



Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent

A. Townsend Peterson, Monica Papeş and Muir Eaton

A. T. Peterson (*town@ku.edu*), M. Papeş and M. Eaton, Natural History Museum and Biodiversity Research Center, The Univ. of Kansas, Lawrence, KS 66045, USA.

We compared predictive success in two common algorithms for modeling species' ecological niches, GARP and Maxent, in a situation that challenged the algorithms to be general – that is, to be able to predict the species' distributions in broad unsampled regions, here termed transferability. The results were strikingly different between the two algorithms – Maxent models reconstructed the overall distributions of the species at low thresholds, but higher predictive levels of Maxent predictions reflected overfitting to the input data; GARP models, on the other hand, succeeded in anticipating most of the species' distributional potential, at the cost of increased (apparent, at least) **commission error**. Receiver operating characteristic (ROC) tests were weak in discerning models able to predict into broad unsampled areas from those that were not. Such transferability is clearly a novel challenge for modeling algorithms, and requires different qualities than does predicting within densely sampled landscapes – in this case, Maxent was transferable only at very low thresholds, and biases and gaps in input data may frequently affect results based on higher Maxent thresholds, requiring careful interpretation of model results.

Ecological niche modeling, or as it is sometimes termed, species' distribution modeling, has seen an impressive growth in attention in recent years. A Web of Science search revealed 42 citations with one of these two terms (and minor variants) in the title or abstract, all in the last 7 yr. Several recent reviews and high-impact papers (Guisan and Zimmermann 2000, Soberón and Peterson 2004, Graham et al. 2004, Thomas et al. 2004, Araújo et al. 2005a, Thuiller et al. 2005, Elith et al. 2006) have served to place this technique among emerging new approaches relevant to ecology, biogeography, and conservation biology.

In spite of the burgeoning attention to and use of the technique, many details remain to be clarified and tested. Recent papers have assessed data needs in terms of sample size and environmental data sets (Peterson and Cohoon 1999, Stockwell and Peterson 2002b), effects of scale and resolution (Thuiller et al. 2004, Araújo et al. 2005b), choice of thresholds (Liu et al. 2005), variability of model projections under changed-climate conditions (Pearson et al. 2006), and effects of sampling bias on model quality (Stockwell and Peterson

2002a). In particular, a suite of recent papers has compared the performance of the many algorithms and approaches to project distributions within the same regions used to train the models (Manel et al. 1999a, b, Elith and Burgman 2002, Elith et al. 2006, Guisan et al. 2007a, b).

Most importantly, a recent working group at the National Center for Ecological Analysis and Synthesis (NCEAS) focused on developing a detailed comparison of modeling methodologies as applied to predicting distributions of species (Graham et al. 2004). The comparisons developed included a broad survey of model performance across very diverse data sets and realms (Elith et al. 2006), as well as a number of detailed comparisons of effects of specific manipulations on model performance (Guisan et al. 2007a, b). This body of work has attracted considerable attention, given its broad scope, in terms of geography, taxonomy, and modeling approaches.

The NCEAS working group analyses, however, focused on only a subset of the challenge that this new field presents. Certainly, one important part of the

challenge is that of predicting variation within a well-sampled region, and estimating the fine details of species' distributions across landscapes – this challenge is what was addressed in the working group's analyses. A very different challenge, however, is that of projecting accurately into new areas, in which sampling is sparse or nonexistent – this challenge of transferability (Thomas and Bovee 1993, Glozier et al. 1997, Schröder and Richter 1999, Kleyer 2002, Randin et al. 2006), however, is critical in several uses of these models, such as predicting effects of global climate change on species' distributional potential, evaluating invasive potential of alien species, and discovery of new populations and species (Soberón and Peterson 2004, Peterson 2006). This distinction has been pointed out in a number of recent publications (Bio et al. 2002, Guay et al. 2003, Randin et al. 2006) – the question was put most clearly in a recent commentary that stated “Does model precision come at the expense of generality, that is, the ability to predict species distributions in different regions or times?” (Araujo and Rahbek 2006).

The purpose of this contribution is to provide a comparison among modeling algorithms regarding the challenge of transferability, or extrapolation in geographic space, interpreted carefully in terms of error component distributions. We do not take on the more complex challenge of true extrapolation (i.e. to new conditions in ecological space), which has been treated elsewhere (Pearson et al. 2006). We chose two popular modeling algorithms, one that performed particularly well in the NCEAS comparisons (Maxent) and one that performed poorly (GARP) (Elith et al. 2006). We presented these two algorithms with a simple task – that of predicting half of the geographic distribution of a species based on dense sampling of the other half of the species' distribution (Peterson and Shaw 2003). As will become clear in this paper, transferability turns out to be quite a different challenge from interpolation.

Methods

Input occurrence data and data preparation

We chose three bird species for analysis, based on data drawn from the North American Breeding Bird Survey (BBS) (Sauer et al. 2001): *Caprimulgus vociferus*, *Coccyzus americanus*, and *Zenaidura macroura*. Species were chosen opportunistically based solely on the criteria of having fairly broad geographic distributions, sample sizes sufficient for analysis (i.e. > 100 presence points for training and testing), and to be good dispersers so that their distributions would be limited largely by current ecological factors, rather than by historical factors (Austin 1985, Soberón and Peterson 2005).

To assure that occurrences used in analyses represented reasonably stable populations, we used only BBS survey routes for each species on which the species had been detected in ≥ 8 yr in 1991–2000. Absences (used in model evaluation only) were defined as those routes on which the species had never been detected, regardless of where the route lies spatially; all other routes (i.e. those with intermediate frequencies of detection across the 10 yr) were omitted from analysis. Overall, we assembled 127, 453, and 1202 presence points, and 2400, 1385, and 6030 absence points for the three species, respectively.

To challenge the niche modeling algorithms to project accurately across unsampled areas of geographic space, we separated available occurrence points for each species into quadrants based on whether their coordinates fell above or below the median longitude and median latitude of occurrence localities for that species. Henceforth, we refer to the NW and SE pair of quadrants as “on-diagonal”, and to the NE and SW pair of quadrants as “off-diagonal” (Fig. 1): sample sizes were 61 in on-diagonal quadrants/66 in off-diagonal quadrants, 210/243, and 620/582 presence points for *Caprimulgus vociferus*, *Coccyzus americanus*, and *Zenaidura macroura*, respectively. Our general approach was that of training models based on one pair of quadrants, and testing them based on the other pair (Peterson and Shaw 2003); this manipulation challenges modeling algorithms to predict into broadly unsampled regions, rather than simply interpolating or filling gaps in a densely sampled landscape. It is important to note that all model development (including, e.g. best subsets filtering in GARP) was carried out on one pair of quadrants, and testing and model evaluation only in the other pair of quadrants. For the purposes of comparison, we also present preliminary models based on randomly splitting occurrence points into equal subsets for one species.

Environmental data sets

We tested predictivity of models based on the 19 “bioclimatic” variables in the 10' WorldClim data set (described in detail in Hijmans et al. 2005), supplemented with information on topographic features summarized in 4 additional raster data layers (elevation, slope, aspect in radial degrees, compound topographic index) from the 1 km resolution Hydro-1K digital elevation model data set (Anon. 2001) that can modify how individual animals experience climates (e.g. mean temperatures are higher on south-facing slopes in the Northern Hemisphere). Because the questions of generality explored in this study are affected by overfitting, which is likely to be more serious in highly dimensional environmental spaces, we used principal components

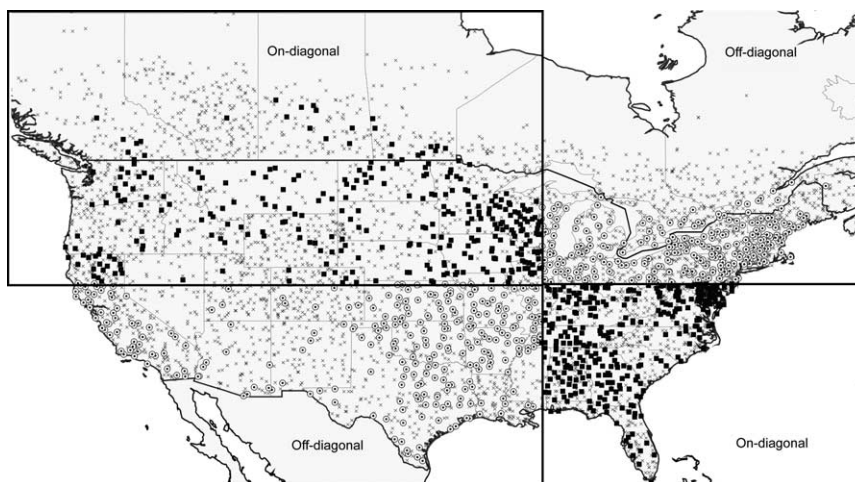


Fig. 1. Example of geographic subsetting for *Zenaida macroura*, in which the median latitude and median longitude are used to divide the available presence data into four quadrants. The points in the on-diagonal quadrants (black squares) were used to train models which were tested using points from the off-diagonal quadrants (dotted circles), and vice versa; absence data are shown as small Xs.

analysis (PCA) to create new axes that summarized the variation in fewer (independent) dimensions; we retained the first 11 components, which together explained >99% of the overall variation in environmental parameters. All data sets were resampled to 10' resolution for analysis to reflect the spatial resolution of the occurrence data, across a region covering almost all of North America (24.3°–76.5°N, 52.0°–169.5°W).

Model development

Several approaches have been used to approximate species' ecological niches (Elith et al. 2006). In this study, we compare a method that performed relatively poorly in that previous study, the Genetic Algorithm for Rule-set Prediction (GARP) (Stockwell and Noble 1992) with one of the top performers, a maximum entropy approach to modeling species' distributions (Maxent) (Phillips et al. 2004, 2006). Both methods use known occurrences and pseudo-absence data resampled from the set of pixels where the species in question is not known to occur.

GARP models were developed using a desktop version that permits flexibility in model development (Pereira 2002). In GARP, occurrence points from the pair of quadrants on which models are to be based are divided evenly into training and "extrinsic test data" sets; the former is again divided evenly into true "training data" (for model rule development) and "intrinsic test data" sets (for model rule evaluation and refinement). GARP models are based on presence-only data, with absence information included via automated, random sampling of 1250 pseudo-absence

points from the set of pixels at which the species has not been detected (note that these data are not equivalent to the true absences derived from the survey data described above). GARP works in an iterative process of rule selection, testing, and incorporation or rejection: first, a method is chosen from a set of possibilities (logistic regression, bioclimatic rules, range rules, negated range rules), and then is applied to the training data and a rule developed; rules may evolve by a number of means (e.g. truncation, point changes, crossing-over among rules) to maximize predictivity. Predictive accuracy (for intrinsic use in model refinement) is then evaluated based on 1250 points resampled from the intrinsic test data and the 1250 pseudo-absence points. The change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the model, and the algorithm runs either 1000 iterations of these processes or until convergence.

The "best subsets" procedure (Anderson et al. 2003) was originally developed to sift through the model-to-model variation generated by the random-walk nature of the GARP algorithm. Hence, we generated 1000 models for each species, and selected the 100 "best" models based on omission (leaving out true potential distributional areas) and commission (including areas not potentially habitable) error statistics. Specifically, we used a soft omission threshold, in which the 20% of the models with lowest omission error were retained for further analysis. We then chose models presenting intermediate levels of commission (i.e. the central 50% of the commission index distribution). It is important to note that we evaluated the omission rates based on extrinsic testing data (note that

these are not the data on which the transferability tests were based), rather than the intrinsic data, as was done in a previous comparison of the two methods (Phillips et al. 2006); this difference presents a more rigorous challenge to GARP models (i.e. to be able to predict novel phenomena), and as such produces models that are more general. The 10 best models were summed in ArcView 3.2 to produce a single final grid.

Maxent models were developed using recently developed software (Phillips 2006) that has been described and tested in detail in recent publications (Phillips et al. 2004, 2006). Maxent focuses on fitting a probability distribution for occurrence of the species in question to the set of pixels across the study region, based on the idea that the best explanation to unknown phenomena will maximize the entropy of the probability distribution, subject to the appropriate constraints. In the case of modeling ecological niches of species, these constraints consist of the values of those pixels at which the species has been detected (Phillips et al. 2004, 2006). In general, we used default parameters for Maxent models, except for experimentation with varying the regularization multiplier value over several orders of magnitude (10^{-5} – 10^2) to see the effects of changes on model quality.

Maxent models produce predictions in the form of real numbers between 0 and 100, representing cumulative probabilities of occurrence (Phillips et al. 2004, 2006). Given that floating-point grids are notoriously difficult to manipulate in GIS programs, we imported them into ArcView as floating-point grids, multiplied by 100, and then converted to integer grids for further analysis, which is equivalent to using Maxent predictions out to hundredths only. Finally, we experimented with training Maxent models with the training region limited to the the spatial subsets of the study region from which training input data were derived, effectively limiting the region from which pseudo-absences could be drawn to the training quadrants.

Evaluation statistics

We focused our model evaluation on the area under the curve (AUC) in receiver operating characteristic analyses (Fielding and Bell 1997), which is the measure generally accepted as best for evaluating model performance, employed in recent comprehensive evaluations (Elith et al. 2006). AUC provides a threshold-independent measure of model performance as compared with that of null expectations, and as such should provide an overall picture of the predictive nature of models (DeLong et al. 1988); AUCs were calculated using a web-based program (Eng 2005).

Results and discussion

As an initial exploration of the data sets, for the purpose of illustration, we mimicked previous model comparison studies by developing models based on simple random samples of half of available occurrences of *Zenaida macroura* (i.e. no spatial stratification). Both algorithms produced maps that coincided well with the known distribution of the species (Fig. 2), although the GARP prediction tended to be overly extensive, and the Maxent model tended to be somewhat under-predicted, particularly in the southwestern United States. AUC scores for these two preliminary models paralleled the results of the previous surveys (Elith et al. 2006) – 0.733 for Maxent, compared with 0.608 for GARP models, suggesting that GARP models are less predictive than Maxent models. As such, the results of challenging models with predicting across densely sampled landscapes is seen to parallel those of previous comparisons (Elith et al. 2006).

Challenging the two algorithms with predicting into unsampled regions via spatially stratified subsampling, however, changed the picture considerably (Fig. 2). GARP models continued to reconstruct much of the species' known distributions. Maxent models, on the other hand, produced an odd pattern coincident with the input data set at higher probability values – reconstructing the on-diagonal quadrants or the off-diagonal quadrants, depending on which were used to train the models (Fig. 2 and 3). AUC values based on independent testing data from the unsampled quadrants for the two techniques overlapped broadly (Table 1), showing no significant differences between the two algorithms (paired t-tests, all $p > 0.05$).

The NCEAS model comparisons ranked GARP among the poorest performing of algorithms, and Maxent among the best (Elith et al. 2006). The NCEAS project, however, focused on the challenge of predicting across densely sampled landscapes (referred to by some as “distribution modeling”). The random-sampling exploration that we conducted at the outset of this study is akin to what was done in the NCEAS project, in that the same landscape was the arena for both training and testing models, and the results were parallel as well in terms of relative ranking of the two methods. In this study, however, we challenged GARP and Maxent models to be general, that is, to be transferable and predict into unsampled regions not represented in the input occurrence data set. This quality of transferability orients models toward capturing the essence of the ecological niche of the species (Soberón and Peterson 2005), rather than just reconstructing the particulars of species' distributions in one region.

Although the quantitative descriptors of model quality showed no significant differences between

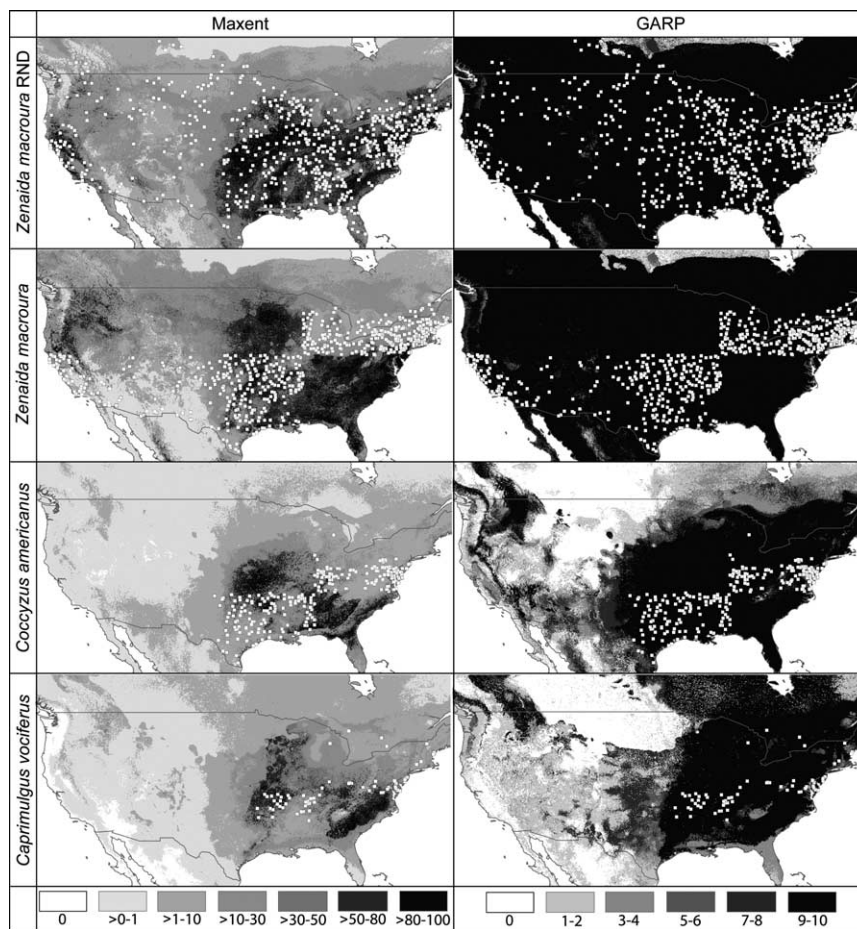


Fig. 2. Example predictions in the test analyses developed in this paper: top: *Zenaida macroura* predictions based on pure (non-stratified) random samples (“RND”) from the species’ known occurrences, showing good overall general performance by both algorithms. Succeeding rows show predictions based on spatially stratified subsamples of available occurrence data for each of the three species; for reasons of space, only predictions based on on-diagonal training data are shown. Independent test occurrence data only are shown, as white squares. Shading indicates levels of model predictions (scale at bottom), assorted into discrete bins for clarity of discrimination.

GARP and Maxent (Maxent models had higher AUC scores in 5 of 6 comparisons; Table 1), the relative inability of the latter to predict into broad unsampled regions – at least at high probability levels – is evident in Fig. 2 and 3. We explored various alternative settings, based in part on the kind counsel of one of the authors of one of the papers documenting the method (R. P. Anderson pers. comm.). First, we explored the effects of changing the regularization multiplier parameter, which affects the degree of generality in the resulting models (Phillips et al. 2006), for *Zenaida macroura*, based on PCA environmental data and the on-diagonal occurrence data set. Across several orders of magnitude of this parameter (10^{-5} – 10^2), we found no improvement – large sectors of the species’ range were still excluded from the higher

levels of model predictions (Fig. 4). At the highest value, however, the model indeed generalized to predict with high probability distribution values most of the southeastern United States, but failed to do the same for the remainder of the range of the species. It appears that the occurrence data were denser in the southeast, which probably led to this odd result. However, in this high regularization model, if the lowest Maxent values were considered, the predictions became too general, covering the whole continent.

We also experimented with training Maxent models based on smaller subsets of the overall landscape of North America. For example, we trained models based only on the quadrants used to build models, and then projected them to the entire region, and we trained models based only on the pixels from which we had

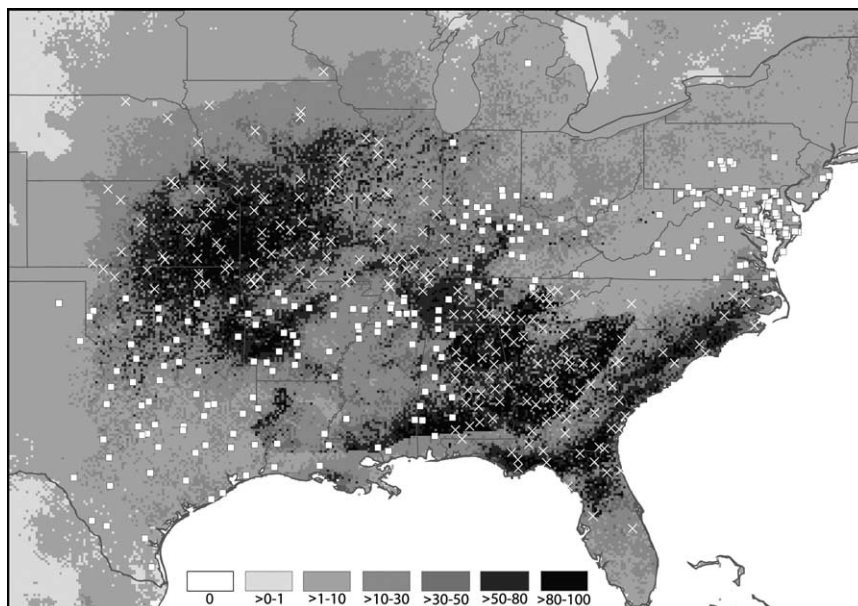


Fig. 3. Detailed view of a typical model output from Maxent, the *Coccyzus americanus* prediction based on on-diagonal training data (white Xs), showing the close match between the probability ramp and the input (training) data, and the discord with independent testing data (white squares). Shading indicates levels of model predictions (scale at bottom), assorted into discrete bins for clarity of discrimination.

presence or absence records, and again projected them to the entire region. In both cases, the general picture of close overfitting to the training region was still observed, suggesting that the phenomena observed in this study will not see any easy fixes.

Maxent models failed to make general predictions unless very low probability value thresholds were considered, in which case they resembled GARP model predictions closely. At higher predictive thresholds, Maxent models reconstructed the general geographic characteristics of the input occurrence data, and did

not generalize much at all, as can be seen in Fig. 2. This behavior has indeed been noted previously – in one of the original descriptions of the algorithm (Phillips et al. 2006), the inability of Maxent (and success of GARP) in identifying other montane areas of potential distribution of *Microrhizomys minutus* was discussed at length. To quote the authors of Maxent (Phillips et al. 2006), “Future research should determine to what degree differences in Maxent’s prediction strength correspond to the relative environmental suitability of the various regions, rather than the

Table 1. Summary of AUC scores for tests of model predictivity: 2 modeling algorithms (GARP and Maxent) \times 3 species (M = *Zenaida macroura*, W = *Caprimulgus vociferus*, Y = *Coccyzus americanus*) \times 2 subsets (models trained with on-diagonal versus off-diagonal quadrants).

Species	Modeling algorithm	Training quadrants	Independent test data
M	GARP	off	0.671
M	Maxent	off	0.683
M	GARP	on	0.520
M	Maxent	on	0.640
W	GARP	off	0.820
W	Maxent	off	0.837
W	GARP	on	0.704
W	Maxent	on	0.808
Y	GARP	off	0.965
Y	Maxent	off	0.950
Y	GARP	on	0.703
Y	Maxent	on	0.845

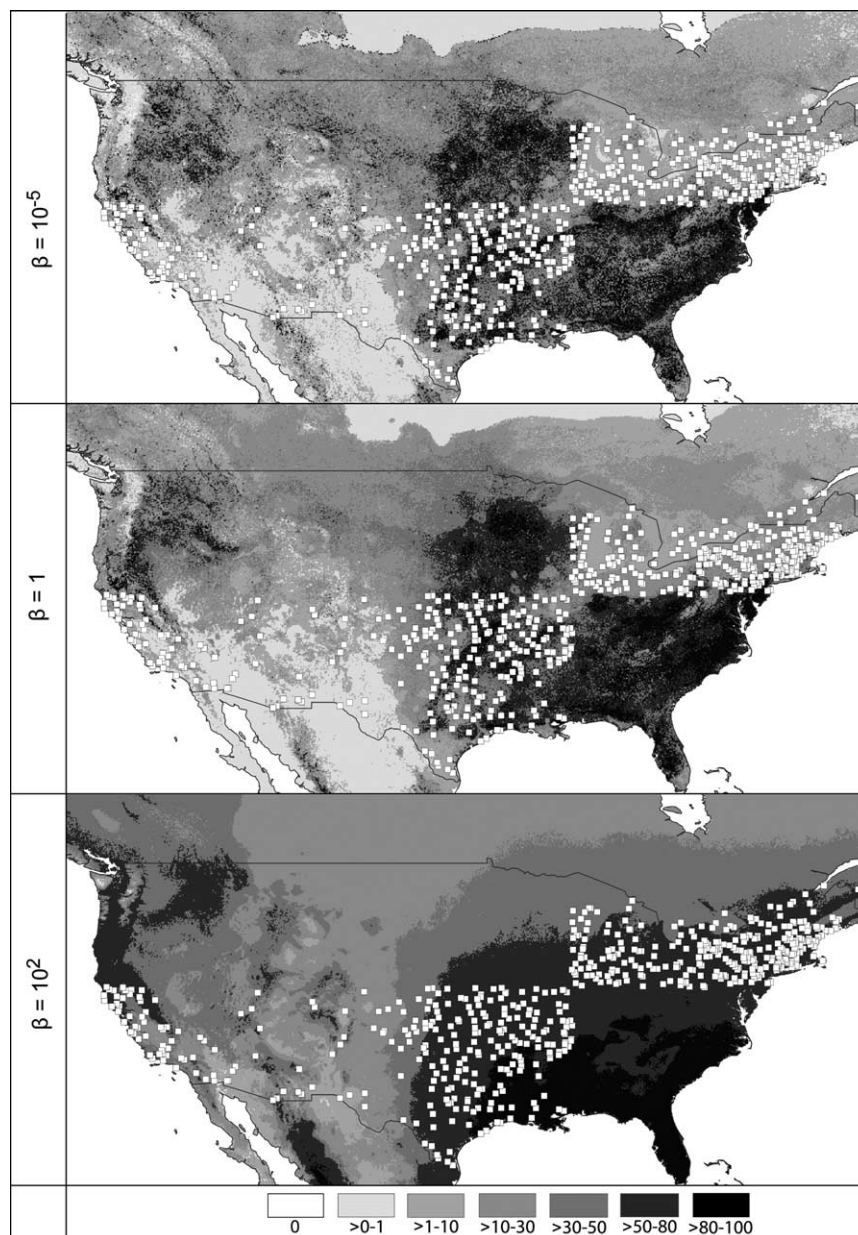


Fig. 4. Exploration of effects of adjusting regularization parameter in Maxent models of *Zenaida macroura*, based on on-diagonal input data (not shown) and principal components-reduced environmental data. Independent testing data are overlain as white squares. Shading indicates levels of model predictions (scale at bottom), assorted into discrete bins for clarity of discrimination.

possibility that they may reflect collection biases.” The analyses presented in this paper suggest strongly that this worry was well-founded – particularly when the challenge is one of transferability, rather than interpolation.

By means of the spatial stratification manipulation used to create training and testing data sets, we have

shown that much of the spectrum of values in Maxent predictions is vulnerable to spatial biases in input data. If the argument is made that differences among the higher values resulting from Maxent analyses are less important than the overall area predicted at some low level (akin to the thresholds we used in our threshold-dependent comparisons), then the default result should

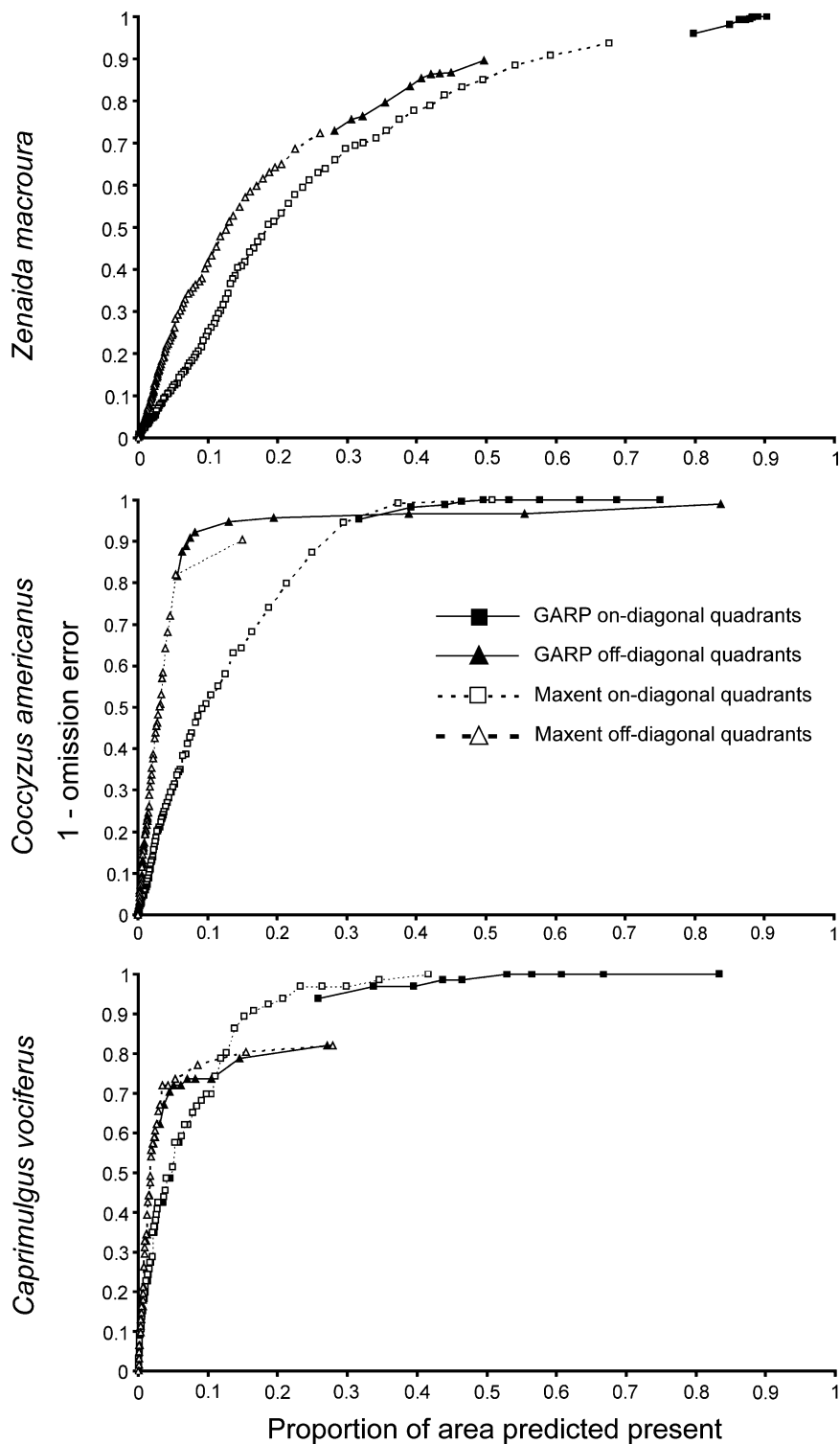


Fig. 5. Illustration of the relationship between area predicted present and (1 – omission error), a graph closely akin to a ROC plot. GARP threshold values (10) are represented by filled symbols, and Maxent values (100) are represented by empty symbols. The two pairs of quadrants are represented by triangles (off-diagonal quadrants) and squares (on-diagonal quadrants), respectively.

probably be changed to a binary output (0 versus greater than some low threshold), to avoid misleading Maxent users into attributing useful information to the detail among higher probabilities. At these lower levels of predictions, Maxent models resemble GARP models closely – broad geographic predictions that may prove somewhat overextensive – but we in general cannot distinguish the useful from the not-useful information above that threshold.

Choosing Maxent thresholds becomes complex precisely because of the contrast between interpolation and extrapolation challenges. Phillips et al. (2006): p. 239) stated that “the binary model obtained by setting a threshold of t will have approximately $t\%$ omission of test localities and minimum predicted area among all such models.” Although this statement was more or less valid regarding the interpolation challenge posed at the outset of this study (random subsetting), it does not hold for the spatially stratified challenge explored herein. In generalization exercises, reasonable thresholds among Maxent values will generate considerably higher omission errors among extrinsic test points than among intrinsic test points (Fig. 2).

Although ROC scores were comparable (i.e. not significantly different statistically) between GARP and Maxent, differences in the resulting predictions are clear. It is normally assumed that discrepancies are the results of model-specific relations between the proportion of area predicted present and omission error. Rather, small geographic areas are predicted present at high Maxent probability values that correspond to high omission errors (Fig. 5); most of the Maxent predictive range is condensed over small percentages of area predicted present, which leave out many of the occurrences, particularly in the independent test data set. Hence, when we take $t = 50$, omission rates based on extrapolative challenges range 72.7–88.4% (average 83.1%) rather than the expected 50%.

More generally, these results place some bounds on the applicability of Maxent in ecological niche modeling. Maxent's developers (Phillips et al. 2004) noted that Maxent is vulnerable to biases in input data sets; the challenge developed here is an extreme sort of bias. GARP performed poorly in comparison with Maxent in the interpolation challenge at the outset of this study, but succeeded in predicting whole species' distributions, rather than just the distribution of the input occurrence data. The broader “failure” of GARP in comparison with other modeling algorithms in distributional modeling challenges (Elith et al. 2006, Phillips et al. 2006) should probably be reexamined before the other, “better” algorithms are used for applications requiring transferability. Maxent results should be interpreted differently in two ways: 1) the threshold is not an indication of likely omission error in situations requiring transferability and prediction

into unsampled regions, and 2) much of the higher portion of the Maxent probability ramp likely represents noise and overfitting to biases in input data. When interpreted and thresholded properly, and the overfit portions of the predictive spectrum are not considered, Maxent models are potentially quite useful (Pearson et al. 2007), as other comparisons of the two methodologies have emphasized – Pearson et al. (2007) found that Maxent achieved better predictive success rates, particularly at small sample sizes, and Phillips et al. (2006) found that Maxent models outperformed GARP models, at least as evaluated using ROC AUC approaches. Still, given the effects in ROC AUC scores illustrated and explored in this study, the details of the best interpretation of this methodology require considerable additional consideration.

The results of this study suggest that different applications present quite-different challenges for modeling algorithms. The transferability challenge places constraints on the algorithms – generality and non-specificity – that may limit the detail that they are able to achieve in interpolation. That a “loser” at one challenge can transform into a winner in the other suggests that much more exploration and testing is warranted.

Finally, we note that the receiver operating characteristic AUC statistic was quite insensitive to the differences noted above. GARP and Maxent performed similarly when AUC statistics were examined, yet they achieved this success in very different ways: Maxent by avoiding commission error, and GARP by avoiding omission error. AUCs capture the general quality of avoiding predictive error, but do not distinguish between useful and non-useful predictions – transferring Maxent models to novel geographic situations requires considerable caution in interpretation as predictions of species' entire distributional areas when sampling is uneven or incomplete across the entire area. The particulars of ROC AUC applications in niche modeling, along with a modification of the testing methodology that limits the extent of the AUC calculations to relevant portions of the spectra of possible omission error and possible proportional areas predicted present, will be presented in a future contribution (Peterson et al. unpubl.).

Acknowledgements – A special thanks to the National Center for Ecological Analysis and Synthesis Species Distribution Modeling Group for numerous debates and discussions of these issues, which have helped greatly to clarify the picture. We also thank Richard Pearson and Robert

Anderson for helpful comments on earlier versions of this manuscript.

References

- Anderson, R. P. et al. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. – *Ecol. Modell.* 162: 211–232.
- Anon. 2001. HYDRO1k Elevation Derivative Database. – U.S. Geological Survey, <<http://edcdaac.usgs.gov/gtopo30/hydro/>>.
- Araújo, M. B. and Rahbek, C. 2006. How does climate change affect biodiversity? – *Science* 313: 1396–1397.
- Araújo, M. B. et al. 2005a. Validation of species-climate impact models under climate change. – *Global Change Biol.* 11: 1504–1513.
- Araújo, M. B. et al. 2005b. Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. – *Global Ecol. Biogeogr.* 14: 17–30.
- Austin, M. P. 1985. Continuum concept, ordination methods, and niche theory. – *Annu. Rev. Ecol. Syst.* 16: 39–61.
- Bio, A. M. F. et al. 2002. Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. – *Biodiv. Conserv.* 11: 2189–2216.
- DeLong, E. R. et al. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. – *Biometrics* 44: 837–845.
- Elith, J. and Burgman, M. 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria. – In: Scott, J. M. et al. (eds), *Predicting species occurrences: issues of scale and accuracy*. Island Press, pp. 303–313.
- Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.
- Eng, J. 2005. ROC analysis: web-based calculator for ROC curves. – <<http://www.rad.jhmi.edu/jeng/javarad/roc/main.html>>.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – *Environ. Conserv.* 24: 38–49.
- Glozier, N. E. et al. 1997. Transferability of habitat suitability curves for a benthic minnow, *Rhinichthys cataractae*. – *J. Freshwater Ecol.* 12: 379–393.
- Graham, C. H. et al. 2004. New developments in museum-based informatics and applications in biodiversity analysis. – *Trends Ecol. Evol.* 19: 497–503.
- Guay, J. C. et al. 2003. Assessment of the transferability of biological habitat models for Atlantic salmon parr (*Salmo salar*). – *Can. J. Fish. Aquat. Sci.* 60: 1398–1408.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – *Ecol. Modell.* 135: 147–186.
- Guisan, A. et al. 2007a. What matters for predicting spatial distributions of trees: techniques, data, or species' characteristics? – *Ecol. Monogr.*, in press.
- Guisan, A. et al. 2007b. Sensitivity of predictive species distribution models to change in grain size. – *Div. Distrib.* 13: 332–340.
- Hijmans, R. J. et al. 2005. WorldClim, ver. 1.3. – Univ. of California, <<http://biogeography.berkeley.edu/worldclim/worldclim.htm>>.
- Kleyer, M. 2002. Validation of plant functional types across two contrasting landscapes. – *J. Veg. Sci.* 13: 167–178.
- Liu, C. et al. 2005. Selecting thresholds of occurrence in the prediction of species distributions. – *Ecography* 28: 385–393.
- Manel, S. et al. 1999a. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. – *J. Appl. Ecol.* 36: 734–747.
- Manel, S. et al. 1999b. Comparing discriminant analysis, neural networks, and logistic regression for predicting species distributions: a case study with a Himalayan river bird. – *Ecol. Modell.* 120: 337–347.
- Pearson, R. G. et al. 2006. Model-based uncertainty in species range prediction. – *J. Biogeogr.* 33: 1704–1711.
- Pearson, R. G. et al. 2007. Predicting species' distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. – *J. Biogeogr.* 34: 102–117.
- Pereira, R. S. 2002. Desktop GARP. – <<http://www.life-mapper.org/desktopgarp/>>.
- Peterson, A. T. 2006. Uses and requirements of ecological niche models and related distributional models. – *Biodiv. Inform.* 3: 59–72.
- Peterson, A. T. and Cohoon, K. C. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. – *Ecol. Modell.* 117: 159–164.
- Peterson, A. T. and Shaw, J. J. 2003. *Lutzomyia* vectors for cutaneous leishmaniasis in southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. – *Int. J. Parasitol.* 33: 919–931.
- Phillips, S. J. 2006. Maxent software for species habitat modeling, ver. 2.2.0. – <<http://www.cs.princeton.edu/~schapire/maxent/>>.
- Phillips, S. J. et al. 2004. A maximum entropy approach to species distribution modeling. – In: Greiner, R. and Schuurmans, D. (eds), *Proc. 21st Int. Conf. on Machine Learning*. ACM press, New York, pp. 655–662.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Modell.* 190: 231–259.
- Randin, C. F. et al. 2006. Are niche-based species distribution models transferable in space? – *J. Biogeogr.* 33: 1689–1703.
- Sauer, J. R. et al. 2001. The North American Breeding Bird Survey, results and analysis 1966–2000, ver. 2001.2. – USGS Patuxent Wildlife Research Center, Laurel, MD.
- Schröder, B. and Richter, O. 1999. Are habitat models transferable in space and time? – *Zeitschr. Ökol. Natursch.* 8: 195–205.

- Soberón, J. and Peterson, A. T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. – *Phil. Trans. R. Soc. B* 359: 689–698.
- Soberón, J. and Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. – *Biodiv. Inform.* 2: 1–10.
- Stockwell, D. R. B. and Noble, I. R. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of analysis. – *Math. Comp. Simul.* 33: 385–390.
- Stockwell, D. R. B. and Peterson, A. T. 2002a. Controlling bias in biodiversity data. – In: Scott, J. M. et al. (eds), *Predicting species occurrences: issues of scale and accuracy*. Island Press, pp. 537–546.
- Stockwell, D. R. B. and Peterson, A. T. 2002b. Effects of sample size on accuracy of species distribution models. – *Ecol. Modell.* 148: 1–13.
- Thomas, C. D. et al. 2004. Extinction risk from climate change. – *Nature* 427: 145–148.
- Thomas, J. A. and Bovee, K. D. 1993. Application and testing of a procedure to evaluate transferability of habitat suitability criteria. – *Reg. Rivers: Res. Manage.* 8: 285–294.
- Thuiller, W. et al. 2004. Do we need land-cover data to model species distributions in Europe? – *J. Biogeogr.* 31: 353–361.
- Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – *Proc. Nat. Acad. Sci. USA* 102: 8245–8250.