices of habitat suitability or habitat use predicted from presence-only SDMs are widely, but incorrectly, interpreted as estimators of the probability of species occurrence (Yackulic et al. in press). For consistency with current literature and for the purposes of comparison with actual probabilities of species occurrence, we refer here to these indices as "probability of species occurrence" or "species occurrence probabilities". However, we agree with Royle et al. (2012) that such indices are not necessarily valid estimators of the probability of species occurrence.

A variety of statistical methods are available for estimating occurrence probabilities from presence-only data (Elith et al. 2006, Franklin 2009), but by far the most widely-used has been Phillips et al.'s (2006) software implementation of MAXENT, a machine-learning algorithm based on principles of maximum entropy (Jaynes 1957). The original paper

describing MAXENT (Phillips et al. 2006) has been cited over 1200 times, with over 300 citations in 2012 alone; Elith et al. (2011) discuss the assumption underlying MAXENT, and provide a series of recipes for using the algorithm.

Royle et al. (2012) reminded ecologists that the habitat suitability indices generated by MAXENT are not direct estimators of the probability of species occurrence, which is typically the key parameter of interest when modeling species distributions. As an alternative, Royle et al. (2012) introduced MAXLIKE, a formal likelihood model that explicitly estimates the probability of species occurrence and the species' prevalence, given presence-only data and a set of environmental covariates measured at each sample location. Royle et al. (2012) also provided an R package (R Development Core Team 2012) to implement MAXLIKE (Chandler and Royle 2012).

To compare the output of MAXLIKE and MAXENT, Royle et al. (2012) used a presence-absence data set based on the occurrence of the Carolina wren (*Thryothorus ludovicianus* (Latham)) in 2222 North American Breeding Bird survey routes censused in 2006. To represent the expected distribution of species occurrence probabilities, they initially fit a logistic regression model to these presence-absence data. They next discarded the absence data, and fit the presence-only records using both MAXLIKE and MAXENT. The continental map of occurrence probabilities generated by MAXLIKE closely resembled the map generated by the logistic regression model. In contrast, the map generated by MAXENT under-estimated the "probability of occurrence" within the geographic range of the Carolina wren, but over-estimated it in areas beyond the geographic range. Royle et al.

(2012) did not report a quantitative evaluation of the predictive performance of the models however.

Royle et al.'s (2012) results suggested that the logistic output of MAXENT may differ substantially from underlying occurrence probabilities, but it is unclear for several reasons whether their results can be generalized to the much larger body of empirical studies that have used MAXENT. First, the sample size in Royle et al.'s (2012) artificial data set was much larger than the sample sizes commonly used by MAXENT practitioners and seen in published studies (e.g., Pearson et al. 2006, Papeş and Gaubert 2007, Wisz et al. 2008). Second, Royle et al. (2012)'s data set encompassed most of the geographic range of the Carolina wren. In contrast, many empirical analyses using MAXENT are based on incomplete censuses that encompass only a portion of the geographic range of the species (e.g., DeMatteo and Loiselle 2008, Trisurat et al. 2011). Finally, to fit structurally-equivalent MAXENT and MAXLIKE models to their data set, Royle et al. (2012) were forced to modify MAXENT'S default settings and disable all feature classes except for "linear" and "quadratic" (see Elith et al. 2011 for details regarding feature classes). Most published analyses use the default settings, which implement multiple feature classes determined by the number of occurrence records. Phillips and Dudík (2008) found that, when analyzing "high-quality" empirical data sets, use of MAXENT's default settings substantially improved model performance.

Other than the Carolina wren data set assembled by Royle et al. (2012), we are not aware of other published comparisons of the performance of MAXENT and MAXLIKE with empirical

data. Such comparisons are important because empirical data sets are often characterized by modest sample sizes, limited geographic coverage, and non-random locations of sampling points. With these kinds of limitations, it is unknown whether MAXENT and MAXLIKE predictions will differ substantially and exhibit the same kinds of differences that emerged in the analyses presented by Royle et al (2012).

In this study, we compared MAXENT and MAXLIKE species distribution models for six species of ants in New England, with occurrence records derived from a recent comprehensive compilation (Ellison et al. 2012). For each of six species, we asked: 1) How do MAXENT and MAXLIKE distribution maps compare for both the mean and the variance of the probability of occurrence? 2) For both sampled and unsampled locations, what is the relationship between the probabilities of occurrence estimated by MAXENT and those predicted by MAXLIKE? 3) How do the mapped predictions of MAXENT and MAXLIKE differ in several goodness-of-fit statistics that are used to quantify model performance? 4) How do the mapped predictions of MAXENT and MAXLIKE compare to expectations based on expert knowledge about the distribution of these species in unsampled areas of New England?

METHODS*Ant occurrence data*

Ant locality records were derived from field collections (Ellison et al. 2002, 2012, Gotelli and Ellison 2002) and museum records with accurate, georeferenced, collection data (Ellison et al. 2012). Each record of a presence (Table 1) represents a collection from a single nest, an individual pitfall trap, or a collection at a single bait. These records encompass collections from a variety of sources and time periods, including museum

128 records and standardized ecological sample surveys and are part of a larger dataset on the
129 distribution of 132 species of ants in New England.
130
131 *Test species*
132 Of the 132 species in the ants of New England dataset, we considered as case studies six
133 species of ants common in the six New England (northeastern U.S.) states (Maine, New
134 Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island), but which differ in their
135 geographic distribution, range size, and number of occurrence records (Table 1). These
136 case studies included a circumboreal species for which New England is in the southern part
137 of its range (*Camponotus herculeanus* (L.)), a southern North American species for which
138 New England is in the northern part of its range (*Prenolepis imparis* (Say)), three
139 widespread, commonly collected North American species for which New England is in the
140 center of its range (*Camponotus novaeboracensis* (Fitch), *Formica integra* (Nylander),
141 *Monomorium emarginatum* (DuBois)), and a North American habitat specialist, the
142 sandplain-inhabiting *Pheidole pilifera* (Roger). See Ellison et al. (2012) for additional details
143 on the natural history of these six ant species and the broader dataset.
144
145 *Environmental data*
146 To avoid over-fitting models with the small number of occurrence records available for
147 some of the study species (Table 1), we used only a small set of 20 potential environmental
148 covariates: elevation (meters above sea level) and 19 bioclimatic variables from the
149 WorldClim database (http://www.worldclim.org, Hijmans et al. 2005) that measure
150 minima, maxima, and seasonality in temperature (°C) and precipitation (mm) at a spatial

resolution of 30 arc-seconds (≈1 × 1 km). We reduced this full set of covariates by removing those covariates that exhibited little spatial variability across the study region (BIO3, BIO8, BIO9, BIO13). We then selected covariates to minimize multicollinearity ($r < 0.7$), but retained correlated pairs of variables that were, in our opinion, biologically informative. This selection process reduced the 20 covariates to three—mean annual temperature (BIO1), mean annual precipitation (BIO12), and elevation—that were used in model fitting and prediction. Temperature is broadly correlated with patterns of ant diversity and abundance (Sanders et al. 2007), elevation is a strong predictor of ant species distribution in the New England region (Gotelli and Ellison 2002), and ant foraging activity in some New England species is associated with precipitation (Nuss et al. 2005). Five of the six ant species we used in our analyses are habitat generalists whose distributions are constrained primarily by these habitat variables; the sixth, *Pheidole pilifera*, is a warm-climate species restricted to sandy soils (Ellison et al. 2012). Prior to analysis, all environmental covariates were standardized to have a mean of zero and unit variance following the recommendations of Royle et al. (2012). Our emphasis in these analyses was not to select the optimal set of variables for modeling ant distributions, but to compare the performance of MAXENT and MAXLIKE with an identical set of predictor variables.

Comparison between MAXENT *and* MAXLIKE

We modeled distributions of each of the six ant species using MAXENT and MAXLIKE and compared the resulting habitat suitability index (MAXENT; logistic output) with estimates of probability of species occurrence (MAXLIKE; $\psi(x)$). Occurrence data for each species were partitioned randomly 50 times into calibration (75%) and evaluation (25%) datasets and

50 MAXENT and MAXLIKE models for each species were fit and evaluated using the same random training and testing datasets. Our primary comparisons involved MAXENT and MAXLIKE models that considered linear effects only and which did not account for sampling bias. However, we also assessed the influences of model complexity and sampling bias on MAXENT performance relative to MAXLIKE. To assess model complexity, we additionally fit MAXENT models using the default settings, which automate the implementation of more complex model feature classes (quadratic, product, hinge, and threshold) depending on the number of occurrence records.

For each type of feature implementation (linear-only and default), we also fit MAXENT models that accounted for sampling bias by selecting background data with the same underlying bias as the ant occurrence data (target group background; Phillips et al. 2009). To generate the sampling bias surface, we totaled the number of ant occurrence records (using the full dataset of 132 species) found within each grid cell and then extrapolated these data across the study region using kernel density estimation as implemented in the sm package (Bowman and Azzalini 2010) of the R statistical language (http://r-project.org/). Lastly, we generated 10,000 background points comprised of random locations weighted by the sampling bias surface (Elith et al. 2010). Otherwise, we fit MAXENT models using the default values as implemented in the dismo package (Hijmans et al. 2012) and MAXENT 3.3.3E. MAXLIKE models were fit using the maxlike package (Chandler and Royle 2012) using the "SANN" method and a maximum of 10,000 iterations to maximize the log-likelihood function. The resulting species distribution maps illustrate the average predicted probability from the 50 models for each species; uncertainty is

illustrated with maps for each species of the standard deviation of the predicted probability from the 50 fitted models. All analyses were performed in R 2.15.1 (R Development Core Team 2012). To provide an independent check of our R-scripts, we also ran analyses using the MAXENT GUI and obtained identical results. All data and code are available through the Harvard Forest Data Archive (http://harvardforest.fas.harvard.edu/data-archive), dataset HF-147.

Model evaluation

We evaluated model outputs in terms of their statistical fit to the training data, their spatial predictions of occurrence relative to testing data, and our professional judgment. To assess the relative goodness of fit of the MAXENT and MAXLIKE models, we used the sample-size corrected Akaike information criteria (AIC_c). For MAXLIKE, AIC_c was calculated directly from the maximized log-likelihood term, whereas for MAXENT we calculated AIC_c using the approached described by Warren and Seifert (2011). Thus, each of the 50 MAXLIKE and MAXENT models implementing linear features and fit using the 50 training datasets for each species had an associated AIC_c, from which we determined the normalized Akaike model selection weight.

The evaluation of the predictive accuracy of presence-only species distribution models is an ongoing challenge; we focused primarily on evaluation criteria that require only information on presence (Franklin 2009). First, we identified the minimum predicted area (MPA; Engler et al. 2004), which is the proportion of the study area predicted as present using the probability threshold required to correctly predict as present a user-defined

proportion of the test data. Here, we set this proportion to 95%. Models that yield a lower MPA are considered superior (Engler et al. 2004, Franklin 2009). In essence, MPA assumes that a good presence-only SDM should predict a spatial distribution that is as small as possible, while correctly predicting a maximum number of observed occurrences as present. In addition to MPA, we compared the mean predicted probability of occurrence from MAXENT and MAXLIKE at known presences and at locations selected at random across New England. We also report AUC (area under the receiver-operator curve (ROC); Fielding and Bell (1997)), which is widely used to evaluate the predictive performance of presence-only SDMs in combination with "background" or pseudo-absence data. However, when used in such contexts, AUC must be interpreted cautiously because it assumes that the costs of over-prediction and under-prediction are equivalent. Because pseudo-absences represent locations where no data are available, not necessarily locations where the species has not been detected, there is little justification for penalizing over- and under-prediction equivalently. In practice, however, presence-only data can inform only under-prediction. Lobo et al. (2008), Peterson et al. (2008), and Jiménez-Valverde (2012) discuss these and other issues arising with the application of AUC to SDMs. Differences in model outputs and evaluation metrics produced by MAXENT and MAXLIKE were tested using Wilcoxon signed-rank tests for related samples.

RESULTS

The number of training records ranged from a maximum of 201 for *Camponotus novaeboracensis* to a minimum of five for *Pheidole pilifera* (mean = 66; Table 1). Model comparison by AIC_c and normalized Akaike model selection weights revealed that for all six

243 ant species, MAXLIKE models were better supported by the data than MAXENT models
244 implementing linear features with or without sampling bias correction (Table 1). However,
245 model evaluation by AUC was inconsistent, with MAXLIKE scoring lower, equal, or greater
246 AUC values than MAXENT, depending on the ant species considered and whether MAXENT
247 models were fit using default settings or restricted to linear features, and whether
248 sampling bias was accounted for or not (Fig. 1). In general, MAXENT models that accounted
249 for sampling bias scored lower or equal AUC values than MAXENT models without bias
250 correction.
251
252 By default, the MAXENT algorithm assumes a baseline species prevalence of 0.5 (Phillips and
253 Dudík 2008), and therefore assigned a probability of occurrence close to 0.5 to most
254 occurrence locations. In contrast, MAXLIKE assigned substantially higher probabilities to
255 locations with recorded presences for five of six species than did any of the
256 implementations of MAXENT (Fig. 2a). For randomly chosen background locations (Fig. 2b),
257 MAXLIKE also tended to generate higher average probabilities of occurrence than MAXENT,
258 although accounting for sampling bias increased average probabilities at random
259 background locations. Randomly-chosen background points also had nearly constant
260 probabilities of occurrence with MAXENT, although the value of the mean probability
261 differed among species; MAXENT models implementing default features tended to generate
262 lower probabilities than MAXENT models implementing only linear features. In contrast,
263 MAXLIKE usually generated a larger range of different probabilities for both occurrence and
264 background locations.
265

For all species except *Prenolepis impairs* and *Monomorium emarginatum*, there were weak correlations between the predictions of species occurrence probabilities from MaxLike and MaxEnt for either occurrence or background locations (Fig. 3). Accounting for sampling bias weakened correlations for all species. Consistent with these findings, mapped predictions from MaxLike (Fig. 4a-f) usually predicted larger areas of higher probability of occurrence than did MaxEnt (Fig. 4g-r). MaxEnt models that accounted for sampling bias tended to increase the area of higher predicted probability of occurrence to some extent (Fig. 4m-r), and, for the two species of *Camponotus* (Fig. 4m, n), largely reversed the south-north trend of increasing occurrence probability predicted by MaxEnt without bias correction (Fig. 4g, h). However, the MaxLike distribution maps also exhibited larger standard deviations in the probability of occurrence and greater uncertainty in predictions across large areas of the study region (Fig. 5a-f). In contrast, MaxEnt had lower standard deviations and uncertainty (Fig. 5g-r).

For all species of ants except *Camponotus herculeanus*, MaxLike models had either a smaller or equivalent mean MPA than MaxEnt, regardless of the feature class implementation and whether sampling bias was accounted for or not, the latter of which tended to increase MPA (Fig. 6a). However, MaxLike exhibited much greater variability than MaxEnt in the probability threshold required to predict 95% of known occurrences as present (Fig. 6b). In instances when differences in probability thresholds between MaxLike and MaxEnt were significant, MaxLike had a higher probability threshold than MaxEnt, except for bias-corrected models for *Monomorium emarginatum*.

DISCUSSION

Our results reinforce Royle et al.'s (2012) comparisons of model output for MAXENT versus MAXLIKE. Specifically, MAXENT tends to under-estimate the probability of occurrence within areas of observed presences, but over-estimates it in unsampled areas beyond the spatial coverage of the data (Fig. 4). Accounting for sampling bias did not fix this issue and, by our measures, tended to result in less robust models, an issue we return to below. In contrast, for 5 of 6 species, MAXLIKE assigned high probabilities of occurrence to areas within the spatial coverage of known occurrence and much lower probabilities elsewhere. Royle et al.'s (2012) example was based on a sample of thousands of presence-absence records measured at a continental scale (see Figure 4 in Royle et al. 2012), but we obtained similar results for more typical small data sets of dozens or hundreds of presence-only records measured over a limited geographic area (Fig. 4).

It is problematic that MAXENT rarely predicts any areas with a high probability of occurrence ($p > 0.80$) and typically generates a relatively narrow distribution of occurrence probabilities of mean $p \cong 0.5$ for recorded presences. These probabilities depend on the assumed value of species prevalence (MAXENT default = 0.5); different values would produce different results, but species prevalence is not estimated from the data by MAXENT nor is there an objective criterion for assuming one value over another. In contrast, MAXLIKE usually generates a broader range of occurrence probabilities, with generally higher occurrence probabilities at observed sample locations compared to randomly chosen background samples (Fig. 2). The fact that the output from MAXENT and MAXLIKE are poorly correlated for most data sets (Fig. 3) suggests that the two models are estimating

312 different quantities. In other words, MAXLIKE estimates probability of occurrence, while
313 MAXENT estimates a relative suitability index that, for five of six species in our study,
314 neither represents probability of occurrence nor is correlated with it.
315
316 Our goodness-of-fit statistics (Table 1) and other evaluation metrics (Fig. 6) generally
317 favored the MAXLIKE formulation, although AUC (Fig. 1) was equivocal. However, given the
318 documented issues with AUC and pseudo-absence data (Lobo et al. 2008, Peterson et al.
319 2008, Jiménez-Valverde 2012), the interpretation of AUC is problematic. In essence,
320 MAXLIKE would have a lower AUC than MAXENT simply because it tends to predict higher
321 probabilities of occurrence across the spatial extent of the occurrence data than MAXENT
322 and therefore will assign higher probabilities to a greater number of pseudo-absence
323 locations. However, an unknown percentage of these pseudo-absences are actually
324 instances of presence and therefore there is little justification for penalizing presumed
325 over-prediction at the same cost as under-predicting known occurrences.
326
327 We also note that, at least for ants of New England, the predicted species distributions from
328 MAXLIKE are more sensible and in line with our expectations from over 15 years of field
329 experience in this system (e.g., Gotelli and Ellison 2002, Ellison et al. 2012). For example,
330 the likely distribution of the widespread carpenter ant, *Camponotus novaeboracensis*, is
331 captured well by MAXLIKE (Fig. 4b), but not nearly as well by MAXENT. In particular, MAXENT
332 without accounting for sampling bias down-weights the probability of occurrence of *C.*
333 *novaeboracensis* in north central New England where it is actually widespread (compare
334 Figs. 4b and 4h). Accounting for sampling bias produces higher predicted probabilities of

occurrence in northern New England, but results in lower predicted probability of occurrence in southern New England (Fig. 4n) and reduced model performance. Similarly, whereas both MAXENT without sampling-bias correction and MAXLIKE inaccurately predict the likely absence of the circumboreal *C. herculeanus* in northern Maine (Fig. 4a, g), the MAXLIKE predictions have much higher uncertainty in this region (Fig. 5a) – which accurately reflects the sparse data – than do the predictions from MAXENT (Fig. 5g, m). As for *C. novaeborancensis*, accounting for sampling bias increases the predicted probability of occurrence of *C. herculeanus* in northern New England, but reduces it in the south (Fig. 4m) to the detriment of model performance.

MAXLIKE is not without its own set of problems, however. For some species, the output from different training and testing partitions of the same data set varied greatly, leading to large standard deviations in mapped probabilities of occurrence, especially in regions where no sample data were recorded (Fig. 5a-f). However, this is perhaps a fair representation of the uncertainty inherent in predicting species distributions to unsampled regions using presence-only data and small sample sizes. In contrast, the MAXENT projections were largely invariant with different data runs and even in unsampled areas of the geographic domain (Fig. 5g-r). This invariance may reflect the precision of the machine-learning algorithm, but yields a greater degree of certainty than perhaps the data warrant. In a few cases, MAXLIKE models generated inappropriately low estimates of occurrence probability for sites that contained occurrence records (e.g., *Formica integra* in Fig. 4c). On the other hand, MAXLIKE accurately identified the climatic envelope of the warm-climate, sandplain specialist *Pheidole pilifera* (Fig. 4e), but in the absence of a data layer for soil type,

358 overpredicted (albeit with little confidence – Fig. 5e) its probability of occurrence in most
359 locations in southern New England. However, MAXENT underpredicted its occurrence in its
360 true range and overpredicted its occurrence further north (Fig. 4k), especially when
361 models accounted for sampling bias (Fig. 4q), and with little uncertainty (Fig. 5k, q).
362
363 Both MAXENT and MAXLIKE assume random sampling, which is rarely possible with species
364 occurrence records. For example, counties throughout central and eastern Massachusetts
365 are more thoroughly sampled for ants than some other areas of New England because of
366 the large number of myrmecologists historically associated with Harvard University
367 (Ellison et al. 2012). Approaches for accounting for such sampling bias, including strategies
368 for the selection of background points, are relatively well developed for MAXENT (e.g.,
369 Phillips et al. 2009, VanDerWal et al. 2009), but remain unexplored for MAXLIKE.
370
371 For MAXENT, a method for accounting for sampling bias involves using all occurrence
372 records for a taxon of interest within a study to estimate relative survey effort and to select
373 background data with same underlying bias present in the occurrence data. This method,
374 known as "target-group background," (Phillips et al. 2009) has been shown to generally
375 improve performance of MAXENT models when averaged across all species (e.g., Mateo et al.
376 2010, Syfert et al. 2013), but not necessarily for all species or regions (Phillips et al. 2009).
377 We found that accounting for sampling bias generally did not improve MAXENT's
378 performance, and in some cases resulted in less robust models (Figures 2, 4). The
379 immediate reasons for the reduction in model performance are not clear, but Phillips et al
380 (2009) found that the improvement in model performance realized when accounting for

sampling bias was positively related to the strength of bias in the target-group presence records. We speculate that the six species we modeled had comparatively little sampling bias relative to that present in the full target group of 132 recorded New England ant species. To investigate this further, we fit additional MAXENT models with a target-group background based only on the six modeled species. We found that model performance declined for two species and marginally improved for three species relative to the full target group. However, these changes were small and model performance still did not exceed that of models without sampling bias correction. How sampling biases influence the relative performance of MAXLIKE and MAXENT is unknown and requires further study.

Finally, it is also unknown how relative performance is affected by variable selection, routines for which are not implemented in the current version of MAXLIKE. Given that several of our study species had few occurrence records and because we wished to emphasize the relative performance of MAXENT and MAXLIKE when both models were given an identical set of environmental variables as input, we were limited to a relatively small set of environmental variables. MAXLIKE projections also will be biased if the relationship between covariates and detection errors differs from the relationship between covariates and the probability of occurrence (Dorazio 2012). This potential issue, and many of these others we have identified, are common to all species distribution models, and are not unique to MAXLIKE. For species distribution modeling, MAXLIKE — and other models that are based on an explicit sampling process (Warton and Shepherd 2010, Dorazio 2012) — should be considered as important alternatives to the widely-used MAXENT framework.

ACKNOWLEDGEMENTS

MCF was supported by US NSF award DEB-1257164 and by funding from UMCES. NJG was supported by US NSF awards DEB-026575, DEB-027478, and DEB 11-36644, and the US Department of Energy award DE-FG02-08ER64510. AME was supported by NSF award DEB 11-36646 and DEB 12-37491, and the US Department of Energy award DE-FG02-08ER64510. This is UMCES Appalachian Laboratory Scientific Contribution No. XXXX and is a contribution from the Harvard Forest Long Term Ecological Research Site. We thank Robert Anderson and two anonymous reviewers for helpful comments.

LITERATURE CITED

Bowman, A. W. and A. Azzalini 2010. sm: Nonparametric smoothing methods. Retrieved from http://CRAN.R-project.org/package=sm.

Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. Ecological Modelling 157:281–300.

Chandler, R. B., and J. A. Royle. 2012. maxlike: Model species distributions by estimating the probability of occurrence using presence-only data. Retrieved from http://CRAN.R-project.org/package=maxlike.

DeMatteo, K. E., and B. A. Loiselle. 2008. New data on the status and distribution of the bush dog (*Speothos venaticus*): Evaluating its quality of protection and directing research efforts. Biological Conservation 141:2494–2505.

Dorazio, R. M. 2012. Predicting the geographic distribution of a species from presence-only data subject to detection errors. Biometrics 68:1303-1312.

426 Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from
427 occurrence data. Ecography 29:129–151.
428 Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and
429 prediction across space and time. Annual Review of Ecology, Evolution, and Systematics
430 40:677–697.
431 Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species.
432 Methods in Ecology and Evolution 1:330–342.
433 Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical
434 explanation of MaxEnt for ecologists. Diversity and Distributions 17:43-57.
435 Ellison, A. M., E. J. Farnsworth, and N. J. Gotelli. 2002. Ant diversity in pitcher-plant bogs of
436 Massachusetts. Northeastern Naturalist 9:267–284.
437 Ellison, A. M., N. J. Gotelli, G. D. Alpert, and E. J. Farnsworth. 2012. A field guide to the ants of
438 New England. Yale University Press, New Haven, Connecticut.
439 Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the
440 distribution of rare and endangered species from occurrence and pseudo-absence data.
441 Journal of Applied Ecology 41:263–274.
442 Ficetola, G. F., W. Thuiller, and C. Miaud. 2007. Prediction and validation of the potential
443 global distribution of a problematic alien invasive species - the American bullfrog.
444 Diversity and Distributions 13:476–485.
445 Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction
446 errors in conservation presence/absence models. Environmental Conservation 24:38–
447 49.

448 Fitzpatrick, M. C., J. F. Weltzin, N. J. Sanders, and R. R. Dunn. 2007. The biogeography of
449 prediction error: why does the introduced range of the fire ant over-predict its native
450 range? Global Ecology and Biogeography 16:24–33.
451 Fitzpatrick, M. C., A. D. Gove, N. J. Sanders, and R. R. Dunn. 2008. Climate change, plant
452 migration, and range collapse in a global biodiversity hotspot: the *Banksia* (Proteaceae)
453 of Western Australia. Global Change Biology 14:1337–1352.
454 Franklin, J. 2009. Mapping species distributions: spatial inference and prediction.
455 Cambridge University Press Cambridge, UK.
456 Gotelli, N. J., and A. M. Ellison. 2002. Biogeography at a regional scale: Determinants of ant
457 species density in New England bogs and forests. Ecology 83:1604–1609.
458 Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution
459 interpolated climate surfaces for global land areas. International Journal of Climatology
460 25:1965–1978.
461 Hijmans, R. J., S. J. Phillips, J. R. Leathwick, and J. Elith. 2012. dismo: Species distribution
462 modeling. R package version 0.7-23. Retrieved from http://CRAN.R-
463 project.org/package=dismo.
464 Jaynes, E. T. 1957. Information theory and statistical mechanics. Physical review 106:620.
465 Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating
466 characteristic curve (AUC) as a discrimination measure in species distribution
467 modelling. Global Ecology and Biogeography 21:498–507.
468 Lawler, J. J., S. L. Shafer, D. White, P. Kareiva, E. P. Maurer, A. R. Blaustein, and P. J. Bartlein.
469 2009. Projected climate-induced faunal change in the Western Hemisphere. Ecology
470 90:588–597.

471 Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the
472 performance of predictive distribution models. Global Ecology and Biogeography
473 17:145–151.
474 Mateo, R. G., T. B. Croat, Á. M. Felicísimo, and J. Muñoz. 2010. Profile or group discriminative
475 techniques? Generating reliable species distribution models using pseudo-absences and
476 target-group absences from natural history collections. Diversity and Distributions
477 16:84-94.
478 Mateo, R. G., Á. M. Felicísimo, J. Pottier, A. Guisan, and J. Muñoz. 2012. Do stacked species
479 distribution models reflect altitudinal diversity patterns? PLoS ONE 7:e32586.
480 Nuss, A. B., D. R. Suiter, and G. W. Bermett. 2005. Continuous monitoring of the black
481 carpenter ant, *Camponotus pennsylvanicus* (Hymenoptera: Formicidae), trail behavior.
482 Sociobiology 45: 597-618.
483 Papeș, M., and P. Gaubert. 2007. Modelling ecological niches from low numbers of
484 occurrences: assessment of the conservation status of poorly known viverrids
485 (Mammalia, Carnivora) across two continents. Diversity and Distributions 13:890–902.
486 Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. Townsend Peterson. 2006. Predicting
487 species distributions from small numbers of occurrence records: a test case using
488 cryptic geckos in Madagascar. Journal of Biogeography 34:102–117.
489 Peterson, A. T., M. Papes, and J. Soberón. 2008. Rethinking receiver operating characteristic
490 analysis applications in ecological niche modeling. Ecological Modelling 213:63–72.
491 Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of
492 species geographic distributions. Ecological Modelling 190:231–259.

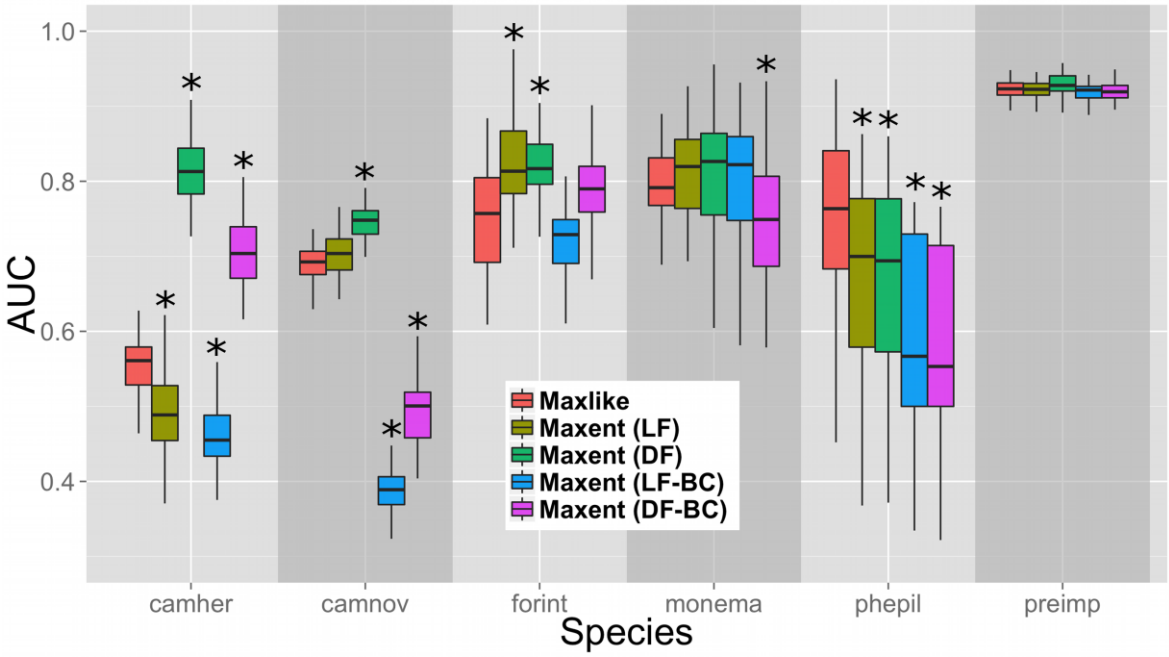
493 Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new
494 extensions and a comprehensive evaluation. Ecography 31:161–175.
495 Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009.
496 Sample selection bias and presence-only distribution models: implications for
497 background and pseudo-absence data. Ecological Applications 19:181–197
498 Pottier, J., A. Dubuis, L. Pellissier, L. Maiorano, L. Rossier, C. F. Randin, P. Vittoz, and A.
499 Guisan. 2012. The accuracy of plant assemblage prediction from species distribution
500 models varies along environmental gradients. Global Ecology and Biogeography 22:52-
501 63.
502 R Development Core Team. 2012. R: a language and environment for statistical computing.
503 R Foundation for Statistical Computing, Vienna, Austria.
504 Royle, J. A., R. B. Chandler, C. Yackulic, and J. D. Nichols. 2012. Likelihood analysis of species
505 occurrence probability from presence-only data for modelling species distributions.
506 Methods in Ecology and Evolution 3:545-554.
507 Sanders, N. J., J. P. Lessard, M. C. Fitzpatrick, and R. R. Dunn. 2007. Temperature, but not
508 productivity or geometry, predicts elevational diversity gradients in ants across spatial
509 grains. Global Ecology and Biogeography 16: 640–649.
510 Svenning, J. C., M. C. Fitzpatrick, S. Normand, C. H. Graham, P. B. Pearman, L. R. Iverson, and
511 F. Skov. 2010. Geography, topography, and history affect realized to potential tree
512 species richness patterns in Europe. Ecography 33:1070-1080.
513 Syfert, M. M., M. J. Smith, and D. A. Coomes. 2013. The effects of sampling bias and model
514 complexity on the predictive performance of MaxEnt species distribution models.
515 PloS one 8:e55158.

516 Thuiller, W., S. Lavorel, M. B. Araujo, M. T. Sykes, and I. C. Prentice. 2005. Climate change
517 threats to plant diversity in Europe. Proceedings of the National Academy of Sciences of
518 the United States of America 102:8245–8250.
519 Trisurat, Y., R. P. Shrestha, and R. Kjelgren. 2011. Plant species vulnerability to climate
520 change in Peninsular Thailand. Applied Geography 31:1106–1114.
521 VanDerWal, J., L. P. Shoo, C. Graham, and S. E. Williams. 2009. Selecting pseudo-absence
522 data for presence-only distribution modeling: How far should you stray from what you
523 know? Ecological Modelling 220:589–594.
524 Ward, D. F. 2007. Modelling the potential geographic distribution of invasive ant species in
525 New Zealand. Biological Invasions 9:723–735.
526 Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: the importance
527 of model complexity and the performance of model selection criteria. Ecological
528 Applications 21:335–342.
529 Warton, D. I., and L. C. Shepherd. 2010. Poisson point process models solve the "pseudo-
530 absence problem" for presence-only data in ecology. The Annals of Applied Statistics
531 4:1383–1402.
532 Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, and A. Guisan. 2008. Effects of
533 sample size on the performance of species distribution models. Diversity and
534 Distributions 14:763–773.
535 Yackulic, C. B., R. Chandler, E. F. Zipkin, J. A. Royle, J. D. Nichols, E. H. Campbell Grant, and S.
536 Veran. *In press*. Presence-only modelling using MAXENT: when can we trust the
537 inferences? Methods in Ecology and Evolution. [doi: 10.1111/2041-210x.12004]

538 Table 1. Modeled ant species, the number of occurrence records that were randomly
539 partitioned into training (75%) and testing (25%) data sets, and comparison of MAXLIKE
540 and MAXENT models implementing linear features without (LF) and with (LF-BC) bias
541 correction using the small sample size correction of Akaike's information criterion (ΔAIC_c,
542 MAXENT – MAXLIKE) and normalized model selection weights (*w*).

Species	Train/Test	ΔAIC_c MAXENT-LF	ΔAIC_c MAXENT-LF-BC	MAXLIKE *w*
Camponotus herculeanus	82/27	731.2	868.1	1
C. novaeboracensis	201/68	1771.7	2064.3	1
Formica integra	32/11	252.0	278.1	1
Monomorium emarginatum	21/7	176.0	194.6	1
Pheidole pilifera	5/2	32.6	32.8	1
Prenolepis imparis	55/26	1610.0	1645.3	1

543 FIGURE 1.



544

545 FIGURE 1. Box plot displaying the 25th and 75th percentiles around the median AUC values

546 of ROC plots for MAXLIKE and MAXENT models implementing linear (LF) or default (DF)

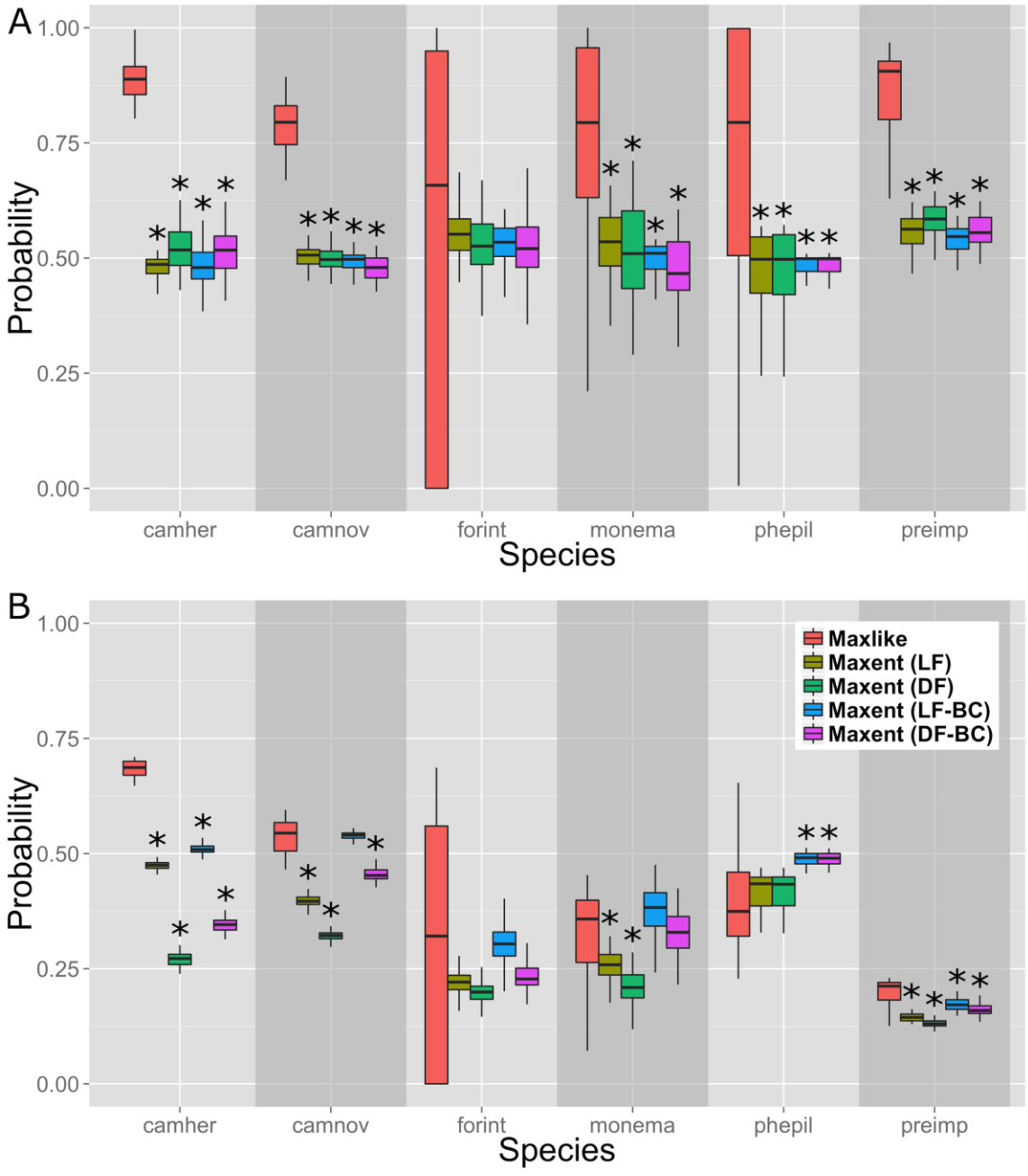
547 features without or with bias correction (BC). *P <0.01 based on Wilcoxon signed-rank

548 tests comparing MAXLIKE to each implementation of MAXENT. camher = *Camponotus*

549 *herculeanus*; camnov = *C. novaeboracensis*; forint = *Formica integra*; monema =

550 *Monomorium emarginatum*; phepil = *Pheidole pilifera*; preimp = *Prenolepis imparis*

551 FIGURE 2.



552

553 FIGURE 2. Box plots displaying the 25th and 75th percentiles around the median predicted

554 probability of presence at (a) test locations and at (b) 10,000 random background points

555 from MAXLIKE and MAXENT models implementing linear (LF) or default (DF) features

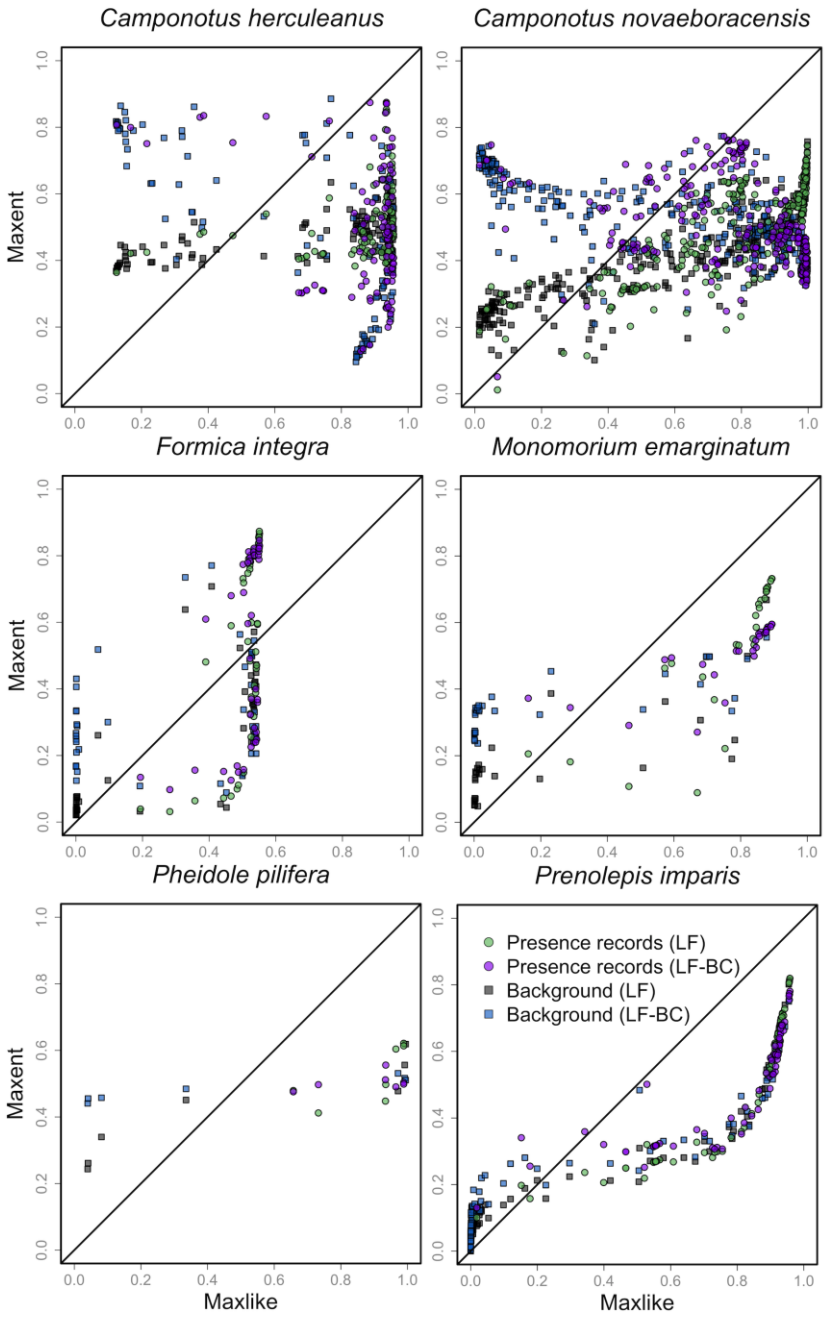
556 without or with bias correction (BC). *P <0.01 based on Wilcoxon signed-rank tests

557 comparing MAXLIKE to each implementation of MAXENT. camher = *Camponotus herculeanus*;

558 camnov = *C. novaeboracensis*; forint = *Formica integra*; monema = *Monomorium*

559 *emarginatum*; phepil = *Pheidole pilifera*; preimp = *Prenolepis imparis*

560 FIGURE 3.



561

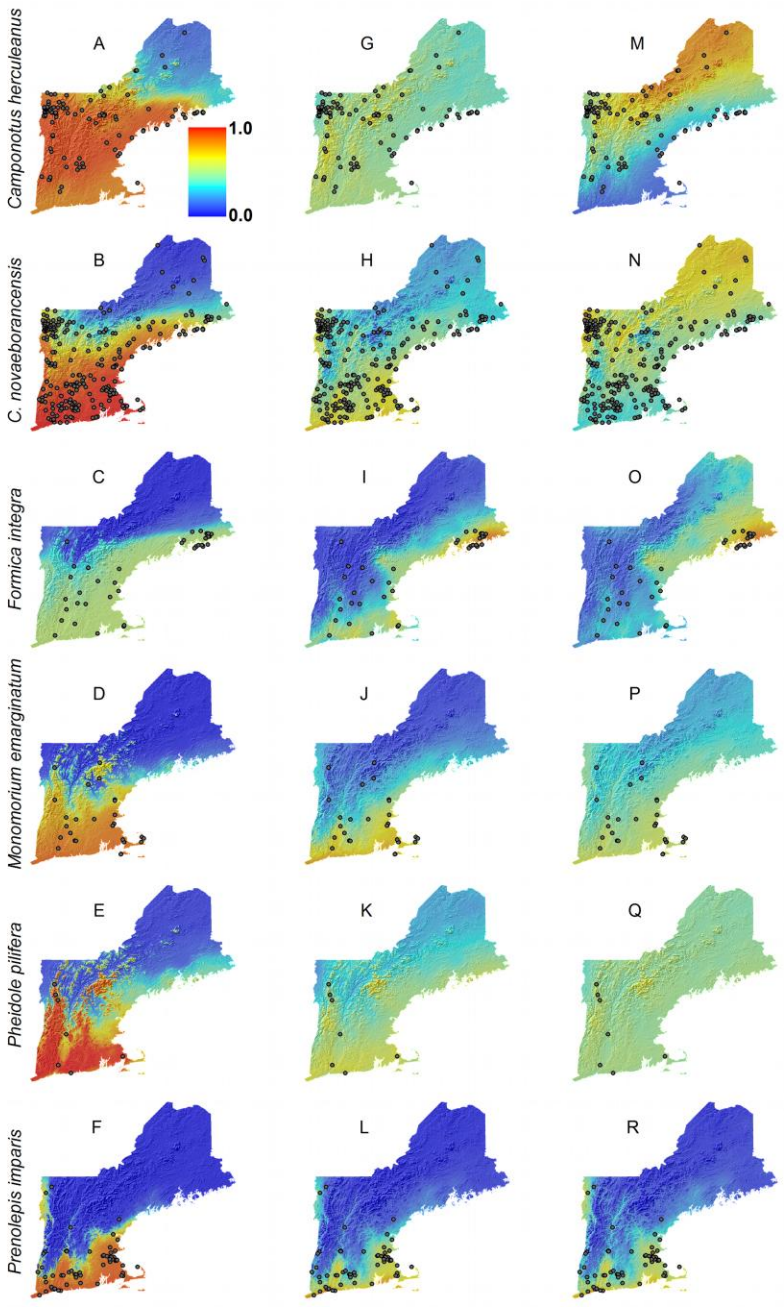
562 FIGURE 3. Probability of occurrence at (circles) presence records and at an equivalent

563 number of randomly selected (squares) background locations from MAXLIKE versus MAXENT

564 implementing linear features (LF) without or with bias correction (BC). The plotted

probabilities at each point indicate the mean of the predictions from the 50 models for each species.

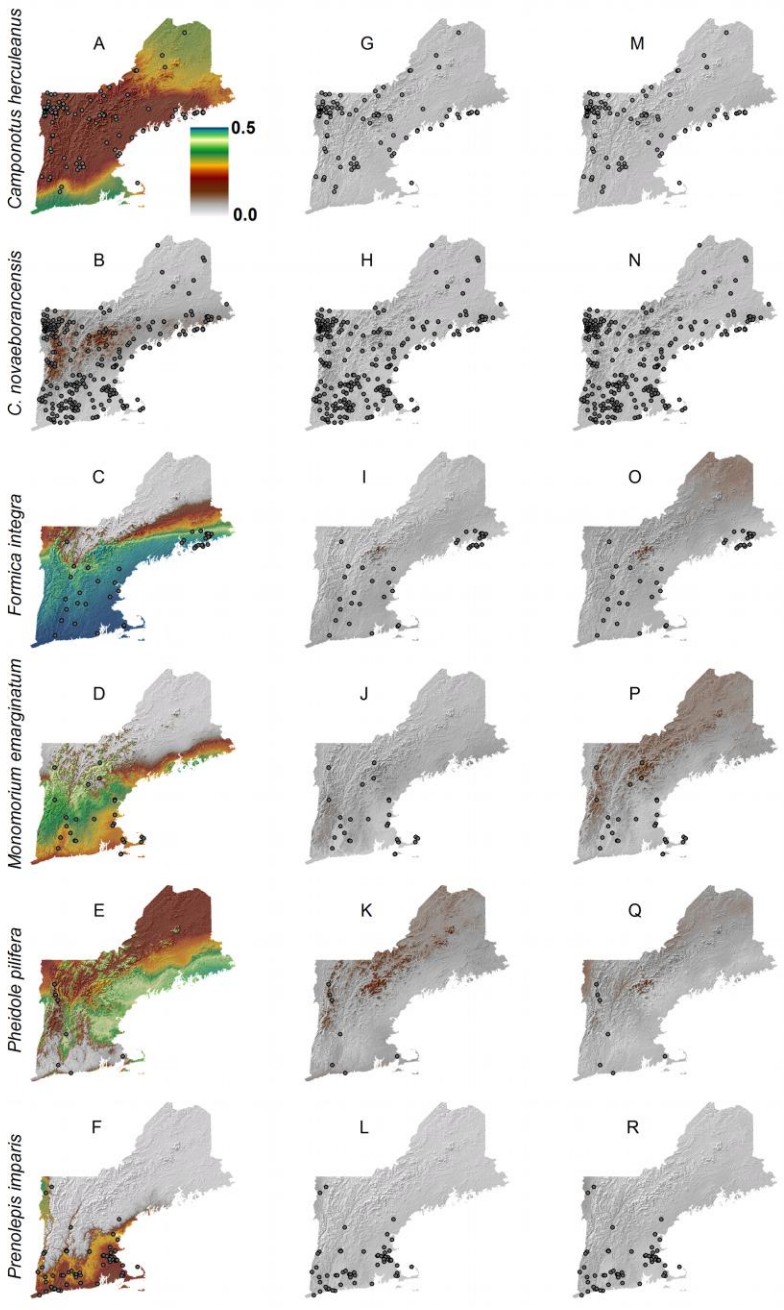
567 FIGURE 4.



568

FIGURE 4. Mean predicted probability of occurrence from MAXLIKE and MAXENT models implementing linear features based on 50 random training/test (75/25%) partitions of occurrence records. (a-f) show predicted probabilities of occurrence from MAXLIKE; (g-l) and (m-q) show logistic output from MAXENT without or with bias correction respectively. Points indicate ant occurrences used to fit models.

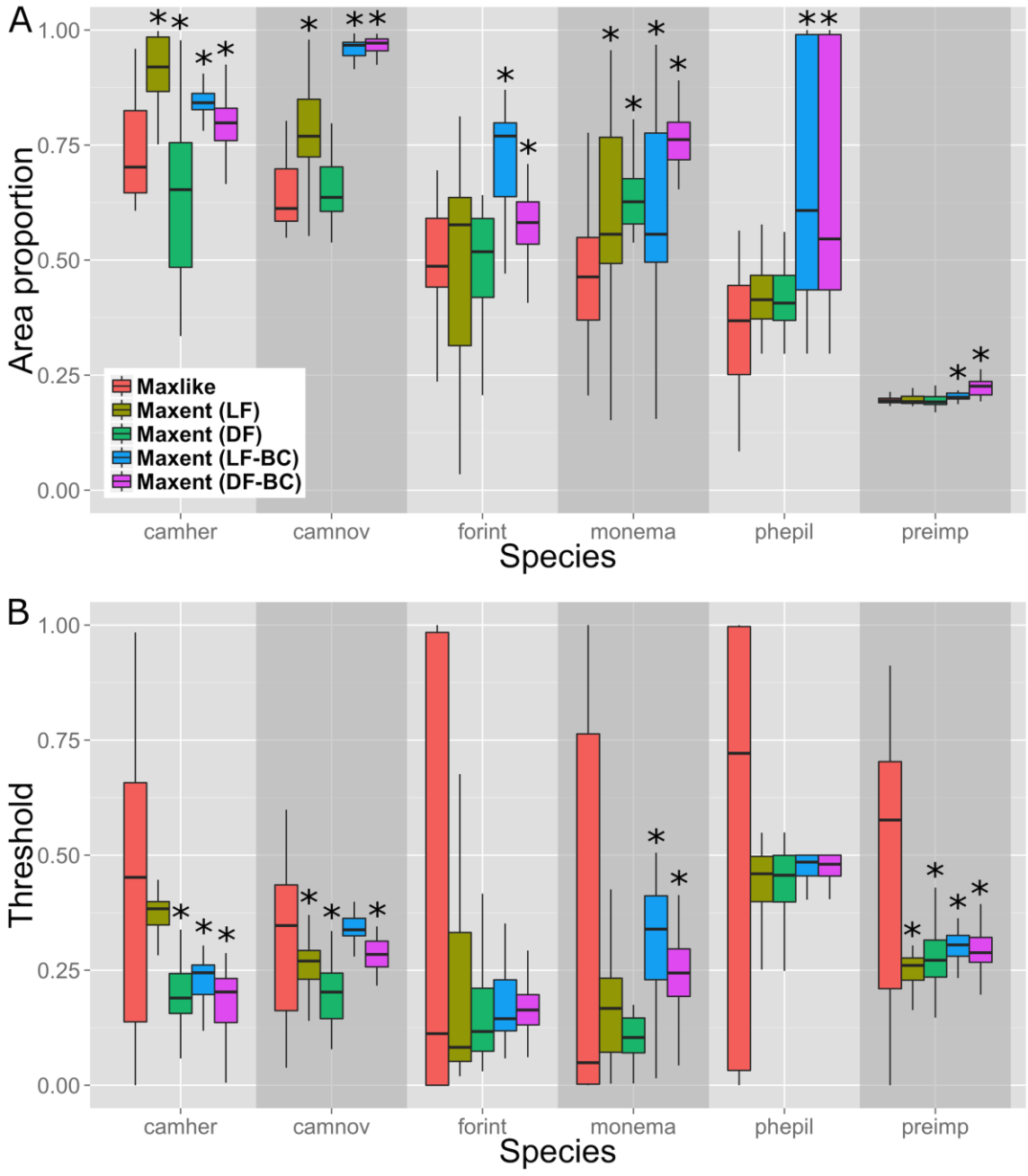
574 FIGURE 5.



575

576 FIGURE 5. Standard deviations of predicted probabilities of occurrence from MAXLIKE and
577 MAXENT models implementing linear features based on 50 random training/test (75/25%)
578 partitions of occurrence records. (a-f) show standard deviations from MAXLIKE; (g-l) and
579 (m-q) show standard deviations from MAXENT without or with bias correction respectively.
580 Points indicate ant occurrences used to fit models.

581 FIGURE 6.



582

583 FIGURE 6. Box plots displaying the 25th and 75th percentiles around the median (a)

584 proportion of the study area predicted as present using (b) the threshold required to

585 correctly predict as present 95% of test occurrences from MAXLIKE and MAXENT models

586 implementing linear (LF) or default (DF) features without or with bias correction (BC). *P

587 <0.01 based on Wilcoxon signed-rank tests comparing MAXLIKE to each implementation of

588 MAXENT. camher = *Camponotus herculeanus*; camnov = *C. novaeboracensis*; forint = *Formica*

589 *integra*; monema = *Monomorium emarginatum*; phepil = *Pheidole pilifera*; preimp =

590 *Prenolepis imparis*