Choice of threshold alters projections of species range shifts under climate change

H.K. Nenzén [a,b,*], M.B. Araújo [b,c]

[a] Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden
[b] Rui Nabeiro Biodiversity Chair, CIBIO, University of Évora, Casa Cordovil, 2° Andar, Rua Dr. Joaquim Henrique da Fonseca, 7000-890 Évora, Portugal
[c] Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Calle José Gutiérrez Abascal, 2, 28006, Madrid, Spain

ARTICLE INFO

ABSTRACT

One of the least explored sources of algorithmic uncertainty in bioclimatic envelope models (BEM) is the selection of thresholds to transform modelled probabilities of occurrence (or indices of suitability) into binary predictions of species presence and absence. We investigate the impacts of such thresholds in the specific context of climate change. BEM for European tree species were fitted combining 9 climatic models and emissions scenarios, 7 modelling techniques, and 14 threshold-setting techniques. We quantified sources of uncertainty in projections of turnover, and found that the choice of the modelling technique explained most of the variability (39%), while threshold choice explained 25% of the variability in the results, and their interaction an additional 19%. Choice of future climates explained 9% of total variability among projections. Estimated species range shifts obtained by applying different thresholds and models were grouped by IUCN-based categories of threat. Thresholds had a large impact on the inferred risks of extinction, producing 1.7- to 9.9-fold differences in the proportions of species projected to become threatened by climate change. Results demonstrate that threshold selection has large – albeit often unappreciated – consequences for estimating species range shifts under climate change.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Bioclimatic envelope models (BEMs) characterise species climatic requirements by relating species occurrences with aspects of climate. These models have been used for a variety of theoretical as well as applied purposes (e.g., Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005). There are a number of algorithmic uncertainties that contribute to uncertainty in projections, apart from uncertainties in the data and biology of the species (for reviews see Fielding and Bell, 1997; Heikkinen et al., 2006; Araújo and New, 2007). One of the least explored sources of uncertainty is the rule to transform probabilities of occurrence (or indices of suitability) produced by models into binary predictions of species presence and absence. There are potentially as many rules for setting thresholds (or cut-offs) as modelling methods and the choice is often made arbitrarily since no guidelines for selection exist. Here, we investigate the impacts of different threshold-optimization methods in the specific context of BEMs used to study the effects of climate change on species range changes.

Possible impacts of threshold selection on species range prediction in BEMs have been discussed previously in the literature. Fielding and Bell (1997) stated that a fixed threshold to transform model output will exaggerate prediction errors. The authors highlighted that choice of modelling method may influence the values of probabilities of occurrence, and thus a single threshold for different models would be unsuitable.

To avoid fixed thresholds many threshold-optimization methods have been proposed (Table 1). We identified three families of approaches: (I) fixed; (II) data-driven – using species data and predicted probability values; and (III) accuracy-based – using the threshold that produces the best agreement between the evaluation data and the original data. The goal of accuracy-based thresholds is to minimize prediction errors across the modelled region: false negatives (modelled absences that are in fact species presences) and false positives (modelled presences that are species absences).

Despite the numerous thresholds available, their effects have rarely been investigated, and never in climate change studies. In previous studies that did investigate thresholds, either the number of threshold-setting methods or the number of modelling methods was low, preventing a thorough evaluation of threshold impacts (Table 2). More importantly, most studies evaluated threshold-setting methods by comparing their ability to discriminate between correctly predicted presences and absences on a random subset of

* Corresponding author at: Rui Nabeiro Biodiversity Chair, CIBIO, University of Évora, Casa Cordovil, 2° Andar, Rua Dr. Joaquim Henrique da Fonseca, 7000-890 Évora, Portugal. Tel.: +351 266 759 350; fax: +351 266 744 969.
E-mail address: hedvig.nenzen@gmail.com (H.K. Nenzén).

Table 1
Fourteen threshold-setting methods for presence–absence data used in this study (abbreviations in parenthesis). Modified after Liu et al. (2005), Pearson (2007), and Freeman and Moisen (2008). Citations include references where the threshold setting methods was initially referred to followed by studies that subsequently used the threshold in models of species distributional changes under climate change. For accuracy-based thresholds we calculated the values in the confusion matrix: TP = true positives, TN = true negatives, FP = false positives, FN = false negatives. Sensitivity = TP/TP + FN. Specificity = TN/FP + TN.

	Threshold	Description	Citations in bioclimatic envelope modelling
1	Fixed	0.5 in this study	Manel et al. (1999), Hijmans and Graham (2006), Buckley et al. (2010)
	Data-driven thresholds		
2	Observed prevalence (ObsPrev)	Using the original species prevalence as the threshold	Cramer (2003), Araújo and Luoto (2007), Baselga and Araújo (2009)
3	Predicted prevalence same as observed (PredPrev = Obs)	Maintain the original prevalence	Hartley et al. (2006), Dormann et al. (2008)
4	Average probability (AvgProb)	Taking the mean of the probabilities of occurrence of occupied locations for presence/absence data as the threshold	Cramer (2003)
5	Mid-point probability (MidptProb)	Taking the median of the probabilities between the occupied and unoccupied sites	Fielding and Haworth (1995)
	Accuracy-based thresholds		
6	Plot based precision-recall (PRplotbased)	Minimize the distance to the 1,1 corner of the precision (TP/TP + FP) against recall (sensitivity) plot	Liu et al. (2005)
7	Minimize precision-recall (PRmin)	Minimize the difference between the precision (TP/TP + FP) and recall (sensitivity)	Schapire et al. (1998)
8	Overall prediction success (OPS)	Maximize OPS (TP + TN/number of presence/absence points)	Fielding and Bell (1997)
9	F	Maximize $F = 1/\alpha/P + (1 - \alpha)/R$ $\alpha = 0.5$ (no preference precision or recall)	Schapire et al. (1998), Drake et al. (2006)
10	Kappa	Maximize Cohen's kappa statistic	Huntley et al. (1995), Berry et al. (2002), Segurado and Araújo (2004), Araújo et al. (2005b), Elith et al. (2006)
11	Maximize sum of sensitivity and specificity (SeSpmax)	Maximize the sum of sensitivity and specificity	Cantor et al. (1999), Manel et al. (2001), Svenning et al. (2008)
12	Equalize sensitivity and specificity (SeSpeql)	Minimize the absolute difference between the sensitivity and specificity	Fielding and Bell (1997), Pearson et al. (2004, 2006)
13	TSS	True skill statistic, sensitivity + specificity-1	Allouche et al. (2006), Keenan et al. (2011)
14	ROC	Minimize the distance to the 0,1 corner of the sensitivity against 1-specificity curve (receiver operating curve)	Cantor et al. (1999), Pearce and Ferrier (2000), Araújo et al. (2005b), Diniz-Filho et al. (2009), Lobo et al. (2008)

the original species distributions data (Table 2). Yet measures of discrimination accuracy obtained with non-independent evaluation data provide limited information about the ability of BEMs to forecast species potential range shifts under climate change scenarios (Araújo et al., 2005a; Araújo and Rahbek, 2006). So, instead of using discrimination accuracy to explore the consequences of

threshold selection under climate change, we employ alternative methods.

In this study, the effects of thresholds on forecasts are analysed on the basis of consistency in (i) temporal species turnover for each location and (ii) modelled range changes for each species. First, to assess sources of uncertainty in turnover results, we quantify the

Table 2
Published studies that explicitly investigated thresholds in bioclimatic envelope models (BEM). Provided are the numbers of BEM, thresholds, and the main conclusions in each study. Note that not all studies investigated performances of thresholds under climate change. See Table 1 for abbreviations.

	Models	Thresholds	Species	Evaluation data	Evaluation method	Conclusions
Manel et al. (2001)	1	2	34 invertebrate families	Cross-evaluation and data from different area	Predictive accuracy	No evidence for effects of prevalence on kappa. ROC proposed as a threshold-setting technique
Thuiller (2004)[a]	4	2	1350 plant species	Turnover results	Component loadings on PCA axes	Impact of threshold-setting method highlighted
Araújo et al. (2005b)[a]	7	2	116 bird species	Cross-evaluation and data from different time	Predictive accuracy	Kappa was more accurate than ROC
Liu et al. (2005)	1	12	2 plant species[b]	Cross-evaluation	Predictive accuracy	ObsPrev, AvgProb, SeSpmax, SeSpeql and ROC were most accurate thresholds
Allouche et al. (2006)	1	2	128 plant species	Independent inventory data from same area	Predictive accuracy	TSS proposed as a threshold-setting technique
Jiménez-Valverde and Lobo (2007)	1	4	1 virtual species[b]	Cross-evaluation	Predictive accuracy	SeSpmax and SeSpeql were most accurate thresholds
Freeman and Moisen (2008)	1	11	13 tree species	Cross-evaluation	Predictive accuracy and change in predicted prevalence	PredPrev = Obs and kappa were most accurate thresholds
Present study[a]	7	14	116 tree species	Species range changes and turnover results	Extinction risk, quantification of sources of uncertainty	Thresholds contributed uncertainty in predictions under climate change

[a] Studies investigated threshold performance under climate change.
[b] Species were sampled at different prevalences.

variability that different thresholds bring to the measures of temporal turnover (e.g., Diniz-Filho et al., 2009). Second, to explore effects of the threshold technique on assessments of species vulnerability to climate change, we associate estimated range changes, to a given category of threat using a simplified interpretation of the IUCN criteria for Red Listing of species. The use of projected range changes from BEMs for grouping species into threat categories has been criticized on theoretical grounds (Akçakaya et al., 2006), and new tools are being devised to couple climatic and population processes thus providing more robust estimates of extinction risk (e.g., Keith et al., 2008). In the meantime, we use the IUCN Red List criteria as a strategy to explore the sensitivity of the range size projections to different thresholds. Note that we are not recommending Red Listing of species directly and exclusively based on range size estimates from BEMs (Akçakaya et al., 2006; Brook et al., 2009).

For this study, BEMs are fitted for 116 European tree species, using seven bioclimatic modelling techniques, 14 threshold-setting methods, and 9 climatic models and scenarios. By using many combinations of modelling algorithms, threshold-setting approaches, and forecasts of future climates we seek to quantify the contribution of different sources of uncertainty in forecasts of species range changes and identify generalities that may arise from assessments of uncertainty. Specifically, we ask: (I) How are assessments of threat and temporal turnover based on BEMs affected by threshold selection? (II) How much variability in turnover values is attributed to the choice of the threshold? (III) How is variability in BEM projections spatially and environmentally distributed?

2. Methods

2.1. Climate data

A set of aggregated climate parameters for the present time period were derived from an updated version of climate data provided by New et al. (2000). The updated data provides monthly values for the years 1901–2000 in a 10′ grid resolution (Mitchell et al., 2004; Schröter et al., 2005, average resolution 18 km), and were aggregated to the species data resolution. We employed the mean values of eleven different bioclimatic parameters for the period of 1961–1991 (referred to as baseline data). To minimize model overfitting and ensure comparability across model projections we selected a smaller set of uncorrelated variables for inclusion in the models, after performing a principal components analysis (PCA) of the climate data using all cells. The PCA identified three axes that together explain 99.6% of the variance in the climate data. We retained the three climatic variables with the highest component loadings in the first (growing degree days, −0.805 component loading), second (annual precipitation, 0.811), and third (mean annual radiation, 0.728) axes (for similar procedure, see also Baselga and Araújo, 2009). These variables are expected to summarise important abiotic factors that directly limit the distributions of plant species (e.g., Woodward, 1987; Sykes et al., 1996). Note that we fitted BEM with the untransformed climatic variables and not the PCA axes.

Nine different predicted climates for 2050 were used to investigate uncertainty arising from models and scenarios (Schröter et al., 2005). The selection included simulated future climates with different general circulation models (GCM) and emissions scenarios (Nakicenovic et al., 2000). We used four general circulation models (GCM each with one to several emissions scenarios, as follows: HadCM version 3 (scenarios A1FI, A2, B1, B2), NCAR parallel climate model (scenarios A1 and B1), MRI coupled general circulation model 2 (scenarios A1 and A2) and CSIRO model (scenario A2). This amounted to nine combinations of climate models and scenarios in total, which will be referred to as 'climates'.

2.2. Species data

For this study, we considered native tree species distributed across Europe. The species presence–absence data are a subset of Atlas Flora Europaeae (AFE, Jalas and Suominen (Eds.)., 1972–1996), which was digitized by Lahti and Lampinen (1999). Only trees were chosen because: (i) their distribution and ecology is relatively well known and the data has been extensively used in other modelling studies (e.g., Thuiller et al., 2005; Baselga and Araújo, 2009; Svenning et al., 2008; Hanspach et al., 2010; Araújo et al., 2011), (ii) they are representative of the overall European plant richness (Spearman correlation $\rho = 0.80$, $P < 0.001$ with the overall richness of the AFE data set, Araújo and Williams, 2000), and (iii) they are long-lived organisms and their distribution is relatively stable in comparison with some other groups. Data were originally located in 4419 UTM (Universal Transverse Mercator) 50 km × 50 km grid cells, but we used only 2122 grid cells, excluding most of the eastern European countries (except for the Baltic States), because of low recording efforts in these areas (Williams et al., 2000). Species occurring in fewer than 25 grid cells were excluded from analyses to avoid problems associated with modelling species with small sample sizes (Stockwell and Peterson, 2002; Wisz et al., 2008); however, with this rule, together with the data splitting procedure used for calibration and evaluation of the models (70% for calibration and 30% for evaluation, see below), the minimum number of records used for calibrating models was 18. The reduced dataset comprised 116 tree species (Supplementary data A). Original prevalence ranged from 1 to 79% of all grid cells.

2.3. Bioclimatic envelope modelling

To characterise species potential distributions we fitted bioclimatic envelope models (BEMs) using BIOMOD-R, which implements multiple model classes in a single platform (Thuiller et al., 2009) using the freeware and open source R package (R Development Core Team, 2010). We employed seven algorithms that generate continuous predictions of probabilities of occurrence as an output: generalized linear models (GLM, linear and quadratic effects), generalized additive models (GAM, polynomial degree 2), multivariate adaptive regression splines (MARS), classification tree analysis (CTA), flexible discriminant analysis (FDA), random forests (RF, 750 trees), and generalized boosted models (GBM, 2000 trees). See specialized literature for description of modelling methods (e.g., Guisan and Zimmermann, 2000; Franklin, 2010 and citations in BIOMOD documentation). We calibrated the models with a random sample of 70% of the original presence records for each species and projected models to the remaining 30% of grid cells (steps 1 and 2 in Fig. 1).

2.4. Threshold optimization

A threshold is required for transforming continuous probabilities of occurrence (or indices of suitability) from the models into binary presence-absence values. The R-script used for threshold optimization is documented in Supplementary data B. First, all thresholds were calculated (Fig. 1, step 3), then the same thresholds were used as cut-offs for transforming probabilities of occurrence into presence and absence of species in the baseline and future climate projections (Fig. 1, steps 4 and 5). A large number of methodologies exist for optimizing thresholds; here we used 14 of those methods (Table 1 and Liu et al., 2005 for further description).

For the fixed and data-driven thresholds (families I and II, Table 1) we set the threshold value directly from the withheld evaluation data (30% of original distribution data) and the predicted probabilities of presence in the same grid cells. The value of the accuracy-based thresholds (III) was found in an iterative
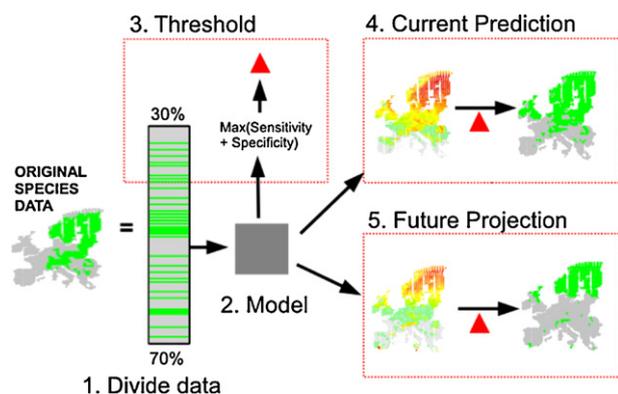
Fig. 1. Flowchart illustrating the steps to optimize one accuracy-based threshold for one species. (1) Original species data were randomly split into calibration and evaluation data. (2) 70% of the data were used to train models (calibration data). (3) The models were projected on the 30% evaluation data. A step-wise sequence of thresholds was applied to predictions to find the threshold (red triangle) that best satisfied the threshold criteria for the selected threshold-setting technique. (4) The model was applied to current climate data and the threshold from step 3 was applied to transform data to binary distribution predictions. (5) The model was applied to future climate data. Figure redrawn from Pearson (2007) with permission. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

fashion. Firstly, we created 1000 testing thresholds from 0 to 1 in successive increments of 0.001 (i.e., 0.001, 0.002, . . ., 0.998, 0.999, 1.000). Secondly, each 'testing' threshold was used as the cut-off to classify the continuous predicted probability values to present or absent – when the values were larger or smaller than the 'testing' threshold respectively. Thirdly, in each grid cell, the predicted range (present/absent) resulting from each 'testing' threshold was compared to the species currently observed range in the AFE data. For any one grid cell, four outcomes were possible: true positives, true negatives, false positives and false negatives. Outcomes for each value of 'testing' threshold over the entire study area were tallied in a confusion matrix (a contingency table of presences and absences in the predicted data and original data, Fielding and Bell, 1997). From the confusion matrix, we were able to calculate the various accuracy-based discrimination indices according to the formulas in Table 1. The value of the 'testing' threshold that generated the prediction that was most accurate and best satisfied the criterion for each method, when compared to the original species distribution, was selected as the final threshold for that method (step 3 in Fig. 1). If several thresholds produced identical accuracy values, we chose the lowest threshold.

Finally, we calibrated models for each species with 100% of the species data so to avoid biases associated with sub-sampling of the data (Araújo et al., 2005b). With the final models, we projected probabilities of occurrence for both the baseline and future climates. Projections were then transformed into binary projections of species presence and absence using previously calculated thresholds. The resulting 882 final projections for each species arise from combinations of 7 modelling techniques, 14 threshold-setting methods, and 9 climates. For the 116 species modelled, 102,312 model outputs were generated.

2.5. Analysis of turnover

In each grid cell, the species temporal turnover between the two time periods was defined as (species gained + species lost)/(species richness + species gained) (e.g., Peterson et al., 2002). This formula measures changes in species composition, which is more informative of range shifts than comparing species richness in two time periods. To distinguish trends in turnover values among the 882

projections for each location, we employed a range of statistical techniques.

First, we performed a principal components analysis (PCA) on the projections of turnover values to determine whether values obtained with certain thresholds and models were more similar than others. It is assumed that if the first PCA axis explains a large proportion of the variation in results, then the axis is close to all values (e.g., Thuiller, 2004). This was not quite true with our results, as the first axis only explained 52% of the variation. Many combinations of methods had a small component loading and there was no clear separation between different threshold-setting methods (maximum loading on the first axis was 0.14).

To examine sources of uncertainty across projections, we partitioned the sources of variance in turnover values resulting from the combination of the three sources of uncertainty in each grid cell. Variance among the 882 turnover values was assessed by three-way ANOVAs with climates, threshold-setting method and model technique as the three factors (Diniz-Filho et al., 2009). The sum of squares of each factor and the interaction of factors was divided by the total sum of squares, to obtain the relative contribution of each factor to the variance.

The relative proportion of variance from models and thresholds was plotted in both geographical and environmental space on the first two axes (annual precipitation and growing degree days). Since climate contributed the least amount to total variance among the three sources of uncertainty (9%), in the further analyses we only retained the projections from one climate (HadCM3 A1FI, which is a fossil fuel intensive scenario).

To further identify the differences in turnover between threshold-setting methods, a clustering technique was used to group the combinations of threshold and modelling methods that generated more similar values of turnover across grid cells. The process involved calculating Euclidean distances between all pairs of threshold-modelling methods. From the Euclidean distance matrix we implemented an agglomerative hierarchical classification (single linkage clustering) to determine the groups of thresholds and models that produced the most similar values of turnover. Non-parametric analysis of similarity (ANOSIM, Clarke and Green, 1988) was then used to test if the resulting clusters were statistically different. A high value of the R global statistic implies that a high degree of separation exists between groups. Permutations of the grouping factor were done 999 times, allowing for the value of the R statistic to be calculated and compared against the null expectation.

2.6. Analysis of species range changes

To assess the consequences of different models and thresholds for conservation prioritization under climate change we calculated potential range changes per species; i.e., the difference between the numbers of grid cells presently predicted to be occupied relative to the number of grid cells lost or gained in the future. Since we were explicitly interested in modelling changes in the potential ranges rather than the actual ones we considered unlimited migration in all calculations.

To assess how different sources of uncertainty affect estimates of species threat under climate change, we calculated how many species would be candidate to IUCN Red Listing if the magnitude of range changes were implemented as a criterion for classifying them into threat categories (e.g., Thomas et al., 2004; Thuiller et al., 2005). Specifically, we used an implementation of the IUCN criterion of population size reduction, projected or suspected to be met within the next 10 years or three generations, whichever is the longer (IUCN, 2001, p. 16). That is, we applied criterion A3 that refers to a decline in the area of occupancy. Hence, to identify a 'Critically Endangered' species we used a range-size reduction of 80%; 'Endangered' 50%; and 'Vulnerable' 30%. Because trees were modelled,
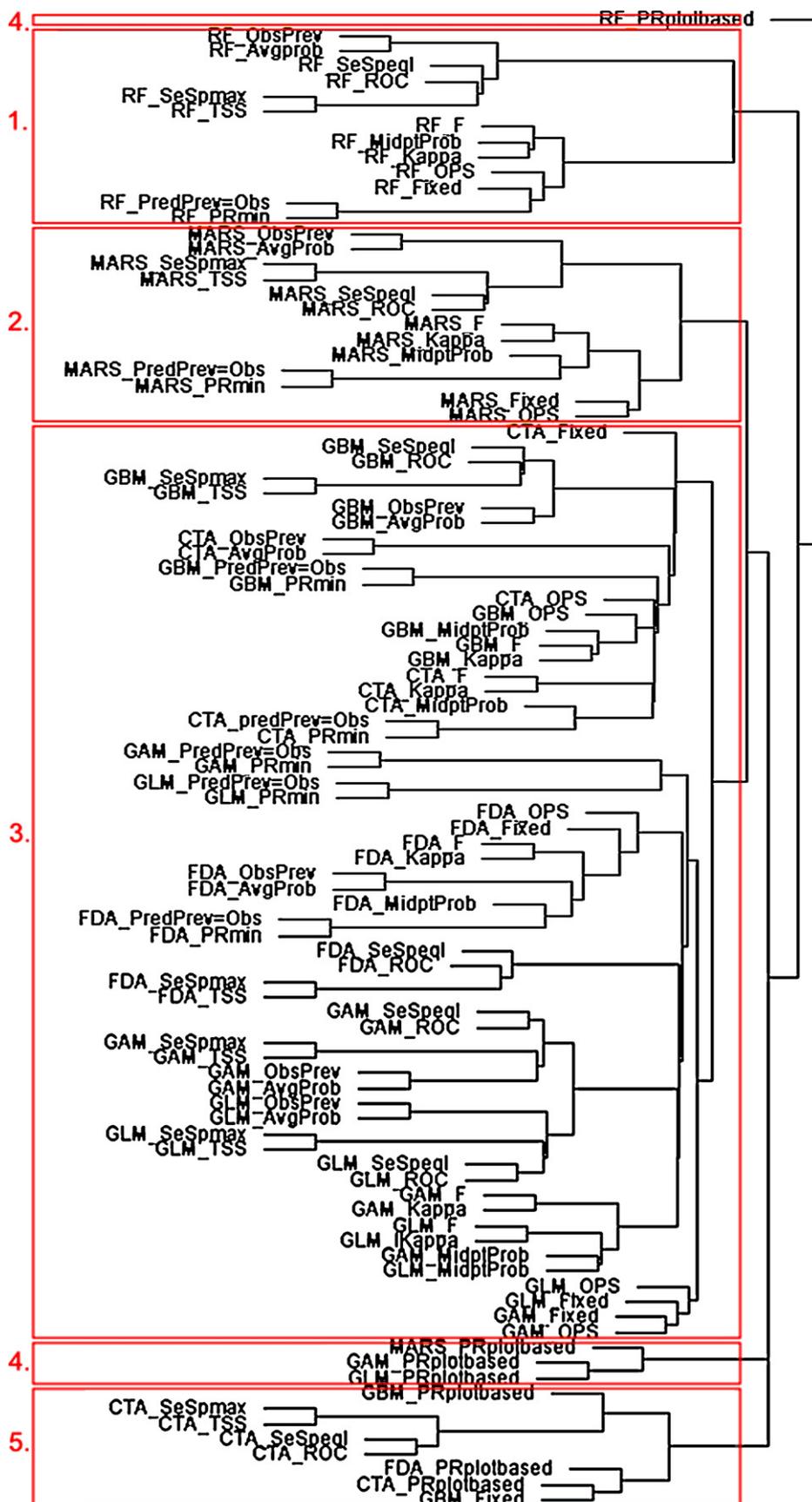
we assumed a long generation time and that range changes from 2000 to 2050 correspond to three generations. However, for some species, the 50 year period considered might correspond to just two generations, so there is a chance that risks are overestimated for those species. Species projected to lose their entire range were considered critically endangered and not extinct, as the European trees modelled could survive outside the study area. To find which source of uncertainty had the largest effect, we estimated the total threat to species as the maximum number of species potentially classified as Red-listed divided by its minimum. We pooled the percentages of threatened species in all threat categories, so that the threat estimate provides an indication of total effects. All analyses were carried out in R statistical environment (R Development Core Team, 2010), using packages BIOMOD (Thuiller et al., 2009) and vegan (Oksanen et al., 2010).

3. Results

3.1. How are assessments of threat and temporal turnover based on BEMs affected by threshold selection?

Choice of models and thresholds both contributed to variation in estimates of temporal turnover values and extinction risk. Because climate contributed a small amount to the total variance (9%), we only show results from one climate model and scenario (HadCM3 A1FI). The clustering dendrogram (Fig. 2) revealed that sources of variation among turnover values were related mainly to the choice of the modelling method (groups 1–2) and the choice of the threshold-setting method (groups 3–4). We separated the following groups: (1) RF modelling method; (2) MARS modelling method; (3) accuracy-based thresholds; (4) PR plot-based thresholds; (5) various thresholds and modelling methods. Group 3 contained all types of threshold setting-methods. Within clusters, further clustering with data-driven or accuracy-based thresholds could be observed. The ANOSIM analysis returned an R statistic of 0.24 ($P = 0.001$), indicating weak grouping among different models and thresholds.

The projections of range changes were greatly affected by the threshold-setting method, as expressed by varying estimates of threat. The percentage of species forecasted to become critically endangered in 2050 ranged from 7% to 69% depending on the modelling method and threshold technique (Table 3). The threshold/model combinations that yielded the highest percentage of threatened species were the GBM with PRmin and PredPrevObs thresholds, which projected that 80% of the European tree species would be candidates for Red Listing under climate change. Applying the same threshold-setting method to different model outputs caused the proportion of threatened species percentages to vary more than 11-fold. Within projections from one modelling method, estimates of threat differed up to 9.9-fold, due to the threshold-setting method. Estimates of threat were, therefore, sensitive to both modelling and threshold-setting method. Apart from these extreme combinations, magnitudes of change were 1.8–2.6 within projections from one modelling method, due to the threshold-setting method, and 2.0–6.7 for the same threshold method applied to different modelling methods.

3.2. How much variance in projections is attributed to the choice of the threshold?

The ANOVA applied to each grid cell indicated that the selection of modelling method contributed to greater variance in projections of species temporal turnover than the selection of future climate and modelling method (Fig. 3). Across all grid cells, the proportion of variance explained by the modelling method contributed to a

Table 3
Percentage of European tree species that became candidates for IUCN Red Listing. The first value in each cell is the percentage of species that could become 'Critically Endangered' under the corresponding modelling class (column, acronyms in Section 2.3)and threshold-setting method (rows, acronyms in Table 1 The second number is the percentage of species that could potentially become 'Endangered'. The third number is the percentage of potentially 'Vulnerable' species. See text and Table 2 for meanings of abbreviations.

	CTA CR	CTA EN	CTA VU	GAM CR	GAM EN	GAM VU	GBM CR	GBM EN	GBM VU	GLM CR	GLM EN	GLM VU	MARS CR	MARS EN	MARS VU	FDA CR	FDA EN	FDA VU	RF CR	RF EN	RF VU	Magnitude of change
Fixed	1	17	14%	27	27	12%	69	2	3%	0	3	8%	5	27	15%	16	21	21%	5	26	15%	6.7
ObsPrev	12	14	15%	3	15	13%	2	14	15%	7	22	16%	10	31	19%	9	31	13%	0	5	14%	3.2
PredPrevObs	7	14	13%	7	23	22%	36	24	20%	1	1	5%	12	31	15%	9	22	21%	5	26	16%	11.4
AvgProb	11	14	17%	3	16	14%	3	19	17%	10	17	14%	11	30	20%	10	30	13%	0	5	14%	3.2
MidptProb	6	21	12%	0	18	23%	28	23	14%	7	7	6%	7	34	22%	5	22	24%	1	22	19%	3.3
PRplotbased	6	7	21%	8	14	15%	21	9	9%	14	29	22%	3	16	19%	7	10	11%	3	25	15%	2.3
PRmin	7	14	12%	7	23	23%	36	24	20%	1	1	5%	12	31	15%	9	22	21%	5	26	16%	11.4
OPS	6	9	18%	24	27	14%	44	23	11%	1	3	9%	5	28	13%	17	21	16%	4	28	10%	6
Fmeasure	7	14	16%	7	25	13%	32	23	10%	9	22	16%	8	33	18%	8	25	21%	3	27	12%	1.8
Kappa	7	11	21%	8	23	15%	34	22	12%	6	16	12%	8	34	17%	8	25	22%	4	30	9%	2
SeSpmax	15	16	20%	3	14	16%	0	23	23%	27	30	12%	8	25	22%	10	24	12%	5	14	14%	2.1
SeSpeql	13	22	16%	0	22	14%	3	22	20%	36	28	3%	6	28	17%	9	26	13%	7	11	13%	2.2
TSS	15	16	20%	3	14	16%	0	23	23%	27	30	12%	8	25	22%	10	24	12%	5	14	14%	2.1
ROC	16	20	20%	2	20	15%	2	22	25%	28	27	9%	7	29	19%	11	25	11%	3	11	13%	2.4
Magnitude of change	1.8			2.1			2.6			9.9			1.7			2.1			2.5			

Fig. 2. Dendrogram from hierarchical single linkage clustering analysis of turnover values, based on the Euclidean distance matrix. At each branch the first acronym is the modelling method (acronyms in Section 2.3) and the second acronym is the threshold-setting method (acroynyms listed in Table 1). We separated the following groups: (1) RF modelling method; (2) MARS modelling method; (3) accuracy-based thresholds; (4) PR plot-based thresholds; (5) various thresholds and modelling methods. The global ANOSIM was 0.24 ($P = 0.001$).
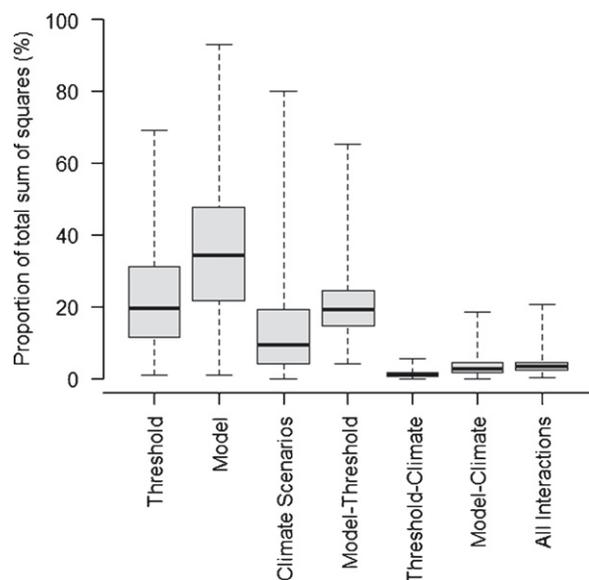
Fig. 3. Sources of variance in species turnover from two-way ANOVAs in all grid cells. Boxes show median, quantiles and complete range of proportions of total sums of squares for each source of uncertainty and their interactions (indicated by hyphens).

median of 34% of variance, while threshold selection contributed 20%. The interaction between thresholds and modelling methods contributed 19%. Choice of climate contributed to 9% of variance.

3.3. How is variability in BEM projections spatially and environmentally distributed?

Variability across projections of species temporal turnover was spatially (Fig. 4, upper panel) and climatically structured (Fig. 4, lower panel). In parts of southern Europe, such as Spain, Greece

and Northern Europe, thresholds represented a higher proportion of the sum of squares thus contributing with most of the variability. In Northern France, Germany, England and Scandinavia, it was the choice of the modelling technique that contributed with more variability, (Fig. 4, upper panel). When the values in climatic space were plotted, model selection explained a greater proportion of the variability in modelled turnover values (Fig. 4, lower panel). In the centre of climatic space where all models produced similar projections, the choice of threshold, in turn, contributed to more variability.

4. Discussion

In this study we evaluated the impact of threshold-setting methods on climate-change induced projections of species range shift using bioclimatic envelope models. We found that the choice of the modelling method and the threshold method altered estimates of species threat that are based on range changes. Model selection generated up to 11.4-fold differences in the proportion of tree species projected to become threatened under climate change (Table 3). For the same threshold method, but with projections from different modelling methods, the largest magnitude of difference produced was 9.9. Most of this variation was from the PRmin threshold and the GBM modelling method. Even the results from most common thresholds such as Kappa, TSS and ROC were sensitive to threshold and model selection, which caused estimates to double in size.

We show that a hitherto unappreciated source of uncertainty – choice of threshold – causes variability in both estimates of species range change and temporal turnover. Most previous studies investigating uncertainty have reported that uncertainty in such estimates was mainly explained by modelling method (Thuiller, 2004; Pearson et al., 2006; Dormann et al., 2008; Diniz-Filho et al., 2009; Buisson et al., 2010; Diniz-Filho et al., 2010; but see Araújo et al., 2005b). One study proposed an approach to break-down results into specific questions relevant to conservation managers
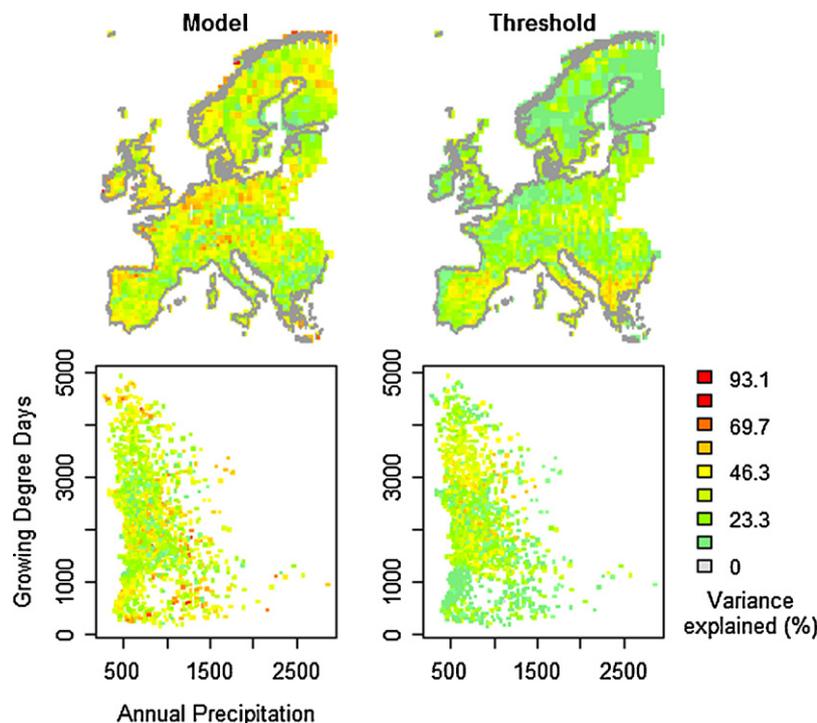
Fig. 4. Distribution of the proportion of variance (in percentages of total sum of squares) for ANOVAs of turnover values among 116 species of European trees. Results are shown in geographical space (upper panel) and environmental space (lower pane, axes represent the first two environmental variables used for modelling in this study).

(e.g., uncertainty in spatial shift of a certain biome, Mbogga et al., 2010). Their method showed that in some cases climate models and emission scenarios might represent a larger source of uncertainty. However, none of the above mentioned studies included effects of thresholds as a source of possible uncertainty.

Variance partitioning also allowed us to identify strong interactions between modelling methods and thresholds, thus highlighting the difficulty of teasing apart the different sources of algorithmic uncertainties (Fig. 3). Such interactions were also reflected in the clustering of range change estimates that were separated by both sources of uncertainty (Fig. 2) and from examining spatial distribution of uncertainty in turnover results (Fig. 4). In most parts of the study area, models produced equivalent projections, so the choice of the thresholds contributed the most uncertainty. At the climatic extremes of the study area, models were more prone to extrapolate beyond the range of climatic values used to calibrate the models, thus contributing to greater variance. This is because models have difficulties in projecting distributions to novel climatic conditions (Thuiller et al., 2004; Fitzpatrick and Hargrove, 2009) and these typically emerge at the climatic edges of the study areas, which, in this particular case, also coincide with the geographical edges (see also Barbet-Massin et al., 2010; Araújo et al., 2011). The size of the study area and the extent of the climatic ranges can also influence model and threshold performance. A smaller area, with reduced climatic gradients, is more likely to have a great proportion of edge climates. This being true, models should also be more likely to extrapolate beyond the climate values used in the calibration thus contributing with more variability. In large study areas including extended climate gradients there is a higher likelihood that models are well trained, thus threshold choice is bound to contribute relatively more uncertainty.

Our study demonstrates that the choice of thresholds has practical consequences for estimating species threat based on projected range shifts under climate change (Table 3). Some thresholds produce less variation in threat estimates, and are less sensitive to modelling method. However, the question of what is the most suitable threshold or, specifically, whether the most stable thresholds are also the better, remains unanswered. Answering such questions would require repeated studies with empirical independent evaluation data, such that is provided by census counts in different periods of time, or using artificial species with a known distribution and response to climate (e.g., Hirzel et al., 2001; Austin et al., 2006).

Several strategies have been proposed to adapt thresholds for different modelling purposes where there are economic and ecological considerations (Fielding and Bell, 1997). For example, studies that use BEMs to assess if a species is threatened should not over-inflate the population range, and could therefore require a high specificity (Freeman and Moisen, 2008). Conversely, when using outputs from BEM to create a sampling design it is more important to include all possible presences, and thus have a very low specificity (Freeman and Moisen, 2008). For studying invasive species, specificity and sensitivity also have to be valued differently. If a species is able to establish in an area where it was not predicted to occur (a false negative), eradication effort would cost more than conducting unnecessary surveys in areas where the species was in fact never predicted to become present (a false positive, Hartley et al., 2006). For climate change modelling it has been suggested that changes in projected probability values should be measured only when they overlay with known occurrences of species (Araújo et al., 2002, 2011). When this approach is used, there is no need to set thresholds to convert probability values into presence and absence as they are 'filtered' by observed records of occurrence instead. However, this strategy is more appropriate when analysing data in well sampled regions with good presence and absence data. When models use a small or limited number of species occurrence

records, restricting analysis of range changes to known occurrence records might reduce their usefulness.

BEMs are tools for forecasting the effects of climate change on species potential distributions and, as such, they may be used for decision-making. However, their usefulness is contingent on a good understanding of the sources of error and uncertainty (Barry and Elith, 2006; Heikkinen et al., 2006; Pereira et al., 2010). Here, we show that threshold selection is a key step in BEM, with important consequences for the resulting range projections. Unless sound justification exists for choosing a particular threshold over the others, there might be advantages in applying various threshold-setting methods within an ensemble forecasting framework and explore the resulting range of uncertainties (Araújo and New, 2007).

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2011.07.011.

References

Akçakaya, H.R., Butchart, S.H.M., Mace, G.M., Stuart, S.N., Hilton-Taylor, C., 2006. Use and misuse of the IUCN Red List Criteria in projecting climate change impacts on biodiversity. Glob. Change Biol. 12, 2037–2043.

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43, 1223–1232.

Araújo, M.B., Luoto, M., 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography 16, 743–753.

Araújo, M.B., Williams, P.H., Fuller, R., 2002. Dynamics of extinction and the selection of nature reserves. Proc. Roy. Soc. Lond. Biol. 269, 1971–1980.

Araújo, M., Williams, P., 2000. Selecting areas for species persistence using occurrence data. Biol. Conserv. 96, 331–345.

Araújo, M., Pearson, R., Thuiller, W., Erhard, M., 2005a. Validation of species-climate impact models under climate change. Glob. Change Biol. 11, 1504–1513.

Araújo, M., Whittaker, R., Ladle, R., Erhard, M., 2005b. Reducing uncertainty in projections of extinction risk from climate change. Glob. Ecol. Biogeogr. 14, 529–538.

Araújo, M.B., Rahbek, C., 2006. How does climate change affect biodiversity? Science 313, 1396–1397.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47.

Araújo, M.B., Alagador, D., Cabeza, M., Nogués-Bravo, D., Thuiller, W., 2011. Climate change threatens European conservation areas. Ecol. Lett. 14, 484–492.

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D., Luoto, M., 2006. Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. Ecol. Model. 199, 197–216.

Barbet-Massin, M., Thuiller, W., Jiguet, F., 2010. How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? Ecography 33, 878–886.

Barry, S., Elith, J., 2006. Error and uncertainty in habitat models. J. Appl. Ecol. 43, 413–423.

Baselga, A., Araújo, M.B., 2009. Individualistic vs community modelling of species distributions under climate change. Ecography 32, 55–65.

Berry, P.M., Dawson, T.P., Harrison, P.A., Pearson, R.G., 2002. Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. Glob. Ecol. Biogeogr. 11, 453–462.

Brook, B., Akcakaya, H., Keith, D., Mace, G., Pearson, R., Araújo, M., 2009. Integrating bioclimate with population models to improve forecasts of species extinctions under climate change. Biol. Lett. 5, 723–725.

Buckley, L.B., Urban, M.C., Angilletta, M.J., Crozier, L.G., Rissler, L.J., Sears, M.W., 2010. Can mechanism inform species' distribution models? Ecol. Lett. 13, 1041–1054.

Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. Glob. Change Biol. 16, 1145–1157.

Cantor, S.B., Sun, C.C., Tortolero-Luna, G., Richards-Kortum, R., Follen, M., 1999. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. J. Clin. Epidemiol. 52, 885–892.

Clarke, K.R., Green, R.H., 1988. Statistical design and analysis for a 'biological effects' study. Mar. Ecol. Prog. Ser. 46, 213–226.

Cramer, J.S., 2003. Logit Models: From Economics and Other Fields. Cambridge University Press.

Diniz-Filho, J., Bini, L., Rangel, T., Loyola, R., Hof, C., Nogues-Bravo, D., Araújo, M., 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. Ecography 32, 897–906.

Diniz-Filho, J.A.F., Nabout, J.C., Bini, L.M., Loyola, R.D., Rangel, T.F.L.B., Nogués-Bravo, D., Araújo, M.B., 2010. Ensemble forecasting and geographic range of *Tropidacris cristata* (Orthoptera: Acridoidea: Romaleidae) under climate change. Insect. Conserv. Diver. 3, 213–221.

Dormann, C.F., Purschke, O., García Márquez, J.R., Lautenbach, S., Schröder, B., 2008. Components of uncertainty in species distribution analysis: a case study of the great grey shrike. Ecology 89, 3371–3386.

Drake, J.M., Randin, C., Guisan, A., 2006. Modelling ecological niches with support vector machines. J Appl Ecology 43, 424–432.

Elith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Peterson, A., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberon, J., Williams, S., Wisz, M., Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Fielding, A., Haworth, P.F., 1995. Testing the generality of bird-habitat models. Conserv. Biol. 9, 1466–1481.

Fielding, A., Bell, J., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.

Fitzpatrick, M., Hargrove, W., 2009. The projection of species distribution models and the problem of non-analog climate. Biodivers. Conserv. 18, 2255–2261.

Franklin, J., 2010. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, UK.

Freeman, E.A., Moisen, G.G., 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol. Model. 217, 48–58.

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

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.

Hanspach, J., Kühn, I., Pompe, S., Klotz, S., 2010. Predictive performance of plant species distribution models depends on species traits. Perspect. Plant Ecol. Evol. Syst. 12, 219–225.

Hartley, S., Harris, R., Lester, P.J., 2006. Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. Ecol. Lett. 9, 1068–1079.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T., 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Prog. Phys. Geog. 30, 751–777.

Hijmans, R.J., Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Glob. Change Biol. 12, 2272–2281.

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

Huntley, B., Berry, P.M., Cramer, W., McDonald, A.P., 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. J. Biogeogr. 22, 967–1001.

IUCN, 2001. IUCN Red List Categories and Criteria: Version 3.1. IUCN Species Survival Commission. IUCN, Gland, Switzerland and Cambridge, UK.

Jalas, J., Suominen, J., 1972. Atlas Flora Europaeae: distribution of vascular plants in Europe, 1. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.

Jalas, J., Suominen, J., 1973. Atlas Flora Europaeae: distribution of vascular plants in Europe, 2. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.

Jiménez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to either – or presence – absence. Acta Oecol. 31, 361–369.

Keenan, T., Serra, J.M., Lloret, F., Ninyerola, M., Sabate, S., 2011. Predicting the future of forests in the Mediterranean under climate change, with niche- and process-based models: CO_2 matters! Glob. Change Biol. 17, 565–579.

Keith, D.A., Akçakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J., Regan, H.M., Araújo, M.B., Rebelo, T.G., 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. Biol. Lett. 4, 560–563.

Lahti, T., Lampinen, R., 1999. From dot maps to bitmaps: Atlas Florae Europaeae goes digital. Acta Bot. Fenn. 162, 5–9.

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

Lobo, J., Jimenez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151.

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

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38, 921–931.

Mbogga, M.S., Wang, X., Hamann, A., 2010. Bioclimate envelope model predictions for natural resource management: dealing with uncertainty. J. Appl. Ecol. 47, 731–740.

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M., New, M., 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1990–2000) and 16 scenarios (2001–2100). Tyndall Centre Working Paper. Tyndal Centre for Climate Change Research, Norwich.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., et al., 2000. Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, US.

New, M., Hulme, M., Jones, P.D., 2000. Representing twentieth century space–time climate variability. Part 2. Development of 1901–96 monthly grids of terrestrial surface climate. J. Climate 13, 2217–2238.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H., Wagner, H., 2010. vegan: Community Ecology Package. R package version 1.17–2. http://CRAN.R-project.org/package=vegan.

Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 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.

Pearson, R., Thuiller, W., Araújo, M., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T., Lees, D., 2006. Model-based uncertainty in species range prediction. J. Biogeogr. 33, 1704–1711.

Pearson, R.G., 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available from: http://ncep.amnh.org.

Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sanchez-Cordero, V., Soberon, J., Buddemeier, R.H., Stockwell, D.R.B., 2002. Future projections for Mexican faunas under global climate change scenarios. Nature 416, 626–629.

Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarrés, J.F., Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L., Cooper, H.D., Gilman, E.L., Guénette, S., Hurtt, G., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R., Walpole, M., 2010. Scenarios for global biodiversity in the 21st century. Science 330, 1496–1501.

R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: http://www.R-project.org.

Schapire, R.E., Singer, Y., Singhal, A., 1998. Boosting and Rocchio applied to text filtering. Proc. ACM SIGIR, 215–223.

Schröter, D., Cramer, W., Leemans, R., Prentice, I., Araújo, M., Arnell, N., Bondeau, A., Bugmann, H., Carter, T., Gracia, C., de la Vega-Leinert, A., Erhard, M., Ewert, F., Glendining, M., House, J., Kankaanpaa, S., Klein, R., Lavorel, S., Lindner, M., Metzger, M., Meyer, J., Mitchell, T., Reginster, I., Rounsevell, M., Sabate, S., Sitch, S., Smith, B., Smith, J., Smith, P., Sykes, M., Thonicke, K., Thuiller, W., Tuck, G., Zaehle, S., Zierl, B., 2005. Ecosystem service supply and vulnerability to global change in Europe. Science 310, 1333–1337.

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

Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148, 1–13.

Svenning, J., Normand, S., Kageyama, M., 2008. Glacial refugia of temperate trees in Europe: insights from species distribution modelling. J. Ecol. 96, 1117–1127.

Sykes, M.T., Prentice, I.C., Cramer, W., 1996. A bioclimatic model for the potential distributions of north European tree species under present and future climates. J. Biogeogr. 23, 203–233.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L.J., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O., Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145–148.

Thuiller, W., 2004. Patterns and uncertainties of species' range shifts under climate change. Glob. Change Biol. 10, 2020–2027.

Thuiller, W., Brotons, L., Araújo, M.B., 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography 27, 165–172.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, I.C., 2005. Climate change threats to plant diversity in Europe. Proc. Natl. Acad. Sci. U.S.A. 102, 8245–8250.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32, 369–373.

Williams, P.H., Humphries, C., Araújo, M.B., Lampinen, R., Hagemeijer, W., Gasc, J.-P., Mitchell-Jones, T., 2000. Endemism and important areas for conserving European biodiversity: a preliminary exploration of Atlas data for plants and terrestrial vertebrates. Belgian J. Entomol. 2, 21–46.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. and NCEAS Predicting Species Distributions Working Group, 2008. Effects of sample size on the performance of species distribution models. Divers. Distrib. 14, 763–773.

Woodward, F.I., 1987. Climate and Plant Distribution. Cambridge University Press, Cambridge.