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

Alberto Jiménez-Valverde*, Jorge M. Lobo

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; Zaniiewski et al., 2002; Engler et al., 2004). However, these methods are useful because available 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 realistic 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 km² (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 absence where actually presences were observed; i.e. the false negative fraction) and omission errors (model predictions of species absence where actually absent; i.e. the false positive 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...
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.

<table>
<thead>
<tr>
<th></th>
<th>Kappa</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F(3,320) )</td>
<td>29.09***</td>
<td>10.97***</td>
<td>4.48*</td>
</tr>
<tr>
<td>( a ) MST</td>
<td>0.734 ± 0.003b,d</td>
<td>0.956 ± 0.002b,d</td>
<td>0.898 ± 0.002c</td>
</tr>
<tr>
<td>( b ) 0.5T</td>
<td>0.599 ± 0.024bc,c</td>
<td>0.799 ± 0.035bc,c</td>
<td>0.894 ± 0.009c</td>
</tr>
<tr>
<td>( c ) MDT</td>
<td>0.766 ± 0.002b,d</td>
<td>0.926 ± 0.002b,d</td>
<td>0.923 ± 0.001b,c,d</td>
</tr>
<tr>
<td>( d ) KMT</td>
<td>0.644 ± 0.016bc,c</td>
<td>0.833 ± 0.028bc,c</td>
<td>0.898 ± 0.008c</td>
</tr>
</tbody>
</table>

---

**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 ≤ 0.003).

<table>
<thead>
<tr>
<th></th>
<th>Kappa</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MST/0.5T</td>
<td>0.06</td>
<td>0.42</td>
<td>0.16</td>
</tr>
<tr>
<td>MST/MDT</td>
<td>0.52</td>
<td>0.40</td>
<td>0.35</td>
</tr>
<tr>
<td>MST/KMT</td>
<td>0.02</td>
<td>0.46</td>
<td>0.18</td>
</tr>
<tr>
<td>0.5T/MST</td>
<td>0.33</td>
<td>0.24</td>
<td>-0.04</td>
</tr>
<tr>
<td>0.5T/KMT</td>
<td>0.93</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>MDT/KMT</td>
<td>0.21</td>
<td>0.27</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

---

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.
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

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.

<table>
<thead>
<tr>
<th></th>
<th>Sample-size Prevalence</th>
<th>% Exp. Dev.</th>
<th>% Exp. Dev.</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kappa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDT-0.5T</td>
<td>0.05</td>
<td>95.94</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td>MDT-KMT</td>
<td>0.24</td>
<td>82.21</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td>MST-0.5T</td>
<td>0.03</td>
<td>95.97</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td>MST-KMT</td>
<td>0.26</td>
<td>85.54</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td>MDT-MST</td>
<td>0.46</td>
<td>7.30</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDT-0.5T</td>
<td>0.08</td>
<td>96.38</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td>MDT-KMT</td>
<td>0.01</td>
<td>93.31</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td>MST-0.5T</td>
<td>0.09</td>
<td>97.75</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td>MST-KMT</td>
<td>0.01</td>
<td>95.19</td>
<td>Cubic</td>
<td></td>
</tr>
<tr>
<td>MDT-MST</td>
<td>&lt;0.01</td>
<td>12.94</td>
<td>Quadratic</td>
<td></td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDT-0.5T</td>
<td>&lt;0.01</td>
<td>74.46</td>
<td>Linear</td>
<td></td>
</tr>
<tr>
<td>MDT-KMT</td>
<td>&lt;0.01</td>
<td>69.44</td>
<td>Quadratic</td>
<td></td>
</tr>
<tr>
<td>MST-0.5T</td>
<td>&lt;0.01</td>
<td>54.93</td>
<td>Linear</td>
<td></td>
</tr>
<tr>
<td>MST-KMT</td>
<td>0.02</td>
<td>40.68</td>
<td>Linear</td>
<td></td>
</tr>
<tr>
<td>MDT-MST</td>
<td>0.38</td>
<td>8.75</td>
<td>Cubic</td>
<td></td>
</tr>
</tbody>
</table>

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.
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

This paper has been supported by a Fundación BBVA project (Diseño de una red de reservas para la protección de la Biodiversidad en América del sur austral utilizando modelos predictivos de distribución con taxones hiperdiversos) and a MEC Project (CGL2004-04309), as well as by a Museo Nacional de Ciencias Naturales/C.S.I.C./Comunidad de Madrid PhD grant.

References


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.
Hugall, A., Moritz, C., Moussalli, A., Stanisic, J., 2002. Reconciling paleodistribution models and comparative phylogeography in the wet Tropics rainforest land snail {G}nara{ph}os{dot}h{um} bellenden{dot}kern{e}n{r}esis (Brazil 1875). Proceedings of the National Academy of Sciences, USA 99, 6112–6117.
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.