

Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change

José Alexandre F. Diniz-Filho, Luis Mauricio Bini, Thiago Fernando Rangel, Rafael D. Loyola, Christian Hof, David Nogués-Bravo and Miguel B. Araújo

J. A. F. Diniz-Filho (diniz@icb.ufg.br), L. M. Bini and R. D. Loyola, Depto de Ecologia, ICB, Univ. Federal de Goiás (UFG), Cx.P. 131, 74001-970 Goiânia, GO, Brasil. – T. F. Rangel, Dept of Ecology and Evolution, Univ. of Connecticut, Storrs, CT 06269-3043, USA. – C. Hof and D. Nogués-Bravo, Depto de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, ES-28006 Madrid, Spain and Center for Macroecology, Evolution and Climate, Dept of Biology, Univ. of Copenhagen, DK-2100 Copenhagen, Denmark. – M. B. Araújo, Depto de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, ES-28006 Madrid, Spain and Cátedra Rui Nabeiro – Biodiversidade, CIBIO, Univ. de Évora, Largo dos Colegiais, PT-7000 Évora, Portugal.

Forecasts of species range shifts under climate change are fraught with uncertainties and ensemble forecasting may provide a framework to deal with such uncertainties. Here, a novel approach to partition the variance among modeled attributes, such as richness or turnover, and map sources of uncertainty in ensembles of forecasts is presented. We model the distributions of 3837 New World birds and project them into 2080. We then quantify and map the relative contribution of different sources of uncertainty from alternative methods for niche modeling, general circulation models (AOGCM), and emission scenarios. The greatest source of uncertainty in forecasts of species range shifts arises from using alternative methods for niche modeling, followed by AOGCM, and their interaction. Our results concur with previous studies that discovered that projections from alternative models can be extremely varied, but we provide a new analytical framework to examine uncertainties in models by quantifying their importance and mapping their patterns.

Environmental niche models, or species distribution models (SDM), are frequently used to forecast shifts in species geographic distributions under climate change (Peterson et al. 2002, Thomas et al. 2004a, Thuiller et al. 2005a, b, Araújo et al. 2006, Lawler et al. 2009). When species ranges closely match their potential niche, associations between species ranges and environmental factors can be reliably used to estimate the ecological requirements of species (Araújo and Guisan 2006, Soberón 2007). Estimated associations can then be utilized to forecast species range shifts and related changes in biodiversity patterns under climate change scenarios.

It is widely acknowledged that SDMs provide a simplified representation of the processes governing the geographic distributions of species (Pearson and Dawson 2003, Guisan and Thuiller 2005). Actually, multiple ecological and evolutionary processes, operating at different spatial and temporal scales, are expected to determine contemporary distributions of most species (Araújo et al. 2008, Pearman et al. 2008), and several of these processes are poorly represented in the models (Guisan and Thuiller 2005). In addition to ecological uncertainties, there are several sources of methodological uncertainty that have been discussed in a number of recent studies (Thuiller et al. 2004, Araújo et al. 2005a, b, Pearson et al. 2006, 2007,

Marmion et al. 2009). Nevertheless, and despite computational and methodological advances, the decision as to which model to use is often ad hoc (Araújo and New 2007), and there is little agreement regarding the relative performance of alternative niche-based techniques and overall modeling strategies for forecasting species distributional changes under climate change (Araújo and Rahbek 2006, Dormann 2007, Peterson et al. 2007, Phillips 2008).

Methodological uncertainties may arise because of differences in data sources and statistical methods used for niche and climate modeling (Heikkinen et al. 2006). If it is not possible to clearly establish which models are more adequate to a particular problem, a potential solution to take inter-model variability into account is to fit multiple models, or ensembles, and combine them into some sort of consensus forecast (for reviews see Araújo and New 2007, Leutbecher and Palmer 2008). Recent studies suggested that improvements in the forecasts could be achieved if ensembles were obtained and the results were appropriately analyzed (Araújo et al. 2005a, 2006, Marmion et al. 2009, Roura-Pascual et al. 2009). For instance, some widely used methods for niche modeling, such as GARP, Neural Networks, and Random Forests do generate multiple projections and combine them into a single consensus solution (Lawler et al. 2009, O'Hanley 2009, see also

Table 1 in Araújo and New 2007). The new version of the BIOMOD software allows different methods to be fitted and projections to be compared and combined (Thuiller et al. 2009).

However, existing approaches sparsely sample all possible uncertainties from models (Araújo and New 2007). For example, it is difficult to fully explore uncertainties arising from data uncertainty or from the large numbers ensembles of AOGCM (Atmosphere-Ocean General Circulation Models) that are currently being generated. There is indeed a possibility that some sources of methodological and technological uncertainty, such as climate models and emission scenarios, might be more important than how the parameters of particular method are estimated (Thomas et al. 2004b, Berthelot et al. 2005, Stainforth et al. 2005). Techniques for handling and combining large ensembles of forecasts are also in their infancy and consensus projections may hide variability arising from disparate sources of uncertainty with existing tools being unable to successfully disentangle them. Thus, a more detailed analysis of the sources of uncertainties and their patterns is important to improve modeling strategies and to define over which sources an ensemble is necessary.

Here, we develop a new quantitative approach to analyze uncertainties in large ensembles of forecasts and disentangle the contribution of individual sources of variation entering the models. This approach is based on a spatially-explicit decomposition of total sum of squares of the ensemble-forecasted values of faunal turnover. Thus, our approach provides maps of uncertainty and allows an investigation of the regions more affected by particular sources of uncertainty. Buisson et al. (2009) proposed an alternative approach to partition the variance in ensembles of forecasts of species distributional changes that also allows an exploration of the geographic components of uncertainty. We applied our approach to understand the uncertainties in forecasts of species turnover maps of New World birds under climate change. Our study provides a comprehensive ensemble forecasting experiment to assess the relative contribution of seven species distribution models, five climate models, and two emission scenarios. Although other sources of uncertainty exist, our approach can be quickly expanded in the future to incorporate other sources of variation and thus it provides a fine perspective that enables new insights on how to better evaluate shifts in

biodiversity patterns and what are the greatest challenges at different levels of the modeling process.

Material and methods

Data

Data on the extent of occurrence (range filling) for 3837 species of the New World birds were downloaded from the NatureServe <www.natureserve.org/getData/birdMaps.jsp> and resampled to a grid of $1^\circ \times 1^\circ$ latitude/longitude. A similar approach for deriving species presence and absence maps from extent of occurrence data was adopted by Lawler et al. (2009). Although we acknowledge that this is not the most commonly used approach to model species distributions (which is usually based on more detailed data of species occurrences and fine scale environmental data), this allows a first understanding of continental patterns in species turnover based on a very large number of species.

Climatic data for species distribution modeling were derived from five coupled Atmosphere-Ocean General Circulation Models (AOGCMs), including CCSM3, CSIRO-Mk3.0, UKMO-HadCM3, ECHAM5/MPI-OM and MIROC. Although other AOGCMs are available, this selection covers a wide range of different predictions and was defined to maximize the different degree of predicted climate warming. The AOGCMs used here have different equilibrium climate sensitivity values ranging from 2.7°C to 4.3°C . Equilibrium climate sensibility is the annual mean surface air temperature change experienced by the climate system after it has attained a new equilibrium in response to a doubling of CO_2 concentration, and are within the range of all AOGCMs available from IPCC. These models also tried to encompass projections with different spatial resolutions, ranging from $1.1^\circ \times 1.1^\circ$ to $3.75^\circ \times 3.75^\circ$ latitude/longitude in the original set. Data were extracted from the World Climate Research Program's (WCRP) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset (Meehl et al. 2007).

Outputs for each model were obtained for two emission scenarios (A1 and B1) that are available for all AOGCMs selected above. In general, scenarios A1 and B1 can be roughly classified as "pessimistic" and "optimistic", respectively, according to the CO_2 emissions. The A1 storyline and scenario family assumes a future of very rapid economic

Table 1. Median proportions of the total sum of squares from the three-way ANOVA performed for each grid cell covering the New World, evaluating the relative contributions of method for niche modeling, AOGCM and emission scenario to the variability in forecasting species turnover. Minimum and maximum values in the maps are also given (see also Fig. 3). Geographical patterns in each variance component are estimated by Moran's I coefficient in the first distance class and by correlogram intercept (in km).

Source	SS(%)		Geographical patterns	
	Median	Min-max	Moran's I*	Correlogram intercept (km)
Method	66.10	3.8–94.7	0.489	4014
AOGCM	13.80	1.2–47.4	0.399	3088
Scenario	0.01	0.0–20.0	0.155	643
Method \times AOGCM	11.80	1.3–40.9	0.259	643
Method \times Scenario	0.70	0.0–15.7	0.219	1646
AOGCM \times Scenario	3.10	0.0–34.3	0.242	1646
Third-order interaction	2.30	0.2–15.2	0.199	643

*All Moran's I significant at $p < 0.01$.

growth and rapid introduction of more efficient technologies, but low population growth. A major underlying theme is a substantial reduction in regional differences in per capita income and, more specifically, the A1 scenario used here assumes a balanced mix of technologies and supply sources, with technology improvements and resource assumptions, including that no single energy source is overly dominant (IPCC 2000). The other scenario used herein (B1), also starts from the same low population growth rate, but it differs from A1 in assuming rapid changes in economic structures toward a service and information economy, with reductions in material intensity, and the introduction of clean and resource-efficient technologies. The emphasis is on global solutions to economic, social, and environmental sustainability, including improved equity, but without additional climate initiatives (IPCC 2000).

For each one of the AOGCMs and emission scenarios, four variables were obtained for both present time (baseline used to calibrate the models, the average values from 1970 to 1999) and future (estimated 2070–2099 interval, 2080 for simplicity hereafter). Variables used were mean annual rainfall and variability, average temperature of the warmest and coldest months. Rather than only using all these variables simultaneously to predict the geographic ranges of all species, the 15 combinations ($2^p - 1$, where $p=4$) of these four variables (Fig. 1) were used in independent models, accounting them for eventual differences in ecological processes driving geographic distributions of species across the New World. These variables encompass the major hypotheses which

are often raised to explain patterns of species richness at global scale (Hawkins et al. 2003).

Modeling species distributions

For each species, data were randomly divided into calibration and validation sets comprising 75 and 25% of the species' range, respectively, and the procedure was repeated 50 times, maintaining the observed prevalence of species in each partition (i.e. for presence only methods, 75% of the cells within the species' range, randomly defined, were used for modeling, whereas for presence-absence methods the analyses were conducted using a random sample of 75% of cells both inside and outside species' range). Beyond creating independent – or at least partially independent – sets for model calibration and validation, partition also allows to take data uncertainty into account, especially considering that range filling maps tend to have larger commission errors (Hawkins et al. 2008). Thus, each calibration dataset was used to project species distributions, according to seven SDMs described below (which were estimated with the 15 combinations of four environmental variables) on each combination of AOGCM and emission scenario previously defined. For species with range size > 10 cells (12% of species), rather than performing a cross-validation 50 times, each of the n cells was deleted once and analyses were repeated n times (roughly equivalent to a Jackknife procedure – Pearson et al. 2007).

We fitted seven species distribution models (SDMs) and projected species potential distributions for baseline and

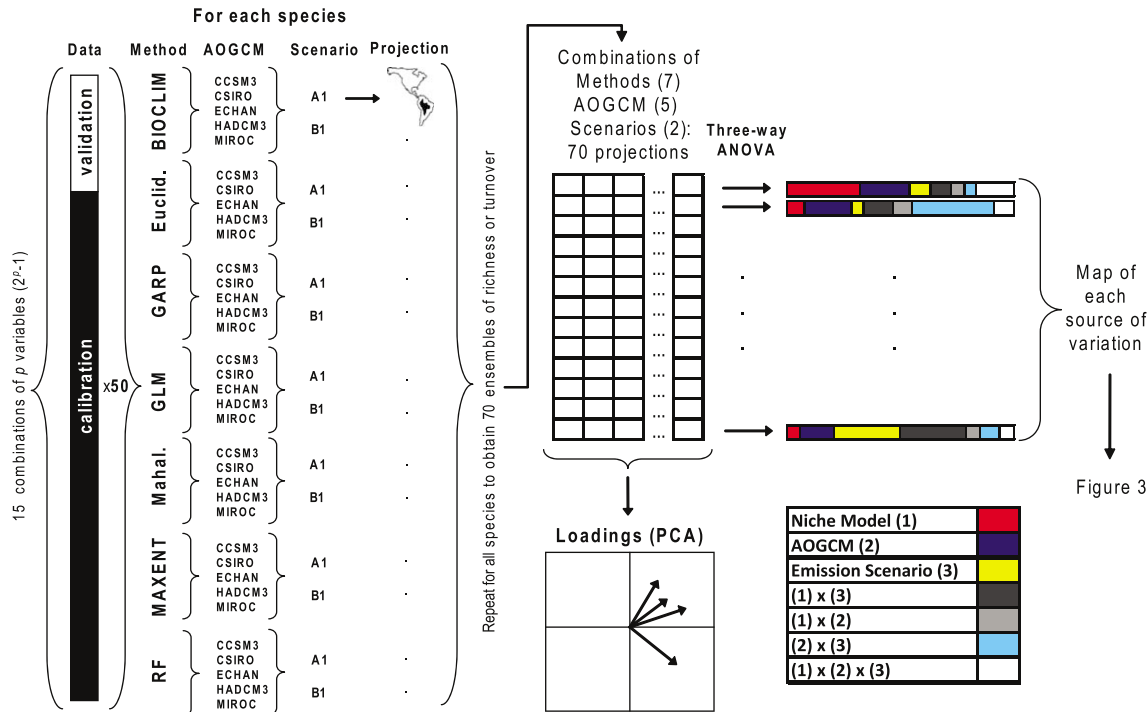


Figure 1. A schematic representation of the analytical framework used to evaluate spatial patterns of uncertainty in ensemble forecasting. Forecasting is generated using 15 combinations of the four bioclimatic variables, based on 50 random replications of calibration/validation datasets, for each method for niche modeling. Projections are based on the five AOGCMs for the two emission scenarios (A1 and B1). Then a three-way ANOVA is applied to each cell and the proportion of the total sum of squares accounted by each source can be mapped. A PCA can be used to evaluate the similarity among the ensemble-based vectors.

future climates (Fig. 1). The modeling methods used included a range of SDMs that are both conceptually and statistically different (Segurado and Araújo 2004, Elith et al. 2006, Tsoar et al. 2007, Phillips and Dudík 2008), such as simple surface range envelope models like BIOCLIM (Busby 1991), and Euclidian and Mahalanobis (EUC and MAHAL) distances (Farber and Kadmon 2003). We also fitted Generalized Linear Models (GLM) (McCullagh and Nelder 1989) and more complex machine learning approaches such as Random Forest (RF) (Breiman 2001), Genetic Algorithm for Rule Set Production (GARP) (Stockwell and Noble 1992), and Maximum Entropy (MAXENT) (Phillips et al. 2006, Phillips and Dudík 2008). We developed new computer software – BioEnsembles – in which all these methods were implemented. This software was designed to optimize and take advantage of high-speed parallel processing, both within (multi-processors computers) and between (grid architecture) computers.

Notice that the fitting and projection of alternative models using data partition and the 15 combinations of variables are used here to explore uncertainties from the initial conditions and model parameterization (*sensu* Araújo and New 2007). However, because of the large number of maps for each species, both the data-splitting procedures and variable selection were not included into the current analysis of sources of uncertainty. Rather, maps resulting from reshuffling data and variables were summed to generate the vector of ensemble frequencies of occurrence of the species for each of the combinations of SDM, AOGCM and emission scenarios (see below).

The True Skill Statistics (TSS) (Allouche et al. 2006), varying between -1 and 1 , was used as a fit statistic. It was calculated for each model based on the confusion matrix expressing matches and mismatches of observed and predicted occurrences in the validation data set. This matrix was computed after using ROC curves to convert continuous predictions into presence-absence. Models with TSS smaller than zero were discarded.

Also, it was not possible to fit all methods for all species using different combinations of variables (for example, lack of convergence in GLM or impossibility of inverting the covariance matrix when computing Mahalanobis distances). Most methods cannot deal with species occurring in a single cell (ca 1.2% of the species). Because of these restrictions, not all 3837 species were used for all SDM methods and AOGCMs (Supplementary material Table S1, Fig. S1). Thus, for each method and AOGCM, up to 750 models (50 dataset partitions modeled using 15 combinations of variables) were generated for each species. Finally, this combination of models generates an ensemble-based frequency of species distributions in the future and species are considered to occur in a given cell if at least 50% of the models predict its occurrence there (i.e. a majority consensus rule) (Araújo et al. 2005a, 2006).

Rather than evaluating each species' ensemble distribution independently, we calculated species turnover for each combination of SDM, AOGCM, and scenario, which was based on the number of potential species gained (G) or lost (L) within each cell and given by $(G+L)/(S+G)$ (Thuiller 2004). Notice that species turnover was calculated by comparing potential ranges as modeled in 2070–2099 and

the modeled ranges on baseline period, rather than observed ranges.

Evaluating the consistency between projections

Species turnover maps ($7 \text{ SDM} \times 5 \text{ AOGCMs} \times 2 \text{ emission scenarios}$, each one based on a maximum number of 750 models for each species) were averaged across each cell, generating a turnover consensus map, as well as standard deviations and coefficients of variation that allow mapping where uncertainty in projections is larger.

To evaluate the origins of variability around this consensus, we submitted the correlation matrix among the 70 projections to a Principal Component Analysis (PCA), as suggested by previous authors (Thuiller 2004, Araújo et al. 2005a, 2006). This allows evaluating the similarity of the projections (values of turnover as predicted by combinations of SDM, AOGCM and emission scenarios) by the loadings (i.e. the Pearson's correlation coefficients between predicted values and the scores) of the interpretable axes, which were defined by the broken-stick criterion (Legendre and Legendre 1998).

The loadings of the PCA can be used to express the relative position (similarity) of the maps from different SDM, AOGCMs and scenarios, whereas the PCA scores are the main directions of covariation among the maps throughout the projections variability. If the first principal component has a large relative eigenvalue, it tends to be highly correlated with the average consensus map and can be used as a consensus map as well (Araújo et al. 2005a, 2006, Marmion et al. 2009).

Mapping the sources of variation around the consensus solution

Although it is trivial to map the mean (consensus) turnover in each cell based on maps generated by different SDM, AOGCMs, and emission scenarios, as well as their variance or coefficient of variation, it is much more difficult to understand the main sources of variation around the consensus. The PCA described above can be used to evaluate similarity of maps, but it does not necessarily allow a formal partition of the sources contributing to the differences among the maps. To address this problem we performed a three-way Analysis of Variance (ANOVA) without replication (Sokal and Rohlf 1995, Legendre and Legendre 1998) for each cell, using species turnover as the response variable and SDM, AOGCM and emission scenarios as factors. We then obtained the sum of squares which can be attributed to each of these sources and their interaction ($\text{SDM} \times \text{AOGCM}$, $\text{SDM} \times \text{emission scenario}$, $\text{AOGCM} \times \text{emission scenario}$ and $\text{SDM} \times \text{AOGCM} \times \text{emission scenario}$). Notice that because this is a three-way ANOVA without replication, it is impossible to disentangle residual variance (i.e. part of variation not explained by SDM, AOGCM and emission scenario) and variance determined by the full (triple) interaction among these three sources. Because the levels in each factor (or source of variation) are a "sample" of possibilities (i.e. other AOGCMs and SDM), we could think in all these factors as random effects, although as pointed out above they were

selected to cover most of the range of variation within each factor.

We estimated the variance components as the simple proportions of the sum of squares attributable to the three sources (and their interaction) in respect to the total sum of squares. As we performed the analyses for each cell in the grid covering the New World (Fig. 1), it was possible to map each variance component and, in this way, identify regions of low and high uncertainty and the main sources accounting for this uncertainty. Notice that ANOVA was applied here to a turnover metrics, which varies between 0 and 1, so that violations in the assumptions of normality are not unlike. This may be not a problem when dealing with other metrics, but it is difficult to check for this problem in every grid cell. However, we believe that our results are robust to these problems and performing a square root/arcsin transformation of turnovers prior to the ANOVA (which are likely to improve the models) did not qualitatively affect the patterns in maps and the relative magnitude of variance components. The correlation between variance components from transformed and untransformed turnover metrics was always higher than 0.95 (median magnitudes all the same up to the second decimal place).

We explicitly quantified the spatial patterns in each variance components using correlograms, based on Moran's I spatial autocorrelation coefficients calculated for 10 geographic distance classes (Legendre and Legendre 1998). Magnitude of spatial pattern was established by Moran's I in the first distance class and by the correlograms' X-intercept (i.e. the distance at which autocorrelation becomes negative). Spatial analyses were performed in SAM (Spatial Analysis in Macroecology) software, freely available at <www.ecoevol.ufg.br/sam> (Rangel et al. 2006).

Results

A consensus map of mean species turnover across the 70 ensembled combinations of SDM, AOGCM and emission scenario (Fig. 2A), shows relatively high turnover (up to 56%) in northern parts of North America, in the Amazon and across the Andean region in South America, in Central America, throughout Mexico and in southeastern US. However, there was a high variation among projections, mainly in north and northwestern North America and parts of the Amazon, with coefficients of variation going up to 90% (Fig. 2B).

The first axis of the principal component analysis applied to the correlation matrix among the 70 turnover maps explained only 29.3% of the correlation structure, whereas the second principal component explained 23.5% of the variation. The first five axes selected by a broken-stick criterion explained 67.5% of the variation among turnover maps, indicating thus a marked level of heterogeneity among them. Spatial patterns and magnitude of turnover were quite different among combinations of SDM, AOGCM and emission scenarios (Supplementary material Table S2). Thus, it is difficult to disentangle by a simple visual inspection of the loading structure which sources of variation contributed more to the variability in the

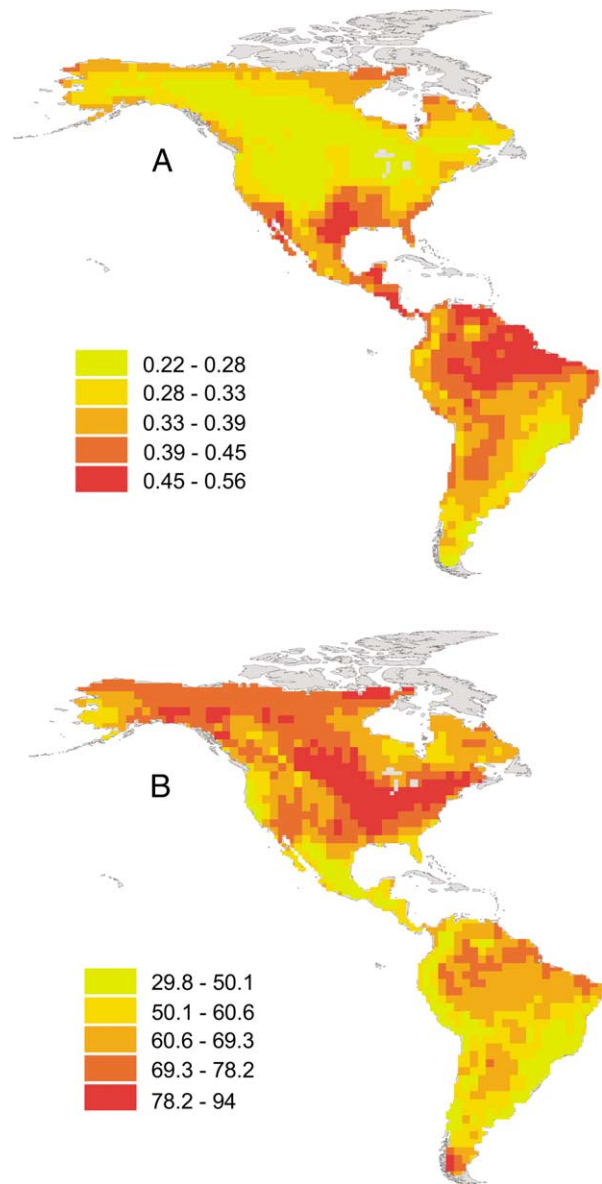


Figure 2. Consensus patterns of species turnover (A) and coefficients of variation, in percentage, among the 70 turnover maps (B) based on 3837 species of New World birds.

ensemble of forecasts (although this may be useful for a posteriori comparisons – see below).

The three-way ANOVA applied to each cell indicated that the distinct sources of variation have different contributions to the geographically-structured variation around the consensus solution observed in Fig. 2A. Out of the main effects, SDM explained a high proportion of the total sum of squares with a median value of 66%, ranging from 4 up to 95% (Table 1). High proportions of the total sum of squares that can be attributable to this factor were found in all North America and in the Amazon (Fig. 3A).

There is a high correlation ($r=0.88$) between the Moran's I coefficients in the first distance class and the median proportion of variation accounted for by each source of uncertainty (Table 1). Thus, factors accounting for most of the variability among the forecasts are also the

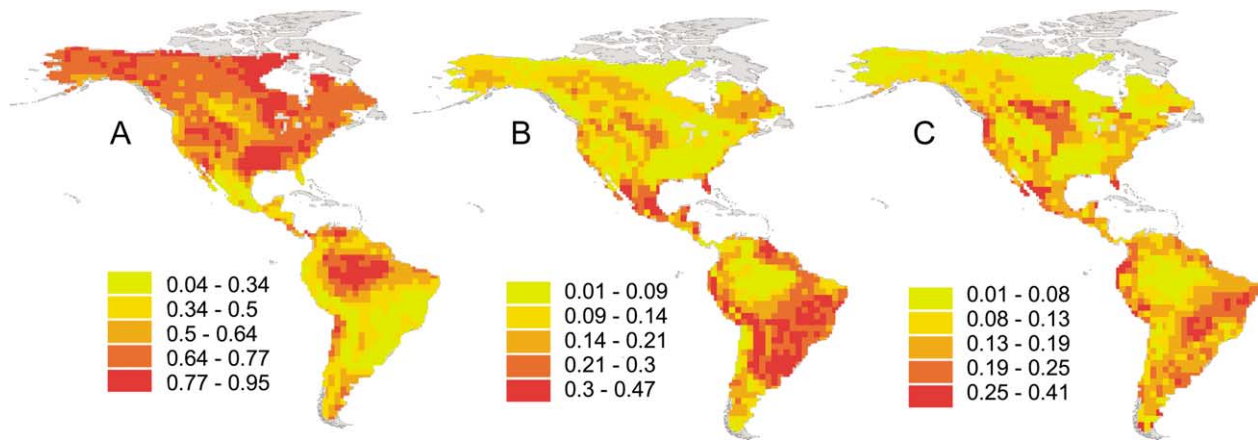


Figure 3. Proportion of the total sum of squares accounted for by SDM (A), AOGCM (B) and the interaction between these factors (C).

ones with the strongest spatial patterns. For instance, the variance component associated to SDM (the main source of variation according to the decomposition of the total sum of squares) is structured at broad geographical scales, with a large Moran's I in the first distance class and positive coefficients extending up to ca 4000 km (Table 1). On the other hand, although AOGCM did not have a very high median value (Table 1), the proportion of the total sum of squares attributable to this source can be as high as 47% in most of South America, as well as in Central

America (Fig. 3B). Geographical patterns expressed in the correlograms are also relatively strong (Moran's I in the first distance class equal to 0.399) and positive autocorrelation can be found up to 3000 km (Table 1). Among the interactions, the most important was the one between SDM and the AOGCM, with proportions ranging from 1.3 to 40.9%, with a median of 11.8% (Table 1). The highest values for this interaction were found in central North America and in the dry regions of eastern part of South America (Fig. 3C). This variance component has

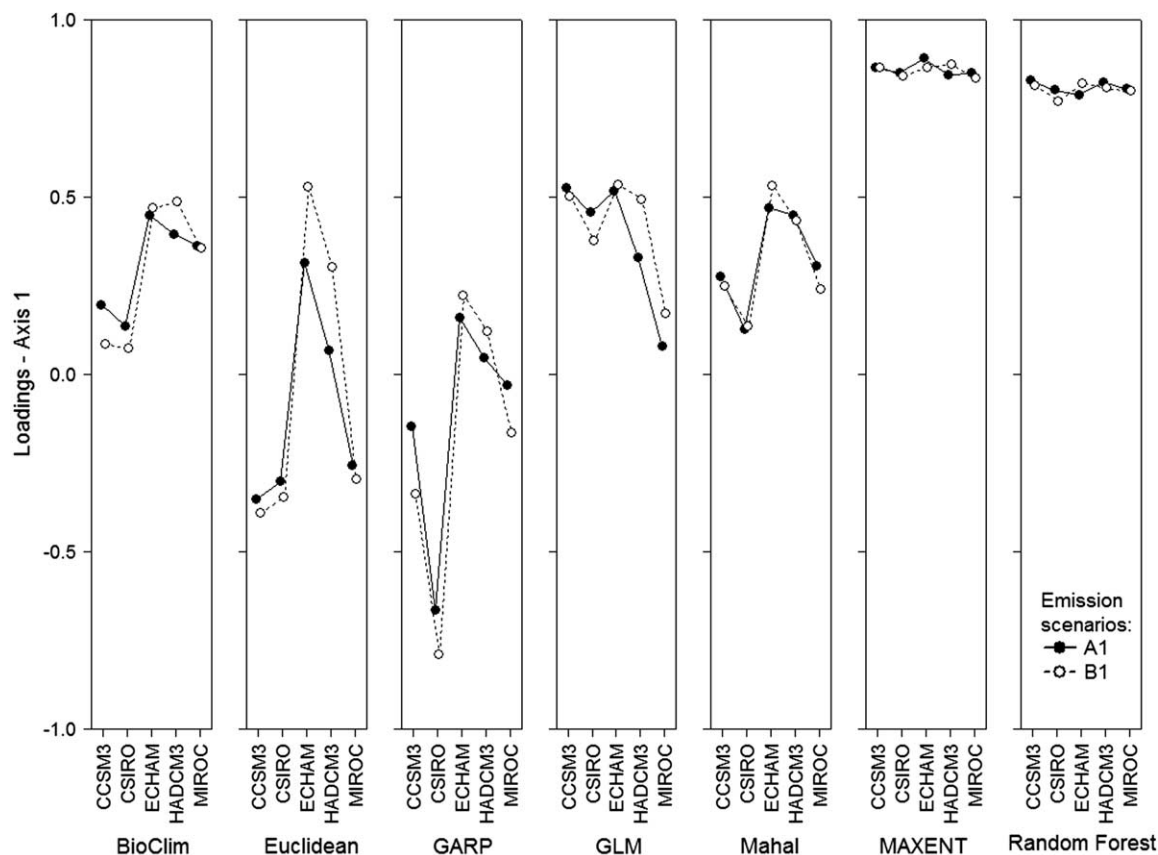


Figure 4. Loadings of the first principal component extracted from a matrix of species turnover forecasted by different methods for niche modeling, AOGCMs and emission scenarios.

spatial autocorrelation only at the smallest distance classes, with positive Moran's I (0.259) in the first distance class (up to 640 km) (Table 1).

The proportion of the total sum of squares revealed wide differences in species turnover patterns derived from distinct SDM and AOGCMs, and these components are geographically structured. These differences in maps can now be analyzed in more detail by examining the loadings of the first principal component. This analysis works here as a multivariate version of "a posteriori" comparisons in ANOVA (Sokal and Rohlf 1995), expressing the relative importance of each vector of prediction to the consensus map (Fig. 4). According to the loadings of the first principal component, the main difference among methods can be seen between a cluster of RF and MAXENT (with higher loadings in PC1) against Euclidean and GARP (with lowest loadings, but with high variation), with BIOCLIM, MAHAL and GLM occupying an intermediate position. All methods, except for RF and MAXENT, produce different results when using different AOGCMs, which explains the interaction term and makes it more difficult to interpret the effects of AOGCM alone (methods for modeling approaches, because of their larger effects, are easier to interpret).

Discussion

Partitioning uncertainties

Our study provides an illustration of how variation in ensembles of forecasts can be partitioned, thus offering a tool for investigating the origins of the uncertainties entering the models. Even if we accept that ensemble forecasts generate more accurate (Araújo et al. 2005a), or at least more conservative projections (Marmion et al. 2009), it is still important to identify the main sources of variation that affect the averaged projections (Brook et al. 2009, Elith and Graham 2009).

Previous studies (Thuiller 2004, Araújo et al. 2005a, 2006) used principal component analysis, or other classification analyses, as exploratory tools to describe the relative similarity among maps produced using different SDM or AOGCMs, allowing then a qualitative assessment of the relative importance of uncertainty sources. However, when several sources of uncertainty are explored, variability among projections might display complex patterns that might be difficult to interpret with the visual inspection of PCA loadings. Furthermore, formal quantitative assessments of uncertainty are necessary if they are to be systematically addressed and conveyed to model users.

Dormann et al. (2008) pioneered the use of ANOVA designs to uncouple uncertainties in SDM. They showed that the variability among statistics used to evaluate models projections of great grey shrike's distributions was, on average, 60% attributable to the use of distinct SDM. Our approach differs from Dormann et al. (2008) in that we quantify variance in the projected distribution maps rather than in the model fit statistics, which have more complex properties and are difficult to interpret in the context of model "transferability" (i.e. projecting the results of SDM in a different region or time; Araújo et al. 2005b, Araújo

and Rahbek 2006, Randin et al. 2006). A manuscript recently accepted for publication in *Global Change Biology* and kindly supplied by one of the authors (Buisson et al. 2009) proposed an alternative approach to partition the variance in ensemble-based forecasts of species turnover that also allows an exploration of the geographic components of uncertainty, as performed here. However, Buisson et al. (2009) used a GLM to evaluate only the main sources of uncertainty, but they did not explore the interactions between SDM and AOGCMs, and indeed our results showed that differences among SDMs are not the same when their projections are based on different AOGCMs (Fig. 4).

It is important to notice that the relative importance of each source of uncertainty depends of the variation among the levels within each factor. For all factors analyzed here, we tried to maximize the variation among levels by selecting SDM, AOGCMs, emission scenarios and data (species) with different characteristics. For example, the PCA reveals that differences among forecasts derived from different SDM are in line with current knowledge on how these methods work and how they are classified (Elith and Graham 2009). It is understood that SDM tend to differ in model fit, but it is less clear what is the link between fit and "transferability" of the models (see below). So, if one uses sophisticated methods such as MAXENT and RF (i.e. assuming that a good fit indicates high transferability – see below), the relative importance of SDM may be reduced (Buisson et al. 2009). Indeed, if only these two methods are used, the median proportion of variation accounted by SDM falls from 66.1 to only 4.2% (whereas the median proportion accounted by AOGCM increases from 13.8 to 53%). Clearly, discussions around the relatively magnitude of these sources of uncertainty (Thomas et al. 2004b, Thuiller et al. 2004) must take into account which amount of variation within a factor the levels used (different SDM or AOGCM) cover in a comparative analysis.

Our results based on the turnover of New World birds clearly showed that the choice of the SDM contributed the most to uncertainty in the range of predictions, when compared with AOGCM and emission scenarios. These results are in line with previous studies showing that different niche modeling approaches may produce markedly different predictions of species range changes under climate change (Thuiller et al. 2004, Araújo et al. 2005a, b, 2006, Araújo and Rahbek 2006, Pearson et al. 2006). Our approach further revealed that the AOGCM would contribute with more uncertainty than the emission scenarios, and its interaction with SDM could be as important as the effects of AOGCM alone.

The discrepancy among the results obtained with different SDM has certainly motivated a large body of literature on comparison of methods that tries to find the "best" predictions (Manel et al. 1999, Thuiller 2003, Brotons et al. 2004, Segurado and Araújo 2004, Elith et al. 2006, Lawler et al. 2006, Meynard and Quinn 2007, Peterson et al. 2007, Tsoar et al. 2007) and discusses how to evaluate modeling approaches, both in terms of model fit (Fielding and Bell 1997, Liu et al. 2005, Lobo et al. 2008, Peterson et al. 2008) and model transferability (Araújo et al. 2005b, Randin et al. 2006, Peterson et al. 2007, Fitzpatrick et al. 2008, Peterson and Nakazawa 2008, Phillips 2008).

It is important to highlight that, in most cases, fit-statistics provide measures of the adjustment of projections to the data used for calibration, but in other cases, fit statistics measure how well model projections fit data sets apart for evaluation. The problem is the lack of independence as evaluation data are frequently spatially and/or temporally autocorrelated with the data used for calibration. Therefore, fit statistics provide an inflated and sometimes spurious measure of the models' suitability to be transferred into independent settings (Araújo et al. 2005b, Araújo and Rahbek 2006, Randin et al. 2006, Peterson et al. 2008).

In addition to such statistical considerations on model fit and transferability, there is also a discussion about the components of the niche that are captured by each model (Araújo and Guisan 2006, Soberón 2007, Jiménez-Valverde et al. 2008), as well as the relative roles of multiple ecological and evolutionary processes driving current species' ranges in deterministic and stochastic ways and how they can be incorporated into the methods (Araújo and Luoto 2007, De Marco et al. 2008, Rickebusch et al. 2008, Anderson et al. 2009).

Thus, supposing that further studies on all these issues will be able to discard some of the SDMs or AOGCMs (considering the criteria of model fit and transferability, or unreliable AOGCMs), we can expect that the uncertainties around the ensemble-based forecasts would be reduced. In a very optimistic (and most unlikely) scenario, if researchers found a definitive solution about which SDM should be selected (the main source of variation according to our results), then the level of uncertainty could be drastically reduced.

Mapping uncertainties

In addition to allowing a quantitative evaluation of the relative importance of different sources of variation around a consensus solution discussed above, our approach also allows mapping the variance components. This can be important because it adds another dimension (geography) to evaluate the uncertainties, giving more information on where the consensus can be achieved with low variation and where more research is needed to minimize variance.

In principle, four independent combinations of the two characteristics of the variance component maps, their geographic structure and their magnitude, could be found, forming a "two-by-two" scheme. Finding which of these combinations exist for a particular analysis possesses some interesting implications for practical decisions regarding forecasting. The first combination is given by a high level of variability which is also geographically-structured. In this case, uncertainty is not spatially random, which can shed light to the problems in each factor generating uncertainty. For example, AOGCMs can give different predictions in regions with a particular environmental characteristic, whereas all methods for SDM can provide similar solutions in a given region and differ in others (Beaumont et al. 2007). These geographically-structured components also show that when analyzing a given region more emphasis can be given in a particular source of uncertainty. Although short-distance spatial autocorrelation in the variance components is inevitable, because of autocorrelation in climate

and distributional data, broad-scale patterns may indicate more complex patterns that require ecological or methodological interpretations.

The second combination can be given by a high, but geographically random variance component. This is the most challenging combination because it will be hard to predict regions of high or low uncertainty or establish which levels within a factor (i.e. SDM) can be used alone without increasing uncertainty. In this case, the differences among models cannot be associated with geographically-structured factors (e.g. environmental characteristics), so that it is more difficult to understand variation among projections. Under this combination, the use of ensemble-based forecasts is probably the best analytical strategies for forecasting.

The third combination, a geographic structure in a variance component with low mean, is unlikely to appear in real data. If the effect is small, there is also small variation among cells and it is unlike that any spatial pattern appears. Finally, the forth combination, given by low variability among results within cells which is also geographically random, indicates that the source of uncertainty is not important at all.

For the New World birds, we found a correlation between geographic structure and the relative proportion of variation accounted for by each source of uncertainty, reinforcing that the two characteristics of the variance component maps (i.e. magnitude and spatial pattern of the component) are not independent, so that the second and third combinations described above are not found.

The map of coefficient of variation in turnover shows that most differences among ensemble-based projections were found in the northern temperate region of the New World, and in the Amazon. A visual inspection of the maps averaged across SDM and AOGCMs further revealed why these differences arise (see also Supplementary material Fig. S2, Fig. S3). For example, some methods, such as RF, MAXENT, EUC and GLM, did not predict high turnovers in central Amazon (although this still depends on AOGCMs for some methods), which explains why the variance component of niche modeling techniques in this region is much higher (up to 90%) than for other regions of the New World (Fig. 3A). In general, MAXENT, RF and GLM predict smaller turnover across the continent than other methods, and the other methods vary a lot in their prediction of turnover in the northern part of the continent. Indeed, Fig. 3 suggest that if predictions of turnover rates are to be made for a few regions in the New World, such as the southeastern coast of Brazil, methods for niche modeling tend to give similar results and their differences are of minor concern. On the other hand, if one is interested in predictions for northern hemisphere of the New World or Amazon, especially northern US and Canada, AOGCMs are not an important source of uncertainty (so any one could be used) and one should focus on why methods are giving different answers. These patterns can also have implications for conservation decisions and it is important to notice, for example, that regions with higher SDM uncertainty are also those with high turnover levels detected by Lawler et al. (2009), based on random forest.

Concluding remarks

Our analyses support previous findings that SDM is the main source of uncertainty in forecasts of species range shifts under climate change and clearly highlights the importance of ensemble forecasting because of the current difficulties in the statistical evaluating model fit and transferability. This conclusion does not oppose the view that reductions of uncertainty in ensembles forecasts still demand a better evaluation of the individual SDM that compose the ensembles. Although the effects of SDM, AOGCMs and emission scenarios have been continuously evaluated in the literature, our approach provides a quantitative evaluation of the magnitude and geographical structure of these sources of uncertainty. Also, it can be easily expanded to encompass more complex designs addressing a larger spectrum of sources of variation in ensemble forecasting. Moreover, mapping uncertainty brings a new avenue for research, as it reveals that even when a given study compares sources of uncertainty they are not necessarily the same across different parts of the globe.

Acknowledgements – We thank Paulo De Marco Jr and Alexandre S. G. Coelho for helpful discussions on the ANOVA method and to Carsten Rahbek, Alan Fielding, Barry Brook and four anonymous reviewers for suggestions that improved initial versions of the manuscript. We also thank Wilfried Thuiller for discussion and for sharing an accepted manuscript on variance partition a couple of days before submission of this manuscript (Buisson et al. 2009). Financial support for this study was provided by the BBVA Foundation to the BioImpact project. Work by J. A. F. Diniz-Filho and L. M. Bini has been continuously supported by productivity grants from CNPq. Work by T. F. Rangel is supported by a CAPES Ph.D. fellowship. CH is supported by a Ph.D. studentship and DNB is supported by a post-doctoral grant of the Danish Center for Macroecology, Evolution and Climate. DNB also wants to thank the Danish National Research Foundation for support to the Center for Macroecology, Evolution and Climate. MBA has also been supported by the FP6 EU funded Ecochange project.

References

- Allouche, O. et al. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). – *J. Appl. Ecol.* 43: 1223–1232.
- Anderson, B. J. et al. 2009. Dynamics of range margins for metapopulations under climate change. – *Proc. R. Soc. B* 276: 1415–1420.
- Araújo, M. B. and Guisan, A. 2006. Five or so challenges for species distribution modelling. – *J. Biogeogr.* 33: 1677–1688.
- Araújo, M. B. and Rahbek, C. 2006. How does climate change affect biodiversity? – *Science* 313: 1396–1397.
- Araújo, M. B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. – *Global Ecol. Biogeogr.* 16: 743–753.
- Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – *Trends Ecol. Evol.* 22: 42–47.
- Araújo, M. B. et al. 2005a. Reducing uncertainty in projections of extinction risk from climate change. – *Global Ecol. Biogeogr.* 14: 529–538.
- Araújo, M. B. et al. 2005b. Validation of species-climate impact models under climate change. – *Global Change Biol.* 11: 1504–1513.
- Araújo, M. B. et al. 2006. Climate warming and the decline of amphibians and reptiles in Europe. – *J. Biogeogr.* 33: 1712–1728.
- Araújo, M. B. et al. 2008. Quaternary climate changes explain diversity among reptiles and amphibians. – *Ecography* 31: 8–15.
- Beaumont, L. J. et al. 2007. Where will species go? Incorporating new advances in climate modeling into projections of species distributions. – *Global Change Biol.* 13: 1368–1385.
- Berthelot, M. et al. 2005. How uncertainties in future climate change predictions translate into future terrestrial carbon flux. – *Global Change Biol.* 11: 959–970.
- Breiman, L. 2001. Random forest. – *Mach. Learn.* 45: 5–32.
- Brook, B. et al. 2009. Integrating bioclimate with population models to improve forecasts of species extinctions under climate change. – *Biol. Lett.* in press, doi: 10.1098/rsbl.2009.0480.
- Brotons, L. et al. 2004. Presence-absence vs presence-only modelling methods for predicting bird habitat suitability. – *Ecography* 27: 165–172.
- Buisson, L. et al. 2009. Uncertainty in ensemble forecasting of species distribution. – *Global Change Biol.*, doi: 10.1111/j.1365-2486.2009.02000.x.
- Busby, J. R. 1991. BIOCLIM: a bioclimate analysis and prediction system. – In: Margules, C. R. and Austin, M. P. (eds), *Nature conservation: cost effective biological surveys and data analysis*. CSIRO, pp. 64–68.
- De Marco, P. et al. 2008. Spatial analysis improves species distribution modelling during range expansion. – *Biol. Lett.* 4: 577–580.
- Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. – *Basic Appl. Ecol.* 8: 387–397.
- Dormann, C. F. et al. 2008. Components of uncertainty on species distribution analysis: a case study of the great grey shrike. – *Ecology* 89: 3371–3386.
- Elith, J. and Graham, C. H. 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. – *Ecography* 32: 66–77.
- Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.
- Farber, O. and Kadmon, R. 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. – *Ecol. Model.* 160: 115–130.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – *Environ. Conserv.* 24: 38–49.
- Fitzpatrick, M. C. et al. 2008. Data set matter, but so do evolution and ecology. – *Global Ecol. Biogeogr.* 17: 562–565.
- Guisan, A. and Thuiller, W. 2005. Predicting species distributions: offering more than simple habitat models. – *Ecol. Lett.* 8: 993–1009.
- Hawkins, B. A. et al. 2003. Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. – *Ecology* 84: 1608–1623.
- Hawkins, B. A. et al. 2008. What do range maps and surveys tell us about diversity patterns? – *Folia Geobot.* 43: 345–355.
- Heikkinen, R. K. et al. 2006. Methods and uncertainties in bioclimatic envelope modeling under climate change. – *Progr. Phys. Geogr.* 30: 751–777.
- IPCC 2000. Emissions scenarios. – Cambridge Univ. Press.
- Jiménez-Valverde, A. et al. 2008. Not as good as they seem: the importance of concepts in species distribution modeling. – *Divers. Distrib.* 14: 885–890.
- Lawler, J. J. et al. 2006. Predicting climate-induced range shifts: model differences and model reliability. – *Global Change Biol.* 12: 1568–1584.

- Lawler, J. J. et al. 2009. Projected climate-induced faunal change in the western Hemisphere. – *Ecology* 90: 588–597.
- Legendre, P. and Legendre, L. 1998. Numerical ecology, 3rd ed. – Elsevier.
- Leutbecher, M. and Palmer, T. N. 2008. Ensemble forecasting. – *J. Comput. Phys.* 227: 3515–3539.
- Liu, C. et al. 2005. Selecting thresholds of occurrence in the prediction of species distributions. – *Ecography* 28: 385–393.
- Lobo, J. M. et al. 2008. AUC: a misleading measure of the performance of predictive distribution models. – *Global Ecol. Biogeogr.* 17: 145–151.
- Manel, S. et al. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distribution: a case study with a Himalayan river bird. – *Ecol. Model.* 120: 337–347.
- Marmion, M. et al. 2009. Evaluation of consensus methods in predictive species distribution modeling. – *Divers. Distrib.* 15: 59–69.
- McCullagh, P. and Nelder, J. A. 1989. Generalized linear models, 2nd ed. – Chapman and Hall/CRC.
- Meehl, G. A. et al. 2007. The WCRP CMIP3 multi-model dataset: a new era in climate change research. – *Bull. Am. Meteorol. Soc.* 88: 1383–1394.
- Meynard, C. N. and Quinn, J. F. 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. – *J. Biogeogr.* 34: 1455–1469.
- O’Hanley, J. R. 2009. NeuralEnsembles: a neural network based ensemble forecasting program for habitat and bioclimatic suitability analysis. – *Ecography* 32: 89–93.
- Pearman, P. B. et al. 2008. Niche dynamics in space and time. – *Trends Ecol. Evol.* 23: 149–158.
- Pearson, R. G. and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelopes useful? – *Global Ecol. Biogeogr.* 12: 361–371.
- Pearson, R. G. et al. 2006. Model-based uncertainty in species range prediction. – *J. Biogeogr.* 33: 1704–1711.
- Pearson, R. G. et al. 2007. Predicting species distributions from small number of occurrence records: a test case using cryptic geckos in Madagascar. – *J. Biogeogr.* 34: 102–117.
- Peterson, A. T. and Nakazawa, Y. 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. – *Global Ecol. Biogeogr.* 17: 135–144.
- Peterson, A. T. et al. 2002. Future projections for Mexican faunas under global climate change scenarios. – *Nature* 416: 626–629.
- Peterson, A. T. et al. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. – *Ecography* 30: 550–560.
- Peterson, A. T. et al. 2008. Rethinking receiver operating characteristic analysis: applications in ecological niche modeling. – *Ecol. Model.* 213: 63–72.
- Phillips, S. J. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. 2007. – *Ecography* 31: 272–278.
- Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – *Ecography* 31: 161–175.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Randin, C. F. et al. 2006. Are niche-based species distribution models transferable in space? – *J. Biogeogr.* 33: 1689–1703.
- Rangel, T. F. L. V. B. et al. 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. – *Global Ecol. Biogeogr.* 15: 321–431.
- Rickebusch, S. et al. 2008. Incorporating the effects of changes in vegetation functioning and CO₂ on water availability in plant habitat models. – *Biol. Lett.* 4: 556–559.
- Roura-Pascual, N. et al. 2009. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. – *Biol. Inv.* 11: 1017–1031.
- Segurado, P. and Araújo, M. B. 2004. An evaluation of methods for modelling species distributions. – *J. Biogeogr.* 31: 1555–1568.
- Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. – *Ecol. Lett.* 10: 1115–1123.
- Sokal, R. R. and Rohlf, F. J. 1995. Biometry, 3rd ed. – W. H. Freeman.
- Stainforth, D. A. et al. 2005. Uncertainty in predictions of the climate responses to rising levels of greenhouse gases. – *Nature* 433: 403–406.
- Stockwell, D. R. B. and Noble, I. R. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. – *Math. Comput. Simul.* 33: 385–390.
- Thomas, C. D. et al. 2004a. Extinction risk from climate change. – *Nature* 427: 145–148.
- Thomas, C. D. et al. 2004b. Biodiversity conservation: uncertainty in predictions of extinction risk/Effects of changes in climate and land use/Climate change and extinction risk (reply). – *Nature* 430, 1 July 2004, doi:10.1038/nature02719.
- Thuiller, W. 2003. BIOMOD: optimising predictions of species distributions and projecting potential future shifts under global change. – *Global Change Biol.* 9: 1353–1362.
- Thuiller, W. 2004. Patterns and uncertainties of species’ range shifts under climate change. – *Global Change Biol.* 10: 2020–2027.
- Thuiller, W. et al. 2004. Biodiversity conservation: uncertainty in predictions of extinction risk. – *Nature* 430, 1 July 2004, doi:10.1038/nature02716.
- Thuiller, W. et al. 2005a. Niche properties and geographical extent as predictors of species sensitivity to climate change. – *Global Ecol. Biogeogr.* 14: 347–357.
- Thuiller, W. et al. 2005b. Climate change threatens plant diversity in Europe. – *Proc. Nat. Acad. Sci. USA* 102: 8245–8250.
- Thuiller, W. et al. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. – *Ecography* 32: 369–373.
- Tsoar, A. et al. 2007. A comparative evaluation of presence only methods for modelling species distribution. – *Divers. Distrib.* 13: 397–405.

Download the Supplementary material as file E6196 from
<www.oikos.ekol.lu.se/appendix>.