



Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods

Modelado de nichos ecológicos y predicción de distribuciones geográficas: comparación de seis métodos

Miguel A. Ortega-Huerta¹ and A. Townsend Peterson²

¹Instituto de Biología, Universidad Nacional Autónoma de México. Estación de Biología Chamela, Apartado postal 21, 48980, San Patricio, Jalisco, Mexico

²Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, Kansas 66045 USA

*Correspondent: maoh@ibiologia.unam.mx

Abstract. Modeling ecological niches of species as a means to predict geographic distributions is a growing field that has been applied to numerous challenges of importance in ecology, systematics, and human well-being. The increasing availability and variety of such predictive algorithms requires testing their performance. In this study, we compare 6 such algorithms (Maxent, BioMapper, DOMAIN, FloraMap, the genetic algorithm GARP, and weights of evidence) as regards their ability to predict the geographic distributions of 10 species of Mexican birds for which ample distributional data are available. The results of this study nevertheless led to reflections on how model quality should be evaluated.

Key words: ecological niche modeling, species' distributions, algorithms, model validation.

Resumen. La predicción de las distribuciones geográficas de las especies obtenida mediante el modelado de sus nichos ecológicos, representa una línea de investigación en expansión, la cual ha sido aplicada en múltiples áreas de conocimiento tales como ecología, sistemática y salud pública. La creciente disponibilidad y variedad de tales métodos y algoritmos de predicción determina su evaluación como necesaria. En este estudio, comparamos 6 algoritmos (Maxent, BioMapper, Domain, FloraMap, GARP, Weights of Evidence) con respecto a su habilidad para predecir las distribuciones geográficas de 10 especies de aves de México, para las cuales se cuenta con suficientes datos distribucionales. No obstante, los resultados de nuestro estudio sugieren la necesidad de elaborar nuevos criterios para la evaluación de modelos.

Palabras clave: modelado de nichos ecológicos, distribuciones geográficas de especies, algoritmos, validación de modelos.

Introduction

A growing field in ecology is that of modeling ecological niches for prediction of geographic distributions of species (Scott et al., 2002). These models permit analysis of a wide variety of biodiversity phenomena, including geographic distributions (Elith and Burgman, 2002), future potential distributions under scenarios of climate change (Thomas et al., 2004), species' invasions (Peterson, 2003), agricultural crop damage by pest organisms (Sánchez-Cordero and Martínez-Meyer, 2000), and priorities for biodiversity conservation (Chen and Peterson, 2002). Given the intense activity in this expanding field, the relative merits of the various methods that have been employed to model ecological niches demand further exploration.

The methods that have been used for modeling ecological niches are diverse, including multiple logistic regression and other forms of general linear models (Austin et al., 1990), set-based approaches characterizing ranges of species along ecological dimensions (Nix, 1986), approaches based on distance measures in ecological space (Carpenter et al., 1993; Hirzel et al., 2002), maximum entropy approaches (Phillips et al., 2006), and genetic algorithms (Stockwell, 1999; Stockwell and Noble, 1992; Stockwell and Peters, 1999), to name a few. Several studies have developed comparisons among methods (Cumming, 2000; Manel et al., 1999; Manel et al., 1999; Tsoar et al., 2007), but only one (Elith et al., 2006) has been at all comprehensive, and many methods remain untested.

A special focus has been on the use of presence-only data for modeling species distributions, even though they face serious challenges in model inference relative

to presence-absence methods (Wintle et al., 2005). Presence-only methods are necessary because absence of species is difficult to demonstrate, and because false absences can decrease the reliability of predictive models (Chefaoui et al., 2005). A species may be recorded as absent at a given location because the species is present but could not be detected, because the species is absent but the habitat is suitable, or because the habitat is truly unsuitable for the species, the former 2 situations can lead to identify false absences (Hirzel et al., 2002). Predicting species' distributions from presence-only data sets and pseudoabsences (i.e., data resampled from areas not holding presences) has the potential to be a useful alternative when presence/absence data are unavailable or impossible to obtain (Zaniewski et al., 2002; Brotons et al., 2004; Graham et al., 2004; Guisan and Thuiller, 2005). Among the attempts to evaluate presence-only models, Hirzel et al., (2006) identified 2 approaches: *a*), generate pseudoabsences and apply standard presence/absence techniques, and *b*), assess how much model predictions are better than random expectations.

Because the challenges involved in modeling ecological niches may vary among regions and taxa, and given the intense efforts focused around understanding biodiversity phenomena in Mexico (CONABIO, 2002), this study was developed to evaluate the behavior of 6 alternative methodologies for Mexican taxa and landscapes. The 6 methods were selected based in the different types of algorithms applied and the potential utility in modeling biodiversity patterns. Three of the methodologies assessed had not been compared with other approaches previously, even in the most comprehensive study to date (Elith et al., 2006).

In this study, the models generated by each method are analyzed as the ecological niche of 10 Mexican bird species for which ample distributional data are available. Our approach follows the relations between niche and species' distribution proposed by Pulliam (2000): the Grinnellian niche concept is that of the set of conditions suitable to the point that the species can maintain populations. This study was neither designed to identify a 'best' method, nor to make comparisons based on exactly the same data. That is, each method was presented with an information base, and was allowed to use whatever portion of that information that it *could*; the relative merits of each method are explored, and the behavior of each method is characterized. In particular, this investigation focused on reconsidering the methods by which we evaluate models to choose the 'best' (= most predictive) model, and on the question about the potential confusing role of statistical significance as a measure of model's predictive ability.

Materials and methods

Input data. Comparisons among methods were based on standard sets of information that were available to each. For occurrence information, we chose 10 bird species that 1), were well-sampled ($N \geq 100$ unique occurrence points in Mexico); 2), that showed a diversity of distributional areas within Mexico [e.g., from relatively restricted (*Thryothorus sinaloa*) to relatively wide (*Myadestes occidentalis*)], and 3), that varied in ecological requirements [e.g., desert (*Campylorhynchus brunneicapillus*), pine forest (*Atlapetes pileata*), tropical forest (*Tityra personata*)]. In each case, we selected 50 unique occurrence points randomly, and set them aside as an independent testing data set; models were developed based on the remaining ≥ 50 occurrence points for each species. (We used single random partitions only for each species because of the computationally intense nature of several of the algorithms explored herein).

Environmental data sets used in these modeling efforts included 10 raster GIS coverages with pixel resolution of 0.05° . Themes included elevation, slope, and aspect (all resampled from 0.01° data sets available from the U.S. Geological Survey Hydro-1K digital elevation model dataset, <http://edcdaac.usgs.gov/gtopo30/hydro/>; climate data including annual mean, maximum, minimum, maximum daily, and minimum daily temperatures, and annual mean precipitation (from the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO): <http://www.conabio.gob.mx>; and potential vegetation (Rzedowski, 1978; also available from the CONABIO website). Different methods used most or all of these data layers, depending on their particular requirements; however, one method, FloraMap, does not allow for user input of environmental data sets, and so used its own native climatic data only (Table 1). Details of our implementation for each method follow.

BioMapper. The ecological niche factor analysis (ENFA) implemented in BioMapper was developed by Hirzel et al. (2002) as a method to calculate habitat suitability maps without the need for data to document species' absences. BioMapper is designed to compute factors that best explain species' ecological distributions. Much as in principal components analysis, factors extracted are by design uncorrelated, but in this case have biological significance. The first factor is the "marginality factor", which describes how far the species' optimum conditions deviate from the conditions dominant in the study area. Next, "specialization factors" are obtained that assess how the species' variance differs from the total variance in the ecological dimensions. Hence, in Biomapper, a relatively few factors explain most of the variation in species' ecological distributions. Main considerations in using Biomapper are as follows:

Table 1. Environmental data coverages used by each algorithm in the development of ecological niche models for distributional predictions. All coverages are continuous unless otherwise noted

<i>Environmental variables</i>	<i>BioMapper</i>	<i>FloraMap</i>	<i>Domain</i>	<i>Weights of evidence</i>	<i>GARP Maxent</i>
Elevation	X		X	X (20 classes)	X
Slope	X		X	X (20 classes)	X
Aspect	X		X	X	X
Annual mean temperature	X		X	X	X
Maximum temperature	X		X	X	X
Minimum temperature	X		X	X	X
Maximum daily temperature	X		X		X
Minimum daily temperature	X		X		X
Annual mean precipitation	X		X	X	X
Potential vegetation (categorical)					X
Monthly rainfall (12 coverages)		X			
Monthly average temperatures (12 coverages)		X			
Diurnal temperature range (12 monthly coverages)		X			

1), the use of categorical variables in a factor analysis is puzzling, so potential vegetation was excluded from analyses; 2), data were normalized, applying the Box-Cox variable transformation, and 3), habitat suitability maps were generated (5 factors, 10 categories) via a series of 1-dimensional histograms.

Domain. Domain (Windows version 1.3) was implemented by the Center for International Forestry Research, Bogor, Indonesia, based on the original program (Carpenter et al., 1993). At its simplest, this algorithm generates maps of ecological similarity or distance (Gower metric) to those sites at which the taxon is known to occur to predict the potential geographic distribution of a species. For any location in the study area, Domain assigns each cell the Gower distance between that cell and the closest point in the training set. If averaging is enabled, the value stored is the average of the *n* nearest cells. Analyses are generally conducted with *n* = 1, but larger values can be useful in reducing effects of outlier training points. Environmental attributes were imported as continuous ASCII files with 3 columns (longitude, latitude, value). Domain was used applying both complete categorical dissimilarity and complete similarity $(1 - D) * 100$. *Weights of evidence.*

Weights of evidence is a quantitative method for combining data-based evidence in support of a hypothesis. The method was originally developed for non-spatial applications in medical diagnosis, in which evidence consisted of a set of symptoms and the hypothesis was of the type “this patient has disease X”. For each symptom, a pair of weights was calculated, 1 for presence of the symptom, and 1 for absence. The magnitude of the weights depended on the association measured between the symptom and the pattern of disease in a large group of patients. These weights could then be used to estimate the probability that a new patient would get the disease, based on the presence or absence of symptoms.

Weights of evidence was adapted in the late 1980’s for mapping mineral deposits with GIS (Raines et al., 2000), which is the implementation used herein. Here, the evidence consists of exploration data sets (= maps), and the hypothesis is “the location is favorable for occurrence of deposit type X”. Weights are estimated from the measured association between known mineral occurrences and values on the maps to be used as predictors. The hypothesis is then evaluated repeatedly for the entire study area using the calculated weights, producing a prediction of the species’

distribution in which evidence from several environmental map layers is combined. This technique belongs to a class of methods suitable for multi-criterion decision-making. Similar to multiple regression in statistics, this approach involves estimation of a response variable from a set of predictor variables based on Bayesian probabilities, with the assumption of conditional independence. For implementing this approach, it was necessary to define area units in km², so we re-projected all environmental layers to a Lambert Conic Conformal projection.

GARP. The *Genetic Algorithm for Rule-set Prediction* is a machine-learning meta-algorithm for ecological niche modeling and distributional prediction. Developed originally in UNIX (Stockwell, 1999; Stockwell and Noble, 1992; Stockwell and Peters, 1999), and now ported to Windows <http://nhm.ku.edu/desktopgarp>, GARP uses known occurrence points and points resampled from the entire map to create populations of presences and pseudoabsences, respectively. Four simple subalgorithms are used to create rules in the form of IF <condition1> <condition2> <condition3> ... THEN <prediction>. These simple initial rules are then optimized via a genetic algorithm, in which particular conditions may be perturbed, combined with conditions from other rules, etc. The end result is a heterogeneous set of 20-50 rules, which in aggregate describe the ecological distribution of a species. In this implementation of GARP, we set the convergence criterion to 0.01, and maximum iterations permitted to 1000. We ran 1000 models for each species, and selected the 10 best (GARP models differ from one another owing to the random-walk nature of the process) using a “best subsets” procedure that separates methods by their omission-commission error characteristics (Anderson et al., 2003).

MaxEnt. Maximum entropy is also a machine-learning general-purpose method used to obtain predictions or make inferences from incomplete information (Phillips et al., 2006). Given a set of samples (i.e., species occurrence) and set of features (environmental variables), MaxEnt estimates niches by finding the distribution of probabilities closest to uniform (maximum entropy), constrained to the fact that feature values match their empirical average (Phillips, 2004). Phillips et al., (2006) document the main features of the MaxEnt software: 1), it uses presence-only data but can also use presence-absence data; 2), environmental data may be both continuous and categorical, and MaxEnt can incorporate interactions between variables; 3), efficient deterministic algorithms make possible estimation of a maximum entropy probability distribution; 4), because of its mathematical definition, it is possible to interpret how environmental variables relate to model suitability; 5), overfitting can be regulated, and 6), continuous output makes possible identification of fine distinctions of model

suitability. The main disadvantage of MaxEnt is the need of further research into issues like regularization and the results produced by the exponential probabilistic model applied.

Even though MaxEnt performs internal model validation tests, we decided to run this software using all training presence sites, evaluating model performance outside MaxEnt, as with the other methods. Main parameters when running MaxEnt software included: feature types = linear, quadratic, product, threshold, and hinge; regularization multiplier = 3.0; regularization values (linear/quadratic/product = 0.050, categorical = 0.050, threshold = 1.000, hinge = 0.500); Maximum iterations = 500, and convergence threshold = 1.0×10^{-5} .

FloraMap. FloraMap (<http://www.floramap-ciat.org/ing/floramap101.htm>) is based on calculations of probabilities that a particular climate condition belongs to a multivariate normal distribution at which a training set of occurrences has been found (Jones and Gladkov, 1999). The methodology may be extended to cover occurrences of any organism with a distribution largely determined by climate parameters.

FloraMap uses a set of interpolated climate surfaces, a method for calculating the probability model, and a method for mapping probabilities over the climate surface. Principal components analysis is used to construct sets of linear combinations of the raw climate variables that maximize the variance in each, are orthogonal to each other, and are uncorrelated. In the end, each pixel is characterized in terms of distance to an n -dimensional probability density function.

Our application of FloraMap used the FloraMap environmental data (per force). We used an 18 x 18 km map resolution, 31 variables in 3 groups (monthly rainfall totals, monthly average temperatures, and monthly average diurnal temperature range). We used power = 0.50, with transformation [rain^a], and all weights set to 1.0. The number of scores was set as $N = 7$, and the lowest probability = 0.0005.

Model evaluation. Two approaches were used to evaluate model performance—1 threshold-independent (Receiver Operating Characteristics plots) and 1 threshold-dependent (chi-square tests)—each has strengths and weaknesses. The threshold-dependent approach is based on coincidence between test occurrence points and model predictions. A 1-tailed chi-square test is built based on observed numbers of correct and incorrect predictions for the test occurrence points, in comparison with expected numbers derived from product of the number of test occurrence points and the proportional area predicted present *versus* absent. With 1 degree of freedom, this approach provides an evaluation of how well the test occurrence points are predicted, taking

into account the proportional area predicted present. When model results are other than binary, however, a necessary step is that of choosing a threshold above which the prediction is considered present. In this study, we chose a threshold for each species and method, based on the level of prediction of the lowest prediction level for any of the input (training) presence points (Pearson et al., 2006).

For a threshold-independent evaluation of model predictivity, we used Receiver Operating Characteristic (ROC) analyses (Fielding and Bell, 1997). This statistic evaluates the sensitivity (absence of commission error) and specificity (absence of omission error) of a diagnostic test in the face of the independent testing dataset. The testing dataset provides a “gold standard” for presence and an equal number of pixels from which the species has not been sampled (pseudoabsences) provide a characterization of absence; each individual model is scored on its ability to predict the new data correctly. These scores are accumulated stepwise, and graphed on an axis of sensitivity (true positive rate of accumulation) and 1 - specificity (true negative rate of accumulation). The result is integrated to produce an area under the curve (AUC) that measures how well the model predicts the new point occurrences. The theoretically perfect result is AUC = 1.0, whereas a test performing no better than random yields AUC = 0.5. The result can be evaluated using a standard normal approximation (*z*-test). All of our ROC analyses were developed using SPSS statistical software v.13.0 (LEAD Technologies Inc.). Pseudoabsence data were generated by selecting an equal population of points ($N = 50$) randomly from those areas documented outside the known distributions of the species. Digital coverages of the species documented geographic distributions were obtained from the project NatureServe (Ridgely et al., 2005). A GIS software (ArcView 3.2) was used to isolate the no-occurrence areas for each species and then to generate 50 random sites within such areas.

Results

Results of the 6 approaches tested herein were variable, with predictions often ranging several-fold in area predicted present among methods for a given species (Fig. 1). All approaches agreed on what could be considered core areas (areas in which species are most likely to be found), in which test occurrence points had a high probability of falling.

However, a spectrum of general tendencies could be distinguished among the different algorithms, ranging from 1), micro-prediction, in which only a core set of points was successfully predicted; 2), generally good prediction, from

which small sets of points were nonetheless left out, and 3), relatively broad predictions including areas larger than the distribution of the test occurrence points. Examples of these patterns can be seen in figure 1, in which weights of evidence produces a micro-prediction, GARP produces a relatively large and inclusive prediction, and the remaining approaches produce generally good predictions, but omit some sets of points (Fig. 1, arrows).

Testing these results across species and modeling approaches using the threshold-dependent chi-square approach, almost all models for all species were seen to be statistically significant (Table 2). Only 3 predictions (1 each for Weights of evidence, Domain, and MaxEnt) were not significantly better than a random prediction. As the chi-square statistic summarizes positive departure from random expectations, its magnitude can be used as an index of model predictive ability (Peterson et al., 1999)—we noted that the average of the chi-square statistic across species was 79.1 for GARP, and lower (41.4-66.8) in the other methods, in spite of the broader areas predicted by GARP.

The threshold-independent ROC approach showed similar trends (Figure 2): most predictions for most species were statistically significant (*z*-tests, $P < 0.05$, Table 3). However, inspecting patterns of failures, FloraMap failed to achieve statistical significance in 2 of 10 species; and Weights of evidence in 1 species. The other 4 methods produced highly statistically significant predictions for all species.

Discussion

This analysis highlights several issues that challenge the growing field of modeling ecological niches and predicting geographic distributions. In particular, questions revolve around issues such as spatial scale, user-friendliness, degree of customization necessary for analysis of a particular species, and computational demands. Some conjunction of these and other considerations will define the ideal method, if 1 is to exist. We consider 2 such questions in detail below.

Ability to use diverse environmental information. One dimension in which we observed strong contrasts among methods was in the types of environmental information that could be used. At one end of the spectrum, FloraMap uses a predetermined set of 36 environmental variables, and does not admit any additional dimensions that a user might wish to include. Weights of evidence as a computational algorithm clearly was near its limits (even on reasonably fast CPUs) with the 10-dimension challenge that our analyses represented: not all of the climate layers

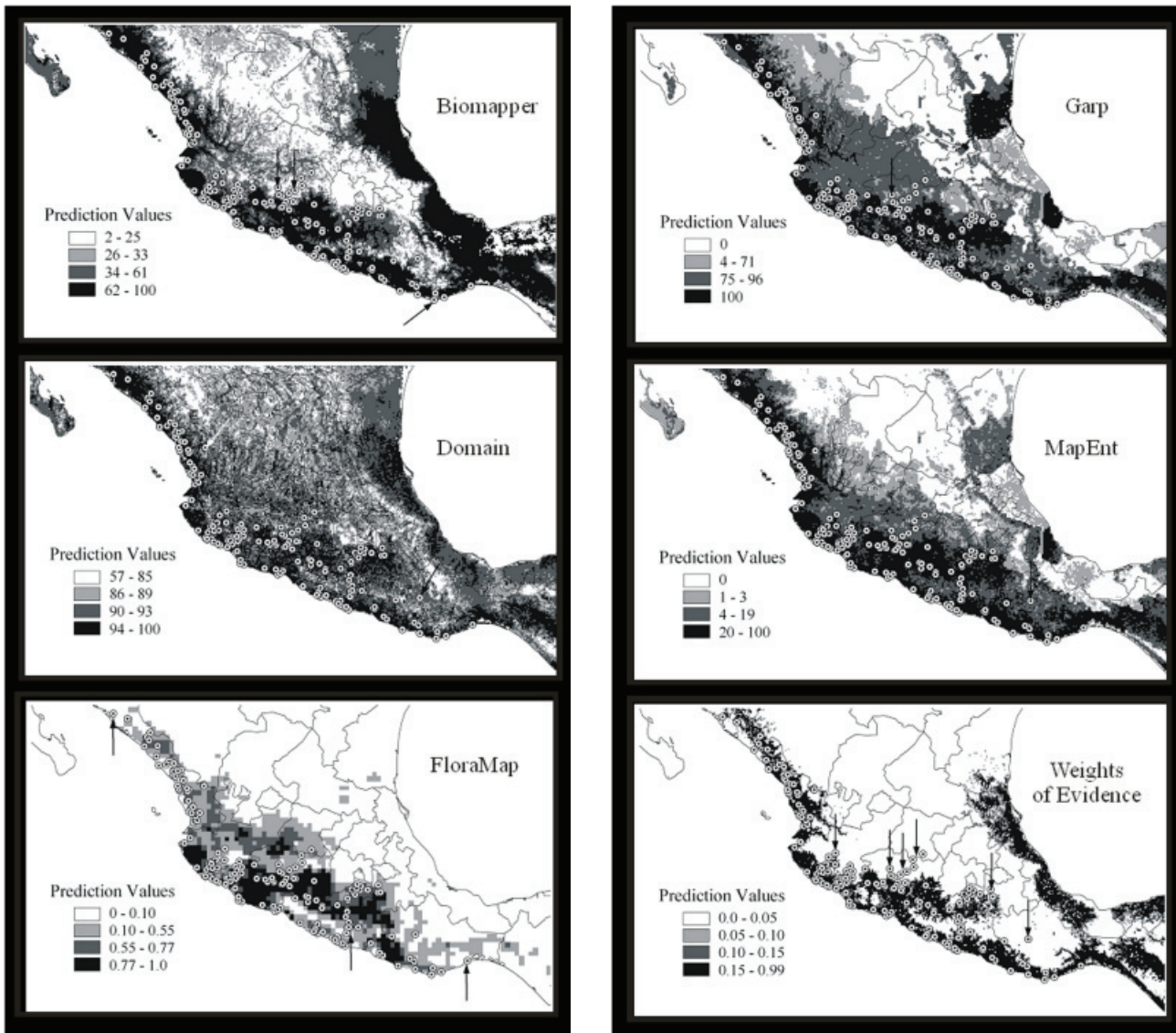


Figure 1. Example results for the Golden-cheeked Woodpecker *Melanerpes chrysogenys*, across the 6 modeling approaches. Shading ramps are arbitrarily chosen, but with every effort to balance across methods. Arrows indicate failures of models to predict known occurrences.

could be used, and those that were used had to be re-classed into 20 discrete, ordered categories. At the other extreme, GARP and MaxEnt were able to use all of the environmental dimensions provided, including even a potential vegetation dataset that was categorical in nature; the other methods were not able to take advantage of this dataset without further data transformations (e.g., using Boolean maps in BioMapper). Hence, 2 of the methods showed significant limitations regarding ability to take advantage of numerous and diverse information sets.

Model validation. The use of statistical significance as a

measure of model validity is generally accepted, and yet is worthy of some discussion (Fielding and Bell, 1997). Previous authors (Anderson et al., 2003) have argued that the “best” models may not be the most significant ones; rather, best models should be identified based on the specific combinations of Type I (omission) and Type II (commission) errors that they present. It is worth noting, for those who have dismissed the best-subsets procedure as a ‘GARP thing,’ that this approach can be used with deterministic algorithms if input occurrence or environmental data are manipulated using a bootstrap or

Table 2. Summary of results of threshold-dependent chi-square tests of model quality for the 6 modeling methods examined. Obs correct = number of points observed to be predicted successfully; Exp correct = number of points expected to be predicted successfully, given the area predicted present. Model failures are indicated in boldface

<i>Method</i>	<i>Species</i>	<i>Obs correct</i>	<i>Exp correct</i>	<i>Chi-square</i>	<i>P</i>
BioMapper	<i>Tyrannus crassirostris</i>	24	10.8	20.4	6.4 x 10 ⁻⁶
	<i>Myadestes occidentalis</i>	23	14.2	7.6	5.7 x 10 ⁻³
	<i>Tityra personata</i>	32	9.1	70.4	4.8 x 10 ⁻¹⁷
	<i>Thryothorus sinaloa</i>	33	12.9	42.1	8.8 x 10 ⁻¹¹
	<i>Sittasomus griseicapillus</i>	24	9.1	29.8	4.8 x 10 ⁻⁸
	<i>Ortalis vetula</i>	28	8.5	54.4	1.6 x 10 ⁻¹³
	<i>Mitrephanes phaeocercus</i>	28	10.3	38.1	6.7 x 10 ⁻¹⁰
	<i>Melanerpes chrysogenys</i>	33	10.6	59.9	10.0 x 10 ⁻¹⁵
	<i>Atlapetes pileata</i>	32	8.8	73.8	8.8 x 10 ⁻¹⁸
	<i>Campylorhynchus brunneicapillus</i>	41	26.4	17.2	3.4 x 10 ⁻⁵
Domain	<i>Tyrannus crassirostris</i>	31	12.7	35.2	3.0 x 10 ⁻⁹
	<i>Myadestes occidentalis</i>	35	29.6	2.4	0.12
	<i>Tityra personata</i>	31	7.2	91.7	10.0 x 10 ⁻²²
	<i>Thryothorus sinaloa</i>	29	9.7	47.9	4.5 x 10 ⁻¹²
	<i>Sittasomus griseicapillus</i>	33	10.4	62.2	3.2 x 10 ⁻¹⁵
	<i>Ortalis vetula</i>	35	11.3	64.2	1.1 x 10 ⁻¹⁵
	<i>Mitrephanes phaeocercus</i>	49	44.1	4.7	0.03
	<i>Melanerpes chrysogenys</i>	43	11.3	115.0	7.8 x 10 ⁻²⁷
	<i>Atlapetes pileata</i>	32	15.2	26.7	2.4 x 10 ⁻⁷
	<i>Campylorhynchus brunneicapillus</i>	46	26.9	29.4	5.9 x 10 ⁻⁸
FloraMap	<i>Tyrannus crassirostris</i>	41	16.7	53.0	3.4 x 10 ⁻¹³
	<i>Myadestes occidentalis</i>	44	27.5	21.9	2.9 x 10 ⁻⁶
	<i>Tityra personata</i>	41	13.1	80.5	3.0 x 10 ⁻¹⁹
	<i>Thryothorus sinaloa</i>	36	9.8	87.7	7.8 x 10 ⁻²¹
	<i>Sittasomus griseicapillus</i>	42	15.5	66.0	4.5 x 10 ⁻¹⁶
	<i>Ortalis vetula</i>	40	15.1	58.7	1.8 x 10 ⁻¹⁴
	<i>Mitrephanes phaeocercus</i>	45	19.5	54.6	1.5 x 10 ⁻¹³
	<i>Melanerpes chrysogenys</i>	42	8.1	168.1	2.0 x 10 ⁻³⁸
	<i>Atlapetes pileata</i>	47	18.4	70.4	4.8 x 10 ⁻¹⁷
	<i>Campylorhynchus brunneicapillus</i>	42	33.0	7.2	7.3 x 10 ⁻³
GARP	<i>Tyrannus crassirostris</i>	43	12.3	101.4	7.5 x 10 ⁻²⁴
	<i>Myadestes occidentalis</i>	44	19.4	51.1	8.8 x 10 ⁻¹³
	<i>Tityra personata</i>	48	12.8	130.8	2.8 x 10 ⁻³⁰
	<i>Thryothorus sinaloa</i>	35	9.6	83.5	6.4 x 10 ⁻²⁰
	<i>Sittasomus griseicapillus</i>	47	15.9	89.4	3.3 x 10 ⁻²¹
	<i>Ortalis vetula</i>	47	17.6	75.8	3.1 x 10 ⁻¹⁸
	<i>Mitrephanes phaeocercus</i>	40	11.5	91.3	1.2 x 10 ⁻²¹
	<i>Melanerpes chrysogenys</i>	31	7.0	95.2	1.8 x 10 ⁻²²
	<i>Atlapetes pileata</i>	45	20.0	51.9	5.9 x 10 ⁻¹³
	<i>Campylorhynchus brunneicapillus</i>	31	16.1	20.4	6.3 x 10 ⁻⁶

Table 2. Continues

<i>Method</i>	<i>Species</i>	<i>Obs correct</i>	<i>Exp correct</i>	<i>Chi- square</i>	<i>P</i>
MaxEnt	<i>Tyrannus crassirostris</i>	43	10.1	133.7	6.5 x 10 ⁻³¹
	<i>Myadestes occidentalis</i>	41	13.6	76.1	2.7 x 10 ⁻¹⁸
	<i>Tityra personata</i>	40	7.0	180.2	4.5 x 10 ⁻⁴¹
	<i>Thryothorus sinaloa</i>	24	7.8	40.0	2.5 x 10 ⁻¹⁰
	<i>Sittasomus griseicapillus</i>	43	12.5	98.5	3.1 x 10 ⁻²³
	<i>Ortalis vetula</i>	36	8.7	104.1	2.0 x 10 ⁻²⁴
	<i>Mitrephanes phaeocercus</i>	41	13.3	78.2	9.5 x 10 ⁻¹⁹
	<i>Melanerpes chrysogenys</i>	40	7.1	176.6	2.7 x 10 ⁻⁴⁰
	<i>Atlapetes pileata</i>	45	10.8	137.3	1.0 x 10 ⁻³¹
	<i>Campylorhynchus brunneicapillus</i>	45	24.9	32.2	1.4 x 10 ⁻⁰⁸
Weights of evidence	<i>Tyrannus crassirostris</i>	43	16.5	63.3	1.8 x 10 ⁻¹⁵
	<i>Myadestes occidentalis</i>	41	13.9	72.9	1.4 x 10 ⁻¹⁷
	<i>Tityra personata</i>	43	11.8	108.5	2.1 x 10 ⁻²⁵
	<i>Thryothorus sinaloa</i>	36	13.7	49.7	1.8 x 10 ⁻¹²
	<i>Sittasomus griseicapillus</i>	49	14.4	116.8	3.1 x 10 ⁻²⁷
	<i>Ortalis vetula</i>	43	14.7	77.3	1.5 x 10 ⁻¹⁸
	<i>Mitrephanes phaeocercus</i>	11	13.8	0.8	0.38
	<i>Melanerpes chrysogenys</i>	34	12.8	47.5	5.6 x 10 ⁻¹²
	<i>Atlapetes pileata</i>	43	12.6	97.5	5.4 x 10 ⁻²³
	<i>Campylorhynchus brunneicapillus</i>	44	27.4	22.2	2.5 x 10 ⁻⁶

jackknife manipulation. Nonetheless, although developed for replicate models from a single algorithm, this schema can be a useful heuristic tool in the present comparisons (Anderson et al., 2003).

The 2 significance tests used herein, however, both balance correct prediction of test points against proportional area predicted present. This balance, at first glance, is beneficial: a “cheating” algorithm might simply predict the entire area present, and thus not fail in predicting presence for a single point. However, with more careful inspection, this balance can distract from true predictivity (in this case, correct prediction of the entire range of distributional possibilities of a species). Consider the equation for the chi-square statistic, which is $(O-E)^2/E$, where O and E are observed and expected values, respectively. This number (and significance) can be maximized in 2 ways: either increase the numerator (= correct prediction of test points) or decrease the denominator (= smaller area predicted). Particularly for species predicted to have wide-ranging distributions (for which E is high), the latter can be much easier: micro-prediction of a subset of test points can be more significant than more complete prediction of the test

points. In this sense, these approaches have the potential to select models that maximize the *wrong* quantity.

Returning to the best-subsets comparison (Anderson et al., 2003), the best approaches to predicting distributions of species would first and foremost minimize omission of the independent test points. Beyond that, their position along the commission axis (area predicted present) is less clear—certainly neither too high (large predicted area) nor too low (small predicted area). In the present study, the results for the 6 modeling approaches spread out with a concave negative relationship (not shown) between omission and commission, just as in the within-model applications (Anderson et al., 2003). Weights of evidence models generally fell at the upper left of this distribution (high omission, smallest area predicted), and GARP models generally fall at the lower right (low omission, larger area predicted).

Presence and absence data for model tests. The tests presented were constrained by the fact that only presence data were available for model development and testing (Elith et al., 2006). However, the effects of historical factors (e.g., limited dispersal, speciation, extinction) must also be

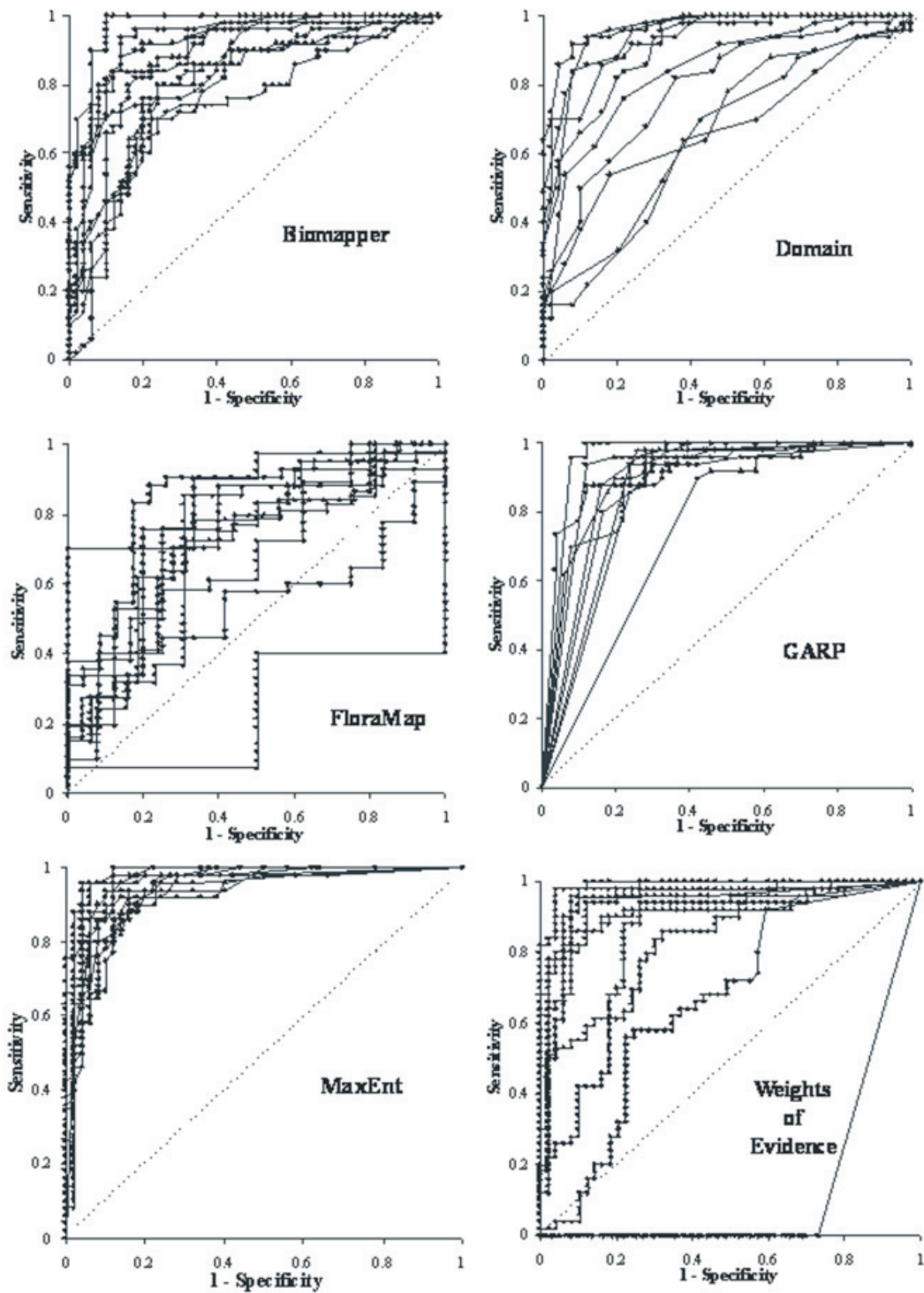


Figure 2. Receiver operating characteristic (ROC) curves for the 10 species analyzed for each algorithm examined in this study. Highest quality models would show a ROC curve that rises quickly towards the upper left corner of the plot. Poor models would either follow the diagonal or fall below the diagonal.

Table 3. Summary of results of the threshold-independent ROC tests of model quality. Model failures are indicated in bold

<i>Species</i>	<i>Statistic</i>	<i>Bio- Mapper</i>	<i>Domain</i>	<i>FloraMap</i>	<i>GARP</i>	<i>MaxEnt</i>	<i>Weights of evidence</i>
<i>Atlapetes pileata</i>	Area under the ROC curve (AUC)	0.845	0.7092	0.7061	0.7502	0.9348	0.9224
	Asymptotic standard error	0.04106	0.05178	0.0744	0.0498	0.0263	0.0301
	P (one-tailed)	2.7×10^{-9}	3.1×10^{-4}	0.0144	1.6×10^{-5}	6.7×10^{-14}	3.3×10^{-13}
<i>Campylorhynchus brunneicapillus</i>	Area under the ROC curve (AUC)	0.9144	0.9614	0.8405	0.9097	0.9250	0.9456
	Asymptotic standard error	0.02769	0.0166	0.0534	0.0316	0.0284	0.0243
	P (one-tailed)	1.2×10^{-12}	2.5×10^{-15}	6.4×10^{-6}	2.1×10^{-12}	4.2×10^{-13}	5.6×10^{-14}
<i>Melanerpes chrysogenys</i>	Area under the ROC curve (AUC)	0.8684	0.9434	0.2380	0.9072	0.9624	0.1354
	Asymptotic standard error	0.0376	0.0213	0.1299	0.0308	0.0178	0.0852
	P (one-tailed)	2.2×10^{-10}	2.1×10^{-14}	0.2151	2.3×10^{-12}	2.2×10^{-15}	0.2113
<i>Myadestes occidentalis</i>	Area under the ROC curve (AUC)	0.7820	0.6192	0.7018	0.8898	0.9468	0.9156
	Asymptotic standard error	0.0459	0.0561	0.0655	0.0351	0.0198	0.0295
	P (one-tailed)	1.2×10^{-6}	0.0399	5.6×10^{-3}	1.8×10^{-11}	1.4×10^{-14}	7.9×10^{-13}
<i>Mitrephanes phaeocercus</i>	Area under the ROC curve (AUC)	0.7373	0.6546	0.5611	0.8671	0.9363	0.6622
	Asymptotic standard error	0.0517	0.0548	0.0752	0.0382	0.0231	0.0558
	P (one-tailed)	4.7×10^{-5}	8.0×10^{-3}	0.5183	3.1×10^{-10}	7.4×10^{-14}	5.4×10^{-3}
<i>Ortalis vetula</i>	Area under the ROC curve (AUC)	0.9276	0.9334	0.8708	0.8624	0.9814	0.9480
	Asymptotic standard error	0.0270	0.0228	0.0630	0.0397	0.0105	0.0231
	P (one-tailed)	1.7×10^{-13}	8.1×10^{-14}	3.7×10^{-3}	4.2×10^{-10}	1.5×10^{-16}	1.2×10^{-14}
<i>Sittasomus griseicapillus</i>	Area under the ROC curve (AUC)	0.9358	0.9062	0.7281	0.9280	0.936	0.9888
	Asymptotic standard error	0.0255	0.0279	0.0804	0.0289	0.0241	0.0074
	P (one-tailed)	5.9×10^{-14}	2.5×10^{-12}	0.0167	1.6×10^{-13}	6.1×10^{-13}	3.6×10^{-17}
<i>Tyrannus crassirostris</i>	Area under the ROC curve (AUC)	0.8304	0.8446	0.7129	0.8550	0.9572	0.8182
	Asymptotic standard error	0.0399	0.0385	0.0935	0.0403	0.0197	0.0448
	P (one-tailed)	1.2×10^{-8}	2.9×10^{-9}	0.0217	9.5×10^{-10}	3.3×10^{-15}	4.2×10^{-8}

considered: these factors act to limit a species' distribution to an area smaller than that in which its ecological needs (e.g., climate, land-cover type, etc.) are met (Soberón and Peterson, 2005). If absence data were available, they could represent absence for reasons of unsuitable ecological conditions, or they could represent absence for reasons of history. In studies at relatively small spatial scales, the latter could perhaps be neglected. However, in applications at scales that include significant topographic and historical complexity (e.g., valleys isolated by mountain ranges, lowland areas separated by rivers, etc.), the latter cause cannot be neglected.

If some absences do not result from ecological causation, then use of absence information to validate models is perilous. A model predicting presence where an absence point is found may actually be correct—the reason being that these procedures are modeling ecological niches, and not geographic distributions (Soberón and Peterson, 2005). The perfect demonstration of this concept is that of species' invasions: species frequently can establish and maintain populations in regions in which they are presently absent, because their ecological niches are nonetheless represented there (Peterson, 2003). Hence, even if absence data were available, their use for model validation at coarse spatial scales would be ill-advised.

GARP and MaxEnt stood out among the 6 algorithms tested—under both sets of validation statistics, as they were the most significant predictions, showing no failures (unlike other approaches). MaxEnt besides generating accurate models, provides an output which identifies, for instance, the role of each environmental variable in the prediction model. On the other hand, GARP unites the best-subsets characteristics of favoring full prediction of test points over micro-prediction of a core set in a much-reduced predicted area. We suspect that this conclusion may be at least in part a consequence of not having absence data available for model evaluations, for reasons explained above. This paper complements a previous analysis (Elith et al., 2006) in that it addresses 3 techniques not included in that study.

Acknowledgments

We thank E. Martínez-Meyer for his assistance at several points with technical issues related to implementation of the tests reported herein. We also thank Ricardo Scachetti-Pereira for assistance with implementation of single algorithms. Mary Wisz provided helpful comments and reflections on an early version of the manuscript. Funding for this study was provided by the U.S. National Science Foundation.

Literature cited

- Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162:211-232.
- Austin, M. P., A. O. Nicholls, and C. R. Margules. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* 60:161-177.
- Brotons, L. W. Thuiller, M. B. Araujo, and A. H. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27:437-448.
- Carpenter, G., A. N. Gillison, and J. Winter. 1993. Domain: a flexible modeling procedure for mapping potential distributions of animals and plants. *Biodiversity and Conservation* 2:667-680.
- Chefaoui, R. M., J. Hortal, and J. M. Lobo. 2005. Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. *Biological Conservation* 122:327-338.
- Chen, G. and A. T. Peterson. 2002. Prioritization of areas in China for the conservation of endangered birds using modelled geographical distributions. *Bird Conservation International* 12:197-209.
- CONABIO. 2002. Red Mexicana de la Información de la Biodiversidad. Comisión Nacional para el Uso y Conocimiento de la Biodiversidad, Mexico City.
- Cumming, G. S. 2000. Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography* 27:441-455.
- Elith, J. and M. Burgman. 2002. Predictions and their validation: Rare plants in the Central Highlands, Victoria. *In* Predicting species occurrences: issues of scale and accuracy, J. M. Scott, P. J. Heglund, and M. L. Morrison (eds.). Island Press, Washington, D.C.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151.
- Fielding, A. H. and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38-49.
- Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and application in biodiversity analysis. *Trends in Ecology and Evolution* 19:497-503.
- Guisan, A. and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993-1009.
- Hirzel, A. H., G. Le Lay, V. Helfer, C. Radin, and A. Guisan.

2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199:142-152.
- Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83:2027-2036.
- Jones, P. G., and A. Gladkov. 1999. FloraMap, version 1: A computer tool for predicting the distribution of plants and other organisms in the wild. Centro Internacional de Agricultura Tropical, Cali, Colombia.
- Manel, S., J. M. Dias, S. T. Buckton, and S. J. Ormerod. 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology* 36:734-747.
- Manel, S., J. M. Dias, and S. J. Ormerod. 1999. Comparing discriminant analysis, neural networks, and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120:337-347.
- Nix, H. A. 1986. A biogeographic analysis of Australian elapid snakes. *In* Atlas of elapid snakes of Australia, R. Longmore (ed.). Australian Government Publishing Service, Canberra. p. 4-15.
- Pearson R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34:102-117.
- Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78:419-433.
- Peterson, A. T., J. Soberón, and V. Sánchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285:1265-1267.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259.
- Phillips, S. J. and M. Dudík. 2004. A maximum entropy approach to species distribution modeling. *Proceedings of the 21st International Conference on Machine Learning*, Baniff, Canada.
- Pulliam, R. H. 2000. On the relationship between niche and distribution. *Ecology Letters* 3:349-361.
- Raines, G. L., G. F. Bonham-Carter, and L. Kemp. 2000. Predictive probabilistic modeling using ArcView GIS. *ArcUser* April-June: 45-48.
- Ridgely, R. S., T. F. Allnut, T. Brooks, D. K. McNicol, D. W. Mehlman, B. E. Young, and J. R. Zook. 2005. Digital distribution maps of the birds of the western hemisphere, version 2.1. NatureServe, Arlington, Virginia.
- Rzedowski, J. 1978. *Vegetación de México*. Limusa, México, D.F. 432 p.
- Sánchez-Cordero, V. and E. Martínez-Meyer. 2000. Museum specimen data predict crop damage by tropical rodents. *Proceedings of the National Academy of Sciences USA* 97:7074-7077.
- Scott, J. M., P. J. Heglund, and M. L. Morrison. 2002. Predicting species occurrences: issues of accuracy and scale. Island, Washington, D.C.
- Soberón, J. and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2:1-10.
- Stockwell, D. R. B. 1999. Genetic algorithms II. *In* Machine learning methods for ecological applications, A. H. Fielding (ed.). Kluwer Academic, Boston. p. 123-144.
- Stockwell, D. R. B. and I. R. Noble. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of analysis. *Mathematics and Computers in Simulation* 33:385-390.
- Stockwell, D. R. B. and D. P. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Systems* 13:143-158.
- Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. Ferreira de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. Van Jaarsveld, G. E. Midgely, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips, and S. E. Williams. 2004. Extinction risk from climate change. *Nature* 427:145-148.
- Tsoar, A. O. Allouche, O. Steinitz, D. Rotem, and R. Kadmon. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and distributions* 13:397-405.
- Wintle, B. A., J. Elith, and J. M. Potts. 2005. Fauna modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecology* 30:719-738.
- Zaniewski, A. E., A. Lehmann, and J. M. Overton. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* 157:261-280.