



# Ensemble modelling of species distribution: the effects of geographical and environmental ranges

Gael Grenouillet, Laetitia Buisson, Nicolas Casajus and Sovan Lek

G. Grenouillet (*gael.grenouillet@cict.fr*), L. Buisson and S. Lek, *Laboratoire Evolution et Diversité Biologique, UMR CNRS 5174, Univ. Paul Sabatier, 118 route de Narbonne, FR-31062 Toulouse cedex 9, France.* – N. Casajus, *Univ. du Québec à Rimouski, 300 allée des Ursulines, Rimouski, QC, G5L 3A1, Canada.*

The aim of this study was to analyse the effects of species geographical and environmental ranges on the predictive performances of species distribution models (SDMs). We explored the usefulness of ensemble modelling approaches and tested whether species attributes influenced the outcomes of such approaches. Eight SDMs were used to model the current distribution of 35 fish species at 1110 stream sections in France. We first quantified the consensus among the resulting set of predictions for each fish species. Next, we created an average model by taking the average of the individual model predictions and tested whether the average model improved the predictive performances of single SDMs. Lastly, we described the ranges of fish species along four gradients: latitudinal, thermal, stream gradient (i.e. upstream-downstream) and elevation. After accounting for the effects of phylogenetic relatedness and species prevalence, these four species attributes were related to the observed variations in both consensus among SDMs and predictive performances by using generalized estimation equations. Our results highlight the usefulness of ensemble approaches for identifying geographical areas of agreement among predictions. Although the geographical extent of species had no effect on the performances of SDMs, we demonstrated that more consensual and accurate predictions were obtained for species with low thermal and elevation ranges, validating the hypothesis that specialist species yield models with higher accuracy than generalist ones. We emphasized that significant improvements in the accuracy of SDMs can be achieved by using an average model. Furthermore, these improvements were higher for species with smaller ranges along the four gradients studied. The geographical extent and ranges of species along environmental gradients provide promising insights into our understanding of uncertainties in species distribution modelling.

Over the last decade, species distribution models (SDMs) have diversified and become widely used in both basic and applied ecology. Dozens of statistical methods are now available and applied routinely (Guisan and Zimmermann 2000, Elith et al. 2006), and an increasing number of studies have compared model performances and predictions across multiple methods (Segurado and Araújo 2004, Elith et al. 2006, Heikkinen et al. 2006, Pearson et al. 2006, Dormann et al. 2008, Roura-Pascual et al. 2009). The techniques implemented have been shown to vary considerably in both performance and spatial predictions of species distributions.

In view of this variability between predictions of SDMs, the emerging recommendation is to simultaneously apply several methods (ensemble modelling, Araújo et al. 2005, Araújo and New 2007) within a consensus modelling framework (Thuiller 2004, Marmion et al. 2009a). Such a modelling framework is attractive as it reduces the predictive uncertainty of single-models by combining their predictions. To date, most available studies have demonstrated that the accuracy of species distribution predictions

could be substantially improved by applying consensus methods (Araújo et al. 2005, Crossman and Bass 2008, Marmion et al. 2009a). Moreover, it has been emphasized that results derived from SDMs are not equally reliable for all species (Luoto et al. 2005) and that the best performing models are not always the same for different species (Segurado and Araújo 2004, Barbet-Massin et al. 2009). Recent literature has thus focused on the effects of the characteristics of species on model performances (see McPherson and Jetz (2007) for a review). Earlier studies have addressed the effects of geographical attributes of species, such as species prevalence (Manel et al. 2001), latitudinal range (Segurado and Araújo 2004), spatial autocorrelation (Boone and Krohn 1999), or rarity (Karl et al. 2000), on the performance of SDMs. The effects of species range along environmental gradients on the performance of SDMs have also been addressed (Brotons et al. 2004, Segurado and Araújo 2004). Although the way SDMs behave when modelling species with different ecological and geographical characteristics can depend on the modelling technique (Marmion et al. 2009b), the

overall message emerging from these studies is that the characteristics of species distribution patterns can significantly influence the behaviour and uncertainty of SDMs (Heikkinen et al. 2006). A general pattern is that species with small geographical extent and strict ecological requirements (i.e. habitat specialists) yield models with higher accuracy than those with larger areas of occupancy and that are habitat generalists (Kadmon et al. 2003, Segurado and Araújo 2004, Hernandez et al. 2006, Tsoar et al. 2007, Franklin et al. 2009).

Despite the growing literature on the ecological characteristics of species that affect the performance of SDMs, how different attributes of species distribution patterns affect the results of ensemble modelling and consensus approaches remains unexplored. There is, to our knowledge, no study addressing the issue of the relationship between species' attributes and the consensus among different modelling methods. If ensemble modelling and consensus approaches are expected to be increasingly used in conservation and management planning studies, thorough examinations of their performances for species with different ecological and geographical characteristics are necessary.

In this study, we modelled the current distribution of stream fish species using eight different SDMs. These eight single-models provided the ensemble of predictions, which contained the eight separate predicted distributions generated for 35 fish species at 1110 stream sections in France. For each fish species, we created an average model by taking the average of the individual model predictions. The main objectives of this study were 1) to test whether the average model improved the predictive performances of single SDMs, and 2) to relate the observed variations in both consensus among SDMs and predictive performances to species ranges along four environmental gradients: latitudinal, thermal, stream gradient (i.e. upstream-downstream) and elevation ranges. Using generalized estimating equations (GEE) to account for the effect of phylogenetic relatedness among species, we tested the hypothesis that restricted-range species are modelled more accurately, and we examined whether the predictive improvements achieved with an average model depended on the attributes of the species distribution patterns.

## Material and methods

### Fish data

Fish data were provided by the Office National de l'Eau et des Milieux Aquatiques (ONEMA), the national fisheries organization in charge of the protection and conservation of freshwater ecosystems in France. A standard electrofishing protocol was carried out during low-flow periods (May–October) to collect information about fish assemblages in a large number stream sections in France. From this database, we extracted a set of 1110 sites selected to be representative of the heterogeneity among French streams and to cover a broad array of environmental conditions (Fig. 1). The presence–absence of the 35 most common species was used in this study.

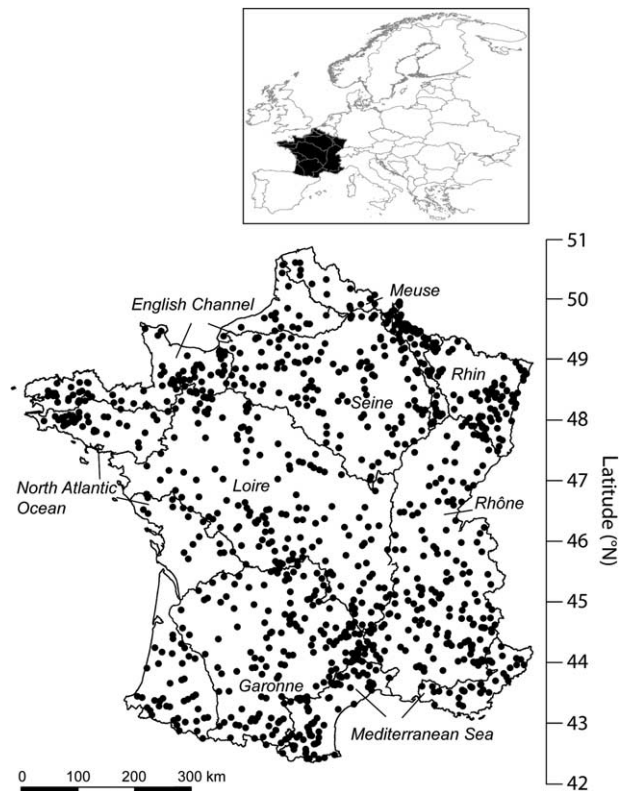


Figure 1. Geographical distribution of fish sampling sites in nine river units in France.

### Climate and environmental data

The CRU CL 2.0 (Climatic Research Unit Climatology 2.0 ver.) dataset (New et al. 2002) at a resolution of  $10' \times 10'$  was chosen to describe current climate. Three variables related to fish ecological requirements were extracted: mean annual precipitation (PAN, mm); mean annual air temperature (TAN, °C); and annual thermal range (TAM, °C), calculated as the difference between the mean air temperature of the warmest month (MTW, °C) and the mean air temperature of the coldest month (MTC, °C). Climate data were obtained from the average of the period 1961–1990. Values of the three climatic variables were extracted for all the grid cells containing at least one of the 1110 studied sites.

Six environmental descriptors were used to describe the 1110 studied sites: 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: a stream gradient (G) and a water velocity index (V). To eliminate the colinearity between DS and SDB which both describe the position of sites along the upstream-downstream gradient, we used a principal component analysis (PCA). The first axis of the PCA, accounting for 93.2% of the total variability, was kept as a synthetic variable describing the stream gradient. The water velocity index was derived from the Chezy formula following Oberdorff et al. (2001):

$$V = \log(\text{WID}) + \log(\text{DEP}) + \log(\text{SLO}) - \log(\text{WID} + 2\text{DEP})$$

To eliminate the colinearity between environmental (i.e. G, V, ELE) and climate, generalized additive models were fitted between each of the three environmental variables and the climate descriptors. Residuals from these three models were used afterwards as individual predictors (see Buisson et al. (2008) for details).

### Species range along environmental gradients

For each fish species, ranges along four environmental gradients were computed: latitudinal, thermal, stream gradient and elevation ranges.

The 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. To define the species thermal range, a PCA was first performed 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. Then, considering only sites where a given species was observed, the species thermal range was calculated as the difference between the average positions of the 10% highest and lowest values along this axis. Lastly, stream gradient and elevation ranges were described as the differences between the average G and ELE values of the 10% most extreme sites where species occurred, respectively.

### Ensemble modelling of species distributions

The presence-absence of species was modelled using eight different statistical methods (reviewed in Heikkinen et al. 2006). These methods included three regression methods (generalized linear models, GLM; generalized additive models, GAM; multivariate adaptive regression splines, MARS), three machine learning methods (artificial neural networks, ANN; random forest, RF; aggregated boosted trees, ABT) and two classification methods (factorial discriminant analysis, FDA; classification tree analysis, CTA). All the different models were calibrated using different packages of the R environment software (R Development Core Team 2009).

For each 35 species, the eight SDMs were built (step 1, Fig. 2) using a random subset of data containing 70% of the sites (i.e. 777 sites). The remaining 30% of the data (i.e. 333 sites) were used to evaluate the predictive performance of the models (step 2, Fig. 2). This split-sample procedure was repeated 100 times, leading to 800 different statistical models calibrated for each species. The predictions of these 800 models were converted into binary values using a threshold maximizing the sum of two measures: sensitivity, which measures the percentage of presence correctly predicted, and specificity, which measures the percentage of absence correctly predicted (Fielding and Bell 1997).

### Consensus among the single-SDM predictions

Following Thuiller (2004), we performed a PCA (step 3, Fig. 2) on the different species distributions predicted by

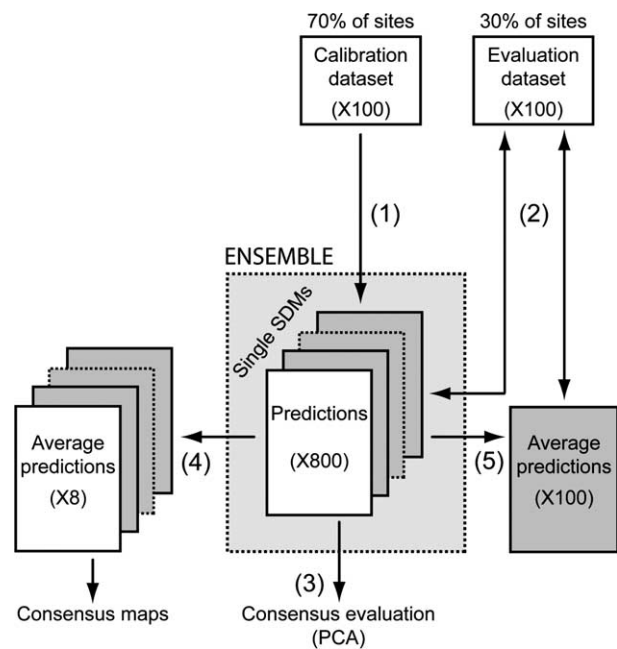


Figure 2. Schematic representation of the modelling design. See Methods for details.

the eight SDMs (i.e. 800 predictions). The first axis of this PCA (hereafter called “consensus axis”) is equivalent to a line that goes through the centroid of all sets of model predictions and minimizes the square of the distance of each set of predictions to that line (Araújo et al. 2005). This consensus axis captures consistent patterns in species distributions and reflects the general trend followed by the different predictions (Marmion et al. 2009a). For each fish species, we thus applied this approach to quantify the consensus among the single-SDM predictions. Consensus was evaluated by calculating the proportion of variance among predictions accounted for by the consensus axis. If all predictions were similar (i.e. no variability across predictions), the consensus axis would explain 100% of the variation. Thus, consensus decreased with increasing variability across the models. In addition to this consensus evaluation, we further analyzed the agreement among models by geographically mapping this agreement for a better visual assessment of spatial patterns (step 4, Fig. 2). To accomplish this, we first averaged the 100 predictions from each of the eight single SDMs (i.e. 100 iterations), and converted the mean probabilities of species occurrence as 0 or 1 using a threshold selected in the same way as described for single models (i.e. maximizing the sum of sensitivity and specificity). Then, we mapped the sum of the single-SDM predictions. This enabled us to visualize geographical areas of agreement for both absence (i.e. sum equals 0) and presence (i.e. sum equals 8), and areas of high disagreement (i.e. intermediate sum values) among SDMs.

### Average model

We combined the single-SDM predictions by computing the mean value of this ensemble (Marmion et al. 2009a). For each of the 100 iterations, the averaged predictions

across the eight statistical modelling techniques resulted in a single prediction at each site for each species (hereafter referred as the “average model”). As for the single models, the predictions of the average model were converted into 0–1 (step 5, Fig. 2).

## Model evaluation

The predictive performances of both single and average models were evaluated on validation datasets at the level of both individual species and assemblages. For species, the predictive performances were evaluated using the area under the curve (AUC) of a receiver operating characteristic plot (Fielding and Bell 1997). AUC values range from 0 to 1, where 0.5 indicates that the model’s discrimination is no better than random sorting, and 1 indicates that the model discriminates perfectly (Swets 1988). We then assessed the predictive accuracy of each model by measuring the percentage of both presence and absence correctly predicted for each species.

At the assemblage level, we analysed the similarity between the observed and predicted species composition at each site by calculating the inverse of the beta-Simpson’s index (Pineda and Lobo 2009). This similarity index (hereafter referred as  $\beta_{sim}$ ) focuses on compositional differences independently of species richness gradients (Koleff et al. 2003), with values ranging from 0 when two species groups have no species in common, to 1 when the two groups have similar composition.

We tested whether the average model increased the performance of single SDMs by comparing its performance (i.e. AUC, accuracy and  $\beta_{sim}$ ) with the values obtained from each single SDMs. Paired t-tests were used as the three modelling measures were normally distributed.

## Species range effects and phylogenetic relatedness

As closely related species often share similar characteristics, species cannot be considered as independent points in comparative analyses (Paradis and Claude 2002). We thus examined whether the geographical and environmental ranges of the species could significantly explain the observed differences in model performance after accounting for phylogenetic relatedness among species.

First, we built the phylogeny of the 35 species studied. We used molecular data to reconstruct phylogenetic relationships based on three mitochondrial genes (cytochrome b, cytochrome oxidase I and ribosomal 16S sub-unit). Sequence data were obtained from Genbank and consisted of 2234 base pairs (i.e. 1124, 651 and 459 for cytochrome b, cytochrome oxidase I and 16S sub-unit, respectively). We reconstructed phylogeny using the Bayesian method under the TVM+I+G substitution model and we implemented the phylogeny estimation with MrBayes and PAUP softwares (Supplementary material Appendix 1).

Then, we related observed differences in modelling accuracy among species to the species ranges by applying generalized estimating equations (GEEs) as implemented in the APE library (Paradis et al. 2004) in the R statistical environment (R Development Core Team 2009). This approach takes into account the phylogenetic relatedness

among fish species by constructing a species-to-species correlation matrix derived from the phylogenetic tree. Unlike the standard independent contrasts method, GEE explicitly incorporates the correlation matrix into the framework of a generalized linear model (GLM), without assuming a significant phylogenetic signal in the studied response. Moreover, the advantages of the GLM fitting is that 1) the response variable can follow several types of distribution belonging to the exponential distribution family (i.e. Gaussian, binomial, Poisson or gamma), and 2) the predictors can be continuous or categorical (Paradis and Claude 2002). As GEE also permits the simultaneous analysis of multiple predictors as covariates in the same model, covariates were included to control for the effects of other factors that are likely to have an impact on the response variable (Willis et al. 2008).

As the effects of the sizes of the species range on model performance have been shown to be largely artefactual and due to sample prevalence (McPherson et al. 2004), we applied GEEs while accounting for the effect of species prevalence (i.e. by including species prevalence as a covariate in the model). We evaluated the effect of each species range on three modelling indices: 1) the predictive performance (i.e. AUC value) from each single-SDM and from the average model, 2) the change in AUC value ( $\Delta_{AUC}$ ) measured by the difference between the AUC value from the average model and the mean value from the eight single-SDMs, and 3) the consensus among the single-SDM predictions.

## Results

The attributes of the species distribution patterns are shown in Supplementary material Appendix 2. Except for the range along the stream gradient, species ranges were all positively correlated (Pearson correlation,  $p < 0.01$ ) with prevalence. Among the cross-species correlations of species ranges, the highest correlation occurred between thermal and elevation ranges (Pearson correlation,  $r = 0.72$ ,  $p < 0.001$ ), while all other pair-wise correlations were weak (Pearson correlation,  $r < 0.4$ ).

### Single-SDM and average model performances

Overall, the eight single-SDMs showed good ability to predict observed distributions, with mean AUC values across all species ranging from 0.72 to 0.86 (mean = 0.82), and accuracy ranging from 0.73 to 0.78 (mean = 0.76). At the assemblage level, the similarity between the observed and predicted species composition (i.e.  $\beta_{sim}$ ) ranged from 0.73 to 0.84 (mean = 0.81). On average, accuracy and  $\beta_{sim}$  values were lowest for CTA (although a large variability between iterations was observed) and highest for RF (Fig. 3a). The relative ranking of SDMs according to their AUC values showed that RF more frequently yielded the models with the highest predictive performance (Fig. 3b). GAM, GLM and ABT came next, frequently among the top four techniques. ANN proved sensitive to iterations, while CART almost always performed the worst.

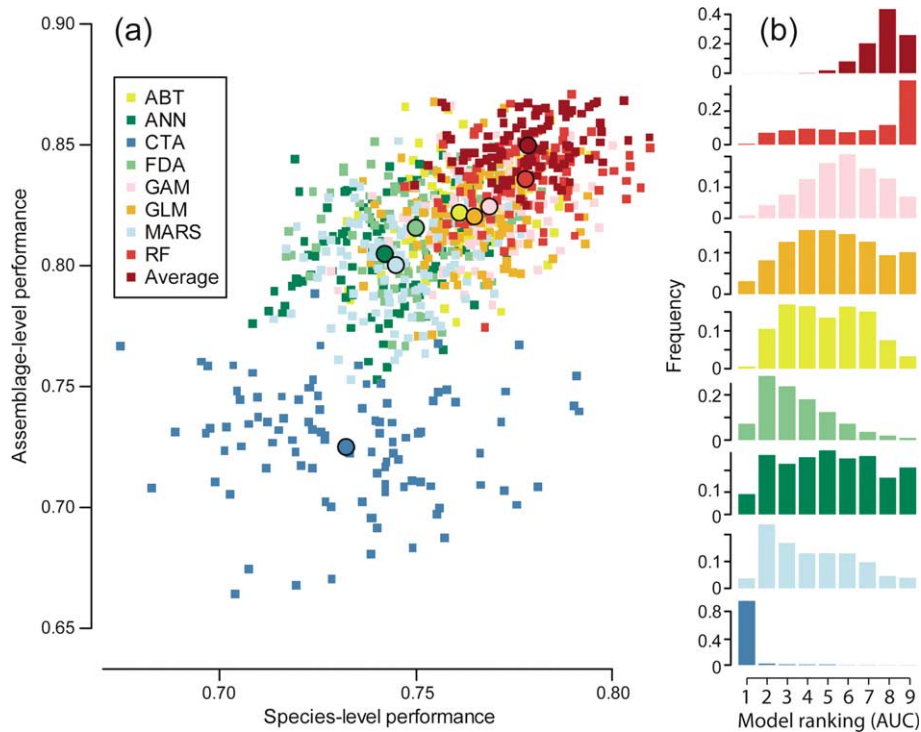


Figure 3. Modelling performances for the eight single-SDMs and the average model combining the predictions from the eight single-SDMs. (a) Species-level (i.e. accuracy values) and assemblage-level (i.e.  $\beta_{sim}$  values) performances. On the x-axis, each square corresponds to the mean accuracy of the models across all species and for one iteration (i.e. one validation dataset), circles indicate mean values across the 100 iterations for each single-SDM. (b) Model ranking according to AUC values across all species. For each iteration, models were ranked from 1 (i.e. lowest AUC) to 9 (i.e. highest AUC). See Methods for details.

Single-SDMs had consistently lower modelling performances than the average model. For accuracy values, only RF did not significantly (paired t-test,  $p = 0.67$ ) differ from the average model. For AUC and  $\beta_{sim}$  values, the average model significantly (paired t-test,  $p < 0.001$ ) outperformed all single-SDMs.

### Consensus among the whole predictions ensemble

The variability accounted for by the first axis of the consensus PCA varied from 69.7 to 94.1% (mean = 84.5%) of the total variability across predictions depending on the fish species considered.

To illustrate geographical areas of agreement between the modelling techniques, we produced maps summing the predictions from the eight individual SDMs for three species (Fig. 4). These three examples showed that the recorded distributions of fish species were well predicted by all single-SDMs, with areas consensually predicted as suitable or unsuitable by all modelling techniques. Visual inspection of these maps revealed that the most notable disagreement between predictions occurred at the edge of the recorded distributions of species.

### Species range effects

When accounting for the effect of phylogenetic relatedness among fish species, the predictive performance (i.e. AUC) of single-SDMs was unrelated to species prevalence, except for ABT which showed a significant decrease in AUC values

with increasing species prevalence (Table 1). According to a GEE analysis, species prevalence was also negatively related to the predictive improvement (i.e.  $\Delta_{AUC}$ ) achieved with the average model, and positively related to the consensus among SDM predictions.

After accounting for both the effects of prevalence and phylogenetic relatedness, the effects of species ranges on modelling performance were highly consistent among single-SDMs. Overall, results from the GEE analysis showed that fish species with small thermal and elevation ranges yielded models with higher AUC, while the latitudinal range had no significant effect. Species with small stream gradient ranges yielded models with higher AUC for CTA, GAM and RF. As for most single-SDMs, the predictive performance of the average model was negatively related to thermal, stream gradient and elevation ranges. The change in AUC value ( $\Delta_{AUC}$ ) was negatively related to all species ranges. Lastly, more consensual predictions among SDMs were observed for fish species with small thermal and elevation ranges, while latitudinal and stream gradient ranges did not influence the consensus among predictions.

### Discussion

Numerous studies have compared the predictive performances of different modelling methods (Segurado and Araújo 2004, Elith et al. 2006, Pearson et al. 2006) and the reasons for the observed differences have long been

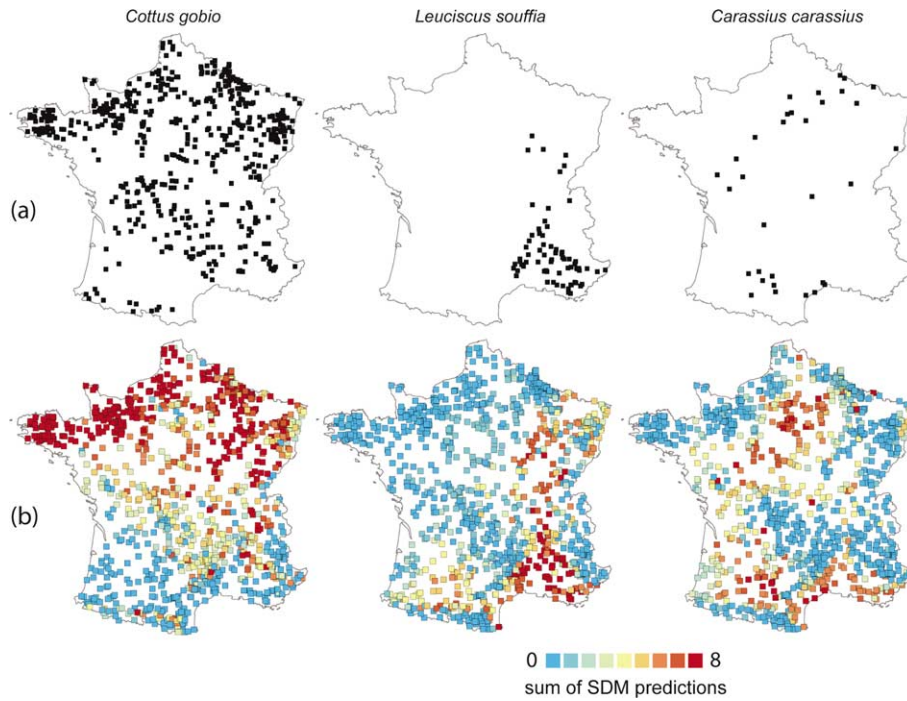


Figure 4. Maps showing three examples of (a) recorded distributions of fish species, and (b) sum of the eight single-SDM predictions of presence versus absence. Agreement among predictions is shown as a slope from blue (all single-SDMs predict absence) to red (all single-SDMs predict presence), while areas of high disagreement among SDMs appear in yellow.

discussed (Guisan and Zimmermann 2000). SDMs may vary in how they model the shape, nature and complexity of species responses (i.e. realized niche), select predictor variables, weight variable contributions, or allow for interactions (Guisan and Zimmermann 2000, Elith et al. 2006, Austin 2007). In this study, our main objective was not to address the differences between single-SDMs, but rather to focus on how to cope with such differences. We thus explored the usefulness of ensemble modelling approaches and we investigated whether species attributes influenced the accuracy of such approaches.

We defined an average model by computing the mean values of the predictions ensemble obtained from eight

statistical methods. Although other methods have been proposed to reduce uncertainty and improve accuracy from ensemble modelling approaches (i.e. Araújo et al. 2005), few ecological studies have compared several consensus methods (but see Marmion et al. 2009a). However, such approaches have been commonly used in other disciplines such as economics, management or meteorology (see Clemen (1989) for a review of the historical development of the combining forecasts literature). The use of a simple average has proven to perform as well as more sophisticated approaches (Armstrong 1989).

In this study, we confirmed that the average model significantly improved the predictive performance of all

Table 1. Relationships among the statistical model outputs and the attributes of fish species' distributions according to generalized estimating equations (GEE). GEE were applied for AUC values from single-SDMs and from the average model, change in AUC ( $\Delta_{AUC}$ ) and consensus among the whole predictions ensemble (%). PV = prevalence; LR = latitudinal range; TR = thermal range; GR = stream gradient range; ER = elevation range. See Methods for details.

Model output	PV	Species ranges			
		LR	TR	GR	ER
ABT AUC	-0.06	ns	-0.029	ns	-0.001
FDA AUC	ns	ns	-0.048	ns	-0.001
ANN AUC	ns	ns	-0.041	ns	-0.001
CTA AUC	ns	ns	-0.034	-0.028	ns
MARS AUC	ns	ns	-0.035	ns	-0.001
GLM AUC	ns	ns	-0.049	ns	-0.001
GAM AUC	ns	ns	-0.040	-0.032	-0.001
RF AUC	ns	ns	-0.031	-0.028	-0.001
Average model AUC	ns	ns	-0.038	-0.038	-0.001
$\Delta_{AUC}$	-0.049	-0.041	-0.081	-0.049	-0.084
Consensus	15.256	ns	-0.423	ns	-0.001

The estimates of coefficients significant after a Bonferroni correction (ns, not significant) are given, with significances based on a deviance change test. The effect of prevalence was tested after accounting for the effect of phylogenetic relatedness, while the effects of the four species ranges were tested after accounting for both the effects of species prevalence and phylogenetic relatedness.

single-SDMs, at both species and assemblage levels. While this technique is based on the outputs of all single-models, other combinative algorithms have been proposed to preselect the single-models based on certain predefined criteria (Marmion et al. 2009a). To date, few authors have addressed the careful selection of the choice of most appropriate models, but available studies have shown that methods based on the average function provided more robust (when significantly different) predictions than the other consensus methods (see Marmion et al. 2009a for an extensive evaluation). Previous ecological studies have explained that the good performance of consensual methods based on average function could be due to the low-pass filtering ability (i.e. “cleaning” effect) of the average function (Marmion et al. 2009a). Based on accuracy values, we found that only RF showed predictive performances equalling those of the average model. As recently noticed by Marmion et al. (2009a), this result is not surprising since RF incorporates the notion of ensemble modelling given that thousands of trees are produced and the predictions aggregated by averaging (Araújo and New 2007).

The percentage of consensus across predictions varied from 69.7 to 94.1% of the total variability depending on the fish species considered. As most other studies dealing with consensual predictions have not directly quantified this consensus (Marmion et al. 2009a, Roura-Pascual et al. 2009), comparisons with other taxonomic groups remain difficult. Nevertheless, some rare studies have described consensus responses in projecting species distributions under environmental changes (e.g. climate change: Thuiller 2004). For instance, Araújo et al. (2006) showed a high degree of consensus in European amphibian and reptile responses to climate change with 80% variability across the projections captured by a consensus PCA axis, whereas only 29.9% of this variability could be summarized for bird species in Great Britain (Araújo et al. 2005). Additional investigations are thus needed to better evaluate both within- and between-taxonomic group variability in consensus across predictions of current species distributions.

Although studies which have used multiple statistical models to predict habitat suitability have identified areas of consistency or divergence by comparing maps of projections (Elith et al. 2006, Crossman and Bass 2008), maps of uncertainty have rarely been provided (Roura-Pascual et al. 2009). Puschendorf et al. (2009) mapped the standard deviations of predictions emanating from different runs of the same statistical method and distinguished areas of higher confidence, but few studies have addressed spatial agreement between maps arising from various statistical methods (but see Brotons et al. 2004, Johnson and Gillingham 2005). In our case, we found that the agreement between SDMs was spatially structured for all 35 species, with the most notable disagreement between predictions occurring at the edge of the recorded distributions of species (data not shown). For mapping purposes, threshold (i.e. cut-off) values are used to derive probability values from SDMs into presence-absence data. This methodological approach is known to lead potentially to contrasting predicted distributions between modelling techniques (Thuiller 2004). The spatial pattern in disagreement between predictions observed at the edge of the distributions could be expected, since 1) the agreement

between presence-absence predictions coming from different statistical methods is expected to be lower for values of probability of occurrence close to the threshold value, and 2) such probability values are typically predicted at the edge of species distribution area (Arntzen 2006). Since we did not model the full extent of species distributions, it is worth noting that national boundaries did not correspond to the edge of species distribution area for 33 out of 35 fish species. Thus, the patterns observed along national boundaries were frequently not characterized by higher disagreement between predictions. To date, much attention has focused on the spatial pattern of the errors of prediction, which may result from both inadequacies in the algorithm or in the data used to do the modelling (Fielding and Bell 1997), and biological processes not included in the models (Fitzpatrick et al. 2007). These errors have also been shown to be spatially structured (Pineda and Lobo 2009), but relating these errors of prediction to the recorded distributions of species has been ignored. Our findings suggest that not only the consensus among SDM predictions, but also the spatial pattern in model agreement, should be addressed when using ensemble modelling of species distribution. We argue that a better understanding of the causes of such spatial patterns could be of considerable interest in forthcoming research as it may have important implications for both conservation planning and ecological management.

Whereas we found that species latitudinal range had no significant effect on model performance, more accurate predictions were obtained for species with low thermal and elevation ranges, thus validating the hypothesis that specialist species yield models with higher accuracy than generalist ones. It is worth noting that some caution should be taken when interpreting the effects of species ranges on model performances. Since our study area was limited by political borders, the measure of geographical (i.e. latitudinal) range reflected the extent of species distribution in the dataset but did not encompass the entire distribution of species. This limitation had probably little effects on the measure of the other environmental ranges because a broad array of environmental conditions was covered by the national database used in this study. Recently, the idea that model performance is not independent of the geographical or environmental distribution of species has been tested for species from a variety of major taxonomic groups, including plants, insects, amphibians, reptiles, birds and mammals (Kadmon et al. 2003, Segurado and Araújo 2004, Luoto et al. 2005, Guisan et al. 2007, McPherson and Jetz 2007, Franklin et al. 2009). These studies have shown that species with smaller areas of occupancy or extent of occurrence, and those found over a restricted range of environmental conditions (habitat specialists), were modelled more accurately than those with larger ranges (habitat generalists). Although several authors have hypothesized a positive relationship between geographical range size and habitat breadth (Pyron 1999), this pattern is not universal. Therefore, disentangling the effects of species range along both geographical and environmental gradients appears crucial when addressing the performances of species distribution models. Our results were consistent with previous studies demonstrating that the differences in the model performance could reflect the variation in the level of habitat specificity between the species studied (Luoto et al. 2005). Our results,

however, did not support the assumption that the accuracy of SDMs is greater for species with small geographical (i.e. latitudinal) ranges (Segurado and Araújo 2004, Hernandez et al. 2006). In these studies, three main explanations have been proposed to explain this assumption. Firstly, species with large geographical range are expected to encompass greater ecological heterogeneity, increasing the likelihood that more factors determine their distributions (Osborne and Suárez-Seoane 2002) or leading to noisier occurrence-environment relationships (McPherson and Jetz 2007). Secondly, since local ecological adaptation by sub-populations is more likely to occur for widely distributed species, differences in ecological preferences between sub-populations could lead to a decrease in model accuracy when the species is modelled as a whole over its entire geographical range (Hernandez et al. 2006). Thirdly, species described as widely distributed might simply not be limited by any of the considered predictive factors at the scale at which models are fitted (Brotons et al. 2004). Here, since the performance of the models was independent of the species latitudinal range after accounting for the effect of species prevalence, none of these three explanations could be verified. It has recently been pointed out that most previous studies have ignored such confounding effects and that their findings might be trivial and potentially reflect statistical artefacts rather than real range size effects (McPherson et al. 2004, Jiménez-Valverde et al. 2008).

More consensual predictions among SDMs were obtained for species with low thermal and elevation ranges. This finding, along with the result previously discussed, could reflect that species more accurately modelled by single-SDMs showed more consensual predictions among statistical methods. However, we suggest that this pattern is not simply trivial since the consensus among prediction increased with species prevalence, while the predictive performance of single-SDMs was unrelated to prevalence. Although some recent studies have investigated how variation in species distributions affects the performance of different modelling techniques, none has, to our knowledge, related the geographical and environmental distributions of species to the consensus among SDM predictions. The finding that specialist species are modelled with more consistency than generalists could have important implications for ensemble modelling of species distributions. In addition to being used to accommodate the increasing number of statistical methods, and thereby the variability between SDMs predictions (Araújo et al. 2005, Araújo and New 2007), such ensemble approaches could actually be very useful to predict the distribution of species with large environmental ranges. Further studies are clearly needed to generalize this finding, but our results caution against applying one single statistical method when modelling habitat generalist species.

Lastly, we emphasized that improvements in the accuracy of SDMs achieved by an average model were higher for species with smaller ranges along both geographical and environmental gradients. This finding is of considerable interest to practitioners, since species with narrow geographical ranges or high constraints in their habitat requirements are typically species of particular concern (e.g. endemic or endangered species) in conservation planning and biodiversity management. Using ensemble

modelling approaches for these species could thus be very helpful to understand better their distribution and to assess the potential impacts of environmental changes.

## Conclusion

Along with some other recent papers, this study strengthens the usefulness of ensemble modelling of species distribution for conservation and biodiversity management. While ensemble modelling has been proposed to reduce the uncertainties in various model predictions, we highlighted that these uncertainties can be related to the geographical extent and ranges of species along environmental gradients. We recommend that the application of a single modelling technique should be avoided especially for species with large environmental ranges. We also encourage researchers to investigate more thoroughly the agreement among predictions obtained from a variety of statistical methods. Although further research is needed to generalize our results, additional studies should now identify species traits linked to the geographical and environmental patterns of their distribution. This could be useful to highlight traits associated with the quality of species distribution models (Pöyry et al. 2008) and would provide promising insights into our understanding of uncertainties in species distribution modelling.

*Acknowledgements* – This research was part of the EU project Euro-Limpacs (contract number GOEC-CT-2003-505540). We are indebted to the Office National de l'Eau et des Milieux Aquatiques (ONEMA) for providing fish data and we thank the many fieldworkers who contributed to the fish records. We also thank Pierre-Jean Malé and Emilie Lecompte, who helped with the phylogenetic analyses, and John Woodley for the English.

## References

- Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – *Trends Ecol. Evol.* 22: 42–47.
- Araújo, M. B. et al. 2005. Reducing uncertainty in projections of extinction risk from climate change. – *Global Change Biol.* 14: 529–538.
- Araújo, M. B. et al. 2006. Climate warming and the decline of amphibians and reptiles in Europe. – *J. Biogeogr.* 33: 1712–1728.
- Armstrong, J. S. 1989. Combining forecasts: the end of the beginning or the beginning of the end? – *Int. J. Forecasting* 5: 585–588.
- Arntzen, J. W. 2006. From descriptive to predictive distribution models: a working example with Iberian amphibians and reptiles. – *Front. Zool.* 2006: 3–8.
- Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. – *Ecol. Model.* 200: 1–19.
- Barbet-Massin, M. et al. 2009. Potential impacts of climate change on the winter distribution of Afro-Palaearctic migrant passerines. – *Biol. Lett.* 5: 248–251.
- Boone, R. B. and Krohn, W. B. 1999. Modeling the occurrence of bird species: are the errors predictable? – *Ecol. Appl.* 9: 835–848.
- Brotons, L. et al. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. – *Ecography* 27: 437–448.



- Buisson, L. et al. 2008. Climate change hastens the turnover of stream fish assemblages. – *Global Change Biol.* 14: 2232–2248.
- Clemen, R. T. 1989. Combining forecasts: a review and annotated bibliography. – *Int. J. Forecasting* 5: 559–583.
- Crossman, N. D. and Bass, D. A. 2008. Application of common predictive habitat techniques for post-border weed risk management. – *Divers. Distrib.* 14: 213–224.
- Dormann, C. F. et al. 2008. Components of uncertainty in species distribution analysis: a case study of the great grey shrike. – *Ecology* 89: 3371–3386.
- Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.
- 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. 2007. The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? – *Global Ecol. Biogeogr.* 16: 24–33.
- Franklin, J. et al. 2009. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. – *Divers. Distrib.* 15: 167–177.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – *Ecol. Model.* 135: 147–186.
- Guisan, A. et al. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? – *Ecol. Monogr.* 77: 615–630.
- Heikkinen, R. K. et al. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. – *Prog. Phys. Geogr.* 30: 751–777.
- Hernandez, P. A. et al. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. – *Ecography* 29: 773–785.
- Jiménez-Valverde, A. et al. 2008. Not as good as they seem: the importance of concepts in species distribution modelling. – *Divers. Distrib.* 14: 885–890.
- Johnson, C. J. and Gillingham, M. P. 2005. An evaluation of mapped species distribution models used for conservation planning. – *Environ. Conserv.* 32: 1–12.
- Kadmon, R. et al. 2003. A systematic analysis of factors affecting the performance of climatic envelope models. – *Ecol. Appl.* 13: 853–867.
- Karl, J. W. et al. 2000. Sensitivity of species habitat-relationship model performance to factors of scale. – *Ecol. Appl.* 10: 1690–1705.
- Koleff, P. et al. 2003. Measuring beta diversity for presence–absence data. – *J. Anim. Ecol.* 72: 367–382.
- Luoto, M. et al. 2005. Uncertainty of bioclimatic envelope models based on the geographical distribution of species. – *Global Ecol. Biogeogr.* 14: 575–584.
- Manel, S. et al. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. – *J. Appl. Ecol.* 38: 921–931.
- Marmion, M. et al. 2009a. Evaluation of consensus methods in predictive species distribution modelling. – *Divers. Distrib.* 15: 59–69.
- Marmion, M. et al. 2009b. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. – *Ecol. Model.* 220: 3512–3520.
- McPherson, J. M. and Jetz, W. 2007. Effects of species' ecology on the accuracy of distribution models. – *Ecography* 30: 135–151.
- McPherson, J. M. et al. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? – *J. Appl. Ecol.* 41: 811–823.
- New, M. et al. 2002. A high-resolution data set of surface climate over global land areas. – *Clim. Res.* 21: 1–25.
- Oberdorff, T. et al. 2001. A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. – *Freshwater Biol.* 46: 399–415.
- Osborne, P. E. and Suárez-Seoane, S. 2002. Should data be partitioned before building large-scale distribution models? – *Ecol. Model.* 157: 249–259.
- Paradis, E. and Claude, J. 2002. Analysis of comparative data using generalized estimating equations. – *J. Theor. Biol.* 218: 175–185.
- Paradis, E. et al. 2004. APE: an R package for analyses of phylogenetics and evolution. – *Bioinformatics* 20: 289–290.
- Pearson, R. G. et al. 2006. Model-based uncertainty in species' range prediction. – *J. Biogeogr.* 33: 1704–1711.
- Pineda, E. and Lobo, J. M. 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. – *J. Anim. Ecol.* 78: 182–190.
- Pöyry, J. et al. 2008. Species traits are associated with the quality of bioclimatic models. – *Global Ecol. Biogeogr.* 17: 403–414.
- Puschendorf, R. et al. 2009. Distribution models for the amphibian chytrid *Batrachochytrium dendrobatidis* in Costa Rica: proposing climatic refuges as a conservation tool. – *Divers. Distrib.* 15: 401–408.
- Pyron, M. 1999. Relationships between geographical range size, body size, local abundance, and habitat breadth in North American suckers and sunfishes. – *J. Biogeogr.* 26: 549–558.
- R Development Core Team 2009. R: a language and environment for statistical computing. – R Foundation for Statistical Computing, Vienna.
- 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. Invasions* 11: 1017–1031.
- Segurado, P. and Araújo, M. B. 2004. An evaluation of methods for modelling species distributions. – *J. Biogeogr.* 31: 1555–1568.
- Swets, K. 1988. Measuring the accuracy of diagnostic systems. – *Science* 240: 1285–1293.
- Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. – *Global Change Biol.* 10: 2020–2027.
- Tsoar, A. et al. 2007. A comparative evaluation of presence-only methods for modelling species distribution. – *Divers. Distrib.* 13: 397–405.
- Willis, C. G. et al. 2008. Phylogenetic patterns of species loss in Thoreau's woods are driven by climate change. – *Proc. Nat. Acad. Sci. USA* 105: 17029–17033.

Download the Supplementary material as file E6152 from  
[www.oikos.ekol.lu.se/appendix](http://www.oikos.ekol.lu.se/appendix).