# Can Species Richness Patterns Be Interpolated From a Limited Number of Well-Known Areas? Mapping Diversity Using GLM and Kriging

Joaquín Hortal[1,2,*] & Jorge M. Lobo[1]

[1]*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain*
[2]*Departamento de Ecologia, Instituto de Ciências Biológicas – ICB, Universidade Federal de Goiás – IFG, Goiânia, GO, Brazil*

## Abstract

The limitations of biodiversity data are commonly overcome by modelling the geographic distribution of species and community characteristics. Here we evaluate two Assemblage-level Modelling (ALM) techniques, General Linear Models (GLM) and kriging, assessing their ability to predict scarab dung beetle richness in the Iberian Peninsula using two different strategies. Calibration Errors (ability to interpolate values within the conditions where the model was built) were assessed by means of a leave-one--out jackknife. Validation Errors (ability to provide partial extrapolations to different environmental conditions within the same geographic domain) were calculated by comparing model projections with an independent dataset. Although the forecasts within the calibration dataset were very good for GLM and extremely good for kriging, both techniques provided surprisingly poor extrapolations. We discuss why such poor performance may be related to non-stationarity in the factors driving diversity patterns, and how ALM may be improved to account for it.

**Key words:** Diversity, Geostatistics, Macroecological Models, Model Performance, Scarabaeidae.

## Introduction

Our knowledge of the geographic distribution of biological diversity is relatively scarce and geographically patchy (i.e. the so-called "Wallacean shortfall"; Lomolino 2004). After more than two centuries of taxonomic tradition, we lack data on the distribution of a large proportion of the diversity of life, with the exception of a few groups (mainly vascular plants and vertebrates) and some geographical regions (particularly central and northern Europe and North America). Typically, the information gathered in biodiversity databases provides an unreliable picture of the distribution of diversity, plagued with geographical and taxonomic gaps, limitations and biases (see Rocchini *et al.* 2011 and references therein). However, such knowledge is needed to both (i) describe and study the causes of the geographic distribution of biodiversity, and to (ii) design effective conservation strategies and protected area networks (Ferrier 2002; Hortal & Lobo 2006) when reliable information on species composition is lacking.

A solution to this lack of information could be the use of predictive models of the distribution of diversity. Many modelling techniques have been proposed so far, mainly grouped in two kinds of approaches; (i) Species Distribution Models (SDM) intend to predict the spatial distribution of single species from data on their occurrences and an array of—often environmental—predictors (Guisan & Zimmermann 2000); (ii) Assemblage-level Models (ALM)—also called Synecological Models (Hortal & Lobo 2006), Community-level Models (Ferrier & Guisan 2006) or Macroecological Models (Guisan & Rahbek 2011)—aim to represent the spatial variations in the diversity of whole assemblages based on data from a few well-known areas and the corresponding set of predictors (Austin *et al.* 1996). While SDM techniques are widely used to project species distributions into different geographical and temporal scenarios (see Lobo *et al.* 2010), ALMs receive less attention as predictive tools, being more used to study the relationships between diversity and environment (*e.g.* Lobo *et al.* 2001; see also Guisan & Rahbek 2011).

Our main aim is to evaluate the capacity of two different ALM techniques—General Linear Models (GLM) and kriging—to map the distribution of species richness from

*****Send correspondence to:** Joaquín Hortal
Departamento de Biodiversidad y Biología Evolutiva,
Museo Nacional de Ciencias Naturales (CSIC),
C/ José Gutiérrez Abascal 2, 28006 Madrid, Spain
e-mail: jhortal@mncn.csic.es

a reduced set of territorial units with complete inventories. These two techniques present fundamental differences in the predictors they use. GLMs are used to model the relationships between the dependent variable and the predictors (Austin *et al.* 1996), providing forecasting models (*sensu* Legendre & Legendre 1998). Kriging is a geostatistical technique that models just the spatial structure in the data from observations of its value at locations placed nearby one from another (Cressie 1993). GLMs are commonly used to model the relationships of species richness with environmental variables and other predictors, either within macroecological analyses or to forecast its geographic distribution (*e.g.* Austin *et al.* 1996; Lobo *et al.* 2001; Lobo & Martín-Piera 2002). Kriging is much less used for this purpose (but see *e.g.* Ter Steege *et al.* 2003; Parmentier *et al.* 2011).

We assess the adequacy and reliability of GLM and kriging for mapping the species richness of scarab dung beetles (Coleoptera, Scarabaeidae) in the Iberian Peninsula. We have chosen species richness because is considered the basic parameter of biodiversity, and although its value for conservation planning is relatively limited (see *e.g.* Ferrier 2002), it provides a first approximation to the geographic variations of species diversity in the absence of good quality data on species distributions or composition, which is more prone to error (see Hortal *et al.* 2007; Pineda & Lobo 2009; Aranda & Lobo 2011 for comparisons of both kinds of data). Being based in a simpler variable, it also provides a simpler way of analyzing the effects of data quality on model reliability; we assume that the results of this study can be extrapolated to ALM applications based on other aspects of biodiversity (see Hortal & Lobo 2006; Guisan & Rahbek 2011). Thus, here we compare the adequacy of GLM and kriging to interpolate richness values. To do this, we use a set of grid cells with well-known inventories to calibrate both modelling techniques, and validate the accuracy of their predictions of the species richness of a completely independent group of cells. Here it is important to note that the boundary between interpolation and extrapolation is not well defined. While interpolations aim to *forecast* the values of the dependent variable to locations placed within the range of the observations, extrapolations try to *predict* these values outside the range of known observations; the latter are subject to a greater rate of uncertainty, and require identifying—either true or assumed—causal relationships between predictors and dependent variable (Legendre & Legendre 1998). Our analysis refers only to the forecast of values within a coherent geographic domain, the Iberian Peninsula.

## Methods

### *Origin and geographic coverage of data*

Dung beetle data comes from BANDASCA (Lobo & Martín-Piera 1991), a database that includes all distributional information available for the 53 species of Scarabaeidae present in the Iberian Peninsula (data available at http://es.mirror.gbif.org/datasets/resource/280/). Records in BANDASCA were referred to the 252 Iberian 50 × 50 km UTM grid squares with more than 15% of land surface (herein called UTM50 or grid cells for short; Figure 1a). Our analysis was based on the same data as Lobo & Martín-Piera (2002), which included the 15,740 records (101,996 individuals) that could be assigned accurately to a single UTM50. We merged the data from three of the 255 grid cells used by them, which had less than 15% of land surface, with their adjacent UTM50. Lobo & Martín-Piera (2002) used species accumulation curves to relate the sampling effort carried out on each cell and the number of species discovered (see also Hortal & Lobo 2005), identifying 82 UTM50 as being well-sampled (Lobo & Martín-Piera 2002). They later excluded seven of these cells, because they either pertained to the Balearic Islands or were identified as outliers due to oversampling (around Madrid and Barcelona). Data on the observed species richness of scarab dung beetles in the remaining 75 UTM50 will be used here as calibration dataset for the Assemblage-level Modelling of species richness (Figure 1b). In a later analysis using an updated version of BANDASCA and a different protocol to assess the quality of the inventories, Lobo (2008) identified 89 well-sampled UTM50. Twenty-two of these grid cells were not included within the 75 used by Lobo & Martín-Piera (2002), and will be used here as independent evaluation dataset, herein called validation dataset (Figure 1b).

We assessed the coverage of the environmental and geographical variability of the Iberian Peninsula provided by the well-sampled UTM50 by means of their ED coverage (that is, coverage of the overall *Environmental Diversity*; see Hortal & Lobo 2005 and references therein). Here, the overall coverage of the environmental and/or spatial variability of a territory provided by a selected subset of areas is calculated as the sum of all the distances from each non-selected area to the selected subset; the lower such sum of distances, the larger is the coverage of the regional variability, in a typically decreasing curve. The environmental variability matrix was calculated as the squared Euclidean distance between UTM50 cells according to their values of 13 variables related to soil, relief and climate (see variable details at Lobo & Martín-Piera 2002). The geographical variability matrix was calculated as the squared Euclidean distances between the centroids of all UTM50.

### *Assemblage-level modelling*

GLM predictions come from Lobo & Martín-Piera (2002) and kriging analyses were developed for this article. Briefly, GLM was used to relate species richness with 24 predictors accounting for soil, relief, climate, land use, habitat diversity and geographical location, assuming a Poisson distribution for species richness and a logarithmic function as link between it and the predictors. The model was built through
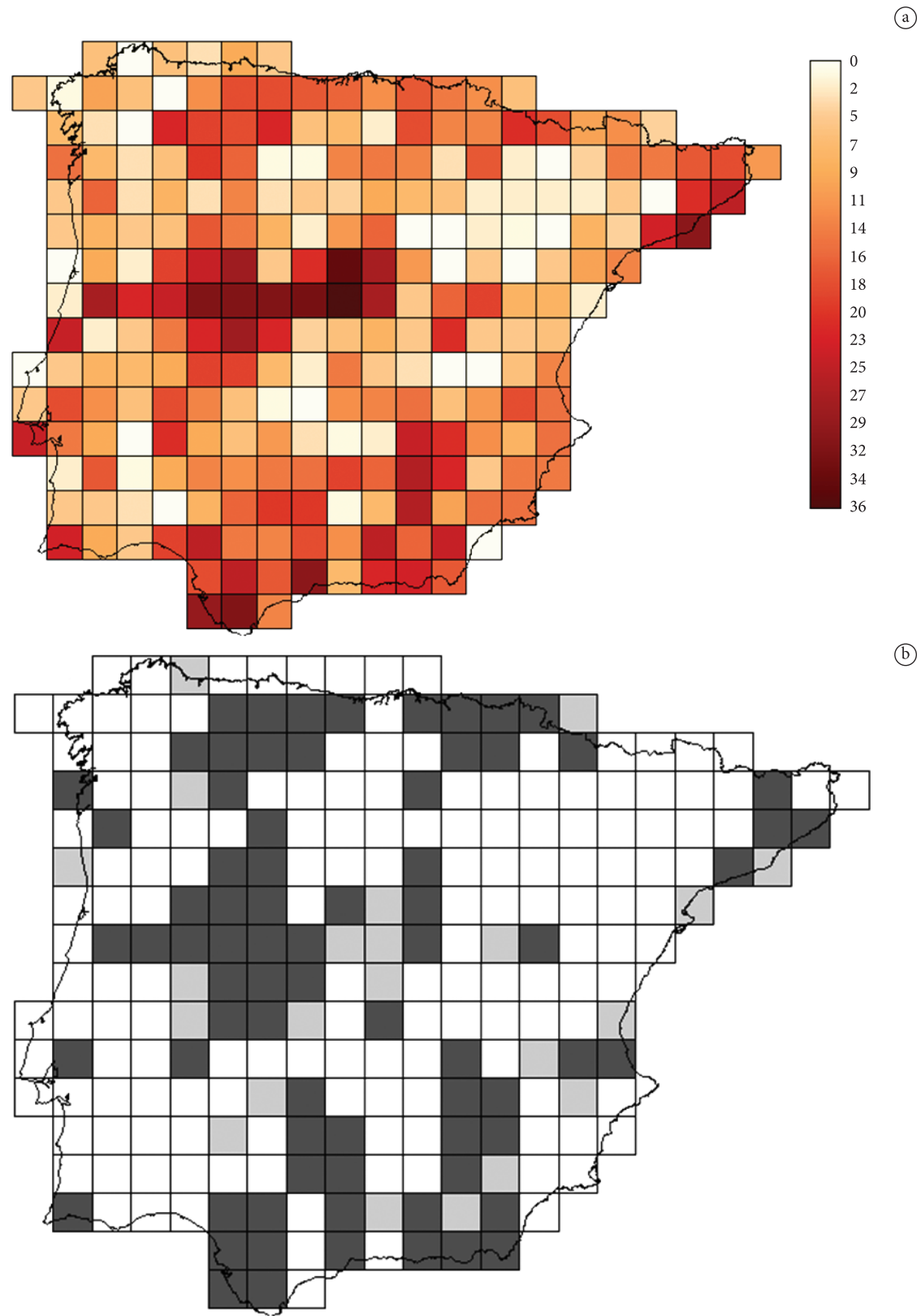
**Figure 1.** Information on dung beetle diversity provided by BANDASCA database for the 252 Iberian 50 × 50 km UTM grid squares used in this work: a) Observed species richness; b) Location of well-sampled UTM50 cells; dark grey – cells identified as well-sampled by Lobo & Martín-Piera (2002), used here as calibration dataset; light grey – additional cells identified as well-sampled by Lobo (2008), used here as empirical validation dataset.

an iterative mixed forward/backward stepwise procedure; in each step a new predictor was added to the model according to their change in deviance and the resulting model was subject to a backward analysis to eliminate the non-significant ones. This process was repeated until no more variables were significant enough to enter into the final model, which was a function of maximum elevation, grassland area, land use diversity, forest area, geological diversity, terrestrial area, surface of calcareous rocks and latitude. See Lobo & Martín-Piera (2002) for further details on the modelling process.

Kriging is the common mapping tool in the toolbox of geostaticians (Legendre & Legendre 1998), because it provides a powerful way to include both local and regional trends in the interpolated maps. It produces a statistically optimum estimate—the mean difference among observed and predicted values is 0, and its variance is minimal—through modelling the spatial dependence among observations (Cressie 1993). Briefly, the spatial dependence among the observations is first identified by a semi-variogram (or simply variogram), a mathematical function that describes the semi-variance at sequential distance classes. Semi-variance is a measure of the variability in the scores of a variable placed at a given distance interval. It thus provides a description of the spatial autocorrelation in the variable alternative to the most-used-in-ecology correlogram, which measures the correlation between values in the focal point and at each distance class. The semi-variance values of the variogram are fitted to a mathematical function, usually to a spherical model, although other models such as linear, exponential or Gaussian may give better results depending on the spatial structure of the data (Rossi *et al.* 1995). The resulting model is used to interpolate the values of the dependent variable. In our case, species richness values were lognormal-transformed to normalize its distribution. According to the variogram (not shown), dung beetle species richness values showed significant spatial dependence through the first 150 km. To account for directional differences in spatial dependence (i.e., non-stationarity), a series of anisotropic variograms were constructed using 150 km as lag distance, to identify the axis of maximum variation. There was little variation in the length of the axes of the anisotropic variograms, so we used an almost isotropic variogram with 150 km of lag distance to fit all models. The spheric model $\mathrm{Ln}(S) = 0.001200 \, \mathrm{Nug}(0) + 0.0693 \, \mathrm{Sph}(148000, 180, 0.13)$ accounted for the highest explained variability, so we used it to predict species richness values. Here, Nug stands out for the *nugget*, the internal variability of the dependent variable within the lag distance (i.e. less than 150 km), and Sph stands out for an spherical model of three parameters defining, respectively, the *range* of significant spatial dependence (in km), the *sill* (i.e. a plateau in the variogram starting at the distance where there is no significant spatial dependence) and the *anisotropy ratio* (i.e. the ratio between the distance of significant spatial dependence in the major and minor anisotropy axes). All geostatistical analyses were conducted in the Idrisi Andes GIS software (Clark Labs 2006).

## *Predictive power and reliability of the models*

We evaluated the reliability of GLM and kriging models by assessing their capacity to: (i) forecast species richness values within the calibration dataset (Calibration Error or CE); and (ii) estimate the values in the UTM50 cells not used in the training process (Validation Error or VE). We calculated CE through a Jackknife procedure (see Lobo & Martín-Piera 2002), where each UTM50 used to fit the model is left out once, and model parameters are recalculated without it. The Prediction Error (PE) for such UTM50 is the percentage of difference between the values observed in the focal cell and the predictions of the model excluding such cell; an CE is then calculated as the average of all PE values from this resubstitution process. Similarly, VE was calculated by comparing model predictions with the 22 cells of the independent validation dataset, which were not used to calibrate the model. Here, PE for each of these cells is calculated as the percentage of difference between the values observed and model predictions, and VE as the average of these values.

## Results

The seventy-five UTM50 of the calibration dataset cover 51.42% of the geographical variability of the Iberian Peninsula and 52.90% of its environmental variability. The twenty-two cells of the validation dataset increased such coverage to 63.02 and 64.61% of the spatial and environmental variability, respectively, showing that although the predictions of the models into these cells are a forecast within the geographical domain of the Iberian Peninsula, they are also in part an extrapolation (i.e. prediction) into new environmental and spatial domains.

The predictions obtained with GLM and kriging were highly concordant in the 75 cells of the calibration dataset (Spearman $R = 0.744$, $p < 0.001$), but such concordance was much smaller in the whole extent of 252 UTM50 (Spearman $R = 0.394$, $p < 0.001$). These inconsistencies reflect that there are important differences in their predictions; although both GLM and kriging provide relatively similar species richness maps, kriging predictions were in general smaller in the Central Iberian Plateaus, parts of the north and most of the eastern coast (Figure 2). GLM tends to overestimate richness values in the poorest UTM50 and underestimate it in the richest ones (Figure 3a), which is somewhat expected because regression methods (such as GLM) estimate central trends in the dependent variable rather than its extreme variations. However, according to the jackknife analyses the Calibration Error of this model is small, with a high correlation between observed and predicted values and a high Predictive Power (Table 1). The performance of kriging in the calibration dataset was nearly perfect, with no apparent bias in the predictions (Figure 3a), and extremely good predictions (Table 1).
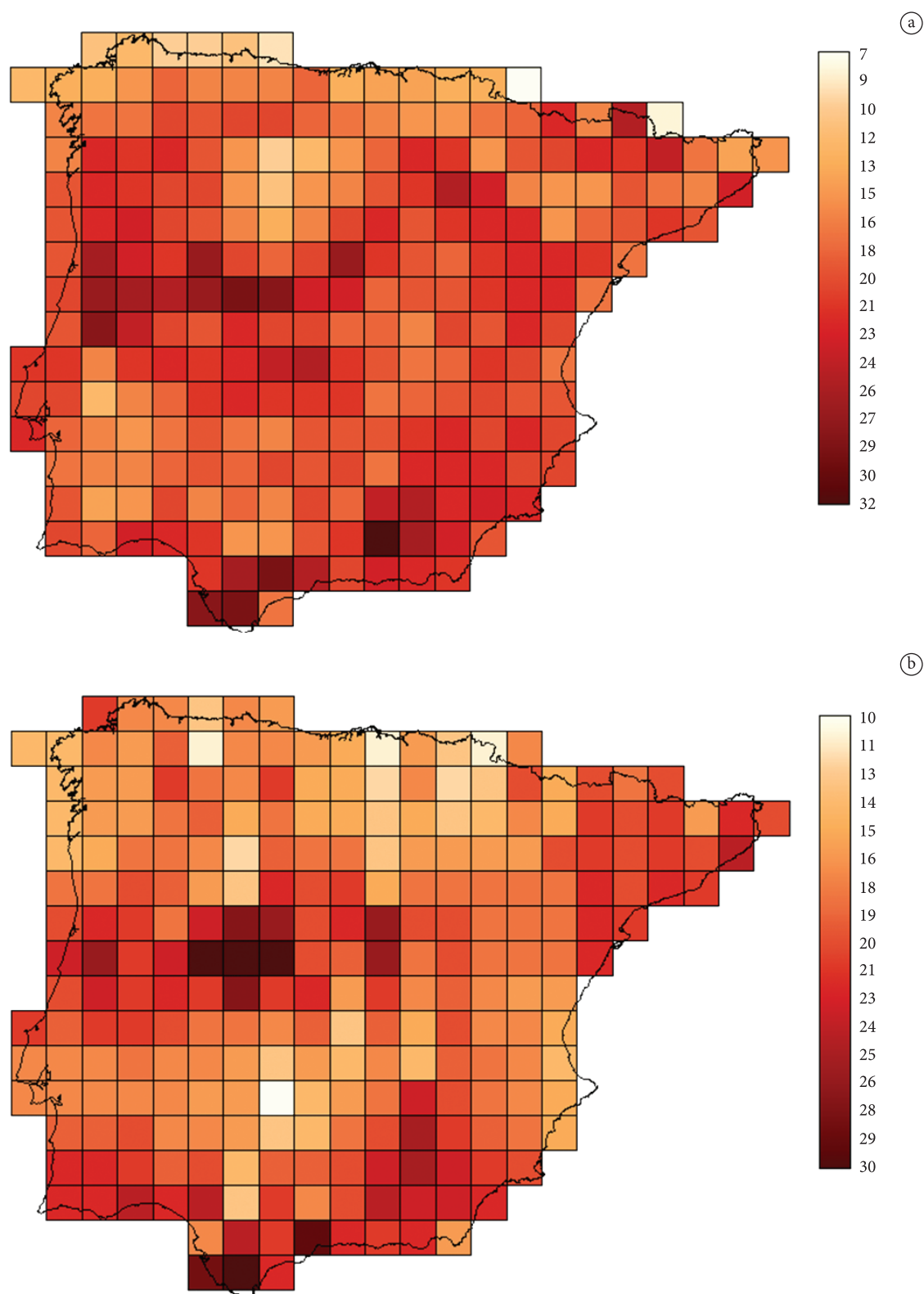
**Figure 2.** Species richness of Scarabaeidae dung beetles on the 252 Iberian 50 × 50 km UTM grid squares used in this work, as predicted by: a) General Linear Models; and b) kriging. See text for details on the interpolation methods and the accuracy of the results.
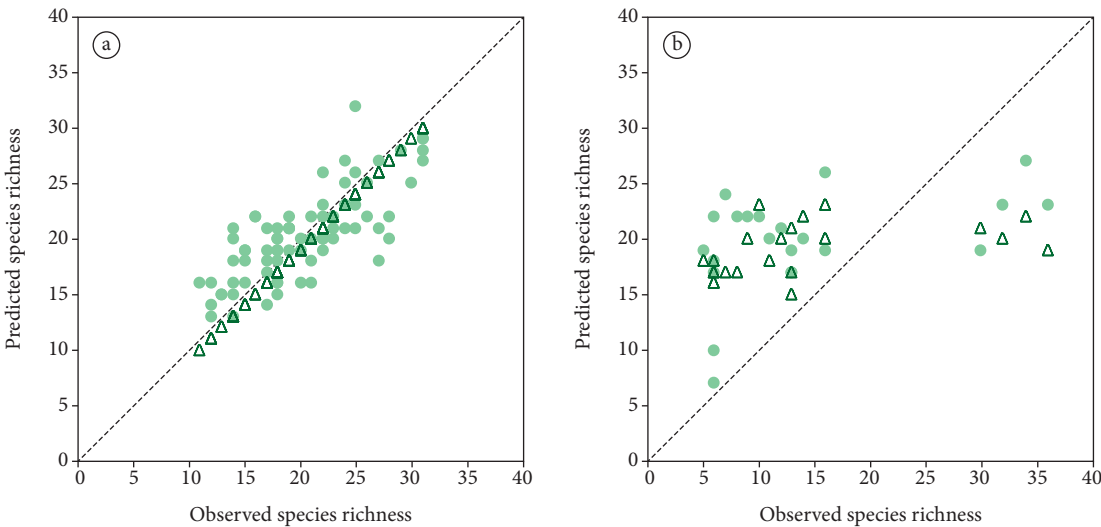
**Figure 3.** Comparison between observed and predicted species richness values for the GLM (green circles) and kriging (empty triangles) models: a) within the 75 cells of the calibration dataset; and b) in the 22 cells of the validation dataset.

**Table 1.** Accuracy of model predictions in the calibration and validation datasets.

| | Obs. vs. pred. | Predictive power | Prediction errors | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | **MPE** | **Min** | **25%Q** | **Median** | **75%Q** | **Max** |
| CE GLM | 0.744*** | 85.78 | 14.22 ± 11.16 | 0 | 6.17 | 11.11 | 17.91 | 50 |
| CE Kriging | 1*** | 94.74 | 5.26 ± 1.41 | 3.23 | 4.17 | 5 | 5.88 | 9.09 |
| VE GLM | 0.442* | -0.26 | 100.26 ± 86.70 | 16.67 | 32.10 | 64.58 | 167.36 | 280 |
| VE Kriging | 0.585** | 0.84 | 99.16 ± 70.38 | 15.38 | 39.06 | 65.15 | 160.71 | 260 |

Calibration Errors (CE) correspond to the results of a leave-one-out jackknife within the 75 UTM 50 x 50 km cells used to calibrate the model. Validation Errors (VE) stand for the comparison between observed and predicted values in the 22 cells of the independent validation dataset. Obs. vs. pred. are the results of Spearman $R$ correlations between observed and predicted values; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Predictive Power and Prediction Errors are calculated as percentages of error from the residuals of comparing observed and predicted values (see text); MPE is Mean Prediction Error (± Std.Dev.), and Min, Max, 25%Q and 75% are the minimum, maximum, first and third quartiles of the prediction errors for each observation, respectively.

In spite of their good results in the interpolation within the calibration dataset, the performance of both techniques was very poor in the (partial) extrapolation to the cells of the validation dataset. In general, both techniques consistently overpredict species richness values in poorer places, and underpredict them in the richest ones (Figure 3b). Although kriging predictions show higher correlation to the observed values than those of GLM, the actual Predictive Power of both techniques is negligible, with validation errors averaging around 100% of the observed richness value and reaching almost three times such difference in the extreme cases (Table 1).

## Discussion

In spite of their high predictive power and correlation with observed richness values within the training dataset, our results clearly show that both GLM and kriging provide surprisingly poor results when they are transferred to localities outside the training dataset. This is in spite that the calibration dataset includes almost 30% of the Iberian

UTM 50 × 50 km grid cells, covering more than half of the environmental and spatial variability of the studied territory. And, strikingly, that the validation dataset represents only a partial extrapolation, because it is located within the same geographic domain and covers only slightly different environmental conditions than those of the cells used to build the model. Also importantly, although the leave-one-out jackknife validation is thought to be unbiased (e.g., Olden & Jackson 2000), its disparity with the independent validation evidences that it does not allow to determine the accuracy of projecting model results to slightly different environmental domains.

Our results are similar to those of Parmentier *et al.* (2011), who found that kriging predictions of rainforest tree diversity were as good as regression-based ones—among other techniques—, but also that the performance of all techniques was good in well-sampled areas and bad in poorly inventoried territories. Kriging and other related techniques such as co-kriging can be used for interpolating spatially autocorrelated data (Legendre & Legendre 1998),

but their success in extrapolating the dependent variable to new geographical domains will depend on how consistent is its spatial structure across the studied region; in our case, on how stationary are the spatial trends in species richness. Arguably, other methods based on modelling the relationship between species richness and environmental predictors, such as GLM, may be better for extrapolation. However, they will provide reliable predictions only if they identify causal relationships that—importantly—are also stationary; in other words, that the predictors affect the dependent variable in the same way that they do in the territory where the model is developed (Lobo & Martín-Piera 2002). However, species richness is the outcome of a number of processes that affect different groups of species differently, depending on their response to environmental gradients—which is mediated by evolutionarily-constrained characteristics and recent adaptations—, together with interactions between species, historical events and environmental changes. This complexity results in non-stationary relationships between species richness and the environment even when the causal relationships have been correctly identified, as with, e.g., European scarab dung beetles and temperature (Hortal *et al.* 2011).

Here we argue that the non-stationary nature of the geographic variations in species richness—and of its relationships with environmental gradients—may prevent reliable estimation of their geographic patterns, unless the sample used provides a complete spatial and environmental coverage of the studied region. Further, the limit between interpolations and extrapolations for ALMs may be closer to the geographic domain defined by the training dataset than is commonly assumed, so many species richness models may have virtually no extrapolation ability. Previous studies have attributed the poor performance of ALMs to predict geographic patterns to the geographical biases and limitations in the data (Hortal *et al.* 2007). However, the bad performance of kriging—that does not need environmental predictors—and the geographic proximity of the cells of calibration and validation datasets (see Figure 1b) suggest that ALMs can fail to provide reliable predictions even when based on relatively good data.

Two different research lines are needed to improve our ability to map diversity gradients in incompletely-known territories: (i) evaluate whether more complex techniques accounting for either non-stationarity—either regression-based (*e.g.* Hortal *et al.* 2011) or geostatistical methods (*e.g.* Fuentes 2001; Hernández-Stefanoni *et al.* 2011)—or both environmental responses and spatial structure altogether (Algar *et al.* 2009) provide better predictions of species richness patterns; and (ii) investigate the processes causing such non-stationarity in detail. It could be argued that the effects of ecological assembly rules may, at least in part, account for the geographical non-stationarity in species richness gradients (Guisan & Rahbek 2011), so an investigation on the relationship between the errors in the extrapolation of species richness and differences in the—phylogenetic, functional or ecological—structure of the assemblages may provide insights on why it is so difficult to predict the geographical patterns of diversity.

## Acknowledgements

## References

Algar AC *et al.*, 2009. Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography,* 32:22-33. http://dx.doi.org/10.1111/j.1600-0587.2009.05832.x

Aranda SC & Lobo JM, 2011. How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. *Ecography,* 34:31-38. http://dx.doi.org/10.1111/j.1600-0587.2010.06134.x

Austin MP *et al.*, 1996. Patterns of tree species richness in relation to environment in southeastern New South Wales, Australia. *Australian Journal of Ecology,* 21:154-164. http://dx.doi.org/10.1111/j.1442-9993.1996.tb00596.x

Clark Labs, 2006. *Idrisi Andes version 15.00.* GIS software package. 14.02th ed. Worcester: Clark Labs, Clark University.

Cressie NA, 1993. *Statistics for Spatial Data*. ed. rev. New York: Wiley.

Ferrier S, 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology,* 51:331-363. http://dx.doi.org/10.1080/10635150252899806

Ferrier S & Guisan A, 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology,* 43:393-404. http://dx.doi.org/10.1080/10635150252899806

Fuentes M, 2001. A high frequency kriging approach for non-stationary environmental processes. *Environmetrics,* 12:469-483. http://dx.doi.org/10.1002/env.473

Guisan A & Zimmermann NE, 2000. Predictive habitat distribution models in ecology. *Ecological Modelling,* 135:147-186. http://dx.doi.org/10.1016/S0304-3800(00)00354-9

Guisan A & Rahbek C, 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, 38(8):1433-1444. http://dx.doi.org/10.1111/j.1365-2699.2011.02550.x

Hernández-Stefanoni JL *et al.*, 2011. Combining geostatistical models and remotely sensed data to improve tropical tree richness mapping. *Ecological Indicators,* 11:1046-1056. http://dx.doi.org/10.1016/j.ecolind.2010.11.003

Hortal J & Lobo JM, 2005. An ED-based protocol for optimal sampling of biodiversity. *Biodiversity and Conservation,* 14:2913-2947. http://dx.doi.org/10.1007/s10531-004-0224-z

Hortal J & Lobo JM, 2006. Towards a synecological framework for systematic conservation planning. *Biodiversity Informatics,* 3:16-45.

Hortal J *et al.*, 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). *Conservation Biology,* 21:853-863. http://dx.doi.org/10.1111/j.1523-1739.2007.00686.x

Hortal J *et al.*, 2011. Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. *Ecology Letters*, 14(8):741-748. http://dx.doi.org/10.1111/j.1461-0248.2011.01634.x

Legendre P & Legendre L, 1998. *Numerical Ecology*. 2th ed. Amsterdam: Elsevier.

Lobo JM, 2008. Database records as a surrogate for sampling effort provide higher species richness estimations. *Biodiversity and Conservation,* 17:873-881. http://dx.doi.org/10.1007/s10531-008-9333-4

Lobo JM & Martín-Piera F, 1991. La creación de un banco de datos zoológico sobre los Scarabaeidae (Coleoptera: Scarabaeoidea) íbero-baleares: una experiencia piloto. *Elytron,* 5:31-38.

Lobo JM & Martín-Piera F, 2002. Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables. *Conservation Biology,* 16:158-173. http://dx.doi.org/10.1046/j.1523-1739.2002.00211.x

Lobo JM *et al.*, 2001. Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. *Biological Journal of the Linnean Society,* 73:233-253. http://dx.doi.org/10.1111/j.1095-8312.2001.tb01360.x

Lobo JM *et al.*, 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography,* 33:103-114. http://dx.doi.org/10.1111/j.1600-0587.2009.06039.x

Lomolino MV, 2004. Conservation Biogeography. In: Lomolino MV & Heaney LR (eds.). *Frontiers of Biogeography*: new directions in the geography of nature. Sunderland: Sinauer Associates, Inc. p. 293-296.

Olden JD & Jackson DA, 2000. Torturing data for the sake of generality: How valid are our regression models? *Écoscience,* 7:501-510.

Parmentier I *et al.*, 2011. Predicting alpha diversity of African rain forests: models based on climate and satellite-derived data do not perform better than a purely spatial model. *Journal of Biogeography,* 38:1164-1176. http://dx.doi.org/10.1111/j.1365-2699.2010.02467.x

Pineda E & Lobo JM, 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology,* 78:182-190. http://dx.doi.org/10.1111/j.1365-2656.2008.01471.x

Rocchini D *et al.*, 2011. Uncertainty in species distribution mapping and the need for maps of ignorance. *Progress in Physical Geography,* 35:211-226. http://dx.doi.org/10.1177/0309133311399491

Rossi JP *et al.*, 1995. Statistical tool for soil biology X. Geostatistical analysis. *European Journal of Soil Biology,* 31:173-181.

Ter Steege H *et al.*, 2003. A spatial model of tree α-diversity and tree density for the Amazon. *Biodiversity and Conservation,* 12:2255-2277. http://dx.doi.org/10.1023/A:1024593414624