

# Uncertainty in ensemble forecasting of species distribution

LAËTITIA BUISSON<sup>\*,†</sup>, WILFRIED THUILLER<sup>‡</sup>, NICOLAS CASAJUS<sup>§</sup>, SOVAN LEK<sup>\*</sup> and GAËL GRENOUILLET<sup>\*</sup>

<sup>\*</sup>Laboratoire Evolution et Diversité Biologique, UMR CNRS 5174, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France, <sup>†</sup>Laboratoire d'Ecologie Fonctionnelle, UMR CNRS 5245, ENSAT, Avenue de l'Agrobiopole, BP 32607, Auzeville-Tolosane, 31326 Castanet-Tolosan, France, <sup>‡</sup>Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France, <sup>§</sup>Université du Québec à Rimouski, 300 allée des Ursulines, Rimouski, QC, G5L 3A1, Canada

## Abstract

Species distribution modelling has been widely applied in order to assess the potential impacts of climate change on biodiversity. Many methodological decisions, taken during the modelling process and forecasts, may, however, lead to a large variability in the assessment of future impacts. Using measures of species range change and turnover, the potential impacts of climate change on French stream fish species and assemblages were evaluated. Our main focus was to quantify the uncertainty in the projections of these impacts arising from four sources of uncertainty: initial datasets (Data), statistical methods [species distribution models (SDM)], general circulation models (GCM), and gas emission scenarios (GES). Several modalities of the aforementioned uncertainty sources were combined in an ensemble forecasting framework resulting in 8400 different projections. The variance explained by each source was then extracted from this whole ensemble of projections. Overall, SDM contributed to the largest variation in projections, followed by GCM, whose contribution increased over time equalling almost the proportion of variance explained by SDM in 2080. Data and GES had little influence on the variability in projections. Future projections of range change were more consistent for species with a large geographical extent (i.e., distribution along latitudinal or stream gradients) or with restricted environmental requirements (i.e., small thermal or elevation ranges). Variability in projections of turnover was spatially structured at the scale of France, indicating that certain particular geographical areas should be considered with care when projecting the potential impacts of climate change. The results of this study, therefore, emphasized that particular attention should be paid to the use of predictions ensembles resulting from the application of several statistical methods and climate models. Moreover, forecasted impacts of climate change should always be provided with an assessment of their uncertainty, so that management and conservation decisions can be taken in the full knowledge of their reliability.

**Keywords:** climate change, ensemble forecasting, environmental gradient, spatial autocorrelation, species distribution, uncertainty

Received 27 March 2009 and accepted 23 May 2009

## Introduction

Recent advances in computing, and the development of large databases, have made species distribution modelling easier, faster, and more widely used in both basic and applied ecology. Species distribution models (SDM) have become increasingly popular in recent years for

predicting the suitability of the current habitat (Guisan & Zimmermann, 2000), evaluating the potential for the establishment and spread of invasive species (e.g., Hartley *et al.*, 2006; Ficetola *et al.*, 2007; Mika *et al.*, 2008; Roura-Pascual *et al.*, 2009) or assessing the magnitude of biological responses to environmental changes such as climate change (e.g., Erasmus *et al.*, 2002; Peterson *et al.*, 2002; Thuiller, 2003; Bomhard *et al.*, 2005; Thuiller *et al.*, 2005; Araujo *et al.*, 2006).

Correspondence: Laëtitia Buisson, tel. +33 5 61 55 69 11, fax + 33 5 61 55 67 28, e-mail: buisson@cict.fr

To date, many statistical methods are available and applied routinely. Facing this wide range of methods, several studies have focused on comparing model performances and ecological predictions. These studies have found that SDM do not perform equally in predicting current species distribution (e.g., Segurado & Araujo, 2004; Elith *et al.*, 2006; Lawler *et al.*, 2006; Pearson *et al.*, 2006; Guisan *et al.*, 2007; Heikkinen *et al.*, 2007) and that SDM give contrasted predictions of habitat suitability (e.g., Elith *et al.*, 2006; Guisan *et al.*, 2007; Evangelista *et al.*, 2008; Roura-Pascual *et al.*, 2009). Predictions have also been shown to be sensitive to diverse steps of the modelling process such as model parameterization or model selection criteria (Araujo & Guisan, 2006; Elith *et al.*, 2006; Heikkinen *et al.*, 2006; Dormann *et al.*, 2008). Some data characteristics, such as sample size, errors in sampling both species distribution and environmental variables, or correlations between environmental descriptors, have also been revealed as important sources of variability (Barry & Elith, 2006; Guisan *et al.*, 2006; Heikkinen *et al.*, 2006; Dormann *et al.*, 2008; Graham *et al.*, 2008).

A growing concern has recently emerged for ensemble forecasting approaches, which fit a number of alternative models (i.e., various initial conditions combined with several statistical methods) and explore the range of resulting projections (Araujo & New, 2007). Most studies that have explored several models in ensemble framework have mainly attempted (i) to identify the 'best' model (i.e., the model with the best predictive performance) among extensive model comparisons (e.g., Segurado & Araujo, 2004; Elith *et al.*, 2006; Lawler *et al.*, 2006; Sharma & Jackson, 2008), or (ii) to use a consensus approach that summarizes the variability within the ensemble of predictions (e.g., Thuiller, 2004; Araujo *et al.*, 2005b, 2006; Crossman & Bass, 2008; Barbet-Massin *et al.*, 2009; Lawler *et al.*, 2009; Marmion *et al.*, 2009b). Variability in projections of such ensembles, however, has rarely been considered itself (but see Hartley *et al.*, 2006, for a special case of one model with several parameterizations), although calls have arisen to improve knowledge of the uncertainty factors that may decrease the reliability of predictions (Thuiller, 2004; Araujo *et al.*, 2005b; Barry & Elith, 2006; Guisan *et al.*, 2006; Heikkinen *et al.*, 2006).

In the context of climate change, uncertainty in projections becomes even more worrying as additional sources of variability arise at two levels with the use of future climate scenarios (Heikkinen *et al.*, 2006; Araujo & New, 2007; Beaumont *et al.*, 2008; Dormann *et al.*, 2008). Firstly, a large number of general circulation models (GCM) have been developed simultaneously, by several meteorological research centres, to represent physical processes in the atmosphere, ocean, cryo-

sphere, and land surface, allowing to simulate the response of the global climate system to increasing greenhouse gas concentrations. Secondly, four storylines, which are alternative images of how the future might unfold assuming a certain level of future greenhouse gas emissions, have been defined (Special Report on Emission Scenarios, Nakicenovic & Swart, 2000). These greenhouse gas emission scenarios (GES) are derived from a complex interplay between demographic and socio-economic developments, as well as technological changes. The outputs of all GCM  $\times$  GES combinations, therefore, cover a large range of potential future climate conditions, thus increasingly emphasizing the need of ensemble forecasting approaches when assessing uncertainties in future species distributions.

While dozens of previous ecological studies have long demonstrated and widely discussed the influence of SDM on future species distributions (e.g., Thuiller, 2003, 2004; Araujo *et al.*, 2005b, Lawler *et al.*, 2006; Pearson *et al.*, 2006), studies exploring a range of climate scenarios have more recently been undertaken, but comparisons between different GES are now commonly addressed (e.g., Peterson *et al.*, 2002; Thuiller, 2004; Thuiller *et al.*, 2005; Araujo *et al.*, 2006; Broennimann *et al.*, 2006; Mika *et al.*, 2008; Barbet-Massin *et al.*, 2009; Lawler *et al.*, 2009). Such comparisons between different GCM, however, are less frequent (but see Tuck *et al.*, 2006; Mika *et al.*, 2008; Barbet-Massin *et al.*, 2009; Bradley, 2009; Durner *et al.*, 2009; Lawler *et al.*, 2009) and as yet remain to be thoroughly considered (Beaumont *et al.*, 2008).

Very little attention has been focused on quantifying (i) the variability between the projections obtained from ensembles combining several sources of uncertainty (but see Hartley *et al.*, 2006), and (ii) the relative contribution to projections of each of these uncertainty components (but see Dormann *et al.*, 2008). As potential impacts of climate change on species distribution should be predicted with an assessment of their uncertainty, so that management and conservation decisions can be taken with full knowledge of the predictions' reliability (Hartley *et al.*, 2006), quantifying the effects of the different sources of uncertainty appears a crucial issue.

In this context, this study was designed to quantify the uncertainty in projections of the impact of climate change on 35 fish species in French streams. We focused on the variability in future projections of climate change impacts on fish species and assemblages due to four uncertainty components: initial dataset, SDM, GCM, and GES. After measuring the variability's extent, we evaluated the relative contribution of each of the four uncertainty components to this variability. Then, we tested whether the magnitude of variability between future projections of species distribution was related to

some species' biological attributes. Finally, the potential occurrence of spatial patterns in the distribution of the uncertainty at the scale of France was analysed.

## Materials and methods

### *Biotic and abiotic data*

**Fish data.** Fish data were provided by the Office National de l'Eau et des Milieux Aquatiques (ONEMA), which is the national fisheries organization in charge of the protection and conservation of freshwater ecosystems in France. A standard electrofishing protocol is conducted yearly during low-flow periods by the ONEMA to collect information on fish assemblages present in a large number of French stream sections (hereafter referred to as 'sites'). We extracted a set of 1110 sites within French boundaries from the ONEMA database. For statistical reasons, the current study was limited to the 35 most common stream fish species (i.e., occurring in >25 out of 1110 sites) among a total number of 55 collected species. Species presence-absence were used.

**Climate data.** Three variables, related to fish ecological requirements, were used to describe climate conditions: mean annual precipitation, mean annual air temperature (TAN, °C), and annual air temperature range (TAM, °C) derived from the difference between mean air temperature of the warmest month (MTW, °C) and mean air temperature of the coldest month (MTC, °C). These values were extracted from the CRU CL 2.0 (Climatic Research Unit Climatology version 2.0) dataset (New *et al.*, 2002) at a resolution of  $10' \times 10'$ , and then averaged for the period 1961–1990 to describe the current climate.

Future climate predictions were averaged for three 30-year periods ending in 2020, 2050, and 2080 (referred to as 2020, 2050, and 2080 scenarios, respectively). They were derived from three GCM, namely CGCM2 (Canadian Centre for Climate Modelling and Analysis), CSIRO2 (Australia's Commonwealth Scientific and Industrial Research Organization), and HadCM3 (Hadley Centre for Climate Prediction and Research's General Circulation Model). These three GCM were selected as they have been commonly used in recent studies dealing with the impacts of climate change on biodiversity (e.g., Thuiller, 2004; Araujo *et al.*, 2006; Tuck *et al.*, 2006; Mika *et al.*, 2008). For each GCM, four GES were examined: A1FI, A2, B1, and B2 (IPCC SRES, Nakicenovic & Swart, 2000). These GES were chosen to capture a range of greenhouse gas emission levels predicted for the 21st century. A1FI was the most pessimistic GES, A2 and B2 were intermediate and B1

was the most moderate one. For each of the 1110 sites, values of the three climatic variables studied were extracted for all GCM  $\times$  GES combinations, thus providing 12 different projections of future climate conditions for each time period.

**Environmental data.** In addition to climate descriptors, six environmental variables were used to describe the 1110 sites studied: surface area of the drainage basin above the sampling site (SDB, km<sup>2</sup>), distance from the headwater source (DS, km), mean stream width (WID, m), mean water depth (DEP, m), river slope (SLO, ‰), and elevation (ELE, m). Five of these six variables were grouped into two synthetic descriptors: (i) a longitudinal gradient *G* derived from the first axis of a principal component analysis (PCA) on DS and SDB describing the position of the sites along the upstream–downstream gradient; and (ii) a local water velocity index *V* summarizing WID, DEP and SLO. Generalized additive models were then fitted between each of the three environmental variables (i.e., *G*, *V*, ELE) and climate because of strong correlations. Residuals from these three models were collected and used as individual predictors independent of climate (see Buisson *et al.*, 2008, for details).

### *Species biological attributes*

For each species, five biological attributes were evaluated from fish, climate, and environmental datasets. Firstly, species prevalence was calculated as the number of sites where the species occurred relative to the total number of sites (1110). Secondly, species latitudinal range was described as the difference between the average latitude of the 10% northernmost and southernmost sites where species occurred in the fish dataset. Then, to define species thermal range, a PCA was conducted on the three thermal variables TAN, MTC, and MTW. The first axis of this PCA, accounting for 90% of the total variability, was kept as a synthetic variable describing thermal conditions. Species thermal range was calculated as the difference between the average of the 10% highest and lowest values along this axis where species was observed. Finally, both longitudinal gradient (hereafter called 'stream gradient') range and elevation range were described as the differences between the average position along the stream gradient/elevation of the 10% most upstream/highest and downstream/lowest sites where species occurred, respectively.

### *Ensemble forecasting of species distribution modelling*

Fish species presence-absence was related individually to the six environmental and climate variables using seven different statistical methods based on

presence-absence input data, which are widely applied to model species distribution (see Heikkinen *et al.*, 2006, for a review; Thuiller *et al.*, 2009). These seven SDM included three regression methods [generalized linear models (GLM), generalized additive models (GAM), multivariate adaptive regression splines], two machine-learning methods [random forests (RF), boosted regression trees (BRT)] and two classification methods [linear discriminant analysis, classification tree analysis]. Depending on the statistical methods, different modelling parameters have been selected (e.g., second-order polynomial in GLM, number of smoothing splines in GAM, number of trees to fit in BRT or RF).

For each of the 35 species, the seven SDM were built using a subset of data containing 70% of the sites selected at random. The remaining 30% of the data was used to evaluate the current predictive performance of the models. This split-sample procedure was repeated 100 times, and thus, 700 different statistical models were calibrated for each species.

Then, current predictions of probability of occurrence of each species were calculated from the calibrated models for the 1110 sites. They were converted into presence-absence using a threshold maximizing the sum of two measures: sensitivity (i.e., the percentage of presence correctly predicted), and specificity (i.e., the percentage of absence correctly predicted) (Fielding & Bell, 1997).

Calibrated models were then used to generate projections of future potential distributions under each 12 GCM  $\times$  GES combinations for 2020, 2050, and 2080 scenarios. These future probabilities of occurrence were transformed into presence-absence values by using the same threshold value as for current predictions.

This procedure led to a full factorial design crossing 100 calibration datasets (Data), seven statistical methods (SDM), three GCM, and four GES, thus resulting in 8400 different projections of future distribution for each species and each of the three time periods. This projections ensemble represented possible states of species distribution in the future.

#### *Assessment of the potential impacts of climate change*

For each of the 8400 projections, a species range change (SRC) was calculated for each of the 35 species as the difference between the number of sites gained by the species (i.e., sites where the species could be present in the future but where it is absent at the moment) and the number of sites lost (i.e., sites where the species could be absent in the future but where it is currently present) relative to the number of sites currently occupied (e.g., Erasmus *et al.*, 2002; Thuiller *et al.*, 2005; Broennimann *et al.*, 2006). In this paper, we assumed that species had unlimited dispersal capacity.

To describe local fish assemblages, the projections for the 35 species were combined for current and future periods, respectively. A species turnover rate was measured at each site for each 8400 projections:

$$\text{Species Turnover} = 100 \times \frac{\text{SG} + \text{SL}}{\text{SR} + \text{SG}}, \quad (1)$$

where SG is the number of species predicted to appear in a site (species gain), SL the number of species predicted to no longer be present in the future (species loss), and SR the current species richness (Peterson *et al.*, 2002). Species turnover indicated the amount of change in the fish assemblages' composition (see Buisson *et al.*, 2008, for details).

#### *Evaluating the relative contribution of uncertainty components in projections*

The 8400 values of SRC (for each species separately) and turnover (for each site separately) for 2020, 2050, and 2080 were related to the four uncertainty factors (Data, SDM, GCM, and GES) using a generalized linear model. Uncertainty factors were considered as categorical predictors coded in 100, seven, three, and four modalities, respectively. SRC were normally distributed and a Gaussian distribution of errors was thus assumed. As turnover rates ranged between 0 and 1, a binomial distribution of errors and a logistic link function were applied.

Then, we partitioned out the variability in projections due to each uncertainty component using the ratio between the deviance explained by one factor and the null deviance. The proportion of deviance explained by each uncertainty component was measured separately for each species, each site, and each time period.

#### *Measuring the uncertainty in future projections*

The variability of both SRC and turnover, between projections, was measured by calculating the standard deviation across the 8400 projections for each species and site, respectively. We also examined the variability of SRC and turnover due to each uncertainty component by calculating the standard deviation across the modalities of one factor after averaging the projections resulting from all the modalities of the three other factors.

#### *Linking uncertainty in projections of species range change to species biological attributes*

Closely related species often share many similar characteristics compared with more distantly related species. Therefore, species cannot be considered as independent points in statistical analyses comparing multiple species (Paradis & Claude, 2002). We thus tested whether species

biological attributes significantly influenced the variability in projections of SRC after accounting for phylogenetic relatedness among fish species.

The first step was to build the phylogeny of the 35 studied fish species. We used molecular data to reconstruct phylogenetic relationships based on three mitochondrial genes (cytochrome *b*, cytochrome oxidase I, and ribosomal 16S subunit). Sequence data were obtained from GenBank and phylogeny was reconstructed using the Bayesian method under the TVM + I + G substitution model. We implemented the phylogeny estimation with MRBAYES and PAUP softwares (see Supporting Information Appendix S1).

Second, we applied generalized estimation equations (GEEs) as implemented in the *ape* package (Paradis *et al.*, 2004) in the R statistical environment (R Development Core Team, 2007, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria) to relate the variability in projections of SRC to species biological attributes. This approach takes into account the phylogenetic relatedness among fish species by constructing a species-to-species correlation matrix derived from the built phylogenetic tree. GEEs also permit to include covariates in a model to control for the effects of other factors that are likely to have an impact on the dependent variable. As species prevalence was expected to be strongly related to the value of SRC, and thus to its standard deviation by a statistical artefact, it was included as a covariate in the GEEs. Therefore, we tested separately the effects of each species' biological attribute on the standard deviation of SRC while accounting for both the effect of prevalence and phylogenetic relatedness.

#### *Spatial patterns in uncertainty of fish assemblages' projections*

To localize geographical areas of high variability between projections, standard deviations of species turnover projections were mapped at the scale of France. Then, to highlight potential spatial patterns in the distribution of uncertainty, two matrices of site-pair distances were computed using normalized Euclidean distance, relating to (i) geographical distance and (ii) dissimilarities in standard deviation of species turnover between each pair of site. Mantel correlograms (Oden & Sokal, 1986; Sokal, 1986) were computed to describe the spatial structure of variability in turnover. The geographical distance matrix was divided into 20 distance classes using Sturge's rule to set the range of pairwise distances in each class (Legendre & Legendre, 1998). Monte Carlo permutations were used to test if the observed values of the Mantel-test statistic ( $r_M$ ) differed from those expected under the null hypothesis (i.e., no spatial arrangement). We evaluated the

significance of  $r_M$  values at each distance class by comparing it with the distribution of 10 000 random permutation values obtained under the null hypothesis (Manly, 1994) using a Bonferroni correction to correct for multiple significance tests.

#### **Results**

Both projected SRC and turnover were highly variable depending on the Data  $\times$  SDM  $\times$  GCM  $\times$  GES combinations (Fig. 1). Overall, SRC increased from  $19.4 \pm 12.7\%$  across the 8400 projections in 2020 to  $48.9 \pm 24.4\%$  in 2050 and  $51.8 \pm 33.7\%$  in 2080 (Fig. 1a). A similar increase was observed for turnover projections, which averaged  $26.9 \pm 7.3\%$  in 2020,  $44.4 \pm 9.0\%$  in 2050, and  $50.3 \pm 9.8\%$  in 2080 (Fig. 1b). SRC projections varied more in 2050 and 2080 than in 2020 whereas variability in turnover projections was more stable over time.

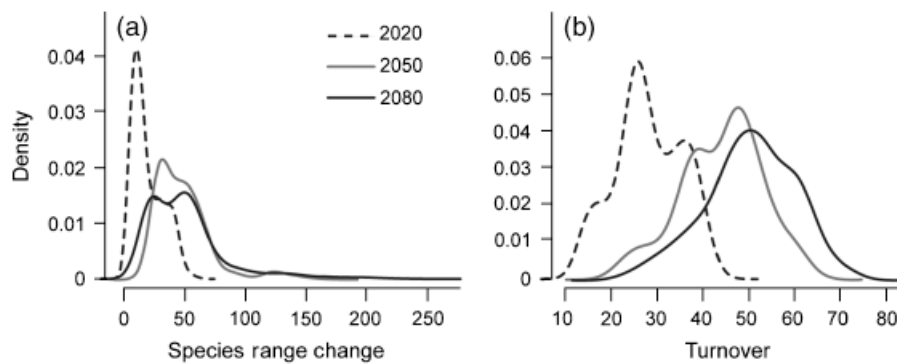
#### *Quantifying the relative contribution of uncertainty components to projections*

Overall, the four uncertainty components studied explained an important part of the total variability for both SRC and turnover projections whatever the time period. The proportion of explained deviance decreased from 70.5% in 2020 to 58.1% in 2080 for SRC projections, and it increased from 51.5% to 61.7% over time for turnover projections.

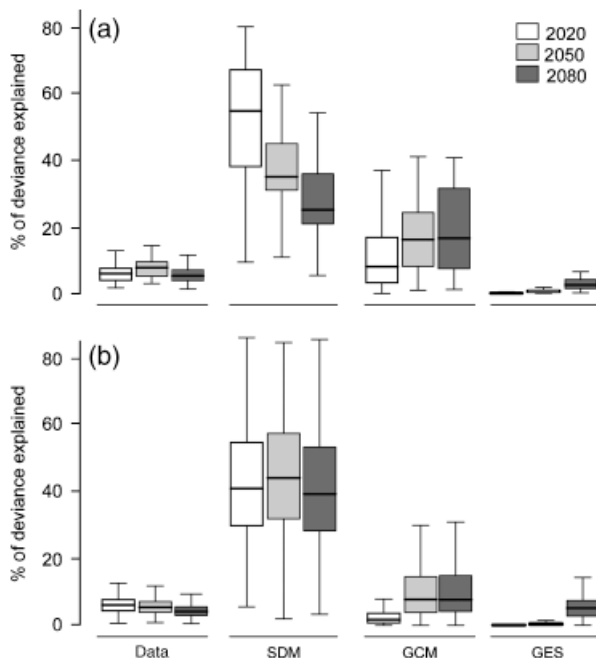
For both SRC and turnover values, SDM explained the largest part of variability in future projections for the three time periods studied (Fig. 2). On average, the relative contribution of SDM decreased from 51.2% in 2020 to 28.9% in 2080 for SRC projections, but remained stable at around 40–45% for turnover projections. Variability in projections of SRC and turnover was secondly explained by GCM. As the time horizon increased, the relative contribution of GCM also increased, ranging on average from 12.6% and 2.7% in 2020 to 19.2% and 10.6% in 2080 for projections of SRC and turnover, respectively (Fig. 2a and b). The proportion of variability in projections explained by Data and GES was small and rarely exceeded 10% for both SRC and turnover (Fig. 2). The effect of Data was relatively stable over time while GES explained a slightly increasing part of variability as the time horizon increased, especially for turnover projections (from 0.1% in 2020 to 5.4% in 2080).

#### *The effects of species biological attributes on uncertainty in projections of species range change*

After accounting for the effects of prevalence and phylogenetic relatedness, all four biological attributes exhibited a significant relationship with the total



**Fig. 1** Probability density functions for (a) species range change and (b) turnover projections based on the 8400 Data  $\times$  SDM  $\times$  GCM  $\times$  GES combinations for 2020 (dashed), 2050 (grey), and 2080 (black) scenarios. The mean value of SRC across the 35 species and the mean value of turnover across the 1110 sites were first calculated for each of the 8400 projections, and their distributions are shown. SDM, species distribution model; GCM, general circulation model; GES, gas emission scenario; SRC, species range change.



**Fig. 2** Percentage of deviance explained by each uncertainty component (Data, SDM, GCM and GES) using a GLM relating the 8400 (a) species range change (35 species), and (b) species turnover (1110 sites) projections to the four uncertainty components for the three time periods: 2020 (white), 2050 (light grey), and 2080 (dark grey). SDM, species distribution model; GCM, general circulation model; GES, gas emission scenario; GLM, generalized linear model.

variability in SRC (Table 1). Standard deviation of SRC was negatively related to species latitudinal and stream gradient ranges and positively related to thermal and elevation ranges. Thus, the whole projections ensemble

of SRC was more consistent for fish species with a large geographical extent, small thermal range or small elevation range.

When separating the variability in SRC due to each uncertainty component, similar results were found (Table 1) except for the standard deviations across GCM, which were positively related to the four environmental ranges studied.

#### *Spatial patterns in uncertainty of fish assemblages' projections*

The variability in projections of turnover depended on the assemblages' location (Fig. 3). Standard deviation for the whole ensemble of projections ranged from 6.6% to 38.4% and half of the assemblages had standard deviations  $>16.2\%$ . The highest values were observed in mountainous and south-western areas, whereas less variability was projected for many assemblages located in the north-eastern France (Fig. 3a). The Mantel correlogram exhibited positive autocorrelation for the smallest distant classes, and negative autocorrelation among the largest distant classes, hence revealing gradual changes in standard deviation values (Fig. 4a).

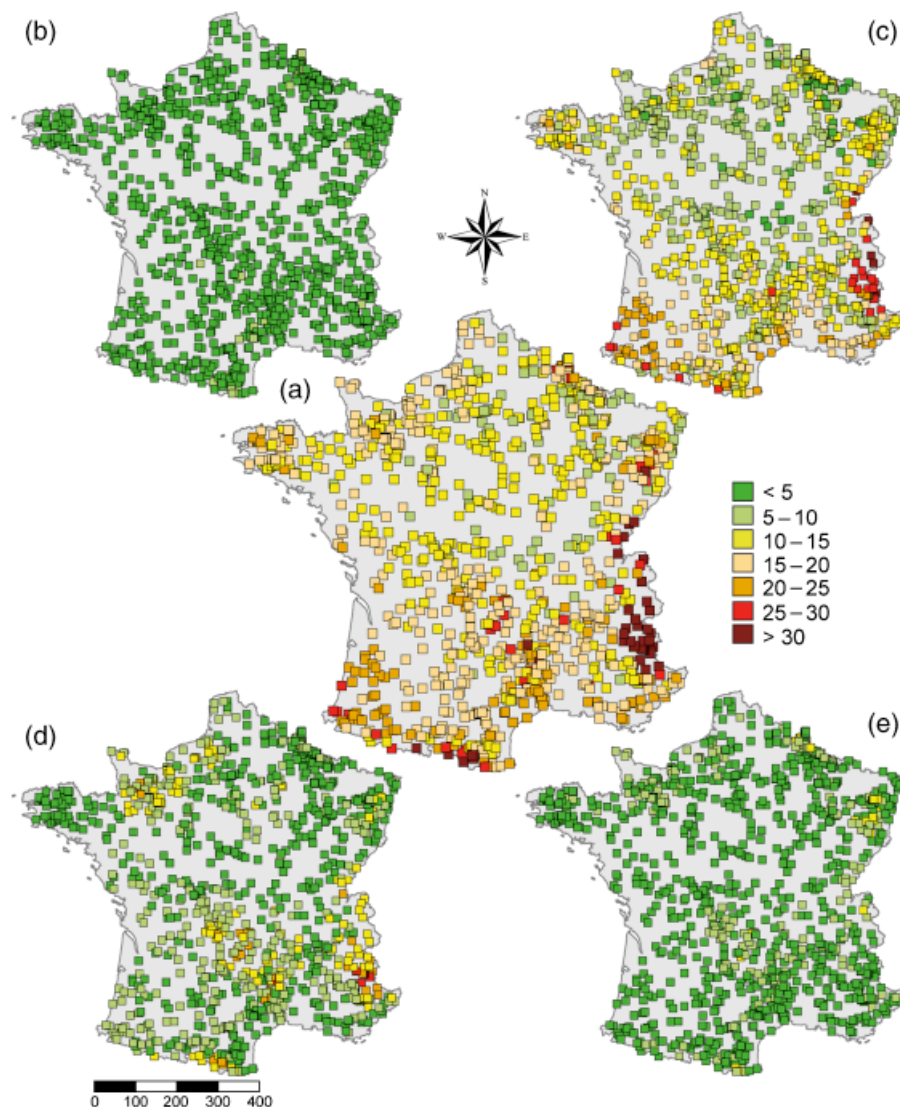
When separating the variability due to each uncertainty component, spatial patterns were contrasted depending on the considered component. Standard deviations of turnover values due to Data and GES were small (on average 3.3% and 4.1%, respectively) and consistent throughout France (Fig. 3b and e). Only the Mantel correlogram for GES indicated a slight spatial structure of variability (Fig. 4e). For variability due to SDM and GCM, although both Mantel correlograms highlighted a spatially structured variability in projections (Fig. 4c and d), divergent patterns were found. Standard deviation values due to SDM, globally

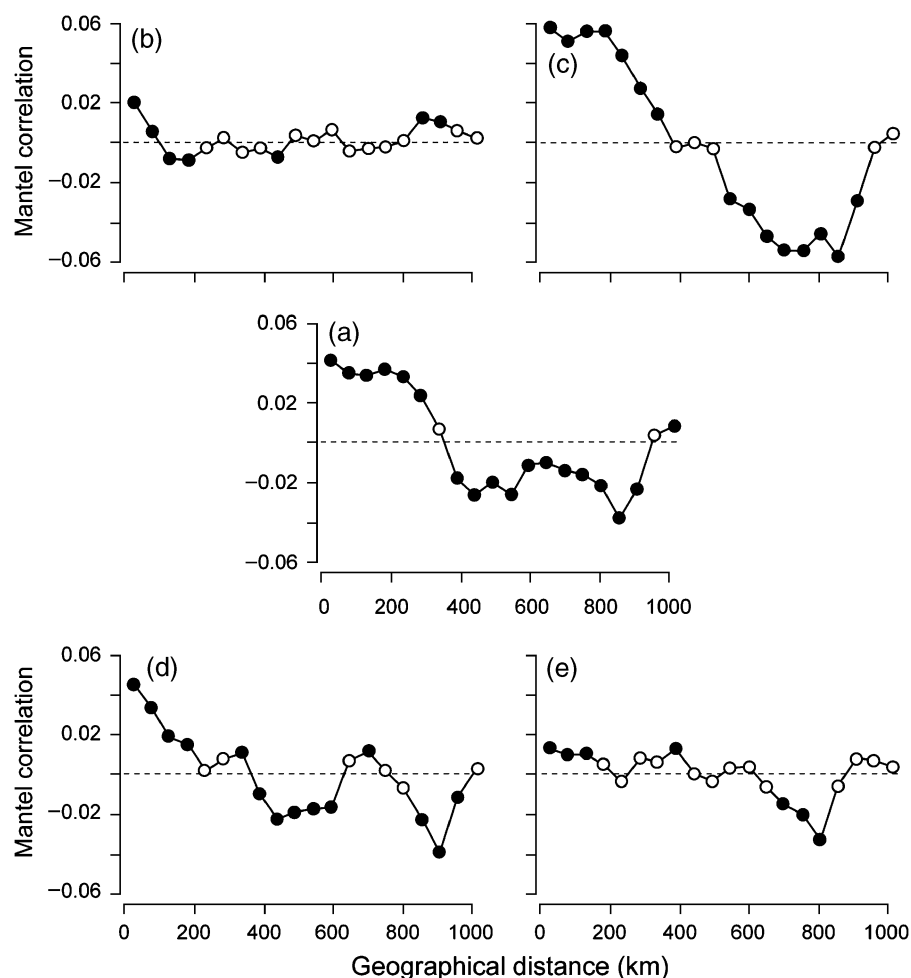
**Table 1** Relationships between the variability of species range change (standard deviations) projected for 2080, and species biological attributes according to generalized estimation equations (GEEs)

	Total	Uncertainty sources			
		Data	SDM	GCM	GES
Latitudinal range	−0.001***	−0.001***	−0.001***	0.001*	−0.001***
Thermal range	0.407***	0.668***	0.188**	0.455***	0.718***
Stream gradient range	−0.185**	−0.292**	−0.287***	0.327**	−0.299*
Elevation range	0.002***	0.003***	0.001***	0.001***	0.003***

SDM, species distribution model; GCM, general circulation model; GES, gas emission scenario.

Given are estimates of coefficients and significance: \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ .

**Fig. 3** Standard deviations of turnover values (%) in 2080 calculated across (a) 8400 projections, and (b) Data, (c) SDM, (d) GCM, (e) GES modalities projections. SDM, species distribution model; GCM, general circulation model; GES, gas emission scenario.



**Fig. 4** Mantel correlograms for spatial autocorrelation of standard deviations of turnover values in 2080 calculated across (a) 8400 projections, and (b) Data, (c) SDM, (d) GCM, (e) GES modalities projections. Dark circles indicate significant correlations (assessed using a Bonferroni correction) between site dissimilarity and geographical distance (upper class limit in km). SDM, species distribution model; GCM, general circulation model; GES, gas emission scenario.

high, showed a strong gradient pattern and were positively autocorrelated for assemblages as far as 350 km apart (Fig. 4c). The highest variability occurred below a straight line joining north-western France to south-eastern France (Fig. 3c). Standard deviation of turnover values due to GCM displayed a patchy spatial structure (Fig. 4d): patches of high variability were located in mountainous areas whereas the distribution of lowest values was more homogeneous (Fig. 3d).

## Discussion

In the present study, we evaluated the variability in projections of climate change impacts on fish species and assemblages. Large divergences in magnitude of species range change and turnover projections were highlighted, especially at long-term, arguing strongly for displaying the full range of projections in climate

change impact studies. These results thus strengthen the need for ensemble forecasting framework combining the outputs of several sources of uncertainty in order to enhance reliability of projected future impacts (Araujo & New, 2007). Advances in computing and the development of platforms or tools such as BIOMOD (Thuiller *et al.*, 2009) or NEURALENSEMBLES (O'Hanley, 2009) will undoubtedly facilitate such approaches.

### *What are the main drivers of uncertainty in future projections?*

Partitioning out the different sources of uncertainty might help in deciding where to focus future research to reduce variability in projections (Guisan *et al.*, 2006). Our results suggested that the choice of the statistical method may strongly determine values of projections because SDMs introduced the largest part of variability



in the short-, medium-, and long-term compared with the other three sources of uncertainty tested in this study. Between one-third and half of the total variability in future projections could be due to the statistical method. This result thus corroborates the recent study by Dormann *et al.* (2008) who found that the different modelling steps, especially the model type, caused far greater variation in projected change of the Great Grey Shrike distribution than differences between climate change scenarios. Their results thus suggested that the methodological uncertainty due to statistical techniques could hide the effects of using different future climate scenarios. Similar trends have been highlighted by other studies although they did not strictly quantify the relative contribution of each uncertainty component (Thuiller, 2004; Araujo *et al.*, 2006).

The important contribution of statistical method was in accordance with the large number of studies, which have compared projections of species distribution using a variety of modelling techniques and found contrasted patterns depending on the method selected (Thuiller, 2003; Lawler *et al.*, 2006; Pearson *et al.*, 2006; Roura-Pascual *et al.*, 2009). Although SDM are all based on a correlative approach, they use different assumptions, mathematical algorithms, and parameterizations. They may vary in how they model the shape, nature, and complexity of species' response, select predictor variables, weight variable contributions, or allow for interactions (Guisan & Zimmermann, 2000; Elith *et al.*, 2006; Austin, 2007). Species ecological niche may therefore be modelled with great discrepancies depending on the statistical method applied. Moreover, all techniques behave differently when extrapolating projections beyond the sampled values of data (Pearson *et al.*, 2006; Elith & Graham, 2009) and may also be a cause of variability in future projections.

In this study, GCM were another important driver of the uncertainty in projections. Unlike SDM, the use of multiple GCM has only recently been considered in climate change impact studies (e.g., Thuiller, 2004; Araujo *et al.*, 2006; Mika *et al.*, 2008; Durner *et al.*, 2009; Lawler *et al.*, 2009), whereas 23 GCM from 18 modelling centres are included in the fourth assessment report of the IPCC. These GCM are very complex systems derived from fundamental physical laws, then subjected to physical and mathematical approximations and able to simulate current climate conditions (Randall *et al.*, 2007). GCM basically differ in initial conditions and mathematical models from which they are built. In addition, at each step of the modelling process, variability is introduced by the way to represent Earth system process (e.g., clouds, water vapour), the number of climate feedback mechanisms operating, the spatial and vertical resolution or the internal climate variability

(Meehl *et al.*, 2007; Beaumont *et al.*, 2008). These climate models may consequently drive to contrasted forecasts of future climate change, and their impacts on species distribution may thus be highly variable (e.g., Tuck *et al.*, 2006; Mika *et al.*, 2008; Barbet-Massin *et al.*, 2009; Bradley, 2009; Durner *et al.*, 2009).

Here, it was also found that GCM introduced increasing variation in future projections as the time horizon increased reaching almost 20% of the total variability of species range change projections in 2080. Uncertainty due to GCM could therefore be as large as uncertainty due to SDM at the end of the 21st century. At a lesser extent, a similar increasing trend was highlighted for the emission scenarios that had no effect on the variability in projections for the first half of the 21st century while they could explain approximately 5% of the variability in turnover at the end of the century. These results thus suggest that the choice of emission scenarios has little effect when projecting future species distribution for the first few decades of the 21st century. This was not surprising as Stott & Kettleborough (2002) have found that there is remarkably little difference between climate predictions from a representative range of emission scenarios until 2040. The relevance of studies comparing future impacts of climate change on biodiversity under several emission scenarios in the short-term thus deserves to be questioned. More attention should rather be paid to the use of different climate models that mainly drive variations in climate projections in the first half of the 21st century. However, as both climate models and emission scenarios are expected to increase their divergences over the 21st century (Stott & Kettleborough, 2002; Meehl *et al.*, 2007), the inclusion of several GCM and GES is particularly relevant for long-term projections in order to sample a sufficient fraction of the uncertainty in future conditions (Beaumont *et al.*, 2008).

*For which species are projections of range change the most consistent?*

Species' ecological characteristics may affect the outputs of SDMs by influencing the quality of available data or by making it difficult to statistically capture the relationship between the species distribution and environmental conditions (McPherson & Jetz, 2007). Here, we focused on the variability in species range change projections as an indicator of the amount of reliability that could be placed in those projections. Outputs of the GEEs suggested that projections of range change were more consistent for species with a large geographical extent (i.e., large distribution along latitudinal or upstream–downstream gradients), or with small thermal or elevation ranges. Greater confidence

should thus be given to the projections of climate change future impacts for species with strict thermal or elevation requirements, which can be coarsely defined as 'specialist' species. On the contrary, projections for species with a large thermal tolerance or a large elevation tolerance or a small geographical extent should be treated with care for conservation purposes.

Recently, an extensive number of studies have investigated the relationship between the performances of SDM at predicting present-day distribution and species ecological characteristics (e.g., Segurado & Araujo, 2004; Luoto *et al.*, 2005; Seoane *et al.*, 2005; Elith *et al.*, 2006; Guisan *et al.*, 2007; McPherson & Jetz, 2007; Evangelista *et al.*, 2008; Poyry *et al.*, 2008; Marmion *et al.*, 2009a). Overall, it appeared that species with limited geographical extent or specialist species with strict ecological requirements yielded models with higher accuracy than generalist species or species with a wide geographical range. Reasons mentioned to explain such findings are diverse, but are mainly based on the ecological niche complexity. It would actually be easier to discriminate suitable from unsuitable habitat for habitat specialist, and thus to best describe their current ecological niche. In the context of climate change, evaluating which species distributions will be projected accurately in the future is more difficult as there are no possible comparisons given that these events have not yet occurred (Araujo *et al.*, 2005a). To date, a few studies have attempted to highlight which species' biological attributes could explain the sensitivity of species to climate modifications (e.g., Thuiller *et al.*, 2005; Broennimann *et al.*, 2006; Poyry *et al.*, 2009). However, no studies have focused on the species' ecological characteristics, which could influence the uncertainty in future projections. A crucial issue would therefore be to test whether species, whose current distribution is predicted with the greatest accuracy, would also be species projected with the greatest consistency in the future. Given our promising finding that specialist species would be more consistently projected than generalists, this hypothesis deserves to be thoroughly tested.

*Where are projections of species turnover the most consistent?*

To date, studies that have used multiple statistical models or climate scenarios to predict habitat suitability or climate change impacts have identified areas of consistency or areas of divergences by comparing maps of projections (e.g., Elith *et al.*, 2006; Crossman & Bass, 2008; Bradley, 2009). However, maps of uncertainty (e.g., confidence intervals around predictions, standard errors) have rarely been provided (Roura-Pascual *et al.*, 2009), whereas they could allow management decisions

to be taken with greater certainty than would be possible from mean predictions alone (Hartley *et al.*, 2006). For instance, it could be informative for a nature reserve manager to learn that 90% of the models project that one particular species will disappear from the reserve boundaries under climate change, whereas only 50% of the models predict the extinction of a second species. Confidence given to these forecasts will probably drive to different priorities in terms of conservation. Moreover, maps of uncertainty could allow the classification of geographical areas as 'certain' or 'uncertain' for predicting the potential impacts of climate change, and thus help to identify areas where every effort should be focused to achieve a better understanding of the future impacts of climate change.

Here, we statistically demonstrated that variability in turnover projections was not randomly distributed and even that uncertainty due to SDM and GCM was spatially structured. We were not surprised that no spatial structure was highlighted in variability due to Data and GES, because these factors entered little variation in projections. Overall, maps of standard deviations indicated that projections of species turnover were not consistent for assemblages located below a straight line joining north-western France to south-eastern France. Within this large area, some patches of highest variability were projected especially in mountainous areas and in south-western France. Understanding the causes of such spatial patterns in uncertainty between projections could be very helpful for future research. For instance, although many features of the simulated climate change are qualitatively consistent among climate models for the majority of Europe, substantial divergences actually remain especially on the magnitude and geographical details of precipitation change (Christensen *et al.*, 2007). Moreover, Giorgi *et al.* (2001) showed that a number of consistent patterns of regional climate change across GCM and scenarios are now emerging, but cases of inconsistency occur in some geographical areas of the world. Thus, it could be interesting to consider the geographical variability in climate data between GCM and GES and its potential link with the uncertainty in turnover projections. Another point would be to look in more details at the composition of assemblages located in the areas of high uncertainty since they probably include species whose future distribution is predicted inconsistently. Errors of prediction of the current species distribution have been shown to be spatially structured (Pineda & Lobo, 2009). Thus, combining the projections of several individual species into turnover values could provide spatially structured patterns of turnover variability. Since the agreement between predictions is expected to be lower for intermediate values of probability of occurrence (i.e.,

at the edge of species distribution area), we could expect that the geographical areas of high uncertainty could also correspond to the environmental range margins of several species. Nevertheless, while these spatial patterns will not be clarified, projections of climate change future impacts in the 'uncertain' geographical areas should be considered with care and further studies conducted to reduce the risk of those projected impacts being misused.

## Conclusion

Given the rate at which studies using SDM expand, and the importance of their potential implications in terms of conservation and management of biodiversity, a better knowledge of the uncertainty associated with the outputs of these models will be one of the main challenges of forthcoming research. Proliferation of distribution modelling studies will be fruitless while uncertainty arising from different steps of the modelling process is not strictly assessed and quantified. Ensemble forecasting approaches appear to be a promising tool to capture a wide range of the variability generated by the modelling procedure. Nevertheless, fundamental research aiming at understanding why outputs of SDM differ should be conducted simultaneously (Elith & Graham, 2009).

To conclude, we recommend paying more attention to the following points when evaluating the impacts of climate change on biodiversity: (i) using several statistical methods in an ensemble forecasting framework, (ii) using climate projections from different GCM to assess impacts in the short- and long-term, (iii) using different climate scenarios only for impacts in the long-term, (iv) always providing maps of uncertainty in conjunction with maps of projected impacts, and (v) taking into account species biological attributes when combining responses of individual species to assess impacts on assemblages. Following these recommendations, efforts could be made to achieve a more realistic understanding of the future impacts of climate change on biodiversity, allowing management and conservation decisions to be taken with awareness of the inherent uncertainty in those impacts.

## Acknowledgements

This research was part of the EU project Euro-Limpacs (contract number GOEC-CT-2003-505540). WT received support from the FP6 EU funded Ecochange (challenges in assessing and forecasting biodiversity and ecosystem changes in Europe, number 066866 GOCE) project. We are indebted to the Office National de l'Eau et des Milieux Aquatiques (ONEMA) for providing fish data. We are grateful to Pierre-Jean Malé and Emilie Lecompte, for help with phylogenetic analyses. Elanor France gave linguistic advice.

## References

- Araujo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- Araujo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.
- Araujo MB, Pearson RG, Thuiller W, Erhard M (2005a) Validation of species–climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.
- Araujo MB, Thuiller W, Pearson RG (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography*, **33**, 1712–1728.
- Araujo MB, Whittaker RJ, Ladle RJ, Erhard M (2005b) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, **14**, 529–538.
- Austin M (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.
- Barbet-Massin M, Walther BA, Thuiller W, Rahbek C, Jiguet F (2009) Potential impacts of climate change on the winter distribution of Afro-Palaearctic migrant passerines. *Biology Letters*, **5**, 248–251.
- Barry S, Elith J (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.
- Beaumont LJ, Hughes L, Pitman AJ (2008) Why is the choice of future climate scenarios for species distribution modelling important? *Ecology Letters*, **11**, 1135–1146.
- Bomhard B, Richardson DM, Donaldson JS *et al.* (2005) Potential impacts of future land use and climate change on the Red List status of the Proteaceae in the Cape Floristic Region, South Africa. *Global Change Biology*, **11**, 1452–1468.
- Bradley BA (2009) Regional analysis of the impacts of climate change on cheatgrass invasion shows potential risk and opportunity. *Global Change Biology*, **15**, 196–208.
- Broennimann O, Thuiller W, Hughes G, Midgley GF, Alkemade JMR, Guisan A (2006) Do geographic distribution, niche property and life form explain plants' vulnerability to global change? *Global Change Biology*, **12**, 1079–1093.
- Buisson L, Thuiller W, Lek S, Lim P, Grenouillet G (2008) Climate change hastens the turnover of stream fish assemblages. *Global Change Biology*, **14**, 2232–2248.
- Christensen JH, Hewitson B, Busuioic A *et al.* (2007) Regional climate projections. In: *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL), pp. 847–940. Cambridge University Press, Cambridge.
- Crossman ND, Bass DA (2008) Application of common predictive habitat techniques for post-border weed risk management. *Diversity and Distributions*, **14**, 213–224.
- Dormann CF, Purschke O, Marquez JRG, Lautenbach S, Schroder B (2008) Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. *Ecology*, **89**, 3371–3386.
- Durner GM, Douglas DC, Nielson RM *et al.* (2009) Predicting 21st-century polar bear habitat distribution from global climate models. *Ecological Monographs*, **79**, 25–58.

- Elith J, Graham CH (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, Graham CH, Anderson RP *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Erasmus BFN, Van Jaarsveld AS, Chown SL, Kshatriya M, Wessels KJ (2002) Vulnerability of South African animal taxa to climate change. *Global Change Biology*, **8**, 679–693.
- Evangelista PH, Kumar S, Stohlgren TJ, Jamevich CS, Crall AW, Norman JB, Barnett DT (2008) Modelling invasion for a habitat generalist and a specialist plant species. *Diversity and Distributions*, **14**, 808–817.
- Ficetola GF, Thuiller W, Miaud C (2007) Prediction and validation of the potential global distribution of a problematic alien invasive species – the American bullfrog. *Diversity and Distributions*, **13**, 476–485.
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Giorgi F, Whetton PH, Jones RG *et al.* (2001) Emerging patterns of simulated regional climatic changes for the 21<sup>st</sup> century due to anthropogenic forcings. *Geophysical Research Letters*, **28**, 3317–3320.
- Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, Loiselle BA (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**, 239–247.
- Guisan A, Lehmann A, Ferrier S, Austin M, Overton JMC, Aspinall R, Hastie T (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, **43**, 386–392.
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT (2007) What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615–630.
- Hartley S, Harris R, Lester PJ (2006) Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters*, **9**, 1068–1079.
- Heikkinen RK, Luoto M, Araujo MB, Virkkala R, Thuiller W, Sykes MT (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 751–777.
- Heikkinen RK, Luoto M, Kuussaari M, Toivonen T (2007) Modelling the spatial distribution of a threatened butterfly: impacts of scale and statistical technique. *Landscape and Urban Planning*, **79**, 347–357.
- Lawler JJ, Shafer SL, White D, Kareiva P, Maurer EP, Blaustein AR, Bartlein PJ (2009) Projected climate-induced faunal change in the Western Hemisphere. *Ecology*, **90**, 588–597.
- Lawler JJ, White D, Neilson RP, Blaustein AT (2006) Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biology*, **12**, 1568–1584.
- Legendre P, Legendre L (1998) *Numerical ecology*. Elsevier, Amsterdam.
- Luoto M, Poyry J, Heikkinen RK, Saarinen K (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography*, **14**, 575–584.
- Manly BFJ (1994) *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London.
- Marmion M, Luoto M, Heikkinen RK, Thuiller W (2009a) The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecological Modelling*, in press.
- Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009b) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.
- McPherson JM, Jetz W (2007) Effects of species' ecology on the accuracy of distribution models. *Ecography*, **30**, 135–151.
- Meeth GA, Stocker TF, Collins WD *et al.* (2007) Global climate projections. In: *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL), pp. 747–846. Cambridge University Press, Cambridge and New York.
- Mika AM, Weiss RM, Olfert O, Hallett RH, Newman JA (2008) Will climate change be beneficial or detrimental to the invasive swede midge in North America? Contrasting predictions using climate projections from different general circulation models. *Global Change Biology*, **14**, 1721–1733.
- Nakicenovic N, Swart R (2000) *Special Report on Emissions Scenarios. Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge.
- New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Climate Research*, **21**, 1–25.
- Oden NL, Sokal RR (1986) Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Systematic Zoology*, **35**, 608–617.
- O'Hanley JR (2009) NeuralEnsembles: a neural network based ensemble forecasting program for habitat and bioclimatic suitability analysis. *Ecography*, **32**, 89–93.
- Paradis E, Claude J (2002) Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology*, **218**, 175–185.
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Pearson RG, Thuiller W, Araujo MB *et al.* (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.
- Peterson A, Ortega-Huerta MA, Bartley J, Sanchez-Cordero V, Soberon J, Buddemeier RH, Stockwell DRB (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature*, **416**, 626–629.
- Pineda E, Lobo JM (2009) Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **79**, 182–190.

- Poyry J, Luoto M, Heikkinen RK, Kuussaari M, Saarinen K (2009) Species traits explain recent range shifts of Finnish butterflies. *Global Change Biology*, **15**, 732–743.
- Poyry J, Luoto M, Heikkinen RK, Saarinen K (2008) Species traits are associated with the quality of bioclimatic models. *Global Ecology and Biogeography*, **17**, 403–414.
- Randall DA, Wood RA, Bony S *et al.* (2007) Climate models and their evaluation. In: *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL), pp. 589–662. Cambridge University Press, Cambridge.
- Roura-Pascual N, Brotons L, Peterson AT, Thuiller W (2009) Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions*, **11**, 1017–1031.
- Segurado P, Araujo MB (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555–1568.
- Seoane J, Carrascal LM, Alonso CL, Palomino D (2005) Species-specific traits associated to prediction errors in bird habitat suitability modelling. *Ecological Modelling*, **185**, 299–308.
- Sharma S, Jackson DA (2008) Predicting smallmouth bass (*Micropterus dolomieu*) occurrence across North America under climate change: a comparison of statistical approaches. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 471–481.
- Sokal RR (1986) Spatial data analysis and historical processes. In: *Data analysis and informatics* (eds Diday E, Escoufier Y, Lebart L, Pages JP, Schektman Y, Tomassone R), North-Holland, Amsterdam.
- Stott PA, Kettleborough JA (2002) Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, **416**, 723–726.
- Thuiller W (2003) BIOMOD – Optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, **9**, 1353–1362.
- Thuiller W (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2020–2027.
- Thuiller W, Lafourcade B, Engler R, Araujo MB (2009) BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.
- Thuiller W, Lavorel S, Araujo MB, Sykes MT, Prentice IC (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8245–8250.
- Tuck G, Glendining MJ, Smith P, House JI, Wattenbach M (2006) The potential distribution of bioenergy crops in Europe under present and future climate. *Biomass and Bioenergy*, **30**, 183–197.

### Supporting Information

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

**Appendix S1.** The phylogenetic tree used in the comparative analysis to calculate the species-to-species correlation matrix.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.