



Uncertainty associated with survey design in Species Distribution Models

Geiziane Tessarolo^{1,2*}, Thiago F. Rangel¹, Miguel B. Araújo^{2,3,4} and Joaquín Hortal^{1,2}

¹Departamento de Ecologia, Instituto de Ciências Biológicas, ICB, Universidade Federal de Goiás, UFG Campus II, Goiânia, GO 74001-970, Brazil, ²Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (CSIC), C/José Gutiérrez Abascal 2, Madrid 28006, Spain, ³Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, SL5 7PY Berks, UK, ⁴Research Network in Biodiversity and Evolutionary Biology (InBIO), Research Center in Biodiversity and Genetic Resources (CIBIO), University of Évora, Largo dos Colegiais, Évora 7000, Portugal

*Correspondence: Geiziane Tessarolo, Departamento de Ecologia, Instituto de Ciências Biológicas, ICB, Universidade Federal de Goiás, UFG Campus II, Goiânia, GO 74001-970, Brazil.
E-mail: geites@gmail.com

ABSTRACT

Aim Species distribution models (SDM) can be used to predict the location of unknown populations from known species occurrences. It follows that how the data used to calibrate the models are collected can have a great impact on prediction success. We evaluated the influence of different survey designs and their interaction with the modelling technique on SDM performance.

Location Iberian Peninsula.

Methods We examine how data recorded using seven alternative survey designs (random, systematic, environmentally stratified by class and environmentally stratified using *P*-median, biased due to accessibility, biased by human density aggregation and biased towards protected areas) could affect SDM predictions generated with nine modelling techniques (BIOCLIM, Gower distance, Mahalanobis distance, Euclidean distance, GLM, MaxEnt, ENFA and Random Forest). We also study how sample size, species' characteristics and modelling technique affected SDM predictive ability, using six evaluation metrics.

Results Survey design has a small effect on prediction success. Characteristics of species' ranges rank highest among the factors affecting SDM results: the species with lower relative occurrence area (ROA) are predicted better. Model predictions are also improved when sample size is large.

Main conclusions The species modelled – particularly the extent of its distribution – are the largest source of influence over SDM results. The environmental coverage of the surveys is more important than the spatial structure of the calibration data. Therefore, climatic biases in the data should be identified to avoid erroneous conclusions about the geographic patterns of species distributions.

Keywords

Climatic bias, environmental niche models, Iberian Peninsula, model performance, relative occurrence area, sample size.

INTRODUCTION

Species distribution models (SDM) relate incomplete information about the occurrence of species (and sometimes about their absence) with environmental predictors, to generate spatially explicit predictions about their geographic distributions (Araújo & Guisan, 2006; Franklin & Miller, 2009; Peterson *et al.*, 2011; Hortal *et al.*, 2012). Despite their increasing use during the last 15 years (Thuiller *et al.*, 2009; Lobo *et al.*, 2010; Varela *et al.*, 2014), SDMs pose many conceptual problems (Araújo & Guisan, 2006; Soberón, 2007, 2010; Jiménez-Valverde *et al.*, 2008; Colwell & Rangel, 2009; Soberón & Nakamura, 2009; Hortal *et al.*, 2012) and

encompass a number of methodological uncertainties (Barry & Elith, 2006; Heikkinen *et al.*, 2006; Rocchini *et al.*, 2011; Beale & Lennon, 2012).

One of the assumptions of SDMs is that the data used for model calibration (i.e. the samples of presence, or presence and absence) are free of bias. When this assumption is violated by biases in the collection of data, the accuracy of model predictions can be affected (Hortal *et al.*, 2008; Lobo, 2008; Loiselle *et al.*, 2008; Rocchini *et al.*, 2011). In fact, the design of the surveys can have a large impact on SDM performance (Hirzel & Guisan, 2002; Edwards *et al.*, 2006; Albert *et al.*, 2010; Braunisch & Suchant, 2010). It follows that an adequate spatial design of the surveys increases the

value of biodiversity data collections to answer different questions in ecology, evolution and biogeography (Hortal & Lobo, 2005; Albert *et al.*, 2010). Many methods to design surveys have been proposed with the objective of maximizing the amount of biodiversity captured, while incorporating time and cost limitations (Austin & Heyligers, 1989; Pereira & Itami, 1991; Hirzel & Guisan, 2002; Funk *et al.*, 2005; Hortal & Lobo, 2005; Medina *et al.*, 2013). However, even planned sample designs can vary in the efficiency with which they detect biodiversity patterns (Sastre & Lobo, 2009).

Perhaps more importantly, good quality data coming from standardized surveys are rare or even lacking for most regions and species (Rocchini *et al.*, 2011). Rather, biodiversity databases include heterogeneous information coming from inventories developed with a variety of objectives (Hortal *et al.*, 2007). The absence of standardized sampling schemes often generates bias in the resulting distributional data. It is well known that the records of occurrence and/or absence of species are more frequent in more accessible locations (i.e. near major road routes, urban areas or the work centres of the taxonomists) and/or classical localities (e.g. national parks or other protected areas) that are repeatedly sampled over time (Dennis & Thomas, 2000; Kadmon *et al.*, 2004; Romo *et al.*, 2006; Hortal *et al.*, 2007). These geographical biases in the survey effort may often result in historical and climatic biases (Hortal *et al.*, 2007, 2008; Lobo *et al.*, 2007a). Such biases can yield incomplete and potentially truncated characterizations of species realized niches (Hortal *et al.*, 2008; Rocchini *et al.*, 2011), with the consequent critical effect on the ability of SDM to describe environmental limits of species' distributions (Austin & Heyligers, 1989; Thuiller *et al.*, 2004; Albert *et al.*, 2010; Hortal *et al.*, 2012).

Despite the potential effects of spatial biases in the design of the surveys on SDM prediction ability (Araújo & Guisan, 2006; Phillips *et al.*, 2009), this issue has not been addressed in the literature. Here, we try to overcome this gap by systematically evaluating the influence of different survey design strategies on the performance of SDM predictions, as well as the potential interactions between survey designs and SDM techniques. Further, besides survey design we evaluate the magnitude of influence on predictive accuracy of other factors that may affect SDM performance, namely sample size, modelling technique and species modelled. To do this, we use distribution data for the 34 Iberian endemic terrestrial vertebrate species, simulating calibration datasets with different survey designs and levels of sample size within the Iberian Peninsula, and evaluating the ability of different SDM techniques to interpolate their geographical distributions based on these datasets.

METHODS

Species data

We used data on the whole extent of the distribution of 34 Iberian endemic terrestrial vertebrate species (15 amphibian,

12 reptile and seven mammal species) in the 5919 UTM cells of 10×10 km that conform the Iberian Peninsula. These data provide accurate representations of current distributions and were compiled from national databases in the context of a recent study examining climate change impacts on Iberian terrestrial fauna (Araújo *et al.*, 2011). Cells where a species is not recorded were considered as a true absence of the species. The complete list of species and information about how they were chosen can be found in Table S1 and Appendix S1.

Environmental data

GIS data on climate and topography were obtained from Worldclim (Hijmans *et al.*, 2005) and Global Resource Information Database – United Nations Environment Programme (UNEP/GRID, <http://www.grid.unep.ch/data/data.php>). To reduce collinearity among variables, we used a principal components analysis (PCA) to guide the selection of a subset of variables among the 29 available (see also Baselga & Araújo, 2009). The first two axes accounted for 73% of total variability; the rest of the axes were not significant according to a broken-stick criterion and were discarded. We therefore selected the two variables that accounted for most of the variability in each one of the first two axes: Mean temperature of the warmest quarter, and precipitation of the driest, coldest and wettest quarters. These four environmental layers were used as the basis for the simulation of the stratified survey designs and the calibration of all SDMs. The location and density of roads was extracted from a commercial database maintained by Tele Atlas (<http://www.tomtom.com/>), and data on human population density (in year 2000) from the Gridded Population of the World Project (CIESIN C.U., 2005). Finally, data on the location and limits of protected areas were extracted from the online database of the Natura2000 European network (<http://ec.europa.eu/environment/nature/nature2000/>) and the World Database on Protected Areas (IUCN, 2009). All these GIS data were reprocessed into the UTM 10×10 km Iberian grid developed by EDIT Geoplatform (<http://edit.csic.es/>; Sastre *et al.*, 2009).

Survey designs

We evaluated the effectiveness of seven alternative survey design strategies. The descriptions of each survey design are as follows (further details in Appendix S1):

- *Random*. Surveys taken without any constraint or bias. Each cell in the study area had equal probability of being sampled.
- *Systematic*. A planned survey free of any bias. For each run, one cell was chosen randomly, and then, the remaining cells were chosen at regular distance intervals starting from the initial cell throughout the study area.
- *Stratified*. Surveys stratified without bias along environmental gradients, following two strategies:
 - (i) *Stratified by class*: Several non-overlapping environmental domains (i.e. environmental classes; 10 groups) were defined

using the environmental data of each cell as descriptors. Then, equal numbers of cells were randomly selected within each environmental class.

(ii) *P-median stratification*: Here, cells were selected to maximize the coverage of the environmental and spatial variation within the study area, as described by the matrices of environmental and spatial distances between cells (Funk *et al.*, 2005; Hortal & Lobo, 2005).

- *Biased*. In biased surveys, some sites are less likely to be selected than others without any systematic criterion. We simulated the bias in the selection of cells produced by three different factors:

(iii) *Accessibility*: Cells close to roads had a higher chance of being surveyed. The probability of selecting a cell depended on its geographic distance to the nearest national road.

(iv) *Aggregation by human density*: Here, we simulate the higher survey effort devoted to the areas nearby urban centres. An initial number of cells (anchor cells) were chosen based on population density. To simulate aggregation of samples, the remaining cells were selected based in their proximity to the anchor cells; being the closer ones the most likely to be selected.

(v) *Protected areas*: Surveys conducted only in protected areas, a common practice on biodiversity research. Cells that had at least 30% of their area protected were chosen (to ensure that protected areas cover the great majority of species present in the cell), and samples were randomly selected among them.

Selection and characteristics of the sample datasets

We extracted samples of five different sizes (1%, 5%, 10%, 20% and 25%) from the full extent of the data domain (5919 10 × 10 km cells in the Iberian Peninsula), using each one of the seven survey designs described above. We ran 50 simulations of each one of the 35 combinations of survey design and sample size (Fig. 1). To describe the quality of the data provided by each class of survey design, we measured the climatic bias of each sample in relation to the environmental conditions of the study area (Kadmon *et al.*, 2003; Hortal *et al.*, 2008). We calculated two characteristics related to the species' distribution: prevalence and Relative Occurrence Area (ROA; Jiménez-Valverde *et al.*, 2008; Lobo, 2008). We also evaluated the number of times that a survey design failed to identify enough occurrences to generate models in each SDM technique. That is, every time that a model was not generated for a given species due to insufficient number of presences. See Appendix S1 for more details.

Species distribution modelling

To make predictions about the distributions of species, we used an ensemble of eight SDM techniques, as well as the combined consensus prediction (Araújo & New, 2007), using the BIOENSEMBLES platform for computer-intensive ensemble forecasting (Diniz-Filho *et al.*, 2009; Rangel *et al.*,

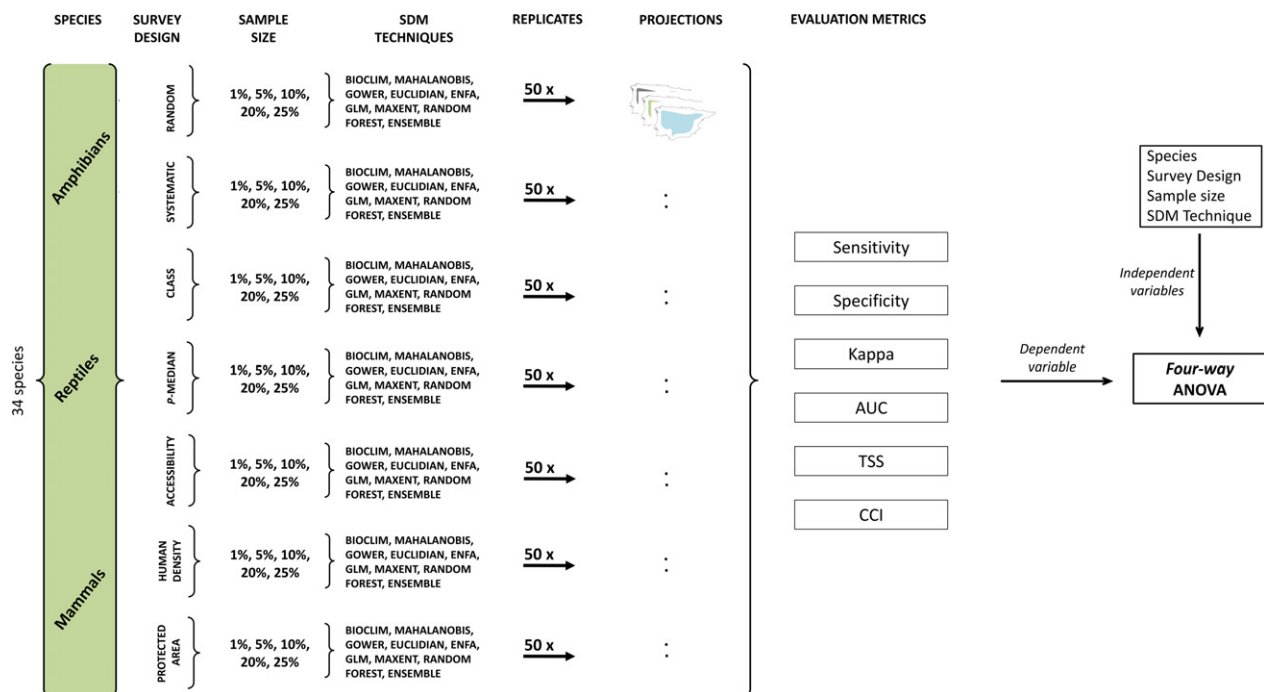


Figure 1 Schematic representation of the protocol used to generate predictions that were used in the evaluation of the effect of survey design on SDM (Species Distribution Models). For each species, we simulated 50 samples based on each survey design and sample size and used these data to generate projections based on each modelling technique, as well as their ensemble. The six assessment metrics calculated for each map were used as dependent variable in six different four-way ANOVAs, using survey design, species, sample size and SDM technique as independent variables.

2009). SDM techniques were chosen to maximize the variety of strategies of statistical adjustment and type of input data currently available and included: BIOCLIM (Busby, 1986), Gower distance (Carpenter *et al.*, 1993), Mahalanobis distance (Farber & Kadmon, 2003), Euclidean distance, Generalized Linear Models (GLM; McCullagh & Nelder, 1989), Maximum Entropy Modelling (MaxEnt; Phillips *et al.*, 2006), Ecological-Niche Factor Analysis (ENFA; Hirzel *et al.*, 2002) and Random Forest (Breiman, 2001). To generate the species distribution maps with these techniques, each combination of survey design and sample size was simulated 50 times (Fig. 1). Each of these simulations consisted in a new selection of calibration dataset following the rules of each type of survey design. See Appendix S1 for more information.

Statistical analyses

We used six metrics to analyse the performance of the different SDM techniques: sensitivity; specificity; kappa (Cohen, 1960), Area under the ROC Curve (AUC; Fielding & Bell, 1997); true skill statistic (TSS; Allouche *et al.*, 2006) and percentage of correctly classified instances (CCI; Fielding & Bell, 1997). These evaluation metrics include the most-used measures of SDM performance. We also evaluated the Klocation metric (Pontius, 2000; Geri *et al.*, 2011), which is similar to kappa but takes into account the spatial location of errors. Given that its results were similar to kappa, we present only the results of this latter technique for the ease of comparison with former studies. All metrics were calculated based on the actual distributions of the species. Here, the evaluation was performed by comparing the binary predictions of presence and absence with the data from the distribution Atlases, after excluding the data used to calibrate the model. The relationships between all evaluation metrics and survey design, species, sample size and modelling technique were assessed using a four-way ANOVA (Fig. 1). Details of the performed ANOVA can be found at Appendix S1.

RESULTS

Factors affecting SDM performance varied across the different evaluation metrics. For five of six evaluation metrics, the species modelled was the most important factor affecting model performance, followed by sample size and SDM technique (Table 1). The contribution of survey design, though

significant, is much less important than any of these three other factors (Table 1). Contrary to our expectations, predictions of species potential distributions generated from different survey designs showed only comparatively small differences in their variance across evaluation metrics, compared with other factors. Interestingly, the degree of climatic bias was generally low for all survey designs. The most biased design was the one that follows human density (bias around 22%), while systematic surveys were the least biased (Fig. 2). Perhaps as a consequence, samples based on designs biased by human density performed worse for sensitivity, kappa, TSS and AUC, while those stratified by groups produced the better survey designs according to the same metrics (Fig. 3; Supporting information Fig. S1). Increasing sample size led to increasing SDM performance according to all metrics, although this pattern was clearer for Sensitivity and TSS (Fig. 4; Supporting information Fig. S2). Given that for most factors Sensitivity and Specificity were the most informative we show their results in the main text. The rest are available at the Supporting information (Figs. S1 to S5).

With regard to SDM techniques, consensus predictions, GLM and MaxEnt ranked nearly always among the highest performing techniques. However, consensus predictions included some of the lowest performing predictions according to AUC, and GLM was below average for specificity (Fig. 5; Supporting information Fig. S3). Random Forest performed the best according to sensitivity and the worst for specificity, confirming a tendency for overfitting of the data. The opposite pattern was recorded for BIOCLIM, also confirming a tendency for underfitting (Supporting information Fig. S3). Random Forest was the technique the most affected by sample size, followed by MaxEnt and GLM, while consensus predictions were the least affected according to Specificity and CCI (Supporting information Fig. S4). Interestingly, the performance of some SDM techniques was unaffected by sample size across all metrics; according to Specificity and CCI, the interaction between sample size and technique causes a decrease in the performance of BIOCLIM, Gower distance and ENFA (Supporting information Fig. S4).

Despite the large influence of the species being modelled on model performance, we were not able to identify a particular group of species for which SDM performance was consistently better or worse than for others (Supporting information Fig. S5). We performed additional correlation analyses between Relative Occurrence Area

Table 1 Fvalues for each factor and assessment metric. The values show the mean of 10,000 ANOVA results per metric

Factor	Evaluation metric					
	Sensitivity	Specificity	Kappa	TSS	CCI	AUC
Species	1893.15	9743.14	21114.05	19954.83	12582.46	3156
SDM technique	1280.90	799.66	1426.80	1389.54	804.46	103578.07
Sample Size	1653.14	607.40	2589.35	4718.27	1010.93	3673.84
Survey Design	1683.43	371.23	798.87	1499.87	150.15	196.26

SDM, Species distribution models; TSS, True skill statistic; CCI, Correctly classified instances; AUC, Area under the ROC curve.

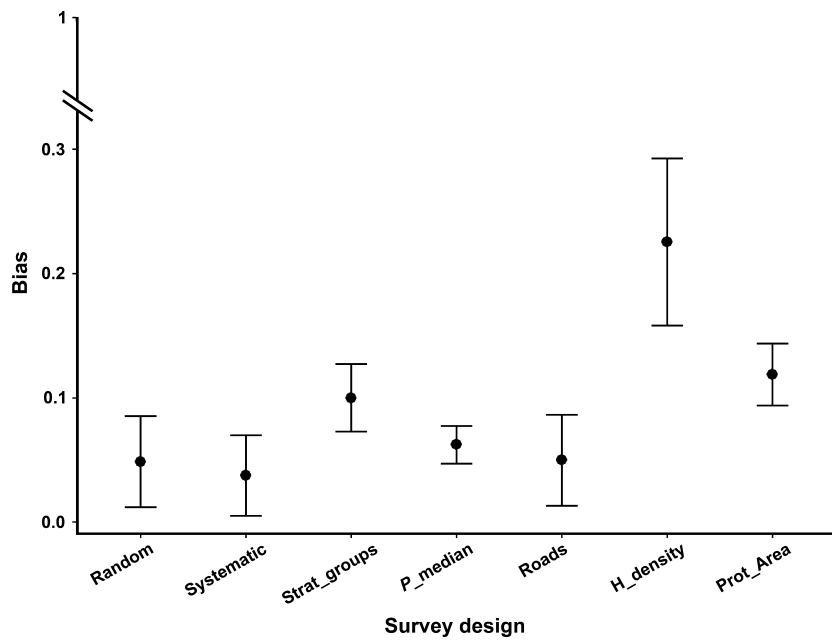


Figure 2 Median and standard deviation of the climatic bias (measured as the sum of differences in the frequencies) in the samples obtained with each survey design. Strat_groups, stratified by groups; P_median, Stratified by P-median; H_density, bias by human density; Prot_Area, bias by protected areas.

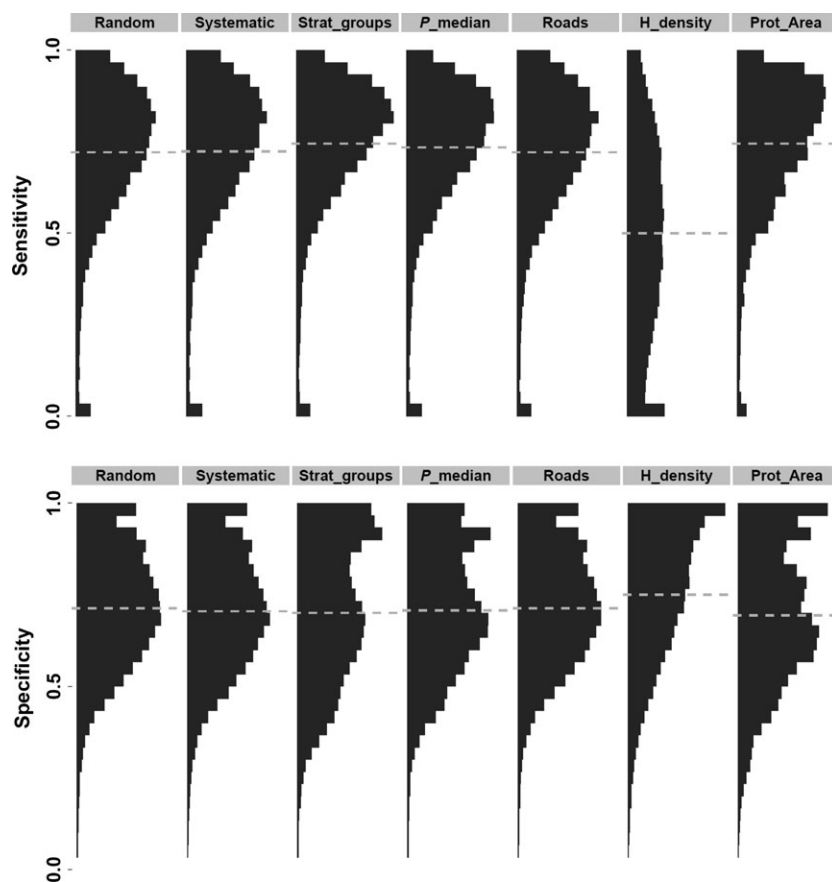


Figure 3 Performance of SDM (Species Distribution Models) predictions generated from data based on different survey designs, according to sensitivity and specificity. Survey design codes as in Fig. 2. The histograms represent the frequency in each class, and the grey lines indicate the mean. For other metrics of SDM performance see supporting information Fig. S1.

(ROA), prevalence of species and all validation metrics (Table 2). Both ROA and prevalence showed negative correlations with all metrics except AUC, so that species with higher prevalences (i.e. more occurrences within the dataset) generated models with lower performance. Specificity, TSS

and CCI detected much stronger negative effects of prevalence and ROA on predictions; this is perhaps not surprising, because these metrics also identify species as the main factor affecting SDM results, thus evidencing that the particular characteristics of each species' geographical range has large

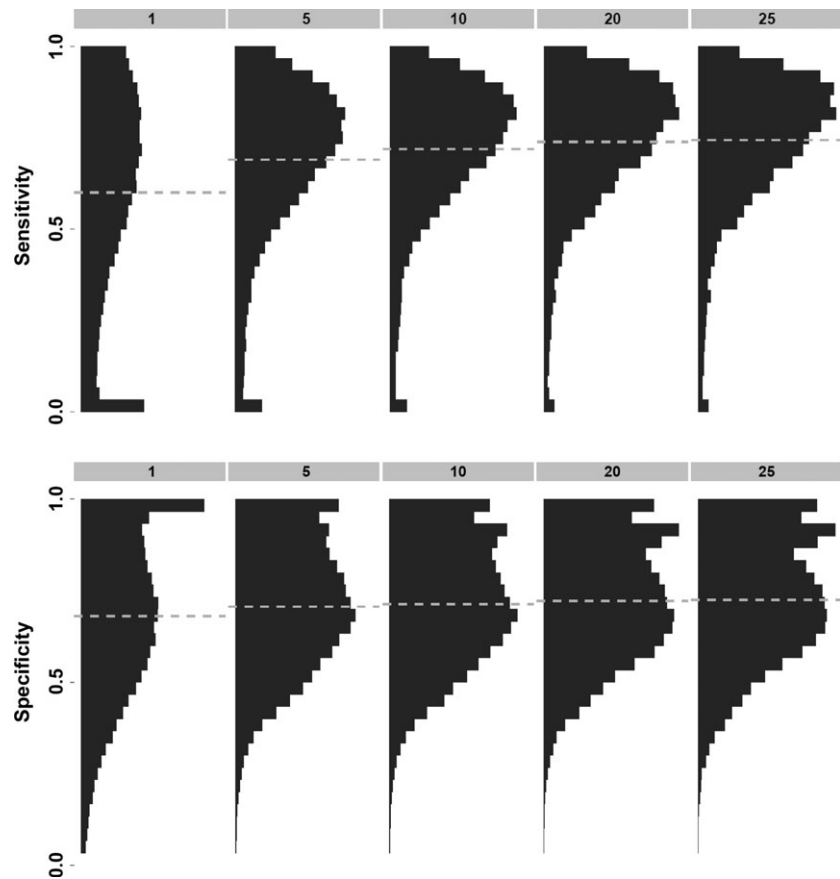


Figure 4 Performance of SDM (Species Distribution Models) predictions generated from datasets of five sample sizes (representing the percentage of cells selected from the whole study area, from 1 to 25%), according to sensitivity and specificity. The histograms represent the frequency in each class and the grey lines indicate the mean. For other metrics of SDM performance see supporting information Fig. S2.

impacts on SDM performance. The number of times that a survey design did not generate sufficient occurrences to allow the modelling decreased with the increase in sample size. In all cases, the values of these failures were highest for the biased human density design, particularly at the smaller sample size; almost 40% of the species did not present enough occurrences to be modelled at such sample size (Supporting information Fig. S6).

DISCUSSION

How data are collected is supposed to be of critical importance for the development of species distribution models (Araújo & Guisan, 2006). The data used to calibrate models are known to lead to differences in SDM predictions thus being expected to affect their performance (e.g. Kadmon *et al.*, 2003; Barry & Elith, 2006; Sánchez-Fernández *et al.*, 2011). However, as shown here for the first time, the importance of data collection is dwarfed when compared with other factors affecting SDM predictions. Indeed, contrary to our own expectations, variation in survey design had relatively small, though significant, effects on the performance of SDM. Instead, species identity ranked highest among factors affecting the models, for all performance metrics (except AUC), followed by sample size and SDM technique.

The main assumption of SDMs is that species distributions are limited by environmental – typically climatic – factors

(Araújo & Peterson, 2012). It follows that restricting the range of environmental variation in which a species occurs could potentially affect the calculation of species-climate response curves which, in turn, could cause projections to be potentially erroneous (Thuiller *et al.*, 2004). Surveys providing a comprehensive coverage of the environmental conditions matching species distributions, like the *P*-median or the stratified sampling by groups are suited for SDM applications because they reduce climatic biases in the data. However, neither the samples taken with these survey strategies showed the lowest values of climatic bias, nor did models based on them present significantly better predictive ability. Rather, most survey designs yielded models with very similar predictive accuracy and showed little differences in climatic bias (except for human density, see below).

One possible explanation for this unexpected result is that the spatial biases simulated in our analyses did not produce large differences in climatic biases. Surveys based in non-random and non-stratified designs are often expected to be spatially and environmentally biased (Hortal *et al.*, 2007; Loiselle *et al.*, 2008), but this is not always the case. In our simulated surveys, sampling along roads had values of climatic bias similar to random samples, and only slightly higher than systematic surveys. This implies that, at the scale and extent analysed, the spatial location of roads in the Iberian Peninsula covers a large climatic variability, even with small sample sizes. Thus, the spatial bias in the surveys due

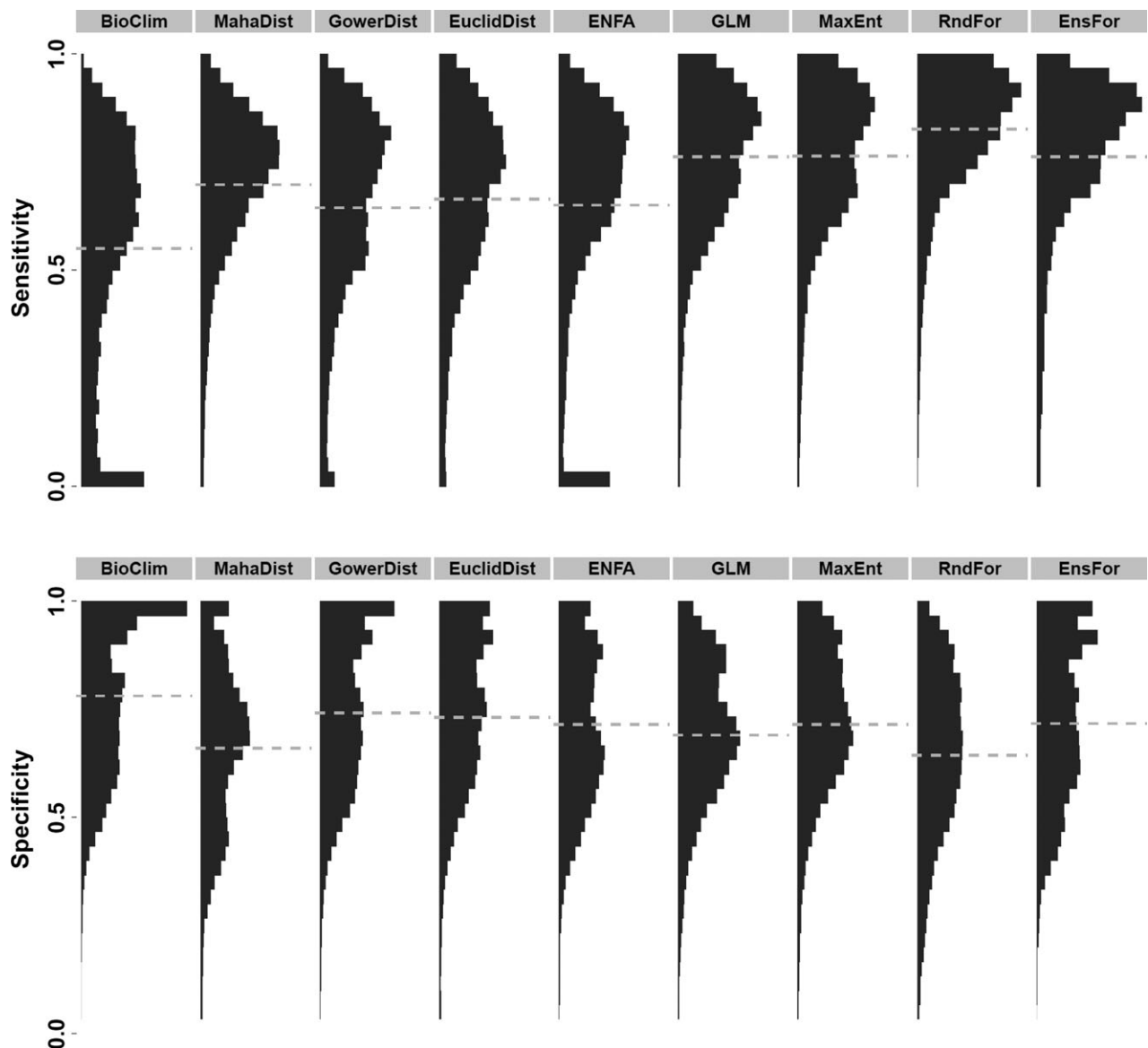


Figure 5 Predictive performance of different SDM (Species Distribution Models) techniques according to sensitivity and specificity. The histograms represent the frequency in each class and the grey lines indicate the mean. For the other metrics of SDM performance see supporting information Fig. S3.

Table 2 Pearson correlations (r) between the evaluation metrics and the prevalence on training dataset and ROA. All correlations were significant at $P > 0.001$

Metric	Prevalence		ROA	
	t	R	t	r
Sensitivity	−693.539	−0.09	−134.36	−0.19
Specificity	−317.905	−0.41	−525.7	−0.6
Kappa	−685.686	−0.097	−165.2	−0.23
TSS	−289.062	−0.38	−516.06	−0.59
CCI	−271.893	−0.36	−544.27	−0.61
AUC	13.58	0.02	21.27	0.03

ROA, Relative occurrence area; TSS, True skill statistic; CCI, Correctly classified instances; AUC, Area under the ROC curve.

to accessibility may not always result in climatic biases (Kadmon *et al.*, 2004; McCarthy *et al.*, 2012).

Importantly, the surveys that followed human density showed substantially higher climatic biases (as well as higher variability in such bias) compared with the other strategies. This is probably due to the fact that, following real situations, this simulation allows the cells that are surveyed at each stage to be located either further away or closer in the climatic space. This increases the variability in the values of climatic bias, and at the same time limits further selections of cells to the neighbours of the already surveyed cells, thus increasing the climatic bias. Not surprisingly, this survey design was the one yielding the worst-performing SDM according to all metrics but specificity. Here, the contrasting behaviour of the values of specificity for the human density

simulation (see Fig. 3) can be due to the characteristics of this metric, which evaluates the capacity of the model to discriminate true absences, so the best models are those presenting less commission errors. Models built with biased climatic data can overfit to the climatic conditions that were effectively sampled (Lobo, 2008; Phillips *et al.*, 2009), hence generating spatially restricted predictions. This increases omission errors (false absences) as well as the number of true absences predicted by the model. Thus, metrics that use the number of true absences to assess SDM performance, such as specificity, will present higher values in climatically biased survey designs, whereas metrics based on the number of omission errors can present low values for the same predictive maps. It is also worth noting that other biased survey designs did not show highly variable climatic bias values because both roads and protected areas are well distributed within the Iberian Peninsula, allowing them to cover the most important environmental gradients within this region. However, in areas where the roads and national parks (or any other spatial attribute influencing the surveys) are not widely distributed, samples following these biased designs can generate larger climatic bias, limiting the predictive ability of SDMs and consequently their utility. Should the climatic biases generated by our surveys be higher, we would have found larger effects of survey design. Having said this, our results point out to that the effects of climatic bias on SDM performance could be limited until a certain level, where model accuracy diminishes dramatically, a potential effect that deserves further investigation.

The importance of the particular species being modelled for SDM performance shown by our results may be due to differences in their geographic distributions. Many studies have highlighted that geographical and ecological species' characteristics affect SDM accuracy (Berg *et al.*, 2004; Segurado & Araújo, 2004; Guisan *et al.*, 2007). Many spatial and ecological characteristics can affect SDM results (Brotons *et al.*, 2004; Segurado & Araújo, 2004; Hernandez *et al.*, 2006; Guisan *et al.*, 2007; McPherson & Jetz, 2007; Newbold *et al.*, 2009; Chefaoui *et al.*, 2011), from which the proportion of the occupied area over the considered territory (i.e. ROA) and the prevalence are frequently reported (Brotons *et al.*, 2004; Luoto *et al.*, 2005; Lobo *et al.*, 2007b; Chefaoui *et al.*, 2011). These variations in performance due to the species characteristics can occur even when using the same SDM technique (Seoane *et al.*, 2005; Hanspach *et al.*, 2010), thus increasing their influence on SDM results. In our case, species with smaller geographical distributions are predicted better, as evidenced by the negative correlation between ROA and the assessment metrics. This is because it is easier for SDMs to capture the climate-distribution relationship for species of restricted range, for they occupy a smaller – and thus easier to classify – environmental domain within the study region (Segurado & Araújo, 2004; Jiménez-Valverde *et al.*, 2008; Lobo, 2008). This implies that relatively small sample sizes but with a fair geographic coverage of surveys can provide data of sufficient quality to avoid spurious

effects on SDM accuracy of geographically restricted species. As a consequence, the importance of survey design for SDM performance may be limited in our analysis because it is restricted to endemic species – which often present limited spatial distributions. It is however likely that the importance of survey design will increase for species with larger ranges, given that the more widely distributed is a species, the more likely is that biased survey designs are not able to include all the climatic conditions it inhabits, hence resulting in less accurate models.

Sample size is also known to have strong effects on SDM predictive accuracy (Hirzel & Guisan, 2002; Reese *et al.*, 2005; Jiménez-Valverde *et al.*, 2009; Chefaoui *et al.*, 2011). For example, Araújo *et al.* (2009) showed that apparent failure of SDM to characterize European bird species-climate relationships in a high profile study (Beale *et al.*, 2008) was due to incomplete coverage of available data and, in another study, it was shown that the ability to predict range shifts of British birds under climate change was highly affected by the completeness of the data used to calibrate the models (Araújo *et al.*, 2005). In our work, increasing sample size led to increases in SDM performance and decreases in variability of model performance; predictions reached a large degree of stability with sample sizes of 10% of the studied area (592 cells) or larger. The measured effect of sample size may be in part related to its interaction with prevalence (Jiménez-Valverde *et al.*, 2009); below a certain threshold (such as the 10% we found), the representation of the distribution of the species may be too poor to produce reliable models. Larger sample sizes would allow more accurate descriptions of the species' response to the environment, thus yielding better predictions (Wisn *et al.*, 2008). Importantly, in our analyses, larger sample sizes provide better coverage of the study area, rather than just increasing numbers of occurrences. Thus, the stability in the predictions can be due to the fair representation of the environmental variation within the region, a key factor for the accuracy of the description of species-environment relationships (Kadmon *et al.*, 2003; Hortal *et al.*, 2008, 2012; Hortal & Lobo, 2011) and therefore of higher importance for SDM performance than sample size *per se* (Newbold *et al.*, 2009). Sastre and Lobo (2009) found that biased survey designs were not efficient in covering the true geographic pattern of species richness within a region (see also Hortal & Lobo, 2011). As showed by our analyses, the failures in covering the species distributions can be accentuated when biased survey designs are associated with smaller sample sizes. These failures are especially problematic when aiming to model the whole biodiversity of a region, and consequently for conservation actions based on many data-intensive strategies.

SDM techniques differ in their capacity to capture the relationship between species distributions and environmental variables. In general, complex techniques generate models with better performance within the training data (Elith *et al.*, 2006; Tsoar *et al.*, 2007; but see Lobo, 2008; Hijmans, 2012). However, the results of model comparisons are conflicting

across studies, due to strong variations in the modelling techniques and their possible interaction with data characteristics (Araújo *et al.*, 2005, 2009). Although no SDM technique performs better in all cases, our results show that the ensemble forecasting approach could be a good strategy to enhance SDM performance, for it consistently ranks within the best-performing technique regardless of the validation metric used. However, in some cases, the consensus predictions yielded models with lower performance than individual SDM, perhaps due to the inclusion of predictions from poorly performing techniques. The accuracy of consensus predictions is highly constrained by the quality of individual predictions (Araújo *et al.*, 2005). Therefore, to improve the performance of ensemble forecasting approaches it is necessary to modify the building-up of the consensus, either removing poor models or weighting individual models based on a previously selected performance metric (Araújo & New, 2007; Marmion *et al.*, 2009; Garcia *et al.*, 2012).

To summarize, our results support the view that the environmental coverage of the surveys is more important than the spatial structure of the calibration data. Spatially biased surveys do not always yield environmentally biased data and therefore can generate models as satisfactory as those using carefully designed surveys. Nevertheless, we contend that the distributional data used for species distribution modelling should be evaluated to characterize their climatic bias and coverage (see Kadmon *et al.*, 2004; Hortal *et al.*, 2008; Hortal & Lobo, 2011). Whenever possible, contrasting estimated realized niches with estimated analogous dimensions of the fundamental niche could provide useful insights as to the possible biases in the estimates with SDM (Araújo *et al.*, 2013). The largest source of variation in SDM predictions is the species being modelled, and more precisely the extent of its distribution within the studied area. Geographically restricted species (i.e. with lower ROA) yield more accurate results, but better predictions are also obtained with larger sample sizes and from survey designs with low climatic biases. Further studies are necessary to reveal the specific geographic or functional traits that can affect the predictive power of SDM. Additionally, it is important to investigate whether there is a minimum threshold of climatic bias behind which the quality of SDM predictions changes dramatically.

ACKNOWLEDGEMENTS

GT was funded by CAPES REUNI and PDSE grants no 11842121. TFR has been supported by CNPq research grants (564718/2010-6, 474774/2011-2, 310117/2011-9). MBA was funded through the FCT PTDC/AAC-AMB/98163/2008 project and the Integrated Program of IC&DT Call no 1/SAE-SCTN/ALENT-07-0224-FEDER-00175. JH was supported by a Spanish DGCyT RyC Fellowship, by a Brazilian CNPq Visiting Researcher grant (400130/2010-6), and by the CSIC-CNPq cooperation project 2011BR0071.

REFERENCES

- Albert, C.H., Yoccoz, N.G., Edwards, T.C., Graham, C.H., Zimmermann, N.E. & Thuiller, W. (2010) Sampling in ecology and evolution - bridging the gap between theory and practice. *Ecography*, **33**, 1028–1037.
- Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.
- Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.
- Araújo, M.B. & Peterson, T.A. (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**, 1527–1539.
- Araújo, M.B., Whittaker, R.J., Ladle, R.J. & Erhard, M. (2005) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, **14**, 529–538.
- Araújo, M.B., Thuiller, W. & Yoccoz, N.G. (2009) Reopening the climate envelope reveals macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences USA*, **106**, E45–E46.
- Araújo, M.B., Guilhaumon, F., Neto, D.R., Ortego, I.P. & Calmaestra, R.G. (2011) *Impactos, vulnerabilidad y adaptación al cambio climático de la biodiversidad española 2. fauna de vertebrados*. Dirección General de Medio Natural y Política Forestal. Ministerio de Medio Ambiente, y Medio Rural y Marino, Madrid.
- Araújo, M.B., Ferri-Yáñez, F., Bozinovic, F., Marquet, P.A., Valladares, F. & Chown, S.L. (2013) Heat freezes niche evolution. *Ecology Letters*, **16**, 1206–1219.
- Austin, M.P. & Heyligers, P. (1989) Vegetation survey design for conservation: gradsect sampling of forests in North-eastern New South Wales. *Biological Conservation*, **50**, 13–32.
- Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.
- Baselga, A. & Araújo, M.B. (2009) Individualistic vs community modelling of species distributions under climate change. *Ecography*, **32**, 55–65.
- Beale, C.M. & Lennon, J.J. (2012) Incorporating uncertainty in predictive species distribution modelling. *Philosophical transactions of the Royal Society of London*, **367**, 247–258.
- Beale, C.M., Lennon, J.J. & Gimona, A. (2008) Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences USA*, **105**, 14908–14912.
- Berg, Å., Gärdenfors, U. & Von Proschwitz, T. (2004) Logistic regression models for predicting occurrence of terrestrial molluscs in southern Sweden - importance of environmental data quality and model complexity. *Ecography*, **27**, 83–93.

- Braunisch, V. & Suchant, R. (2010) Predicting species distributions based on incomplete survey data: the trade-off between precision and scale. *Ecography*, **33**, 826–840.
- Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.
- Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **4**, 437–448.
- Busby, J.R. (1986) A biogeographical analysis of *Notophagus cunninghamii* (Hook.) in south-eastern Australia. *Australian Journal of Ecology*, **11**, 1–7.
- 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.
- Chefaoui, R.M., Lobo, J.M. & Hortal, J. (2011) Effects of species' traits and data characteristics on distribution models of threatened invertebrates. *Animal Biodiversity and Conservation*, **34**, 229–247.
- CIESIN C.U. & CIAT (2005) *Gridded Population of the World, Version 3 (GPWv3)*. Socioeconomic Data and Applications Center (SEDAC), Columbia University – Center for International Earth Science Information Network (CIESIN), Columbia University and Centro Internacional de Agricultura, Tropical (CIAT), Palisades, NY.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.
- Colwell, R.K. & Rangel, T.F. (2009) Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences USA*, **106**, 19651–19658.
- Dennis, R.L.H. & Thomas, C.D. (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation*, **4**, 73–77.
- Diniz-Filho, J.A.F., Bini, L.M., Fernando Rangel, T., Loyola, R.D., Hof, C., Nogués-Bravo, D. & Araújo, M.B. (2009) Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, **32**, 897–906.
- Edwards, T. Jr, Cutler, D., Zimmermann, N., Geiser, L. & Moisen, G. (2006) Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling*, **199**, 132–141.
- Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Farber, O. & Kadmon, R. (2003) Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, **160**, 115–130.
- 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**, 38–49.
- Franklin, J. & Miller, J.A. (2009) *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press Cambridge, UK, New York, New York, USA.
- Funk, V.A., Richardson, K.S. & Ferrier, S. (2005) Survey-gap analysis in expeditionary research: where do we go from here? *Biological Journal of the Linnean Society*, **85**, 549–567.
- García, R.A., Burgess, N.D., Cabeza, M., Rahbek, C. & Araújo, M.B. (2012) Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology*, **18**, 1253–1269.
- Geri, F., Amici, V. & Rocchini, D. (2011) Spatially-based accuracy assessment of forestation prediction in a complex Mediterranean landscape. *Applied Geography*, **31**, 881–890.
- Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007) What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615–630.
- Hanspach, J., Kühn, I., Pompe, S. & Klotz, S. (2010) Predictive performance of plant species distribution models depends on species traits. *Perspectives in Plant Ecology, Evolution and Systematics*, **12**, 219–225.
- Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 751–777.
- Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.
- Hijmans, R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679–688.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331–341.
- Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-Niche Factor Analysis: how to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027–2036.
- Hortal, J. & Lobo, J.M. (2005) An ED-based protocol for optimal sampling of biodiversity. *Biodiversity and Conservation*, **14**, 2913–2947.
- Hortal, J. & Lobo, J.M. (2011) Can species richness patterns be interpolated from a limited number of well-known areas? Mapping diversity using GLM and kriging. *Natureza & Conservação*, **9**, 200–207.
- Hortal, J., Lobo, J.M. & Jiménez-Valverde, A. (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology*, **21**, 853–863.
- Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M. & Baselga, A. (2008) Historical bias in biodiversity inventories

- affects the observed environmental niche of the species. *Oikos*, **117**, 847–858.
- Hortal, J., Lobo, J.M. & Jiménez-Valverde, A. (2012) Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. *Natureza e Conservação*, **10**, 108–118.
- IUCN & UNEP (2009) *The World Database on Protected Areas (WDPA)*. UNEP-WCMC, Cambridge. Available at: <http://www.wdpa.org/> (accessed 01 October 2012).
- Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, **14**, 885–890.
- Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2009) The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology*, **10**, 196–205.
- Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, **13**, 853–867.
- Kadmon, R., Farber, O. & Danin, A. (2004) Effects of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, **14**, 401–413.
- Lobo, J.M. (2008) More complex distribution models or more representative data? *Biodiversity Informatics*, **5**, 14–19.
- Lobo, J.M., Baselga, A., Hortal, J., Jiménez-Valverde, A. & Gómez, J.F. (2007a) How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? *Diversity and Distributions*, **13**, 772–780.
- Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2007b) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.
- Lobo, J.M., Jiménez-Valverde, A. & Hortal, J. (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, **33**, 103–114.
- Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G. & Montiel, O.M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, **35**, 105–116.
- Luoto, M., Pöyry, J., Heikkinen, R. & Saarinen, K. (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography*, **14**, 575–584.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.
- McCarthy, K., Fletcher, R.J., Rota, C.T. & Hutto, R.L. (2012) Predicting species distributions from samples collected along roadsides. *Conservation Biology*, **26**, 68–77.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*. Chapman and Hall, London.
- McPherson, J.M. & Jetz, W. (2007) Effects of species' ecology on the accuracy of distribution models. *Ecography*, **30**, 135–151.
- Medina, N.G., Lara, F., Mazimpaka, V. & Hortal, J. (2013) Designing bryophyte surveys for an optimal coverage of diversity gradients. *Biodiversity and Conservation*, **22**, 3121–3139.
- Newbold, T., Reader, T., Zalat, S., El-Gabbas, A. & Gilbert, F. (2009) Effect of characteristics of butterfly species on the accuracy of distribution models in an arid environment. *Biodiversity and Conservation*, **18**, 3629–3641.
- Pereira, J.M.C. & Itami, R.M. (1991) GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. *Photogrammetric Engineering & Remote Sensing*, **57**, 1475–1486.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Phillips, S., Anderson, R. & Schapire, R. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.
- Pontius, R.G. (2000) Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering & Remote Sensing*, **66**, 1011–1016.
- Rangel, T.F.L.V.B., Diniz-Filho, J.A.F. & Araújo, M.B. (2009) *BIOENSEMBLES 1.0. Software for Computer Intensive Ensemble Forecasting of Species Distributions Under Climate Change*. Privately Distributed, Madrid, Goiânia, Évora.
- Reese, G.C., Wilson, K.R., Hoeting, J.A. & Flather, C.H. (2005) Factors affecting species distribution predictions: a simulation modeling experiment. *Ecological Applications*, **15**, 554–564.
- Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, **35**, 211–226.
- Romo, H., Garcia-Barros, E. & Lobo, J.M. (2006) Identifying recorder-induced geographic bias in an Iberian butterfly database. *Ecography*, **29**, 873–885.
- Sánchez-Fernández, D., Lobo, J.M. & Hernández-Manrique, O.L. (2011) Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. *Diversity and Distributions*, **17**, 163–171.
- Sastre, P. & Lobo, J. (2009) Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation*, **142**, 462–467.
- Sastre, P., Roca, P. & Lobo, J.M. (2009) A Geoplatform for improving accessibility to environmental cartography. *Journal of Biogeography*, **36**, 568.
- Segurado, P. & Araújo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555–1568.

- Seoane, J., Bustamante, J. & Díaz-Delgado, R. (2005) Effect of expert opinion on the predictive ability of environmental models of bird distribution. *Conservation Biology*, **19**, 512–522.
- Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.
- Soberón, J.M. (2010) Niche and area of distribution modeling: a population ecology perspective. *Ecography*, **33**, 159–167.
- Soberón, J. & Nakamura, M. (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences USA*, **106**, 19644–19650.
- Thuiller, W., Brotons, L., Araújo, M.B. & Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165–172.
- Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.
- Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007) A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**, 397–405.
- Varela, S., Anderson, R.P., García-Valdés, R. & Fernández-González, F. (2014) Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography*, **37**, 1–8.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H. & Guisan, A. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.

SUPPORTING INFORMATION

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

Appendix S1 Details about methods.

Table S1 List of species used for the analyses, with numbers of presences and absences, degree of endemism and species code in the analysis.

Figure S1 Performance of SDM predictions from datasets based on different survey designs.

Figure S2 Performance of predictions from datasets of five sample sizes.

Figure S3 Predictive performance of SDM techniques according to Kappa, AUC, TSS and CCI validation metrics.

Figure S4 Interaction between sample size and model for specificity and CCI metrics.

Figure S5 SDM performance for all studied species according to Kappa, AUC, TSS and CCI validation metrics.

Figure S6 Number of times that a survey design does not get enough occurrences to generate models.

BIOSKETCH

Geiziane Tessarolo is a PhD student based at the Universidade Federal de Goiás, Brazil. Her research focuses on the assessment and evaluation of sources of uncertainty in species distribution modelling, particularly on data quality.

Thiago F. Rangel, Miguel B. Araújo and Joaquín Hortal are biogeographers studying the geographic distributions of species, among many other things.

Author contributions: G.T., J.H. and T.F.R. conceived the study; A.M.B. provided the data; G.T., J.H. and T.F.R. designed the analyses, with A.M.B.; G.T., J.H. and T.F.R. analysed the data; G.T., J.H., T.F.R. and A.M.B. wrote the paper.

Editor: Linda Beaumont