ELSEVIER

Maximum entropy modeling of species geographic distributions

Steven J. Phillips[a],*, Robert P. Anderson[b,c], Robert E. Schapire[d]

[a] *AT&T Labs-Research, 180 Park Avenue, Florham Park, NJ 07932, USA*
[b] *Department of Biology, City College of the City University of New York, J-526 Marshak Science Building,*
Convent Avenue at 138th Street, New York, NY 10031, USA
[c] *Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West at 79th Street,*
New York, NY 10024, USA
[d] *Computer Science Department, Princeton University, 35 Olden Street, Princeton, NJ 08544, USA*

Abstract

The availability of detailed environmental data, together with inexpensive and powerful computers, has fueled a rapid increase in predictive modeling of species environmental requirements and geographic distributions. For some species, detailed presence/absence occurrence data are available, allowing the use of a variety of standard statistical techniques. However, absence data are not available for most species. In this paper, we introduce the use of the maximum entropy method (Maxent) for modeling species geographic distributions with presence-only data. Maxent is a general-purpose machine learning method with a simple and precise mathematical formulation, and it has a number of aspects that make it well-suited for species distribution modeling. In order to investigate the efficacy of the method, here we perform a continental-scale case study using two Neotropical mammals: a lowland species of sloth, *Bradypus variegatus*, and a small montane murid rodent, *Microryzomys minutus*. We compared Maxent predictions with those of a commonly used presence-only modeling method, the Genetic Algorithm for Rule-Set Prediction (GARP). We made predictions on 10 random subsets of the occurrence records for both species, and then used the remaining localities for testing. Both algorithms provided reasonable estimates of the species' range, far superior to the shaded outline maps available in field guides. All models were significantly better than random in both binomial tests of omission and receiver operating characteristic (ROC) analyses. The area under the ROC curve (AUC) was almost always higher for Maxent, indicating better discrimination of suitable versus unsuitable areas for the species. The Maxent modeling approach can be used in its present form for many applications with presence-only datasets, and merits further research and development.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Maximum entropy; Distribution; Modeling; Niche; Range

* Corresponding author. Tel.: +1 973 360 8704;
fax: +1 973 360 8871.
E-mail addresses: phillips@research.att.com
(S.J. Phillips), anderson@sci.ccny.cuny.edu (R.P. Anderson),
schapire@cs.princeton.edu (R.E. Schapire).

1. Introduction

Predictive modeling of species geographic distributions based on the environmental conditions of sites of known occurrence constitutes an important tech-

nique in analytical biology, with applications in conservation and reserve planning, ecology, evolution, epidemiology, invasive-species management and other fields (Corsi et al., 1999; Peterson and Shaw, 2003; Peterson et al., 1999; Scott et al., 2002; Welk et al., 2002; Yom-Tov and Kadmon, 1998). Sometimes both presence and absence occurrence data are available for the development of models, in which case general-purpose statistical methods can be used (for an overview of the variety of techniques currently in use, see Corsi et al., 2000; Elith, 2002; Guisan and Zimmerman, 2000; Scott et al., 2002). However, while vast stores of presence-only data exist (particularly in natural history museums and herbaria), absence data are rarely available, especially for poorly sampled tropical regions where modeling potentially has the most value for conservation (Anderson et al., 2002; Ponder et al., 2001; Soberón, 1999). In addition, even when absence data are available, they may be of questionable value in many situations (Anderson et al., 2003). Modeling techniques that require only presence data are therefore extremely valuable (Graham et al., 2004).

1.1. Niche-based models from presence-only data

We are interested in devising a model of a species' environmental requirements from a set of occurrence localities, together with a set of environmental variables that describe some of the factors that likely influence the suitability of the environment for the species (Brown and Lomolino, 1998; Root, 1988). Each occurrence locality is simply a latitude–longitude pair denoting a site where the species has been observed; such georeferenced occurrence records often derive from specimens in natural history museums and herbaria (Ponder et al., 2001; Stockwell and Peterson, 2002a). The environmental variables in GIS format all pertain to the same geographic area, the study area, which has been partitioned into a grid of pixels. The task of a modeling method is to predict environmental suitability for the species as a function of the given environmental variables.

A niche-based model represents an approximation of a species' ecological niche in the examined environmental dimensions. A species' fundamental niche consists of the set of all conditions that allow for its long-term survival, whereas its realized niche is that subset of the fundamental niche that it actually occupies

(Hutchinson, 1957). The species' realized niche may be smaller than its fundamental niche, due to human influence, biotic interactions (e.g., inter-specific competition, predation), or geographic barriers that have hindered dispersal and colonization; such factors may prevent the species from inhabiting (or even encountering) conditions encompassing its full ecological potential (Pulliam, 2000; Anderson and Martínez-Meyer, 2004). We assume here that occurrence localities are drawn from source habitat, rather than sink habitat, which may contain a given species without having the conditions necessary to maintain the population without immigration; this assumption is less realistic with highly vagile taxa (Pulliam, 2000). By definition, then, environmental conditions at the occurrence localities constitute samples from the realized niche. A niche-based model thus represents an approximation of the species' realized niche, in the study area and environmental dimensions being considered.

If the realized niche and fundamental niche do not fully coincide, we cannot hope for any modeling algorithm to characterize the species' full fundamental niche: the necessary information is simply not present in the occurrence localities. This problem is likely exacerbated when occurrence records are drawn from too small a geographic area. In a larger study region, however, spatial variation exists in community composition (and, hence, in the resulting biotic interactions) as well as in the environmental conditions available to the species. Therefore, given sufficient sampling effort, modeling in a study region with a larger geographic extent is likely to increase the fraction of the fundamental niche represented by the sample of occurrence localities (Peterson and Holt, 2003), and is preferable. In practice, however, the departure between the fundamental niche (a theoretical construct) and realized niche (which can be observed) of a species will remain unknown.

Although a niche-based model describes suitability in ecological space, it is typically projected into geographic space, yielding a geographic area of predicted presence for the species. Areas that satisfy the conditions of a species' fundamental niche represent its potential distribution, whereas the geographic areas it actually inhabits constitute its realized distribution. As mentioned above, the realized niche may be smaller than the fundamental niche (with respect to the environmental variables being modeled), in which

case the predicted distribution will be smaller than the full potential distribution. However, to the extent that the model accurately portrays the species' fundamental niche, the projection of the model into geographic space will represent the species' potential distribution.

Whether or not a model captures a species' full niche requirements, areas of predicted presence will typically be larger than the species' realized distribution. Due to many possible factors (such as geographic barriers to dispersal, biotic interactions, and human modification of the environment), few species occupy all areas that satisfy their niche requirements. If required by the application at hand, the species' realized distribution can often be estimated from the modeled distribution through a series of steps that remove areas that the species is known or inferred not to inhabit. For example, suitable areas that have not been colonized due to contingent historical factors (e.g., geographic barriers) can be excluded (Peterson et al., 1999; Anderson, 2003). Similarly, suitable areas not inhabited due to biotic interactions (e.g., competition with closely related morphologically similar species) can be identified and removed from the prediction (Anderson et al., 2002). Finally, when a species' present-day distribution is desired, such as for conservation purposes, a current land-cover classification derived from remotely sensed data can be used to exclude highly altered habitats (e.g., removing deforested areas from the predicted distribution of an obligate-forest species; Anderson and Martínez-Meyer, 2004).

There are implicit ecological assumptions in the set of environmental variables used for modeling, so selection of that set requires great care. Temporal correspondence should exist between occurrence localities and environmental variables; for example, a current land-cover classification should not be used with occurrence localities that derive from museum records collected over many decades (Anderson and Martínez-Meyer, 2004). Secondly, the variables should affect the species' distribution at the relevant scale, determined by the geographic extent and grain of the modeling task (Pearson et al., 2004). For example, using the terminology of Mackey and Lindenmayer (2001), climatic variables such as temperature and precipitation are appropriate at global and meso-scales; topographic variables (e.g., elevation and aspect) likely affect species distributions at meso- and topo-scales; and land-cover variables like percent canopy cover

influence species distributions at the micro-scale. The choice of variables to use for modeling also affects the degree to which the model generalizes to regions outside the study area or to different environmental conditions (e.g., other time periods). This is important for applications such as invasive-species management (e.g., Peterson and Robins, 2003) and predicting the impact of climate change (e.g., Thomas et al., 2004). Bioclimatic and soil-type variables measure availability of the fundamental primary resources of light, heat, water and mineral nutrients (Mackey and Lindenmayer, 2001). Their impact, as measured in one study area or time frame, should generalize to other situations. On the other hand, variables representing latitude or elevation will not generalize well; although they are correlated with variables that have biophysical impact on the species, those correlations vary over space and time.

A number of other serious potential pitfalls may affect the accuracy of presence-only modeling; some of these also apply to presence–absence modeling. First, occurrence localities may be biased. For example, they are often highly correlated with the nearby presence of roads, rivers or other access conduits (Reddy and Dávalos, 2003). The location of occurrence localities may also exhibit spatial auto-correlation (e.g., if a researcher collects specimens from several nearby localities in a restricted area). Similarly, sampling intensity and sampling methods often vary widely across the study area (Anderson, 2003). In addition, errors may exist in the occurrence localities, be it due to transcription errors, lack of sufficient geographic detail (especially in older records), or species misidentification. Frequently, the number of occurrence localities may be too low to estimate the parameters of the model reliably (Stockwell and Peterson, 2002b). Similarly, the set of available environmental variables may not be sufficient to describe all the parameters of the species' fundamental niche that are relevant to its distribution at the grain of the modeling task. Finally, errors may be present in the variables, perhaps due to errors in data manipulation, or due to inaccuracies in the climatic models used to generate climatic variables, or interpolation of lower-resolution data. In sum, determining and possibly mitigating the effects of these factors represent worthy topics of research for all presence-only modeling techniques. With these caveats, we proceed to introduce a modeling approach that may prove use-

ful whenever the above concerns are adequately addressed.

1.2. Maxent

Maxent is a general-purpose method for making predictions or inferences from incomplete information. Its origins lie in statistical mechanics (Jaynes, 1957), and it remains an active area of research with an Annual Conference, Maximum Entropy and Bayesian Methods, that explores applications in diverse areas such as astronomy, portfolio optimization, image reconstruction, statistical physics and signal processing. We introduce it here as a general approach for presence-only modeling of species distributions, suitable for all existing applications involving presence-only datasets. The idea of Maxent is to estimate a target probability distribution by finding the probability distribution of maximum entropy (i.e., that is most spread out, or closest to uniform), subject to a set of constraints that represent our incomplete information about the target distribution. The information available about the target distribution often presents itself as a set of real-valued variables, called "features", and the constraints are that the expected value of each feature should match its empirical average (average value for a set of sample points taken from the target distribution). When Maxent is applied to presence-only species distribution modeling, the pixels of the study area make up the space on which the Maxent probability distribution is defined, pixels with known species occurrence records constitute the sample points, and the features are climatic variables, elevation, soil category, vegetation type or other environmental variables, and functions thereof.

Maxent offers many advantages, and a few drawbacks; a comparison with other modeling methods will be made in Section 2.1.4 after the Maxent approach is described in detail. The advantages include the following: (1) It requires only presence data, together with environmental information for the whole study area. (2) It can utilize both continuous and categorical data, and can incorporate interactions between different variables. (3) Efficient deterministic algorithms have been developed that are guaranteed to converge to the optimal (maximum entropy) probability distribution. (4) The Maxent probability distribution has a concise mathematical definition, and is therefore amenable to analysis. For example, as with generalized linear and generalized additive models (GLM and GAM), in the absence of interactions between variables, additivity of the model makes it possible to interpret how each environmental variable relates to suitability (Dudík et al., 2004; Phillips et al., 2004). (5) Over-fitting can be avoided by using ℓ_1-regularization (Section 2.1.2). (6) Because dependence of the Maxent probability distribution on the distribution of occurrence localities is explicit, there is the potential (in future work) to address the issue of sampling bias formally, as in Zadrozny (2004). (7) The output is continuous, allowing fine distinctions to be made between the modeled suitability of different areas. If binary predictions are desired, this allows great flexibility in the choice of threshold. If the application is conservation planning, the fine distinctions in predicted relative environmental suitability can be valuable to reserve planning algorithms. (8) Maxent could also be applied to species presence/absence data by using a conditional model (as in Berger et al., 1996), as opposed to the unconditional model used here. (9) Maxent is a generative approach, rather than discriminative, which can be an inherent advantage when the amount of training data is limited (see Section 2.1.4). (10) Maximum entropy modeling is an active area of research in statistics and machine learning, and progress in the field as a whole can be readily applied here. (11) As a general-purpose and flexible statistical method, we expect that it can be used for all the applications outlined in Section 1 above, and at all scales.

Some drawbacks of the method are: (1) It is not as mature a statistical method as GLM or GAM, so there are fewer guidelines for its use in general, and fewer methods for estimating the amount of error in a prediction. Our use of an "unconditional" model (cf. advantage 8) is rare in machine learning. (2) The amount of regularization (see Section 2.1.2) requires further study (e.g., see Phillips et al., 2004), as does its effectiveness in avoiding over-fitting compared with other variable-selection methods (for alternatives, see Guisan et al., 2002). (3) It uses an exponential model for probabilities, which is not inherently bounded above and can give very large predicted values for environmental conditions outside the range present in the study area. Extra care is therefore needed when extrapolating to another study area or to future or past climatic conditions (for example, feature values outside the range of values in the study area should be "clamped", or reset to the appropriate upper or lower bound). (4) Special-purpose

software is required, as Maxent is not available in standard statistical packages.

1.3. Existing approaches for presence-only modeling

Many methods have been used for presence-only modeling of species distributions, and we only attempt here to give a broad overview of existing methods. Some methods use only presences to derive a model. BIOCLIM (Busby, 1986; Nix, 1986) predicts suitable conditions in a "bioclimatic envelope", consisting of a rectilinear region in environmental space representing the range (or some percentage thereof) of observed presence values in each environmental dimension. Similarly, DOMAIN (Carpenter et al., 1993) uses a similarity metric, where a predicted suitability index is given by computing the minimum distance in environmental space to any presence record.

Other techniques use presence and background data. General-purpose statistical methods such as generalized linear models (GLMs) and generalized additive models (GAMs) are commonly used for modeling with presence–absence datasets. Recently, they have been applied to presence-only situations by taking a random sample of pixels from the study area, known as "background pixels" or "pseudo-absences", and using them in place of absences during modeling (Ferrier and Watson, 1996; Ferrier et al., 2002). A sample of the background pixels can be chosen purely at random (sometimes excluding sites with presence records, Graham et al., 2004), or from sites where sampling is known to have occurred or from a model of such sites (Zaniewski et al., 2002; Engler et al., 2004). Similarly, a Bayesian approach (Aspinall, 1992) proposed modeling presence versus a random sample. The Genetic Algorithm for Rule-Set Prediction (Stockwell and Noble, 1992; Stockwell and Peters, 1999) uses an artificial-intelligence framework called genetic algorithms. It produces a set of positive and negative rules that together give a binary prediction; rules are favored in the algorithm according to their significance (compared with random prediction) based on a sample of background pixels and presence pixels. Environmental-Niche Factor Analysis (ENFA, Hirzel et al., 2002) uses presence localities together with environmental data for the entire study area, without requiring a sample of the background to be treated like

absences. It is similar to principal components analysis, involving a linear transformation of the environmental space into orthogonal "marginality" and "specialization" factors. Environmental suitability is then modeled as a Manhattan distance in the transformed space.

As a first step in the evaluation of Maxent, we chose to compare it with GARP, as the latter has recently seen extensive use in presence-only studies (Anderson, 2003; Joseph and Stockwell, 2002; Peterson and Kluza, 2003; Peterson and Robins, 2003; Peterson and Shaw, 2003 and references therein). While further studies are needed comparing Maxent with other widely used methods that have been applied to presence-only datasets, such studies are beyond the scope of this paper.

2. Methods

2.1. Maxent details

2.1.1. The principle

When approximating an unknown probability distribution, the question arises, what is the best approximation? E.T. Jaynes gave a general answer to this question: the best approach is to ensure that the approximation satisfies any constraints on the unknown distribution that we are aware of, and that subject to those constraints, the distribution should have maximum entropy (Jaynes, 1957). This is known as the maximum-entropy principle. For our purposes, the unknown probability distribution, which we denote π, is over a finite set X, (which we will later interpret as the set of pixels in the study area). We refer to the individual elements of X as points. The distribution π assigns a non-negative probability $\pi(x)$ to each point x, and these probabilities sum to 1. Our approximation of π is also a probability distribution, and we denote it $\hat{\pi}$. The entropy of $\hat{\pi}$ is defined as

$$H(\hat{\pi}) = - \sum_{x \in X} \hat{\pi}(x) \ln \hat{\pi}(x)$$

where ln is the natural logarithm. The entropy is non-negative and is at most the natural log of the number of elements in X. Entropy is a fundamental concept in information theory: in the paper that originated that field, Shannon (1948) described entropy as "a measure

of how much 'choice' is involved in the selection of an event". Thus a distribution with higher entropy involves more choices (i.e., it is less constrained). Therefore, the maximum entropy principle can be interpreted as saying that no unfounded constraints should be placed on $\hat{\pi}$, or alternatively,

The fact that a certain probability distribution maximizes entropy subject to certain constraints representing our incomplete information, is the fundamental property which justifies use of that distribution for inference; it agrees with everything that is known, but carefully avoids assuming anything that is not known (Jaynes, 1990).

2.1.2. A machine learning perspective

The maximum entropy principle has seen recent interest in the machine learning community, with a major contribution being the development of efficient algorithms for finding the Maxent distribution (see Berger et al., 1996 for an accessible introduction and Ratnaparkhi, 1998 for a variety of applications and a favorable comparison with decision trees). The approach consists of formalizing the constraints on the unknown probability distribution π in the following way. We assume that we have a set of known real-valued functions $f_1, \ldots, f_n$ on X, known as "features" (which for our application will be environmental variables or functions thereof). We assume further that the information we know about π is characterized by the expectations (averages) of the features under π. Here, each feature f_j assigns a real value $f_j(x)$ to each point x in X. The expectation of the feature f_j under π is defined as $\sum_{x \in X} \pi(x) f_j(x)$ and denoted by $\pi[f_j]$. In general, for any probability distribution p and function f, we use the notation $p[f]$ to denote the expectation of f under p.

The feature expectations $\pi[f_j]$ can be approximated using a set of sample points $x_1, \ldots, x_m$ drawn independently from X (with replacement) according to the probability distribution π. The empirical average of f_j is $\frac{1}{m} \sum_{i=1}^{m} f_j(x_i)$, which we can write as $\tilde{\pi}[f_j]$ (where $\tilde{\pi}$ is the uniform distribution on the sample points), and use as an estimate of $\pi[f_j]$. By the maximum entropy principle, therefore, we seek the probability distribution $\hat{\pi}$ of maximum entropy subject to the constraint that each feature f_j has the same mean under $\hat{\pi}$ as observed empirically, i.e.

$$\hat{\pi}[f_j] = \tilde{\pi}[f_j], \quad \text{for each feature } f_j \tag{1}$$

It turns out that the mathematical theory of convex duality can be used (Della Pietra et al., 1997) to show that this characterization uniquely determines $\hat{\pi}$, and that $\hat{\pi}$ has an alternative characterization, which can be described as follows. Consider all probability distributions of the form

$$q_\lambda(x) = \frac{e^{\lambda \cdot f(x)}}{Z_\lambda} \tag{2}$$

where λ is a vector of n real-valued coefficients or feature weights, f denotes the vector of all n features, and Z_λ is a normalizing constant that ensures that q_λ sums to 1. Such distributions are known as Gibbs distributions. Convex duality shows that the Maxent probability distribution $\hat{\pi}$ is exactly equal to the Gibbs probability distribution q_λ that maximizes the likelihood (i.e., probability) of the m sample points. Equivalently, it minimizes the negative log likelihood of the sample points

$$\tilde{\pi}[-\ln(q_\lambda)] \tag{3}$$

which can also be written $\ln Z_\lambda - \frac{1}{m} \sum_{i=1}^{m} \lambda \cdot f(x_i)$ and termed the "log loss".

As described so far, Maxent can be prone to overfitting the training data. The problem derives from the fact that the empirical feature means will typically not equal the true means; they will only approximate them. Therefore the means under $\hat{\pi}$ should only be restricted to be close to their empirical values. One way this can be done is to relax the constraint in (1) above (Dudík et al., 2004), replacing it with

$$|\hat{\pi}[f_j] - \tilde{\pi}[f_j]| \leq \beta_j, \quad \text{for each feature } f_j \tag{4}$$

for some constants β_j. This also changes the dual characterization, resulting in a form of ℓ_1-regularization: the Maxent distribution can now be shown to be the Gibbs distribution that minimizes

$$\tilde{\pi}[-\ln(q_\lambda)] + \sum_j \beta_j |\lambda_j| \tag{5}$$

where the first term is the log loss (as in (3) above), while the second term penalizes the use of large values for the weights λ_j. Regularization forces Maxent to focus on the most important features, and ℓ_1-

regularization tends to produce models with few non-zero λ_j values (Williams, 1995). Such models are less likely to overfit, because they have fewer parameters; as a general rule, the simplest explanation of a phenomenon is usually best (the principle of parsimony, Occam's Razor). Note that ℓ_1 regularization has also been applied to GLM/GAMs, and is called the "lasso" in that context (Guisan et al., 2002 and references therein).

This maximum likelihood formulation suggests a natural approach for finding the Maxent probability distribution: start from the uniform probability distribution, for which $\boldsymbol{\lambda} = (0, \ldots, 0)$, then repeatedly make adjustments to one or more of the weights λ_j in such a way that the regularized log loss decreases. Regularized log loss can be shown to be a convex function of the weights, so no local minima exist, and several convex optimization methods exist for adjusting the weights in a way that guarantees convergence to the global minimum (see Section 2.2 for the algorithm used in this study).

The above presentation describes an "unconditional" maximum entropy model. "Conditional" models are much more common in the machine learning literature. The task of a conditional Maxent model is to approximate a joint probability distribution $p(\boldsymbol{x}, y)$ of the inputs $\boldsymbol{x}$ and output label y. Both presence and absence data would be required to train a conditional model of a species' distribution, which is why we use unconditional models.

2.1.3. Application to species distribution modeling

Austin (2002) examines three components needed for statistical modeling of species distributions: an ecological model concerning the ecological theory being used, a data model concerning collection of the data, and a statistical model concerning the statistical theory. Maxent is a statistical model, and to apply it to model species distributions successfully, we must consider how it relates to the two other modeling components (the data model and ecological model). Using the notation of Section 2.1.2, we define the set X to be the set of pixels in the study area, and interpret the recorded presence localities for the species as sample points $x_1, \ldots, x_m$ taken from an unknown probability distribution π. The data model consists of the method by which the presence localities were collected. One idealized sampling strategy is to pick a random pixel,

and record 1 if the species is present there, and 0 otherwise. If we denote the response variable as y, then under this sampling strategy, π is the probability distribution $p(x|y = 1)$. By applying Bayes' rule, we get that π is proportional to probability of occurrence, $p(y = 1|x)$, although with presence-only data we cannot determine the constant of proportionality.

However, most presence-only datasets derive from surveys where the data model is much less well-defined that the idealized model presented above. The various sampling biases described in Section 1 seriously violate this data model. In practice, then, π (and $\hat{\pi}$) can be more conservatively interpreted as a relative index of environmental suitability, where higher values represent a prediction of better conditions for the species (similar to the relaxed interpretation of GLMs with presence-only data in Ferrier et al. (2002)).

The critical step in formulating the ecological model is defining a suitable set of features. Indeed, the constraints imposed by the features represent our ecological assumptions, as we are asserting that they represent all the environmental factors that constrain the geographical distribution of the species. We consider five feature types, described in Dudík et al. (2004). We did not use the fourth in our present study, as it may require more data than were available for our study species.

1. A continuous variable f is itself a "linear feature". It imposes the constraint on $\hat{\pi}$ that the mean of the environmental variable, $\hat{\pi}[f]$, should be close to its observed value, i.e., its mean on the sample localities.

2. The square of a continuous variable f is a "quadratic feature". When used with the corresponding linear feature, it imposes the constraint on $\hat{\pi}$ that the variance of the environmental variable should be close to its observed value, since the variance is equal to $\hat{\pi}[f^2] - \hat{\pi}[f]^2$. It models the species' tolerance for variation from its optimal conditions.

3. The product of two continuous environmental variables f and g is a "product feature". Together with the linear features for f and g, it imposes the constraint that the covariance of those two variables should be close to its observed value, since the covariance is $\hat{\pi}[fg] - \hat{\pi}[f]\hat{\pi}[g]$. Product features therefore incorporate interactions between predictor variables.

4. For a continuous environmental variable f, a "threshold feature" is equal to 1 when f is above a given

threshold, and 0 otherwise. It imposes the following constraint: the proportion of π that has values for f above the threshold should be close to the observed proportion. All possible threshold features for f together allow Maxent to model an arbitrary response curve of the species to f, as any smooth function can be approximated by a linear combination of threshold functions.

5. For a categorical environmental variable that takes on values $v_1 \ldots v_k$, we use k "binary features", where the ith feature is 1 wherever the variable equals v_i, and 0 otherwise. As with threshold features, these binary features constrain the proportion of $\hat{\pi}$ in each category to be close to the observed proportion.

For each of these feature types, the corresponding regularization parameter β_j governs how close the expectation under $\hat{\pi}$ is required to be to the observed value; without regularization, they are required to be equal (Section 2.1.2). The above list of features types is not exhaustive, and additional feature types could be derived from the same environmental variables. The features used should be those that likely constrain the geographic distribution of the species.

The applicability of the maximum entropy principle to species distributions is supported by thermodynamic theories of ecological processes (Aoki, 1989; Schneider and Kay, 1994). The second law of thermodynamics specifies that in systems without outside influences, processes move in a direction that maximizes entropy. Thus, in the absence of influences other than those included as constraints in the model, the geographic distribution of a species will indeed tend toward the distribution of maximum entropy.

2.1.4. Relationships to other modeling approaches

Maxent has strong similarities to some existing methods for modeling species distributions, in particular, generalized linear models (GLMs), generalized additive models (GAMs) and machine learning methods such as Bayesian approaches and neural networks. GLMs, GAMs, Bayesian approaches and neural networks are all broad classes of techniques, and we refer here only to the way they have been applied to presence-only modeling of species distributions. Similarly, Maxent generally refers to a class of techniques, but we

restrict our discussion to ℓ_1-regularized unconditional Maxent, as described in Section 2.1.2.

Theoretically, Maxent is most similar to GLMs and GAMs. In what follows, we use the terminology of Yee and Mitchell (1991). A frequently-used GLM is the Guassian logit model, in which the logit of the predicted probability of occurrence is

$$\alpha + \beta_1 f_1(x) + \gamma_1 f_1(x)^2 + \ldots + \beta_n f_n(x) + \gamma_n f_n(x)^2 \tag{6}$$

where the f_j are environmental variables, α, β_j and γ_j are fitted coefficients, and the logit function is defined by $\mathrm{logit}(p) = \ln(\frac{p}{1-p})$. The expression in (6) is the same form as the log (rather than logit) of the probability of the pixel x in a Maxent model with linear and quadratic features. A common method for modeling interactions between variables in a GLM is to create product variables, which is analogous to the use of product features in Maxent.

In the same way, if probability of occurrence is modeled with a GAM using a logit link function, the logit of the predicted probability has the form

$$g_1(f_1(x)) + \ldots + g_n(f_n(x))$$

where the f_i are again environmental variables. The g_i are smooth functions fit by the model, with the amount of smoothing controlled by a width parameter. This is the same form as the log probability of the pixel x in a Maxent model with threshold features, and regularization has an analogous effect to smoothing on the otherwise arbitrary functions g_i. In both cases, the shape of the response curve to each environmental variable is determined by the data.

Despite these similarities, important differences exist between GLM/GAMs and Maxent, causing them to make different predictions. When GLM/GAMs are used to model probability of occurrence, absence data are required. When applied to presence-only data, background pixels must be used instead of true absences (Ferrier and Watson, 1996; Ferrier et al., 2002). However, the interpretation of the result is less clear-cut—it must be interpreted as a relative index of environmental suitability. In contrast, Maxent models a probability distribution over the pixels in the study region, and in no sense are pixels without species records interpreted as absences. In addition, Maxent

is a generative approach, whereas GLM/GAMs are discriminative, and generative methods may give better predictions when the amount of training data is small (Ng and Jordan, 2001). For a joint probability distribution $p(\boldsymbol{x}, y)$, a discriminative classifier models the posterior probability $p(y|\boldsymbol{x})$ directly, in order to choose the most likely label y for given inputs $\boldsymbol{x}$. Typically, a generative classifier models the distribution $p(\boldsymbol{x}, y)$ or $p(\boldsymbol{x}|y)$, and relies on Bayes' rule to determine $p(y|\boldsymbol{x})$. Our unconditional Maxent models are generative: we model a distribution $p(\boldsymbol{x}|y = 1)$.

Maxent shares with other machine learning methods an emphasis on probabilistic reasoning. Regularization, which penalizes the use of large values of model parameters, can be interpreted as the use of a Bayesian prior (Williams, 1995). However, Maxent is quite different from the particular Bayesian species modeling approach of Aspinall (1992). The latter approach is known as "naive Bayes" in the machine learning literature, and assumes independence of the environmental variables. This assumption is frequently not met for environmental data.

The data requirements of Maxent are closest to those of environmental niche factor analysis (ENFA), which also uses presence data in combination with environmental data for the whole study area (although both could use only a random sample of background pixels to improve running time).

2.2. A Maxent implementation for modeling species distributions

In order to make the Maxent method available for modeling species geographic distributions, we implemented an efficient algorithm together with a choice of feature types that are well suited to the task. Our implementation uses a sequential-update algorithm (Dudík et al., 2004) that iteratively picks a weight λ_j and adjusts it so as to minimize the resulting regularized log loss. The algorithm is deterministic, and is guaranteed to converge to the Maxent probability distribution. The algorithm stops when a user-specified number of iterations has been performed, or when the change in log loss in an iteration falls below a user-specified value (convergence), whichever happens first.

As described in Section 2.1, Maxent assigns a non-negative probability to each pixel in the study area. Because these probabilities must sum to 1, each prob-

ability is typically extremely small. Although these "raw" probabilities are an optional output, by default our software presents the probability distribution in another form that is easier to use and interpret, namely a "cumulative" representation. The value assigned to a pixel is the sum of the probabilities of that pixel and all other pixels with equal or lower probability, multiplied by 100 to give a percentage. The cumulative representation can be interpreted as follows: if we resample pixels according to the modeled Maxent probability distribution, then $t\%$ of the resampled pixels will have cumulative value of t or less. Thus, if the Maxent distribution $\hat{\pi}$ is a close approximation of the probability distribution π that represents reality, 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 (cf. the "minimal predicted area" evaluation measure of Engler et al. (2004)). This provides a theoretical foundation that aids in the selection of a threshold when a binary prediction is required.

Our Maxent implementation has a straightforward graphical user interface (Fig. 1). It also has a command-line interface, allowing it to be run automatically from scripts for batch processing. It is written in Java, so it can be used on all modern computing platforms, and is freely available on the world-wide web at http://www.cs.princeton.edu/~schapire/maxent. The user-specified parameters and their default values (which we used in all runs described below) are: convergence threshold $= 10^{-5}$, maximum iterations $= 1000$, regularization value $\beta = 10^{-4}$, and use of linear, quadratic, product and binary features. The first two parameters are conservative values that allow the algorithm to get close to convergence. The small value of β has minimal effect on the prediction but avoids potential numerical difficulties by keeping λ values from tending to infinity; how to choose the best regularization parameters is a topic of ongoing research (see Dudík et al. (2004)).[1]

---

[1] A later version of the software, Version 1.8.1, was posted on the web site during review of this paper. It allows each β_j to depend on observed variability in the corresponding feature, as described in Dudík et al. (2004). The recommended regularization is now obtained by setting the regularization parameter to "auto", allowing the program to select an amount of regularization that is appropriate for the types of features used and the number of sample localities. The version of the software used in the present study (Version 1.0, also available on the web site) uses the same value β for all β_j.
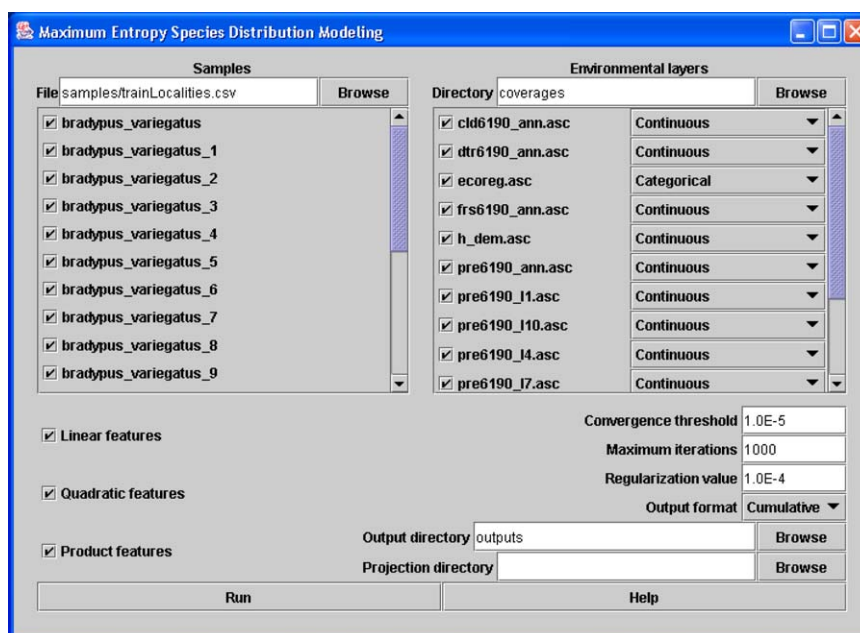
Fig. 1. User interface for the Maxent application (Version 1.0) for modeling species geographic distributions using georeferenced occurrence records and environmental variables. The interface allows for the use of both continuous and categorical environmental data, and linear, quadratic, and product features. See Section 2 for further documentation.

2.3. GARP

In its simplest form, GARP seeks a collection of rules that together produce a binary prediction. Positive rules predict suitable conditions for pixels satisfying some set of environmental conditions; similarly, negative rules predict unsuitable conditions. Rules are favored in the algorithm according to their significance (compared with random prediction) based on a sample of 1250 presence pixels and 1250 background pixels, sampled with replacement. Some pixels may receive no prediction, if no rule in the rule-set applies to them, and some may require resolution of conflicting predictions. A genetic algorithm is used to search heuristically for a good rule-set (Stockwell and Noble, 1992).

There is considerable random variability in GARP predictions, so we implemented the best-subset model selection procedure as follows, similar to Peterson and Shaw (2003) and following the general recommendations of Anderson et al. (2003). First, we generated 100 binary models, with pixels that did not received a prediction interpreted as predicted absence, using GARP version 1.1.3 with default values for its parameters

(0.01 convergence limit, 1000 maximum iterations, and allowing the use of atomic, range, negated range and logit rules). We then eliminated all models with more than 5% intrinsic omission (of training localities). If at most 10 models remained, they then constituted the best subset (this happened 4 out of 44 times, yielding best subsets with 5, 7, 8 and 9 models). In all other cases, we determined the median value of the predicted area of the remaining models, and selected the 10 models whose predicted area was closest to the median. Finally, we combined the best-subset models to make a composite GARP prediction, in which the value of a pixel was equal to the number of best-subset models in which the pixel was predicted present (0–10).

2.4. Data sources

2.4.1. Study species
The brown-throated three-toed sloth *Bradypus variegatus* (Xenarthra: Bradypodidae) is a large arboreal mammal (3–6 kg) that is widely distributed in the Neotropics from Honduras to northern Argentina. It is found primarily in lowland areas but also ranges up to

middle elevations. It has been documented in regions of deciduous forest, evergreen rainforest and montane forest, but is absent from xeric areas and non-forested regions (Anderson and Handley, 2001). Three other species are known in the genus. *B. pygmaeus* is endemic to Isla Escudo on the Caribbean coast of Panama, and two species have geographic distributions restricted to South America: *B. tridactylus* in the Guianan region and *B. torquatus* in the Atlantic forests of Brazil. The latter two species show geographic distributions that likely come into contact (or did historically) with that of *B. variegatus*, but areas of sympatry are apparently minimal.

Microryzomys minutus (Rodentia: Muridae) is a small-bodied rodent (10–20 g) known from middle-to-high elevations of the Andes and associated mountain chains from Venezuela to Bolivia (Carleton and Musser, 1989). It occupies an elevational range of approximately 1000–4000 m and has been recorded primarily in wet montane forests, although sometimes in mesic *páramo* habitats above treeline (in the *páramo*-forest ecotone). A congeneric species, *M. altissimus*,

occupies generally higher elevations in much of this region, but occasionally the two have been found in sympatry. *M. minutus* has not been encountered in lowland regions (below approximately 1000 m). Likewise, it is apparently absent from open *páramo* far from forests, dry *puna* habitat above treeline, and obviously from permanent glaciers on the highest mountain peaks.

These two species hold several characteristics conducive to their use in evaluating the utility of Maxent in modeling species distributions. First of all, they show widespread geographic distributions with clear ecological/environmental patterns. Secondly, they have been the subject of recent taxonomic revisions by specialists. Finally, those revisions provide a reasonable number of georeferenced occurrence localities for each species based on confirmed museum specimens (128 for *B. variegatus*, Anderson and Handley, 2001; 88 for *M. minutus*, Carleton and Musser, 1989; Fig. 2).

2.4.2. Environmental variables

We examine the species' potential distributions in the Neotropics from southeastern Mexico to Argentina
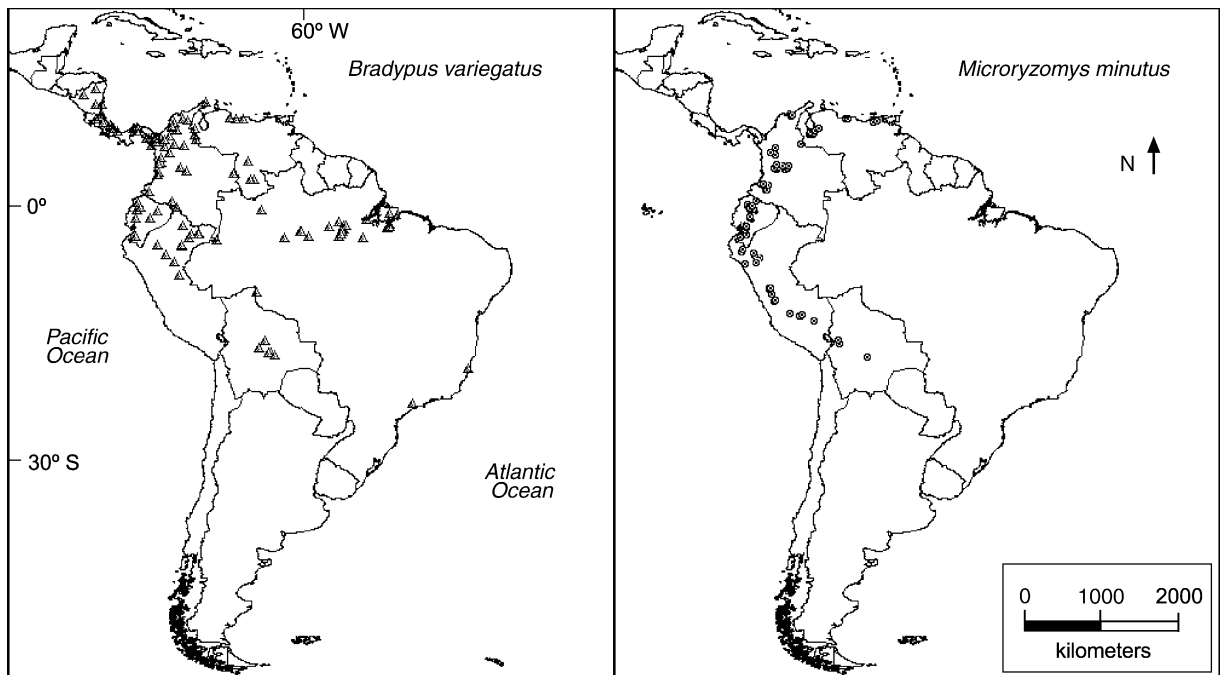
Fig. 2. Occurrence records for *Bradypus variegatus* (triangles; left, 116 records) and *Microryzomys minutus* (circles; right, 88 records) used in this study. Data derive from vouchered museum specimens reported in recent taxonomic revisions (Anderson and Handley, 2001; Carleton and Musser, 1989).

(23.55° N – 56.05° S, 94.8° W – 34.2° W), including the Caribbean from Cuba southward. The environmental variables fall into three categories: climate, elevation and potential vegetation. All variables are recorded at a pixel size of 0.05° by 0.05°, yielding a 1212 × 1592 grid, with 648,658 pixels containing data for all variables.

The climatic variables derive from data provided by the Intergovernmental Panel on Climate Change (IPCC; New et al., 1999). The original variables have a resolution of 0.5° by 0.5°, and were produced using thin-plate spline interpolation based on readings taken at weather stations around the world from 1961 to 1990. They describe mean monthly values of various variables, which we processed to convert to ascii raster grid format, as required by GARP and Maxent. From these monthly data, we also created annual variables by averaging or taking the minimum or maximum as appropriate.

Of the many monthly and annual variables available, we selected the following twelve, based on our assessment that they would likely have relevance for the species being modeled (see also Peterson and Cohoon, 1999): annual cloud cover; annual diurnal temperature range; annual frost frequency; annual vapor pressure; January, April, July, October and annual precipitation; and minimum, maximum and mean annual temperature. We used bilinear interpolation to resample to a pixel size of 0.05° by 0.05°. Although this resampling clearly does not actually increase the resolution of the data, bilinear interpolation is likely more realistic than simply using nearest-neighbor interpolation.

Two other variables were used in addition to the climatic data. An elevation variable was derived from USGS HYDRO1k data (USGS, 2001) by resampling from the original finer resolution (1 km pixels) to 0.05° by 0.05°. Finally, we used a potential vegetation variable, consisting of a partition of Latin America and the Caribbean into "major habitat types", produced as part of a terrestrial conservation assessment (Dinerstein et al., 1995). This variable does not take into account historical (contingent) biogeographic information or human-induced changes, and represents a reconstruction of original vegetation types in the region. We used digital data on 15 major habitat types in a vector coverage (shape file), which we converted to a grid with resolution of 0.05° by 0.05° coincident with the climatic and elevational variables. The digital

data differed slightly from the description and map in Dinerstein et al. (1995) by having 15 rather than 11 major habitat types. The differences arise from the addition of a snow/ice/glaciers/rock category, a tundra category and a water category; deletion of the *restingas* category; splitting of grassland savannas and shrublands into temperate versus tropical/subtropical categories; and splitting of temperate forests into temperate coniferous and temperate broadleaf and mixed forests. The processed climatic variables (at the original resolution), all resampled variables, and the occurrence localities are available at http://www.cs.princeton.edu/~schapire/maxent.

2.5. Model building

For each species, we made 10 random partitions of the occurrence localities. Each partition was created by randomly selecting 70% of the occurrence localities as training data, with the remaining 30% reserved for testing the resulting models. Twelve of the original 128 localities for *B. variegatus* lay in coastal areas or on islands that were missing data for one or more of the environmental variables, and were excluded from this study. Each partition for *B. variegatus* thus held 81 training localities and 35 test localities, and those for *M. minutus* held 61 training localities and 27 test localities.

We made 10 random partitions rather than a single one in order to assess the average behavior of the algorithms, and to allow for statistical testing of observed differences in performance (via Wilcoxon signed-rank tests). In addition, the algorithms were also run on the full set of occurrence localities, taking advantage of all available data to provide best estimates of the species' potential distributions for visual interpretation.

The algorithms (Maxent and GARP) were run with two suites of environmental variables: first with only climatic and elevational data, and then with those variables plus potential vegetation. The reasons for treating potential vegetation separately are three-fold: (1) climatic and elevational data are readily available for the whole world (whereas potential vegetation is not), and we wished to determine whether good models can be created using uniformly available data. (2) The potential vegetation coverage is rather subjective, whereas the others are objectively produced from measured data. (3) Potential vegetation is the only categorical variable, and the potential existed for the algorithms

to respond differently to categorical versus continuous data.

2.6. Model evaluation

The first step in evaluating the models produced by the two algorithms was to verify that both performed significantly better than random. For this purpose, we first used a (threshold-dependent) binomial test based on omission and predicted area. However, it does not allow for comparisons between algorithms, as the significance of the test is highly dependent on predicted area. Indeed, comparison of the algorithms is made difficult by the fact that neither gives a binary prediction. Hence, we also used two comparative statistical tests that employ very different means to overcome this complication. First, we employed a new threshold-dependent method of model evaluation, which we term the "equalized predicted area" test, whose purpose is to answer the following question: at the commonly used thresholds representing the extremes of the GARP prediction, how does Maxent compare? Second, we used (threshold-independent) receiver operating characteristic (ROC) analysis, which characterizes the performance of a model at all possible thresholds by a single number, the area under the curve (AUC), which may be then compared between algorithms.

2.6.1. Threshold-dependent evaluation

After applying a threshold, model performance can be investigated using the *extrinsic omission rate*, which is the fraction of the test localities that fall into pixels not predicted as suitable for the species, and the *proportional predicted area*, which is the fraction of all the pixels that are predicted as suitable for the species. A low omission rate is a necessary (but not sufficient) condition for a good model (Anderson et al., 2003). In contrast, it might be necessary to predict a large proportional area to model the species' potential distribution adequately.

A one-tailed binomial test can be used to determine whether a model predicts the test localities significantly better than random (Anderson et al., 2002). Say there are t test localities, the omission rate is r, and the proportional predicted area is a. The null hypothesis states that the model is no better than one randomly selected from the set of all models with proportional predicted area a. It is tested using a one-tailed binomial test to determine

the probability of having at least $t(1 − r)$ successes out of t trials, each with probability a. Although the probabilities for such tests are often approximated using a χ^2 or z test (for large sample sizes), we calculated exact probabilities for the binomial test using Minitab (1998).

The binomial test requires that thresholds be used, in order to convert continuous Maxent and discrete GARP predictions into binary predictions delimiting the suitable versus unsuitable areas for the species. A good general rule for determining an appropriate threshold would depend at least on the following factors: the predicted values assigned to the training localities, the number of training localities and the context in which the prediction is to be used. Nevertheless, for each run of each algorithm, we simply used the minimum predicted value assigned to any of the training localities as the threshold. However, for four of the twenty GARP runs, such a threshold would cause the whole study area to be predicted (as some training localities fell in pixels not predicted by any of the best-subset models). In those cases, we used the smallest non-zero predicted value among the training localities.

Because this omission test is highly sensitive to the proportional predicted area (Anderson et al., 2003), it cannot be used to compare model performance between two algorithms directly. Hence, we propose an "equalized predicted area" test, which chooses thresholds so that the two binary models have the same predicted area, allowing direct comparison of omission rates. Here, composite GARP models have little flexibility in the choice of threshold. On the other hand, Maxent predictions, being continuous, can be thresholded to obtain any desired predicted area. So, we set a threshold for each Maxent prediction to give the same predicted area as the corresponding GARP prediction. A two-tailed Wilcoxon signed-rank test (a non-parametric equivalent of a paired *t*-test) can then be used to determine whether the observed difference in omission rates between the two algorithms at the given predicted area is statistically significant. We used this test to compare Maxent predictions with two thresholds of the composite GARP predictions, namely 1 (any best-subset model) and 10 (all best-subset models; see Anderson and Martínez-Meyer, 2004). These are natural thresholds for GARP that are frequently used in practice, so for reasons of conciseness, we do not consider intermediate thresholds. For some data partitions for *B. var-*

iegatus, the maximum value of the composite GARP model was less than 10 (because fewer than 10 GARP models met the best-subset criteria), in which case we used the maximum predicted value instead of 10.

The thresholds and resulting predicted areas used above are not necessarily optimal for either algorithm. Rather, they were chosen to facilitate statistical analysis of the algorithms. Note that we are not suggesting that GARP should or need be used in general to select a threshold for Maxent predictions when binary predictions are desired. Rather, we took advantage of the flexibility of Maxent's continuous outputs to allow direct comparisons of omission rates between it and GARP. Determining optimal thresholds for Maxent models remains a topic of future research. In practice, thresholds would currently be chosen by hand, since no general-purpose thresholding rule has been developed yet for either algorithm (but see Section 2.2 for theoretical expectations for Maxent).

2.6.2. Threshold-independent evaluation

A second common approach compares model performance using receiver operating characteristic (ROC) curves. ROC analysis was developed in signal processing and is widely used in clinical medicine (Hanley and McNeil, 1982, 1983; Zweig and Campbell, 1993). The main advantage of ROC analysis is that area under the ROC curve (AUC) provides a single measure of model performance, independent of any particular choice of threshold. ROC analysis has recently been applied to a variety of classification problems in machine learning (for example Provost and Fawcett, 1997) and in the evaluation of models of species distributions (Elith, 2002; Fielding and Bell, 1997).

Here we will first describe ROC curves in general terms, following Fawcett (2003), before demonstrating how they apply to presence-only modeling. Consider a classification problem, where each instance is either positive or negative. A classifier assigns a real value to each instance, to which a threshold may be applied to predict class membership; for clarity we use labels {Y, N} for the class predictions. The *sensitivity* of a classifier for a particular threshold is the fraction of all positive instances that are classified Y, while *specificity* is the fraction of all negative instances that are classified N. Sensitivity is also known as the true positive rate, and represents absence of omission error. The quantity

1–specificity is also known as the false positive rate, and represents commission error.

The ROC curve is obtained by plotting sensitivity on the *y* axis and 1–specificity on the *x* axis for all possible thresholds. For a continuous prediction, the ROC curve typically contains one point for each test instance, while for a discrete prediction, there will typically be one point for each of the different predicted values, in addition to the origin. The area under the curve (AUC) is usually determined by connecting the points with straight lines; this is called the trapezoid method (as opposed to parametric methods, which fit a curve to the points). The AUC has an intuitive interpretation, namely the probability that a random positive instance and a random negative instance are correctly ordered by the classifier. This interpretation indicates that the AUC is not sensitive to the relative numbers of positive and negative instances in the test data set.

When only presence data are available, it would appear that ROC curves are inapplicable, since without absences, there seems to be no source of negative instances with which to measure specificity (see Section 1.1, and the discussion of real and apparent commission error in Anderson et al. (2003) and Karl et al. (2002)). However, we can avoid this problem by considering a different classification problem, namely, the task of distinguishing presence from random, rather than presence from absence. More formally, for each pixel x in the study area, we define a negative instance x_{random}. Similarly, for each pixel x that is included in the species' true geographic distribution, we define a positive instance x_{presence}. A species distribution model can then make predictions for the pixels corresponding to these instances, without seeing the labels `random` or `presence`. Thus, we can make predictions for both a sample of positive instances (the presence localities) and a sample of negative instances (background pixels chosen uniformly at random, or according to another background distribution as described in Section 1.3). Together these are sufficient to define an ROC curve, which can then be analyzed with all the standard statistical methods of ROC analysis. This process can be interpreted as using pseudo-absence in place of absence in the ROC analysis, as is done in Wiley et al. (2003). However, we believe that the observation that the statistical methods of ROC analysis can be applied without prejudice to presence/random data is new.

The above treatment differs from the use of ROC analysis on presence/absence data in one important respect: with presence-only data, the maximum achievable AUC is less than 1 (Wiley et al., 2003). If the species' distribution covers a fraction a of the pixels, then the maximum achievable AUC can be shown to be exactly $1 - a/2$. Unfortunately, we typically do not know the value of a, so we cannot say how close to optimal a given AUC value is. Nevertheless, we can still use standard methods to determine statistical significance of the AUC, and to distinguish between the predictive power of different classifiers. We note that random prediction still corresponds to an AUC of 0.5.

We used AccuROC Version 2.5 (Vida, 1993) for the ROC analyses. AccuROC uses the trapezoid method, as described above. To test if a prediction is significantly better than random, AccuROC uses a ties-corrected Mann–Whitney-U statistic, which it approximates using a z-statistic. It uses a non-parametric test (DeLong et al., 1988) to determine whether one prediction is significantly better than another when using correlated samples (i.e., with both predictions evaluated on the same test instances), and reports the result as a χ^2 statistic and corresponding p value. For each ROC analysis, we used all the test localities for the species as presence instances, and a sample of 10,000 pixels drawn randomly from the study region as random instances.

3. Results

3.1. Quantitative results

3.1.1. Threshold-dependent omission tests

Both algorithms consistently produced predictions that were better than random. Using the simple threshold rule (Section 2.6.1), the binomial omission test was highly significant ($p < 0.001$, one-tailed) for both algorithms on all data partitions for each species (see Table 1 for details on runs with the climatic and elevational variables; results on the variable suite including potential vegetation were similar). For Maxent, the thresholds determined by the simple threshold rule ranged from 0.022 to 2.564 for *B. variegatus* and 0.543 to 3.822 for *M. minutus*. For GARP, the thresholds ranged from 1 to 7 for *B. variegatus* and 2 to 10 for *M. minutus*. In addition to statistical

significance, omission rates were consistently low or zero, never exceeding 17% (Table 1).

The results of the equalized predicted area test differed between the species (Tables 2 and 3). For *B. variegatus*, the omission rates of the two algorithms were lower for Maxent in 16 cases, equal in 15 cases, and lower for GARP in 9 cases. However, two-tailed Wilcoxon signed-rank tests did not reveal a significant difference in median omission rates for either threshold or either variable suite ($p = 0.178$ and 0.314 for thresholds of 1 and 10, respectively, with climatic and elevational variables; $p = 0.371$ and 0.155 for thresholds of 1 and 10, respectively, with addition of the potential vegetation variable).

Maxent almost always had equal or lower omission than GARP for *M. minutus* (19 out of 20 models). The difference in median omission rates was significant at both thresholds on runs with climatic and elevational variables ($p = 0.036$ and $p = 0.014$ for thresholds of 1 and 10, respectively; two-tailed Wilcoxon signed-rank test). When the potential vegetation variable was added, the difference in median omission rates was highly significant for a threshold of 10, but not for a threshold of 1 ($p = 0.009$ and 0.345, respectively), largely because Maxent had greater omission than before on data partition 2, discussed below (Section 4.3).

3.1.2. Threshold-independent tests

For all partitions of the occurrence data, the AUC values (calculated on extrinsic test data) were highly statistically significant for both algorithms and variable suites ($p < 0.0001$), again indicating better-than-random predictions. The Maxent AUC was significantly greater than that of GARP ($p < 0.05$; two-tailed non-parametric test of DeLong et al., 1988; see Methods) in all data partitions except *B. variegatus*-4 and *B. variegatus*-8 for models using the climatic and elevational variables, and *B. variegatus*-8 and *M. minutus*-2 when potential vegetation was added (Table 4).

Addition of the potential vegetation variable should increase the AUC, since there is more information available to the classifier. This was true in general for Maxent and in some cases for GARP (Table 4). For Maxent on *B. variegatus*, the overall increase in median AUC approached significance ($p = 0.093$, one-tailed Wilcoxon signed rank test). However, for GARP the test was not significant ($p = 0.949$); indeed, the AUC generally decreased. For *M. minutus*, the AUC

Table 1
Results of the threshold-dependent binomial tests of omission

Data partition	Maxent		GARP	
	Area	Omission rate	Area	Omission rate
Bradypus variegatus-1	0.51	0.03	0.41	0.11
B. variegatus-2	0.66	0	0.56	0.06
B. variegatus-3	0.80	0	0.61	0.03
B. variegatus-4	0.42	0.17	0.51	0
B. variegatus-5	0.75	0.03	0.57	0.06
B. variegatus-6	0.62	0	0.54	0
B. variegatus-7	0.59	0	0.53	0
B. variegatus-8	0.59	0.06	0.62	0
B. variegatus-9	0.69	0	0.66	0
B. variegatus-10	0.62	0.06	0.44	0.06
Average	0.626	0.034	0.545	0.031
Microryzomys minutus-1	0.03	0.11	0.06	0.15
M. minutus-2	0.04	0.11	0.06	0.15
M. minutus-3	0.03	0.11	0.07	0.15
M. minutus-4	0.04	0.04	0.08	0.04
M. minutus-5	0.03	0.04	0.06	0.15
M. minutus-6	0.04	0.15	0.06	0.11
M. minutus-7	0.05	0	0.09	0.07
M. minutus-8	0.04	0.04	0.10	0
M. minutus-9	0.03	0.07	0.10	0
M. minutus-10	0.03	0.11	0.08	0.07
Average	0.035	0.078	0.075	0.089

Area (proportion of the study area predicted) and extrinsic omission rate (proportion of the test localities falling outside the prediction) are given for each of 10 random data partitions for Maxent and GARP. For both *B. variegatus* and *M. minutus*, the binomial test was highly significant for all partitions ($p < 0.001$, one-tailed). Models were derived using the climatic and elevational variables for each random partition of occurrence records, and area and omission rates were calculated using simple threshold rules based on the training localities (see Section 2). The results for models made with the addition of the potential vegetation variable were similar but are not shown here (see Section 3). The omission rates should not be compared between algorithms, as they are strongly affected by differences in predicted area. The simple threshold rule used here for Maxent is not recommended for general use in practice; in this case, it gives too high a threshold for Maxent on *B. variegatus*-4, causing a high omission rate, and too low a threshold on *B. variegatus*-3, resulting in too much predicted area.

usually increased for both Maxent and GARP, with results significant or nearly so for both ($p = 0.051$ and 0.033, respectively, although performance was poorer for Maxent on data partition 2; see Section 4.3). While the differences in AUC values are very small, the changes may still be meaningful biologically. For example, the largest visual effect of adding potential vegetation for Maxent was to (correctly) exclude some non-forested areas from the prediction for *B. variegatus* (Section 3.2.2). However, because of the small geographic extent of those areas, the effect on AUC values was small.

The ROC curves for the two algorithms showed distinct patterns, evident in the curves for the first random data partition for each species, for models made using climatic and elevational variables (Fig. 3). In the case

of *M. minutus*, the performance of Maxent was better across the entire spectrum: for any given omission rate, Maxent achieved that rate with a lower false positive rate (1–specificity, which is almost identical to proportional predicted area, see Section 2). The results with *B. variegatus* were more complex. There is a point where the ROC curves for the two algorithms intersect, corresponding to a sensitivity of 0.83 (omission rate of 0.17) and a false positive rate of 0.27. At that point, therefore, the performance of the two algorithms was the same. A small component of the higher AUC for Maxent was due to the lower omission rate it achieved to the right of that point. However, most of Maxent's higher AUC occurred to the left of that point, where many test localities fell in small areas very strongly predicted by Maxent. In contrast, GARP did not differ-

Table 2
Results of the equalized predicted area tests of omission for *B. variegatus* and *M. minutus* produced with Maxent and GARP using the climatic and elevational variables

Data partition	GARP threshold = 1			GARP threshold = 10		
	Area	Maxent omission	GARP omission	Area	Maxent omission	GARP omission
B. variegatus-1	0.59	0.03	0.03	0.27	0.17	0.17
B. variegatus-2	0.56	0.03	0.06	0.34	0.11	0.17
B. variegatus-3	0.61	0.06	0.03	0.33	0.09	0.17
B. variegatus-4	0.63	0.14	0	0.40	0.17	0.06
B. variegatus-5	0.67	0.03	0	0.36	0.11	0.26
B. variegatus-6	0.69	0	0	0.29	0.14	0.11
B. variegatus-7	0.74	0	0	0.31	0.03	0.14
B. variegatus-8	0.69	0	0	0.33	0.17	0.11
B. variegatus-9	0.72	0	0	0.36	0.06	0.11
B. variegatus-10	0.61	0.06	0.03	0.34	0.14	0.17
Average	0.652	0.034	0.014	0.333	0.120	0.149
M. minutus-1	0.12	0	0.07	0.06	0.04	0.15
M. minutus-2	0.10	0	0.07	0.06	0.04	0.15
M. minutus-3	0.16	0	0.04	0.07	0.07	0.15
M. minutus-4	0.17	0	0.04	0.08	0.04	0.04
M. minutus-5	0.12	0	0.07	0.06	0	0.15
M. minutus-6	0.12	0	0.04	0.06	0.07	0.11
M. minutus-7	0.16	0	0	0.09	0	0.07
M. minutus-8	0.17	0	0	0.09	0	0
M. minutus-9	0.17	0	0	0.09	0	0.04
M. minutus-10	0.18	0	0	0.08	0	0.07
Average	0.146	0	0.033	0.073	0.026	0.093

Area (proportion of the study area predicted by GARP with the indicated threshold) and extrinsic omission rate (proportion of test localities falling outside the prediction) for each algorithm are given for each random partition of occurrence records under two threshold scenarios. Thresholds were set for the extremes of the GARP predictions: any GARP model predicting presence (GARP threshold = 1) and all 10 GARP models predicting presence (GARP threshold = 10). To allow for direct comparison of omission rates between the algorithms, thresholds were then set for each Maxent model to yield a binary prediction with the same area as the corresponding GARP prediction.

entiate environmental quality to the left of that point, as all pixels there were predicted by all 10 of the best-subset models. Results for other data partitions were roughly similar (not shown).

3.2. Visual interpretation

The output format differs dramatically between Maxent and GARP, so care must be taken when making comparisons between them. Maxent produces a continuous prediction with values ranging from 0 to 100, whereas GARP, as used here, yields a discrete composite prediction with integer values from 0 to 10. Visual inspection of the Maxent predictions for both species indicated that a low threshold was appropriate, and in general terms, pixels with predicted values of at least 1 may be interpreted as a reasonable approximation of the species' potential distribution. This is in concor-

dance with the thresholds obtained in Section 3.1.1, and the theoretical expectation that the omission rate for a thresholded cumulative prediction will be approximately equal to the threshold value (see Section 2.2). For GARP, visual inspection suggested a higher threshold in the range 5–10 was appropriate for approximating the species' potential distribution. In the following sections, we interpret the models under this framework.

3.2.1. Models derived from climatic and elevational variables

When using the full set of occurrence localities for each species, the two algorithms produced broadly similar predictions for the potential geographic distribution of *B. variegatus* (Fig. 4). For this species, both algorithms indicated suitable conditions throughout most of lowland Central America, wet lowland areas of northwestern South America, most of the Amazon

Table 3

Results of the equalized predicted area tests of omission for *B. variegatus* and *M. minutus* produced with Maxent and GARP using the climatic, elevational and potential vegetation variables

Data partition	GARP threshold = 1			GARP threshold = 10		
	Area	Maxent omission	GARP omission	Area	Maxent omission	GARP omission
B. variegatus-1	0.57	0.03	0.03	0.28	0.20	0.23
B. variegatus-2	0.58	0	0.06	0.29	0.11	0.29
B. variegatus-3	0.67	0	0.03	0.33	0.14	0.11
B. variegatus-4	0.67	0	0	0.42	0.06	0.11
B. variegatus-5	0.67	0.03	0.03	0.36	0.14	0.17
B. variegatus-6	0.71	0	0	0.28	0.17	0.17
B. variegatus-7	0.74	0	0	0.33	0.06	0.20
B. variegatus-8	0.67	0	0	0.34	0.20	0.17
B. variegatus-9	0.78	0	0	0.39	0.03	0.06
B. variegatus-10	0.67	0	0	0.36	0.14	0.17
Average	0.672	0.006	0.014	0.337	0.126	0.169
M. minutus-1	0.12	0	0.04	0.06	0.04	0.15
M. minutus-2	0.11	0.11	0.04	0.06	0.15	0.19
M. minutus-3	0.13	0	0.04	0.07	0.04	0.15
M. minutus-4	0.15	0	0.04	0.08	0.04	0.04
M. minutus-5	0.12	0	0.07	0.06	0	0.15
M. minutus-6	0.14	0	0	0.05	0.04	0.11
M. minutus-7	0.16	0	0.04	0.08	0	0.07
M. minutus-8	0.16	0	0	0.08	0	0.04
M. minutus-9	0.16	0	0	0.08	0	0.07
M. minutus-10	0.17	0	0	0.07	0	0.04
Average	0.142	0.011	0.026	0.070	0.030	0.100

Area (proportion of the study area predicted by GARP with the indicated threshold) and extrinsic omission rate (proportion of test localities falling outside the prediction) for each algorithm are given for each random partition of occurrence records under two threshold scenarios. Thresholds were set for the extremes of the GARP predictions: any GARP model predicting presence (GARP threshold = 1) and all 10 GARP models predicting presence (GARP threshold = 10). To allow for direct comparison of omission rates between the algorithms, thresholds were then set for each Maxent model to yield a binary prediction with the same area as the corresponding GARP prediction.

basin, large areas of Atlantic forests in southeastern Brazil, and most large Caribbean islands in the study area. The species was generally predicted absent from high montane areas, temperate areas in southern South America, and non-forested areas of central Brazil (e.g., *cerrado*). The algorithms differed in their predictions for non-forested savannas in northern South America. The composite GARP model indicated the species' potential presence there, but Maxent excluded some non-forested savannas in Venezuela (*llanos*) and the Guianas.

In contrast, the algorithms yielded quite different predictions for *M. minutus* (Fig. 4). Maxent indicated suitable conditions for the species in the northern and central Andes (and associated mountain chains) from Bolivia and northern Chile to northern Colombia and Venezuela. It also included highland areas in Jamaica, the Dominican Republic and Haiti, as well as very

small highland areas in Brazil, southeastern Mexico, Costa Rica and Panama. In contrast, GARP predicted a much more extensive potential distribution for the species. In addition to a broad highland prediction in the northern and central Andes and the Caribbean, the composite GARP prediction also included areas of the southern Andes as well as extensive highland regions in Mesoamerica, the Guianan-shield region and southeastern Brazil. The prediction in the Brazilian highlands extended into adjacent lowland areas of that country as well as into Uruguay and northern Argentina.

3.2.2. Addition of potential vegetation variable

The two algorithms responded differently to the inclusion of the potential vegetation variable (Fig. 5). The Maxent prediction with potential vegetation for *B. variegatus* was generally similar to the original one,

Table 4

Results of threshold-independent receiver operating characteristic (ROC) analyses for *B. variegatus* and *M. minutus* produced with Maxent and GARP using the climatic and elevational variables (left) and climatic, elevational and potential vegetation variables (right)

Data partition	Without potential vegetation			With potential vegetation		
	Maxent AUC	GARP AUC	*p*	Maxent AUC	GARP AUC	*p*
B. variegatus-1	0.889	0.807	<0.01	0.879	0.793	<0.01
B. variegatus-2	0.892	0.765	<0.01	0.899	0.769	<0.01
B. variegatus-3	0.872	0.779	0.01	0.887	0.790	<0.01
B. variegatus-4	0.819	0.789	0.51	0.858	0.757	<0.01
B. variegatus-5	0.868	0.740	<0.01	0.885	0.753	<0.01
B. variegatus-6	0.881	0.818	<0.01	0.868	0.812	0.03
B. variegatus-7	0.902	0.812	<0.01	0.919	0.784	<0.01
B. variegatus-8	0.839	0.807	0.34	0.829	0.786	0.13
B. variegatus-9	0.903	0.794	<0.01	0.897	0.784	<0.01
B. variegatus-10	0.866	0.779	0.01	0.879	0.769	<0.01
Average	0.873	0.789		0.880	0.780	
M. minutus-1	0.985	0.926	0.01	0.986	0.946	0.02
M. minutus-2	0.987	0.931	0.02	0.932	0.943	0.75
M. minutus-3	0.985	0.938	<0.01	0.987	0.939	<0.01
M. minutus-4	0.983	0.938	<0.01	0.984	0.941	<0.01
M. minutus-5	0.988	0.926	0.02	0.990	0.926	0.01
M. minutus-6	0.983	0.947	0.05	0.986	0.966	<0.01
M. minutus-7	0.989	0.950	<0.01	0.988	0.936	<0.01
M. minutus-8	0.988	0.954	<0.01	0.989	0.956	<0.01
M. minutus-9	0.989	0.952	<0.01	0.990	0.955	<0.01
M. minutus-10	0.985	0.955	<0.01	0.987	0.961	<0.01
Average	0.986	0.942		0.982	0.947	

For each random partition of occurrence records, the area under the ROC curve (AUC) is given for Maxent and GARP, as well as the probability of the observed difference in the AUC values between the two algorithms (under a null hypothesis that the true AUCs are equal). All AUC values for both algorithms were significantly better than a random prediction ($p < 0.0001$; individual p values not shown). AUC values are given to three decimal places to reveal small changes under addition of the potential vegetation coverage.

but now indicated unsuitable conditions for the species in the *llanos* of Colombia and Venezuela and in other non-forested areas in Bolivia and Brazil. On the contrary, the composite GARP prediction with potential vegetation included was very similar to the original prediction, still indicating suitable environmental conditions for the species in non-forested areas of Colombia, Venezuela, Guyana, Brazil, Paraguay and Bolivia.

Addition of the potential vegetation variable changed the Maxent and GARP predictions for *M. minutus* only minimally. The Maxent prediction with potential vegetation differed principally by a sharp reduction in the area predicted for the species along the western slopes of the Andes in central and southern Peru and in northern Chile. The composite GARP prediction with potential vegetation differed from the original one mainly by indicating a smaller area of suitable environmental conditions for the species in central Chile and in central-eastern Argentina.

4. Discussion and conclusions

4.1. Statistical tests

Both algorithms consistently performed significantly better than random, and Maxent frequently achieved better results than GARP. Threshold-dependent binomial tests (Table 1) showed low omission of test localities and significant predictions for both algorithms across the board. The equalized predicted area test generally indicated better performance for Maxent on *M. minutus*, but the test did not detect a significant difference between the two algorithms for *B. variegatus* (Tables 2 and 3). Threshold-independent ROC analysis also showed significantly better-than-random performance for both algorithms. The area under the ROC curve (AUC) was significantly higher for Maxent on almost all data partitions for both species (Table 4). Use of the categorical potential vegetation
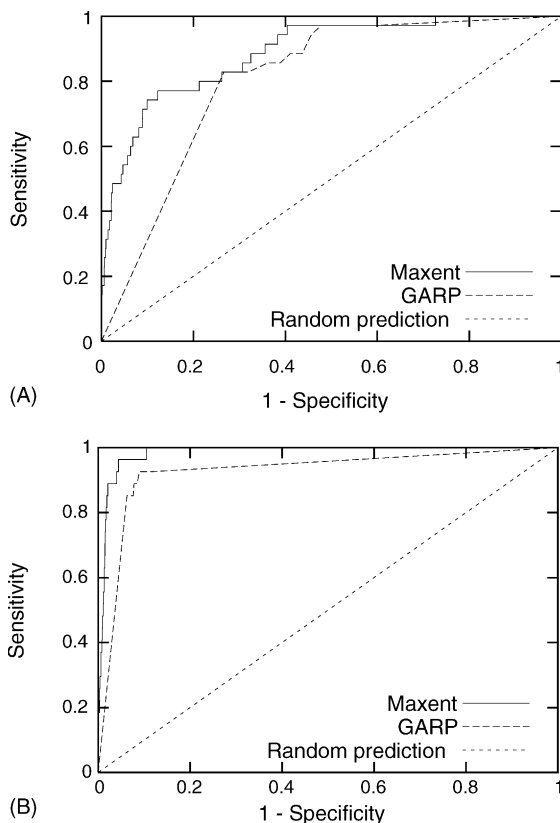
Fig. 3. Extrinsic receiver operating characteristic (ROC) curves for Maxent and GARP on the first random partition of occurrence records of *B. variegatus* (left) and *M. minutus* (right). Models were produced using the climatic and elevational variables. Sensitivity equals the proportion of test localities correctly predicted present (1–extrinsic omission rate). The quantity (1–specificity) equals the proportion of all map pixels predicted to have suitable conditions for the species. Note that both algorithms perform much better than random, and that Maxent is generally superior to GARP; see Table 4 for results of statistical analyses. *B. variegatus* is a wider-ranging species than *M. minutus*, so it has a smaller maximum achievable AUC in these ROC analyses performed without true absence data (see Section 2.6.2). The curves therefore do not necessarily imply that the algorithms are performing better on *M. minutus*.

4.2. Biological interpretations

Both algorithms produced reasonable predictions of the potential distribution for *B. variegatus*. The areas predicted by 5–10 GARP models were similar geographically to those areas predicted with a value of at least 1 (out of 100) for Maxent. Although much research addressing the issue of operationally determining an optimal threshold remains for both algorithms, these thresholds produce good maps of the species' potential distributions (areas of suitable environmental conditions). In particular, the models perform far superior to the shaded outline maps available in standard field guides, (e.g., Eisenberg and Redford, 1999; Emmons, 1997), and in digital compilations of species ranges designed for use in conservation biology and macroecological studies (Patterson et al., 2003). Most strikingly, the models correctly indicate an expansive region of unsuitable environmental conditions for *B. variegatus* in the non-forested *cerrado* of Brazil, whereas the shaded outline maps indicate continuous distribution for the species from Amazonian forests to coastal Atlantic forests. Although GARP has the capacity to consider categorical variables, the inclusion of the potential vegetation variable did not rectify the deficiencies seen in the original composite GARP prediction for *B. variegatus*. In contrast, Maxent successfully integrated this additional information. This is most evident in close-up images in Fig. 4.2, where GARP (incorrectly) predicted suitable conditions for the species in the non-forested *llanos* of Colombia and Venezuela.

In contrast to the generally similar predictions for *B. variegatus*, different deficiencies were evident in the predictions produced by the two algorithms for *M. minutus*. Maxent produced an impressive prediction within the species' known range. The Maxent prediction lay almost entirely within the Andes. However, wet montane forests also exist in Mesoamerican, Guianan and Brazilian highlands. Those areas likely contain conditions suitable for the species, but it apparently has not been able to colonize them due to geographic barriers. We investigated possible reasons for Maxent's low prediction in these areas, by examining environmental characteristics of six classical highland sites that are well sampled for small mammals: Monteverde and Cerro de la Muerte in Costa Rica, Auyan-tepui and Mount Duida in southern Venezuela, and Itatiaia and

variable (in addition to the continuous climatic and elevational variables) generally improved performance for both algorithms on *M. minutus* and for Maxent on *B. variegatus*, but the changes had limited statistical significance, likely due to the small amount of data.
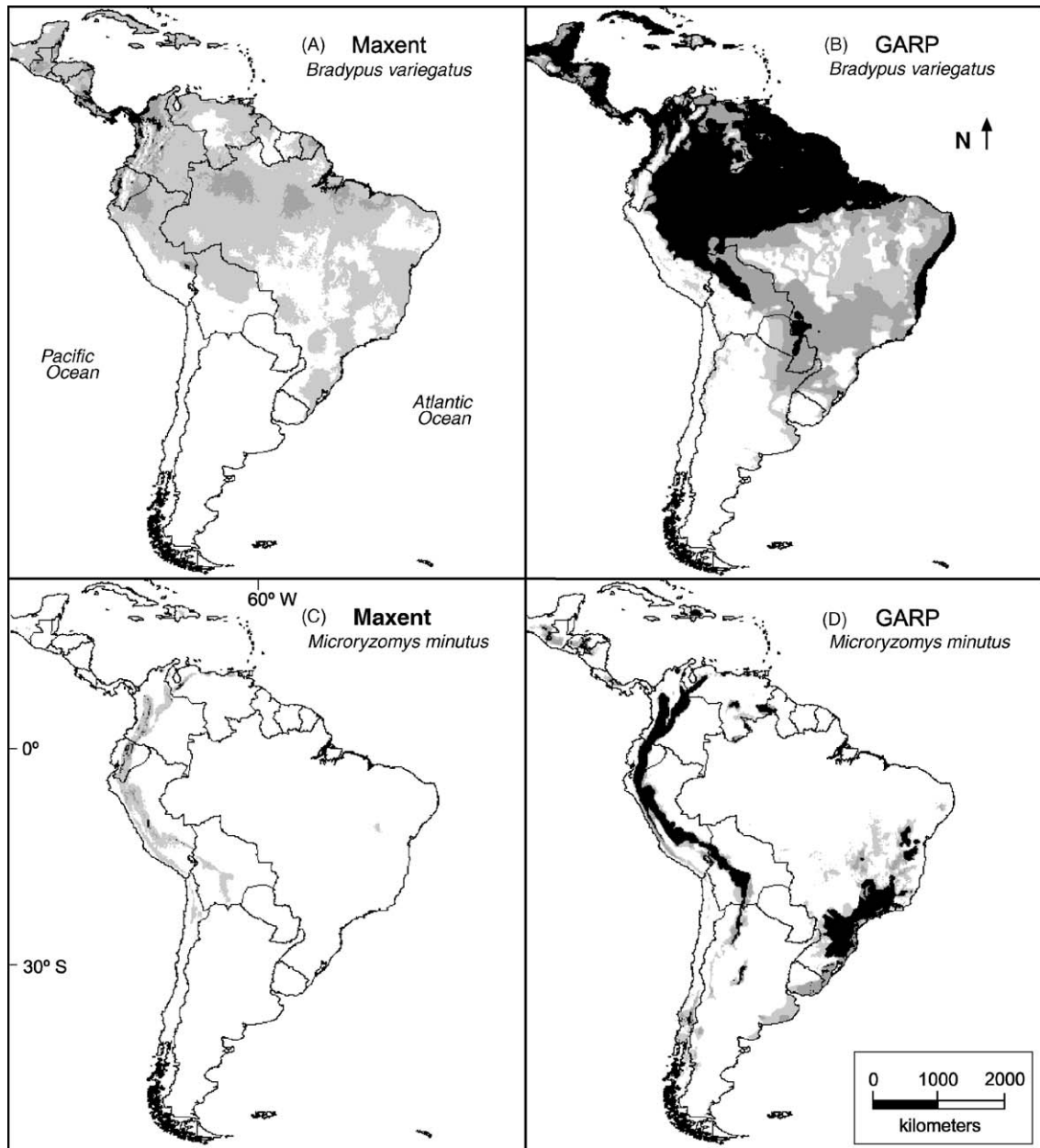
Fig. 4. Predicted potential geographic distributions for *B. variegatus* (top) and *M. minutus* (bottom) made using all occurrence records and the climatic and elevational variables. Results are given for Maxent (left) and GARP (right). Four colors are used to indicate the strength of the prediction for each individual map pixel. Maxent produces a continuous prediction with values ranging from 0 to 100; the values are depicted here using white = [0,1); pale grey = [1,34); dark grey = [34,66); black = [66,100]. The best-subsets selection procedure employed here for GARP yields a discrete prediction with integer values from 0 to 10, depicted here using white = 0; pale grey = 1–4; dark grey = 5–9; black = 10. The strength of the predictions thus cannot be compared directly. All areas with a Maxent prediction of 1 or greater likely represent suitable environmental conditions for the species; in contrast, areas with a GARP prediction of 5–10 probably indicate suitable conditions (see Section 3.2).
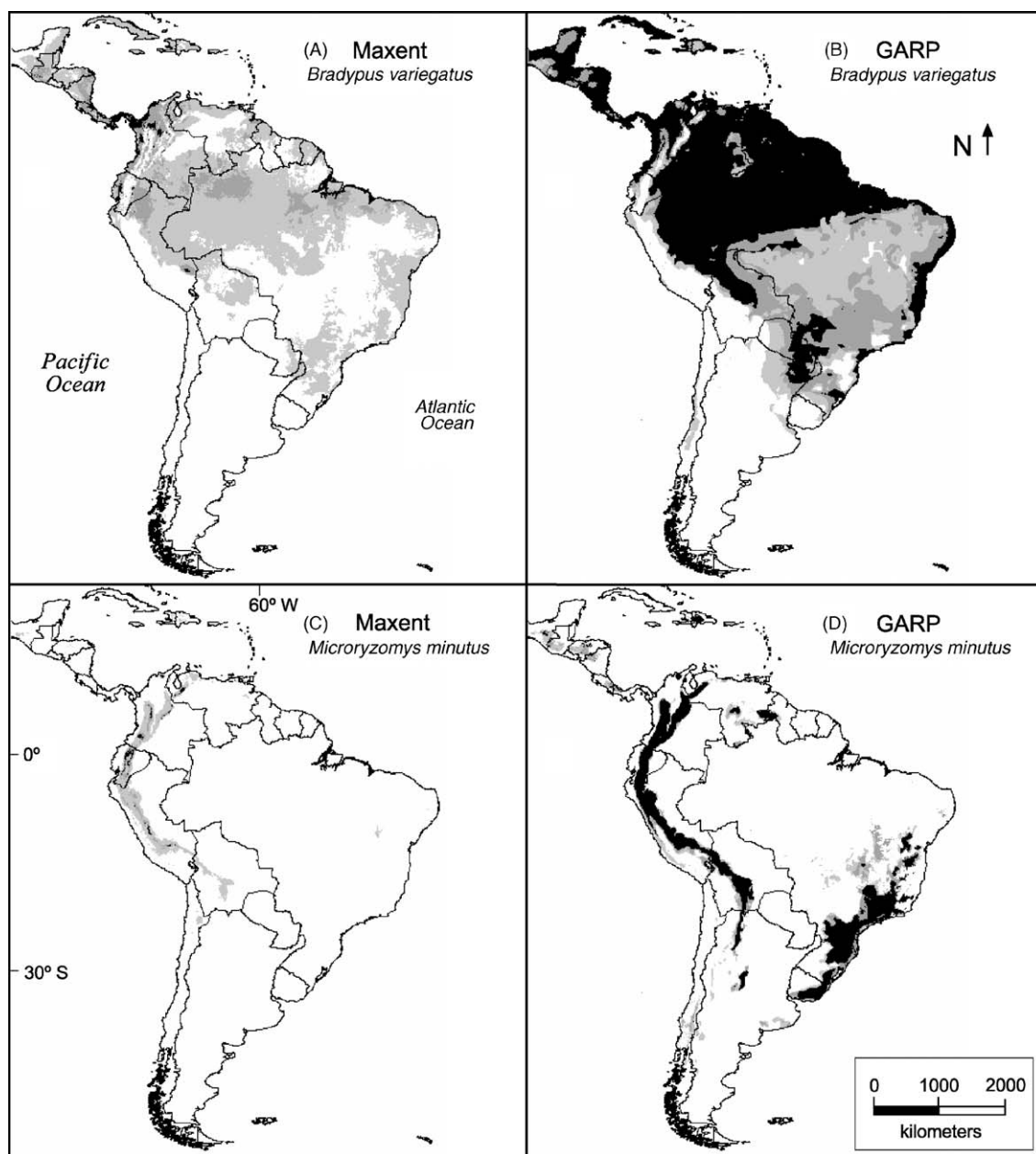
Fig. 5. Predicted potential geographic distributions for *B. variegatus* (top) and *M. minutus* (bottom) made using all occurrence records and climatic, elevational and potential vegetation variables. Results are given for Maxent (left) and GARP (right). Four colors are used to indicate the strength of the prediction for each map pixel. Maxent produces a continuous prediction with values ranging from 0 to 100; the values are depicted here using white = [0,1); pale grey = [1,34); dark grey = [34,66); black = [66,100]. The best-subsets selection procedure employed here for GARP yields a discrete composite prediction with integer values from 0 to 10, depicted here using white = 0; pale grey = 1–4; dark grey = 5–9; black = 10. The strength of the predictions thus cannot be compared directly. All areas with a Maxent prediction of 1 or greater likely represent suitable environmental conditions for the species; in contrast, areas with a GARP prediction of 5–10 probably indicate suitable conditions (see Section 3.2).

Caparaó in Brazil (Gouvêa and de Ávila Pires, 1999; Hershkovitz, 1998; McPherson, 1985; Paynter, 1982; Tate, 1939). For the first four sites, July precipitation values were at least 5 standard deviations higher than the mean of the *M. minutus* occurrence localities. In addition, the annual maximum temperature at those sites was 1.37–3.23 standard deviations higher than the mean of the occurrence localities. In contrast, the Brazilian sites had July precipitation within the same range as the occurrence localities, but the January precipitation was much higher for both (by 3.12 and 1.84 standard deviations, respectively), and maximum temperature was much higher for Caparaó (by 1.95 standard deviations). Thus, Maxent's behavior given the data provided is correct and reasonable. However, despite the differences in some environmental variables, the forests in the six sites are probably functionally similar to those inhabited by *M. minutus*. This situation highlights the difficulty of extrapolating from a species' realized distribution, and emphasizes that the variables used should be chosen with care. For *M. minutus*, better extrapolation might be achieved using derived climatic parameters that are more relevant for the species, for example, precipitation of wettest month (Busby, 1986), rather than values for specific months (see Section 1.1). Quite the opposite to the Maxent predictions, extensive areas of potential distribution indicated in Mesoamerican, Guianan and Brazilian highland regions by GARP surely overestimate the extent of suitable environmental conditions for the species there. In particular, the vast majority of the pixels predicted by all 10 models in southeastern Brazil lie below 1000 m, where the species' presence is quite unlikely.

4.3. Spatial context of errors

The performance of Maxent on *M. minutus* when the potential vegetation variable was used warrants some discussion. The AUC for the second random data partition was notably lower than for the other partitions, and for the model run on the same partition without potential vegetation. Most of the occurrence localities for the species are contained in the "tropical and subtropical moist broadleaf forest" and "tropical and subtropical dry broadleaf forest" classes of potential vegetation. However, two of them fall within the "montane grasslands" class (the species indeed can inhabit this vegetation type in mosaic habitats along the ecotone with

forested regions below; Carleton and Musser, 1989). For data partition 2, both of those latter two localities fell in the test dataset (i.e., not the training set). Accordingly, Maxent's prediction strongly avoided the "montane grasslands" class. The pixels corresponding to those two test localities thus had very low predicted value, bringing down the AUC for that partition. This is an artifact of under-regularization. More regularization for categorical features would allow some prediction in classes with no presence records, especially if the total number of presence records is small (Haffner et al., in preparation, and implemented in later versions of Maxent).

The behavior of Maxent is in fact reasonable in this case, as the training data do not cover the range of vegetation classes that the species can inhabit. Furthermore, it is better than the statistics would suggest, as the occurrence localities falling in montane grasslands both lie on the border with pixels of one of the other two classes inhabited by the species, and are therefore close to highly predicted areas. Their omission should thus be penalized less than other test localities (Fielding and Bell, 1997). Indeed, smoothing the prediction by twice applying a simple 3×3 smoothing convolution with the following weights as a low-pass filter (Jensen, 1996)

$$\begin{pmatrix} 0.05 & 0.05 & 0.05 \\ 0.05 & 0.6 & 0.05 \\ 0.05 & 0.05 & 0.05 \end{pmatrix}$$

increases the AUC to 0.98 for that partition, which is in line with those of the other random partitions, and causes very little visible change to the prediction. Such post-processing may be of general utility when spatial error is known to exist in the data, for example due to errors in site localities or boundaries of polygons representing categorical variables.

4.4. Advantages of Maxent

Maxent exhibits a number of inherent advantages (see Section 1). In addition, visual inspection of the models indicates two further possible advantages. In these analyses, areas predicted by 5–10 of the best-subset GARP models generally showed a reasonable prediction of the species' geographic ranges (see above). Most of those areas were predicted by all 10
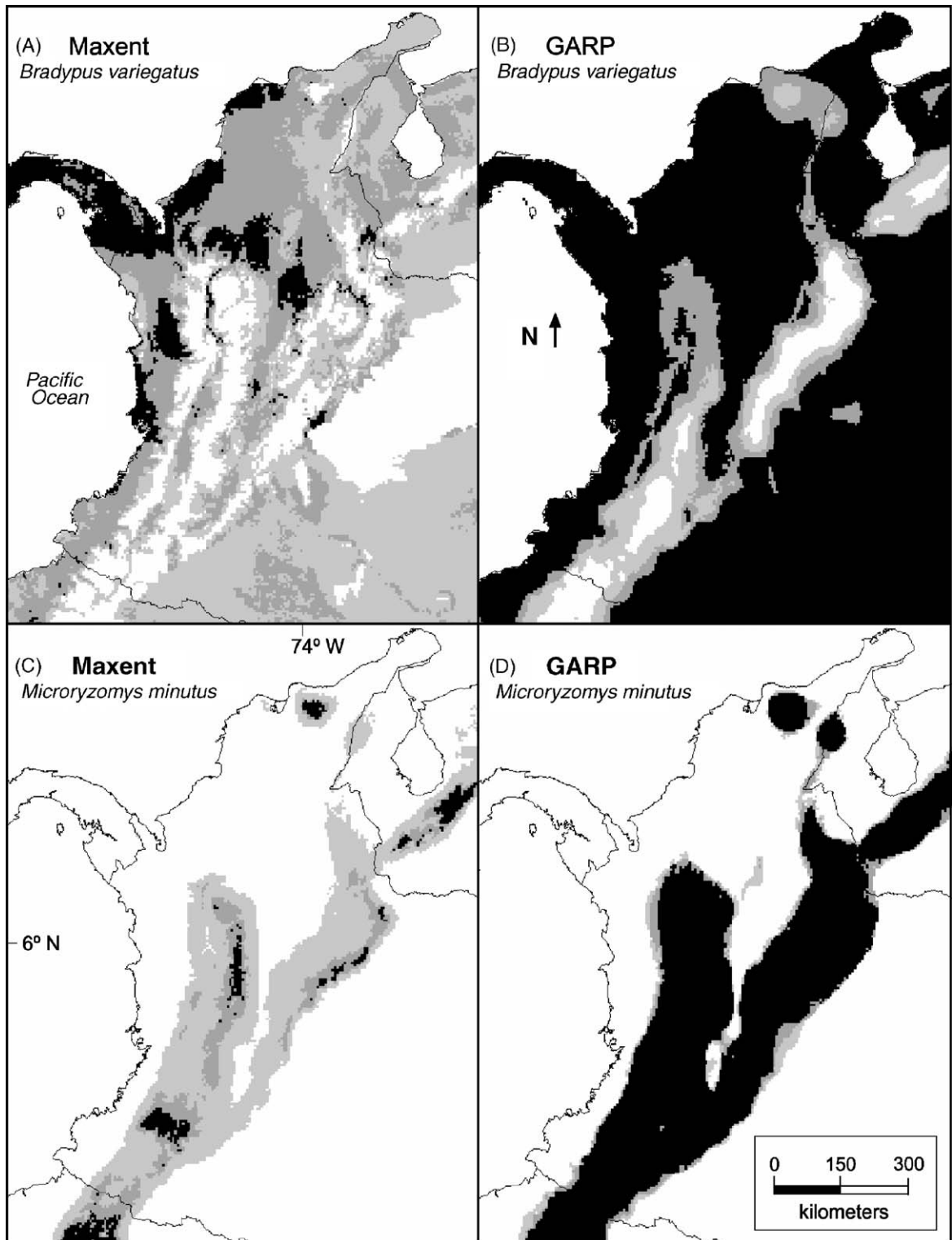
models. In contrast, the Maxent prediction is continuous, and within those areas suitable for each species, it further distinguishes between those with a marginally (but sufficiently) strong prediction versus those with increasingly stronger predictions. This represents an important advantage for Maxent, and explains part of its consistently higher AUC values. The AUC for GARP could potentially be improved by attempting to increase the resolution at the left end of the ROC curve, namely by creating more original binary GARP models (say 1000) and choosing a larger best subset (say 100). We tried this for both species using all occurrence localities and all variables, and found that the predictions were virtually unchanged (in comparison to a best subset of 10 out of 100 models). We also note that even with 100 total models, GARP was already testing the limits of the computers we used (processing all 22 datasets produced almost 20 GB of output, compared with 285 MB for Maxent). Apart from output size, the computational requirements of the two algorithms were similar in this study; GARP averaged 1.95 h to produce a single prediction (best-subset composite derived from 100 models), compared with 2.27 h for Maxent, both on an 850 MHz Pentium 3 processor. Later versions of Maxent available on the website use a faster algorithm (Haffner and Phillips, in preparation); Version 1.8.1 takes a total of 70 min to process all 22 datasets on the above-mentioned computer, or 20 min on a newer 3.2 GHz Intel Xeon computer.

Secondly, Maxent more successfully integrated fine topographic data for both species, producing more detailed (finer-grained) predictions (see close-up images in Fig. 4.2). We propose that this is true, at least in part, because the Maxent model exhibits additivity (while GARP does not), with the contribution of all the variables being added at each pixel (see Eq. (2) in Section 2.1.2). We tried two approaches in an attempt to get GARP to make finer predictions. First, we examined the composite models derived from much larger numbers of GARP models (described above), but the resolution did not increase noticeably. Second, we decreased the convergence limit, allowing GARP to refine its predictions and potentially make more complex models. Again using the full datasets, we reduced the convergence limit from 0.01 to 0.0001, which increased the running time five-fold. Decreasing the convergence limit may result in overfitting in some circumstances; however, we saw no indication of that here. In fact, it improved the prediction for *B. variegatus* somewhat (for example, reducing overprediction in some highland areas), but it did not increase the apparent resolution of the predictions.

4.5. Future work

Much work can be done to refine the use of Maxent for modeling species geographic distributions. Research should determine the number of occurrence localities needed to make an adequate prediction, and to determine how much regularization is needed to avoid overfitting when the number of occurrence localities is small; preliminary results regarding these issues are presented by Dudík et al. (2004) and Phillips et al. (2004). Regarding the quality of the inputs to Maxent, the effect of non-uniform sampling of species localities should be also investigated, building on Zadrozny (2004), with an eye to estimating and limiting the impact of sampling bias (Reddy and Dávalos, 2003). For example, selection of background points taking into account which sites have been sampled (rather than simply at random) can ameliorate the effects of sampling bias in some cases (Zaniewski et al., 2002). As

Fig. 6. Close-up of northwestern South America for the predicted potential geographic distributions of *B. variegatus* (top) and *M. minutus* (bottom) made using all occurrence records and climatic, elevational and potential vegetation variables. Results are given for Maxent (left) and GARP (right). For both species, note the finer grain of the Maxent prediction. For *B. variegatus*, the Maxent prediction correctly indicated unsuitable conditions in the non-forested tropical savannas (*llanos*) of eastern Colombia, but the GARP prediction continued to predict presence there (even with the inclusion of the potential vegetation variable). Four colors are used to indicate the strength of the predictions. Maxent produces a continuous prediction with values ranging from 0 to 100, depicted here by white = [0,1); pale grey = [1,34); dark grey = [34,66); black = [66,100]. The best-subsets selection procedure employed for GARP yields a discrete composite prediction with integer values from 0 to 10, depicted here using white = 0; light grey = 1–4; dark grey = 5–9; black = 10. The strength of the predictions thus cannot be compared directly. All areas with a Maxent prediction of 1 or greater likely represent suitable environmental conditions for the species; in contrast, areas with a GARP prediction of 5–10 probably indicate suitable conditions (see Section 3.2). Note that among areas predicted as suitable for the species for each algorithm, Maxent indicates areas of successively stronger predictions, whereas GARP assigns a maximal value (10) to most such areas (see Section 4.4).

described in Section 4.3, smoothing a prediction may be a useful general method of reducing the negative impact of spatial errors in localities and environmental variables. Additionally, before modeling the species' requirements, smoothing could also be applied to any variables that are suspected of having spatial errors, but it is far from a complete approach to error management. Another possibility, which may improve performance even in the absence of errors in the input data, would be to use bilinear (rather than nearest-neighbor) interpolation to obtain values for the environmental variables at the training localities. Thus, training localities near the boundary between two pixels would receive a combination of the values of the two pixels; for categorical variables, training localities very close to the boundary between two classes would have partial membership in both classes. Alternatively, rather than using a binary feature to represent membership in each class, a continuous feature representing distance from the class could be used.

Research is also called for regarding the use and application of Maxent predictions. First, a good rule needs to be developed to set a threshold operationally using intrinsic data (when a binary prediction is desired). 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 possibility that they may reflect collection biases (areas with many occurrence records). Additional feature types should also be considered, for example, threshold features (see Dudík et al., 2004) and log-linear features (the logarithm of continuous environmental variables; Holdridge et al., 1971).

The great potential utility of data from natural history museums and herbaria, as well as the difficulty of making such data readily available, have been clear for many years (Anderson, 1963; Suarez and Tsutsui, 2004). Recent advances in distributed databases and biodiversity informatics facilitate information retrieval (Baker et al., 1998; Bisby, 2000; Pérez-Hernández et al., 1997; Soberón, 1999; Soberón and Peterson, 2004; Stein and Wieczorek, 2004). Furthermore, fine-resolution environmental data are becoming increasingly available. The success of tools such as Maxent and GARP in modeling species distributions provides increased impetus for such efforts, especially given the gravity of many environmental issues that may be addressed using these techniques.

References

Anderson, R.P., 2003. Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. J. Biogeogr. 30, 591–605.

Anderson, R.P., Gómez-Laverde, M., Peterson, A.T., 2002. Geographical distributions of spiny pocket mice in South America: insights from predictive models. Global Ecol. Biogeogr. 11, 131–141.

Anderson, R.P., Handley Jr., C.O., 2001. A new species of three-toed sloth (Mammalia: Xenarthra) from Panamá, with a review of the genus *Bradypus*. Proceedings of the Biological Society of Washington 114, 1–33.

Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol. Model. 162, 211–232.

Anderson, R.P., Martínez-Meyer, E., 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. Biol. Conser. 116, 167–179.

Anderson, R.P., Peterson, A.T., Gómez-Laverde, M., 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 98, 3–16.

Anderson, S., 1963. Problems in the retrieval of information from natural history museums. In: Proceedings of the Conference on Data Acquisition and Processing in Biology and Medicine, Pergamon Press, New York, pp. 55–57.

Aoki, I., 1989. Hological study of lakes from an entropy viewpoint – Lake Mendota. Ecol. Model. 45, 81–93.

Aspinall, R., 1992. An inductive modeling procedure based on Bayes' theorem for analysis of pattern in spatial data. Int. J. Geogr. Inform. Syst. 6 (2), 105–121.

Austin, M., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Model. 157, 101–118.

Baker, R.J., Phillips, C.J., Bradley, R.D., Burns, J.M., Cooke, D., Edson, G.F., Haragan, D.R., Jones, C., Monk, R.R., Montford, J.T., Schmidly, D.J., Parker, N.C., 1998. Bioinformatics, museums, and society: integrating biological data for knowledge-based decisions. Occasional Papers Museum Texas Tech Univ. 187, 1–4.

Berger, A.L., Della Pietra, S.A., Della Pietra, V.J., 1996. A maximum entropy approach to natural language processing. Comput. Linguist. 22 (1), 39–71.

Bisby, F.A., 2000. The quiet revolution: biodiversity informatics and the internet. Science 289, 2309–2312.

Brown, J.H., Lomolino, M.V., 1998. Biogeography, 2nd ed.. Sinauer Associates, Sunderland, Massachusetts.

Busby, J.R., 1986. A biogeographical analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. Aust. J. Ecol. 11, 1–7.

Carleton, M.D., Musser, G.G., 1989. Systematic studies of oryzomyine rodents (Muridae, Sigmodontinae): a synopsis of *Microryzomys*. Bull. Am. Museum Nat. History 119, 1–83.

Carpenter, G., Gillison, A.N., Winter, J., 1993. DOMAIN: a flexible modeling procedure for mapping potential distributions of plants, animals. Biodivers. Conserv. 2, 667–680.

Corsi, F., de Leeuw, J., Skidmore, A., 2000. Modeling species distribution with GIS. In: Boitani, L., Fuller, T. (Eds.), Research Techniques in Animal Ecology. Columbia University Press, New York, 389–434.

Corsi, F., Duprè, E., Boitani, L., 1999. A large-scale model of wolf distribution in Italy for conservation planning. Conserv. Biol. 13, 150–159.

Della Pietra, S., Della Pietra, V., Lafferty, J., 1997. Inducing features of random fields. IEEE Trans. Pattern Anal. Mach. Intell. 19 (4), 1–13.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. Biometrics 44, 837–845.

Dinerstein, E., Olson, D.M., Graham, D.J., Webster, A.L., Primm, S.A., Bookbinder, M.P., Ledec, G., 1995. Ecoregions of Latin America and the Caribbean (inset map). In: A Conservation Assessment of the Terrestrial Ecoregions of Latin America and the Caribbean, The World Bank, Washington, DC.

Dudík, M., Phillips, S.J., Schapire, R.E., 2004. Performance guarantees for regularized maximum entropy density estimation. In: Proceedings of the 17th Annual Conference on Computational Learning Theory, ACM Press, New York, pp. 655–662.

Eisenberg, J.F., Redford, K.H., 1999. Mammals of the Neotropics. vol. 3, the central Neotropics: Ecuador, Peru, Bolivia, Brazil. University of Chicago Press, Chicago.

Elith, J., 2002. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Ferson, S., Burgman, M. (Eds.), Quantitative Methods for Conservation Biology. Springer-Verlag, New York, 39–58.

Emmons, L.H., 1997. Neotropical Rainforest Mammals: A Field Guide, 2nd ed.. University of Chicago Press, Chicago.

Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J. Appl. Ecol. 41, 263–274.

Fawcett, T., 2003. ROC graphs: notes and practical considerations for data mining researchers. Technical Report HPL-2003–4, Palo Alto, CA: HP Laboratories.

Ferrier, S., Watson, G., 1996. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. Canberra, Australia: NSW National Parks and Wildlife Service. http://www.deh.gov.au/biodiversity/publications/technical/surrogates.

Ferrier, S., Watson, G., Pearce, J., Drielsma, M., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. 1. Species-level modeling. Biodivers. Conserv. 11, 2275–2307.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Env. Conserv. 24, 38–49.

Gouvêa, E., de Ávila Pires, F.D., 1999. Mamíferos do Parque Nacional do Itatiaia. Revista Científica do Centro Universitário do Barra Mansa, UBM 1 (2), 11–26.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C., Peterson, A.T., 2004. New developments in museum-based informatics and ap-

plications in biodiversity analysis. Trends Ecol. Evol. 19 (9), 497–503.

Guisan, A., Edwards Jr., T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157, 89–100.

Guisan, A., Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186.

Hanley, J., McNeil, B., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Hanley, J., McNeil, B., 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148, 839–843.

Hershkovitz, P., 1998. Report on some sigmodontine rodents collected in southeastern Brazil with descriptions of a new genus and six new species. Bonner zoologische Beiträge 47, 193– 256.

Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 87, 2027–2036.

Holdridge, L., Grenke, W., Hatheway, W., Liang, T., Tosi Jr., J., 1971. Forest Environments In Tropical Life Zones: A Pilot Study. Pergamon Press, New York.

Hutchinson, G.E., 1957. Concluding remarks. In: Cold Spring Harbor Symposia on Quantitative Biology 22, 415–427.

Jaynes, E.T., 1957. Information theory and statistical mechanics. Phys. Rev. 106, 620–630.

Jaynes, E.T., 1990. Notes on present status and future prospects. In: Grandy Jr., W.T., Schick, L.H. (Eds.), Maximum Entropy and Bayesian Methods. Kluwer, Dordrecht, The Netherlands, 1–13.

Jensen, J.R., 1996. Introductory Digital Image Processing, A Remote Sensing Perspective. Prentice Hall, Upper Saddle River, NJ.

Joseph, L., Stockwell, D.R.B., 2002. Climatic modeling of the distribution of some Pyrrhura parakeets of northwestern South America with notes on their systematics and special reference to Pyrrhura caeruleiceps Todd, 1947. Ornitología Neotropical 13, 1–8.

Karl, J.W., Svancara, L.K., Heglund, P.J., Wright, N.M., Scott, J.M., 2002. Species commonness and the accuracy of habitat-relationship models. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, 573–580.

Mackey, B.G., Lindenmayer, D.B., 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. J. Biogeogr. 28, 1147–1166.

McPherson, A., 1985. A biogeographical analysis of factors influencing the distribution of Costa Rican rodents. Brenesia 23, 97–273.

Minitab, 1998. Minitab, Release 12.1. Minitab, Inc., State College, PA.

New, M., Hulme, M., Jones, P., 1999. Representing twentieth-century space-time climate variability. Part 1. Development of a 1961–90 mean monthly terrestrial climatology. J. Climate 12, 829–856. Data available at http://ipcc-ddc.cru.uea.ac.uk/cru_data/examine/cru_climate.html.

Ng, A.Y., Jordan, M.I., 2001. On discriminative versus generative classifiers: a comparison of logistic regression and naive Bayes. Adv. Neural Inform. Process. Syst. 14, 605–610.

Nix, H., 1986. A biogeographic analysis of Australian elapid snakes. Atlas of Elapid Snakes of Australia. Australian Government Publishing Service, Canberra, Australia, 4–15.

Patterson, B.D., Ceballos, G., Sechrest, W., Tognelli, M.F., Brooks, T., Luna, L., Ortega, P., Salazar, I., Young, B.E., 2003. Digital Distribution Maps of the Mammals of the Western Hemisphere, Version 1.0. NatureServe, Arlington, VA.

Paynter Jr., R.A., 1982. Ornithological Gazetteer of Venezuela. Museum of Comparative Zoology, Harvard University, Cambridge, MA.

Pearson, R.G., Dawson, T.P., Lin, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography 27, 285–298.

Pérez-Hernández, R., Colomine, G., Villarroel, G., 1997. Los museos de historia natural vinculados con la universidad venezolana y sus perspectivas hacia el siglo XXI. Acta Científica Venezolana 48, 177–181.

Peterson, A.T., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecol. Model. 117, 154–164.

Peterson, A.T., Holt, R.D., 2003. Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. Ecol. Lett. 6, 774–782.

Peterson, A.T., Kluza, D.A., 2003. New distributional modeling approaches for gap analysis. Anim. Conserv. 6, 47–54.

Peterson, A.T., Robins, C.R., 2003. Using ecological-niche modeling to predict barred owl invasions with implications for spotted owl conservation. Conserv. Biol. 17, 1161–1165.

Peterson, A.T., Shaw, J., 2003. Lutzomyia vectors for cutaneous leishmaniasis in southern Brazil: ecological niche models, predicted geographic distribution, and climate change effects. Int. J. Parasitol. 33, 919–931.

Peterson, A.T., Soberón, J., Sánchez-Cordero, V., 1999. Conservatism of ecological niches in evolutionary time. Science 285, 1265–1267.

Phillips, S.J., Dudík, M., Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. In: Proceedings of the 21st International Conference on Machine Learning, ACM Press, New York, pp. 655–662.

Ponder, W.F., Carter, G.A., Flemons, P., Chapman, R.R., 2001. Evaluation of museum collection data for use in biodiversity assessment. Conserv. Biol. 15, 648–657.

Provost, F.J., Fawcett, T., 1997. Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. Knowledge Discovery and Data Mining. ACM Press, New York, 43–48.

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecol. Lett. 3, 349–361.

Ratnaparkhi, A., 1998. Maximum entropy models for natural language ambiguity resolution. Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA.

Reddy, S., Dávalos, L.M., 2003. Geographical sampling bias and its implications for conservation priorities in Africa. J. Biogeogr. 30, 1719–1727.

Root, T., 1988. Environmental factors associated with avian distributional boundaries. J. Biogeogr. 15, 489–505.

Schneider, E., Kay, J., 1994. Life as a manifestation of the second law of thermodynamics. Math. Comput. Model. 19 (6–8), 25–48.

Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), 2002. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423, 623–656.

Soberón, J., 1999. Linking biodiversity information sources. Trends Ecol. Evol. 14, 291.

Soberón, J., Peterson, A.T., 2004. Biodiversity informatics: managing and applying primary biodiversity data. Philos. Trans. Royal Soc. Lond. B 359, 689–698.

Stein, B.R., Wieczorek, J., 2004. Mammals of the world: MaNIS as an example of data integration in a distributed network environment. Biodivers. Inform. 1 (1), 14–22.

Stockwell, D., Peters, D., 1999. The GARP modeling system: problems and solutions to automated spatial prediction. Int. J. Geograph. Inform. Sci. 13, 143–158.

Stockwell, D.R.B., Noble, I.R., 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. Math. Comput. Simul. 33, 385–390.

Stockwell, D.R.B., Peterson, A.T., 2002a. Controlling bias in biodiversity data. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC 537–546.

Stockwell, D.R.B., Peterson, A.T., 2002b. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148, 1–13.

Suarez, A.V., Tsutsui, N.D., 2004. The value of museum collections for research and society. BioScience 54 (1), 66–74.

Tate, G.H.H., 1939. The mammals of the Guiana region. Bull. Am. Museum Nat. History 76, 151–229.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145–148.

USGS, 2001. HYDRO 1k, elevation derivative database. United States Geological Survey, Sioux Falls, South Dakota. Available at http://www.edcdaac.usgs.gov/gtopo30/hydro.

Vida, S., 1993. A computer program for non-parametric receiver operating characteristic analysis. Comput. Meth. Prog. Biomed. 40, 95–101.

Welk, E., Schubert, K., Hoffmann, M.H., 2002. Present and potential distribution of invasive mustard (*Alliara petiolata*) in North America. Divers. Distributions 8, 219–233.

Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R., Stewart, A.M., 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography 16 (3), 120–127.

Williams, P.M., 1995. Bayesian regularization and pruning using a Laplace prior. Neural Comput. 7 (1), 117–143.

Yee, T.W., Mitchell, N.D., 1991. Generalized additive models in plant ecology. J. Veg. Sci. 2, 587–602.

Yom-Tov, Y., Kadmon, R., 1998. Analysis of the distribution of insectivorous bats in Israel. Divers. Distributions 4, 63–70.

Zadrozny, B., 2004. Learning and evaluating classifiers under sample selection bias. In: Proceedings of the 21st International Conference on Machine Learning, pp. 903–910.

Zaniewski, A.E., Lehmann, A., Overton, J.M., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol. Model. 157, 261–280.

Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. 39 (4), 561–577.