

RESEARCH
PAPER



Species richness, area and climate correlates

David Nogués-Bravo^{1*} and Miguel B. Araújo^{2,3}

¹Macroecology and Conservation Unit, University of Évora, Estrada dos Leões, 7000-730 Évora, Portugal, ²Department of Biodiversity and Evolutionary Biology, National Museum of Natural Sciences, CSIC, C/José Gutiérrez Abascal, 2, 28006 Madrid, Spain and ³Biodiversity Research Group, Oxford University Centre for the Environment, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK

ABSTRACT

Aim Species richness–area theory predicts that more species should be found if one samples a larger area. To avoid biases from comparing species richness in areas of very different sizes, area is often controlled by counting the numbers of co-occupying species in near-equal area grid cells. The assumption is that variation in grid cell size accrued from working in a three-dimensional world is negligible. Here we provide a first test of this idea. We measure the surface area of *c.* 50 × 50 km and *c.* 220 × 220 km grid cells across western Europe. We then ask how variation in the area of grid cells affects: (1) the selection of climate variables entering a species richness model; and (2) the accuracy of models in predicting species richness in unsampled grid cells.

Location Western Europe.

Methods Models are developed for European plant, breeding bird, mammal and herptile species richness using seven climate variables. Generalized additive models are used to relate species richness, climate and area.

Results We found that variation in the grid cell area was large (50 × 50 km: 8–3311 km²; 220 × 220: 193–55,100 km²), but this did not affect the selection of variables in the models. Similarly, the predictive accuracy was affected only marginally by exclusion of area within models developed at the *c.* 50 × 50 km grid cells, although predictive accuracy suffered greater reductions when area was not included as a covariate in models developed for *c.* 220 × 220 km grid cells.

Main conclusions Our results support the assumption that variation in near-equal area cells may be of second-order importance for models explaining or predicting species richness in relation to climate, although there is a possibility that drops in accuracy might increase with grid cell size. The results are, however, contingent on this particular data set, grain and extent of the analyses, and more empirical work is required.

Keywords

Area, climate correlate, Europe, grid cells, GAMs, grain, model accuracy, species richness, variable selection, 3D area.

*Correspondence: David Nogués-Bravo, Macroecology and Conservation Unit, University of Évora, Estrada dos Leões, 7000-730 Évora, Portugal.
E-mail: dnoques@uevora.pt

INTRODUCTION

One of the oldest and best-documented patterns in ecology is the species–area relationship. It follows from this relationship that more species should be found if one samples a larger area (for review see Rosenzweig, 1995). The generality of this pattern has led to the formulation of the ‘geographical-area hypothesis’, whereby the high species richness of the tropics is expected because of the greater geographical area of tropical biome (for

discussion see Gaston & Blackburn, 2000; Hawkins & Porter, 2001; Turner, 2004). To control for ‘area effects’ in studies investigating correlates of species richness, researchers have often superimposed grids on equal or near-equal area projections and counted the number of species co-occupying grid cells. One of the problems is that grids are not strictly equal in area because of the range of altitude encompassed and because water bodies reduce the area covered by land. For example, a 45° average slope would increase the surface area of a planimetric grid by up to

40% (Turner, 2004). When using non-equal area projections there is the additional problem that grid cells may follow a reduction in area from the equator to the pole. For example, Jetz & Rahbek (2001) reported a 20% change in grid cell size when studying species richness using 1-degree latitude longitude cells in Africa. Because of the species–area relationship, the combined effects of topography, water and geographical position are likely to produce biases in the numbers of species co-occurring in grid cells that are not strictly of equal area (e.g. Connors & McCoy, 1979). Although some authors have acknowledged this problem (e.g. Whittaker *et al.*, 2001), to our knowledge no study has investigated systematically the effect of varying the surface area among near-equal area grid cells on perceived species richness–climate relationships. In this study we attempt such an investigation and ask how variation in surface area among near-equal area grid cells affects the explanatory and predictive power of models relating species richness with climate.

A number of studies have discussed the implications of area in patterns of species richness (e.g. Whittaker *et al.*, 2001; Rahbek, 2005; Whittaker *et al.* in preparation). Typically, two distinctions are made: the effects associated with the ‘grain’, or the size of individual sampling units; and the effects associated with the ‘extent’, or the geographical space over which comparisons are made. In this study we are concerned with small variations of grain on species richness or, put another way, the effects of using non-standardized measurements of area when using a fixed grain. Typically, studies addressing this problem have explored how a large variation in grain affects correlations between controlling factors and species richness (e.g. Kaufman & Willig, 1998; Lyons & Willig, 1999; Rahbek & Graves, 2000, 2001; Hurlbert & White, 2005). These studies have highlighted the fact that grain can influence the perception of species richness patterns and thus the magnitude and significance of species richness correlates. There are at least four reasons why grain may affect perceived species richness patterns and the species richness relationship with climate. The first is the controlling effect of area over the numbers of species found within grid cells (Rosenzweig, 1995). The second is the habitat hypothesis (Williams, 1964), whereby larger areas are postulated to contain more species because of their increased number of habitats and resources available for species. However, the extent to which a positive relationship between habitat diversity and species richness is an artefact of the species–area relationship remains unclear (e.g. Gaston & Blackburn, 2000). The third is the idea that variations in grain can reveal different patterns in the distribution of individual species, with coarse grids merging to form range maps, whereas finer grids reveal broad-scale habitat distributions, regional ubiquity, and local population patterns (Kunin, 1998). The extent to which biases in individual species maps affect overall patterns of species richness is still unknown (but see Hurlbert & White, 2005), but an analysis of modelled data indicates that errors, or biases, may propagate exponentially (Araújo *et al.*, 2005b). Finally, there is the ‘reddened spectrum’ of climate heterogeneity, i.e. climate variables are heterogeneous at different scales of analysis and this may have marked effects on measured species richness–climate relationships (e.g. O’Brien *et al.*, 2000).

In order to avoid biases from comparing species richness among areas of very different sizes, researchers have utilized a variety of approaches to control for area (for a review see Table 1). These approaches include the use of rules for dealing with grid cells near the coast; the inclusion of area, or surrogates of area, as a covariate in statistical analyses; the fitting of species richness–area curves (SAR) to estimate species numbers within grid cells; the use of equal-area projections; and combinations of different approaches (e.g. O’Brien *et al.*, 2000; Tognelli & Kelt, 2004; Field *et al.*, 2005). In spite of the many attempts to control for area in models of species richness, to our knowledge there have been no attempts to measure variations in the surface area of near-equal area grids and to assess the impact of this source of bias on perceived species richness–climate relationships. Here, we provide the most accurate measurement of surface area within *c.* 50 × 50 km and 220 × 220 km grid cells in Europe available to date and investigate how variations in the surface area of near-equal grid cells affect: (1) the selection of climate variables entering a generalized additive model of species richness; and (2) the accuracy of generalized additive models in predicting species richness in unsampled grid cells. Analyses are carried out using two grid cell sizes because there is a possibility that the results are contingent on grain. Furthermore, the two grid sizes are among the most commonly used in studies analysing species richness patterns at the macro scale (see Table 1 and references therein). Models are developed for European plant, breeding bird, mammal and herptile species richness using seven climate variables.

MATERIALS AND METHODS

Species data

Species locality data comprised 187 species of mammals (Mitchell-Jones *et al.*, 1999), 445 species of breeding birds (Hagemeijer & Blair, 1997), 149 species of amphibians and reptiles (Gasc *et al.*, 1996), and 2362 species of plants (Jalas & Suominen, 1972–96). Data varied with regard to taxonomic coverage. Terrestrial vertebrates comprise all known species, whereas plants comprise *c.* 20% of the European flora. The grid used was based on the Common European Chorological Grid Reference System (CGRS). The chorological data were inserted into a *c.* 50 × 50 km grid map (2500 km²), based on the Universal Transverse Mercator (UTM) projection and the Military Grid Reference System (MGRS). In this study the mapped area included western, northern and southern Europe, but excluded most of the eastern European countries (except for the Baltic States), where recording effort was both less uniform and less intensive (for more details on data conversion see Williams *et al.*, 2000).

Climate data

Climate data included seven variables developed for Europe in the context of the EC-funded ATEAM project (<http://www.pik-postdam.de/ateam>). Data were averaged for the period 1961–90 and included mean annual temperature (°C), mean temperature of the coldest month per year (°C), mean annual precipitation

Table 1 Approaches to deal with biases introduced when using sampling units of varying sizes in studies of species richness: (1) combination of coastal grid cells with neighbour terrestrial grid cells and removal of grid cells under certain thresholds of land area; (2) inclusion of area in statistical analyses of species richness; (3) fitting of species richness–area curves (SAR) to estimate an expected value of species richness for every grid cell; and (4) use of equal-area projections. We utilized the ‘web of science’ to undertake this survey and restricted the analyses to studies at continental or global scales. Because of the large sample size involved we selected one representative study per lead author

Approach	Authors	Spatial units	Size	Extent
1	Araújo (2003)	Grid cells	50 km	Europe
	Badgley & Fox (2000)	Grid cells	150 miles	North America
	Bini <i>et al.</i> (2004)	Grid cells	220 km	South America
	Diniz-Filho <i>et al.</i> (2004)	Grid cells	220 km	South America
	H-Acevedo & Currie (2003)	Grid cells	2°	North and Central America
	Hawkins & Porter (2003)	Grid cells	220 km	Palaeartic
	Pearson & Carroll (1999)	Grid cells	275–325 km	North America, India, Australia
	Kaufman & Willig (1998)	Grid cells, latitudinal bands	2.5°; 5°	New World
	Jetz & Rahbek (2001)	Grid cells	1°	Africa
	Taplin & Lovett (2003)	Grid cells	1°	Africa
2	Bárcena <i>et al.</i> (2004)	Hydrological basins		Europe
	Lyons & Willig (1999)	Nested grid cells	1000–25,000 km ²	New World
	Patten (2004)	Administrative units	North American states	USA
	McKinney (2002)	Administrative units	North American states	USA
	Kerr (1999)	Grid cells	2.5°	North America
	Currie (1991)	Grid cells	2.5°	North America
	Ribera <i>et al.</i> (2003)	Administrative units	European countries	Europe
3	Meliadou & Troumbis (1997)	Grid cells	2.5°	Europe
4	Aava (2001)	Grid cells	200 km	Australia
	Tognelli & Kelt (2004)	Grid cells	100 km	South America
	Rangel & Diniz-Filho (2003)	Grid cells	350 km	World

sum (mm), mean annual winter precipitation sum (mm), mean annual summer precipitation sum (mm), mean annual growing degree days (> 5°), and the mean ratio of annual actual evapotranspiration over annual potential evapotranspiration. These climate variables are postulated to act as controlling factors of the physiological processes limiting the spatial distribution of species, especially among plants (e.g. Prentice *et al.*, 1992).

Measurement of area

Surface area is defined as the sum of the area of all of the faces or surfaces of an object. Here, surface area of every grid cell was measured as a function of three main factors: (1) the area related to the position of each cell in the UTM reticule; (2) the area covered by the sea within every grid cell; and (3) the area added by variation in topographic ranges. These three factors were taken into account and measured for each grid cell (2431 grid cells) using surface areas and ratios from elevation grid version 1.2, SAREG (Jenness, 2004); methodological details of how surface area is measured are provided in http://www.jennessent.com/arcview/surface_areas.htm. A digital elevation model (DEM) of Europe with 1 km pixel resolution (GTOPO30, freely available at <http://edcdaac.usgs.gov/gtopo30/gtopo30.asp>), was used to calculate a three-dimensional surface area for each grid cell. The SAREG program converts the raster information of the DEM into a triangulated irregular network structure to calculate the surface

area of the grid. To assess the effect of topography contributing to area of grid cells, a surface ratio index (surface area/planimetric area) was used. We also measured the planimetric area of the cells and the planimetric area minus the area occupied by the sea in every grid cell. Finally, to summarize the percentage change in grid cell sizes we ranked cells as having: (1) deviation larger than 50% in relation to the planimetric area; (2) deviation from 50% to 25%; and (3) deviation from 25% to 5%.

Ideally, a finer resolution DEM should have been used to calculate surface area, as these measurements are bound to be sensitive to the resolution of input data. However, the processing power required for such analysis is beyond our current capacity. Logically, we would expect that the greater the resolution of the DEM, the greater the estimated surface area for any given grid cell. However, the shape of the response curve between surface area and the resolution of the DEM is unknown. In order to investigate this question we used a 100-m resolution DEM available for a 10 × 10 km square located in the Pyrenees (Instituto Geográfico Nacional, www.ign.es). The DEM was then resampled to resolutions of 250, 500 and 1000 m and surface area calculated for each one. In this particular case the estimated surface area increased exponentially with increasing pixel resolution (Fig. 1). The results of this meta-analysis may not be representative of all grid cells in Europe, as they are more likely to reflect patterns observed in mountains. Nevertheless, they invite the interpretation that the results of this study, using a 1-km DEM resolution,

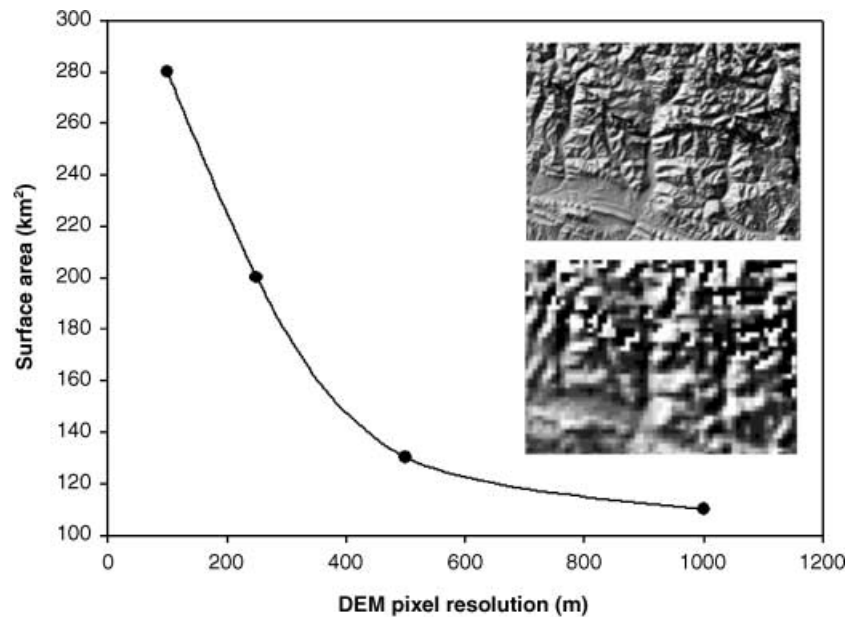


Figure 1 Relationship between pixel resolution and surface area in a $10 \times 10 \text{ km}^2$ plot in the Spanish Pyrenees. A 100-m pixel resolution DEM was resampled to 250 m, 500 m, and 1000 m and surface area measured for each resolution. Two hill-shade surfaces derived from the DEM of the area are also shown (top: 100 m; bottom: 1000 m).

might be conservative in relation to the real effect of variation in surface area of grid cells due to changes in topography.

Statistical analyses

Generalized additive models (GAMs) (Hastie & Tibshirani, 1990) were used to assess the effect of area of grid cells on variable selection and the model's predictive accuracy. GAMs are non-parametric extensions of generalized linear models that apply nonparametric smoothers to each predictor and calculate additively the component response. Being non-parametric and data-driven they circumvent the problem of making a priori assumptions on the species richness–area relationship. A GAM is expressed by:

$$g(E(Y)) = \alpha + s_1(X_{1i}) + s_2(X_{2i}) + \dots + s_p(X_{pi}) \quad (1)$$

where g is the link function that relates the linear predictor with the expected value of the response variable Y , X_{pi} is a predictor variable and s_p a smoothing function. A Poisson distribution was selected and log as link function. This is recommended, as species richness is often considered as a form of count data (Crawley, 1993). Explained deviance and chi-square tests were calculated to summarize the ability of selected variables to explain the spatial pattern of species richness and to provide statistical significance estimates of these relationships. The splines used to relate each climate variable to species richness were simplified to a maximum of 4 degrees of freedom (Wood & Augustin, 2002). The objective was to reduce complexity, to prevent possible over-estimations and to avoid hardly interpretable splines. The selection of predictive variables was based on a forward–backward selection method. The aim of this approach is to select a number of variables that maximize the explained deviance while reducing the GCV value (general cross-validation; see Wood & Augustin, 2002).

Four GAMs were fitted for each taxa. First, species richness of every group was regressed against the surface area of the grid cells. Secondly, species richness was regressed against climate variables using a stepwise variable selection procedure. Thirdly, species richness was regressed against area of the cells and climate variables using a stepwise variable selection procedure. Fourthly, the residuals of the first model (using area as a covariate) were regressed against climate variables; residuals were then regressed against climate variables using the calibration set and predicted values compared with original residual scores in the validation set. These analyses allowed us to investigate the effect of grid cell size on variable selection and model accuracy. We used $P > 0.005$ as a threshold to eliminate variables in the stepwise selection.

Models were calibrated on 70% random sample of the original data and evaluated against the remaining 30%. Willmott's D (Willmott, 1982) was selected as a measure of model accuracy. Alternative measures would include the mean absolute error or root mean square error. However, these measures give estimates of the average error, but fall short in providing information about the relative magnitude of the average difference. To overcome this problem Willmott proposed an 'index of agreement' of the form:

$$D = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P'_i| + |O'_i|)^2} \quad (2)$$

where N = number of observations, O = observed value, \bar{O} = mean of observed values, P = predicted value, I = counter for individual observed and predicted values, $P'_i = P_i - \bar{O}$ and $O'_i = O_i - \bar{O}$. Willmott's D varies from 0 to 1 (1 means a perfect prediction). Willmott's D was obtained to assess changes in model accuracy before and after controlling for the effects of area in the models.

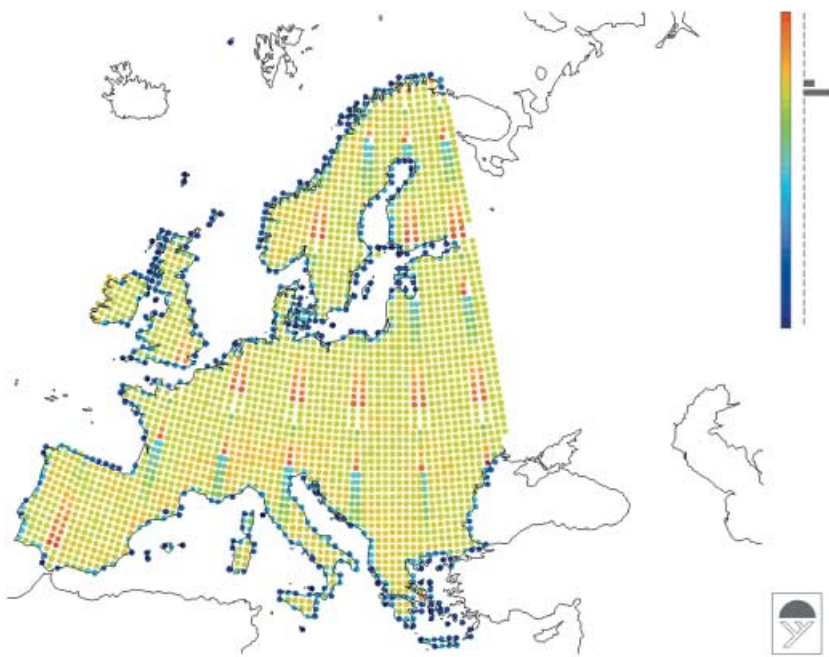


Figure 2 Surface area (km²) of the c. 50 × 50 km (2500 km²) grid cells used commonly in macroecological studies in western Europe. Scores are divided into 33 equal interval colour-scales, with maximum area scores shown in red (larger grid cells) and minimum area scores shown in blue (smaller grid cells). 57% of grid cells are included in the 2500 ± 125 km² range. WORLDMAP software (Williams, 1999) was used to produce this map.

RESULTS

Variation in grid cell size

The smallest grid cell in this study had a land area of 8 km² and was located in the south-west coast of England; the largest cell was 3311 km² and was located between the 32 and 33 UTM zone in the Italian Alps. The range of variation in 50 × 50 km grid cell size was 3303 km². Approximately 43% of the grid cells did not have the expected area of 2500 ± 125 km². As expected, variation in grid cell size was spatially structured. Larger grid cells were located in the intersection of UTM zones and in the mountains; smaller grids were found near the coasts (Fig. 2). The most important source of variation in grid cell size was the presence of the sea in coastal grids, followed by the geographical position of grid cells in the coordinate system used (Table 2). The effect of topography was small and affected only a small proportion of cells; an increase of up to 9% of grid cell area was recorded in some mountain regions. Variation in 220 × 220 km grid cell size was 54,900 km² and 82.5% of the cells did not have the expected area. The presence of the sea on coastal grids is the most important factor, followed by the geographical position of grid cells and the effect of topography (see Table 2 for a summary of results).

The effect of area on the selection of climate variables

For 50 × 50 km grid cells, the highest correlation between area and a climate variable was recorded with mean temperature of the coldest month ($r = 0.30$). This correlation did not affect variable selection with GAM: selected climate variables were identical for models with and without including area as a covariate (see Appendix S1 in Supplementary Material). This pattern was consistent for all groups. The *P*-values of climate variables did

Table 2 Proportion of grid cells (%) deviating from the expected size (2500 ± 125 km² for 50 × 50 km grids; 48,400 ± 2420 km² for 220 × 220 km grids)

	Sea	UTM position	Topography	All effects
(a) 50 × 50 km grid				
> 50% area deviation	13.5	0.0	0.0	13.3
50–25% area deviation	4.0	10.3	0.0	12.4
25–5% area deviation	4.6	12.5	2.5	17.9
Sum	22.1	22.8	2.5	43.6
(b) 220 × 220 km grid				
> 50% area deviation	22.2	0.0	0.0	32.7
50–25% area deviation	15.3	8.2	0.0	17.8
25–5% area deviation	31.3	9.0	6.0	31.9
Sum	68.8	17.2	6.0	82.5

not exceed the selected threshold ($P > 0.005$) when area was included in the stepwise process. The same patterns were recorded when the 220 × 200 km grid cells were considered, i.e. no changes in variable selection were recorded across all groups (see Appendix S2 in Supplementary Material).

The effect of area on model accuracy

Models including area as a covariate had small to moderate increases in predictive accuracy (Willmott's *D*-values) in the validation set in comparison with models that did not include area as a covariate: plants 0.78–0.81 ($\Delta = 0.03$); birds 0.68–0.74 ($\Delta = 0.06$); mammals 0.70–0.79 ($\Delta = 0.09$); and herptiles 0.82–0.83 ($\Delta = 0.01$). Similarly, models that controlled for area by regressing residuals of a species–area model had a decrease in

predictive accuracy on the validation set compared to models that only included climate variables. The highest reduction in model accuracy was recorded for plants (0.78–0.75; $\Delta = 0.03$) and birds (0.68–0.65; $\Delta = 0.03$), followed by mammals (0.70–0.68; $\Delta = 0.02$). These reductions are a consequence of eliminating the predictive power of area over species richness. Predictions for herptile species richness did not record reductions in accuracy on the validation set (0.82 before and after discounting the effect of area).

Reductions in model accuracy from not including area as a covariate in the regression model were slightly more important when the 220 × 220 km grid cells were considered: plants 0.84–0.93 ($\Delta = 0.09$); mammals 0.84–0.91 ($\Delta = 0.07$); birds 0.75–0.90 ($\Delta = 0.15$); and herptiles 0.82–0.92 ($\Delta = 0.10$). Models that controlled for area by regressing residuals of the species–area model had a decrease in predictive accuracy on the validation set compared to models that only included climate variables: birds (0.84–0.76; $\Delta = 0.08$), plants (0.84–0.78; $\Delta = 0.06$), followed by mammals (0.75–0.70; $\Delta = 0.05$) and herptiles (0.75–0.73; $\Delta = 0.02$).

We also assessed the ability of surface area to predict species richness. Unsurprisingly, the ability to which models calibrated with area alone predicted species richness in the validation set was low when using the 50 × 50 km grid cells (Willmott's *D*-values for birds = 0.46, herptiles = 0.11, mammals = 0.41, plants = 0.10; and 0.11, 0.03, 0.10 and 0.07, respectively, in terms of r^2), but was greater when using 220 × 220 km grid cells (Willmott's *D*-value for birds = 0.82, herptiles = 0.62, mammals = 0.78, plants = 0.73, and 0.57, 0.25, 0.39 and 0.46, respectively, in terms of r^2).

DISCUSSION

Studies investigating correlates of species richness often use near-equal area cells to control for the species–area relationship. It is implicit that the variation in surface area among grid cells is not sufficiently important to compromise the robustness of the analyses. Here we provide a first test of this idea using species richness scores for different groups of terrestrial vertebrates and plants superimposed onto European *c.* 50 × 50 km and *c.* 220 × 220 km grid systems that are often used to investigate species–distribution patterns in Europe (e.g. Huntley *et al.*, 1995; Lathi & Lampinen, 1999; Araújo & Williams, 2000; Williams *et al.*, 2000; Araújo *et al.*, 2001, 2004; Araújo, 2003; Araújo & Pearson, 2005; Thuiller *et al.*, 2005). We found that 43% of the 50 × 50 km cells and 82.5% of the 220 × 220 km cells in this study did not have the expected $2500 \pm 125 \text{ km}^2$ and $48,400 \pm 2420 \text{ km}^2$, respectively. The influence of coastal areas extended over 22.1% of the *c.* 50 × 50 km European grid, 13.5% of which had a reduction in area greater than 50%; when larger grid cells were considered (i.e. 220 × 220 km), coastal cells also exerted a greater influence (68.8%), 22.2% of which had a reduction of area greater than 50%. The geographical position of grid cells along the UTM grid was responsible for changes in area by 22.8% for the 50 × 50 km grid cells and 17.2% for the 220 × 220 km grid cells, although these changes in area were less pronounced than for coastal areas. Finally, variations in altitudinal range affected only 2.5% of the

50 × 50 km grid cells and 6.0% of the 220 × 220 km grid cells. In both cases, the reduction in surface area of the grid cells was less than 25% of the expected area (Table 2). Despite considerable variation in the surface area of grid cells, the selection of climate variables entering models of species richness was not affected by area in both the 50 × 50 km and the 220 × 220 km grid cell resolution (see Appendices S1 and S2 in Supplementary Material). We also found that the accuracy of models on the validation set was affected only moderately by variation in the size of the grid cells although the effects were slightly greater in the case of 220 × 200 km grid cells. There was some cross-taxa variation, with models of herptile species richness not being affected by variations in surface area and models for mammal species richness being relatively more affected by area. Despite cross-taxa variation, our results provide provisional support for the assumption that no major biases in species richness are incurred from using near-equal area instead of truly equal-area grids.

However, there is a possibility that this reduced effect of variation in grid cell size might be magnified when (1) larger grids are considered, or (2) when finer-resolution DEMs are used to calculate surface area. The possibility that controlling for area might be more important with larger grids was illustrated by our results, where models fitted at the larger 220 × 220 km resolution had consistent albeit small reductions in accuracy when area was not explicitly accounted for within models. These results provide moderate support for studies using known relationships between climate and species richness for predictive purposes, although there is evidence that increasing the grid cell size might result in a loss of model's predictive ability unless surface area is explicitly controlled for (see also Rahbek & Graves, 2001). The possibility that varying the resolution of DEM might affect the perceived relationships between species richness and climate variables also needs further investigation. As shown in Fig. 1, surface area increases with DEM resolution exponentially. This is an analogous result to that obtained for the Mandelbrot coast, where the length of UK coast increased with spatial resolution. Therefore, we anticipate that the effect of topography on the area of the grid cells may be more important when using finer-resolution DEM; and expectedly, this effect should be greater in mountainous areas. In the particular case of our analyses, we have shown that changes in surface area of grid cells were due mainly to the area-reducing effect of coastal areas, while, contrary to our initial expectations, the area-increasing effect of topography was more limited. For example, for the 48,400 km² grid cells (220 × 220 km), the presence of coastal areas reduced grid cell area by more than 99% (the smallest coastal 220 × 220 km cell was 193 km²), whereas topography was responsible only for increases in area of 9% in some mountainous zones when using 1-km pixel resolution DEM. Although larger increases in estimated surface area as a function of topographic variation can be expected with increasing resolution of DEMs, as we show here in Fig. 1 (and see Turner, 2004), it is unlikely that the impact of topography on model accuracy will be as great as the reduction in area through overlap of coastal cells with the sea.

The recorded stability of variable selection is encouraging for studies investigating species richness correlates. Most studies

using similar grid cell systems do not measure variation in the surface area of near-equal area grids, and when this is attempted studies fail to account for variations in area associated with changes in topography (for an exception see O'Brien *et al.*, 2000), a shortcoming that is shown here to be possibly negligible in most circumstances. Naturally, our results are contingent on the particular data set, extent and grain used. Using grid cells with grains of different sizes, or study areas with different extent, could cause the results of this study to change to a considerable, albeit unknown, extent (see Kaufman & Willig, 1998; Rahbek & Graves, 2001). Although the range of variation in the surface area of near-equal area grid cells was greater for cells with larger grain size, i.e. 55,000 km² for the 220 × 220 km grid compared with 3303 km² for the 50 × 50 km grid (for more details see Table 2b), no changes in variable selection were reported at either resolution.

In any case, it is important to bear in mind that our results provide a necessarily overly optimistic assessment of the predictive ability of the models as a function of the nature of the validation set. We use a random subset of the calibration set for validation and the extent to which this provides an independent, i.e. not autocorrelated, set for validation is debatable (e.g. Araújo *et al.*, 2005a). Previous studies assessing accuracy of species richness models did not attempt to split calibration and validation sets (e.g. Lobo *et al.*, 2002; Hortal *et al.*, 2004), hence results from these studies are not comparable to ours. Nevertheless, there are good reasons to consider that most studies of species richness correlates have an explanatory value rather than a predictive one. Correlation does not imply causation and the debate over mechanisms driving global patterns of variation in species richness is still a matter of contentious debate (e.g. Hawkins *et al.*, 2003; Colwell *et al.*, 2004; Currie & Francis, 2004; Qian & Ricklefs, 2004; Field *et al.*, 2005). Using regression-type approaches to make predictions of species richness on non-independent validation sets will not bring much light to this debate.

ACKNOWLEDGEMENTS

We thank the national atlas representatives and the many volunteer fieldworkers who contributed the atlas records; J.P. Gasc (herptiles), W.J.M. Hagemeyer (breeding birds), Raino Lampinen (plants), and A.J. Mitchell-Jones (mammals) for kindly supplying European species distribution data; Joaquin Hortal, Niall O'Dea, Robert J. Whittaker and two anonymous referees for discussion and comments on the manuscript. This research is partly funded by the European Commission's FP6 ALARM (GOCE-CT-2003-506675) project.

REFERENCES

- Aava, B. (2001) Can resource use be the link between productivity and species richness in mammals? *Biodiversity and Conservation*, **10**, 2011–2022.
- Araújo, M.B. (2003) The coincidence of people and biodiversity in Europe. *Global Ecology and Biogeography*, **12**, 5–12.
- Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.
- Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.
- Araújo, M.B., Humphries, C.J., Densham, P.J., Lampinen, R., Hagemeyer, W.J.M., Mitchell-Jones, A.J. & Gasc, J.P. (2001) Would environmental diversity be a good surrogate for species diversity? *Ecography*, **24**, 103–110.
- Araújo, M.B., Cabeza, M., Thuiller, W., Hannah, L. & Williams, P. (2004) Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, **10**, 1618–1626.
- Araújo, M.B., Pearson, R.J., Thuiller, W. & Erhard, M. (2005a) Validation of species–climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.
- Araújo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2005b) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, **14**, 17–30.
- Badgley, C. & Fox, D.L. (2000) Ecological biogeography of North American mammals: species density and ecological structure in relation to environmental gradients. *Journal of Biogeography*, **27**, 1437–1467.
- Bárcena, S., Real, R., Olivero, J. & Mario Vargas, J. (2004) Latitudinal trends in breeding waterbird species richness in Europe and their environmental correlates. *Biodiversity and Conservation*, **13**, 1997–2014.
- Bini, L.M., Diniz-Filho, J.A. & Hawkins, B.A. (2004) Macroecological explanations for differences in species richness gradients: a canonical analysis of South American birds. *Journal of Biogeography*, **31**, 1819–1827.
- Colwell, R.K., Rahbek, C. & Gotelli, N.J. (2004) The mid-domain effect and species richness patterns: what have we learned so far? *The American Naturalist*, **163**, E1–E23.
- Connors, E.F. & McCoy, E.D. (1979) The statistics and biology of the species–area relationships. *The American Naturalist*, **113**, 792–833.
- Crawley, M.J. (1993) *GLIM for Ecologists*. Blackwell Scientific Publications, Oxford.
- Currie, D.J. (1991) Energy and large-scale patterns of animal and plant species richness. *The American Naturalist*, **137**, 27–49.
- Currie, D.J. & Francis, A.P. (2004) Regional versus climatic effect on taxon richness in angiosperms: reply to Qian and Ricklefs. *The American Naturalist*, **163**, 780–785.
- Diniz-Filho, J.A., Thiago, F.L., Rangel, V.B. & Hawkins, B.A. (2004) A test of multiple hypotheses for the species richness gradient of South American owls. *Oecologia*, **140**, 633–638.
- Field, R., O'Brien, E.M. & Whittaker, R.J. (2005) Global models for predicting woody plant richness from climate: development and evaluation. *Ecology*, **86**, 2263–2277.
- Gasc, J.-P., Cabela, A., Crnobrnja-Isailovic, J., Dolmen, D., Grossenbacher, K., Haffner, P., Lescure, J., Martens, H., Martinez-Rica, J.P., Maurin, H., Oliveira, M.E., Sofianidou, T.S., Veith, M. & Zuiderwijk, A. (1996) *Atlas of amphibians and reptiles in Europe*. Societas Europaea Herpetologica & Museum National d'Histoire Naturelle, Paris.

- Gaston, K.J. & Blackburn, T.M. (2000) *Pattern and process in macroecology*. Blackwell Science, Oxford.
- H-Acevedo, D.H. & Currie, D.J. (2003) Does climate determine broad-scale patterns of species richness? A test of the causal link by natural experiment. *Global Ecology and Biogeography*, **12**, 461–473.
- Hagemeijer, W.J.M. & Blair, M.J. (1997) *The EBCC atlas of European breeding birds, their distribution and abundance*. Poyser, London.
- Hastie, T. & Tibshirani, R. (1990) *Generalized additive models*. Chapman & Hall, London.
- Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.-F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.
- Hawkins, B.A. & Porter, E.E. (2001) Area and the latitudinal diversity gradient for terrestrial birds. *Ecology Letters*, **4**, 595–601.
- Hawkins, B.A. & Porter, E.E. (2003) Water–energy balance and the geographic pattern of species richness of western Palearctic butterflies. *Ecological Entomology*, **28**, 678–686.
- Hortal, J., García-Pereira, P. & García-Barros, E. (2004) Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns. *Ecography*, **27**, 68–82.
- Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography*, **22**, 967–1001.
- Hurlbert, A.H. & White, E.P. (2005) Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecology Letters*, **8**, 319–327.
- Jalas, J. & Suominen, J. (1972–96) *Atlas Florae Europaeae*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo.
- Jenness, J.S. (2004) Calculating landscape surface area from digital elevation models. *Wildlife Society Bulletin*, **32**, 829–839.
- Jetz, W. & Rahbek, C. (2001) Geometric constraints explain much of the species richness pattern in African birds. *Proceedings of the National Academy of Sciences USA*, **98**, 5661–5666.
- Kaufman, D.M. & Willig, M.R. (1998) Latitudinal patterns of mammalian species richness in the New World: the effects of sampling method and faunal group. *Journal of Biogeography*, **25**, 795–805.
- Kerr, J.T. (1999) Weak links: 'Rapoport's rule' and large-scale species richness patterns. *Global Ecology and Biogeography*, **8**, 47–54.
- Kunin, W.E. (1998) Extrapolating species abundance across spatial scale. *Science*, **281**, 1513–1515.
- Lathi, T. & Lampinen, R. (1999) From dot maps to bitmaps: *Atlas Florae Europaeae* goes digital. *Acta Botanica Fennica*, **162**, 5–9.
- Lobo, J.M., Lumaret, J.P. & Jay-Robert, P. (2002) Modelling the species richness distribution of French dung beetles (Coleoptera, Scarabaeidae) and delimiting the predictive capacity of different groups of explanatory variables. *Global Ecology and Biogeography*, **11**, 265–277.
- Lyons, S.K. & Willig, M.R. (1999) A hemispheric assessment of scale dependence in latitudinal gradients of species richness. *Ecology*, **80**, 2483–2491.
- McKinney, M.L. (2002) Do human activities raise species richness? Contrasting patterns in United States plants and fishes. *Global Ecology and Biogeography*, **11**, 343–348.
- Meliadou, A. & Troumbis, A.Y. (1997) Aspects of heterogeneity in the distribution of diversity of the European herpetofauna. *Acta Oecologica*, **18**, 393–412.
- Mitchell-Jones, A.J., Amori, G., Bogdanowicz, W., Krystufek, B., Reijnders, P.J.H., Spitzenberger, F., Stubbe, M., Thissen, J.M.B., Vohralik, V. & Zima, J. (1999) *Atlas of European mammals*. Academic Press, London.
- O'Brien, E.M., Field, R. & Whittaker, R.J. (2000) Climatic gradients in woody plant (tree and shrub) diversity: water–energy dynamics, residual variation, and topography. *Oikos*, **89**, 588–600.
- Patten, M.A. (2004) Correlates of species richness in North American bat families. *Journal of Biogeography*, **31**, 975–985.
- Pearson, D.L. & Carroll, S.S. (1999) The influence of spatial scale on cross-taxon congruence patterns and prediction accuracy of species richness. *Journal of Biogeography*, **26**, 1079–1090.
- Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R. & Monserud, R.A. & Solomon, A.M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117–134.
- Qian, H. & Ricklefs, R.E. (2004) Taxon richness and climate in angiosperms: is there a globally consistent relationship that precludes region effects? *The American Naturalist*, **163**, 773–779.
- Rahbek, C. (2005) The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, **8**, 224–239.
- Rahbek, C. & Graves, G.R. (2000) Detection of macro-ecological patterns in South American hummingbirds is affected by spatial scale. *Proceedings of the Royal Society London Series B*, **267**, 2259–2265.
- Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences USA*, **98**, 4534–4539.
- Rangel, T.F.L.V.B. & Diniz-Filho, J.A. (2003) Spatial patterns in species richness and the geometric constraint simulation model: a global analysis of mid-domain effect in Falconiformes. *Acta Oecologica*, **24**, 203–207.
- Ribera, I., Foster, G.N. & Vogler, A.P. (2003) Does habitat use explain large scale species richness patterns of aquatic beetles in Europe? *Ecography*, **26**, 145–152.
- Rosenzweig, M.L. (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge.
- Taplin, J. & Lovett, J.C. (2003) Continental scale patterns of biodiversity: can higher taxa accurately predict African plant distributions? *Botanical Journal of the Linnean Society*, **142**, 187–197.
- Thuiller, W., Lavorel, S. & Araújo, M.B. (2005) Niche properties as predictors of species sensitivity to climate change. *Global Ecology and Biogeography*, **14**, 347–357.

- Tognelli, M. & Kelt, D.A. (2004) Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography*, **27**, 427–436.
- Turner, J.R.G. (2004) Explaining the global biodiversity gradient: energy, area, history and natural selection. *Basic and Applied Ecology*, **5**, 435–448.
- Whittaker, R.J., Willis, K.J. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453–470.
- Williams, C.B. (1964) *Patterns in the balance of nature*. Academic Press, New York.
- Williams, P.H. (1999) *WORLDMAP 4 Windows: software and help document 4.2*. Distributed privately, London.
- Williams, P.H., Humphries, C., Araújo, M.B., Lampinen, R., Hagemeyer, W., Gasc, J.-P. & Mitchell-Jones, T. (2000) Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian Journal of Entomology*, **2**, 21–46.
- Willmott, C. (1982) Some comments on the evaluation of model performance. *American Meteorological Society*, **63**, 1309–1313.
- Wood, S.N. & Augustin, N.H. (2002) GAMs with integrated model selection using penalized regression splines applications to environmental modelling. *Ecological Modelling*, **157**, 157–177.

BIOSKETCHES

David Nogués Bravo has a PhD from the University of Zaragoza. He is a post-doctoral research fellow at the University of Évora, and a visiting postdoctoral research fellow at the University of Copenhagen. His research focuses on the relationships between climate, species distribution and human impacts at large scales, including future impacts of climate and land-uses changes on biodiversity.

Miguel B. Araújo has a PhD from the University of London. He is Research Fellow at the National Museum of Natural Sciences in Madrid, a senior research associate of the Oxford University Centre for the Environment, and a visiting associate professor at the University of Copenhagen. His main interests include an investigation of the processes affecting the occurrence and persistence of species at different spatial and temporal scales. He is also interested in the development of effective conservation-planning tools to prevent local extinctions of species.

Editor: David Currie

SUPPLEMENTARY MATERIAL

The following material is available online at www.blackwell-synergy.com/loi/geb

Appendix S1 Summary of GAM results for data at 50 × 50 km grid resolution.

Appendix S2 Summary of GAM results for data at 220 × 220 km grid resolution.