



Sensitivity of predictive species distribution models to change in grain size

Antoine Guisan^{1*}, Catherine H. Graham², Jane Elith³, Falk Huettmann⁴ and the NCEAS Species Distribution Modelling Group†

¹Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland;

²Department of Ecology and Evolution, 650 Life Sciences Building, Stony Brook University, NY 11794, USA; ³School of Botany, The University of Melbourne, Parkville, Victoria, 3010 Australia;

⁴EWHALE Lab — Biology and Wildlife Department, Institute of Arctic Biology, University of Alaska Fairbanks, AK 99775-7000, USA

*Correspondence: Antoine Guisan, Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland.

Fax: +41 21 692 42 65;

E-mail: antoine.guisan@unil.ch

†Miro Dudík, Simon Ferrier, Robert Hijmans, Anthony Lehmann, Jin Li, Lúcia G. Lohmann, Bette Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Scachetti-Pereira, Robert E. Schapire, Stephen E. Williams, Mary S. Wisz, Niklaus E. Zimmermann

ABSTRACT

Predictive species distribution modelling (SDM) has become an essential tool in biodiversity conservation and management. The choice of grain size (resolution) of environmental layers used in modelling is one important factor that may affect predictions. We applied 10 distinct modelling techniques to presence-only data for 50 species in five different regions, to test whether: (1) a 10-fold coarsening of resolution affects predictive performance of SDMs, and (2) any observed effects are dependent on the type of region, modelling technique, or species considered. Results show that a 10 times change in grain size does not severely affect predictions from species distribution models. The overall trend is towards degradation of model performance, but improvement can also be observed. Changing grain size does not equally affect models across regions, techniques, and species types. The strongest effect is on regions and species types, with tree species in the data sets (regions) with highest locational accuracy being most affected. Changing grain size had little influence on the ranking of techniques: boosted regression trees remain best at both resolutions. The number of occurrences used for model training had an important effect, with larger sample sizes resulting in better models, which tended to be more sensitive to grain. Effect of grain change was only noticeable for models reaching sufficient performance and/or with initial data that have an intrinsic error smaller than the coarser grain size.

Keywords

Environmental grain, niche-based modelling, natural history collections, presence-only data, resolution, spatial scale, sample size, species distribution modelling, model comparison, predictive performance.

INTRODUCTION

Predictive species distribution models (SDMs; Guisan & Zimmermann, 2000) have become essential tools in biodiversity conservation and management (Côté & Reynolds, 2002). Fitting an SDM involves a series of steps, each requiring a number of choices and well-justified decisions (Ferrier *et al.*, 2002b; Guisan & Thuiller, 2005). Grain (resolution) is one important factor that may affect predictions. It is, together with study extent, a component of spatial scale (Wiens, 2002) and a major feature in ecology (e.g. Holling, 1992) and ecological modelling (e.g. Huettmann & Diamond, 2006). Important questions therefore include: Is there an optimal grain for fitting SDMs? What is the effect of changing grain size on SDM performance?

Choosing a grain size for modelling is partly a technical issue. For instance, grain size is related to the grid cell size of available environmental data (Graham *et al.*, 2004), characteristics of the species data (e.g. geographical accuracy, sample size, field survey constraints, or autocorrelation structure; Guisan & Hofer, 2003;

Gottschalk *et al.*, 2005; Linke *et al.*, 2005; Huettmann & Diamond, 2006) or computer power (i.e. too many cells may require too demanding computer resources). Grain size is also a crucial ecological as well as management issue. Changing the grain size can influence the perception of a phenomenon, such as patterns of presence or abundance (Johnson *et al.*, 2002; Tobalske, 2002; Wiens, 2002; Graham & Hijmans, 2006), or affect the relevance of the output for management applications (Araújo *et al.*, 2005). Working at the wrong scale can be very inefficient.

Data from natural history collections often have significant error, making them difficult to use with fine-grained environmental data (Graham *et al.*, 2004). However, current georeferencing initiatives (e.g. GBIF, MANIS error calculator¹, OBIS², and OBIS-SEAMAP³) include an estimate of the maximum error distance

¹<http://manisnet.org>

²<http://www.iobis.org/welcome.htm>

³<http://seamap.env.duke.edu>

(location uncertainty) with each occurrence. Hence, the modeller can determine what data are appropriate at a given grain size and can choose to increase the grain of environmental data. At larger grains, more occurrence records may be available because the limit on locational accuracy is relaxed to match the new grain size (see, e.g. Engler *et al.*, 2004). Generally, more accurate predictions can be made with larger numbers (Stockwell & Peterson, 2002; Hernandez *et al.*, 2006), but also more accurate occurrences (e.g. for plants, Engler *et al.*, 2004). Hence, it is essential to evaluate the trade-off between the number of occurrence samples available for modelling and the grain size of environmental data.

If species records are inaccurate, a set of predictors available at a too fine grain may need to be aggregated to a coarser grain. For example, in recent projections of plant distributions in Europe (Thuiller *et al.*, 2005), the grain of the input species data was 50×50 km and the grain of available explanatory climatic maps was around 16×16 km ($10' \times 10'$). Models were first fitted by aggregating climatic map to the coarser grain and then re-projected to the finer climatic grain. Various other approaches have also been proposed to downscale atlas data using species distribution models (McPherson *et al.*, 2006). Araiyo *et al.* (2005) found a reasonable agreement of downscaled predictions with real patterns of occurrences obtained from fine-scale breeding birds atlas data, but they did not formally compare coarse- to fine-grain models.

There are many reasons why changing grain size could have an effect on the performance of SDMs. For instance, at a fine grain, the risk that a wrong geographical location of a species record samples a cell representing a different habitat than the one where the species actually occurred increases, while the opposite will be observed when aggregating data towards coarser grains. However, when coarsening the grain, the risk of a forced-matching between environmental conditions that do not occur together but nearby in the field increases and can make the model identify spurious combinations of suitable environmental conditions for a species. This is likely most important for sessile organisms (Guisan & Thuiller, 2005).

One way to assess the importance and effects of grain on SDM performance is to conduct analyses where grain is changed, and the qualitative and quantitative effects on models are measured. Yet, surprisingly few examples exist (e.g. Boyce *et al.*, 2003), suggesting that the effect of coarsening the grain of predictor variables on SDM performance has not been sufficiently tested. By changing grain size when fitting logistic models for the green woodpecker, Tobalske (2002) observed a clear model improvement at their coarse (400 ha) compared to their fine-grain (100 ha) predictions. Ferrier & Watson (1997) tested the effect of grain size on SDMs on 55 species using four techniques and found, on the contrary, degraded model performances at coarser grain.

A study was started in 2003 to compare a large range of existing techniques for predicting species distributions from presence-only data and assess the effects of several intrinsic and extrinsic factors on model performance (Elith *et al.*, 2006). Here, we analyse (1) the effect of a 10-fold coarsening of grain size on the performance of SDMs and (2) whether any observed effects are dependent on the type of region, modelling technique, or organism considered. In the absence of any well-documented effect, a 10-fold

change was considered a realistic assumption, large enough for most studies and beyond what is otherwise used for aggregations (Huettmann & Diamond, 2006).

METHODS

Species and environmental data

The data used for this modelling study are a subset of the data described in Elith *et al.* (2006). Species and environment data sets, including climatic, topographical, and soil data, are briefly described, and the detailed names and characteristics are provided in Appendices S1 and S2 in Supplementary Material. These are: 10 bird species and 11 GIS environmental predictors from Ontario (CAN); 6 plant species and 4 vertebrate species and 13 GIS environmental predictors from New South Wales (NSW, Australia); 10 plant species (shrubs and trees) and 13 GIS environmental predictors from New Zealand (NZ); 10 plant species and 10 GIS environmental predictors from South America (SA); and 10 tree species and 13 GIS environmental predictors from Switzerland (SWI). Ten species were selected from each larger data set following three broad criteria: (1) a range of sample sizes, (2) a range of geographical distributions, and, where possible, (3) a range of biological groups or life-forms.

Model fitting

Ten predictive techniques were used to fit the species distribution models. These were: (1) the DIVA-GIS implementation of BIOCLIM (Busby, 1991), (2) DOMAIN (Carpenter *et al.* 1993), (3) GLM (generalized linear model; McCullagh & Nelder, 1989), (4) GAM (generalized additive model; Hastie & Tibshirani, 1986), (5) BRUTO (Hastie *et al.* 1994), (6) MARS (multivariate adaptive regression splines; Friedman, 1991), (7) BRT (boosted regression tree; Friedman, 2001), (8) OM-GARP (genetic algorithm for rule-based predictions; a new but as yet unreleased version of the one developed by Stockwell & Peters, 1999), (9) GDMSS (the single species version of the generalized dissimilarity model of Ferrier *et al.*, 2002a), and (10) MAXENT-T (maximum entropy; Phillips *et al.*, 2004, 2006). The first two are profile techniques, the next four belong to the large family of generalized regression techniques, and the last four are all very distinct approaches. All modelling techniques and specific fitting detail are described in more detail in Elith *et al.* (2006). The models were fitted by experienced modellers, usually by those among the authors who knew the technique best. As the training sets only contained presence data, pseudo-absences were generated for those which techniques required them (all except BIOCLIM and DOMAIN), by taking a random sample of 10,000 sites (Elith *et al.*, 2006). All techniques were implemented as described in Elith *et al.* (2006).

Model evaluation

An independent presence-absence data set was available for each study area, which usually had greater locational accuracy than the presence data used to train the models and allowed for a

Table 1 Grid cell size per data set at the original (fine) and 10 times coarse resolutions.

Region	Fine	Coarse
CAN	1 km	10 km
NSW	100 m	1 km
NZ	100 m	1 km
SA	1 km	10 km
SWI	100 m	1 km

Please see Methods for definitions of codes used.

robust test of model performance across grain sizes (Elith *et al.*, 2006). The quantitative evaluation was performed by comparing the predictions to the observations in the independent test data set using the area under the curve (AUC) of a receiver-operating characteristic (ROC) plot as the measure of prediction success (Fielding & Bell, 1997). AUC values range from 0 to 1, where 0.5 indicates that the model has no predictive power, 1 signifies a perfect model, and values below 0.5 would indicate a relationship worse than expected by chance. According to Swets (1988), models providing values > 0.9 are considered highly accurate, those providing values in the range 0.7–0.9 ‘useful’, and those lower than 0.7 ‘poorly accurate’. We also investigated performance with maximized Kappa (Cohen, 1960). However, because this measure yielded similar results to the AUC we only present the latter.

The 10-fold grain experiment

To test the effect of grain on model performance, the initial resolution of each GIS environmental layer (Table 1) was coarsened by a factor 10 times. The species data were then related to the coarser environmental grids to fit a second series of models of distribution for each species, which could be compared to the first series fitted at the initial resolution (initial models vs. coarse-grain models). In any one coarse cell, there may have been more than one species record. Rather than reducing these to one record per cell, we retained all records so that sample size was consistent between the initial and the coarse-grain scenarios. This avoids confounding the effects of change in grain with change in sample size. We also retained all records in the evaluation data set. Modellers were asked to predict with their model to each independent evaluation site, blind to the true presence–absence evaluation observations occurring there, using both the fine and the coarse environmental data.

Aggregating environmental data was performed in the ARCMAP GIS (© Environmental Science Research Institute Inc., Redlands, CA, USA). Continuous predictors were upscaled with the ‘aggregate’ function using the mean function. Categorical predictors were upscaled with the ‘zonalmajority’ function, thus attributing the most represented class (majority) to each coarser cell.

Comparing performance at the two grain sizes

Model performance at the two grain sizes was assessed across (1) the regions (data sets with distinct climate, geographical

position on Earth, extent, etc.), (2) the modelling techniques, and (3) the species types or characteristics, such as group membership (e.g. plant vs. animal species) or sample size. Because the different groups show large differences in the number of their representative species in the data sets for the different regions, we only compared plants and birds which were the two most represented groups. It is also important to note that a confounding of species group by region may occur, as only plant species were present in the SWI, NZ, and SA data sets. An additional characteristic considered for each species was its sample size, measured as the absolute number of occurrences in the training set.

Mean AUC performances per region, technique, or species characteristics were provided at each grain size, together with their relative ranking. Standard statistical tests were used to quantify the differences in AUC performances and allow identification of significant explanatory factors. Paired Wilcoxon tests were used as the data were generally non-normally distributed. For each comparison, we additionally calculated the average direction of change, to assess whether the models were degraded or improved by coarsening the grain.

RESULTS

Comparing the performance of fine-grain models

The performance of the fine resolution models was on average not equal across regions, techniques, and species types (Fig. 1a–c, Table 2a–c). Within regions (Fig. 1a, Table 2a), CAN had significantly lower average performance compared to all other data sets, with the following overall ranking: SWI > SA > NZ > NSW > CAN. Within techniques (Fig. 1b, Table 2b), the two profile techniques — BIOCLIM and DOMAIN — and the genetic algorithm OM-GARP yielded the models with lowest average performance, whereas BRT and MAXENT, and secondarily GAM and other regression techniques, yielded the best performance on average. Within species types (Fig. 1c, Table 2c), PLANTS are better modelled than BIRDS. Model performance slightly increased on average with the number of occurrences in the training set, while at the same time its variance decreases (Fig. 1d, Table 2d, first two columns).

Overall effects of coarsening the grain

Differences in AUC scores between the initial fine-grain models and those with coarsened grain are on average significant when assessed across all regions, techniques, and species types (overall P -value < 0.00001). However, the ranking of data sets, techniques, or species characteristics was not affected by coarsening grain size. Most significant trends are towards model degradation (Table 2). The next paragraphs report on more details patterns observed at the levels of regions, techniques, or species characteristics.

Effects per data sets

Within each data set (Table 2a), differences in AUC between fine- and coarse-grain models reveal a significant effect only for SA and SWI. Changes in AUC in NZ are relatively small (test is not

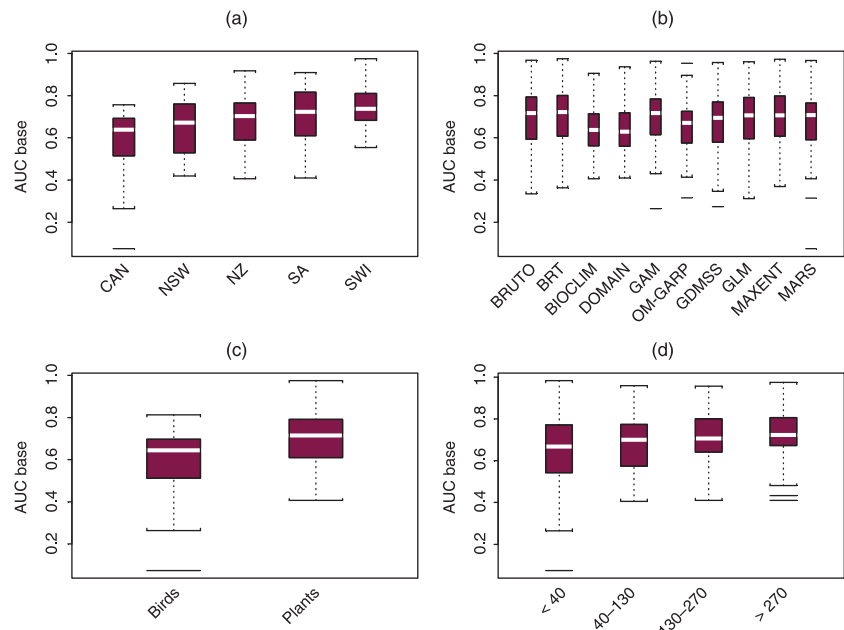


Figure 1 Boxplot of AUC of base models displayed by: (a) data sets, (b) methods, (c) species types, and (d) classes of absolute number of occurrences in the training set. The line across the box indicates the median. The box boundaries show the interquartile range. Whiskers identify extreme data points that are no more than 1.5 times the interquartile range on both sides. Horizontal lines outside the square brackets picture outliers.

Table 2 Mean AUC for the base and 10 times grain experiment (columns 1–4), with *P*-values of paired Wilcoxon tests of the AUC differences between base and grain experiments equal to zero (column 5). Models are classified: (a) per data set, (b) per method, (c) per species types, and (d) per number of occurrences in the training set (NOCC fit). See Figs 1–3. The null hypothesis H_0 is that the difference in model performance is on average not significantly different than zero (bilateral). *P*-values lower than 0.05 signify rejection of H_0 . Values in italics are not significant after Bonferroni correction for multiple testing. Values in bold are significant at the $P < 0.01$ level.

	Mean AUC	Rank base	SD AUC	Mean AUC 10×	Rank 10×	SD AUC 10×	<i>P</i> -value (AUC diff.)
(a) Data sets							
CAN	0.5834	5	0.1343	0.5829	5	0.1427	0.2834
NSW	0.6511	4	0.1258	0.6477	4	0.1227	0.2638
NZ	0.6823	3	0.1193	0.6783	3	0.1177	0.0115
SA	0.7123	2	0.1197	0.7037	2	0.1130	0.0082
SWI	0.7482	1	0.1011	0.7233	1	0.1013	0.0000
(b) Techniques							
BRUTO	0.6928	4	0.1306	0.6859	2	0.1291	0.0067
BRT	0.7040	1	0.1321	0.6993	1	0.1299	0.2060
BIOCLIM	0.6391	9	0.1074	0.6313	10	0.1134	0.0078
DOMAIN	0.6383	10	0.1248	0.6337	9	0.1262	0.1248
GAM	0.6939	3	0.1338	0.6844	3	0.1287	0.0004
OM-GARP	0.6605	8	0.1284	0.6542	8	0.1129	0.8963
GDMSS	0.6726	7	0.1390	0.6731	6	0.1216	0.1476
GLM	0.6838	5	0.1352	0.6546	7	0.1447	0.0049
MAXENT	0.6954	2	0.1241	0.6793	4	0.1216	0.0008
MARS	0.6744	6	0.1583	0.6760	5	0.1543	0.4602
(c) Species types							
Birds	0.6330	2	0.1490	0.6328	2	0.1534	0.5143
Plants	0.7044	1	0.1159	0.6931	1	0.1114	0.0000
(d) NOCC FIT							
< 40	0.6522	4	0.1577	0.6408	4	0.1566	0.0012
40–130	0.6807	3	0.1285	0.6795	3	0.1286	0.2298
130–270	0.7074	2	0.1123	0.7078	2	0.1072	0.0611
> 270	0.7264	1	0.1125	0.7096	1	0.1067	0.0000

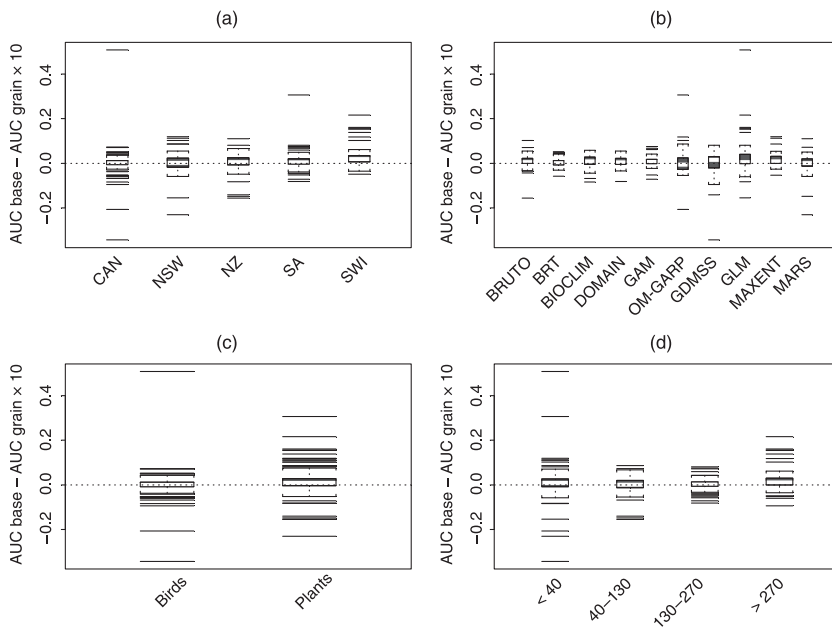


Figure 2 Boxplot of differences in AUC displayed by: (a) data sets, (b) methods, (c) species types, and (d) classes of absolute number of occurrences in the training set. See legend of Fig. 1 for details on boxplots.

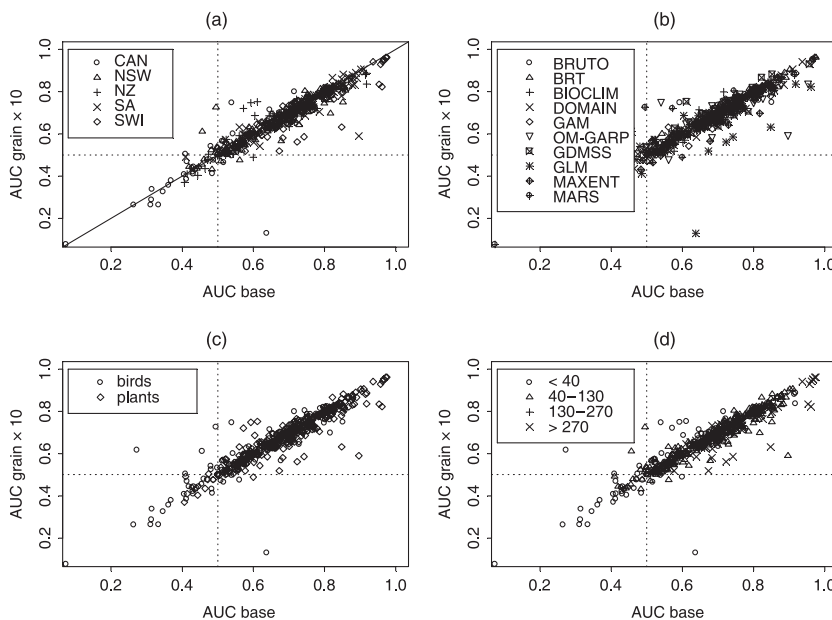


Figure 3 AUC of base models plotted against AUC of 10 times coarsened models, with labels for: (a) data sets, (b) methods, (c) species types, and (d) classes of absolute number of occurrences in the training set. Cases of model degradation are below the 1 : 1 line; cases of improvement are above it. Models below 0.5 on both axes (dotted lines) predict worse than expected by chance. Only outlying points are informative. Points grouped on the main cloud along the 1 : 1 line represent models not affected by grain change that do not need to be individually distinguished.

significant after Bonferroni correction) and there is no evidence for effects in CAN and NSW.

As expected from these results, data sets where models appear most degraded are SWI, SA, and NZ (Fig. 3a). SA and SWI have a large number of points below the 1 : 1 line. Data sets with outlying points on the improvement side include NSW, CAN, and NZ. CAN is the only data set on both sides.

Effects per techniques

Difference in AUC scores between fine- and coarse-grain models only reveal slight to moderate changes across techniques (Table 2b). BIOCLIM, BRUTO, GAM, GLM, and MAXENT all

show some decrease in mean AUC, but only GAM, GLM, and MAXENT are significant with Bonferroni correction (Table 2b). From the *P*-values, MAXENT and GAM are the most sensitive and additionally exhibit low variance across differences (Fig. 2b). Nonetheless, both remain among the four best techniques at coarser grain size. There is no evidence of sensitivity to change in grain size for BRT, DOMAIN, OM-GARP, GDMSS, and MARS. BRT is the best technique at the coarse resolution.

No clear trend can be detected in the scatter plot labelled by techniques (Fig. 3b). Techniques that undergo on average model degradation with increased grain size are OM-GARP, GLM, and GAM. Those techniques that undergo on average model improvement are GDMSS and MARS, but although significant, the trend

is weak in the latter cases. Hence, GAM-type techniques (GAM, BRUTO, and MARS) can both yield improvement or degradation. Based on the observation of outlying points on the graphs, techniques that are associated with more extreme changes are OM-GARP, GDMSS, and GLM.

Effects per species types and sample size

Among species types, change in grain size affects model performance for plants, whereas no significant differences exist for birds (Table 2c). Plants have outlying models towards degradation, whereas bird models exhibit more outliers towards improvement (Fig. 3c).

The number of occurrences in the training set shows a clear relationship with model performance, with performance decreasing linearly from the largest to the smallest class (Table 2d). The effect of change in grain size is more variable for these classes. The variance in performance change is very large in the lowest class and decreases in the three other classes (Fig. 2d). The highest class (> 270) shows a highly significant change in model performance (Table 2d), towards model degradation (Fig. 2d), even after Bonferroni correction. The lowest class (< 40) also shows a significant change in performance (Table 2d), also towards model degradation, but a very large variance overall (Fig. 2d). A gradient is observed for the number of occurrences in the training set, with classes with fewer occurrences being farther away from the 1 : 1 line, and the higher classes being successively closer, following their ordering (3, 2, 1), towards either model improvement or degradation (Fig. 3d).

DISCUSSION

This study represents a broad comparative exploration of the effect of grain on the performance of 10 predictive modelling techniques for predicting species distribution, conducted across five data sets and 50 species spread across different species groups. As presence-only data were used to train the models, this also constitutes a valuable assessment of the use of natural history collection data for fitting biogeographical models (Graham *et al.*, 2004). Our results are primarily valid for discussing a 10-fold grain increase on model performance, starting from the initial resolutions of 100 m and 1 km.

The assessment of model performance on the full sample of species available over all data sets (226 species) is presented elsewhere (Elith *et al.*, 2006). Results from this subset of 50 species did not substantially change the overall ranking of techniques or data sets evaluated by Elith *et al.* (2006). Boosted regression tree (BRT) and the maximum entropy approach (MAXENT) still prove among the top three techniques, and techniques of the regression family — GLM, GAM, BRUTO, and MARS — come next and perform very similarly to each other (ranks 3–6). One interesting approach — generalized dissimilarity modelling, single-species implementation (GDMSS) — is ranked seventh here, whereas it was ranked fourth in the general assessment. The ranking of the data sets is exactly the same as in Elith *et al.* (2006), with SWI first and CAN last. Therefore, we refer to the more com-

plete Elith *et al.* (2006) assessment for comparison of techniques and data sets, and put our main emphasis on grain change.

Overall, changing grain size did not change model performance as much as one might have expected. Although significantly degrading the models on average, across all data sets, techniques, and species, it only proved very significant for one data set (SWI) and just or marginally significant for two additional ones (SA and NZ). As these three data sets also had the best models in term of predictive performance (SWI > SA > NZ), this could suggest that models of sufficient predictive power are required to display such scale effect. That is, models that already lack predictive power do not include detail that might be degraded by an increase in grain. Unfortunately, we do not have quantification of the locational accuracy of the modelling records in this study, but we suspect that the SWI records are more accurately located than others. Further research is required on this topic, for example, examining changes in model performance across a wider range of grain sizes, controlling for prevalence, and considering the types of environmental variables and the scale at which they vary in the landscape. A more detailed biological interpretation of species requirements would also be useful. It is also worth noting here — as reported elsewhere (Elith *et al.*, 2006; A.G., unpublished data) — that differences between species were, on average, much higher than between techniques.

Concerning techniques, BRT was the least affected by grain change. BRT, MAXENT, GAM, and BRUTO were the four best techniques at coarser grain size, but unlike BRT, the other three degraded significantly between the fine and coarse-grain treatments. OM-GARP performed far less well but shared with BRT the same overall insensitivity to grain change. Other techniques, such as BRUTO, MAXENT, GAM, and GLM, also perform well but prove more sensitive to grain change. BIOCLIM performed the worst and was sensitive to grain change.

Models with a large number of occurrences in the training set performed, on average, better and had smaller variances than models built with few occurrences. These results indicate that poorly performing models are less likely to be fitted for species with a sufficient number of occurrences. Models for species with highest sample size in the training set are also the most sensitive to grain change. This might be because such an effect is mainly detectable for models with sufficient predictive performance; otherwise, model instability may hide such a response.

Our main result — i.e. a slight average trend towards model degradation at coarser grain size — can be compared to the results of others. By changing the grain size but not the extent when fitting logistic models for the green woodpecker (*Picus viridis*), Tobalske (2002) observed a model improvement at the coarser grain (400 ha) compared to the finer grain (100 ha). While this contrasts to the overall trend of degradation in our study, there were examples of some species that did show improved prediction at coarser grain. Tobalske (2002) only used a single species in a small study area (c. 32,000 ha) in Switzerland, applying a 2-fold grain change, whereas we tested a 10-fold change on 50 species over five much larger areas (from national to continental extents). Hence, the Tobalske result could be viewed as one example of the range of responses that are possible with change in grain.

Similarly, Johnson *et al.* (2002) compared models fitted with various sets of predictors at three scales (site-specific, local landscape — 2-km radius, broad landscape — 10-km radius). They found contrasting results, with each of the three scales proving potentially the best for predicting occurrence of a different species. Thomas *et al.* (2002) compared models fitted with fine- and coarse-resolution predictors. They showed that both sets of predictors performed equally well for fitting vegetation models in the Mojave Desert. However, totally distinct sets of predictors were used at the two resolutions (field vs. GIS) and thus no aggregation of the predictors was performed in their study. Ferrier & Watson (1997) used similar aggregation procedure as used here to coarsen the resolution from 200 m to 5 km grid cells and found no significant effect on the performance of SDMs fitted on presence-only data. In contrast, effects for presence-absence data were significant. They conclude that improving the resolution of environmental data will not necessarily improve the performance of presence-only models if biological data are not also georeferenced at a comparably fine resolution.

Thus it seems from our results and from other published studies that all three effects — improvement, no change, or degradation of models — can be obtained when changing grain size, and other confounding factors must play a role. Previous studies suggested the configuration of the study area and species identity (e.g. type of organism or rarity) as important causes of discrepancy between models, possibly yielding differential responses to a change in grain size of similar amplitude (e.g. 2× or 5×).

CONCLUSION AND PERSPECTIVES

1 Change in grain size does not have a substantial effect on species distribution models. The trend is overall weakly significant towards degradation of model performance, but improvement can also be observed for some species.

2 Changing grain size does not equally affect model performance across regions, techniques, and species; the strongest effect is on regions and species types, with plant species in Switzerland and South America being most affected. Bird models are more similar at the two resolutions.

3 Modelling techniques are also differentially affected, with MAXENT and GAM models being most degraded. However, changing grain size does not change much the ranking of techniques, with BRT, MAXENT, and GAM being the best performers at both resolutions.

4 Overall, the effect of change in grain size seems mostly noticeable for models: (1) that reach sufficient performance, (2) based on sufficient sample size, or (3) fitted with fine-grain initial data that have an intrinsic error lower than the coarser grain size.

Our study and results also highlighted some of the difficulties in evaluating the effect of changing grain size with real data, and some controlled studies with simulated data might provide a useful complementary view (see Austin *et al.*, 2006). Future steps would also be to look more carefully at the shapes of predictors' response curves and the variables selected in the models at the various grain sizes, and testing a larger range of changes in grain size.

ACKNOWLEDGEMENTS

We thank the National Center for Ecological Analysis and Synthesis (NCEAS; <http://nceas.ucsb.edu>) in Santa Barbara for supporting the three workshops that generated these results (project 'Testing alternative methodologies for modelling species' ecological niches and predicting geographical distributions', granted to A.T. Peterson and C. Moritz) and Jorge Soberón, John Leathwick, Robert Anderson, and Alexandre Hirzel for useful comments during some of the workshops. We further thank the following people and organizations for supplying species data: the Royal Ontario Museum (ROM), M. Peck and M. Cadman at Bird Studies Canada for CAN data; the National Vegetation Survey Databank and the Allan Herbarium for NZ data; Robert Magill, Trisha Consiglio, and the Missouri Botanical Garden for SA data; and T. Wohlgemuth and U. Braendli for the SWI NFI and forest plots data. Comments from Wilfried Thuiller and an anonymous reviewer helped to improve the manuscript. We finally thank the Universidad Internacional de Andalucía, sede Antonio Machado for organizing the workshop 'Predictive modelling of species distributions, new tools for the XXI century', which led to the special issue including this paper.

REFERENCES

- Araújo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2005) Downscaling European species atlas distribution to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, **14**, 17–30.
- Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. & Luoto, M. (2006) Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. *Ecological Modelling*, **199** (2), 197–216.
- Boyce, M., Mao, J.S., Merrill, E.H., Fortin, D., Turner, M.G., Fryell, J. & Turchin, P. (2003) Scale and heterogeneity in habitat selection by elk in Yellowstone National Park. *Ecology*, **84** (4), 421–431.
- Busby, J.R. (1991) BIOCLIM — A bioclimate analysis and prediction system. *Nature conservation: cost effective biological surveys and data analysis* (ed. by C.R. Margules and M.P. Austin), pp. 64–68. CSIRO, Melbourne.
- Carpenter, G., Gillison, A.N. & Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, **2**, 667–680.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20** (1), 37–46.
- Côté, I.M. & Reynolds, J.D. (2002) Predictive ecology to the rescue? *Science*, **298** (5596), 1181–1182.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S., Richardson, K., Schachetti Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., Wisz, M. & Zimmermann, N.E. (2006) Novel methods improve predictions

- of species' distributions from occurrence data. *Ecography*, **29** (2), 129–151.
- Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263–274.
- Ferrier, S. & Watson, G. (1997) *An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity*. Environment Australia, Canberra.
- Ferrier, S., Drielsma, M., Manion, G. & Watson, G. (2002a) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, **11** (12), 2309–2338.
- Ferrier, S., Watson, G., Pearce, J. & Drielsma, M. (2002b) Extended statistical approaches to modelling spatial patterns in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation*, **11** (12), 2275–2307.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24** (1), 38–49.
- Friedman, J. (1991) Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, **19** (1), 1–141.
- Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, **29**, 1189–1232.
- Gottschalk, T., Huettmann, F. & Ehlers, M. (2005) Thirty years of analysing and modelling avian habitat relationships using satellite imagery data: a review. *International Journal of Remote Sensing*, **26** (12), 2631–2656.
- Graham, C.H., Ferrier, S., Huettmann, F., Moritz, C. & Peterson, A.T. (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, **19** (9), 497–503.
- Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species richness. *Global Ecology and Biogeography Letters*, **15**, 578–587.
- Guisan, A. & Hofer, U. (2003) Predicting reptile distributions at the mesoscale: relation to climate and topography. *Journal of Biogeography*, **30** (8), 1233–1243.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135** (2–3), 147–186.
- Hastie, T. & Tibshirani, R. (1986) Generalized additive models. *Statistical Sciences*, **1** (3), 297–318.
- Hastie, T., Tibshirani, R. & Buja, A. (1994) Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association*, **89** (428), 1255–1270.
- Hernandez, P.A., Graham, C.H., Master, M.M. & Albert, D. (2006) The effect of sample size on the performance of species distribution models. *Ecography*, **29**, 773–785.
- Holling, C.S. (1992) Cross-Scale morphology, geometry, and dynamics of ecosystems. *Ecological Monographs*, **62**, 447–502.
- Huettmann, F. & Diamond, A.W. (2006) Large-scale effects on the spatial distribution of seabird in the Northwest Atlantic. *Landscape Ecology*, **21** (7), 1089–1108.
- Johnson, C.M., Johnson, L.B., Richard, C. & Beasley, V. (2002) Predicting the occurrence of amphibians: an assessment of multiple-scale models. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, P.J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael and B. Wall), pp. 157–170. Island Press, Covelo, California.
- Linke, J., Franklin, S.E., Huettmann, F. & Stenhouse, G.B. (2005) Seismic cutlines, changing landscape metrics and grizzly bear landscape use in Alberta. *Landscape Ecology*, **20**, 811–826.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*, 2nd edn. Chapman & Hall, London.
- McPherson, J.M., Jetz, W. & Rogers, D.J. (2006) Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions — possibilities and limitations. *Ecological Modelling*, **192**, 499–522.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190** (3–4), 231–259.
- Phillips, S.J., Schapire, R.E. & Dudik, M. (2004) A maximum entropy approach to species distribution modelling. *Machine Learning, Proceedings of the Twenty-First International Conference (ICML 2004)* (ed. by C.E. Brodley), pp. 655–662. ACM, Banff, Alberta, Canada.
- Stockwell, D. & Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13** (2), 143–158.
- Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148** (1), 1–13.
- Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- Thomas, K.A., Keeler-Wolf, T. & Franklin, J. (2002) A comparison of fine- and coarse-resolution environmental variables toward predicting vegetation distribution in the Mojave desert. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, P.J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael and B. Wall), pp. 133–139. Island Press, Covelo, California.
- Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T. & Prentice, I.C. (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the USA*, **102** (23), 8245–8250.
- Tobolske, C. (2002) Effects of spatial scale on the predictive ability of habitat models for the green woodpecker in Switzerland. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, P.J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael and B. Wall), pp. 197–204. Island Press, Covelo, California.
- Wiens, J.A. (2002) Predicting species occurrences: progress, problems, and prospects. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 739–749. Island Press, Covelo, California.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1 Selected species and their occurrence in each data set at the original, fine resolution

Appendix S2 Environmental GIS predictors associated to each data set.

This material is available as part of the online article from:
<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1472-4642.2007.00342.x>
(This link will take you to the article abstract).

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