MaxEnt versus MaxLike: empirical comparisons with ant species distributions

Citation

Published Version
doi:10.1890/ES13-00066.1

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:11327042

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
MAXENT VS. MAXLIKE: Empirical Comparisons with Ant Species Distributions

Matthew C. Fitzpatrick\textsuperscript{1,†}, Nicholas J. Gotelli\textsuperscript{2}, Aaron M. Ellison\textsuperscript{3}

\textsuperscript{1}Appalachian Lab, University of Maryland Center for Environmental Science, Frostburg, Maryland, 21502 USA
\textsuperscript{2}Department of Biology, University of Vermont, Burlington, Vermont, 05405 USA
\textsuperscript{3}Harvard Forest, Harvard University, Petersham, Massachusetts, 01366 USA

\textsuperscript{†}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.
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, Papeš 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
records and standardized ecological sample surveys and are part of a larger dataset on the

Test species

Of the 132 species in the ants of New England dataset, we considered as case studies six
species of ants common in the six New England (northeastern U.S.) states (Maine, New
Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island), but which differ in their
geographic distribution, range size, and number of occurrence records (Table 1). These
case studies included a circumboreal species for which New England is in the southern part
of its range (*Camponotus herculeanus* (L.)), a southern North American species for which
New England is in the northern part of its range (*Prenolepis imparis* (Say)), three
widespread, commonly collected North American species for which New England is in the
center of its range (*Camponotus novaeboracensis* (Fitch), *Formica integra* (Nylander),
*Monomorium emarginatum* (DuBois)), and a North American habitat specialist, the
sandplain-inhabiting *Pheidole pilifera* (Roger). See Ellison et al. (2012) for additional details
on the natural history of these six ant species and the broader dataset.

Environmental data

To avoid over-fitting models with the small number of occurrence records available for
some of the study species (Table 1), we used only a small set of 20 potential environmental
covariates: elevation (meters above sea level) and 19 bioclimatic variables from the
WorldClim database ([http://www.worldclim.org](http://www.worldclim.org), Hijmans et al. 2005) that measure
minima, maxima, and seasonality in temperature (°C) and precipitation (mm) at a spatial
resolution of 30 arc-seconds ($\approx 1 \times 1$ 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 AICc and normalized Akaike model selection weights revealed that for all six
ant species, MaxLike models were better supported by the data than MaxEnt models implementing linear features with or without sampling bias correction (Table 1). However, model evaluation by AUC was inconsistent, with MaxLike scoring lower, equal, or greater AUC values than MaxEnt, depending on the ant species considered and whether MaxEnt models were fit using default settings or restricted to linear features, and whether sampling bias was accounted for or not (Fig. 1). In general, MaxEnt models that accounted for sampling bias scored lower or equal AUC values than MaxEnt models without bias correction.

By default, the MaxEnt algorithm assumes a baseline species prevalence of 0.5 (Phillips and Dudík 2008), and therefore assigned a probability of occurrence close to 0.5 to most occurrence locations. In contrast, MaxLike assigned substantially higher probabilities to locations with recorded presences for five of six species than did any of the implementations of MaxEnt (Fig. 2a). For randomly chosen background locations (Fig. 2b), MaxLike also tended to generate higher average probabilities of occurrence than MaxEnt, although accounting for sampling bias increased average probabilities at random background locations. Randomly-chosen background points also had nearly constant probabilities of occurrence with MaxEnt, although the value of the mean probability differed among species; MaxEnt models implementing default features tended to generate lower probabilities than MaxEnt models implementing only linear features. In contrast, MaxLike usually generated a larger range of different probabilities for both occurrence and background locations.
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
dissimilar quantities. In other words, MAXLIKE estimates probability of occurrence, while MAXENT estimates a relative suitability index that, for five of six species in our study, neither represents probability of occurrence nor is correlated with it.

Our goodness-of-fit statistics (Table 1) and other evaluation metrics (Fig. 6) generally favored the MAXLIKE formulation, although AUC (Fig. 1) was equivocal. However, given the documented issues with AUC and pseudo-absence data (Lobo et al. 2008, Peterson et al. 2008, Jiménez-Valverde 2012), the interpretation of AUC is problematic. In essence, MAXLIKE would have a lower AUC than MAXENT simply because it tends to predict higher probabilities of occurrence across the spatial extent of the occurrence data than MAXENT and therefore will assign higher probabilities to a greater number of pseudo-absence locations. However, an unknown percentage of these pseudo-absences are actually instances of presence and therefore there is little justification for penalizing presumed over-prediction at the same cost as under-predicting known occurrences.

We also note that, at least for ants of New England, the predicted species distributions from MAXLIKE are more sensible and in line with our expectations from over 15 years of field experience in this system (e.g., Gotelli and Ellison 2002, Ellison et al. 2012). For example, the likely distribution of the widespread carpenter ant, Camponotus novaeboracensis, is captured well by MAXLIKE (Fig. 4b), but not nearly as well by MAXENT. In particular, MAXENT without accounting for sampling bias down-weights the probability of occurrence of C. novaeboracensis in north central New England where it is actually widespread (compare 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,
overpredicted (albeit with little confidence – Fig. 5e) its probability of occurrence in most
locations in southern New England. However, MAXENT underpredicted its occurrence in its
ture range and overpredicted its occurrence further north (Fig. 4k), especially when
models accounted for sampling bias (Fig. 4q), and with little uncertainty (Fig. 5k, q).

Both MAXENT and MAXLIKE assume random sampling, which is rarely possible with species
occurrence records. For example, counties throughout central and eastern Massachusetts
are more thoroughly sampled for ants than some other areas of New England because of
the large number of myrmecologists historically associated with Harvard University
(Ellison et al. 2012). Approaches for accounting for such sampling bias, including strategies
for the selection of background points, are relatively well developed for MAXENT (e.g.,
Phillips et al. 2009, VanDerWal et al. 2009), but remain unexplored for MAXLIKE.

For MAXENT, a method for accounting for sampling bias involves using all occurrence
records for a taxon of interest within a study to estimate relative survey effort and to select
background data with same underlying bias present in the occurrence data. This method,
known as “target-group background,” (Phillips et al. 2009) has been shown to generally
improve performance of MAXENT models when averaged across all species (e.g., Mateo et al.
2010, Syfert et al. 2013), but not necessarily for all species or regions (Phillips et al. 2009).
We found that accounting for sampling bias generally did not improve MAXENT’s
performance, and in some cases resulted in less robust models (Figures 2, 4). The
immediate reasons for the reduction in model performance are not clear, but Phillips et al
(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


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

<table>
<thead>
<tr>
<th>Species</th>
<th>Train/Test</th>
<th>ΔAICc MaxEnt-LF</th>
<th>ΔAICc MaxEnt-LF-BC</th>
<th>MaxLike w</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Camponotus herculeanus</em></td>
<td>82/27</td>
<td>731.2</td>
<td>868.1</td>
<td>1</td>
</tr>
<tr>
<td><em>C. novaeboracensis</em></td>
<td>201/68</td>
<td>1771.7</td>
<td>2064.3</td>
<td>1</td>
</tr>
<tr>
<td><em>Formica integra</em></td>
<td>32/11</td>
<td>252.0</td>
<td>278.1</td>
<td>1</td>
</tr>
<tr>
<td><em>Monomorium emarginatum</em></td>
<td>21/7</td>
<td>176.0</td>
<td>194.6</td>
<td>1</td>
</tr>
<tr>
<td><em>Pheidole pilifera</em></td>
<td>5/2</td>
<td>32.6</td>
<td>32.8</td>
<td>1</td>
</tr>
<tr>
<td><em>Prenolepis imparis</em></td>
<td>55/26</td>
<td>1610.0</td>
<td>1645.3</td>
<td>1</td>
</tr>
</tbody>
</table>
FIGURE 1. Box plot displaying the 25\textsuperscript{th} and 75\textsuperscript{th} percentiles around the median AUC values of ROC plots for MAXLIKE and MAXENT models implementing linear (LF) or default (DF) features without or with bias correction (BC). *$P < 0.01$ based on Wilcoxon signed-rank tests comparing MAXLIKE to each implementation of MAXENT. camher = Camponotus herculeanus; camnov = C. novaeboracensis; forint = Formica integra; monema = Monomorium emarginatum; phepil = Pheidole pilifera; preimp = Prenolepis imparis
FIGURE 2. Box plots displaying the 25th and 75th percentiles around the median predicted probability of presence at (a) test locations and at (b) 10,000 random background points from MAXLIKE and MAXENT models implementing linear (LF) or default (DF) features without or with bias correction (BC). *P < 0.01 based on Wilcoxon signed-rank tests.
comparing MAXLIKE to each implementation of MAXENT. camher = *Camponotus herculeanus*;
camnov = *C. novaeboracensis*; forint = *Formica integra*; monema = *Monomorium emarginatum*; phepil = *Pheidole pilifera*; preimp = *Prenolepis imparis*
FIGURE 3. Probability of occurrence at (circles) presence records and at an equivalent number of randomly selected (squares) background locations from \texttt{MaxLike} versus \texttt{MaxEnt} 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.
FIGURE 4.
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.
FIGURE 5.
FIGURE 5. Standard deviations of predicted probabilities of occurrence from \textsc{MaxLike} and \textsc{MaxEnt} models implementing linear features based on 50 random training/test (75/25\%) partitions of occurrence records. (a-f) show standard deviations from \textsc{MaxLike}; (g-l) and (m-q) show standard deviations from \textsc{MaxEnt} without or with bias correction respectively. Points indicate ant occurrences used to fit models.
FIGURE 6. Box plots displaying the 25\textsuperscript{th} and 75\textsuperscript{th} percentiles around the median (a) proportion of the study area predicted as present using (b) the threshold required to correctly predict as present 95\% of test occurrences from MAXLIKE and MAXENT models implementing linear (LF) or default (DF) features without or with bias correction (BC). *P
<0.01 based on Wilcoxon signed-rank tests comparing MAXLIKE to each implementation of MAXENT.
camher = Camponotus herculeanus; camnov = C. novaeboracensis; forint = Formica integra; monema = Monomorium emarginatum; phepil = Pheidole pilifera; preimp = Prenolepis imparis