Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models

Sara Varela, Robert P. Anderson, Raúl García-Valdés and Federico Fernández-González

Ecological niche models represent key tools in biogeography but the effects of biased sampling hinder their use. Here, we address the utility of two forms of filtering the calibration data set (geographic and environmental) to reduce the effects of sampling bias. To do so we created a virtual species, projected its niche to the Iberian Peninsula and took samples from its binary geographic distribution using several biases. We then built models for various sample sizes after applying each of the filtering approaches. While geographic filtering did not improve discriminatory ability (and sometimes worsened it), environmental filtering consistently led to better models. Models made with few but climatically filtered points performed better than those made with many unfiltered (biased) points. Future research should address additional factors such as the complexity of the species’ niche, strength of filtering, and ability to predict suitability (rather than focus purely on discrimination).

Ecological niche models (ENMs) – also called species distribution models or climatic envelope models (Araújo and Peterson 2012) – have been widely used to predict the potential geographic ranges of species. Publications using ENMs have doubled in the last 5 yr, while citations of those papers have increased over 5000% (data: ISI Web of Knowledge; search: ‘ecological niche model’; May 2013). Out of this boom, a variety of specific software tools and new ecological models have appeared (Elith et al. 2010, 2011).

Researchers have focused their attention on detecting and analyzing the differences in predictions of these diverse models (Elith and Graham 2009, Lobo et al. 2010, Ashcroft et al. 2011, Lobo and Tognelli 2011, Nenzén and Araújo 2011). Different numerical tools and evaluation strategies have been analyzed, and there is an on-going debate about model-prediction accuracy (Peterson et al. 2011, Anderson 2012). Here we aim to contribute to the current improvement in model predictions by applying a new perspective focused on the problems associated with sampling bias, which can hinder the production of high-quality models (Wintle et al. 2005, Araujo and Guisan 2006, Anderson and Gonzalez Jr 2011).

Biodiversity or citizen science databases offer the possibility of using thousands of species records to calibrate models and map species distributions (Guralnick et al. 2007). However, data contained in these databases are highly heterogeneous. Distribution data sets include information from museums, herbaria, university databases or amateur field work, and usually compile hundreds of different surveys, each one designed with a different goal. As a consequence of this they accumulate taxonomic and geographic sampling biases, which often result in environmental biases as well (Hortal et al. 2008, Boakes et al. 2010, Newbold 2010). For instance, taxonomic biases or uneven sampling effort affect current biodiversity databases (Loiselle et al. 2008) and taphonomic biases or dating biases affect fossil databases (Varela et al. 2011).

Despite these biases, it is essential that we take advantage of the huge quantity of accumulated data. Until now ENMs have been typically calibrated without explicit steps to reduce the effects of sampling bias, which is not a desirable methodological procedure because biased data sets can produce poor predictions (Kadmon et al. 2003, Barry and Elith 2006, Loiselle et al. 2008, Varela et al. 2009, Lobo and Tognelli 2011). As model predictions are affected by spatial and/or temporal biases in the calibration data sets, it is highly desirable to find methods to filter the data sets...
and determine an appropriate subsample for calibrating ENMs, regardless of the initial bias of the raw data.

Sampling is a sensitive step for any ecological analysis (Albert et al. 2010). However, few studies attempt to correct the sampling bias of species records when constructing ENMs. One of these attempts was done using Maxent model, biasing the background data sample with the same bias of the occurrence records (Phillips et al. 2009). Other papers have filtered occurrence records in geographic space (Hidalgo-Mihart et al. 2004, Iguchi et al. 2004, Anderson and Raza 2010). Both approaches hold promise, but their efficacies remain poorly documented.

Here we analyze the performance of two different types of filters for selecting calibration data when constructing an ENM. These were a geographic filter and an environmental filter (specifically, a climatic filter). Geographic filters have already been used as a tool to improve ENMs (Hidalgo-Mihart et al. 2004, Iguchi et al. 2004, Anderson and Raza 2010, Hijmans 2012, Rodríguez-Castañeda et al. 2012), while climatic filters remain generally unexplored. We aim to assess whether simple filtering rules in environmental space can improve model predictions. Our goal is to develop a procedure for improving model predictions that would work for many different kinds of sampling bias and across wide ranges of sample sizes.

To test the performance of this approach, we create a virtual (simulated) species with a geographic distribution related to three climatic variables. Virtual species have been used to test different methodological aspects of ENMs (Hirzel et al. 2001, Jimenez-Valverde and Lobo 2007, Meynard and Kaplan 2013). In our case, the virtual species allows us to circumvent complications regarding dispersal limitations and biotic interactions inherent to the studies that use real species. We generate different geographically biased data samples to illustrate several common biases in biological databases (e.g., distance to roads). Subsequently, we apply geographic and environmental filters to the biased data sets to obtain different subsamples that we use to calibrate the models. After that, model results are evaluated against the real distribution of the virtual species.

Model results allow us to analyze three different aspects. First, we test the difference in performance between the two filters. We hypothesize that geographic filters could fail to select the optimal calibration data set if they discard aggregated points with unique climatic conditions. On the other hand, climatic filters might select optimal calibration data sets by removing redundant information (points with similar climatic conditions). Second, we investigate whether filters work regardless of the initial bias in the data sets. Clearly, we desire methods that are robust and are not affected by the initial bias of the data set. Finally, we address the issue of sample size. We hypothesize that small but unbiased (or less biased) data samples should produce better predictions than large but strongly biased data samples.

Thus, for this first exploration we assess the robustness of the filters to changes in sampling biases, and the sensitivity of the filters in relation to sample size. The current study does not address sensitivity of the filters to variation in species’ niches (for instance, complex/simple relationship with variables, number of variables, or broad/narrow niches), or the performance of filters when using different modelling approaches and algorithms. These and other questions should be addressed in future works.

Material and methods

Defining the virtual species

We defined a virtual species that is not resistant to extreme environmental conditions. Cold temperatures, arid conditions, and very warm or wet conditions are avoided. The survival of this species depends on three variables: maximum temperature of the warmest month (Bio5), minimum temperature of the coldest month (Bio6), and precipitation of the driest month (Bio14). All variables were downloaded from <www.worldclim.org> (Hijmans et al. 2005). The species’ relationship with each variable was defined using normal curves (Bio5: mean = 20°C and standard deviation = 10°C; Bio6: mean = 10°C, standard deviation = 10°C; Bio14: mean = 20 mm, standard deviation = 10 mm; Fig. 1). The species’ overall climatic suitability was defined as the multiplication of each single variable’s suitability. R packages base, stats (R Core Team) and raster (Hijmans and Etten 2013) were used to construct the virtual species.

After that, we projected the virtual species’ requirements onto the Iberian Peninsula to obtain a distribution map. The species’ highest suitability (and defined geographic range, see below) covers Atlantic and mild Mediterranean areas (Fig. 2A). An arbitrary suitability value of 0.2 was chosen as the survival threshold of the species. Thus, Fig. 2B represents the distribution of the species following this threshold. We do not include biotic interactions or dispersal limitations in the analysis. We assume that the species’ distribution is in equilibrium with climate and that the species inhabits all climatically suitable areas.

Figure 1. Climatic suitability of the virtual species in relation to precipitation of the driest month (Bio14) and maximum temperature of the warmest month (Bio5) (<www.worldclim.org>). When maximum temperature is 20°C, minimum temperature is 10°C, and precipitation in the driest month is 20 mm, the species’ climatic suitability is 1 (maximum). Suitability for the species decreases dramatically as temperature approaches the following extreme values: 15°C and 25°C for maximum temperature; 5°C and 15°C for minimum temperature; 15 mm and 25 mm for precipitation of the driest month.
Furthermore, we sample from the species’ distribution (after applying the survival threshold) without consideration of the relative suitability above that threshold. Such a simplification may not necessarily translate well to real biological species, but research on the associations among abundance, detection probability, and modelled suitability remains preliminary (Brown et al. 1995, Jiménez-Valverde et al. 2009, Meynard and Kaplan 2013). Therefore, the present study is the most relevant to research aimed at identifying the areas suitable for a species (rather than reconstructing a gradient of suitability).

Data sets

The species’ distribution (Fig. 2B) was sampled using the sample function in R-base package (<www.r-project.org/>). This function has a specific parameter to include weights in the sampling process. Here, weights are related to four different geographic distances: distance to roads (<www.diva-gis.org/datadown>), distance to nature reserves (<www.europarc.org/home/>, distance to populated areas (Iberian Peninsula population in 2000), and a combination of the three (simple multiplication of the former three biases) (Fig. 3). Additionally, a random sample was also generated. We created distance layers using the Idrisi Kilimanjaro distance algorithm; application defaults were used (Eastman 2003). For each bias treatment, as well as for the random sample, we selected an initial data set of 10 000 points, which were later subjected to filtering (see below). GIS layers used to generate biases were downloaded from the Laboratorio de Biogeografía Informática of the Museo Nacional de Ciencias Naturales in Madrid, Spain (<www2.mncn.csic.es/LBI/Recursos.htm>).

Filtering the raw data samples

We filtered the initial raw biased data samples using two filters, one geographic and the other environmental (climatic). Non-filtered (i.e. random) subsamples of the same sample size were also generated to allow comparisons. We used the gridSample function of the R dismo package for the filtering process (Hijmans et al. 2012). This function selects points from an x-y layer, using a defined grid as stratification. In the geographic filter, the x-axis is longitude, the y-axis is latitude, and repetitive clumped occurrence records are discarded; therefore, the filtered resamples lack geographically aggregated data (Fig. 4A). In the climatic filter, the x-axis is maximum temperature of the warmest month, and the y-axis is precipitation of the driest month. Repetitive occurrences under similar climatic conditions are discarded. Hence, the climatically stratified resamples will maximize the climatic coverage evenness of the Iberian Peninsula (Fig. 4B). We used two variables instead of all three climatic variables to mimic the reality of studies with real species, where we do not have (or even know) all of the relevant environmental variables. The geographic grid and climatic grid were adjusted to the actual geographic extent or climatic range for each subsample set. The grid resolution for filtering was then set at 0.1, which means that the value of the grid cells is 10 times smaller than the actual range of the variables. Hence, here the filtering led to a pool of unique cells for each filtering treatment (maximum of 100 here, although with more axes or finer grid resolution many more would be possible).

We applied the filters to the raw biased data sets, obtaining filtered pools of points, and then selected samples of different sizes: 5, 10, 15, 25, 50, 75, 100, 250, 500 and 1000 points. This process was repeated 100 times (and the mean was calculated as the final result for that size, bias and filter). Thus, we obtained 100 samples for every combination of sample size, bias, and filter; and overall, 5000 calibration data samples for each filter. In total, we generated 15 000 models and map predictions. Results show the mean value of each experiment.

Ecological niche models

Several methods are available for constructing ENMs, and they have been compared widely (Austin et al. 2006, Tsoar et al. 2007). As our work is not focused on model comparison, we decided to select one method. We chose Maxent because of its high performance (Elith et al. 2010) and common use in biogeography (Joppa et al. 2013).

Models were constructed using the R-package dismo (Hijmans et al. 2012). We selected the maxent function and used the default settings of this model to build our ENMs (Phillips et al. 2006, 2009, Elith et al. 2011). All three causal climatic variables were used as predictors.

We calculated the performance for each model using the Area Under the Curve (AUC) of the Receiver Operating Characteristic plot. For calculating AUC, we drew a new
sample of presence and absence points from the binary map of the species’ true occupied and unoccupied areas, respectively (Fig. 2B). All models have the same geographic extent, so comparisons among treatments are appropriate (Lobo et al. 2008, Peterson et al. 2011). Shapiro–Wilk tests indicated that the distribution of the AUC results was not Gaussian, so we used non-parametric Wilcoxon–Mann–Whitney tests (also called the Mann–Whitney U tests), for exploring the observed differences between paired models. We conducted these tests to compare unfiltered vs geographic filters, unfiltered vs climatic filters and geographic vs. climatic filters. We calculated W, which is the absolute value of the sum of the signed ranks, and the p-value for each respective analysis. Conservatively, we implemented these tests with two-tailed null hypotheses and inspected the direction of observed differences. We first performed comparisons for all experiments (pooled). Then, to determine whether the results were consistent across treatments, we repeated the tests individually for each sample size and each bias treatment. To compare the change in AUC we used a Kruskal–Wallis rank test, because the distributions of the subsamples were not Gaussian and we had more than two samples to compare. All models, statistical analysis, and plots were made in R (<www.r-project.org/>).

Results

Effects of geographic and climatic filters

Experiments showed stark contrasts between the geographic and environmental filters when all data sets were compared (Table 1). Geographic filters generally decreased the predictive power of models slightly when compared with non-filtered (random) data sets (Table 1 and 2). On the other hand, climatic filtering improved model results for all experiments, compared with non-filtered results (Table 3), often substantially. Concomitantly, the AUC after climatic filtering was significantly higher than that after geographic filtering.

Influence of the type of bias

Differences in AUC between geographic and climatic filters were consistent across different bias treatments, with experiments for all sample sizes pooled (Table 1). Thus, all bias treatments were better using the climatic filter compared with the geographically filtered results (Supplementary material Appendix 1, Fig. A1). Finally, there was no evidence of differences in the change in AUC among biases after using either filter (Table 1 and Supplementary material Appendix 1).

Sample size

Differences in AUC between geographic and climatic filters were rather consistent across the 10 different sample sizes (n = 5, because we used 5 different biases), with the climatic filter leading to better results than the geographic

Figure 3. Geographic pattern of the selected biases in the Iberian Peninsula: (A) random, (B) distance to roads, (C) distance to nature reserves, (D) distance to populated areas (Iberian Peninsula population in 2000), and (E) a combination of all previous biases.
First, results indicate better performance for the climatic filter than for the geographic one. The geographic filter did not increase model performance and even decreased it in some circumstances (Fig. 5). The Iberian Peninsula is environmentally highly heterogeneous (Gallardo et al. 2012). Thus, by using the geographic filter with the present data set, we discarded points with relevant (non-repetitive) climatic information (but aggregated in geographic space), and instead selected points tending to have similar climatic conditions (but located more distantly in geography) (Fig. 4). Therefore, the results suggest that species living in patchy or spatially heterogeneous environments could be negatively affected by using this kind of filter. Nevertheless, the intuitive reasoning of discarding geographically aggregated points might work only in some situations, and is likely related to the spatial distribution of the environmental variables and the manner in which the occurrence records were sampled (here, randomly from the binary map of the species’ distribution).

Second, the climatic filter did improve model results. The models had higher discriminatory power when filter (Table 1). The observed pattern of improvement in AUC was related to sample size. On one hand, the geographically filtered results showed a very low AUC (relative to no filter) when the sample size was small, but performance for this filtering treatment increased dramatically as the sample size increased above 15 points (although, on average, it stayed slightly lower than for no filter). On the other hand, the climatic filter obtained relatively high model performance at small sample sizes, with the mean performance of the climatically filtered data sets varying remarkably little across sample sizes.

Despite the observed stability of mean AUC values, sample size did affect model results. Sample size was related to the standard deviation of AUC values: small sample sizes had very large standard deviations, and large sample sizes small ones. A linear regression of the logarithmic transformation of standard deviation onto sample size was significant for both data sets, with negative slopes (geographic and climatic; r-squared = 0.72 and 0.64, respectively, \(p < 0.005\); Supplementary material Appendix 1, Fig. A2). Thus, the differences among experiments decreased when the calibration sample size increased.

### Discussion

First, results indicate better performance for the climatic filter than for the geographic one. The geographic filter did not increase model performance and even decreased it in some circumstances (Fig. 5). The Iberian Peninsula is environmentally highly heterogeneous (Gallardo et al. 2012). Thus, by using the geographic filter with the present data set, we discarded points with relevant (non-repetitive) climatic information (but aggregated in geographic space), and instead selected points tending to have similar climatic conditions (but located more distantly in geography) (Fig. 4). Therefore, the results suggest that species living in patchy or spatially heterogeneous environments could be negatively affected by using this kind of filter. Nevertheless, the intuitive reasoning of discarding geographically aggregated points might work only in some situations, and is likely related to the spatial distribution of the environmental variables and the manner in which the occurrence records were sampled (here, randomly from the binary map of the species’ distribution).

Second, the climatic filter did improve model results. The models had higher discriminatory power when

### Table 1

Mean AUC values for filtering treatments (no filter, geographic filter, and climatic filter) for all models pooled and for subsets (by bias treatment and by sample size). The last two columns provide the W test statistic and \(p\) values of Wilcoxon (Mann–Whitney) tests comparing results of the geographic and climatic filters.

<table>
<thead>
<tr>
<th></th>
<th>No filter</th>
<th>Geographic filter</th>
<th>Climatic filter</th>
<th>Geographic minus no filter</th>
<th>Climatic minus no filter</th>
<th>W</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All experiments</td>
<td>0.92</td>
<td>0.91</td>
<td>0.98</td>
<td>−0.01</td>
<td>0.06</td>
<td>2438</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Random</td>
<td>0.91</td>
<td>0.91</td>
<td>0.98</td>
<td>0.00</td>
<td>0.07</td>
<td>98</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Roads</td>
<td>0.90</td>
<td>0.91</td>
<td>0.98</td>
<td>0.01</td>
<td>0.08</td>
<td>100</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Population</td>
<td>0.93</td>
<td>0.91</td>
<td>0.99</td>
<td>−0.02</td>
<td>0.06</td>
<td>100</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Reserves</td>
<td>0.90</td>
<td>0.91</td>
<td>0.98</td>
<td>0.01</td>
<td>0.08</td>
<td>90</td>
<td>&lt;0.00015</td>
</tr>
<tr>
<td>All biases</td>
<td>0.95</td>
<td>0.92</td>
<td>0.99</td>
<td>−0.03</td>
<td>0.04</td>
<td>100</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>5</td>
<td>0.81</td>
<td>0.25</td>
<td>0.99</td>
<td>−0.56</td>
<td>0.18</td>
<td>25</td>
<td>0.007</td>
</tr>
<tr>
<td>10</td>
<td>0.91</td>
<td>0.81</td>
<td>0.97</td>
<td>−0.10</td>
<td>0.06</td>
<td>20</td>
<td>0.15</td>
</tr>
<tr>
<td>15</td>
<td>0.93</td>
<td>0.97</td>
<td>0.98</td>
<td>−0.04</td>
<td>0.05</td>
<td>24</td>
<td>0.015</td>
</tr>
<tr>
<td>25</td>
<td>0.94</td>
<td>0.88</td>
<td>0.98</td>
<td>−0.06</td>
<td>0.04</td>
<td>25</td>
<td>0.007</td>
</tr>
<tr>
<td>50</td>
<td>0.93</td>
<td>0.90</td>
<td>0.98</td>
<td>−0.03</td>
<td>0.05</td>
<td>25</td>
<td>0.007</td>
</tr>
<tr>
<td>75</td>
<td>0.93</td>
<td>0.91</td>
<td>0.98</td>
<td>−0.02</td>
<td>0.05</td>
<td>25</td>
<td>0.007</td>
</tr>
<tr>
<td>100</td>
<td>0.93</td>
<td>0.92</td>
<td>0.98</td>
<td>−0.01</td>
<td>0.05</td>
<td>25</td>
<td>0.007</td>
</tr>
<tr>
<td>250</td>
<td>0.94</td>
<td>0.91</td>
<td>0.98</td>
<td>−0.03</td>
<td>0.04</td>
<td>25</td>
<td>0.007</td>
</tr>
<tr>
<td>500</td>
<td>0.94</td>
<td>0.92</td>
<td>0.98</td>
<td>−0.02</td>
<td>0.04</td>
<td>25</td>
<td>0.007</td>
</tr>
<tr>
<td>1000</td>
<td>0.94</td>
<td>0.92</td>
<td>0.98</td>
<td>−0.02</td>
<td>0.04</td>
<td>25</td>
<td>0.007</td>
</tr>
</tbody>
</table>
environmental biases were avoided by the use of climatic filters. This positive result suggests that at least under some circumstances researchers may be able to increase model performance by filtering by environmental variables. Here, climatic filters were effective in reducing redundant sampling, likely especially those caused by biased sampling, without unduly removing the signal of the species’ niche (Fig. 4). Even when we did not use all three important variables for the species in the filtering, the results clearly demonstrate good performance of climatic filters. The efficacy of environmental filtering under other circumstances (e.g. for species with more complex niches, species with wide/narrow distributions, or when species records are more likely in increasingly more suitable areas) remains to be explored. In the future, the robustness of this method should be tested under different and more complex circumstances, including the possibility of the researcher adding variables to the filters.

Additionally, the filters led to similar results for all five initial biases. Three of the treatments had a strong geographic pattern (distance to road, distance to nature reserves and distance to populated areas), while the other two had a much more diffuse geographic pattern (random and all biases together) (Fig. 3). The increase in model predictive power by using climatic filters was independent of the initial bias of the data sets (Supplementary material Appendix 1, Fig. A2). The minimum number of points needed to achieve maximal performance varied between filters (Table 1 and 2, and Fig. 5). Using a climatic filter, the optimal performance appeared with as few as 5 points, while when using non-filtered data 50 were necessary and using geographically filtered data required 100. Here, we defined optimal performance as that reached when increases in AUC were less than 0.01 after adding more points to the calibration data set. Although the specific number of points necessary to achieve optimal performance surely depends on the complexity of the niche and the number of environmental variables used, we predict that the overall pattern (of smaller, unbiased data sets outperforming larger, biased ones) will hold.

Finally, the study shows notable results with regard to sample size. Models calibrated with few climatically filtered data points produced better results (on average) than did models calibrated with large biased data sets (Fig. 5). Biodiversity databases have large and typically biased data sets of species records. The present results indicate that it can be better to calibrate models using a climatically filtered subsample of those occurrences, than using the whole set of available species records. Conversely, real occurrence records are scarce for some endangered and/or rare species. Here, we show that small data sets might be able to produce good predictions (at least for species with simple niches), as long as records were satisfactory representations of the species’ environmental requirements.

On the other hand, large data sets indeed produced model results that were more consistent, with smaller standard deviations between experimental replicates than did small data sets (Supplementary material Appendix 1, Fig. A2). The minimum number of points needed to achieve maximal performance varied between filters (Table 1 and 2, and Fig. 5). Using a climatic filter, the optimal performance appeared with as few as 5 points, while when using non-filtered data 50 were necessary and using geographically filtered data required 100. Here, we defined optimal performance as that reached when increases in AUC were less than 0.01 after adding more points to the calibration data set. Although the specific number of points necessary to achieve optimal performance surely depends on the complexity of the niche and the number of environmental variables used, we predict that the overall pattern (of smaller, unbiased data sets outperforming larger, biased ones) will hold.

Table 2. Change in AUC when using a geographic filter (mean value after 100 iterations). Experiments differ in the initial biases of the raw data sets (rows) and the selected sample size for calibrating the model (columns). Results show that there is no improvement in model discriminative power when using a geographic filter, compared with using no filter. Only one experiment has a positive value, the one with sample size 10 and the initial raw data set biased by the distance to nature reserves (in italics). Bold text highlights the experiments with largest change in AUC (here, decreases).

<table>
<thead>
<tr>
<th>Biases</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Random</td>
<td>0.67</td>
</tr>
<tr>
<td>Roads</td>
<td>−0.46</td>
</tr>
<tr>
<td>Population</td>
<td>−0.49</td>
</tr>
<tr>
<td>Reserves</td>
<td>0.52</td>
</tr>
<tr>
<td>All</td>
<td>−0.65</td>
</tr>
</tbody>
</table>

Table 3. Change in AUC after using a climatic filter (mean value after 100 iterations). Experiments differ in the initial biases of the raw data sets (rows) and the selected sample size for calibrating the model (columns). Results show an improvement in model discriminative power in all experiments. Bold text highlights experiments with largest change in AUC (here, increases).

<table>
<thead>
<tr>
<th>Biases</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Random</td>
<td>0.01</td>
</tr>
<tr>
<td>Roads</td>
<td>0.24</td>
</tr>
<tr>
<td>Population</td>
<td>0.24</td>
</tr>
<tr>
<td>Reserves</td>
<td>0.22</td>
</tr>
<tr>
<td>All</td>
<td>0.09</td>
</tr>
</tbody>
</table>
models could be evaluated against the known suitability of points taken probabilistically vs. unsuitable areas (Fig. 2B). Future research should conduct parallel experiments where points are taken in model accuracy with increasing sample sizes. Small samples were the most sensitive to filtering. Interestingly, small samples using a climatic filter produced better models than did large non-filtered data sets.

**Conclusions and future directions**

This exploration of geographic and climatic filtering of biased data sets with a virtual species allows several conclusions and points to various avenues for future research. Clearly, climatic filtering can improve model results, and here the improvement was independent of the initial biases. Furthermore, it allowed calibration of accurate models even when using extremely small data sets. In contrast, geographic filters generally did not improve model results. Nevertheless, the results presented here should be taken as a preliminary attempt to explore viable solutions for optimizing selection of calibration data when points are known or suspected to suffer from bias. Further investigation is necessary to reach general conclusions and produce guidelines regarding many issues, including: the optimal grain of the climatic and geographic filters, the optimal number of variables included in the filters, the level of sampling bias, and the filtering performance for different kinds of species (e.g. species with broad/narrow niches) or different modelling algorithms. Furthermore, the current experiments took points from the binary map of suitable vs. unsuitable areas (Fig. 2B). Future research should conduct parallel experiments where points are taken probabilistically from the suitability surface (e.g. Fig. 2A). Such models could be evaluated against the known suitability value of the virtual species, rather than compared with its binary geographic distribution, as here.

**Acknowledgements** – We thank Catherine H. Graham, Mariano Soley-Guardia, Robert A. Boria and Peter J. Galante for their helpful suggestions to improve the manuscript. This research was made possible by funding from the Education for Competitiveness Operational Programme (ECOP) project ‘Support of establishment, development and mobility of quality research teams at the Charles Univ.’ (CZ.1.07/2.3.00/30.0022, funded by the European Science Foundation and Czech Republic; SV); the project ‘Potential effects of climate change on Natura 2000 conservation targets in Castilla-La Mancha (CitUCHR)’ (Ref. no.: POIC10-0311-0585), funded by the Regional Government of Castilla-La Mancha (Spain; SV, RG-V, and FF-G); and the U. S. National Science Foundation (NSF DEB-0717357 and DEB-1119915; RPA).

**References**


Supplementary material (Appendix ECOG-00441 at www.oikosoffice.lu.se/appendix). Appendix 1.