Journal of Biogeography (J. Biogeogr.) (2004) **31**, 1555–1568

ORIGINAL ARTICLE

An evaluation of methods for modelling species distributions

Pedro Segurado[1] and Miguel B. Araújo[1,2]*

[1]*Unidade de Macroecologia e Conservação, Universidade de Évora, Estrada dos Leões Évora, Portugal and* [2]*Centre d'Ecologie Fonctionnelle et Evolutive, CNRS 1919 Route Mende 34293 Montpellier Cedex 5, France*

*Correspondence: Miguel B. Araújo, Unidade de Macroecologia e Conservação, Universidade de Évora, Estrada dos Leões 7000-730 Évora, Portugal. E-mail: mba@uevora.pt

ABSTRACT

Aim Various statistical techniques have been used to model species probabilities of occurrence in response to environmental conditions. This paper provides a comprehensive assessment of methods and investigates whether errors in model predictions are associated to specific kinds of geographical and environmental distributions of species.

Location Portugal, Western Europe.

Methods Probabilities of occurrence for 44 species of amphibians and reptiles in Portugal were modelled using seven modelling techniques: Gower metric, Ecological Niche Factor Analysis, classification trees, neural networks, generalized linear models, generalized additive models and spatial interpolators. Generalized linear and additive models were constructed with and without a term accounting for spatial autocorrelation. Model performance was measured using two methods: sensitivity and Kappa index. Species were grouped according to their spatial (area of occupancy and extent of occurrence) and environmental (marginality and tolerance) distributions. Two-way comparison tests were performed to detect significant interactions between models and species groups.

Results Interaction between model and species groups was significant for both sensitivity and Kappa index. This indicates that model performance varied for species with different geographical and environmental distributions. Artificial neural networks performed generally better, immediately followed by generalized additive models including a covariate term for spatial autocorrelation. Non-parametric methods were preferred to parametric approaches, especially when modelling distributions of species with a greater area of occupancy, a larger extent of occurrence, lower marginality and higher tolerance.

Main conclusions This is a first attempt to relate performance of modelling techniques with species spatial and environmental distributions. Results indicate a strong relationship between model performance and the kinds of species distributions being modelled. Some methods performed generally better, but no method was superior in all circumstances. A suggestion is made that choice of the appropriate method should be contingent on the goals and kinds of distributions being modelled.

Keywords

Classification and regression trees, conservation planning, Ecological Niche Factor Analysis, generalized additive models, generalized linear models, Gower metric, neural networks, Portugal, spatial interpolators.

INTRODUCTION

Models exploring the relationships between species' occurrences and sets of predictor variables produce two kinds of useful outputs. The first are estimates of the probability that species might occur at given unrecorded locations. The second are estimates of an area's suitability for species. The use of empirical models of occurrence in conservation planning has been increasingly advocated (e.g. Margules & Nicholls, 1987; Cocks & Baird, 1989; Araújo & Williams, 2000; Williams & Araújo, 2000, 2002; Polasky & Solow, 2001; Araújo *et al.*, 2004). For example, Polasky & Solow (2001) suggested that probabilities of occurrence could be used as attributes for species-based reserve selection. These probabilities would be regarded as estimates of the likelihood that a species might occur at a given unrecorded location. From another perspective, Araújo & Williams (2000) suggested that probabilities of occurrence could be interpreted as estimates of the probability that species might find suitable habitat in a given area. They also argued that suitability estimates from probabilistic models would be inversely correlated to probabilities of extinction in a near future. Observations of local extinctions in British breeding birds in relation to estimates of probability of occurrence have provided empirical support for this idea (e.g. Gates & Donald, 2000; Donald & Greenwood, 2001; Araújo *et al.*, 2002). Nevertheless the question remains as to what method to choose and whether the choice of modelling technique should depended on the goals, region or species being modelled.

There are many modelling techniques available to explore the correlation between occurrence of a species and sets of predictor variables and there is a possibility that techniques may differ in their ability to summarize useful relationships between response and predictor variables. But model performance may also be contingent on the goals, assumptions and data used for particular analysis. If the goal of a model is to predict occurrences of species outside their known range, interest could be focused on methods that optimize the overall fit, i.e. the balance between false positives (absent, but predicted to be present) and false negatives (present, but predicted to be absent) (Fielding & Bell, 1997). However, if the goal is to minimize unexplained variation by models (i.e. minimize false negatives), then a measure of performance would be required that accounted for the proportion of false negatives alone (e.g. Araújo & Williams, 2000).

Although some studies have compared the performance of different modelling techniques (Walker, 1990; Bio *et al.*, 1998; Manel *et al.*, 1999; Araújo & Williams, 2000; Pearce & Ferrier, 2000; Vayssières *et al.*, 2000; Vetaas, 2000; Thuiller *et al.*, 2003), none has, to our knowledge, investigated systematically how variation in species geographical and environmental distributions affect model' performance. Furthermore, the number of techniques compared has often been small (four or less techniques compared at once) and the number of species studied in such comparisons limited (Table 1). Here, the performance of seven techniques in modelling the occurrence of 44 reptile and amphibian species in Portugal was compared using nine model runs for each species. The aim was to investigate whether observed errors are consistently related to descriptions of spatial and environmental distributions of species. The initial idea was that the frequency and magnitude of errors in the models might not be independent of the kinds of distributions species have within geographical and environmental space. If the behaviour of error could be predicted from analysis of species distribution alone, then modellers could determine the circumstances where empirical models of occurrence were likely to provide adequate outputs. The results should also indicate which models would be more robust for specific sets of data, and when results should be generally less likely to be useful for specific conservation applications. This could provide users of models with a better understanding of model uncertainty and further insights into how to reduce it.

Table 1 Previous studies comparing model performance, number of species modelled, number and description of techniques used

Reference to paper	Number of species modelled	Number of models compared	Models used*
Walker (1990)	3	2	GLM, CART
Pereira & Itami (1991)	1	3	GLM, SI, BM
Lek *et al.* (1996)	1	2	GLM, NNETW
Bio *et al.* (1998)	156	2	GLM, GAM
Franklin (1998)	20	3	GLM, GAM, CART
Manel *et al.* (1999)	6	3	GLM, DA, NNETW
Mastrorillo *et al.* (1997)	3	2	NNETW, DA
Araújo & Williams (2000)	187	2	GLM, SI
Pearce & Ferrier (2000)	24	2	GLM, GAM
Vayssières *et al.* (2000)	3	2	GLM, CART
Hirzel *et al.* (2001)	1 (virtual)	2	EE, GLM
Dettmers *et al.* (2002)	6	4	GLM, DA, MD, CART
Elith & Burgman (2002)	8	4	GLM, GAM, BP, GARP
Fertig & Reiners (2002)	1	2	GLM, CART
Thuiller *et al.* (2003)	4	3	GLM, GAM, CART

* DA, discriminant analysis; BM, Bayesian Model; BP, bioclimatic profiles; CART, classification and regression trees; GAM, generalized additive models; GLM, generalized linear models; GARP, genetic algorithm for rule-set production; MD, mahalanobis distance method; NNETW, neural networks; SI, spatial interpolation.

DATA

Species data

Data included 9939 occurrence records for 44 species of reptiles and amphibians in Portugal (Table 2). Data were located in 993 Universal Transverse Mercator (UTM) 10 by 10 km grid cells, and were compiled from a recently updated Atlas of the Portuguese mainland herpetofauna (Godinho *et al.*, 1999). The minimum number of records for a species was four, the median number was 184, and the maximum number was 787.

Environmental and habitat data

Sixteen variables were compiled from various digital sources and treated as predictor variables in the models (Table 3). Environmental variables were compiled from available digital

Table 2 List of species modelled, prevalence, area of occupancy, extent of occurrence, marginality and tolerance

Species	Prevalence	Occupancy*	Extent of occurrence (km)	Marginality	Tolerance
Acanthodactylus erythrurus	0.062	62	595.06	0.589	0.624
Alytes cisternasii	0.271	270	590.93	0.75	0.713
Alytes obstetricans	0.198	198	446.54	0.896	0.699
Anguis fragilis	0.097	97	568.59	0.945	0.557
Blanus cinereus	0.126	126	608.36	0.617	0.754
Bufo bufo	0.541	540	630.71	0.317	0.922
Bufo calamita	0.275	274	617.41	0.45	0.872
Chaemaleo chaemaleo	0.022	22	104.40	1.016	0.172
Chalcides bedriagai	0.056	56	594.81	0.595	0.505
Chalcides striatus	0.166	166	613.27	0.579	0.868
Chioglossa lusitanica	0.207	207	304.63	1.025	0.538
Coluber hippocrepis	0.167	167	572.71	0.552	0.814
Coronella austriaca	0.031	31	272.03	0.939	0.314
Coronella girondica	0.136	136	604.15	0.607	0.81
Discoglossus galganoi	0.211	211	608.36	0.443	0.909
Elaphe scalaris	0.295	294	621.69	0.403	0.906
Emys orbicularis	0.065	65	610.98	0.706	0.584
Hemidactylus turcicus	0.031	31	252.39	0.779	0.378
Hyla arborea	0.156	156	568.59	0.479	0.78
Hyla meridionalis	0.117	117	416.77	0.739	0.632
Lacerta lepida	0.423	422	635.06	0.485	0.909
Lacerta monticula	0.004	4	14.14	0.756	0.102
Lacerta schreiberi	0.291	290	590.93	0.852	0.768
Macroprotodon cucullatus	0.061	61	542.31	0.751	0.484
Malpolon Monspessulanus	0.492	491	630.71	0.347	0.94
Mauremys leprosa	0.436	435	613.27	0.637	0.838
Natrix maura	0.492	491	622.41	0.401	0.934
Natrix natrix	0.192	192	622.41	0.607	0.848
Pelobates cultripes	0.245	245	599.33	0.656	0.826
Pelodytes punctatus	0.104	104	523.93	0.808	0.624
Pleurodeles waltl	0.220	220	554.71	0.686	0.776
Podarcis bocagei	0.078	78	583.10	1.003	0.516
Podarcis hispanica	0.368	367	626.10	0.572	0.909
Psammodromus algirus	0.591	590	630.71	0.325	0.943
Psammodromus hispanicus	0.141	141	604.15	0.582	0.804
Rana iberica	0.264	263	399.62	0.923	0.681
Rana perezi	0.788	786	626.50	0.27	0.939
Salamandra salamandra	0.484	483	621.69	0.445	0.928
Terentola mauritanica	0.176	176	581.81	0.678	0.816
Triturus boscai	0.369	368	617.41	0.546	0.881
Triturus helveticus	0.019	19	230.87	0.776	0.313
Triturus marmoratus	0.382	381	621.69	0.531	0.907
Vipera latastei	0.088	88	617.41	0.744	0.688
Vipera seoanei	0.014	14	136.01	0.812	0.148

* Number of 10 km grid squares.

Table 3 Predictor variables in the models

Variables	Units	Parameters	Type
Environmental			
Drainage	mm	Mode	Ordinal
Acidity	pH	Mode	Ordinal
Evapo-transpiration	mm	Mode	Ordinal
Humidity	%	Mode	Ordinal
Precipitation	mm	Mode	Ordinal
Solar radiation	kcal	Mode	Ordinal
Temperature	°C	Mode	Ordinal
Socio-economic			
Human population density	n/km^2	Mode	Ordinal
Land use			
Urban areas	–	% Area	Ordinal
Annual crops	–	% Area	Ordinal
Permanent crops	–	% Area	Ordinal
Pastures	–	% Area	Ordinal
Heterogeneous crops	–	% Area	Ordinal
Florests	–	% Area	Ordinal
Schrub lands	–	% Area	Ordinal
Bear soil	–	% Area	Ordinal

layers of the 'Atlas do Ambiente' (http://www.dga.pt). Habitat data included eight variables selected from the Portuguese EU CORINE land-cover data base (http://www.snig.cnig.pt). Human-population data were also compiled from published sources (CNA, 1983). Data were converted from available vector maps and resampled to raster UTM 10 by 10 km grid cell maps, using zonal functions in ArcView version 3.2 (ESRI, 1999) to calculate the mode of each variable in each cell.

MODELLING TECHNIQUES

There is a broad range of modelling techniques available to explore the correlation between response and predictor variables (Guisan & Zimmermann, 2000). These techniques include the Gower-similarity model (e.g. Carpenter *et al.*, 1993), Ecological Niche Factor Analysis (ENFA) (e.g. Hirzel *et al.*, 2001), classification trees (e.g. Breiman *et al.*, 1984; Clark & Pregibon, 1992), neural networks (e.g. Mastrorillo *et al.*, 1997; Manel *et al.*, 1999; Özesmi & Özesmi, 1999), generalized linear models (McCullagh & Nelder, 1989), generalized additive models (Hastie & Tibshirani, 1990), and spatial interpolation techniques (e.g. Bailey & Gatrell, 1995). Here the performance of seven modelling techniques was compared and applied to the same data. Nine models were obtained for each of the 44 reptiles and amphibian species. Overall, 396 models were compared. To make outputs from models comparable, the whole set of variables in all models was used (e.g. Fielding & Haworth, 1995).

Gower-similarity models

Modelling approaches based on similarities between data points only use presence data (ignoring absences) to create species' environmental envelopes. The Gower-similarity approach, as implemented in DOMAIN V1.3 for Windows (http://www.cifor.cgiar.org/domain/index.htm), was used. The DOMAIN algorithm assigns each cell in the output layer an average multivariate distance, termed the Gower metric, between that cell and the closest presence cell in the training set (Carpenter *et al.*, 1993). These distance values were rescaled from zero to one (value − minimum/maximum − minimum) to provide values comparable with that obtained with probability-based techniques.

Ecological Niche Factor Analysis

A related approach, using presence data alone, is based on ordination of data in a multivariate space of environmental variable (ENFA; Perrin, 1984). This technique is based on the computation of the factors explaining the major part of species environmental distribution. Extracted factors are uncorrelated and have biological significance: the first factor is the *marginality* factor, which describes how far the species optimum is from the mean environmental profile in the study area; the second is the *tolerance* factor, which is sorted by decreasing amount of explained variance and describe how specialized the species is by reference to the available range of environments in the study area (Hirzel, 2001). This approach was implemented using BIOMAPPER (Hirzel, 2001) software to produce habitat suitability maps. A Habitat Suitability Index (HSI) of each cell is a value inversely proportional to the weighted mean distance of the cell to the median of each ENFA factor. This value is normalized in such a way that the suitability index ranges from zero to one (Hirzel, 2001). The number of retained factors for computing habitat suitability maps were chosen using a 'broken stick advice' implemented in BIOMAPPER (Hirzel *et al.*, 2001). Here, the distribution of the eigenvalue of each factor is compared with the distribution of MacArthur's (1957) broken-stick, which is the expected distribution when breaking a stick randomly. The eigenvalues that are larger than expected according to the broken stick distribution may be considered 'significant'.

Classification trees

Classification trees (TREE; Breiman *et al.*, 1984; Clark & Pregibon, 1992) consist of recursive partitions of the dimensional space defined by the predictors into groups that are as homogeneous as possible in terms of the response. The tree is built by repeatedly splitting the data into two exclusive groups, defined by a simple rule based on a single explanatory variable at each step. Classification trees were fitted using the RPART library of tree routines developed by Therneau & Atkinson (1997, 2000) for S-Plus (Statistical Sciences, 1999). The RPART function generates pruned trees based on the results of cross-validations (Therneau & Atkinson, 1997, 2000). For each tree, a series of 50 10-fold cross-validations were run and the most frequently occurring tree size was chosen using the 1-SE rule (De'Ath & Fabricius, 2000).

Neural networks

Artificial neural networks were computed using the NNET library of S-Plus (NNETW; Venables & Ripley, 2002). NNETW is a feed-forward neural network, parameterized using seven hidden units in a single layer (selected by cross-validation), with a weight decay equal to 0.03. Given the heuristic nature of NNETW, each simulation gives slightly different results. For this reason, NNETW solutions were calculated 10 times for each species and the mean was used to provide predictions. The procedure described is commonly used for modelling presence/absence data using neural networks (e.g. Mastrorillo *et al.*, 1997; Manel *et al.*, 1999; Özesmi & Özesmi, 1999; Thuiller, 2003).

Generalized linear models

Generalized linear models (GLM) (McCullagh & Nelder, 1989) assuming a binomial error distribution – logistic regression – were adjusted using S-Plus functions and routine facilities (Statistical Sciences, 1999). In order to account for autocorrelation in the observations, models were also fitted in which contagion (see below: spatial interpolators) was included as an autocovariate term in the initial variable set (AGLM). These models are termed autologistic (Smith, 1994; Augustin *et al.*, 1996; Araújo & Williams, 2000).

Generalized additive models

To allow consideration of more complex response shapes than those possible through the linear responses of GLM, generalized additive models (GAM; Hastie & Tibshirani, 1990) were fitted using logit as the link function and binomial error distribution. As for GLM, a set of models were fitted that included an autocovariate term in the initial set of variables. Hence, two GAM models were obtained for each species (GAM and AGAM).

Spatial interpolators

Measures of aggregation for point and lattice data, such as Kernel estimation and nearest neighbour measures (e.g. Bailey & Gatrell, 1995), can be used to model species' probabilities of occurrence. This uses the idea of positive spatial autocorrelation (Legendre, 1993), in which the occurrence of a species in one area is expected to be more likely if the species occurs in many surrounding areas (Araújo & Williams, 2000, 2001; Araújo *et al.*, 2002). A measure of contagion (CONT) for each cell, based on a two-order neighbourhood, was used to estimate a distance-based probability of occurrence. Contagion is measured as a weighted average of the number of occupied grid-cells among a set of k_a neighbours of a central grid-cell y_a, so that

$$\text{Contagion} = \left(\frac{\sum_{b=1}^{k_a} w_{ab} y_b}{\sum_{b=1}^{k_a} w_{ab}} \right) \tag{1}$$

where the weight given to the grid-cell y_b is $w_{ab} = 1/d_{ab}$, and d_{ab} is the distance between grid-cells y_a and y_b. Two orders of neighbours, assigning a weight of $d = 1$ to the first-order and a weight of $d = 2$ to the second-order neighbours were used. Neighbours in the first order were the eight adjacent cells touching the central cell along the edges and at the corners within a rectangular grid. The second-order neighbours were the next group of cells concentric to the first order with 16 grid cells.

MEASURES OF MODEL PERFORMANCE

Models can be assessed qualitatively, or quantitatively. Qualitative assessments measure how well models fit the data, while quantitative assessments measure how well models predict real events (Myers, 1997). In most circumstances, model performance can only be estimated through qualitative assessments (but for example of quantitative assessment see Feria & Peterson, 2002). Hence, because models are optimized for the training set, the 'goodness-of-fit' to these data is likely to be an over-optimistic estimate of predictive skill outside the training set (Beutel *et al.*, 1999). Here model' performance was assessed with two measures: sensitivity and the Kappa statistic.

Sensitivity is based on the concept of true-presences misclassification (false negatives or type II error), and is calculated as 1 – percentage of false negatives (Fielding & Bell, 1997; Cumming, 2000). The number of false negatives is particularly useful because it measures the number of residuals, or amount of unexplained variation in the data; the greater the number of false negatives, the more models are likely to be unrealistic. The main problem using standard indexes of model performance, such as sensitivity, is that a cut-off level of probability has to be defined *a priori*. Some authors have used the questionable rule of thumb of a 0.5 cut-off probability (e.g. Austin *et al.*, 1996; Franklin, 1998; Manel *et al.*, 1999; Manel *et al.*, 2001), but this has been shown to be inadequate because variation in prevalence (i.e. the number of occurrences in relation to the number of samples) makes optimal cut offs for species vary (e.g. Manel *et al.*, 2001). In this paper sensitivity was based on the cut-off level for which the Cohen's Kappa was maximum. The Kappa statistic assesses the extent to which models predict occurrence at a rate higher than expected by chance (Monserud & Leemans, 1992). Hence, it was used both as a performance measure and as a criterion for cut-off level selection.

Since the goal of this study was to investigate variation of model performance in relation to attributes of species' distribution, the adoption of standard procedures across species to assess model performance was required. A bootstrap procedure (Efron & Tibshirani, 1993) was used to estimate the standard error of sensitivity as a measure of model stability (Guisan & Zimmermann, 2000). Ideally this should be performed in a training set checked against a validation set. However, the number of records for some species was so low that it was unfeasible to split the records into training and validation sets for a great proportion of them. An alternative would have been to exclude some species with restricted ranges from the analysis (e.g. Thuiller, 2003; Araújo *et al.*, 2004). The consequence would be that some of the most important species for conservation would be excluded. In order to avoid this problem, model performance was estimated using the training set alone. This represents a severe limitation for models trying to predict distributions outside known ranges. However for models trying to explain current occurrences and assign suitability scores to areas this limitation is less severe.

SPATIAL AND ENVIRONMENTAL DISTRIBUTIONS

When comparing the performance of models for large numbers of species it becomes difficult to make interpretations on a species-by-species basis. In such circumstances, it might be useful to group species, for example, according to their spatial or environmental distributions. These groupings are expected to mirror ecological and historical factors. Hence, they are likely to provide useful insights for *post-hoc* interpretation of results. Here spatial distribution of species was described with two measures: (1) area of occupancy, given by the number of 10 km grid cells where the species occur; and (2) extent of occurrence, given by the straight-line distance between the two most distant occupied grid cells (Gaston, 1996). Species-environmental distributions were described with two measures: (1) marginality (or niche position), which reflects how far the species optimum is from the mean environmental conditions in the study area; and (2) tolerance (or niche breath), which describes how variable the species association to environmental factors is with reference to the available range in the study area. These two environmental measures were calculated using the ENFA approach implemented in BIOMAPPER (Hirzel *et al.*, 2001).

Species were then grouped according to their distributional profile, using hierarchical agglomerative clustering (group average). The statistical significance of the resulting groups was tested with ANOSIM (analysis of similarity), which performs permutation tests for the null hypothesis that there are no assemblage differences between groups (Clarke & Gorley, 2001). Both cluster analysis and ANOSIM were performed using PRIMER for Windows, version 5.2.0 (Clarke & Gorley, 2001).

TESTING DIFFERENCES IN MODEL PERFORMANCE

Since the same measurement of model performance was calculated several times (different modelling procedures) on each subject (species), a repeated measures factorial design was used for testing differences in model performance. Differences between overall performances of each modelling procedure, assuming no interaction between model procedure and species groups, were tested using Friedman test, which is a non-parametric version of one-way repeated ANOVA measures (Sprent & Smeeton, 2001). Dunn tests (Klockars & Sax, 1986) were then performed for *post-hoc* multiple comparisons.

In order to test differences between groups of species, two-way repeated measures ANOVA-tests were also performed using the modelling procedure as the within-subject factor and the species group as the between-subject factor. The main goal of this test was to check for interactions between modelling procedures and species groups, that is, to test if distinct modelling procedures performed differently for distinct groups of species. Data were normalized using an arcsine transformation (Sokal & Rohlf, 1995). Tests were carried out independently for each partition of species and for each performance measure, i.e. for the Kappa index and sensitivity. The same analyses were carried out to test for differences in model stability, measured as the bootstrap standard error estimates, between species groups and modelling techniques.

RESULTS

Comparison of the overall model performance

The results of Friedman tests (Table 4) showed that there is significant effect of the modelling technique in model performance for both measures considered (Kappa index and sensitivity). These differences were lower for sensitivity. The rank order of model performance did not vary much with the measure used (Fig. 1a,b). On the other hand, not all rank order differences were significant according to multiple comparisons Dunn tests (Tables 5–7). Neural networks (NNETW) provided the best results as measured both with Kappa and sensitivity indices ($P < 0.001$) (Table 5). However, NNETW performance as measured with Kappa index was not significantly different to the results achieved with generalized additive models with a covariate term accounting for spatial autocorrelation (AGAM). AGAM also showed better performances in terms of Kappa index ($P < 0.001$, except for GAM where $P < 0.05$) than the remaining models. However, AGAM distribution of sensitivity was no different than GAM, without a covariate term accounting for spatial auto-correlation, generalized linear model (AGLM), classification tree analysis (TREE) and contagion (CONT). GAM, AGLM and TREE had similar performances showing no significant differences amongst them. The poorest performing methods in terms of Kappa index and sensitivity measurements were DOMAIN and BIOMAP. There were no significant differences of model performance among these methods. GLM also performed poorly, showing no significant differences from DOMAIN and BIOMAP in terms of the Kappa index, although being significantly better

Table 4 Results of the Friedman tests for global differences between the performance measures of the modelling procedures

Performance measure	F	P-level
Kappa	292.54	0.0000
Sensitivity	145.84	0.0000
SE of sensitivity	107.67	0.0000

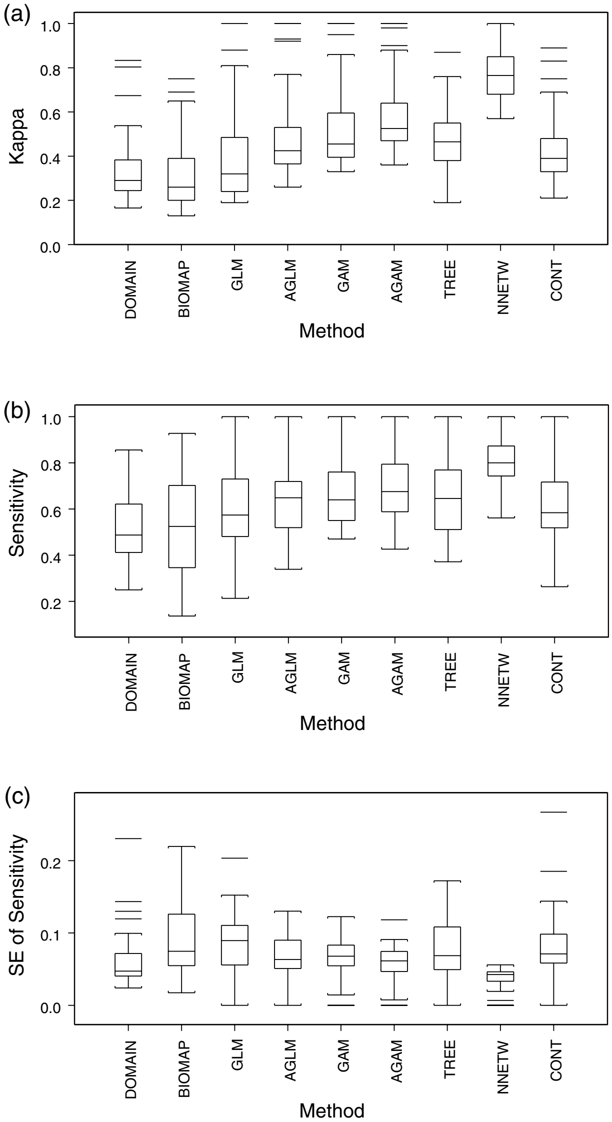

Figure 1 Boxplots of overall performance measures for each 12 modelling procedures (a, Kappa; b, sensitivity). Abbreviations: DOMAIN, Gower metric; BIOMAP, Ecological Niche Factor Analysis; GLM, general linear model; AGLM, autologistic general linear model, GAM, general additive model; AGAM , autologistic general additive model; TREE, classification tree; NNETW, neural network; CONT, contagion.

Table 5 Multiple comparisons of model's Kappa index

	DOMAIN	BIOMAP	GLM	AGLM	GAM	AGAM	TREE	NNETW
BIOMAP	n.s.							
GLM	n.s.	n.s.						
AGLM	***	***	**					
GAM	***	***	n.s.	n.s.				
AGAM	***	***	***	***	*			
TREE	***	***	***	n.s.	n.s.	***		
NNETW	***	***	***	***	***	n.s.	***	
CONT	*	*	n.s.	n.s.	**	***	n.s.	***

Dunn tests; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; n.s., $P > 0.05$.

Table 6 Multiple comparisons of model's sensitivity

	DOMAIN	BIOMAP	GLM	AGLM	GAM	AGAM	TREE	NNETW
BIOMAP	n.s.							
GLM	*	n.s.						
AGLM	***	*	n.s.					
GAM	***	*	n.s.	n.s.				
AGAM	***	***	*	n.s.	n.s.			
TREE	***	**	n.s.	n.s.	n.s.	n.s.		
NNETW	***	***	***	***	***	***	***	
CONT	**	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***

Dunn tests; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; n.s., $P > 0.05$.

Table 7 Multiple comparisons of model's standard errors of sensitivity

	DOMAIN	BIOMAP	GLM	AGLM	GAM	AGAM	TREE	NNETW
BIOMAP	**							
GLM	***	n.s.						
AGLM	n.s.	n.s.	n.s.					
GAM	n.s.	n.s.	n.s.	n.s.				
AGAM	n.s.	n.s.	**	n.s.	n.s.			
TREE	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.		
NNETW	*	***	***	***	***	***	***	
CONT	**	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	***

Dunn tests; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; n.s., $P > 0.05$.

($P < 0.01$) than the worse method in terms of sensitivity, i.e. BIOMAP. Spatial interpolation with CONT had a performance similar to GLM but, unlike GLM, did not show significant differences in the Kappa index from AGLM and TREE.

The results also show that model stability, measured by bootstrap estimates of sensitivity standard errors, was also affected by modelling technique. NNETW showed the lowest bootstrap standard errors (Fig. 1c) and according to Dunn tests (Table 7) were significantly different from all the other methods ($P < 0.001$, except for DOMAIN where $P < 0.05$). DOMAIN also showed significantly lower bootstrap standard errors than BIOMAP ($P < 0.01$), GLM ($P < 0.001$) and CONT ($P < 0.01$). GLM had the lowest overall model stability, although differences were only significant in relation to AGAM ($P < 0.01$) and NNETW ($P < 0.001$).

Model performance between methods and species groupings – spatial distributions

Four groups were created on the basis of species-spatial distribution (Fig. 2a) and they were all significantly different from each other (ANOSIM, $P < 0.05$): (1) narrowly distributed species, i.e. species with low area of occupancy and narrow extent of occurrence; (2) species with aggregated

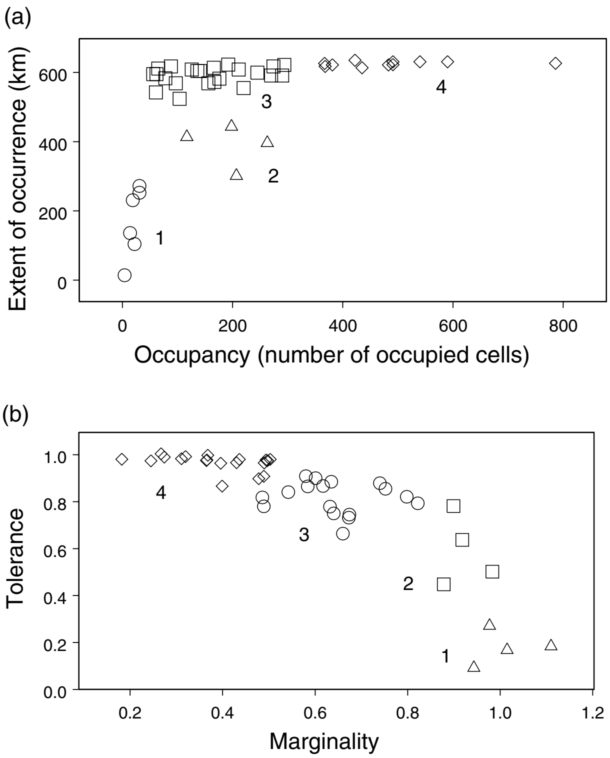

Figure 2 Representation of species and groups according to their geographical attributes (a) and their distribution in the environmental space (b).

distributions, i.e. species with both intermediate areas of occupancy and extent of occurrence; (3) widespread but locally rare species, i.e. species with low to intermediate areas of occupancy, but with large extent of occurrence; and (4) truly widespread species, i.e. species with high area of occupancy and large extent of occurrence.

Model performance varied consistently between species with different spatial distributions (ANOVA $P < 0.001$; Fig. 3a,b). Differences in standard error of sensitivity estimates and species groupings were also significant at $P < 0.05$. Variations in Kappa index revealed a clear trend towards increasing model performance with the restricted-range species of group 1, i.e. species with low area of occupancy and narrow extent of occurrence (Fig. 3a). At the opposite extreme we found truly widespread species of group 4 to have the lowest performance from the models. Variations in sensitivity values were similar to Kappa index, except for widespread species of group 4 that had generally good performances with sensitivity (Fig. 3b). This discrepancy is due to the numbers of false positives – included in Kappa index measurement, but not in sensitivity – that are particularly high among widespread species models causing Kappa index values to drop.

Interactions between the effects of method and species groupings on measures of model performance and stability were also highly significant (ANOVA, $P < 0.00001$). The relative performances of each method varied between species with different spatial distributions (Figs 4a, 5a & 6a): (1) lower relative performance and stability measures of DOMAIN and BIOMAP were more evident for species of group 1; (2) highest performances of NNETW are more evident for species of groups 3 and 4 (more marked for the Kappa index); (3) AGAM showed performances comparable with those of NNETW for species of groups 1 and 2.

Model performance between methods and species groupings – environmental distributions

As for spatial groupings, there were also four groups created on the basis of species' environmental distribution (Fig. 2b) and they were also significantly different from each other (ANOSIM, $P < 0.05$): (1) species with high marginality and low tolerance; (2) species with high marginality and intermediate levels of tolerance; (3) species with intermediate marginality and high tolerance; (4) species with low marginality and high tolerance.

Model performance also varied consistently between species with different environmental distributions (ANOVA, $P < 0.01$; Fig. 3c,d). Differences in standard error of sensitivity estimates and species groupings were also significant at $P < 0.05$. Model performance was higher for specialist species of group 1, i.e. species with high environmental marginality and low tolerance, and lower for generalist species of group 4, i.e. species with low marginality and high tolerance (Fig. 3c). Variations

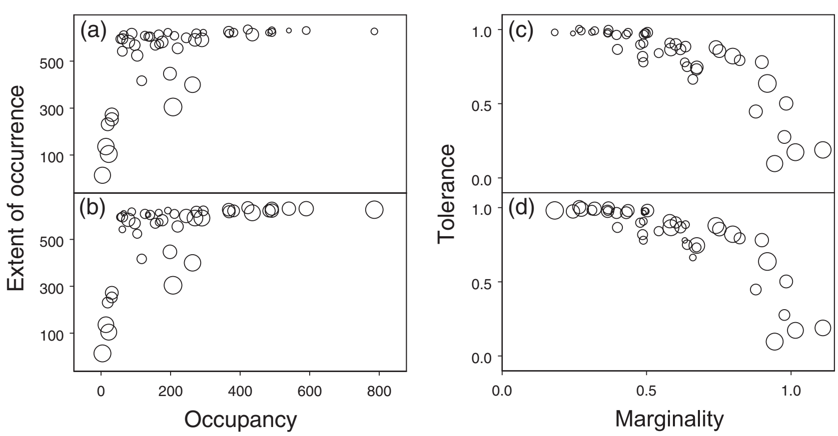

Figure 3 Representation of the mean values of Kappa index (a and c) and Sensitivity (b and d) of each species according to the species distribution in the geographical (a and b) and the environmental spaces.

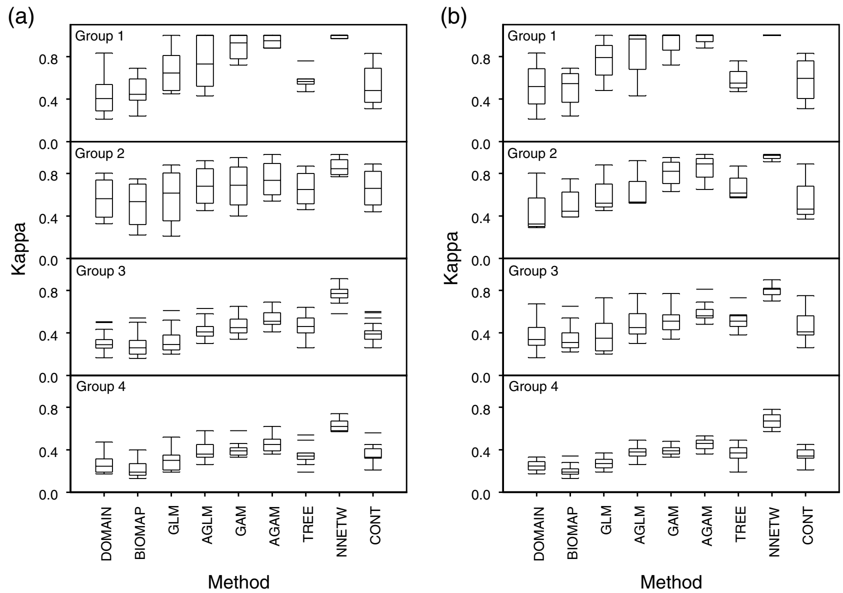

Figure 4 Boxplots of Kappa index for each model procedure and each species according to the species distribution in (a) geographical space and (b) environmental space (DOMAIN, Gower metric; BIOMAP, Ecological Niche Factor Analysis; GLM, general linear model; AGLM, with a covariate term accounting for spatial autocorrelation general linear model, GAM, general additive model; AGAM, with a covariate term accounting for spatial autocorrelation general additive model; TREE, classification tree; NNETW, neural network; CONT, contagion).

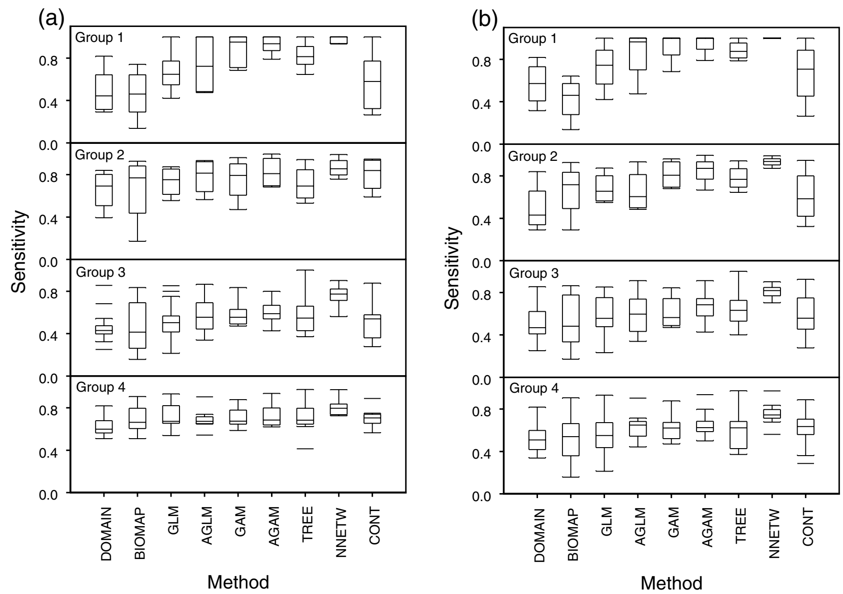

Figure 5 Boxplots of sensitivity for each model procedure and each species according to the species distribution in (a) geographical space and (b) environmental space. Abbreviations: DOMAIN, Gower metric; BIOMAP, Ecological Niche Factor Analysis; GLM, general linear model; AGLM, general linear model, with a covariate term accounting for spatial autocorrelation GAM, general additive model; AGAM, with a covariate term accounting for spatial autocorrelation general additive model; TREE, classification tree; NNETW, neural network; CONT, contagion.

in sensitivity values were similar to Kappa index, except for species of group 4 that had generally good performances with sensitivity (Fig. 3d).

Interactions between the effects of method and species groupings on measures of model performance and stability were also highly significant (ANOVA, $P < 0.00001$). The relative performances of each method varied between species with different environmental distributions (Figs 4b, 5b & 6b). As for spatial distributions, the same trend was observed for environmental distributions: (1) lower measures of performance and stability of DOMAIN and BIOMAP were more evident for species of group 1; (2) highest performances of NNETW were more evident for species of groups 3 and 4; (3) AGAM showed performances comparable to those of NNETW for species of groups 1 and 2.

DISCUSSION

Most studies assessing probabilistic models of occurrence for species have compared the performance of techniques for one or few species (e.g. Walker, 1990; Pereira & Itami, 1991; Manel *et al.*, 1999; Dettmers *et al.*, 2002; Thuiller *et al.*, 2003). Amongst these, only a few have investigated how variation in predictor (e.g. environmental) variables affected the ability of the models to predict correctly variation in the response (e.g. species) variable (e.g. Manel *et al.*, 1999). There have also been attempts to investigate how variation in the response variable would affect the outcome of the models (Araújo & Williams, 2000; Pearce & Ferrier, 2000; Manel *et al.*, 2001; Pearce *et al.*, 2001; Hepinstall *et al.*, 2002; Karl *et al.*, 2002). This latter approach was investigated further and the error in models

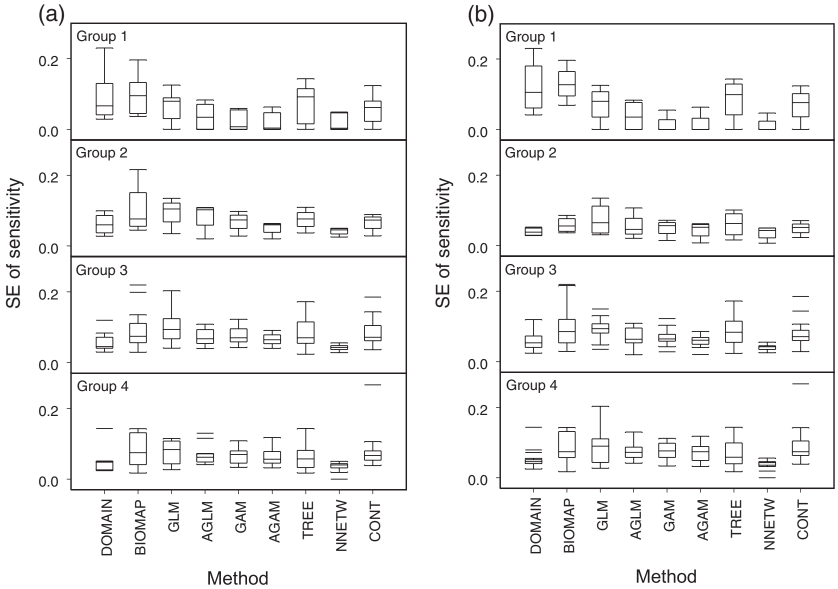

Figure 6 Boxplots of the standard error of sensitivity for each model procedure and each species according to the species distribution in (a) geographical space and (b) environmental space. Abbreviations: DOMAIN, Gower metric; BIOMAP, Ecological Niche Factor Analysis; GLM, general linear model with a covariate term accounting for spatial autocorrelation; AGLM, general linear model; GAM, general additive model; AGAM, general additive model with a covariate term accounting for spatial autocorrelation; TREE, classification tree; NNETW, neural network; CONT, contagion.

shown to be associated consistently with simple quantitative descriptors of the response variable, i.e. species distributions. In other words, the original idea that model performance was not independent of the kinds of geographical or environmental distributions of species was shown to be correct.

Generally, species with large areas of occupancy and great extents of occurrence (i.e. truly widespread) had greater overall errors (i.e. Kappa statistic), although the amount of unexplained variation (i.e. sensitivity) was not larger for species of this group than for species with other distribution profiles. This coincides with observations made by Araújo & Williams (2000), who explored the relationship between model performance (sensitivity and specificity) and the total number of records per species (i.e. area of occupancy). They found that sensitivity (proportion of false negatives) was higher for widespread and lower for restricted-range species, while specificity (proportion of false positives) was lower for widespread species and higher for restricted-range species. Similarly, Manel *et al.* (2001) investigated how model performance varied with the number of records per species relative to the number of samples (i.e. prevalence). They found sensitivity to be related positively to prevalence and negatively to specificity. Since the number of samples was identical for all species, in both studies, area of occupancy and prevalence were perfectly covariant. Karl *et al.* (2002) related avian species commonness and model accuracy using field and simulated data and concluded that habitat relationships for many rare species were likely to be as accurate as for common species, despite increasing error estimates with decreasing sample sizes. Similarly, Elith & Burgman (2002) did not find clear associations between modelling success and species characteristics such as rarity.

In this study, species with low marginality and high tolerance had lower overall performances. They were also generally better modelled with non-parametric techniques and benefited most from the inclusion of a term accounting for spatial autocorrelation. This is because autocorrelation reduces the amount of false positives (specificity), which in turn contributes to a reduction in the Kappa statistic. Hepinstall *et al.* (2002) also observed that performance of avian habitat models was negatively correlated with the proportion of habitats used by a species, although they used total correct predictions as the accuracy measure, and did not take marginality into account. They also noted an opposite trend with measures of sensitivity but in the current study no clear trend was observed for this measure.

Overall, the NNETW showed the highest model performance whereas similarity and ordination-based models (DOMAIN and BIOMAP, respectively) showed the lowest performances. While some authors (e.g. Mastrorillo *et al.*, 1997; Pearson *et al.*, 2002) also consider NNETW to be advantageous to model species occurrences, these observations are not supported by other studies where NNETW showed overall performances comparable to GLM (Manel *et al.*, 1999). In the few studies where species distribution models using GAM were compared with their GLM counterparts, the former approach also resulted in better model performances (e.g. Bio *et al.*, 1998; Franklin, 1998; Pearce & Ferrier, 2000; Thuiller *et al.*, 2003). Studies comparing TREE and GLM reached variable conclusions. For example, Dettmers *et al.* (2002) found that TREE and GLM produced comparable results. Franklin (1998) showed TREE to perform better than GLM, while Thuiller *et al.* (2003) showed GLM to perform better than TREE. Since these studies did not always use the same parameterization for TREE and GLM, they are, however, not fully comparable. Other studies also showed that similarity and ordination-based methods perform less well than other techniques, namely GAM and GLM (Elith & Burgman, 2002). Hirzel *et al.* (2001) compared the performance of GLM and BIOMAP using virtual species and concluded, as here, that GLM produced better results for overabundant species.

These apparently divergent patterns of model' performance are likely to be related to variations in the methods' abilities to recover useful relationships between species with different distributions and environmental factors with different strengths and lengths of gradients. Different modelling techniques have distinct approaches regarding adjustment to data (Guisan & Zimmermann, 2000). Indeed, modelling techniques used in this paper can be grouped into four categories according to their kinds of adjustment to data: (1) techniques that give priority to the empirical behaviour of species' presence/absence response to environmental variables (e.g. non-parametric models such as GAM, classification trees and neural networks); (2) techniques that focus on general trends of species' presence/absence response (e.g. parametric models such as GLM); (3) techniques that use only species presence data to seek relationships with the environmental predictors (e.g. Gower-similarity and ENFA); (4) techniques that use only species presence data and their geographic positions to develop predictions of species occurrence (e.g. spatial interpolators). The first and fourth approaches are expected to provide better models for species with complex distribution patterns, i.e. where occurrences do not respond to environmental variables according to a predefined 'shape'. This is frequently the case among widespread species that do not appear to respond to clear environmental gradients, although studies at varying spatial scales might provide further insights on species' governing factors (Thuiller *et al.*, 2003). Conversely, the second approach is expected to provide reasonable models for species responding to environmental gradients as predicted by simple response curves. The third and fourth approaches were both expected to provide models with high sensitivity (low misclassification of true presences). However, they were also expected to yield models with low overall performances, since their formulation disregards the response of absence data to environmental variables. The third approach, in particular, should be considered when no reliable data on species absences are available.

Factors other than species ecological attributes can also be a source of variation in model performance. For example, environmental variables may be adequate for some species but not for others. Data quality can also strongly affect model performance. This may include imprecise location of species occurrences, but also error propagation during generation of response variables in GIS environments (Corsi *et al.*, 2000). Moreover, if another spatial scale or resolution is adopted, then a different pattern of model performances might be observed (Mackey & Lindenmayer, 2001; Thuiller *et al.*, 2003).

It is unlikely that a single best habitat modelling procedure will ever be identified. Different methods have different strengths and weaknesses and the choice of the appropriate method depends on the data, assumptions and goals of the exercise. This study provides some insights into the structure of the error in the models and how they vary for species with different kinds of distribution. Although NNETW and AGAM provided generally better results than other methods tested, an unequivocal pattern of model performance that worked for all species was not found. This leaves the modeller with two choices. Either (1) to use expert systems (e.g. GARP, Stockwell & Peters, 1999; BIOMOD, Thuiller, 2003) that compare methods in an automatic fashion and choose the best method for each species run, or else choose a single method that is seen to be generally robust (e.g. NNETW and AGAM) or; (2) choose a method thought to be particularly robust to the kinds of data and goals pursued (e.g. DOMAIN and ENFA when limited records of presence, alone, are available). The first strategy seeks to optimize the fit of the models given available data. It is completely data driven and no *a priori* hypotheses are made about data and the nature of species' responses to gradients. The second option ensures reasonable results while preserving accountability, given that model assumptions are made clear and applied for all species modelled. This last option would exclude NNETW because ecological principles underlying the fit of these models are poorly understood, although progress is being made towards eliminating this limitation (e.g. Olden & Jackson, 2002). From the relatively extensive literature on the subject, it becomes clear that model adequacy depends greatly on each particular situation. Studies that attempt to evaluate and compare model performances for various techniques should always consider the effect of the many factors involved, such as geographical scale, adequacy of predictor variables and, as shown in this paper, the spatial and environmental distribution of species.

ACKNOWLEDGMENTS

We thank the many volunteer fieldworkers who contributed to atlas records; Alexander Hirzel for support with ENFA; Wilfried Thuiller, Richard Pearson, Susana Rosa and two anonymous referees for comments on the manuscript. This work benefited from research grants by the Portuguese FCT/ICN research programme (PNAT/1999/BIA/15051/P) and by the European ATEAM project (EVK2-CT-2000-00075); MBA and PS were also funded by FCT (SFRH/BPD/5547/2001 and SFRH/BD/8493/2002, respectively).

REFERENCES

Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96,** 331–345.

Araújo, M.B. & Williams, P.H. (2001) The bias of complementarity hotspots toward marginal populations. *Conservation Biology*, **15,** 1710–1720.

Araújo, M.B., Williams, P.H. & Fuller, R.J. (2002) Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society London*, **B269,** 1971–1980.

Araújo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2004) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, doi: 10.1111/j.1466-822x.2004.00128.x.

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (1996) An autologistic model for spatial distribution of wildlife. *Journal of Applied Ecology*, **33,** 339–347.

Austin, G.E., Thomas, C.J., Houston, D.C. & Thompson, D.B.A. (1996) Predicting the spatial distribution of buzzard

Buteo buteo nesting areas using a Geographical Information System and remote sensing. *Journal of Applied Ecology*, **33,** 1541–1550.

Bailey, T.C. & Gatrell, A.C. (1995) *Interactive spatial data analysis*. Longman, London.

Beutel, T.S., Beeton, R.J.S. & Baxter, G.S. (1999) Building better wildlife-habitat models. *Ecography*, **22,** 219–223.

Bio, A.M.F., Alkemande, R. & Barendregt, A. (1998) Determining alternative models for vegetation response analysis – a non-parametric approach. *Journal of Vegetation Science*, **9,** 5–16.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) *Classification and regression trees*. Wadsworth International Group, Belmont, CA.

CNA (1983) *Atlas do Ambiente*. Direcção Geral do Ambiente. Ministério do Ambiente e dos Recursos Naturais, Lisbon.

Carpenter, G., Gillison, A.N. & Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, **2,** 667–680.

Clark, L.A. & Pregibon, D. (1992) Tree-based Models. *Statistical Models in S*. (ed. by J.M. Chambers and T.J. Hastie), pp. 377–419. Chapman and Hall, New York.

Clarke, K.R. & Gorley, R.N. (2001) *PRIMER v5: users manual/ tutorial*. PRIMER-E Ltd, Plymouth, UK.

Cocks, K.D. & Baird, I.A. (1989) Using mathematical programming to address the multiple reserve selection problem: an example from Eyre Peninsula, South Australia. *Biological Conservation*, **49,** 113–130.

Corsi, F., de Leeuw, J. & Skidmore, A. (2000) Modeling Species Distribution with GIS. *Research techniques in animal ecology. Controversies and consequences* (ed. by L. Boitani and T.K. Fuller), pp. 389–434. Columbia University Press, New York.

Cumming, G.S. (2000) Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography*, **27,** 441–455.

De'Ath, G. & Fabricius, K.E. (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, **81,** 3178–3192.

Dettmers, R., Buehler, D.A. & Bartlett, J.G. (2002) A test and comparison of wildlife-habitat modelling techniques for predicting bird occurrence at a regional scale. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, M.L. Morrison and P.J. Heglund), pp. 607–615. Island Press, Covelo, CA.

Donald, P.F. & Greenwood, J.J.D. (2001) Spatial patterns of range contraction in British breeding birds. *IBIS*, **143,** 593–601.

Efron, B. & Tibshirani, R.J. (1993) *An introduction to the bootstrap*. Chapman & Hall, San Francisco, CA.

Elith, J. & Burgman, M. (2002) Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, M.L. Morrison and P.J. Heglund), pp. 303–313. Island Press, Covelo, CA.

ESRI (1999) *ArcView GIS Help (v3.2)*. Environmental Systems Research Institute, Inc., NewYork, USA.

Feria, T.P. & Peterson, A.T. (2002) Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments. *Diversity & Distributions*, **8,** 49–56

Fertig, W. & Reiners, W.A. (2002) Predicting presence/absence of plant species for range mapping: a case study from Wyoming. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, M.L. Morrison and P.J. Heglund), pp. 483–489. Island Press, Covelo, CA.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/ absence models. *Environmental Conservation*, **24,** 38–49.

Fielding, A.H. & Haworth, P.F. (1995) Testing the generality of bird-habitat models. *Conservation Biology*, **9,** 1466–1481.

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.

Gaston, K.J. (1996) Species-range-size distributions: patterns, mechanisms and applications. *Trends in Ecology and Evolution*, **11,** 197–201.

Gates, S. & Donald, P.F. (2000) Local extinction of British farmland birds and the prediction of further loss. *Journal of Applied Ecology*, **37,** 806–820.

Godinho, R., Teixeira, J., Rebelo, R., Segurado, P., Loureiro, A., Alvares, F., Gomes, N., Cardoso, P., Camilo-Alves, C. & Brito, J.C. (1999) Atlas of the continental Portuguese herpetofauna: an assemblage of published and new data. *Revista Española de Herpetologia*, **13,** 61–82.

Guisan, A., Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135,** 147–186.

Hastie, T. & Tibshirani, R. (1990) *Generalised additive models*. Chapman and Hall, London.

Hepinstall, J.A., Krohn, B. & Sader, S.A. (2002) Effects of niche width on the performance and agreement of avian habitat models. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, M.L. Morrison and P.J. Heglund), pp. 593–606. Island Press, Covelo, CA.

Hirzel, A. (2001) When GIS come to life. Linking landscape- and population ecology for large population management modelling: the case of Ibex (Capra ibex) in Switzerland, PhD Thesis. University of Lausanne, Lausanne.

Hirzel, A., Hausser, J. & Perrin, N. (2001) *Biomapper 1.0. Lausanne, laboratory for conservation biology*. URL: http:// www.unil.ch/biomapper.

Karl, J.W., Bomar, L.K., Heglund, P.J., Wright, N.M. & Scott, M. (2002) Species commonness and the accuracy of habitat-relationship models. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, M.L. Morrison and P.J. Heglund), pp. 573–580. Island Press, Covelo, CA.

Klockars, A. & Sax, G. (1986) *Multiple Comparisons*. Quantitative Applications in the Social Sciences series. Sage Publications, Thousand Oaks, CA.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm. *Ecology*, **74,** 1659–1673.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S. (1996) Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, **90,** 39–52.

MacArthur, R.H. (1957) On the relative abundance of bird species. *Proceedings of the National Academy of Sciences*, **43,** 293–295.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*. Chapman & Hall, London.

Mackey, B.G. & Lindenmayer, D.B. (2001) Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography*, **28,** 1147–1166.

Manel, D., Dias, J.M., Buckton, S.T. & Ormerod, S.J. (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36,** 734–747.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38,** 921–931.

Margules, C.R. & Nicholls, A.O. (1987) Assessing the conservation value of remnant 'islands': mallee patches on western Eyre Peninsula, South Australia. *Nature conservation: the role of remnants of native vegetation* (ed. by D.A. Saunders, G.W. Arnold, A.A. Burbidge and A.J.M. Hopkins), pp. 89–102, Surrey Beatty and Sons, Sidney.

Mastrorillo, S., Lek, S., Dauba, F. & Belaud, A. (1997) The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology*, **38,** 237–246.

Monserud, R.A. & Leemans, R. (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, **62,** 275–293.

Myers, J.C. (1997) *Geostatistical error management: quantifying uncertainty for environmental sampling and mapping*. Van Nostrand Reinhold, New York.

Olden, J.D. & Jackson, D.A. (2002) Illuminating the "black box": Understanding variable contributions in artificial neural networks. *Ecological Modelling*, **154,** 135–150.

Özesmi, S.L. & Özesmi, U. (1999) An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling*, **116,** 15–31.

Pearce, J. & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128,** 127–147.

Pearce, J, Ferrier, S. & Scotts, D. (2001) An evaluation of the predictive performance distributional models for flora and fauna in north-east New South Wales. *Journal of Environmental Management*, **62,** 171–184.

Pearson, R.G., Dawson, T.P., Berry, P.M. & Harrison, P.A. (2002) Species: a spatial evaluation of climate impact on the envelope of species. *Ecological Modelling*, **154,** 289–300.

Pereira, J.M.C. & Itami, R.M. (1991) GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham Red Squirrel. *Photogrammetric Engineering and Remote Sensing*, **57,** 1475–1486.

Perrin, N. (1984) Contribution à l'écologie du genre Cepaea (Gastropoda): Approche descriptive et expérimentale de l'habitat et de la niche écologique, PhD Thesis. University of Lausanne, Lausanne.

Polasky, S. & Solow, A.R. (2001) The value of information in reserve site selection. *Biodiversity and Conservation*, **10,** 1051–1058.

Smith, P.A. (1994) Autocorrelation in logistic regression modelling of species distributions. *Global Ecology and Biogeography Letters*, **4,** 47–61.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: the principles and practice of statistics in biological research*. W.H. Freeman, New York.

Sprent, P. & Smeeton, N.C. (2001) *Applied nonparametric statistical methods. Texts in statistical science*. Chapman & Hall/CRC, Boca Raton.

Statistical Sciences (1999) *S-Plus, version 2000 for Windows*. Mathsoft Inc., Seattle, WA.

Stockwell, D.R.B. & Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13,** 143–158.

Therneau, T.M. & Atkinson, E.J. (1997) *An introduction to recursive partitioning using the RPART routines*. URL: http://www.mayo.edu/hsr/techrpt/61.pdf.

Therneau, T.M. & Atkinson, E.J. (2000) *An introduction to recursive partitioning using the RPART routines*. URL: http://www.mayo.edu/hsr/techrpt/rpartmini.pdf.

Thuiller, W. (2003) BIOMOD – Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, **9,** 1353–1362.

Thuiller, W., Araújo, M.B. & Lavorel, S. (2003) Generalised models versus classification tree analysis: a comparative study for predicting spatial distributions of plant species at different spatial scales. *Journal of Vegetation Science*, **14,** 669–680.

Vayssières, M.P., Plant, R.E. & Allen-Diaz, B.H. (2000) Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, **11,** 679–694.

Venables, W.N. & Ripley, B.D. (2002) *Modern applied statistics with S-PLUS*. Springer-Verlag, New York.

Vetaas, O.R. (2000) Comparing species temperature response curves: population density versus second-hand data. *Journal of Vegetation Sciences*, **11,** 659–666.

Walker, P.O. (1990) Modelling wildlife distribution using a geographic information system: kangaroos in relation to climate. *Journal of Biogeography*, **17,** 279–289.

Williams, P.H. & Araújo, M.B. (2000) Using probability of persistence to identify important areas for biodiversity conservation. *Proceedings of the Royal Society London B*, **267,** 1959–1966.

Williams, P.H. & Araújo, M.B. (2002) Apples, oranges, and probabilities: integrating multiple factors into biodiversity conservation with consistency. *Environmental Modeling and Assessment*, **7,** 139–151.

BIOSKETCHES

Pedro Segurado is a PhD student at the University of Évora, Portugal, and University of Leeds, UK. His main interests include herpetology, conservation biology and macroecology.

Miguel Araújo has a PhD from the University of London. His research focuses on the development of quantitative reserve selection methods that are more robust with regard to local extinctions of species. An important part of his work also includes macroecological analysis of the patterns and processes affecting occurrence and persistence of species at different temporal scales.