Original article

Threshold criteria for conversion of probability of species presence to either–or presence–absence

Alberto Jiménez-Valverde*, Jorge M. Lobo

Departamento Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC),
c/ José Gutiérrez Abascal 2, 28006 Madrid, Spain

ARTICLE INFO

ABSTRACT

For many applications the continuous prediction afforded by species distribution modeling must be converted to a map of presence or absence, so a threshold probability indicative of species presence must be fixed. Because of the bias in probability outputs due to frequency of presences (prevalence), a fixed threshold value, such as 0.5, does not usually correspond to the threshold above which the species is more likely to be present. In this paper four threshold criteria are compared for a wide range of sample sizes and prevalences, modeling a virtual species in order to avoid the omnipresent error sources that the use of real species data implies. In general, sensitivity–specificity difference minimizer and sensitivity–specificity sum maximizer criteria produced the most accurate predictions. The widely-used 0.5 fixed threshold and Kappa-maximizer criteria are the worst ones in almost all situations. Nevertheless, whatever the criteria used, the threshold value chosen and the research goals that determined its choice must be stated.

1. Introduction

Species distributions are increasingly being modeled in ecology and conservation research. Prediction of species geographic distribution, based on known occurrences, is now possible due to both Geographic Information Systems (GIS) and statistical quantification of species–environment relationships (Guisan and Zimmermann, 2000; Lehmann et al., 2002; Rushton et al., 2004). Species distribution model predictions help to delve into questions of biogeography and evolution (Peterson et al., 1999; Anderson et al., 2002a,b; Hugall et al., 2002; Peterson and Holt, 2003; Wiens and Graham, 2005), to search for biological indicators (Bonn and Schröder, 2001), to

study the effect of climate warming on species distribution (Teixeira and Arntzen, 2002; Thuiller et al., 2005; Araújo et al., 2006), and to develop management decisions and conservation strategies (Godown and Peterson, 2000; Schadt et al., 2002; Barbosa et al., 2003; Meggs et al., 2004; Russell et al., 2004; Chefaoui et al., 2005).

Prediction methods currently available to scientists can be divided, roughly, into those that use only presence data (profile techniques, e.g. environmental envelopes, see Elith et al., 2006 for an updated review) and those that also incorporate absence data (group discrimination techniques, e.g. generalized regression, see Guisan and Zimmermann, 2000; Scott et al., 2002). Profile techniques tend, in general, to overestimate distributions

due to the lack of absence data, which would otherwise restrict predictions where needed (Ferrier and Watson, 1997; Zaniewski et al., 2002; Engler et al., 2004). However, these methods are useful because reliable absence data are seldom available (MacKenzie et al., 2002; Wintle et al., 2005; but see, for example, Engler et al., 2004; Lobo et al., 2006, for ways of generating probable absences). When available, reliable absence data should be treated with group discrimination techniques, capable of accounting for more real relationships between species and environmental and historical factors (Hirzel et al., 2001; Brotons et al., 2004; Segurado and Araújo, 2004).

Although the continuous map of the probability of presence produced by distribution models is itself useful for many conservation applications (e.g., Araújo et al., 2002; Wilson et al., 2005), it is often converted into an either–or presence/absence map. The resulting categorical map is used for practical applications (e.g., Araújo et al., 2004, 2006; Luoto et al., 2006), as well as for evaluating model prediction reliability, involving comparison with the inherently either–or presence/absence data using a confusion matrix (Fielding and Bell, 1997). Although threshold-independent accuracy measures are now widely preferred, such as the area under the Receiver Operating Characteristic curve (AUC; Swets, 1988; Fielding and Bell, 1997), these techniques have started to be seriously criticized (Austin, in press; Termansen et al., 2006). The conversion of the continuous map into a categorical one involves adopting a threshold probability indicative of species presence (Fielding and Bell, 1997) which will determine model output, as it will condition the cases assigned to each category (Fielding and Bell, 1997; Manel et al., 1999b; Pearce and Ferrier, 2000). However, logistic regression probabilities are biased toward the highest number of either presences or absences, where they differ (Hosmer and Lemeshow, 1989; Cramer, 1999). Because of this bias, due to prevalence (the proportion of presence cases), the intuitive threshold value of 0.5 (e.g. Li et al., 1997) may not actually correspond to the threshold above which the species is more likely to be present. For example, where a large number of target-species absence observations bias probabilities toward zero, a threshold of 0.5 will lead to absence predictions for sites with known presences (high omission error rate), reduce sensitivity (true predicted presences) and increase specificity (true predicted absences). Lowering the threshold from 0.5 would increase sensitivity, at the expense of decreased specificity. What is the best threshold for binary data with a dissimilar number of presences and absences?

The choice of threshold criteria can depend on the role of commission (false positive) and omission (false negative) errors (Fielding and Bell, 1997; Fielding, 2002; Pearson et al., 2004). The degree to which these errors are minimized depends on how the model will be used (Loiselle et al., 2003; Rondinini et al., 2006). However, models usually are designed to discriminate as reliably as possible between presence and absence sites. There has been little effort aim at comparing model results applying different threshold criteria, and the 0.5 point threshold or the one which maximized the Kappa statistic (Monserud and Leemans, 1992) are actually widely used in order to generate binary distribution maps (see, for example, Li et al., 1997; Guisan et al., 1998; Manel et al., 1999a; Fleishman et al., 2001; Thuiller, 2003; Berg et al., 2004; Meggs

et al., 2004; Segurado and Araújo, 2004; Araújo et al., 2005; Luoto et al., 2005). Manel et al. (2001) examined a large set of species and concluded that results from a threshold which maximized the sensitivity–specificity sum (following Zweig and Campbell, 1993) were superior to results from a threshold of 0.5. Liu et al. (2005) compared 12 threshold decision criteria using data of two plant species in Europe modeled with neural networks. They concluded that fixed thresholds and those based on the Kappa statistic work worse that those accounting, directly or indirectly, for prevalence.

The present study compares model outputs obtained from varying prevalence data, modeled with logistic regression (LR), a widely-used modeling technique (Guisan et al., 2002; Lehmann et al., 2002), and four threshold criteria: the widely-used 0.5 and Kappa maximization thresholds, and the sensitivity–specificity difference minimizer (e. g. Bonn and Schröder, 2001; Barbosa et al., 2003) and sensitivity–specificity sum maximizer (e. g. Manel et al., 2001), two of the best criteria found by Liu et al. (2005). The general aim of this paper is to find the optimum threshold criteria for a wide range of model specifications. In order to achieve this objective real data are not used to avoid the frequently error sources that their use implies; instead, a distribution of a virtual species was postulated.

2. Methods

2.1. The virtual species

As has been recommended (Allredge and Ratti, 1986; Austin, in press), predictions derived from four threshold criteria were compared using a postulated species distribution with known environmental influence. This procedure has been employed by other researchers (Hirzel et al., 2001; Reese et al., 2005) to avoid complications from natural variation. Specifically, we have used this approach in order to:

(i) Avoid misclassification error associated with real presence–absence data.
(ii) Be sure that the modeling technique (LR) can correctly predict species distribution while avoiding the bias due to contingent unaccounted-for or unknown factors. Both virtual species and distribution model predictions are based on the same environmental variables.
(iii) Be completely confident about models accuracy.

The virtual species distribution was mapped at a spatial resolution of 0.04° for the European region (−13° to 35° longitude, and 34° to 72° latitude). The total area of the region studied measured 6,576,424 km^2 (510,440 squares). For this region, four environmental variables were extracted from WORLD-CLIM interpolated map database (version 1.3; see http://biogeo.berkeley.edu/worldclim/worldclim.htm and Hijmans et al., 2005): monthly total precipitation, precipitation of the warmest quarter, monthly mean maximum temperature and monthly mean minimum temperature. Box-Cox normalized environmental variables were standardized to 0 mean and 1 standard deviation to eliminate measurement-scale effects. Principal Component Analysis (PCA; see Legendre

and Legendre, 1998) was performed to obtain two reduced non-correlated environmental variables, able to explain 92.6% of the environmental variation across the European region. The first axis was positively correlated with temperature variables, while the second was correlated with precipitation variables. The mean scores of these two environmental factors were calculated. The environment range inhabited by the species was set to the mean ± S.D. of each factor. All cells falling within these intervals for both factors were selected as the true distribution range of the virtual species in Europe (presences; $n = 91,144$), while the remaining cells were considered as true absences ($n = 419,296$). All geographic analysis were done with IDRISI Kilimanjaro software (Clark Labs, 2003). The geographical distribution of this ''central'' European virtual species is completely determined by well-known environmental factors.

2.2. The modeling process

Previous analysis showed that a sample size lower than around 70 observations decreases model performance (A. Jiménez-Valverde, J.M. Lobo, J. Hortal, submitted; see also Kadmon et al., 2003). So, in order to explore a possible effect of sample size on threshold selection, we created a series of datasets with sample sizes greater than 70 observations. Nine categories of presence numbers ($n = 91, 456, 911, 4557, 9114, 22,786, 45,572, 68,358$ and $91,144$) were randomly selected from the distribution. Absences were also randomly selected in the same number categories as presences. Thus 81 models (9 categories of presences × 9 categories of absences) were designed, which vary both in the number of observations (minor value = 182, major value = 182,288, based on the 9 categories above) and in the prevalence or proportion of presences (minor value = 0.001, major value = 0.999, based on the 9 categories above). All models were designed using logistic regression analysis (Generalized Linear Models with binomial distribution and logit-link function; McCullagh and Nelder, 1997). The linear, quadratic and cubic functions of the two environmental factors, together with their interaction, were used as explanatory variables; models were selected by a backward stepwise procedure (Harrell, 2001). Models were fitted using the STATISTICA program (StatSoft Inc, 2001).

Statistics to compare the accuracy of two binary maps are derived from a cross-tabulated matrix of the number of observed presence and absence cases against the predicted presences and absences (confusion matrix). Commission errors (model predictions of species presence where not actually observed, i.e. the false positive fraction) and omission errors (model predictions of species absence where actually observed; i.e. the false negative fraction) are determined by the number of cases correctly and incorrectly assigned to presences and absences. Specificity is calculated as the ratio of correctly predicted absences to the total number of absences, and sensitivity as the ratio of correctly predicted presences to their total number. The confusion matrix was calculated for the training data. The models were then projected onto the whole European territory and their probability scores converted into a binary variable (presence/absence) by applying the threshold criteria explained in the next section, based on the confusion matrix. Predicted and real virtual maps were compared by

calculating the sensitivity and specificity as well as the frequently-used Kappa statistic (Monserud and Leemans, 1992; Fielding and Bell, 1997; Pontius, 2000) that takes into consideration both commission and omission errors.

2.3. Threshold criteria

Model extrapolations were converted into presence/absence maps by selecting threshold probabilities above which presence was established, according to the following criteria:

– 0.5T criteria: a value of 0.5 was the threshold above which presence was assigned (Li et al., 1997; Manel et al., 1999a).
– KMT criteria (Kappa-maximized threshold): Kappa scores were calculated for 100 threshold values (in 0.01 increments) and the one which provides maximum Kappa became the accepted threshold (Guisan et al., 1998; Thuiller, 2003). When multiple thresholds had the same Kappa value, the mean threshold value was selected.
– MDT criteria (minimized difference threshold): difference between sensitivity and specificity was calculated for the same 100 threshold values and the one which minimized that difference was selected (Rojas et al., 2001).
– MST criteria (maximized sum threshold): sum of sensitivity and specificity was calculated for the same 100 threshold values and the one which maximized that sum was selected (Manel et al., 2001).

3. Results

Sample size was uncorrelated with the thresholds selected by MDT, MST and KMT criteria (Spearman rank correlation coefficients, $Rs = -0.03, -0.02$ and 0.04, respectively). On the contrary, prevalence was significantly and positively correlated with the thresholds selected by those three criteria. Both MDT and MST thresholds were linearly related with prevalence (Fig. 1), so frequency of presence data alone could be used to select the most appropriate threshold. KMT thresholds increased rapidly with either low or high prevalence values, remaining relatively constant (around 0.5) for a wide range of prevalence values. The thresholds from the KMT, MDT and MST criteria were also correlated with each other, especially those of the two latter criteria (see Fig. 1).

The four threshold criteria produced significantly different mean values of sensitivity, specificity and Kappa (Table 1). Mean sensitivity and Kappa values were significantly higher for MST and MDT, while they were significantly lower for 0.5T and KMT. Mean specificity values were significantly higher only for MDT (Table 1).

Kappa and specificity values obtained across all sample size and prevalence treatments with MST and MDT were significantly correlated, as were those obtained with 0.5T and KMT thresholds; the latter producing the highest correlation scores (Table 2). Sensitivity values were also positively correlated for the pairs MST/KMT and MST/0.5T, although again 0.5T and KMT producing the highest correlation score. Sensitivity scores from 0.5T and KMT, which were linearly related (Fig. 2), were extremely variable in comparison with those from MST and MDT, which, while correlated, were always

	Prevalence	MDT	KMT
MDT	0.99	-	-
KMT	0.73	0.73	-
MST	0.99	0.99	0.74
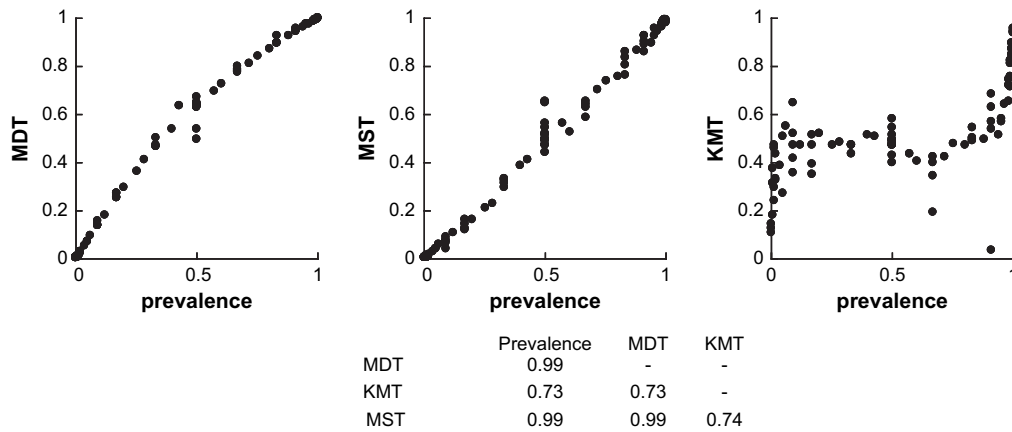
Fig. 1 – Relationship between prevalence of occurrence and threshold scores for the three non-fixed threshold criteria. In the table, Spearman correlation scores are shown between prevalence and threshold values and between threshold scores of the three criteria.

higher than 0.8. The pattern for the Kappa statistic was quite similar, while specificity values were high for the four criteria.

Kappa, sensitivity and specificity score differences derived from the various threshold criteria (relative performance) could not be explained by variation in sample size, but could be explained by variation in prevalence (Table 3), which explained (with the exception of MDT/MST) 41% to 98% of their deviance. The reliability of MDT- or MST-criteria designed presence–absence maps proved to be independent of the frequency of presence points. However, 0.5T and KMT predicted presences were relatively reliably in cases of high prevalence scores, and absences were relatively reliably at low prevalence scores (Fig. 3). In such cases 0.5T and KMT criteria superiority over MDT and MST was negligible. It is interesting to highlight the MDT- and MST-criteria significant superiority over the other two in predicting presences when the prevalence was low (Fig. 3B); no such pattern was observed in the prediction of absences.

4. Discussion

Prediction reliability from models is particularly sensitive to threshold criteria applied in model derivation. Our results,

derived from a wide range of conditions, provide some guidance to the choice of threshold criteria. Above all, threshold criteria should be dependent on prevalence. For the same geographical extent, mean LR probability magnitudes, biased by prevalence, tend toward zero for rare species (narrow geographic range, i.e. low occurrence rate and, so, low prevalence scores) and toward one for common species (widespread, i.e. high occurrence rate and high prevalence scores). Thus, as shown in the present work, a threshold fixed a priori yields a binary model that is not biologically meaningful. Of the KMT, MDT and MST criteria tested, the ones which maximize the sensitivity–specificity sum or minimize their difference (MST and MDT, respectively) are the most linearly related with prevalence, a desirable property since the meaningfulness of model probabilities depends on the maximum value obtained (Pontius and Batchu, 2003). The strong correlation between the threshold values from these two criteria and prevalence obtained by us supports the recent proposal of using prevalence values themselves as threshold decision criteria (Liu et al., 2005; Jiménez-Valverde and Lobo, 2006), as previously recognized by statisticians (Cramer, 1999).

These two prevalence-dependent thresholds are strongly correlated and always score high in accuracy. KMT criteria produce quite variable accuracy scores that are highly correlated with 0.5T scores, a consequence of the stability of the

Table 1 – Mean values of the three accuracy measures (Kappa statistic, sensitivity and specificity) ± SE for the four threshold criteria, and ANOVA results (***$P < 0.001$, *$P < 0.005$). Pairwise significant differences were determined using a Tukey test (HDS; $P < 0.05$) and are shown with superscript letters

	Kappa	Sensitivity	Specificity
$F_{(3,320)}$	29.09***	10.97***	4.48*
[a]MST	0.734 ± 0.003[b,d]	0.956 ± 0.002[b,d]	0.898 ± 0.002[c]
[b]0.5T	0.599 ± 0.024[a,c]	0.799 ± 0.035[a,c]	0.894 ± 0.009[c]
[c]MDT	0.766 ± 0.002[b,d]	0.926 ± 0.002[b,d]	0.923 ± 0.001[a,b,d]
[d]KMT	0.644 ± 0.016[a,c]	0.833 ± 0.028[a,c]	0.898 ± 0.008[c]

Table 2 – Spearman correlation scores between threshold criteria for the three accuracy measures, applying Bonferroni correction for multiple comparison test. Significant pairwise are marked in bold ($P \leq 0.003$)

	Kappa	Sensitivity	Specificity
MST/0.5T	0.06	**0.42**	0.16
MST/MDT	**0.52**	**0.40**	**0.35**
MST/KMT	0.02	**0.46**	0.18
0.5T/MDT	0.33	0.24	−0.04
0.5T/KMT	**0.93**	**0.98**	**0.98**
MDT/KMT	0.21	0.27	−0.07
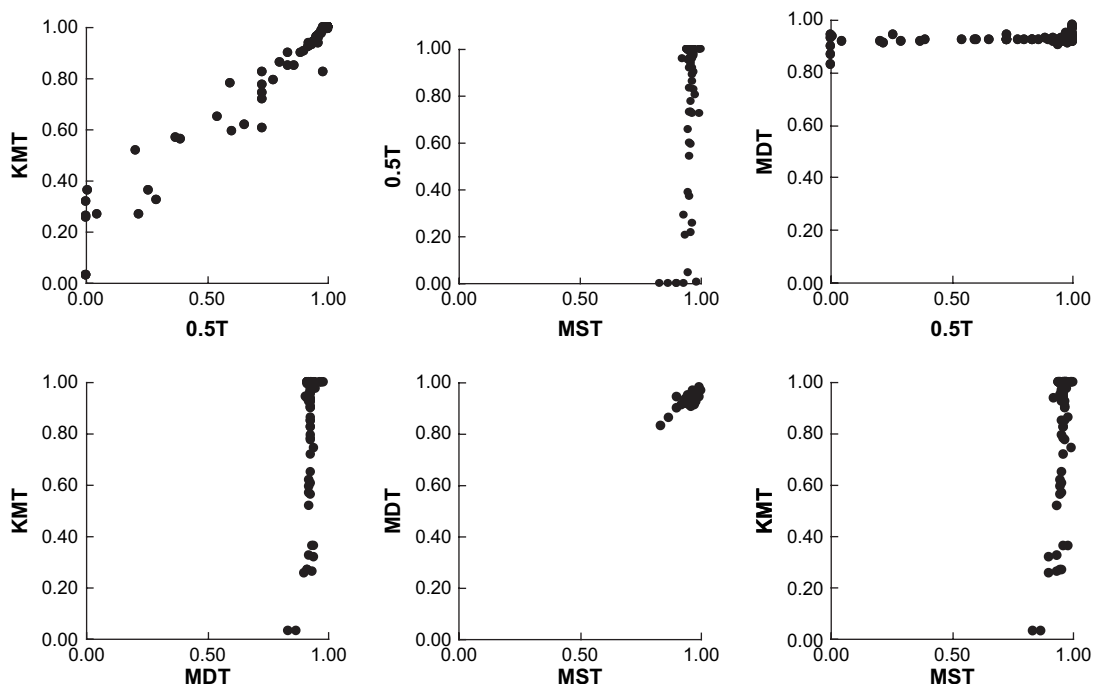
Fig. 2 – Sensitivity scores obtained using different threshold criteria. While those with 0.5T and KMT criteria are highly variable, sensitivity values obtained with MDT and MST criteria show always quite high values.

Table 3 – The relative difference in accuracy using different threshold criteria was modeled using Generalized Linear Models (GLMs) with a log-link function and a normal distribution. The independent variables (sample size and prevalence of presence data) were included in the model considering their cubic, quadratic and linear functions, and the adequacy of the models was tested by means of the change in explained deviance (% Expl. Dev.) from a null model in which the difference in relative performance is modeled with no explanatory variables

	Sample-size	Prevalence	
	% Exp. Dev.	% Exp. Dev.	Function
Kappa			
MDT-0.5T	0.05	95.94	Cubic
MDT-KMT	0.24	82.21	Cubic
MST-0.5T	0.03	95.97	Cubic
MST-KMT	0.26	85.54	Cubic
MDT-MST	0.46	7.30	Cubic
Sensitivity			
MDT-0.5T	0.08	96.38	Cubic
MDT-KMT	0.01	93.31	Cubic
MST-0.5T	0.09	97.75	Cubic
MST-KMT	0.01	95.19	Cubic
MDT-MST	<0.01	12.94	Quadratic
Specificity			
MDT-0.5T	<0.01	74.46	Linear
MDT-KMT	<0.01	69.44	Quadratic
MST-0.5T	<0.01	54.93	Linear
MST-KMT	0.02	40.68	Linear
MDT-MST	0.38	8.75	Cubic

threshold value around 0.5 in a wide range of prevalence conditions. Although MDT- and MST-criteria model predictions are, in general, significantly more accurate, KMT and 0.5T can be used in some circumstances: (i) when accurateness in predicting presences is the objective and the prevalence is high, and (ii) when we want to predict absences but the prevalence is low. However, we must consider that, in the first case, the distribution of the species will be overestimated while, in the second case, it will be underestimated. If the Kappa statistic is used to measure model accuracy, then KMT and 0.5T criteria outperform in a prevalence interval of 0.1–0.5, approximately. Nevertheless, their performance differs only negligibly with respect to MDT and MST.

The measurement and meaningfulness of accuracy estimations depends on the purpose of the research, leading to varying concerns about accuracy. For example, a threshold optimizing species absences may lead to a suboptimal classification when omission errors are undesirable. While it is frequently assumed that commission and omission errors are equal costwise, in conservation it is probably more costly to classify a recognized presence site as absence than vice versa (Fielding, 2002; but see Loiselle et al., 2003). Omission errors should therefore be avoided and sensitivity favored; in this situation, MDT and MST are the threshold criteria that should be employed. On the contrary, if commission errors are considered more costly, MDT is the only criteria which produce higher specificity values. Hence, we recommend the MDT threshold criteria as the one of more general use (see also Liu et al., 2005; Jiménez-Valverde and Lobo, 2006).

It is important to highlight MDT- and MST-criteria superiority over the other two in predicting presences when the prevalence of presences is low. Low prevalence scores usually
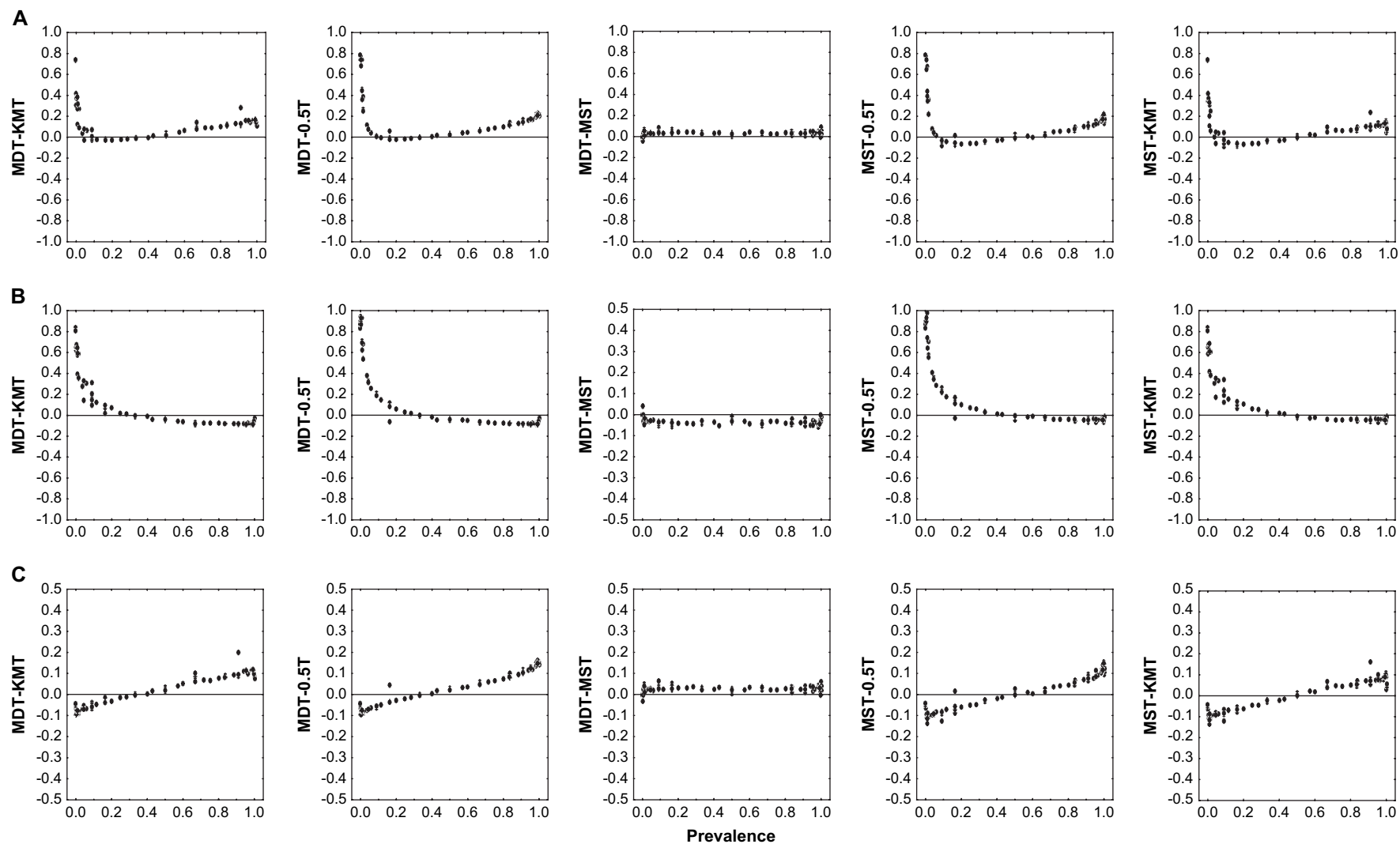
Fig. 3 – Relative different performance between good-performance (MDT and MST) and bad-performance (KMT and 0.5T) criteria along the prevalence of occurrence gradient (A, Kappa statistic; B, sensitivity; C, specificity).

characterize cases in which the distribution area of the species is significantly narrower than the whole spatial extent under study. This situation is a common pattern in real situations. However, MDT- and MST-criteria apparently do not show such superiority when predicting absences at high prevalence scores. Our virtual species has 4.6 times more absences than presences and, so, there are many absences far from the mean environmental conditions of the species. This implies that many absences will have very low logistic probability scores. Consequently, any threshold will be able to correctly classify a great proportion of absences. We suggest than in more balanced situations between the extent analyzed and the species distribution, MDT and MST criteria would considerably outperformed 0.5T and KMT when predicting absences at high prevalence scores.

Manel et al. (2001) questioned the reliability of predictions when the MST criterion is used. They extrapolated their models for aquatic invertebrates to Himalayan regions different from those where the models were trained and observed that potential distributions of rare species were overestimated. However, we think such overestimation is probably not due to the threshold criteria chosen, but to the loss of accuracy of models extrapolated to areas different from those used in model design (Fielding and Haworth, 1995; Marsden and Fielding, 1999).

As reliable absence data are scarcely available, presence-only modeling is now deserving special attention (Elith et al., 2006). In such cases, background data are used for absence, randomly selecting absences from all the locations lacking presence data (see Elith et al., 2006). When this method is used, false absences will likely be included in the dependent variable. If the true distribution of the focus species is directly or indirectly related with the predictors chosen, and if the amount of false absences is low, these false absences will have a mean logistic probability higher that the one assigned to the true absences. Then, the application of a threshold such as MDT will correctly classify most of those false absences as presences when randomly distributed (J.M. Lobo, A. Jiménez-Valverde, J. Hortal, submitted). However, if the amount of false data is high or if the predictors are able to account for those false absences, no threshold will be able to correctly classify them as presences (J.M. Lobo, A. Jiménez-Valverde, J. Hortal, submitted).

On occasion, some authors have failed to point out the threshold used (e.g. Teixeira and Arntzen, 2002), a practice which should be avoided. The criteria employed for deciding the threshold, whatever it is, as well as the threshold value itself, should be specified so that readers can reach their own conclusions.

Acknowledgments

REFERENCES

Allredge, J.R., Ratti, J.T., 1986. Comparison of some statistical techniques for analysis of resource selection. Journal of Wildlife Management 50, 157–165.

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

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

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 - Biological Series B269, 1971–1980.

Araújo, M.B., Cabeza, M., Thuiller, W., Hannah, L., Williams, P.H., 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection method. Global Change Biology 10, 1618–1626.

Araújo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. Global Change Biology 11, 1504–1513.

Araújo, M.B., Thuiller, W., Pearson, R.G., 2006. Climate warming and the decline of amphibians and reptiles in Europe. Journal of Biogeography 33, 1712–1728.

Austin, M.P., in press. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecologival Modelling.

Barbosa, A.M., Real, R., Olivero, J., Vargas, J.M., 2003. Otter (*Lutra lutra*) distribution modeling at two resolution scales suited to conservation planning in the Iberian Peninsula. Biological Conservation 114, 377–387.

Berg, Å., Gärdenfors, U., von Proschwitz, T., 2004. Logistic regression models for predicting occurrence of terrestrial molluscs in southern Sweden—importance of environmental data quality and model complexity. Ecography 27, 83–93.

Bonn, A., Schröder, B., 2001. Habitat models and their transfer for single and multi species groups: a case study of carabids in an alluvial forest. Ecography 24, 483–496.

Brotons, L., Thuiller, W., Araújo, M.B., Hirzel, A.H., 2004. Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27, 437–448.

Chefaoui, R.M., Hortal, J., Lobo, J.M., 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.

Clark Labs, 2003. Idrisi Kilimanjaro. GIS software package. Clark Labs, Worcester, MA.

Cramer, J.S., 1999. Predictive performance of binary logit model in unbalanced samples. *Journal of the Royal Statistical Society: Series D (The Statistician)* 48, 85–94.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

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

Ferrier, S., Watson, G., 1997. An evaluation of the effectiveness of environmental surrogates and modeling techniques in

predicting the distribution of biological diversity. Environment Australia, Canberra, Australia.

Fielding, A.H., 2002. What are the appropriate characteristics of an accuracy measure? In: Scott, J.M., Heglund, P.J., Haufler, J.B., Morrison, M., Raphael, M.G., Wall, W.B., Samson, F. (Eds.), Predicting Species Occurrences. Issues of Accuracy and Scale. Island Press, Covelo, CA, pp. 271–280.

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.

Fleishman, E., Mac Nally, R., Fay, J.P., Murphy, D.D., 2001. Modeling and predicting species occurrence using broad scale environmental variables: an example with butterflies of the Great Basin. Conservation Biology 15, 1674–1685.

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

Guisan, A., Theurillat, J.-P., Kienast, F., 1998. Predicting the potential distribution of plant species in an alpine environment. Journal of Vegetation Science 9, 65–74.

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

Godown, M., Peterson, A.T., 2000. Preliminary distributional analysis of US endangered bird species. Biodiversity and Conservation 9, 1313–1322.

Harrell, F.E.J., 2001. Regression modelling strategies: with applications to linear models, logistic regression, and survival analysis. Springer, New York.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978.

Hirzel, A.H., Helfer, V., Metral, F., 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling 145, 111–121.

Hosmer, D.W., Lemeshow, S., 1989. Applied Logistic Regression. Wiley, New York.

Hugall, A., Moritz, C., Moussalli, A., Stanisic, J., 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail Gnarosophia bellendenkerensis (Brazier 1875). Proceedings of the National Academy of Sciences, USA 99, 6112–6117.

Jiménez-Valverde, A., Lobo, J.M., 2006. The ghost of unbalanced species distribution data in geographical model predictions. Diversity and Distributions 12, 521–524.

Kadmon, R., Farber, O., Danin, A., 2003. A systematic analysis of factors affecting the performance of climatic envelope models. Ecological Applications 13, 853–867.

Legendre, P., Legendre, L., 1998. Numerical Ecology. Elsevier, Amsterdam.

Lehmann, A., Overton, J.McC., Austin, M.P., 2002. Regression models for spatial prediction: their role for biodiversity and conservation. Biodiversity and Conservation 11, 2085–2092.

Li, W., Wang, Z., Ma, Z., Tang, H., 1997. A regression model for the spatial distribution of red-crown crane in Yancheng Biosphere Reserve, China. Ecological Modelling 103, 115–121.

Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28, 385–393.

Lobo, J.M., Verdu, J.R., Numa, C., 2006. Environmental and geographical factors affecting the Iberian distribution of flightless Jekelius species (Coleoptera: Geotrupidae). Diversity and Distributions 12, 179–188.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., Williams, P.H., 2003. Avoiding pitfalls of using

species distribution models in conservation planning. Conservation Biology 17 (6), 1591–1600.

Luoto, M., Pöyry, J., Heikkinen, R.K., Saarinen, K., 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. Global Ecology and Biogeography 14, 575–584.

Luoto, M., Heikkinen, R.K., Pöyry, J., Saarinen, K., 2006. Determinants of the biogeographical distribution of butterflies in boreal regions. Journal of Biogeography 33, 1764–1778.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83 (8), 2248–2255.

Manel, S., Dias, J.M., Buckton, S.T., Ormerod, S.J., 1999a. Alternative methods for predicting species distributions: an illustration with Himalayan river birds. Journal of Applied Ecology 36, 734–747.

Manel, S., Dias, J.M., Buckton, S.T., Ormerod, S.J., 1999b. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecological Modelling 120, 337–347.

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.

Marsden, S., Fielding, A., 1999. Habitat associations of parrots on the Wallacean islands of Buru, Seram and Sumba. Journal of Biogeography 26, 439–446.

McCullagh, P., Nelder, J.A., 1997. Generalized Linear Models: Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Meggs, J.M., Munks, S.A., Corkrey, R., Richards, K., 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.

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

Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133, 225–245.

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

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

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

Pontius, R.G., 2000. Quantification error versus location error in comparison of categorical maps. Photogrammetric Engineering and Remote Sensing 66, 1011–1016.

Pontius, R.G., Batchu, K., 2003. Using the Relative Operating in prediction of location of land cover change in India. Transactions in GIS 7, 467–484.

Reese, G.C., Wilson, K.R., Hoeting, J.A., Flather, C.H., 2005. Factors affecting species distribution predictions: a simulation modeling experiment. Ecological Applications 15, 554–564.

Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H., Possingham, P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecology Letters 9, 1136–1145.

Rojas, A.B., Cotilla, I., Real, R., Palomo, L.J., 2001. Determinación de las áreas probables de distribución de los mamíferos

terrestres en la provincia de Málaga. Galemys 13 (n° especial), 217–229.

Rushton, S.P., Ormerod, S.J., Kerby, G., 2004. New paradigms for modelling species distributions? Journal of Applied Ecology 41, 193–200.

Russell, K.R., Mabee, T.J., Cole, M.B., 2004. Distribution and habitat of Columbia Torrent Salamanders at multiple spatial scales in managed forests of Northwestern Oregon. Journal of Wildlife Management 68, 403–415.

Schadt, S., Revilla, E., Wiegand, T., Knauer, F., Kaczensky, P., Breitenmoser, U., Bufka, L., Červený, J., Koubek, P., Huber, T., Staniša, C., Trepl, L., 2002. Assessing the suitability of central European landscapes for the reintroduction of Eurasian lynx. Journal of Applied Ecology 39, 189–203.

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

Segurado, P., Araújo, M.B., 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography 31, 1555–1568.

StatSoft Inc, 2001. STATISTICA (data analysis software system). www.statsoft.com.

Swets, K., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

Teixeira, J., Arntzen, J.W., 2002. Potential impact of climate warming on the distribution of the Golden-striped salamander, Chioglossa lusitanica, on the Iberian Peninsula. Biodiversity and Conservation 11, 2167–2176.

Termansen, M., McClean, C.J., Preston, C.D., 2006. The use of genetic algorithms and Bayesian classification to model species distributions. Ecological Modelling 192, 410–424.

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

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, I.C., 2005. Climate change threats to plant diversity in Europe. Proceedings of the National Academy of Sciences, USA 102, 8245–8250.

Wiens, J.J., Graham, C.H., 2005. Niche conservatism: integrating Evolution, Ecology and conservation Biology. Annual Review of Ecology. Evolution and Systematics 36, 519–539.

Wilson, K.A., Westphal, M.I., Possingham, H.P., Elith, J., 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation 122, 99–112.

Wintle, B.A., Kavanagh, R.P., McCarthy, M.A., Burgman, M.A., 2005. Estimating and dealing with detectability in occupancy surveys for forest owls and arboreal marsupials. Journal of Wildlife Management 69, 905–917.

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

Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 39, 561–577.