



# Spatial range shape drives the grain size effects in species distribution models

Christine Lauzeral, Gaël Grenouillet and Sébastien Brosse

C. Lauzeral ([christine.lauzeral@univ-tlse3.fr](mailto:christine.lauzeral@univ-tlse3.fr)), G. Grenouillet and S. Brosse, Laboratoire Evolution et Diversité Biologique, U.M.R. 5174, C.N.R.S. – Univ. Paul Sabatier, 118 route de Narbonne, FR-31062 Toulouse, France, and C.N.R.S., U.M.R. 5174 EDB, FR-31062 Toulouse, France.

Species distribution models (SDMs) link species occurrences to environmental descriptors using species and environmental data that are often recorded at different grain sizes. The upscaling process implied by grain size matching between species data and environmental data may affect the observed species distribution and thus might also modify the SDM-derived species distribution. Here we used five virtual species with differing range sizes to determine the effects of grain size on SDM-derived distribution area. We showed that the increase of SDM-derived distribution area with grain size is mainly due to the geometric increase of the area of the observed distribution range used to build the SDMs. Models based on presence–absence data that were built using the initial prevalence in the calibration dataset and the maximization of TSS or Kappa as cut-off threshold accurately predicted the observed area whatever the grain size and species range size. In addition we found that the commonly used evaluation measures (AUC, TSS and Kappa) cannot be used to evaluate the accuracy of SDM-derived distribution areas. Thus, the grain size of the data used to feed SDMs has to be chosen carefully, depending on the data quality and the goals of the study.

The grain size of the data used to feed species distribution models (SDMs) varies with the environmental and climate variables considered as well as with the source of occurrence data, the extent, and the location of the region considered (Guisan and Thuiller 2005). Previous studies have focused on the effects of grain size on model prediction performance (Guisan et al. 2007), but the influence of grain size on the SDM-derived distribution area (i.e. the surface of the predicted distribution range) remains a poorly addressed question. Seo et al. (2009) first devoted a paper entirely to this question and showed that the SDM-derived distribution areas of nine tree species could undergo up to a four-fold increase when increasing the grain size from  $1 \times 1 \text{ km}^2$  to  $64 \times 64 \text{ km}^2$ . They however did not identify the sources of this increase. Similarly, Hu and Jiang (2010) modelled the potential distribution of a rare gazelle and noticed a sharper increase in the predicted species distribution area (more than 15-fold increase) when increasing the grain size from  $1 \times 1 \text{ km}^2$  to  $32 \times 32 \text{ km}^2$  despite a slight decrease in model quality through upscaling. The consistency of these trends for various organisms including both plants (Seo et al. 2009) and animals (Hu and Jiang 2010) suggests that the species distribution area increase through grain size increase could have its root in methodological sources.

Increasing the operational grain size obviously induces a geometric increase of the observed species distribution area. This geometrical area increase is sensitive to the spatial

range shape measured as the mean number of empty neighbours around occupied cells (i.e. the fragmentation level of the distribution, Fig. 1). Indeed, if a species is present in only one out of four adjacent cells, merging these four cells when upscaling will lead to a four-fold area increase, whereas merging the cells will not affect the area if the species is present in the four cells. The increase of the observed area with grain size will in turn affect the SDM-derived species distribution area, explaining why Seo et al. (2009) found that the predicted species distribution area increase depended on the species range size. Nevertheless, the relationship between observed species distribution area and SDM-predicted area through upscaling, as well as the sources of variation of SDM-derived distribution areas through upscaling, remains to be clarified.

Here, we tested how increasing the grain size affects observed species distribution areas and in turn SDM-predicted areas. Five virtual species with known range size were created to measure how the observed and SDM-derived distribution areas vary through the upscaling process. We first assessed how the spatial range shape of the distribution range affected the observed species distribution area increase through upscaling. We then determined how the SDM-derived distribution areas were linked to the observed areas and how this relationship was affected by species range size, species prevalence in the calibration dataset, cut-off threshold choice and grain size. Finally,

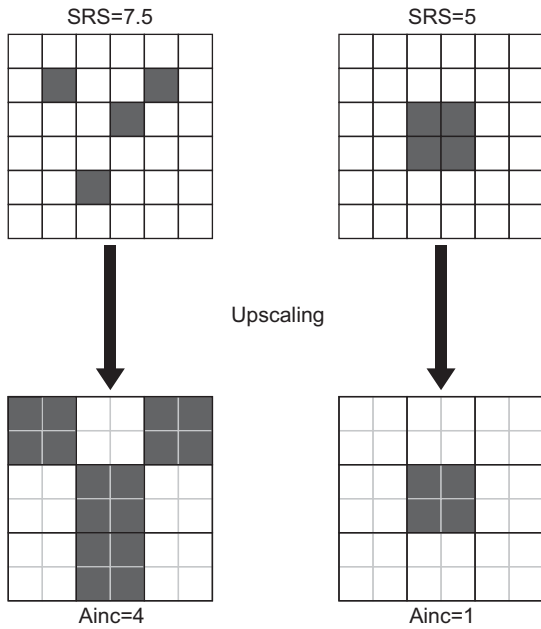


Figure 1. Conceptual representation of the spatial range shape (SRS, measured as the mean number of empty neighbours around occupied cells) and of the area increase (Ainc) through upscaling. Two initial range shapes with the same initial occupied area are represented. Area increase through upscaling was four-fold for the left panels, whereas area did not change through upscaling for the right panels.

we investigated how grain size affected model accuracy using model quality indices commonly found in the literature.

## Material and methods

### Virtual ecological niches and predictor variables

The effect of grain size on model predictions was tested using virtual species for which the presence only depended on climate variables. The procedure, although somewhat unrealistic from an ecological point of view, was designed to avoid 1) biases due to unmeasured environmental factors and 2) misvaluation of model quality associated to misclassification of sites (false absences).

The five virtual species were mapped over France at a  $30'' \times 30''$  resolution. Six climate variables were extracted over France from the  $30'' \times 30''$  resolution WorldClim layers for the period 1961–1990 (Hijmans et al. 2005): 1) precipitation in the driest quarter of the year and 2) in the wettest quarter; 3) average monthly precipitation; 4) mean temperature of the coldest quarter and 5) of the warmest quarter; and 6) annual mean temperature. These variables were chosen as they are related to the ecological requirements of numerous species, and have often been used in SDMs (Thuiller et al. 2005, Marini et al. 2009, Buisson et al. 2010).

The virtual species distributions were delineated by constructing two synthetic climate variables (Jimenez-Valverde and Lobo 2007), since the six variables show clear correlation structure. A normalized principal component analysis (PCA) was computed on the six climate variables and the first two axes of the PCA, which accounted for

93% of the total variance, were kept as synthetic variables. In the two-dimensional space created by these two orthogonal axes summarizing climatic variation across France, the virtual species niches were defined as discs centred on coordinates (0,0). All geographic cells falling within the disc for the pair of climate variables were considered as the observed distribution range of the virtual species in France. Using five different disc radii, five virtual species were created, covering 1, 5, 15, 30 and 60%, respectively, of all cells available in France (Fig. 2), so as to cover species with differing range sizes.

### Sampling of the virtual species

Among the 912730  $30'' \times 30''$  cells covering the area of France, 5000 cells were sampled according to a spatially random selection. This operation was repeated 100 times, giving rise to 100 datasets. As a presence–absence ratio of 1:1 (i.e. a prevalence of 0.5) has been recommended in some studies whereas others used the initial prevalence (Jimenez-Valverde and Lobo 2006, Buisson et al. 2008) i.e. the presence–absence ratio as obtained through the random sampling process, both were considered in this study. In the initial prevalence all the cells were kept whatever was the prevalence. In the fixed prevalence of 0.5, all presence cells were kept and absence cells were selected randomly from the remaining sampling cells. When insufficient absence cells were available, all absence cells were kept and presence cells were randomly selected to balance the prevalence.

### Grain sizes and upscaling

Seven grain sizes were selected:  $30'' \times 30''$ ,  $1' \times 1'$ ,  $2' \times 2'$ ,  $4' \times 4'$ ,  $8' \times 8'$ ,  $16' \times 16'$  and  $32' \times 32'$  (approximately

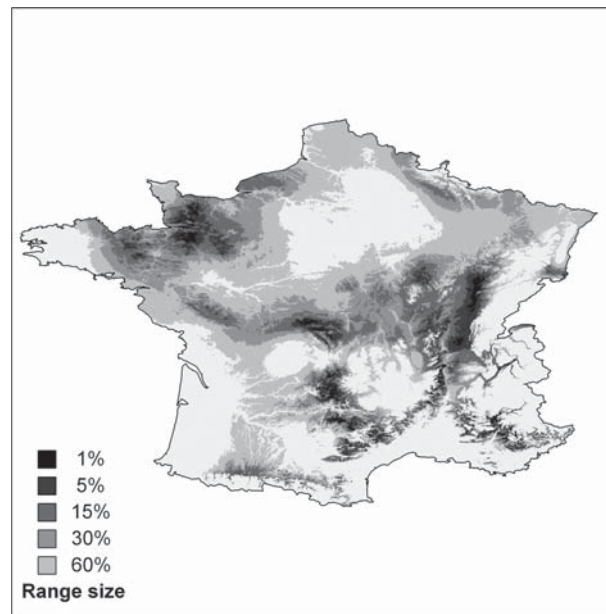


Figure 2. The geographic niches of the five species over France. Each niche contains the smaller ones. Grey scale represents different species range sizes expressed as proportion covered relative to the study area.

from  $1 \times 1 \text{ km}^2$  to  $60 \times 60 \text{ km}^2$ ) (Seo et al. 2009, Hu and Jiang 2010). Climate variables and species presence or absence were upscaled from the original  $30''$  resolution into these coarser resolution grids. Predictor variables were upscaled using the mean value of the merged predictor variables. A 'presence' was assigned to a coarser cell when one or more presence records were found in the merged cells.

In the sampling dataset, a cell belonged to the dataset if it contained at least one cell occurring in the dataset at the finest grain size. Sample size therefore decreased through upscaling as an upscaled cell contains one or more initial sampling cells. In the fixed prevalence of 0.5, the prevalence was balanced at each grain size as it was done at the smallest one. This gave rise to a total of 7000 distribution datasets ( $100 \text{ sampling datasets} \times 5 \text{ species} \times 7 \text{ grain sizes} \times 2 \text{ presence-absence ratio}$ ). Each of these 7000 datasets was randomly split into two parts: two-thirds of the data were used to calibrate the SDMs and the remaining third was used as a test set.

## Models

We used an ensemble modelling approach (Araújo and New 2007). Five classical SDMs (i.e. generalized linear models (GLM); generalized additive models; classification and regression trees; discriminant factorial analysis (DFA) and Random Forest) (Buisson et al. 2010) were run for each species, at each grain size, using the six climate variables. For GLMs and DFAs, squared variables were included in the model to deal with non-linearity. The probability of presence of each species was predicted across the whole region. Then a mean ensemble model was built by averaging the probabilities of presence from the five models and the resulting probability of presence was converted into presence-absence data using three different cut-off thresholds commonly employed (Liu et al. 2005, Jimenez-Valverde and Lobo 2007): 1) by maximizing the percentage of presences and absences correctly predicted i.e. by maximizing the TSS, 2) by maximizing the Kappa, 3) by using the prevalence of the calibration dataset. We thus obtained 600 ( $100 \text{ datasets} \times 2 \text{ presence-absence ratios in the calibration and test sets} \times 3 \text{ cut-off thresholds}$ ) predicted niches at each grain size and for each species.

## Effect of grain size on observed area

All the distribution areas were evaluated as the number of occupied  $30'' \times 30''$  cells (or pixels) included in the considered range. We first evaluated the observed distribution areas on each of the seven grain sizes and verified that observed areas increased exponentially with grain size. We then tested in which way this increase from one grain size to the following was linked to the spatial range shape of the observed niche. For each species and each grain size (except the coarsest one), we calculated the mean number of empty cells ( $N_{\text{emp}}$ ) adjacent to the occupied cells (i.e. the number of empty cells among the eight nearest cells). We then calculated the species distribution area increase ( $A_{\text{inc}}$ ) as the ratio between the distribution areas measured for two successive grains and assessed the relationship between  $N_{\text{emp}}$

and  $A_{\text{inc}}$  to test whether the increase of species distribution areas was dependent on the spatial range shape.

## Effect of grain size on SDM-derived distribution area

Regarding SDM outputs, we first tested the model's ability to predict the observed area. We evaluated the correlation between observed and SDM-derived distribution areas. The ratio between observed and SDM-derived distribution areas was then calculated for each grain size. This enabled us to view the effect of grain size on the accuracy of the SDM-predicted distribution area after removing the geometric increase of the observed area. We assessed the results for each species, each prevalence in the calibration dataset and each cut-off threshold. Then, for each species, we used a linear model to partition out the variability in the ratio between observed and predicted areas due to each of the other parameters (grain size, prevalence in the calibration dataset and cut-off threshold) by using the ratio between the variance explained by one factor and the total variance (Dormann et al. 2008).

## Model accuracy

To assess spatial congruence between the observed and the predicted niche, model accuracy was evaluated using three commonly used metrics: the area under the ROC curve (AUC), a threshold-independent measure (Hanley and McNeil 1982); Kappa (Cohen 1960) and TSS (Allouche et al. 2006), that are two threshold-dependent measures. We also plotted a map of omission and commission errors. For each species, we counted, over the 100 models built using the 100 different calibration datasets, the percentage of mispredicting models in each cell. This was done at each grain size.

## Results

Upscaling the grain size from  $30'' \times 30''$  to  $32' \times 32'$  caused a sharp increase in observed species distribution areas (a 1.8-fold increase for the common species, and up to 22-fold for the rarest species) (Fig. 3A). The species distribution area increased exponentially with grain size (Pearson correlation,  $r = 0.99$ ,  $p < 0.001$ ). Through upscaling, the increase of species distribution areas (log-transformed) was significantly linked to the spatial range shape (Fig. 3B) as it was strongly positively correlated to the number of empty cells around occupied cells (Pearson correlation  $r = 0.99$ ,  $p < 0.001$ ). The increase of species distribution areas was therefore more pronounced for rare species, which distribution range was more fragmented (2.9 empty neighbours for the rarest species vs 0.3 for the common species at the finest grain size, Fig. 3B).

The total number of points in the calibration dataset exhibited a slight decrease for the four smaller grain size changes and then sharply decreased, from 4200 to 260 sampling points (Fig. 4A). Upscaling slightly affected the occurrence of the species in the sampling (calibration and test) datasets except for the largest grain sizes (i.e. from

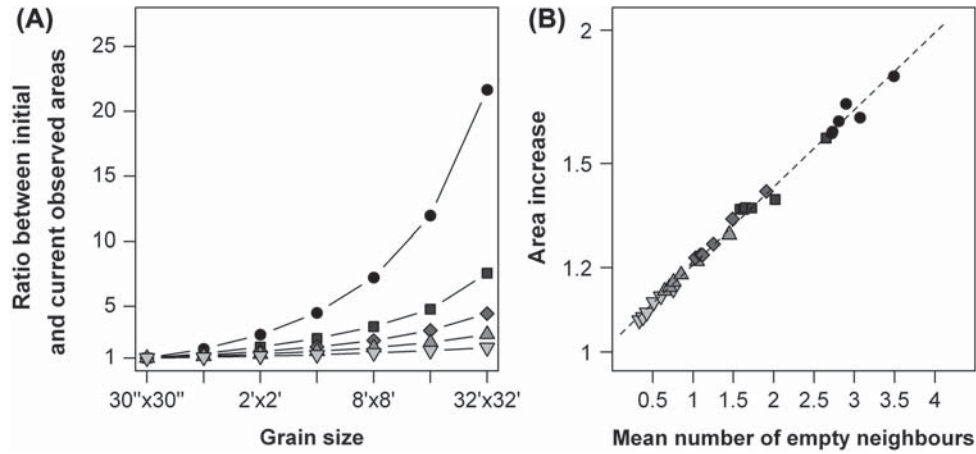


Figure 3. Observed distribution area increase through upscaling. (A) Ratio between the observed distribution area at the 30'' × 30'' grain size and the six larger grain sizes. (B) Relationship between the observed distribution area increase (i.e. the ratio between the area measured on two successive grains, log-scaled) and the mean number of empty neighbours adjacent to occupied cells (for the smaller grain size). The dashed line represents the linear relationship between the area increase and the number of empty neighbours ( $r = 0.99$ ,  $p < 0.001$ ). Symbols represent virtual species range sizes expressed as proportion covered relative to the study area. Circles: 1%; squares: 5%; diamonds: 15%; up triangles: 30%; down triangles: 60%. Grey scale as in Fig. 2.

8' × 8' to 32' × 32', Fig. 4A). Whatever the species, occurrence first slightly increased with grain size (from 30'' × 30'' to 4' × 4'), as increasing the cell size also increased the number of cells intersecting with the niche. For larger grain sizes (8' × 8' to 32' × 32'), the occurrence decreased due to a decline of the total number of cells in the datasets, although rare species had more stable occurrences than abundant ones. Meanwhile, the prevalence of the species in the sampling dataset increased with grain size, especially for rare species (Fig. 4B). In addition, for the five species, grain size increase through upscaling gradually modified the shape of the environmental niche. At smaller grain sizes the niche was compact but shifted towards a more dispersed pattern at higher grain sizes (Supplementary material Appendix 1, Fig. A1).

Most of the SDM-derived distribution areas increased exponentially with grain size. The Pearson correlation coefficients between grain size and the log of the area, as well as between SDM-derived and observed distribution areas

were highly significant for most species whatever the species range size and the prevalence in the calibration dataset ( $r > 0.93$ ,  $p < 0.001$  and  $r > 0.80$ ,  $p < 0.05$ , respectively). The only exceptions were when a prevalence of 0.5 was used for both the rarest species (1% of the study area) particularly using Kappa maximisation as cut-off, and the most abundant species (60% of the study area) using TSS maximisation as cut-off (Supplementary material Appendix 1, Table A1, A2). The ratio between observed and SDM-derived distribution area varied slightly, and the variance was mostly explained by the cut-off threshold and the grain size (Table 1).

Whatever the prevalence in the calibration dataset and cut-off threshold considered, SDM predictions of common species (15, 30 and 60% of the study area) showed an increase in SDM-derived distribution area ranging from 1.5 to 5-fold through upscaling (Fig. 5, Supplementary material Appendix 1, Fig. A2). More precisely, all SDMs accurately predicted the common species distribution areas,

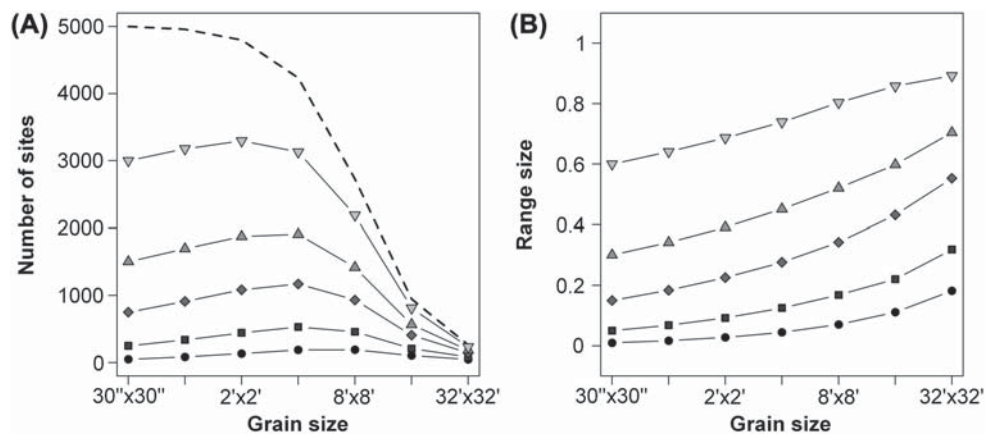


Figure 4. (A) Number of species occurrences in response to the upscaling process. The mean number of available sampling cells through upscaling is indicated by a dashed line. (B) Species range size expressed as proportion covered relative to the study area. Symbols represent virtual species range sizes at the smallest grain size. Circles: 1%; squares: 5%; diamonds: 15%; up triangles: 30%; down triangles: 60%. Grey scale as in Fig. 2.

Table 1. Analysis of variance (ANOVA) of the ratio between observed and SDM-derived range sizes. Using a linear model, the variability was partitioned out in the ratio between observed and predicted areas due to each of the other parameters (grain size, prevalence in the calibration dataset and cut-off threshold) by using the ratio between the variance explained by one factor and the total variance.

	Species range size				
	1%	5%	15%	30%	60%
Grain size	22.28	9.08	18.59	40.99	27.57
Prevalence	11.96	9.1	4.84	0.14	9.27
Cut-off threshold	7.35	40.47	36.86	13.8	6.28
Grain size × prevalence	8.68	0.08	1.15	3.26	8.5
Grain size × cut-off threshold	4.69	2.9	3.15	4.75	5.48
Prevalence × cut-off threshold	6.82	16.61	6.58	0.12	3.37

as they predicted an area, which was on average 1.02 ( $\pm$  0.11) times larger than the observed one measured on the same grain size (Fig. 5, Supplementary material Appendix 1, Fig. A2).

With regard to rare species (1 or 5% of the study area), SDM predictions were strongly affected by the prevalence (initial or 0.5) and the cut-off threshold (Kappa, TSS or prevalence). Using a prevalence of 0.5, SDM-derived distribution area increased through upscaling between 3 and 9-fold (Supplementary material Appendix 1, Fig. A2). These models strongly overpredicted the rare species distribution areas (except at very large grain size). They predicted a species distribution area 5.06 ( $\pm$  3.90) times larger than

the observed one considered on the same grain size (on average over grain sizes) and 12.26 ( $\pm$  7.35) at the smallest grain size for the rarest species (Supplementary material Appendix 1, Fig. A2).

Using an initial prevalence in the calibration dataset, the increase of SDM-derived distribution area for rare species through upscaling was strongly influenced by cut-off threshold selection (Fig. 5). The prevalence cut-off provided results similar to those obtained using a prevalence of 0.5. The SDM-derived distribution area showed an 5-fold increase through upscaling. Ratio analysis showed that the SDM-derived distribution area was on average 4.83 ( $\pm$  2.02) times larger (7.91 ( $\pm$  0.86) times larger at the smallest grain size for the rarest species). On the contrary, the models using Kappa and TSS cut-off as thresholds produced SDM-derived distribution area increases through upscaling similar to those observed for the observed niche (Fig. 5). The two cut-off threshold models accurately predicted both species distribution areas, as they predicted areas, which were on average 1.40 ( $\pm$  0.51) times larger than the observed ones measured on the same grain size (Fig. 5).

Concerning model predictive accuracy, the AUC slightly decreased with grain size (Fig. 6, Supplementary material Appendix 1, Fig. A3). This decrease was greater for rare species and for large grain sizes. The pattern did not differ among prevalence in the calibration and test sets. A similar tendency was recorded for Kappa, TSS, sensitivity and specificity (Fig. 7, Supplementary material Appendix 1, Fig. A4). Only the Kappa of the rarest species models built

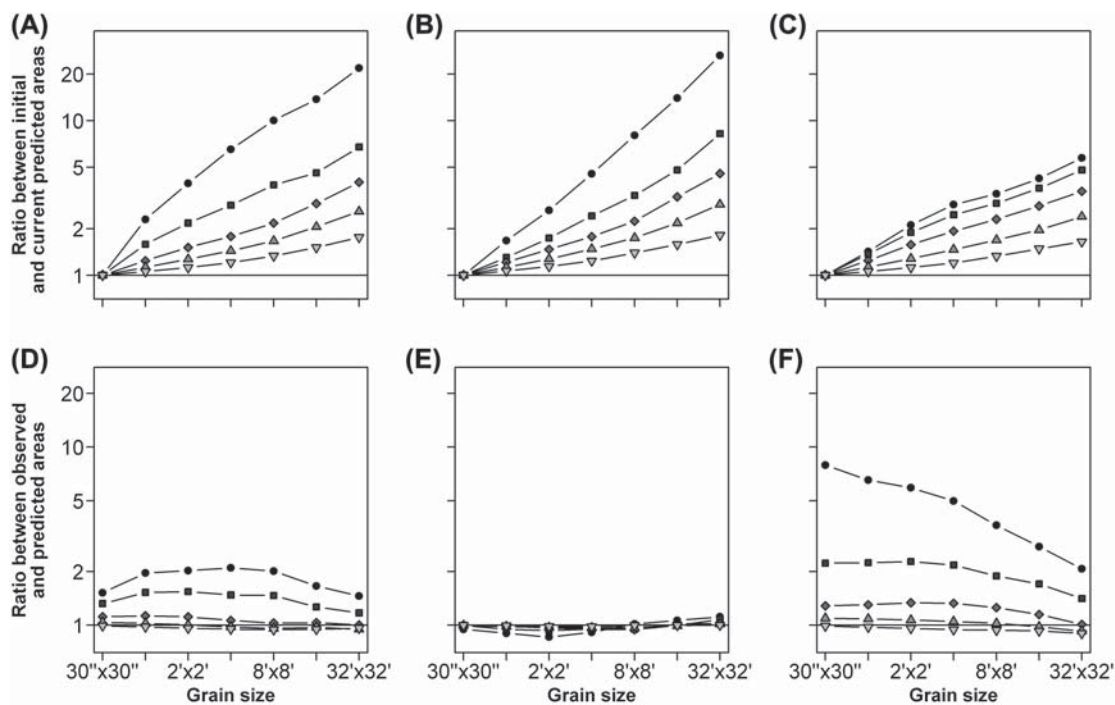


Figure 5. SDM predicted distributions areas through upscaling using initial prevalence in the calibration dataset. (A, B, C) Mean ratio between the predicted distribution areas at the 30'' × 30'' grain and the six other grains sizes (log-scaled). (D, E, F) Mean ratio between the observed and the SDM-derived distribution areas (log-scaled). In each case, the SDM-derived area was measured on the grain size at which the model was built and compared to the observed area measured on the same grain size. Cut-off thresholds are the maximisation of the TSS (A, D); the maximisation of the Kappa (B, E); the prevalence in the calibration dataset (C, F). Symbols represent virtual species range sizes expressed as proportion covered relative to the study area. Circles: 1%; squares: 5%; diamonds: 15%; up triangles: 30%; down triangles: 60%. Grey scale as in Fig. 2.

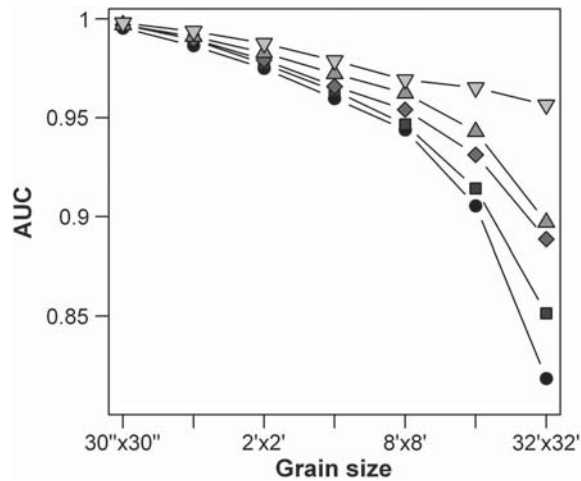


Figure 6. Effect of grain size on AUC (mean value over the 100 test sets) for the five virtual species. All the models were built using initial prevalence in the calibration and test sets. Symbols represent virtual species range sizes expressed as proportion covered relative to the study area. Circles: 1%; squares: 5%; diamonds: 15%; up triangles: 30%; down triangles: 60%. Grey scale as in Fig. 2.

using initial presence–absence ratio and prevalence as threshold and the sensitivity of rare species models using initial presence–absence ratio and Kappa or TSS as threshold increased with grain size.

Although quality indices were slightly affected by upscaling, the geographical distribution of model omission errors of rare species dramatically varied through the upscaling process. At small grain size, models predicting the rare species distribution often omitted a large part of the geographical distribution while omission errors of models built at larger grain sizes were more uniformly distributed (Fig. 8, Supplementary material Appendix 1, Fig. A5). In contrast, model commission errors were mainly located at the edge of the observed distribution, whatever the grain size.

## Discussion

Work focusing on grain size effects has provided mixed results: although Guisan et al. (2007) showed that SDM performance was not greatly affected by a 10-fold change in grain size, Seo et al. (2009) demonstrated that changing grain size dramatically increased the SDM-derived species distribution area. As reported by Seo et al. (2009) using real species, in the present study we observed that the SDM-derived distribution areas of five virtual species increased exponentially with grain size. This increase was primarily due to the increase of the observed distribution area caused by coarsening grain size, mostly at the edge of the distribution due to the presence of the species in cells adjacent to empty cells, revealing a strong effect of the spatial range shape. Moreover, its magnitude was strongly influenced by species rarity, with the rarest species experiencing the largest distribution area increase through the upscaling process. This pattern differed from that reported by Seo et al. (2009) as they observed a greater increase for the species of intermediate range size (10–20% of the

study area). The link between distribution range shape and area increase may explain this difference. Our five virtual species were built in such a way that the spatial distributions of rare species were more fragmented than those of more common ones. This is not a general rule as some rare species can have clustered distributions and in this case should experience low area increase through upscaling. On the contrary, more common species with a fragmented distribution should experience a higher area increase. This probably explains why Hu and Jiang (2010) observed a sharper area increase than Seo et al. (2009) as they modelled a notably fragmented species. The area increase of observed distributions might hence be amplified in the case of real species with highly fragmented distributions. Such a situation is especially relevant to endangered species, since these often show fragmented distribution ranges due to habitat destruction or other human disturbances (Ewers and Didham 2006). Particular caution should therefore be taken when choosing grain size to measure distribution areas of species whose populations are fragmented.

Our results obtained from virtual species based on simulated niches reveal a similar exponential increase of range size with grain size to that reported in studies conducted on real species (Seo et al. 2009, Hu and Jiang 2010), but the coupled effect of range size and distribution shape on observed and predicted distributions through upscaling remains to be tested in more detail. Apart from the geometric effect influencing the observed area and hence the SDMs outputs, other parameters can affect the predicted species distribution areas, i.e. the choice of the sampling dataset (number of sampling sites, occurrence of the species), the species range size, the species prevalence in the calibration dataset and the cut-off threshold. In our study the size of the sampling dataset remained sufficient to build relevant models whatever the grain size as little change was observed in models outputs despite a marked decline of the number of sampling sites.

Concerning the three remaining parameters (species range size, species prevalence in the calibration dataset and cut-off threshold), the distribution areas of species having a range size higher than 10% of the study area were accurately predicted whatever the prevalence in the calibration and test set and the threshold used. This contrasts with rare species for which area prediction accuracy depended on both the prevalence in the calibration dataset and the cut-off threshold selection. For these species, a strong area overprediction occurred in four out of the six presence–absence ratio and cut-off combinations, thus paralleling the overpredicted prevalence observed by Manel et al. (2001) on rare invertebrate species. Considering the prevalence in the calibration dataset for rare species, the initial prevalence preserved a better representation of the distribution area, and should hence be preferred to a fixed prevalence of 0.5. It should also be noted that the models built using a prevalence of 0.5 tended to underpredict the distribution area of the most abundant species (prevalence 60%), especially for the largest grain sizes. In both cases, the use of a prevalence of 0.5 in the calibration dataset introduced a bias when the prevalence of the species in the observed niche notably differed from 0.5, such as for the rarest species at the smallest grain size (1% of the study area)

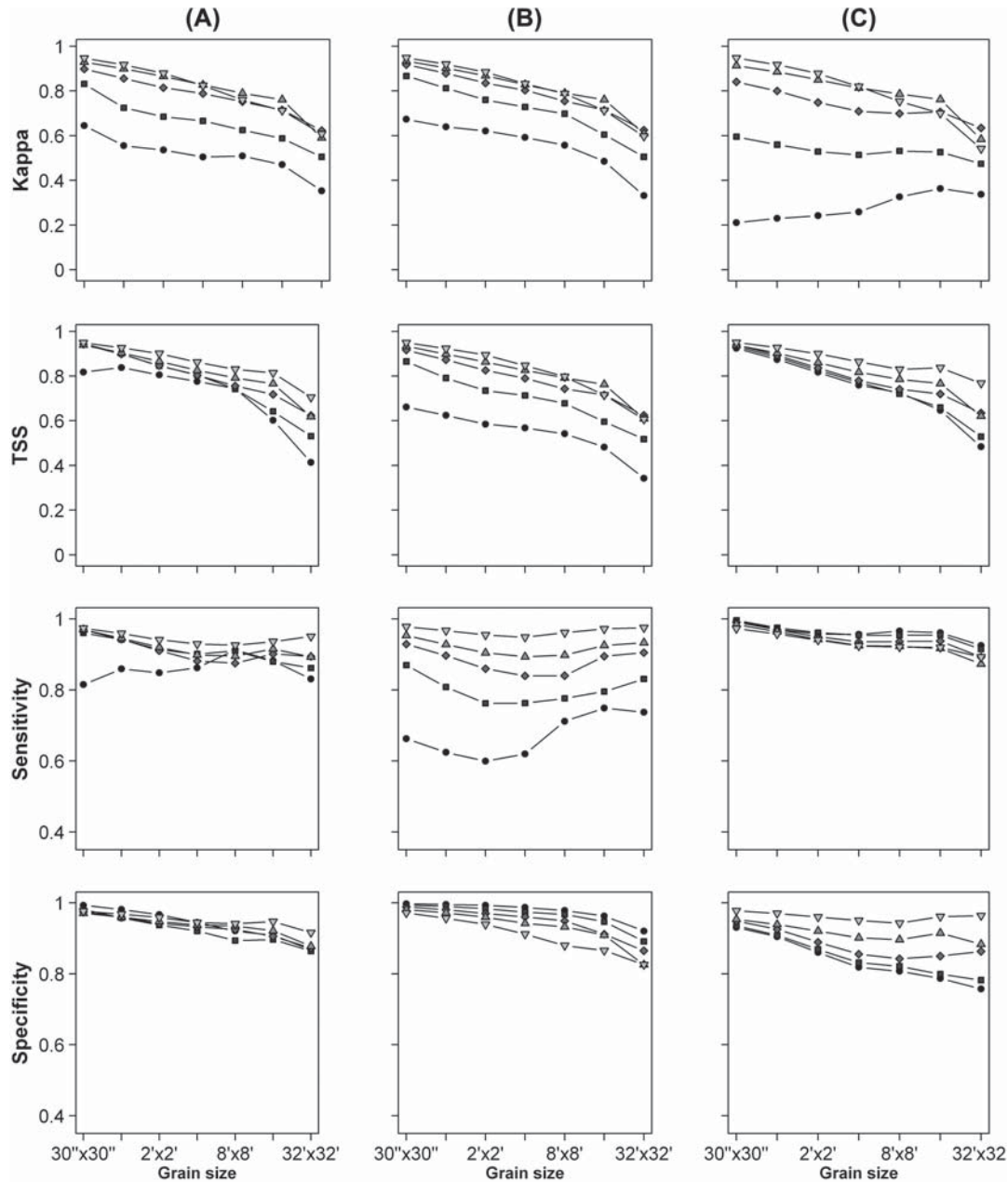


Figure 7. Effect of grain size on model accuracy using Kappa, TSS, sensitivity and specificity (mean value over the 100 test sets) for the five virtual species. All the models were built using initial prevalence in the calibration and test sets. Cut-off thresholds are the maximisation of the TSS (A); the maximisation of the Kappa (B) and the prevalence in the calibration dataset (C). Symbols represent virtual species range sizes expressed as proportion covered relative to the study area. Circles: 1%; squares: 5%; diamonds: 15%; up triangles: 30%; down triangles: 60%. Grey scale as in Fig. 2.

or for the most abundant species at the largest grain size (90% of the study area). Indeed, resampling in order to obtain a prevalence of 0.5 yields a loss of information (Jimenez-Valverde and Lobo 2006), especially in environmental conditions unsuitable for the rare species, leading to overprediction of the SDM-derived distribution area. We hence recommend designing SDMs built on presence-absence data to use initial prevalence in the calibration dataset, and either a Kappa or a TSS maximisation as cut-off threshold, as these two indices minimize area overprediction. Such a procedure is however sometimes difficult to apply, as absence data are often rare or unreliable. In that

case, the use of a large number of pseudo-absences is often encouraged (Wisn and Guisan 2009, Stokland et al. 2011), but their influence on model accuracy remains controversial. Indeed, Stokland et al. (2011) argued that the number of pseudo-absences had little effect on model accuracy (measured using AUC) and distribution area, whereas Lobo and Tognelli (2011) showed that the number of pseudo-absences significantly affected model accuracy (measured using AUC, sensitivity and specificity). Lobo and Tognelli (2011) did not test the effect of pseudo-absence number on the predicted area. Here we show that the ratio between presence and absence indeed affects the predicted

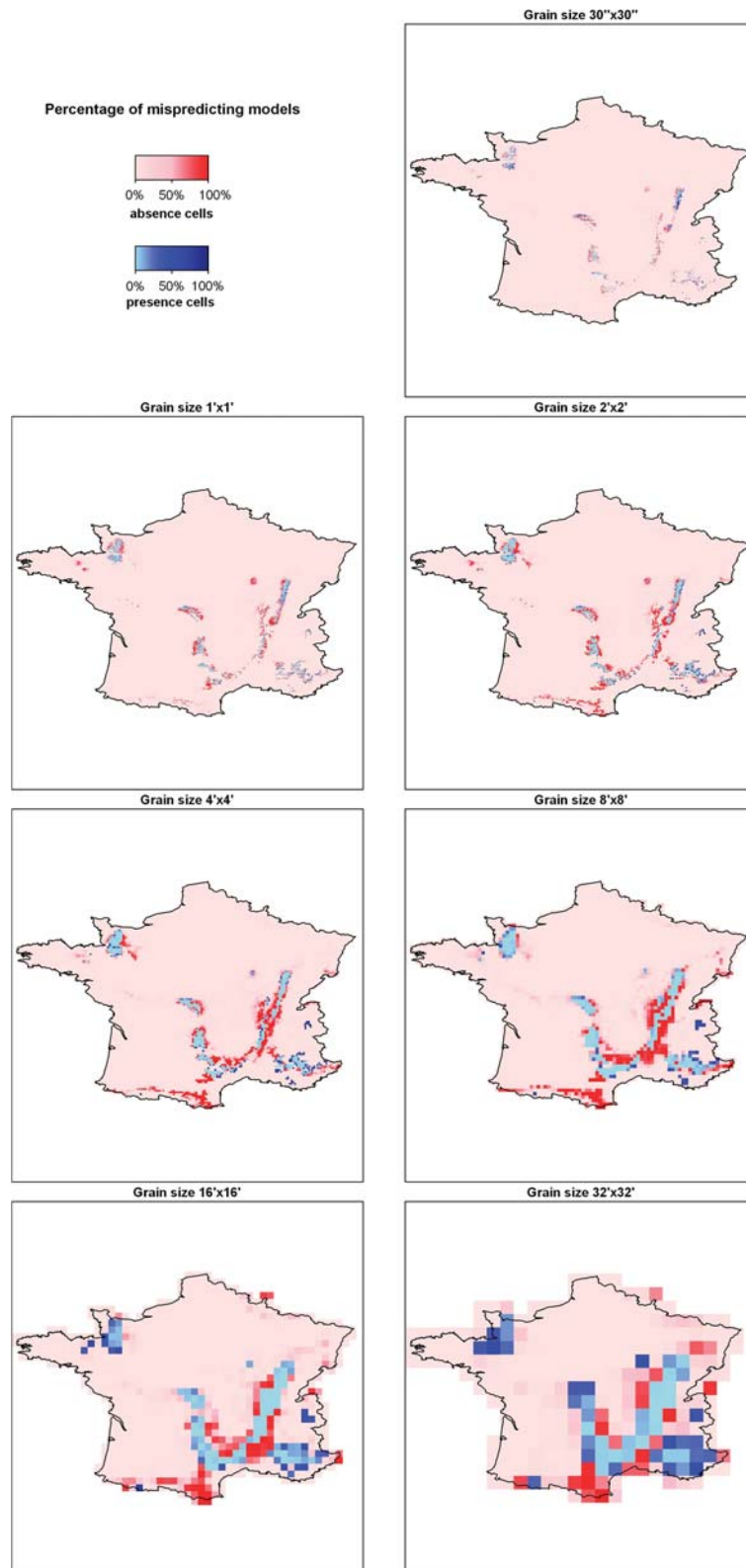


Figure 8. The predicted distribution of the rare species (1% of the study area) at each grain size. Models were built using initial prevalence in the calibration datasets and TSS maximisation as cut-off threshold. The 100 models based on the 100 different calibration datasets were used and we evaluated the percentage of mispredicting models in each cell. The darkest cells are the most often mispredicted. Note that the model built at the smallest grain omitted a large geographic part of the niche (the south-eastern part).



distribution area, suggesting that the number of pseudo-absences might matter. We therefore recommend that forthcoming studies evaluate the effect of pseudo-absence number, threshold selection and grain size not only on model accuracy but also on the ratio between observed and predicted distribution areas.

Considering model predictive accuracy, the slight decrease of most of the quality indices with grain size is consistent within the findings of previous studies (Guisan et al. 2007, Hu and Jiang 2010). However, Guisan et al. (2007) observed that this trend might be taxon-dependent. Indeed, plant distribution was better predicted using small grain sizes, whereas model quality was improved by upscaling for some bird species. This improvement was mostly observed for a grain size increase from 0.1 to 1 km<sup>2</sup>. Such an increase of model quality was also observed on green woodpecker *Picus viridis* distribution (Tobalske and Tobalske 1999) when increasing the grain size from 1 to 4 km<sup>2</sup>. These contrasting results between plants and animals might be related to discrepancies in displacement abilities of organisms. Birds are mobile and might therefore evaluate the quality of their niche at a larger spatial grain than plants that are strictly subjected to the environmental conditions occurring at the exact place they live. A relevant grain size to model a species distribution might therefore reflect the spatial grain at which the individuals exploit their environment. It might hence increase with the organism movement ability and the size of its living area. This hypothesis however remains to be tested on a much more varied and abundant sample of real species.

Above all, our results emphasise that the indices classically used in the evaluation of model accuracy (AUC, Kappa, TSS) cannot be used to evaluate the accuracy of SDM-derived distribution areas. Some models, mainly using a prevalence of 0.5, that failed to reliably predict distribution area (Supplementary material Appendix 1, Fig. A2) were considered as reliable according to TSS, Kappa and AUC indices (Supplementary material Appendix 1, Fig. A3, A4). In addition, Kappa is known to be highly sensitive to prevalence in the test dataset (Allouche et al. 2006), the choice of a initial prevalence in the test dataset could lead to smaller values of Kappa at fine grain size. But this dependence reflects the chance corrected nature of Kappa (Santika 2011). In our case, Kappa calculated on models built on initial prevalence by TSS or Kappa maximization slightly varied with grain size although upscaling dramatically increased prevalence in the test dataset. The Kappa increase was only observed when prevalence was used as a cut-off. It was linked to the decrease of rare species area overprediction. The decrease of the number of predicted presence cells decreased the probability of correct presence prediction by chance. Moreover, the decrease of area overprediction did not affect the TSS measure due to the rarity of the species. Our study thus supports the interest of using Kappa to measure model quality.

The Kappa maximisation cut-off threshold that is already known to give accurate prevalence predictions (Freeman and Moisen 2008), also gives the most accurate area predictions, particularly for rare species. But this is done at the expense of model sensitivity as models predicting rare species omitted a large part of the observed

distribution. From a practical point of view, the cut-off threshold thus has to be chosen depending on the study goal. If false absences in distribution predictions have to be restricted, TSS should be preferred. This is for example the case for invasive species studies, where the model aims at identifying potential invasion areas. Indeed from an ecosystem management point of view, overpredicting the invasion area is better than omitting potential invasion sites (Mack et al. 2000, Leprieur et al. 2009). In contrast, Kappa should be preferred when searching to predict potential prevalence or range size. It could be of interest to design conservation strategies and to set-up protected areas for endangered species, as socioeconomic constraints often strongly limit the extent of the protected areas (Klein et al. 2008). Hence, a reliable identification of the favourable sites to protect using highly specific models (and Kappa cut-off threshold) should be preferred over more sensitive models that will overpredict potentially favourable niches.

In accordance with Guisan et al. (2007), here we confirmed that grain size slightly affects model performance measures, but we also demonstrated that, using initial prevalence in the calibration dataset and Kappa or TSS maximisation as cut-off threshold, range size predictions were hardly affected once the effect of observed area increase was removed. However, grain size affected the geographic distribution of omission errors of models predicting rare species distribution. Indeed, at fine grain size, SDMs failed to predict presence in large geographical regions at the edge of the environmental niche. Such a bias occurred less frequently at coarser grain size. However, the model outputs at the edge of the niche may be affected by the definition of the virtual species, as this definition considered a binomial response to environmental variables instead of a continuous gradient of occurrence probability. Moreover, real species parameters like topography or land cover can increase the uncertainty of projections, especially at fine grain size (Wiens and Bachelet 2009). The effect of grain size on the geographic distribution remains to be tested in more detail on real species. Our results however suggest that an optimal grain size probably does not exist and that it has to be selected depending on the data quality, the ecology of the species and the goals of the study.

*Acknowledgements* – This study was supported by the BIOFRESH European project (FP7-ENV-2008, contract no. 226874). EDB is part of the 'Laboratoire d'Excellence' (LABEX) entitled TULIP (ANR-10-LABX-41).

## References

- Allouche, O. et al. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). – *J. Appl. Ecol.* 43: 1223–1232.
- Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – *Trends Ecol. Evol.* 22: 42–47.
- Buisson, L. et al. 2008. Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. – *Ecol. Freshwater Fish* 17: 244–257.
- Buisson, L. et al. 2010. Uncertainty in ensemble forecasting of species distribution. – *Global Change Biol.* 16: 1145–1157.

- Cohen, J. 1960. A coefficient of agreement for nominal scales. – *Educ. Psychol. Meas.* 20: 37–46.
- 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.
- Ewers, R. M. and Didham, R. K. 2006. Confounding factors in the detection of species responses to habitat fragmentation. – *Biol. Rev.* 81: 117–142.
- Freeman, E. A. and Moisen, G. G. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. – *Ecol. Model.* 217: 48–58.
- Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – *Ecol. Lett.* 8: 993–1009.
- Guisan, A. et al. 2007. Sensitivity of predictive species distribution models to change in grain size. – *Divers. Distrib.* 13: 332–340.
- Hanley, J. A. and McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. – *Radiology* 143: 29–36.
- Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – *Int. J. Climatol.* 25: 1965–1978.
- Hu, J. and Jiang, Z. 2010. Predicting the potential distribution of the endangered Przewalski's gazelle. – *J. Zool.* 282: 54–63.
- Jimenez-Valverde, A. and Lobo, J. M. 2006. The ghost of unbalanced species distribution data in geographical model predictions. – *Divers. Distrib.* 12: 521–524.
- Jimenez-Valverde, A. and Lobo, J. M. 2007. Threshold criteria for conversion of probability of species presence to either- or presence-absence. – *Acta Oecol.* 31: 361–369.
- Klein, C. J. et al. 2008. Striking a balance between biodiversity conservation and socioeconomic viability in the design of marine protected areas. – *Conserv. Biol.* 22: 691–700.
- Leprieur, F. et al. 2009. Scientific uncertainty and the assessment of risks posed by non-native freshwater fishes. – *Fish Fish.* 10: 88–97.
- Liu, C. R. et al. 2005. Selecting thresholds of occurrence in the prediction of species distributions. – *Ecography* 28: 385–393.
- Lobo, J. M. and Tognelli, M. F. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. – *J. Nat. Conserv.* 19: 1–7.
- Mack, R. N. et al. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. – *Ecol. Appl.* 10: 689–710.
- Manel, S. et al. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. – *J. Appl. Ecol.* 38: 921–931.
- Marini, M. A. et al. 2009. Predicted climate-driven bird distribution changes and forecasted conservation conflicts in a Neotropical savanna. – *Conserv. Biol.* 23: 1558–1567.
- Santika, T. 2011. Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. – *Global Ecol. Biogeogr.* 20: 181–192.
- Seo, C. et al. 2009. Scale effects in species distribution models: implications for conservation planning under climate change. – *Biol. Lett.* 5: 39–43.
- Stokland, J. N. et al. 2011. Species distribution modeling – effect of design and sample size of pseudo-absence observations. – *Ecol. Model.* 222: 1800–1809.
- Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – *Proc. Natl Acad. Sci. USA* 102: 8245–8250.
- Tobalske, C. and Tobalske, B. W. 1999. Using atlas data to model the distribution of woodpecker species in the Jura, France. – *Condor* 101: 472–483.
- Wiens, J. A. and Bachelet, D. 2009. Matching the multiple scales of conservation with the multiple scales of climate change. – *Conserv. Biol.* 24: 51–62.
- Wisz, M. S. and Guisan, A. 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. – *BMC Ecol.* 9: 8.

Supplementary material (Appendix E7696 at <[www.oikosoffice.lu.se/appendix](http://www.oikosoffice.lu.se/appendix)>). Appendix 1.