

The impact of modelling choices in the predictive performance of richness maps derived from species-distribution models: guidelines to build better diversity models

Blas M. Benito^{1*}, Luis Cayuela² and Fabio S. Albuquerque¹

¹Department of Ecology, University of Granada, Centro Andaluz de Medio Ambiente, Av. del Mediterráneo, s/n, 18006 Granada, Spain; and ²Departamento de Biología y Geología, Universidad Rey Juan Carlos, Calle Tulipán s/n, E-28933 Móstoles, Madrid, Spain

Summary

1. The stacking of species-distribution models (S-SDMs) is receiving attention by conservation researchers because this approach is capable of simultaneously predicting species richness and composition. However, the steps required to build S-SDMs implies at least two choices that influence its predictive performance which have not been extensively assessed: the selection of the modelling algorithm and the application of a threshold to transform the species-distribution models into binary maps to be added together to build the final S-SDM. Our goal was to provide guidelines concerning the best combinations of modelling algorithms and thresholds with which to build more accurate S-SDMs.

2. We generated 380 S-SDMs of 1224 tree species in Mesoamerica by combining 19 distribution modelling methods with 20 different thresholds using presence-only data from the Global Biodiversity Information Facility. We compared the predicted richness and composition with inventory data obtained from the BIOTREE-NET forest plot database. We designed two indicators of predictive performance that were based on the diversity factors used to measure species turnover: *a* (shared species between the observed and predicted compositions), *b* and *c* (the exclusive species of the predicted and observed compositions respectively) and compared them with the Sorensen and Beta-Simpson turnover measures.

3. Our proposed indexes and the Sorensen index proved suitable as indicators of predictive performance for S-SDMs, whereas the Beta-Simpson turnover measure presented issues that would prevent its application to evaluate S-SDMs.

4. Some modelling methods – especially machine learning and ensemble model forecasting methods performed significantly better than others in minimizing the error in predicted richness and composition. Our results also points out that restrictive thresholds (with high omission errors) lead to more accurate S-SDMs in terms of species richness and composition. Here, we demonstrate that particular combinations of modelling methods and thresholds provide results with higher predictive performance.

5. These results provide clear modelling guidelines that will help S-SDM modellers to select the appropriate combination of modelling methods and thresholds to build more accurate S-SDMs, and therefore will have a positive impact on the quality of the diversity models used to assist conservation planning.

Key-words: BIOTREE-NET, species composition, species richness, stacked species-distribution models

Introduction

Data on biodiversity distribution to be used in decision making, conservation and the management of natural resources have been, and are being extensively gathered around the world. There are collaborative initiatives to increase the public availability of biodiversity data via World Wide Web, such as GBIF (Global Biodiversity Information Facility; URL: www.gbif.org;

over than 330 million georeferenced records), the TROPICOS database (Missouri Botanical Garden; URL: www.tropicos.com; over than 4 million records) and the Atlas of Living Australia (URL: <http://www.ala.org.au/>; more than 20 million georeferenced records). Some issues, however, are limiting its application in conservation planning, including: coarse spatial resolution, geographical gaps, insufficient budget to execute field work and the lack of qualified taxonomists (Yesson *et al.* 2007; Cayuela *et al.* 2009).

The lack of reliable biodiversity data for conservation planning has led to the development of modelling methods based

*Corresponding author. E-mail: blasbp@ugr.es

on ecological theory and GIS technologies. Species-distribution models (*sensu* Guisan & Zimmermann 2000; SDMs hereafter) emerged in this context as tools to generate continuous maps of habitat suitability or presence probability from continuous or discrete environmental variables and incomplete presence data. After years of ongoing development and testing, SDMs are now offering a valuable contribution to the design of management plans for endangered species worldwide, climate-change assessment, invasive-species management or the elucidation of paleo-distributions (for some examples of such applications, see Williams *et al.* 2009; Skov & Svenning 2004; Richardson *et al.* 2010; Lorenzen *et al.* 2011).

SDMs have also been used to predict species richness and composition based on the assemblage of single predicted species ranges, following the 'predict first, assemble later' principle proposed by Ferrier & Guisan (2006). This approach has been recently referred to as 'stacked species-distribution models' (S-SDM hereafter) by Guisan & Rahbeck (2011) because it involves the aggregation (stacking) of SDMs from different species inhabiting the same geographical region. The rationale behind S-SDMs relies on the Gleason's individualistic concept of continuum (Gleason 1926), in which the abundance optima of a given set of co-occurring species are independently distributed along different environmental gradients. At least four steps are required to build an S-SDM for a given pool of target species: (1) selection of an SDM method (Elith *et al.* 2006); (2) SDM calibration for each target species; (3) conversion of each single SDM from a continuous suitability map to a presence/absence binary map using a threshold criterion (Liu *et al.* 2005; Jiménez-Valverde & Lobo 2007; Freeman & Moisen 2008) and (4) summation of the binary maps to build a richness model, and formulation of a presence/absence matrix to represent species composition.

S-SDMs do not account for biotic interactions or any constraints in the maximum number of species potentially occurring in the same geographical unit. Consequently, a significant level of overprediction error in species richness should be expected (Gioia & Pigott 2000; Algar *et al.* 2009; Newbold *et al.* 2009; Dubuis *et al.* 2011; Mateo *et al.* 2012; but see Lehmann, Leathwick & Overton 2002). However, there is also evidence that the choice of one or another SDM method and the selection of the threshold to transform SDMs into binary maps significantly affect the overprediction error of S-SDMs. Regarding the selection of SDM methods, Mateo *et al.* (2012) generated two different kinds of S-SDMs, one derived from the ensemble of six SDM methods and the other derived from an ensemble of a subset of six such SDM methods based on their higher AUC values, and found that the results of the ensemble of all SDM methods were more reliable than the ensemble of the best SDM methods. Regarding threshold selection, Pineda & Lobo (2009) have found that the representation of species richness provided by an S-SDM varied along different thresholds, and that the fine tuning of thresholds for each single species significantly reduced the overprediction error.

The general goal of the present study is to contribute to the research on the performance of S-SDMs and to provide guide-

lines for selecting the best SDM algorithms and the most suitable threshold criteria to build more accurate S-SDMs, considering two outcomes: species richness and composition. For this, we generated 380 S-SDMs based on the predicted ranges of 1224 tropical tree species in Mesoamerica by combining 19 SDM and ensemble methods with 20 threshold criteria using presence-only data from the Global Biodiversity Information Facility (GBIF), and we compared the outcomes in species richness and composition with ground-truth information provided by the BIOTREE-NET forest plot database (Cayuela *et al.* 2012). Specifically, we aimed to answer the following question: Which combination of SDM algorithms (or ensemble method) and threshold criteria maximize the performance of S-SDMs in terms of species richness and composition?

Materials and methods

STUDY AREA, PRESENCE DATA AND PREDICTIVE VARIABLES

Our study area comprised the Mesoamerican region (southern Mexico and Central America), within the limits 2° 30' N; 7° 10' N; 102° 15' W; 77° 10' W (geographical coordinates, datum WGS84). The target region is known to be one of the most important biodiversity hotspots on earth, with almost 2900 endemic plant species (Conservation International 2011), but is threatened because of habitat fragmentation and forest degradation (Chacon 2005).

We compiled a list of 2793 target tree species from the BIOTREE-NET forest plot database (Cayuela *et al.* 2012). The species names were standardized using The Plant List (<http://www.theplantlist.org/>) as a reference checklist. We searched the GBIF database for such species and all their synonyms and downloaded 742 385 records (Fig. 1). We selected GBIF as presence-data source because it includes the records of the Missouri Botanical Garden's TROPICOS database, that has been applied before to model tree distribution in Mesoamerica (Golicher *et al.* 2012a; Golicher, Cayuela & Newton 2012b). Despite its obvious sampling bias in tropical areas (Feeley & Silman 2011), GBIF

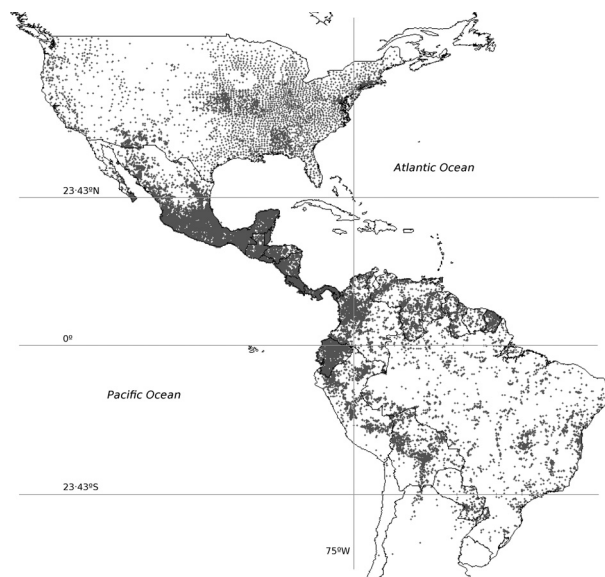


Fig. 1. Modeling area for the target species. The gray dots represents the processed GBIF presence records of the modeled 1224 species.

provides a reasonable snapshot of biodiversity distribution in Mesoamerica, without the commission error that should be expected from the application of species range maps (La Sorte & Hawkins 2007). To prepare a reliable presence dataset, we cleaned the data following these steps: (1) deletion of duplicate records; (2) reduction of spatial aggregation by ensuring a minimum distance of 30 km between consecutive presence records and (3) rejection of species with less than 30 presence records. Finally we obtained data for 1669 species. To build an evaluation dataset, we computed the observed composition and richness for 250 cells of 10 × 10 km resolution where data from the BIOTREE-NET database were available (Fig. 2). To avoid overlap between GBIF and BIOTREE data, the presence records overlapping such 250 cells were not used as input for SDMs.

Following the guidelines provided by Williams *et al.* (2012), we selected a comprehensive set of environmental variables which have been shown to influence the distribution of tropical trees in the study area (Golicher *et al.* 2012a; Golicher, Cayuela & Newton 2012b). In the process of variable selection, we set the maximum correlation between variables at 0.5 Pearson's correlation index. The variables selected were: mean diurnal temperature range, minimum temperature of coldest month, precipitation of wettest month, precipitation of driest month and precipitation seasonality, taken from the Worldclim database (Hijmans *et al.* 2005); human footprint (Sanderson *et al.* 2002); average and standard deviation of the normalized difference vegetation index (NDVI; Tucker, Pinzon & Brown 2004) and topographic diversity, derived from the Shuttle Radar Topography Mission elevation model (USGS 2004) in the GRASS GIS environment (GRASS Development Team 2011).

SDM CALIBRATION AND EVALUATION

The GBIF presence data and the predictor variables were used to calibrate 13 modelling methods and five ensembles (see Araújo & New 2007) computed by the arithmetic average for each species (see Table 1 for further information). To capture the complete response curves of the species to the environmental variables, we modelled the distribution of the target species for the whole American continent. The modelling area was within the limits 72° 00' N; 55° 45' S; 141° W; 34° 45' W and comprised 590 998 cells at 10 km spatial resolution. Once the continental models were generated, we clipped the Mesoamerican region, constituting 12 249 cells at the same resolution as the continental models.

To evaluate the predictive performance of the SDMs we applied the area under the ROC curve method (AUC; Fielding & Bell 1997), which is a suitable method to compare models for the same species and study



Fig. 2. Target study area. The grey dots represents the GBIF presence records. The Boxes represents the BIOTREENET inventory plots used to evaluate the S-SDMs.

area, executed with different algorithms (Lobo, Jiménez-Valverde & Real 2008). The AUC values for each model were computed by k-fold validation (five groups) with the 'evaluate' function of the R package 'dismo' (Hijmans *et al.* 2012), using a set of 5000 random points as pseudo-absences for each species, avoiding overlap with the presence points. To ensure comparable S-SDMs for all the modelling methods in Table 1, we used only the species with all their SDMs and ensembles with AUC values higher than 0.65.

S-SDM CALIBRATION AND EVALUATION

We built 380 S-SDMs grouped in 19 'S-SDM series'. Each S-SDM series was the result of stacking all the SDMs of each of the 18 modelling or ensemble methods presented in Table 1, plus another built by stacking the best SDM for each species, based on AUC values. Hereafter, we will use the respective acronym in Table 1 to refer to each S-SDM series and BES for the one built with the best SDMs. Each S-SDM series was composed of 20 S-SDMs built with successive threshold criteria based on a range of omission percentages ('predetermined sensitivities' according to Fielding & Bell 1997). For example, a 10% omission-percentage threshold implies that the suitable habitat of the resulting binary map contained 90% of the species presence records (Fielding & Bell 1997). We built the 20 S-SDMs of each S-SDM series by applying a range of thresholds, from 0% to 95% omission error, in steps of 5%.

To compare the performance of the S-SDM predictions against the observed data of BIOTREE-NET, we applied the rationale underlying the assessment of species turnover (see Lennon *et al.* 2001 and Koleff, Gaston & Lennon 2003). Such rationale is based on the definition of the diversity components *a*, *b* and *c* for two given locations, being *a* the number of species shared by both locations, *b* the number of exclusive species of the first location and *c* the number of exclusive species of the second location (Lennon *et al.* 2001). We computed *a*, *b* and *c* for the observed 250 BIOTREE inventory locations and the 380 predicted compositions with the *betadiver* function included in the R package *vegan* (Oksanen *et al.* 2011). To analyse the results, we slightly changed the interpretation of the diversity components explained above. Hereafter, *a* represents the number of species common to the observed and predicted composition (i.e. the number of species correctly predicted), whereas *b* and *c*, respectively, represent the number of exclusive species of the simulated (i.e. the number of overpredicted species, or commission error) and observed compositions (i.e. the number of underpredicted species or omission error). Considering such an interpretation, we designed two simple indexes to assess S-SDM performance in predicting richness and composition: *R_{pp}* (richness predictive performance) and *C_{pp}* (composition predictive performance) respectively. The former is presented in Eq. 1 and represents the absolute difference between the predicted and the observed richness. The latter is presented in Eq. 2 and represents the number of incorrectly predicted species for each correctly predicted species. Both indexes present a lower limit of zero and an upper limit at the maximum number of modelled species, with lower values indicating better predictive performance. The results of both indexes can be easily expressed relatively to the observed number of species, that may be useful to compensate differences in species richness between locations and obtain more reliable assessments, or converted to proportions or percentages, to allow the comparison between different studies.

$$R_{pp} = |(a + b) - (a + c)| \quad \text{eqn 1}$$

$$C_{pp} = (b + c)/a \quad \text{eqn 2}$$

Table 1. Modelling algorithms and their ensembles stacked by summation to generate the S-SDMs analysed. The 'Parameters' column represents the configuration of the most important parameters used to generate the SDMs. The R version 2.14.2 was used to calibrate the models depending on R packages

Family	Algorithm	Acronym	Software	Reference	Parameters		
Distance-based	Chebyshev	CHE	Open Modeller	Muñoz <i>et al.</i> (2009)	nearest n points = 3		
	Euclidean	EUC					
	Manhattan	MAN					
Machine learning	Mahalanobis	MAH	R package Dismo	Hijmans <i>et al.</i> (2012)	tree complexity = 2 max trees = 1000 default settings		
	Boosted	BRT					
	Regression Trees						
	Maximum Entropy	MAX				MaxEnt	Phillips, Anderson & Schapire (2006)
	Artificial Neural Networks	ANN				R package nnet	Venables & Ripley (2002)
Regression	Random Forests	RFR	R package random Forest	Breiman (2001); Liaw & Wiener (2002)	5000 pseudo-absences avoiding spatial overlap with presences. family = binomial link = logit no interactions		
	Support Vector Machines	VMO VMR	Open Modeller R package kernlab	Muñoz <i>et al.</i> (2009) Karatzoglou <i>et al.</i> (2004)			
	GLM	LOG	R core	R Development Core Team (2012)			
	GAM	GAM	R package gam	Hastie (2011)			
	MARS	MAR	R package earth	Milborrow (2012)			
Ensembles	All algorithms	EN1	GRASS GIS	GRASS Development Team (2011)			
	Best algorithms (Elith <i>et al.</i> 2006)	EN2					
	Distance-based	EN3					
	Machine learning	EN4					
	Regression-based	EN5					

To evaluate the suitability of Cpp as a measure of predictive performance in composition for S-SDMs, we compared it with other measures of similarity in composition. We examined Koleff, Gaston & Lennon (2003) searching for indexes with different properties and selected two of them: (1) the Sorensen similarity index (β_{sor}), presented by Lennon *et al.* (2001), is a measure of continuity that depends on the variation of the matching component a . To allow the comparison of β_{sor} with the other indexes, we transformed its values into dissimilarity by subtracting the equation in Koleff, Gaston & Lennon (2003) by one (see Eq. 3). According to Eq. 3, β_{sor} has its upper limit at one (strong dissimilarity between compositions) and the lower limit at zero (strong similarity between compositions); (2) the symmetric form of the Simpson's dissimilarity index (β_{sim} , Eq. 4), that was also presented in Lennon *et al.* (2001), derived from the one proposed by Simpson (1943), and reformulated by Koleff, Gaston & Lennon (2003). β_{sim} is a measure of turnover focused on compositional differences instead of differences in species richness. The upper β_{sim} value is found when the observed and predicted compositions have no species in common, and the lower value is zero, when both compositions are the same. Previously, β_{sim} have been used in the sphere of S-SDMs by Pineda & Lobo (2009) to assess the predictive performance of richness models of amphibians in Mexico.

$$\beta_{\text{sor}} = 1 - (2a/(2a + b + c)) \quad \text{eqn 3}$$

$$\beta_{\text{sim}} = \min(b, c)/(\min(b, c) + a) \quad \text{eqn 4}$$

Also, we computed the mean a , b , c , Rpp, Cpp, β_{sor} and β_{sim} between the observed and predicted compositions for each S-SDM to graphically analyse the behaviour of each index along thresholds.

ANALYSIS

We approached the question 'Which combination of SDM algorithms (or ensemble method) and threshold criteria maximizes the performance of S-SDMs in terms of species richness and composition?' from two viewpoints. Firstly, we analysed the performance of each S-SDM series without considering threshold criteria. To do so, we applied the Friedman nonparametric method with a *post hoc* test to search for significant differences in the raw values of Rpp and Cpp between pairs of the 19 groups of S-SDMs, each consisting of 5000 cases: 20 thresholds \times 250 inventory locations (Hollander & Wolfe 1973). Then we ranked each S-SDM series according to its performance for Rpp and Cpp and the statistical differences between them. At this stage of the analysis, we did not consider thresholds as a factor. Secondly, we computed the mean values for Rpp and Cpp for every combination of threshold and S-SDM series and scaled the values between 0 and 0.5 to weight equally predictive performance in richness and composition. We added together the scaled indexes into a single combined performance index in the interval [0, 1] (the lesser, the better) that was plotted using the *heatmap.2* function of the *gplots* R library (Warnes 2012). The resulting plot was intended to be used as a tool to select the methods and thresholds that build the most accurate S-SDMs.

Results

From the initial 1669 species used to build SDMs, only 1224 presented all of their models with an AUC higher than 0.65. Therefore, each of the final 380 S-SDMs was built by summation of 1224 SDMs (see Fig. 3). The BES series, which was the

only one mixing different SDM methods, was composed mainly of EN1 (311 species), RFR (310) and NNR (279), followed by EN5 (68), VMO (57) and GAM (49).

The trends of *a*, *b* and *c* along thresholds were similar in shape between the different S-SDM series (see upper plot at Fig. 4). The diversity component *b* (exclusive species of the simulated composition; commission error) increased exponentially while the threshold decreased, whereas the component *c* (exclusive species of the observed composition; omission error) increased linearly but at a much lower rate along thresholds. The trend of shared species between the observed and simulated composition was symmetric with respect to *c*. Rpp was strongly influenced by the values of *b* for the lower thresholds, but reaching its minimum values when *b* and *c* reached the same value. Cpp and β_{sor} were symmetric, showing a high and significant correlation (mean R^2 across S-SDM series = 0.96; p-value < 0.005), with low similarity values when the thresholds were very low or very high, and optimum values that varied between S-SDM series (see lower plot at Fig. 4). On the other hand, β_{sim} showed a contradictory trend, indicating high similarity between the observed and simulated composition when the thresholds were very low or very high. Considering the strong correlation between Cpp and β_{sor} and the consistent behaviour of Rpp, hereafter, we will report our results referring only to Rpp and Cpp.

The Friedman test confirmed that there were significant differences between S-SDM series for both indexes (p-value < 2.2e-16). When richness and composition were predicted, RFR (see Fig. 5a and 5b) emerged as the most accurate SDM method to build S-SDMs, significantly differing from the other methods (p-value < 2.2e-16). The heatmap (Fig. 6) shows the values of the combined performance index (the scaled sum of

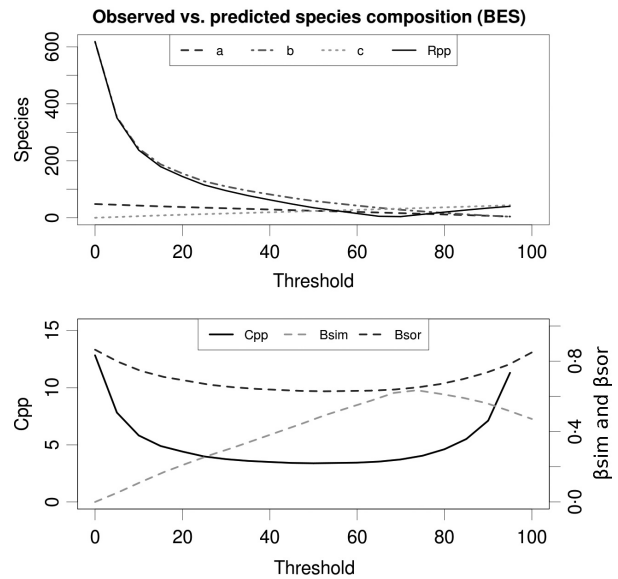


Fig. 4. Trends of the mean *a*, *b*, *c*, Rpp, Cpp, β_{sor} , and β_{sim} along thresholds for the BES S-SDM series. Both plots share the same x axis.

Cpp and Rpp) for each combination of S-SDM series and threshold. In the combined performance index, lower values represent higher predictive performances for species richness and composition. This heatmap shows intuitively the best combinations of modelling method and threshold to build the most accurate S-SDMs.

Discussion

The aim of this study was to assess how the choice of SDM method and threshold selection affects the predictive perfor-

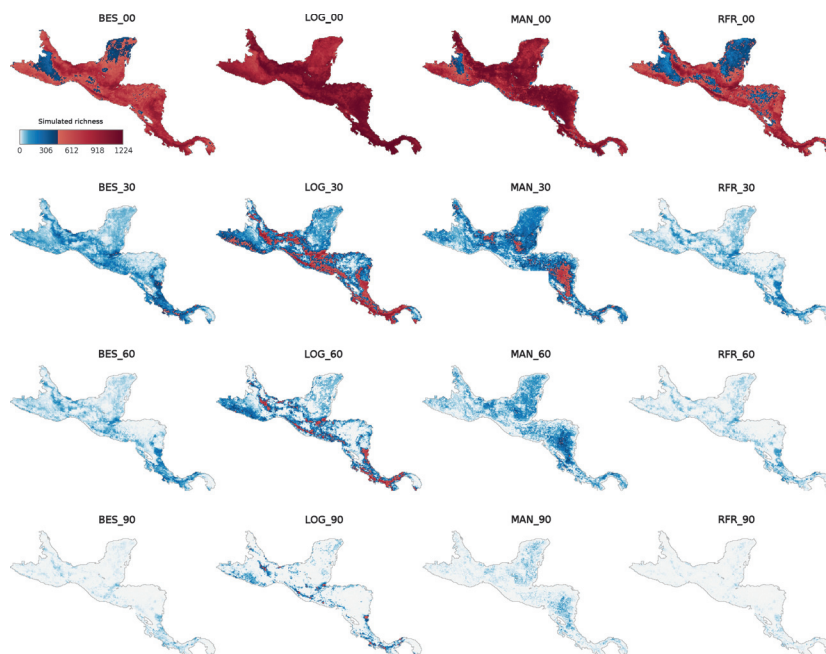


Fig. 3. A fraction of the generated S-SDMs. Columns represent methods (BES, LOG, MAN, RFR; see Table 1) whereas rows represent thresholds (0, 30, 60, 90). The colour scale, which represents species richness, is blue until 440, which was the maximum observed richness in a BIOTREE-NET inventory plot. The red colours representss simulated richness between 441 and 1224 species.

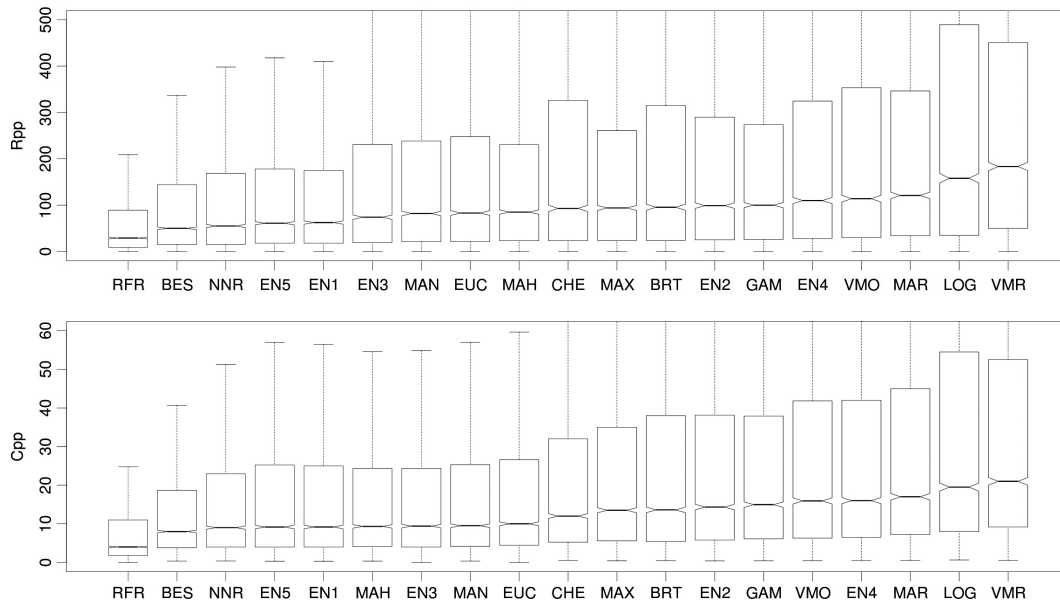


Fig. 5. Ordered boxplots showing the performance of each S-SDM series in terms of species richness (Rpp, plot a) and species composition (Cpp, plot b). Lower values indicate better predictive performances. The y axis was clipped to the higher quartiles to provide a better visualization.

mance of S-SDMs in terms of species richness and composition. Different SDM methods are known to provide results of varying predictive performance (Elith *et al.* 2006), whereas the threshold selection has a major impact on the omission and commission errors in the model outcome (Liu *et al.* 2005). We found that S-SDMs built with different SDM methods present different predictive accuracies, and that threshold selection strongly influences the reliability of the predicted richness and composition.

PERFORMANCE INDEXES

We tested several S-SDM performance measures regarding accuracy in species richness and composition. The richness predictive-performance index (Rpp) has a mathematical formulation similar to the root mean square deviation (RMSD), but adapted to be computed from the diversity components a , b and c . This enabled us to properly identify the SDM methods and thresholds that provided the minimum error in predicted richness. We applied this index instead of the Spearman's rank correlation coefficient applied by Pineda & Lobo (2009) because the latter is a measure of dependence between the observed and simulated richness patterns, but does not provide the overall difference between the observed and predicted species richness, which was one of the target values to minimize in our study.

Regarding the indexes to compare predicted and observed composition, our analysis showed that β_{sim} provided misleading results, especially at lower thresholds, at which its values indicated high similarity between observed and predicted compositions. At these thresholds the commission error was actually high – there were many more species predicted in the simulation than present in the observed composition – and therefore the actual similarity between predicted and observed

compositions must be very low. This behaviour can be explained for β_{sim} considering that its computation relies on the term ' $\min(b, c)$ ', which at each threshold automatically selects the diversity component of lower value, which turns out to be c at lower thresholds. As Koleff, Gaston & Lennon (2003) stated, β_{sim} is very sensitive to subtle variations in b or c when the values of a and either b or c are low. In the light of our findings, and considering the nature of β_{sim} , we cannot consider β_{sim} as a reliable measure of composition similarity to compare the outcomes of S-SDMs generated with different threshold criteria, especially when such thresholds are low. On the other hand, β_{sor} and Cpp were highly correlated and they were readily interchangeable.

There are other alternative possibilities to evaluate the predictive performance of S-SDMs that also considers a fourth biodiversity component named d , which represents the species that are both observed and predicted as absent. The four biodiversity components taken together can be used to fill a confusion matrix and compute different accuracy measures like specificity, sensitivity or kappa (Pottier *et al.* 2012).

BEST SDM METHODS TO BUILD S-SDMS

We found a consistent pattern of predictive accuracy in richness and composition when the median of Rpp and Cpp was analysed for the S-SDM series without considering thresholds. The rankings for both accuracy indexes were very similar, with RFR, BES, NNR, EN5 and EN1 in the first five positions. The RFR S-SDM series was generated with random forests, a machine-learning method able to fit complex nonlinear surfaces from high-dimensional input data (Cutler *et al.* 2007). The use of random forest to calibrate SDMs is increasing in the literature, outperforming other modelling methods such as logistic regression, maximum entropy, artificial neural

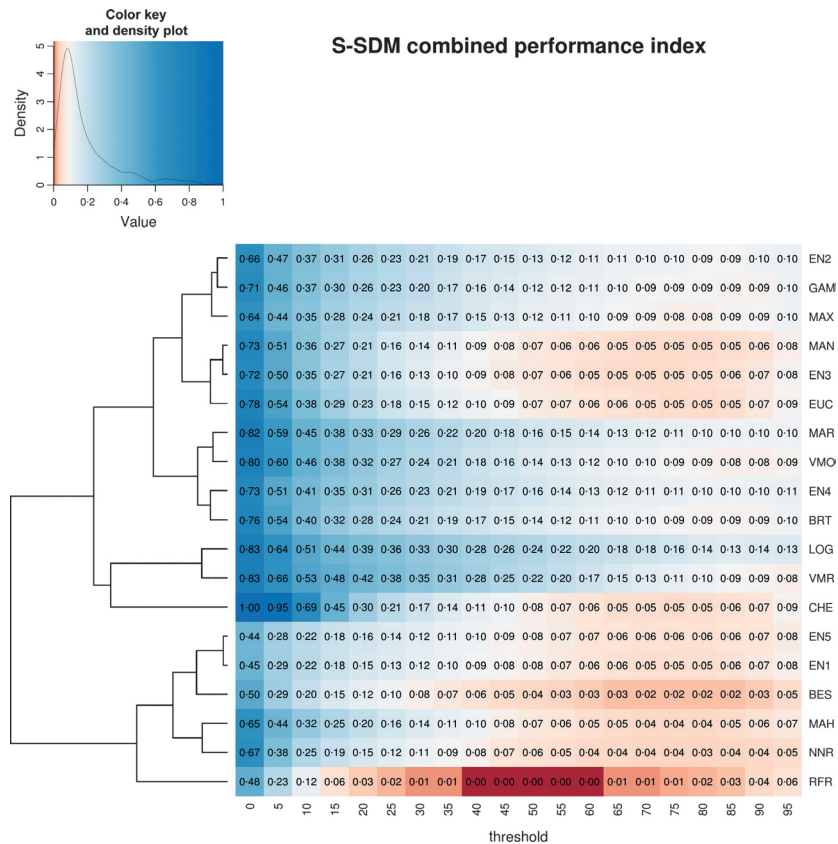


Fig. 6. Heatmap representing the combined index of predictive performance in species richness and composition for each combination of SDM method and threshold. In the colour scale, blue and red (lower and higher index values) respectively represents good and poor predictive performance, whereas the transition in between (white) was situated in the overall mean of all the combinations. The cluster represents performance similarities between SDM methods based on its mean values for the combined index. The upper left panel represents the statistical distribution of the combined index.

networks or support vector machines (Cutler *et al.* 2007; Williams *et al.* 2009; Bisrat *et al.* 2012). Such good results indicate that random forest is a promising algorithm to generate S-SDMs, but taking into account that the algorithm results are prone to overfitting, and show poor performance when transferred across time or space (Heikkinen, Marmion & Luoto 2012). The second position was occupied by the BES S-SDM, which was built with the best SDM method or ensemble for each species according to AUC values, EN1 (an ensemble of all the SDM methods), RFR and NNR being the most frequent SDM methods selected. The good results of BES to predict richness and composition were expected and easily explained: when comparing competing SDMs, AUC was higher in models with less commission error. Therefore, the BES S-SDM series was built with the single models with less commission error – that is, the error of higher magnitude in S-SDMs. However, BES also presents a significant drawback, because it requires substantial computing effort, plus the necessary know-how to run different families of models. Consequently, the selection of other methods with better or similar predictive accuracies, like RFR or NNR, would be preferred by most modellers. The third position was occupied by three methods without significant differences in performance for richness and composition: NNR, EN5 and EN1. The Artificial

Neural Networks (NNR) method has been applied to build SDMs ever since the works of Manel *et al.* (1999a), Manel, Dias & Ormerod (1999b), but its appearance in the specialized literature is scarce when compared with methods like GLM or MaxEnt, probably because Artificial Neural Networks are perceived by biologists as a black-box modelling method (but see Benitez, Castro & Requena 1997; and Gahegan 2003). Also, it has been reported that SDMs built with artificial neural networks present good transferability in space or time (Heikkinen, Marmion & Luoto 2012). In the light of our results, NNR should be considered a good choice to build S-SDMs. Regarding EN5 (the ensemble of regression-based methods) and EN1 (the ensemble of all methods), both S-SDM series performed very well, considering that some of the S-SDMs built with fundamental SDM methods performed poorly when predicting richness and composition. This case is a clear example of the advantages of ensemble model forecasting (Araújo & New 2007).

THRESHOLDS IN S-SDMS

The selection of thresholds to transform a continuous SDM surface into a binary presence/absence map have important effects on SDM outcomes, and has been widely discussed in

the SDM literature (Liu *et al.* 2005; Jiménez-Valverde & Lobo 2007; Freeman & Moisen 2008). However, such literature is oriented mainly to the selection of thresholds when the input data are presence/absence, and therefore the threshold criteria for presence-only based SDMs are not well-developed. A potential approach proposed by Fielding & Bell (1997) is the use of thresholds based on predetermined sensitivities – or predefined percentages of omission error – but these kinds of criteria have been criticized because of their arbitrariness (Liu *et al.* 2005). Considering this body of work, we followed the thresholding approach previously applied by Pineda & Lobo (2009), which we found to be the most suitable to explore the consequences of changing thresholds in the S-SDM outcomes. Despite that we applied predefined omission errors as threshold criteria to build S-SDMs, we did so only for testing purposes. To build the most accurate S-SDMs oriented to conservation, we recommend two alternative approaches: (1) the fine tuning of the threshold for each individual species based on ground-truth information, proposed by Pineda & Lobo (2009); (2) the exploration of a subset of all the potential thresholds available to build the S-SDM to find the one that maximizes the predictive performance of the model in terms of species richness and composition.

GUIDELINES TO BUILD S-SDMS

Our extensive predictive performance analysis, for richness and composition, of combinations of S-SDM methods and thresholds provided a noteworthy result (Fig. 6). The heatmap of the combined index offered clear guidelines concerning the best choices, in terms of S-SDM methods and thresholds, for building more accurate S-SDMs. Particularly, we found that, for at least half of the S-SDM series, the thresholds providing the most accurate results were those ensuring omission percentages between 50 and 85. The exception was RFR, which performed better in the range of 40–60, but offered good results beyond that range, also. On the other hand, the thresholds based on very low omission errors, of between 0 and 20, which has been rarely applied in the literature of S-SDMs (but see Algar *et al.* 2009 and Mateo *et al.* 2012) performed very poorly for all S-SDM series, and thus cannot be recommended for application in future studies.

Acknowledgements

The authors are indebted to David Nesbitt for the English revision and to the Associate Editor and two anonymous referees, whose comments improved the quality of the article. Funding for BMB was from the 'Consejería de Economía, Innovación y Ciencia, Junta de Andalucía' project RNM-6734 (MIGRAME). LC and FSA were supported by project BIOTREE-NET (BIOCON08_044), funded by Fundación BBVA. We thank the data providers of the BIOTREE-NET database, and the Global Biodiversity Information Facility and all contributing herbaria for making their data publicly available.

References

Algar, A.C., Kharouba, H.M., Young, E.R. & Kerr, J.T. (2009) Predicting the future of species diversity, macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, **32**, 22–33.

- Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.
- Benitez, J.M., Castro, J.L. & Requena, I. (1997) Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks*, **8**, 1156–1164.
- Bisrat, S.A., White, M.A., Beard, K.H. & Richard Cutler, D. (2012) Predicting the distribution potential of an invasive frog using remotely sensed data in Hawaii. *Diversity and Distributions*, **18**, 648–660.
- Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.
- Cayuela, L., Golicher, D.J., Newton, A.C., Kolb, M., de Albuquerque, F.S., Arets, E.J.M.M., Alkemade, J.R.M. & Pérez, A.M. (2009) Species distribution modeling in the tropics, problems, potentialities, and the role of biological data for effective species conservation. *Tropical Conservation Science*, **2**, 319–352.
- Cayuela, L., Gálvez-Bravo, L., Pérez Pérez, R., Albuquerque, F.S., Golicher, D.J., Zahawi, R.A., Ramírez-Marcial, N., Garibaldi, C., Field, R., Rey Benayas, J.M., González-Espinosa, M., Balvanera, P., Castillo, M.A., Figueroa-Rangel, B.L., Griffith, D.M., Islebe, G.A., Kelly, D.A., Olvera-Vargas, M., Schnitzer, S.A., Velázquez, E., Williams-Linera, G., Brewer, S.W., Camacho-Cruz, A., Coronado, I., de Jong, B., del Castillo, R., de la Cerda, I., Fernández, J., Fonseca, W.G., Galindo-James, L., Gillespie, T.W., González-Rivas, B., Gordon, J.E., Hurtado, J., Linares, J., Letcher, S.G., Mangan, S.C., Meave, J.A., Méndez, E.V., Meza, V., Ochoa-Gaona, S., Peterson, C.J., Ruiz-Gutierrez, V., Snarr, K.A., Tun Dzul, F., Viergever, K.M., White, D.A., Williams, J.N., Bonet, F.J. & Zamora, R. (2012) The tree biodiversity network (BIOTREE-NET), prospects for biodiversity research and conservation in the tropics. *Biodiversity and Ecology*, **4**, 211–224.
- Chacon, C. (2005) Fostering conservation of key priority sites and rural development in Central America: the role of private protected areas. *Parks*, **15**, 39–47.
- Conservation International. 2011. Biodiversity Hotspots - Mesoamerica. URL: <http://www.biodiversityhotspots.org/xp/hotspots/mesoamerica/pages/default.aspx#indepth>.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J. (2007) Random forests for classification in ecology. *Ecology*, **88**, 2783–2792.
- Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.P. & Guisan, A. (2011) Predicting spatial patterns of plant species richness. A comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions*, **17**, 1122–1131.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettman, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, M.C.J., Peterson, A.T., Phillips, S.J., Richardson, K., Pereira, R.S., Schapire, R.E., Soberón, J., Williams, S.E., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Feeley, K.J. & Silman, M.R. (2011) The data void in modeling current and future distributions of tropical species. *Global Change Biology*, **17**, 626–630.
- Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence. *Environmental Conservation*, **24**, 38–49.
- 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. *Ecological Modelling*, **217**, 48–58.
- Gahegan, M. (2003) Is inductive machine learning just another wild goose (or might it lay the golden egg)? *International Journal of Geographical Information Science*, **17**, 69–92.
- Gioia, P. & Pigott, P.J. (2000) Biodiversity assessment: a case study in predicting richness from the potential distributions of plant species in the forests of southwestern Australia. *Journal of Biogeography*, **27**, 1065–1078.
- Gleason, H.A. (1926) The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, **53**, 7–26.
- Golicher, D.J., Cayuela, L. & Newton, A.C. (2012b) Effects of Climate Change on the Potential Species Richness of Mesoamerican Forests. *Biotropica*, **44**, 284–293.
- Golicher, D.J., Cayuela, L., Ford, A. & Newton, A.C. (2012a) Pseudo-absences, pseudo-models and pseudo-niches, pitfalls of model selection based on the Area Under the Curve. *International Journal of Geographical Information Science*, **26**, 2049–2063.
- GRASS Development Team (2011) Geographic Resources Analysis Support System (GRASS) Software, Version 6.4.1. Open Source Geospatial Foundation. <http://grass.osgeo.org>.
- Guisan, A. & Rahbeck, C. (2011) SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hastie, T. (2011). *gam: Generalized Additive Models*. R package version 1.06.2, <http://CRAN.R-project.org/package=gam>.
- Heikkinen, R.K., Marmion, M. & Luoto, M. (2012), Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, **35**, 276–288.
- 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.
- Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2012) *dismo: Species distribution modeling*. R package version 0.6-10, <http://CRAN.R-project.org/package=dismo>.
- Hollander, M. & Wolfe, D.A. (1973) *Nonparametric Statistical Methods*. John Wiley & Sons, New York.
- Jiménez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, **31**, 361–369.
- Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. (2004) kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, **11**, 1–20.
- Koleff, P., Gaston, K.J. & Lennon, J.J. (2003) Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, **72**, 367–382.
- La Sorte, F.A. & Hawkins, B.A. (2007) Range maps and species richness patterns: errors of commission and estimates of uncertainty. *Ecography*, **30**, 649–662.
- Lehmann, A., Leathwick, J.R. & Overton, J.M. (2002) Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation*, **11**, 2217–2238.
- Lennon, J.J., Koleff, P., Greenwood, J.J.D. & Gaston, K.J. (2001) The geographical structure of British bird distributions, Diversity, spatial turnover and scale. *Journal of Animal Ecology*, **70**, 966–979.
- Liaw, A. & Wiener, M. (2002) Classification and Regression by randomForest. *R News*, **2**, 18–22.
- 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., Jiménez-Valverde, A. & Real, R. (2008) AUC, A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.
- Lorenzen, E.D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K.A., Ugan, A., Borregaard, M.K., Gilbert, M.T.P., Nielsen, R., Ho, S.Y.W., Goebel, T., Graf, K.E., Byers, D., Stenderup, J.T., Rasmussen, M., Campos, P.F., Leonard, J.A., Koepfli, K.-P., Froese, D., Zazula, G., Stafford Jr, T.W., Aaris-Sørensen, K., Batra, P., Haywood, A.M., Singarayer, J.S., Valdes, P.J., Boeskorov, G., Burns, J.A., Davydov, S.P., Haile, J., Jenkins, D.L., Kosintsev, P., Kuznetsova, T., Lai, X., Martin, L.D., McDonald, H.G., Mol, D., Meldgaard, M., Munch, K., Stephan, E., Sablin, M., Sommer, R.S., Sipko, T., Scott, E., Suchard, M.A., Tikhonov, A., Willerslev, R., Wayne, R.K., Cooper, A., Hofreiter, M., Sher, A., Shapiro, B., Rahbek, C. & Willerslev, E. (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, **479**, 359–364.
- Manel, S., Dias, J.M. & 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., Dias, J.M., Buckton, S.T. & Ormerod, S.J. (1999a) Alternative methods for predicting species distribution, An illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734–747.
- Mateo, R.G., Felicísimo, A.M., Pottier, J., Guisan, A. & Muñoz, J. (2012) Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS ONE*, **7**, e32586.
- Milborrow, S. (2012) *Earth, Multivariate Adaptive Regression Spline Models*. R package version 3.2-2, <http://CRAN.R-project.org/package=earth>.
- Muñoz, M.E.S., Giovanni, R., Siqueira, M.F., Sutton, T., Brewer, P., Pereira, R.S., Canhos, D.A.L. & Canhos, V.P. (2009) openModeller: a generic approach to species' potential distribution modelling. *Geoinformatica*, **15**, 111–135.
- Newbold, T., Gilbert, F., Zalata, S., El-Gabbas, A. & Reader, T. (2009) Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, **36**, 2085–2095.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2011) *Vegan, Community Ecology Package*. R package version 1.17-10, <http://CRAN.R-project.org/package=vegan>.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Pineda, E. & Lobo, J.M. (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182–190.
- Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P. & Guisan, A. (2012) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography*, **22**, 52–63.
- R Development Core Team (2012). *R, A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Richardson, D.M., Iponga, D.M., Roura-Pascual, N., Krug, R.M., Milton, S.J., Hughes, G.O. & Thuiller, W. (2010) Accommodating scenarios of climate change and management in modelling the distribution of the invasive tree *Schinus molle* in South Africa. *Ecography*, **33**, 1049–1061.
- Sanderson, E.W., Jaiteh, M., Levy, M.A., Redford, K.H., Wannebo, A.V. & Woolmer, G. (2002) The human footprint and the last of the wild. *BioScience*, **52**, 891–904.
- Simpson, G.G. (1943) Mammals and the nature of continents. *American Journal of Science*, **241**, 1–31.
- Skov, F. & Svenning, J.C. (2004) Potential impact of climatic change on the distribution of forest herbs in Europe. *Ecography*, **27**, 366–380.
- Tucker, C.J., Pinzon, J.E. & Brown, M.E. (2004) *Global Inventory Modeling and Mapping Studies, Global Land Cover Facility*, University of Maryland, College Park, Maryland.
- USGS (2004) *Shuttle Radar Topography Mission, Global Land Cover Facility*, University of Maryland, College Park, Maryland.
- Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S*. 4th edn. Springer, New York.
- Warnes, G.R. (2012). *gplots: Various R programming tools for plotting data*. R package version 2.11.0, <http://CRAN.R-project.org/package=gplots>.
- Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M. & Schwartz, M.W. (2009) Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, **15**, 565–576.
- Williams, K.J., Belbin, L., Austin, M.P., Stein, J.L. & Ferrier, S. (2012) Which environmental variables should I use in my biodiversity model? *International Journal of Geographical Information Science*, **26**, 2009–2047.
- Yesson, C., Brewer, P.W., Sutton, T., Caithness, N., Pahwa, J.S., Burgess, M., Gray, W.A., White, R.J., Jones, A.C., Bisby, F.A. & Culham, A. (2007) How Global Is the Global Biodiversity Information Facility? *PLoS ONE*, **2**, e1124.

Received 10 October 2012; accepted 15 November 2012

Handling Editor: Robert B. O'Hara

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Richness models: the stacked species distribution models (S-SDMs) of 1224 tree species analysed in this paper.

Appendix S2. Complete list of species used in this paper.