RESEARCH ARTICLE

An evaluation of methods for modelling distribution of Patagonian insects

Una evaluación de los métodos para modelizar la distribución de insectos patagónicos

MARCELO F. TOGNELLI[1], SERGIO A. ROIG-JUÑENT[1], ADRIANA E. MARVALDI[1], GUSTAVO E. FLORES[1]
& JORGE M. LOBO[2,*]

[1] Laboratorio de Entomología, IADIZA-CONICET, CCT-Mendoza, CC 507, 5500 Mendoza, Argentina
[2] Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC)
c/ José Gutiérrez Abascal 2, 28006 Madrid, Spain
* Corresponding author: mcnj117@mncn.csic.es

ABSTRACT

Various studies have shown that model performance may vary depending on the species being modelled, the study area, or the number of sampled localities, and suggest that it is necessary to assess which model is better for a particular situation. Thus, in this study we evaluate the performance of different techniques for modelling the distribution of Patagonian insects. We applied eight of the most widely used modelling methods (artificial neural networks, BIOCLIM, classification and regression trees, DOMAIN, generalized additive models, GARP, generalized linear models, and Maxent) to the distribution of ten Patagonian insect species. We compared model performance with five accuracy measures. To overcome the problem of not having reliable absence data with which to evaluate model performance, we used randomly selected pseudo-absences located outside of the polygon area defined by taxonomic experts. Our analyses show significant differences among modelling methods depending on the chosen accuracy measure. Maxent performed the best according to four out of the five accuracy measures, although its accuracy did not differ significantly from that obtained with artificial neural networks. When assessed on per species basis, Maxent was also one of the strongest performing methods, particularly for species sampled from a relatively low number of localities. Overall, our study identified four groups of modelling techniques based on model performance. The top-performing group is composed of Maxent and artificial neural networks, followed closely by the DOMAIN technique. The third group includes GARP, GAM, GLM, and CART, and the fourth best performer is the BIOCLIM technique. Although these results may allow obtaining better distributional predictions for reserve selection, it is necessary to be cautious in their use due to the provisional nature of these simulations.

Key words: Expert opinion, model evaluation, Patagonia, pseudo-absence data, species distribution modelling.

RESUMEN

Varios estudios han mostrado que el desempeño de los modelos de distribución puede variar dependiendo de la especie modelizada, el área de estudio o el número de localidades de presencia utilizadas, sugiriendo que es necesario evaluar cuál es la mejor técnica de modelización en cada situación concreta. En este estudio evaluamos distintas técnicas de modelización para la distribución de los insectos patagónicos. Hemos aplicado ocho de los métodos más ampliamente usados (redes neuronales, BIOCLIM, árboles de clasificación y regresión, DOMAIN, Modelos Aditivos Generalizados, GARP, Modelos Lineares Generalizados y Maxent) a los datos de distribución de diez especies de insectos patagónicos, comparando su efectividad mediante cinco medidas diferentes. Para evitar el problema de la carencia de datos de ausencia fiables con los que evaluar los modelos, hemos utilizado pseudoausencias seleccionadas al azar fuera de un área poligonal definida por taxónomos expertos. Nuestros análisis muestran diferencias significativas entre los distintos métodos de modelización dependiendo de la medida de validación utilizada. Maxent es el método que ofrece mejores resultados para cuatro de las cinco medidas de validación utilizadas, aunque su precisión no difiere de la obtenida con redes neuronales. Cuando se examina la efectividad para cada una de las especies, Maxent resultó también uno de los métodos más fiables, especialmente en el caso de aquellas especies con un pequeño número de localidades. En conjunto, este estudio identifica cuatro grupos de técnicas de modelización. El de

mayor fiabilidad sería el compuesto por Maxent y las redes neuronales, seguido de cerca por DOMAIN. El tercer grupo incluiría GARP, GAM, GLM y CART, mientras que el cuarto estaría formado por BIOCLIM. Aunque estos resultados pueden permitir obtener mejores predicciones de distribución con capacidad para ser utilizadas en la planificación de reservas, es necesario ser cauto en su utilización debido a la naturaleza provisional de estas simulaciones.

INTRODUCTION

One of the central problems in ecology is understanding how organisms are distributed on earth. In the absence of a complete inventory of where species occur, predictive models of species distribution are an alternative that is increasingly being explored to produce detailed distribution and habitat suitability maps. Species distribution models examine associations between general environmental characteristics and the known occurrences of a particular species (Guisan & Zimmermann 2000, Scott et al. 2002, Guisan & Thuiller 2005). These models allow us to project the geographic distribution of a species into unexplored regions, or into scenarios of future or past climatic conditions. In recent years, this approach has been widely applied to address issues in ecology (Anderson et al. 2003, Vetaas 2002), biogeography (Coudun et al. 2006, Luoto et al. 2006), evolution (Peterson et al. 1999, Graham et al. 2004, Martínez-Meyer & Peterson 2006), conservation biology (Ferrier 2002, Araújo et al. 2004, Cabeza et al. 2004; Sanchez-Cordero et al. 2005), species invasion (Peterson 2005, Fitzpatrick et al. 2006), and the effects of climate change on species (Skov & Svenning 2004, Thomas et al. 2004, Thuiller 2004, Araújo et al. 2006, Thuiller et al. 2006). However, many studies have not expressly addressed a very important part of the modelling process: selecting the most appropriate modelling method to address a particular question (Pearson et al. 2006).

Numerous modelling methods and tools have been developed in the last decade (Guisan & Thuiller 2005, Elith et al. 2006). These can be roughly divided into those that only use records of species presence (e.g., bioclimatic envelopes and distance-based measures) and those that use presence-absence data (e.g., general linear and additive models and decision trees; Guisan & Zimmermann 2000). As absence data are frequently lacking, an alternative to presence-absence techniques involves generating "pseudo-absences" from background areas (Zaniewski et al. 2002, Engler et al. 2004). Generally, pseudo-absences are randomly selected from the background environment (Stockwell & Peterson 2002) or by using environmentally weighted random sampling (Zaniewski et al. 2002, Engler et al. 2004). The comparative performance of different modelling methods has been assessed with varying results (Muñoz & Felicísimo 2004, Segurado & Araújo 2004, Elith et al. 2006, Hernandez et al. 2006, Pearson et al. 2006). Overall, results show that model performance varies according to the taxa being modelled, the studied region, and the sample size (i.e., number of locality records). Thus, before generating predictive distribution models for a specific taxonomic group and region, it is necessary to first examine the comparative performance of the different possible techniques.

Predictive models of species distribution have been generally developed for plants and vertebrate species (see Guisan & Thuiller 2005 and references therein). Only a few studies have modelled the potential distribution of insect species, and these have dealt mostly with vectors of human diseases (Komar et al. 2005, López-Cárdenas et al. 2005, Peterson et al. 2005) and introduced species (Roura-Pascual et al. 2005, Fitzpatrick et al. 2006). Although arthropod species comprise approximately 80 % of total species (Rupert et al. 2004), the lack of reliable taxonomic and distributional information for insects has prevented the use of species distribution models for conservation purposes (but see Meggs et al. 2004, Chefaoui et al. 2005). As conservation prioritization is generally based on higher-level taxa (i.e., vertebrates), although they may not be effective surrogates for invertebrates, our general aim is obtaining guidelines for the further use of

distribution models in the design of conservation plans for Patagonian insects. Using the data of ten species of insects belonging to different families of southern South America for which we only have presence data, we assessed here the comparative performance of eight of the most widely used modelling techniques. Predictions of profile presence-only methods and discrimination modelling techniques which use environmentally weighted pseudo-absences were compared. As absence information is unreliable, model predictions were evaluated using absence areas based on expert opinions.

METHODS

Species and climate data

We studied a region of southern South America extending from 46°15' to 75°66' W and 23°98' to 55°98' S. Species locality data included presence data for 10 insect species (Table 1). These species were selected because they are good representatives of their taxonomic groups in the studied region. The minimum number of presence localities for a species was 19, the maximum number was 94, and the median was 39.

A set of 25 climatic variables was initially considered (Table 2). Twenty-one climatic

variables were extracted from the WorldClim database (Hijmans et al. 2005), and four variables were extracted from the Climate Research Unit (New et al. 2002). The spatial resolution of environmental variables was 0.04° (approximately 4.6 x 4.6 km). These variables were standardized to eliminate measurement-scale effects (with a mean of 0 and a standard deviation of 1). To select variables that better represent the environment of the region, we used the so-called Jolliffe's principal component method (Rencher 2002). First, a Principal Component Analysis (PCA) was carried out including all variables, and five non-correlated factors with eigenvalues ≥ 1 were obtained that explained 87.13 % of the climatic variation across the region. For each one of the five PCA factors, the variable with the highest factor loadings (which measure the correlations between the original variables and the factor axes) was selected (> 0.8). The five selected variables were annual mean temperature, isothermality, mean diurnal range, precipitation during the driest month of the year, and precipitation during the wettest quarter of the year. Relative humidity was also included as a predictor variable because it was the only one that was not significantly correlated with any of the formerly mentioned PCA factors. In total, these six variables were considered to be the most representative of the climate in southern South America.

TABLE 1

List of species used for modelling including a unique identifier and the number of presence localities where they have been collected.

Lista de especies usadas para modelizar incluyendo un identificador único y el número de localidades de presencia donde fueron colectadas.

ID	Sample size	Species	Family
1	19	*Aegorhinus superciliosus* (Guerin)	Curculionidae
2	24	*Chirodamus kingii* Haliday	Pompilidae
3	30	*Cnemalobus obscurus* (Brullé)	Carabidae
4	31	*Migadops latus* (Guerin)	Carabidae
5	34	*Aegorhinus nodipennis* (Hope)	Curculionidae
6	44	*Mitragenius araneiformis* Curtis	Tenebrionidae
7	60	*Baripus clivinoides* Curtis	Carabidae
8	86	*Rhyephenes maillei* (Gay and Solier)	Curculionidae
9	89	*Epipedonota cristallisata* (Lacordaire)	Tenebrionidae
10	94	*Creobius eydouxii* Guérin-Ménéville	Carabidae

Predictor variables from which six variables
(marked with an asterisk) were selected (see
text for details) for developing models.
Numbers indicate the origin of the variables as
follows: 1 WorldClim database; 2 CRU
database.

Variables predictivas a partir de las cuales se
seleccionaron seis variables (marcadas con un asterisco)
para generar los modelos (ver texto para mayor detalle).
Los números indican el origen de las variables de la
siguiente manera: 1 base de datos WorldClim; 2 base de
datos CRU.

Predictor variable	Source
Annual mean temperature *	1
Annual precipitation	1
Frost days frequency	2
Isothermality *	1
Maximum annual temperature	1
Maximum temperature of warmest month	1
Mean diurnal range *	1
Mean temperature of coldest quarter	1
Mean temperature of driest quarter	1
Mean temperature of warmest quarter	1
Mean temperature of wettest quarter	1
Minimum annual temperature	1
Minimum temperature of coldest month	1
Precipitation of coldest quarter	1
Precipitation of driest month *	1
Precipitation of driest quarter	1
Precipitation of warmest quarter	1
Precipitation of wettest month	1
Precipitation of wettest quarter *	1
Precipitation seasonality	1
Relative humidity *	2
Sunshine duration	2
Temperature annual range	1
Temperature seasonality	1
Wind speed	2

Modelling methods and evaluation

We tested eight of the most commonly used
modelling methods. The type of data these
techniques use distinguishes them: those that use
presence-only data (BIOCLIM, DOMAIN,
GARP, and Maxent), and those that require
presence-absence data (Artificial Neural
Networks, Classification and Regression Trees,
Generalized Additive Models, and Generalized
Linear Models). Within the first group, two of the
methods (GARP and Maxent) use a sample of the
background environment as pseudo-absences.
BIOCLIM and DOMAIN were run with DIVA-
GIS (www.diva-gis.org) using default
specifications. GARP models were run using the
best subset (Anderson et al. 2003) module of
OpenModeller (www.openmodeller.sf.net), and
maximum entropy models were run with Maxent
(Phillips et al. 2006). Artificial Neural Networks
(ANN), Classification and Regression Trees
(CART), Generalized Additive Models (GAM),
and Generalized Linear Models (GLM) were
fitted in R (www.r-project.org) using the mgcv,
nnet and tree packages.

For each species, we randomly separated the
occurrence localities into 10 partitions. Each
partition was created by selecting 50 % of the
data to train the models, and the remaining 50 %
to test the models. For those modelling methods
that required presence-absence data to run, we
generated environmentally weighted pseudo-
absences (Lobo et al. 2006). These pseudo-
absences were selected calculating the maximum
and minimum scores of the six selected
environmental variables for the presence
localities of each species. Then, we calculated
the multidimensional envelope defined by the
scores of the localities in which the species was
recorded. Pseudo-absences were then selected
outside the environmentally suitable region (10
times the number of training presences).

Several discrimination indices are
frequently used to test model performance
(Fielding & Bell 1997). These are usually
derived from the two by two confusion matrix,
which describes the frequency of correctly and
incorrectly predicted data from the known
presences and absences. However, when
reliable absence data are lacking and pseudo-
absences selected across environmentally
unsuitable regions are used, model absence
predictions should not be validated using these
pseudo-absences; high absence success rates
would only indicate successful forecasting of
the locations under unfavourable environmental
regions. Thus, when only presence data are
available, commonly applied indices, such as
Kappa and Area Under the Curve (AUC) of the
ROC plot, cannot be calculated. In this
instance, to overcome the problem of not
having reliable absence data with which to
evaluate model performance, we used the
taxonomic expertise of some of the authors

(SARJ, AEM, and GEF), who are specialists in the taxa being modelled. These experts defined a polygon area based on their collecting experience and knowledge of where the species might occur, including those probable areas inhabited by the species for which no presence information exist. We then generated 10,000 random pseudo-absences (validation pseudo-absences or VSA) outside the polygon area defined by the taxonomic experts to calculate performance evaluation indices.

To assess the agreement between the presence-VSA data and the predictions obtained by the different modelling methods, we calculated the AUC, maximum Kappa value, and the true skill statistic (TSS; Allouche et al. 2006). Area Under the ROC Curve (AUC) is one of the most widely used approaches to evaluate model performance (Fielding & Bell 1997) whose results should not be used in species' model comparisons when the ratios between the extent of occurrence and the whole extent of the territory under study differ (Lobo et al. 2008). In our case, we have decided to maintain AUC results because all the species considered differ slightly in their occurrence area, and the analyzed territory is the same. AUC measures the ability of a model to discriminate between sites where a species is present and sites where a species is absent. Values of AUC range from 0.5 for models with predictive discrimination abilities no better than random to 1 for models with perfect predictive ability (Fielding & Bell 1997). The Kappa statistic measures the proportion of correctly predicted sites after accounting for the probability of chance agreement (Moisen & Frescino 2002). It requires a suitability cut-off threshold, which is generally arbitrarily selected. Alternatively, one can choose the maximum value for the Kappa score obtained from varying the threshold from 0 to 1 (Guisan et al. 1998). We calculated and used this maximum score (max Kappa) for each modelling method. The True Skill Statistic (TSS) has been recently introduced to ecology as an alternative measure of model accuracy (Allouche et al. 2006). In addition to having the advantages of Kappa (i.e., takes into account omission and commission errors and corrects for chance agreement), TSS also does not depend on prevalence (Allouche et al. 2006). It is

generally used in weather forecasting and compares the number of correctly classified forecasts, excluding those attributable to random guessing, to that of a hypothetical set of perfect forecasts (Allouche et al. 2006). TSS ranges from -1 to +1, where values of 0 or less indicate a model performance no better than random, and a value of +1 indicates perfect performance (Allouche et al. 2006). In addition to these three accuracy measures, we calculated sensitivity (the proportion of true positives correctly predicted) and specificity (the proportion of true negatives correctly predicted). To calculate these two last accuracy measures we used the point of the ROC curve where the sum of sensitivity and specificity is maximized as a cut-off criterion to convert continuous model predictions to binary classifications (presence/absence). This threshold has the advantage of giving equal weights to the probability of success of both presences and absences (Manel et al. 2001). It is one of the most appropriate methods to correctly derive a binary variable from continuous probabilities when species presence-absence distribution data are unbalanced (Liu et al. 2005, Jiménez-Valverde & Lobo 2006).

We used the Friedman test, which is a non-parametric version of the one-way repeated measures ANOVA (Sprent & Smeeton 2001), to test for differences in modelling performance among the different modelling methods. We then performed post hoc tests of multiple comparisons (Dunn test) to determine which methods differed from each other. In addition, we assessed model performance on a per species basis and evaluated effects due to sample sizes Models for species 1 to 5 were trained with < 18 sites, whereas models for species 6 to 10 were trained with 22 to 47 sites (Table 1).

<center>RESULTS</center>

Comparisons across modelling methods

Performance differed significantly among modelling methods for all accuracy measures considered (AUC, max Kappa, TSS, sensitivity, and specificity; Table 3). In general, the rank order of model performance varied considerably

for each measure assessed (Fig. 1a-e). However, multiple comparisons tests revealed that not all rank-order differences were significant (Table 4). Overall, we found that three methods, ANN, DOMAIN and Maxent, consistently performed better than the remaining techniques in four of the five accuracy measures. However, when it came to predict percentage of absence evaluation points, they lagged behind BIOCLIM in accuracy (Fig. 1E; Table 4). GARP was the next best ranked method, performing very well in three out five accuracy measures. The remaining modelling methods were less consistent in their performance across accuracy measures, and also showed higher variability within accuracy measures.

Comparisons across species

In general, model performance at the species level showed similar trends to the pooled species models. ANN, DOMAIN, and Maxent performed very well on a per-species basis for all accuracy measures, except for Specificity, followed by GARP, which performed well in three of the five accuracy measures (Fig. 2A-E).

To assess how model performance varied with sample size (i.e., the number of presence records), we plotted the median and interquantile range values for each species, and for each modelling method (Fig. 3). In general, there was not a clear trend between sample size and model performance. Only a few modelling methods showed a trend for some, but not all, of the accuracy measures. For instance, BIOCLIM and GAM improved their performance with larger sample sizes when assessed by AUC, TSS, and Sensitivity, whereas GARP showed this trend for AUC, TSS, and Specificity (Fig. 3A-E). ANN, DOMAIN, Maxent, and GARP showed relative low within and across species variation when assessed by AUC, TSS, and Sensitivity.

The lack of a clear trend between sample size and model performance is also reinforced by correlation analyses between sample size and each accuracy measure. Although significant, Spearman's rank correlations were very low (Table 5), and showed only a slight increase in the values of performance measures with sample size, with the exception of specificity, which decreased with sample size.
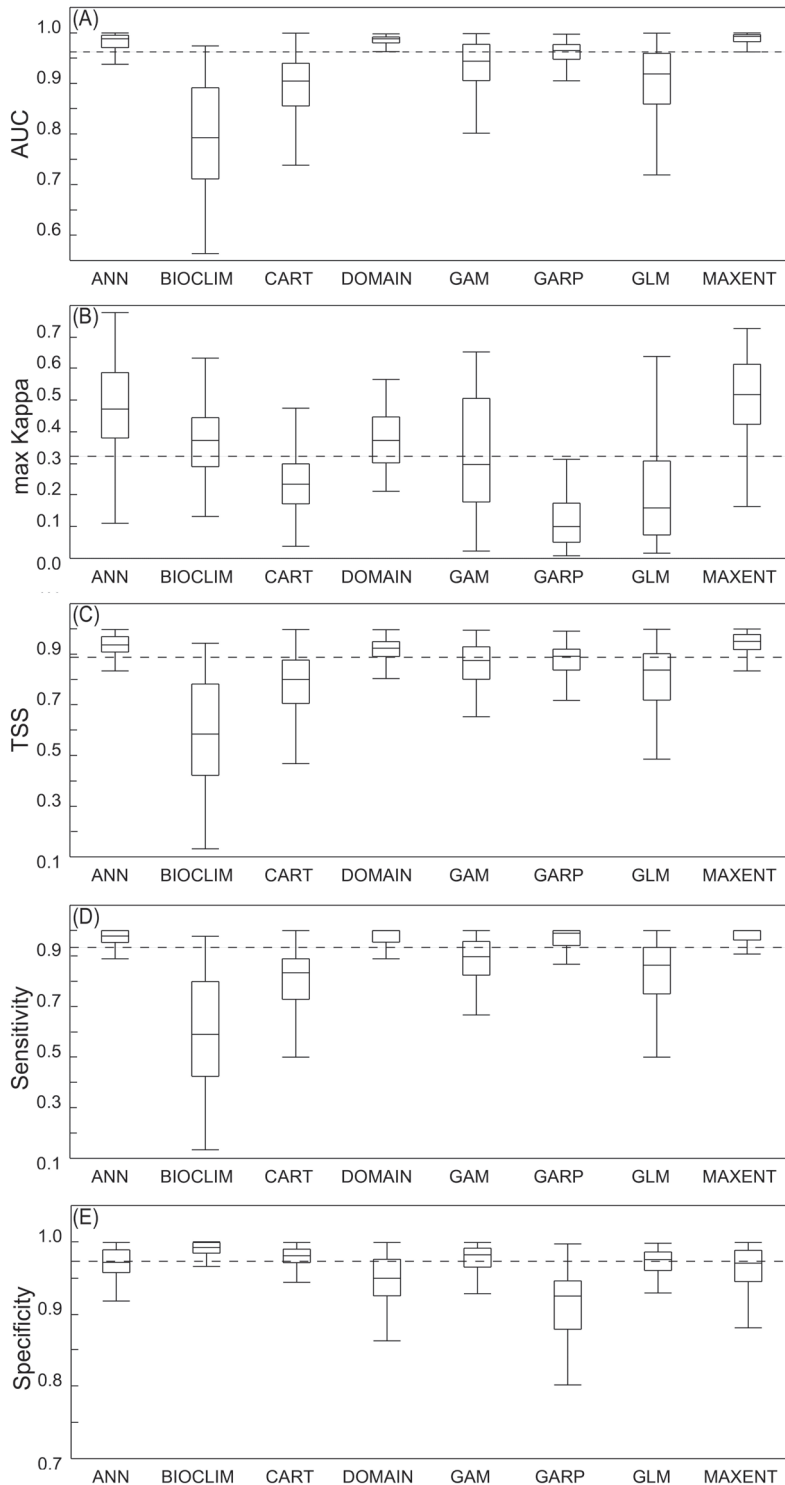
TABLE 3

Results of the Friedman test for differences among performance measures of the different modelling methods.

Resultados de la prueba de Friedman para establecer diferencias entre las medidas de precisión de los distintos métodos de modelización.

Performance measure	F	P level
AUC	460.2	< 0.0001
max Kappa	395.5	< 0.0001
TSS	388.6	< 0.0001
Sensitivity	452.4	< 0.0001
Specificity	300.4	< 0.0001

TABLE 4

Multiple comparisons of modelling methods' performance for each of the five accuracy measures (post hoc tests of multiple comparisons; Dunn test). P < 0.05 indicates significant differences. Accuracy measures of modelling methods represented with the same letter do not significantly differ.

Comparaciones múltiples del desempeño de los métodos de modelización para cada una de las cinco medidas de precisión usada (prueba a posteriori de comparaciones múltiples; Dunn test). P < 0.05 indica diferencias significativas. Medidas de precisión de los métodos de modelización representados por la misma letra no difieren significativamente.

	AUC	Kappa	TSS	Sensitivity	Specificity
ANN	a	a	a	a	b
BIOCLIM	d	b	d	d	a
CART	c	c	c	c	b
DOMAIN	a	b	a	a	c
GAM	b	b	b	b	b
GARP	b	e	b	a	d
GLM	c	c	b c	b c	b
MAXENT	a	a	a	a	b

Fig. 1: Box plots displaying the median, interquantile range, and maximum and minimum values of accuracy measures (AUC, max Kappa, TSS, Sensitivity, and Specificity) for each modelling method. Dashed horizontal line indicates the grand median. Note that Y axes have different scales and range.
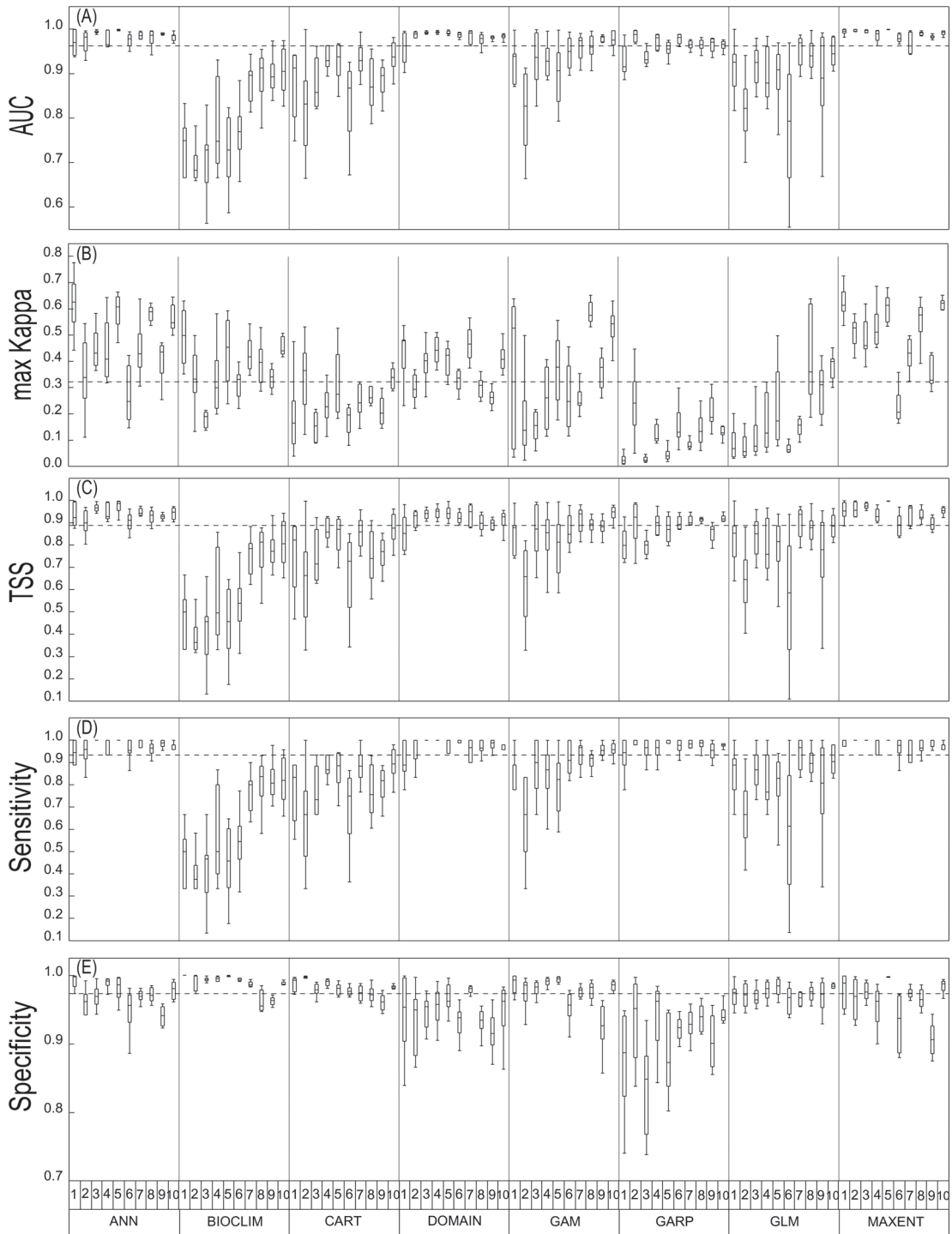
Gráficos mostrando la mediana, rango intercuartil y los valores máximos y mínimos de las medidas de precisión (AUC, max Kappa, TSS, Sensitivity, y Specificity) para cada método de modelización. La línea punteada horizontal indica la gran mediana. Notar que los ejes Y tienen diferentes escalas y rangos.
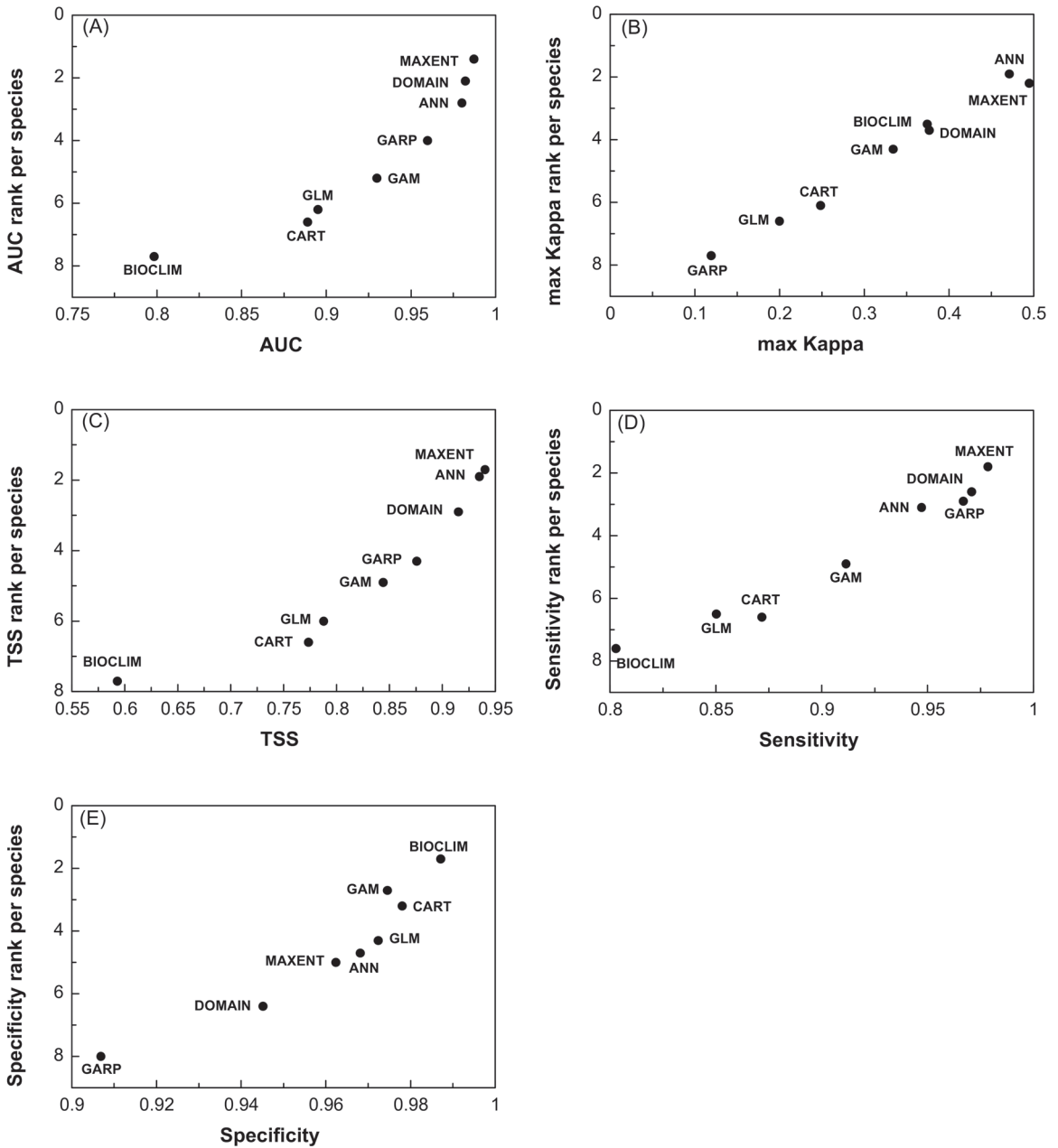
Fig. 2: Mean value of performance measure versus the rank of that measure when assessed on a per species basis. Note that the Y axis is reversed in all plots, so that methods with better performance are found in the upper right corner of the plots.

Valor promedio de la medida de desempeño versus el ranking de esa medida evaluada para cada especie. Notar que el eje Y está revertido en todos los gráficos para que los métodos con mejor desempeño se ubiquen en la esquina superior derecha del gráfico.

Fig. 3: Box plots displaying the median, interquantile range, and maximum and minimum values of accuracy measures (AUC, max Kappa, TSS, Sensitivity, and Specificity) for each modelling method and for each species. Species numbers are as in Table 1. Dashed horizontal line indicates the grand median. Note that Y axes have different scales and range.

Gráficos mostrando la mediana, rango intercuartil y los valores máximos y mínimos de las medidas de precisión (AUC, max Kappa, TSS, Sensitivity, y Specificity) para cada método de modelización y para cada especie. Los números de las especies corresponden a los de la tabla 1. La línea punteada horizontal indica la gran mediana. Notar que los ejes Y tienen diferentes escalas y rangos.

TABLE 5

Spearman's rank correlations (r_s) between sample size and the different performance measures.

Correlaciones por rangos de Spearman (r_s) entre tamaño de muestra y las distintas medidas de precisión.

Performance measure	r_s	P value
AUC	0.1092	< 0.01
max Kappa	0.1297	< 0.01
TSS	0.1290	< 0.01
Sensitivity	0.1014	< 0.01
Specificity	-0.2201	< 0.01

DISCUSSION

Previous studies have shown that model performance varies among different modelling methods, depending on the species modelled, the study area, and the sampling size (Manel et al. 1999, Dettmers et al. 2002, Moisen & Frescino 2002, Thuiller et al. 2003, Segurado & Araújo 2004, Elith et al. 2006, Hernandez et al. 2006, Pearson et al. 2006, Tsoar et al. 2007). Not surprisingly, our analyses also showed significant differences among predictions generated by various modelling techniques. However, since the goal of this study is to use predictive models of species distributions for prioritizing areas for conservation, it is crucial that we evaluate which modelling method is better for the taxa being studied and for our area of study. Although museum records combined with predictive models of species distribution have great potential value for the conservation of biodiversity (Graham et al. 2004), uncritical use of these models may misdirect conservation actions. Poor model choice may lead to misrouted conservation efforts (Loiselle et al. 2003, Johnson & Gillingham 2004).

Model performance not only differed among different modelling methods but also within methods used to evaluate different species. This variability in predictions makes it difficult to identify the 'best' modelling technique (Segurado & Araújo 2004, Pearson et al. 2006). It has been suggested that the best modelling methods are those that reduce the omission error rate (Anderson et al. 2003). The argument

is that predicting unsuitable habitat where a species is known to be present is a clear error, whereas predicting suitable habitat where there is no record of a species' presence may be due to insufficient sampling or other non-climatic factors that limit its distribution (Anderson et al. 2003, Pearson et al. 2006). This approach, however, may be adequate when predicting the ranges of invasive species that have not yet colonized all suitable environments (Peterson 2003). In contrast, Loiselle et al. (2003) suggest that if models are used to predict the distribution of a particular species for conservation purposes, they should not over-predict areas of distribution so that unsuitable sites for the protection of the species are included. Thus, model assessments should weigh the costs of making a false positive prediction versus the costs of making a false negative prediction depending on the intended use of the model (Fielding & Bell 1997). Multiple measures of model accuracy are needed to evaluate relative model performance, besides separately assessing their ability to predict presences and absences (Fielding 2002, Segurado & Araújo 2004, Bulluck et al. 2006, Hernandez et al. 2006).

In all, our results showed that Maxent was one of the best performing methods, as was ANN. Maxent ranked first in four out the five measures used to assess model accuracy; although, its performance was not significantly different from that of ANN. When assessed on a per species basis, Maxent outperformed the other methods in three out of the five accuracy measures. In addition to its strong performance, with the exception of max Kappa values, it was very stable across species with different sample sizes. Most importantly, Maxent performed well with small sample sizes, which constituted the majority of the species records in our database. These results support the findings of previous studies (Elith et al. 2006, Hernandez et al. 2006, Pearson et al. 2007) that compared several modelling techniques and found that Maxent was one of the best performing methods, even with sample sizes as low as five positive observations (Hernandez et al. 2006, Pearson et al. 2007).

ANN also performed well in this study. However, other researchers comparing the performance of different modelling methods have found contradictory results regarding the

accuracy of ANN. Whereas Segurado & Araújo (2004) found ANN to be the highest-performing method, Manel et al. (1999) found ANN model performance to be comparable to that of GLM. Unfortunately, in the most comprehensive model comparison study to date in which the predictive performance of 16 modelling techniques was assessed (Elith et al. 2006), ANN was not included. Besides performing slightly better, the advantage of Maxent over ANN is that the former method does not need absence data to generate a predictive distribution model. It has been suggested that the success of Maxent may be due to its regularization procedure that prevents or allows overfitting depending on whether a few or many training data points are used, respectively, to build the model (Dudik et al. 2004, Phillips et al. 2004, Phillips et al. 2006, Hernandez et al. 2006). Thus, given its strong performance and simplicity of use, we believe that Maxent is the best option for our larger project.

Another method that performed relatively well in this study was DOMAIN. Its performance was among the best, as judged by three of the five measures of accuracy. However, although it correctly predicted a high proportion of presences, it did not predict absences as well. A similar situation occurred with GARP, which had the lowest value for specificity. This method performed relatively well in three of the five accuracy measures, but it performed the worst according to the remaining two. In general, GARP performed somewhat better than GAM, and the latter outperformed GLM and CART. Previous model comparison studies have found GAM to perform better than GLM (Franklin 1998, Pearce & Ferrier 2000, Thuiller et al. 2003). Segurado & Araújo (2004) found CART to perform better than GLM, but not significantly better than GAM. Finally, although BIOCLIM correctly predicted a high proportion of the absences, in general, it was the poorest performing method, particularly for species with a low number of locality records. This finding is consistent with other studies in which BIOCLIM had a poor performance compared to other methods (Elith et al. 2006, Hernandez et al. 2006).

Overall, we can broadly identify four groups of modelling techniques in our study of model performance. First, the top-performing group is composed of Maxent and ANN, followed closely by the second group, DOMAIN. The third group includes GARP, GAM, GLM, and CART which is followed by BIOCLIM. These results should be viewed as recommendations only applicable to the taxa modelled (insects), and for the study area under consideration. Importantly, model techniques cannot replace good data when the purpose is to delimit the realized distributions of species (Lobo 2008). Distribution models are rarely used in the case of invertebrates, but this group of species harbour most of the recognized species and their data are rarely used for conservation purposes. Lack of data, survey bias, and restricted distributions are common characteristics in the case of insects and distribution simulations, although of provisional nature, are the only available procedure able to quickly provide relatively reliable distributional estimations that can be validated in the future.

However, the recommendations provided by our study should be considered with caution because some caveats can influence our results. First and importantly, modelling methods differ in their capacity to represent the potential-realized distribution gradient (Jiménez-Valverde et al. 2008) and comparisons among techniques must consider both this question and the quality of the data used as dependent variable. When reliable absences are lacking modelling exercises generate geographical representations located between the potential and realized distribution of species. The models obtained that require presence-absence data (i.e. ANN, CART, GAM, and GLM) were calculated using pseudo-absences outside the environmental domain of presences. This procedure artificially increase the statistical power of group discrimination techniques (Guisan & Zimmermann 2000, Graham et al. 2004), allowing obtaining geographic representations closer to the potential distribution of species (Jiménez-Valverde et al. 2008) and so, able to include the species in some not colonized localities. This can lead to high values of AUC and specificity but also high commission errors (Lobo 2008). Second, our models have been validated using a different kind of pseudo-absences previously defined by an expert taxonomist. Ideally,

validation data should be independent of the data used to generate the model (Fielding & Bell 1997), although this is practically impossible for Natural History Collection data (Graham et al. 2004). Our procedure is not exempt of error since the suggestions based on the knowledge and collecting experience of taxonomists can be biased (see Seoane et al. 2005). These drawbacks outline the fact that distributions models are simply hypotheses of the real distributions of the species and cannot replace a well designed field collection of species data.

ACKNOWLEDGEMENTS

LITERATURE CITED

ALLOUCHE O, A TSOAR & R KADMON (2006) Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43: 1223-1232.

ANDERSON RP, D LEW & AT PETERSON (2003) Evaluating predictive models of species' distributions: Criteria for selecting optimal models. Ecological Modelling 162: 211-232.

ARAÚJO MB, M CABEZA, W THUILLER, L HANNAH & PH WILLIAMS (2004) Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. Global Change Biology 10: 1618-1626.

ARAÚJO MB, W THUILLER & RG PEARSON (2006) Climate warming and the decline of amphibians and reptiles in Europe. Journal of Biogeography 33: 1712-1728.

BULLUCK L, E FLEISHMAN, C BETRUS & R BLAIR (2006) Spatial and temporal variations in species occurrence rate affect the accuracy of occurrence models. Global Ecology and Biogeography 15: 27-38.

CABEZA M, MB ARAÚJO, RJ WILSON, CD THOMAS, MJR COWLEY & A MOILANEN (2004) Combining probabilities of occurrence with spatial reserve design. Journal of Applied Ecology 41: 252-262.

CHEFAOUI RM, J HORTAL & JM 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.

COUDUN C, JC GEGOUT, C PIEDALLU & JC RAMEAU (2006) Soil nutritional factors improve models of plant species distribution: An illustration with *Acer campestre* (L.) in France. Journal of Biogeography 33: 1750-1763.

DETTMERS R, DA BUEHLER & JG BARLETT (2002) A test and comparison of wildlife-habitat modelling techniques for predicting bird occurrence at a regional scale. In: Scott JM, PJ Heglund, ML Morrison, JB Haufler, MG Raphael, WA Wall & FB Samson (eds) Predicting species occurrences: issues of accuracy and scale: 607-615. Island Press, Washington D.C.

ELITH J, CH GRAHAM, RP ANDERSON, M DUDIK, S FERRIER et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151.

ENGLER R, A GUISAN & L RECHSTEINER (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41: 263-274.

FERRIER S (2002) Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? Systematic Biology 51: 331-363.

FIELDING AH (2002) What are the appropriate characteristics of an accuracy measure? In: Scott JM, PJ Heglund, ML Morrison, JB Haufler, MG Raphael, WA Wall & FB Samson (eds) Predicting species occurrences: issues of accuracy and scale: 271-280. Island Press, Washington D.C.

FIELDING AH & JF BELL (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49.

FITZPATRICK MC, JF WELTZIN, NJ SANDERS & RR DUNN (2006) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? Global Ecology and Biogeography 16: 24-33.

FRANKLIN J (1998) Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. Journal of Vegetation Science 9: 733-748.

GRAHAM CH, S FERRIER, F HUETTMAN, C MORITZ & AT PETERSON (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology and Evolution 19: 497-503.

GUISAN A, JP THEURILLAT & F KIENAST (1998) Predicting the potential distribution of plant species in an Alpine environment. Journal of Vegetation Science 9: 65-74.

GUISAN A & W THUILLER (2005) Predicting species distribution: Offering more than simple habitat models. Ecology Letters 8: 993-1009.

GUISAN A & NE ZIMMERMANN (2000) Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

HERNANDEZ PA, CH GRAHAM, LL MASTER & DL ALBERT (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773-785.

HIJMANS RJ, SE CAMERON, JL PARRA, PG JONES & A JARVIS (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

JIMÉNEZ-VALVERDE A & JM LOBO (2006) The ghost of unbalanced species distribution data in geographical model predictions. Diversity and Distributions 12: 521-524.

JIMÉNEZ-VALVERDE A, JM LOBO & J HORTAL (2008) Not as good as they seem: the importance of concepts in species distribution modelling. Diversity and Distributions 14: 885-890.

JOHNSON CJ & MP GILLINGHAM (2005) An evaluation of mapped species distribution models used for conservation planning. Environmental Conservation 32: 117-128.

KOMAR O, BJ O'SHEA, AT PETERSON & AG NAVARRO-SIGÜENZA (2005) Evidence of latitudinal sexual segregation among migratory birds wintering in Mexico. Auk 122: 938-948.

LIU C, PM BERRY, TP DAWSON & RG PEARSON (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385-393.

LOBO JM (2008) More complex distribution models or more representative data? Biodiversity Informatics 5: 14-19.

LOBO JM, JR VERDÚ & C NUMA (2006) Environmental and geographical factors affecting the Iberian distribution of flightless *Jekelius* species (Coleoptera: Geotrupidae). Diversity and Distributions 12: 179-188.

LOBO JM, A JIMÉNEZ-VALVERDE & R REAL (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17: 145-151.

LOISELLE BA, CA HOWELL, CH GRAHAM, JM GOERCK, T BROOKS, KG SMITH, & PH WILLIAMS (2003) Avoiding Pitfalls of Using Species Distribution Models in Conservation Planning. Conservation Biology 17: 1591-1600.

LÓPEZ-CÁRDENAS J, FE GONZALEZ BRAVO, PM SALAZAR SCHETTINO, JC GALLAGA SOLORZANO, ER BARBA et al. (2005) Fine-scale predictions of distributions of Chagas disease vectors in the state of Guanajuato, Mexico. Journal of Medical Entomology 42: 1068-1081.

LUOTO M, RK HEIKKINEN, J POYRY & K SAARINEN (2006) Determinants of the biogeographical distribution of butterflies in boreal regions. Journal of Biogeography 33: 1764-1778.

MANEL S, JM DIAS, ST BUCKTON & SJ ORMEROD (1999) Alternative methods for predicting species distribution: An illustration with Himalayan river birds. Journal of Applied Ecology 36: 734-747.

MANEL S, HC WILLIAMS & SJ ORMEROD (2001) Evaluating presence-absence models in ecology: The need to account for prevalence. Journal of Applied Ecology 38: 921-931.

MARTÍNEZ-MEYER E & AT PETERSON (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. Journal of Biogeography 33: 1779-1789.

MEGGS JM, SA MUNKS, R CORKREY & K RICHARDS (2004) Development and evaluation of predictive habitat models to assist the conservation planning of a threatened lucanid beetle, *Hoplogonus simsoni*, in north-east Tasmania. Biological Conservation 118: 501-511.

MOISEN GG & TS FRESCINO (2002) Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling 157: 209-225.

MUÑOZ J & AM FELICÍSIMO (2004) Comparison of statistical methods commonly used in predictive modelling. Journal of Vegetation Science 15: 285-292.

NEW M, D LISTER, M HULME & I MAKIN (2002) A high-resolution data set of surface climate over global land areas. Climate Research 21: 1-25.

PEARCE J & S FERRIER (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecological Modelling 128: 127-147.

PEARSON RG, W THUILLER, MB ARAÚJO, E MARTÍNEZ-MEYER, L BROTONS et al. (2006) Model-based uncertainty in species range prediction. Journal of Biogeography 33: 1704-1711.

PEARSON RG, CJ RAXWORTHY, M NAKAMURA & AT PETERSON (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography 24: 102-117.

PETERSON AT (2005) Predicting potential geographic distributions of invading species. Current Science 89: 9.

PETERSON AT, C MARTÍNEZ-CAMPOS, Y NAKAZAWA & E MARTÍNEZ-MEYER (2005) Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. Transactions of the Royal Society of Tropical Medicine and Hygiene 99: 647-655.

PETERSON AT, J SOBERÓN & V SÁNCHEZ-CORDERO (1999) Conservatism of ecological niches in evolutionary time. Science 285: 1265-1267.

PHILLIPS SJ, RP ANDERSON & RE SCHAPIRE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231-259.

RENCHER, AC (2002) Methods of Multivariate Analysis. Second edition. John Wiley & Sons, New York. 715 pp.

ROURA-PASCUAL N, AV SUAREZ, C GÓMEZ, P PONS, Y TOUYAMA, AL WILD & AT PETERSON (2004) Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. Proceedings of the Royal Society of London B 271: 2527-2534.

RUPPERT E, RD BARNES & RS FOX (2004) Invertebrate Zoology. A functional evolutionary approach. Seventh Edition. Brooks/Cole Thompson Learning, Belmont, California. 989 pp.

SÁNCHEZ-CORDERO V, P ILLOLDI-RANGEL, M LINAJE, S SARKAR & AT PETERSON (2005) Deforestation and extant distributions of Mexican endemic mammals. Biological Conservation 126: 465-473.

SCOTT JM, PJ HEGLUND, ML MORRISON, JB HAUFLER, MG RAPHAEL, WA WALL & FB SAMSON (2002) Predicting species occurrences: issues of accuracy and scale. Island Press, Washington D.C. 868 pp.

SEGURADO P & MB ARAÚJO (2004) An evaluation of methods for modelling species distributions. Journal of Biogeography 31: 1555-1568.

SEOANE J, J BUSTAMANTE & R DÍAZ-DELGADO (2005) Effect of expert opinion on the predictive ability of environmental models of bird distribution. Conservation Biology 19: 512-522.

SKOV F & JC SVENNING (2004) Potential impact of climatic change on the distribution of forest herbs in Europe. Ecography 27: 366-380.

SPRENT P & NC SMEETON (2001) Applied nonparametric statistical methods. Texts in

statistical science. Third Edition. Chapman & Hall/CRC, Boca Raton, Florida. 461 pp.

STOCKWELL DRB & AT PETERSON (2002) Effects of sample size on accuracy of species distribution models. Ecological Modelling 148: 1-13.

THOMAS CD, A CAMERON, RE GREEN, M BAKKENES, LJ BEAUMONT et al. (2004) Extinction risk from climate change. Nature 427: 145-148.

THUILLER W (2004) Patterns and uncertainties of species' range shifts under climate change. Global Change Biology 10: 2020-2027.

THUILLER W, MB ARAÚJO & S LAVOREL (2003) Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. Journal of Vegetation Science 14: 669-680.

THUILLER W, S LAVOREL, MT SYKES & MB ARAÚJO (2006) Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. Diversity and Distributions 12: 49-60.

TSOAR A, O ALLOUCHE, O STEINITZ, D ROTEM & R KADMON (2007) A comparative evaluation of presence-only methods for modelling species distribution. Diversity and Distributions 13: 397-405.

VETAAS OR (2002) Realized and potential climate niches: A comparison of four *Rhododendron* tree species. Journal of Biogeography 29: 545-554.

ZANIEWSKI AE, A LEHMANN & JM OVERTON (2002) Predicting species spatial distributions using presence-only data: A case study of native New Zealand ferns. Ecological Modelling 157: 261-280.