ECOGRAPHY 27: 68–82, 2004

Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns

Joaquín Hortal, Patrícia Garcia-Pereira and Enrique García-Barros

Hortal, J., Garcia-Pereira, P. and García-Barros, E. 2004. Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns. – Ecography 27: 68–82.

A three-step protocol described elsewhere is used to obtain a map of butterfly species density in Portugal on a 50 × 50 km grid. First, all available faunistic information was compiled and analysed to explore the historic patterns of butterfly sampling in Portugal, and to determine which grid cells are sufficiently prospected to produce reliable estimates of species richness. Then, we relate the estimated species richness scores from these areas to a set of environmental and spatial variables by means of General Linear Models, obtaining a function to extrapolate of species density scores to the rest of Portugal. Finally, the model is validated, results explored and outliers identified and deleted. Any spatial autocorrelation remaining in the residuals is examined. Lastly, model parameters are recalculated in absence of deleted outliers, and the resulting function is used to predict species richness scores throughout mainland Portugal. A highly-predictive function based on some variables previously related to butterfly composition at macro-scale, such as number of sunny days per year, temperature or environmental heterogeneity, was obtained. However, in Portugal those variables are highly spatially structured along a steep latitudinal gradient, leading to difficulty in ascertaining if the latitudinal gradient detected by our analysis is due to macroecological or historic effects. Information on European and Iberian butterfly assemblages and causal processes are discussed in the light of the patterns observed. Then, previous information obtained on Portuguese scarabs is added to identify conservation areas, biogeographically important for both insect groups. Finally, the main drawbacks and advantages of this approach to mapping biodiversity for conservation are discussed briefly.

J. Hortal (mcnjh521@mncn.csic.es), Dept Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, C/José Gutiérrez Abascal 2, E-28006 Madrid, Spain. – P. Garcia-Pereira and E. García-Barros, Dept Biología (Zoología), Fac. de Ciencias, Univ. Autónoma de Madrid, E-28049 Madrid, Spain.

The design of effective conservation policies requires a detailed knowledge of the spatial distribution of organisms (Miller 1994, Dennis and Williams 1995). Computing tools and information-storage systems, facilitating 1) the selection of biodiversity hotspots within national territories, and 2) the identification of sets of areal units that would maximise the number and diversity of effectively protected species (e.g. Araújo 1999), can help in the design of protected-area networks effective for all species within the territory considered.

However, even after >250 yr of accumulating distribution and taxonomic data, there is no complete inventory of all organisms inhabiting any single locality. The spatial distribution of the majority of the species, is yet unknown, as is their total number (Purvis and Hector 2000). After pioneering experiences in countries such as the United Kingdom (Lawton et al. 1994, Griffiths et al. 1999), obtaining and compiling of distributional information on large databases has received strong international support in the form of the

Accepted 2 July 2003

ISSN 0906-7590

Global Biodiversity Information Facility (GBIF; see <http://www.gbif.org> and Edwards et al. 2000). Although the rate of creation of such databases is accelerating, the information gathered is still scarce and in the Mediterranean region is biased to certain countries and regions (Dennis and Williams 1995, Ramos et al. 2001).

Fortunately, computational power provided by personal computers, statistical packages and available high-quality environmental GIS information (Johnston 1998) facilitate analyses of distributional information. Spatial distribution models for biodiversity could help compensate, quickly and economically, for our present lack of knowledge (Scott 1998, Lobo 2000). Models of biodiversity attributes (such as species richness or rarity), obtained from data gathered from well-inventoried areas, can predict scores for sites where faunistic or floristic knowledge is poor (Hortal and Lobo 2002).

Hortal et al. (2001) used data from the Iberian species of a dung beetle family (Coleoptera, Scarabaeinae) to identify the squares of the UTM 50×50 km grid with reliable inventories within mainland Portugal. Next, they used several environmental and spatial variables (latitude and longitude), to obtain a predictive model for species richness per grid square (i.e. species density) and extrapolated the scores for the rest of Portugal. In this study, we present an improved version of this procedure, following the standard protocol defined elsewhere (Hortal and Lobo 2002). This protocol is applied here to model butterfly (Lepidoptera, Papilionoidea and Hesperiidae) species richness across mainland Portugal.

More precisely, our aims are to: 1) determine the spatial distribution, based on current knowledge, of Portuguese butterflies, and identify areas that need further inventory effort, 2) produce a map of predicted species richness for Portuguese mainland butterflies.

Subsequently, the results for Portuguese Scarabaeinae (Hortal et al. 2001), based on the same procedures, geographic units and location, are compared with those derived for butterflies, in order to: 3) recommend improvements in Portuguese regional conservation policies based on species richness.

Finally, potential applications are discussed for maps of predicted biodiversity, combined with accurate biogeographic and ecological knowledge, to large-scale conservation planning.

Methods

The methods used to model the distribution of butterfly species richness follow the protocol discussed in Hortal and Lobo (2002). Briefly, they describe a consecutive three-stage process: a) Compilation and analysis of information, where: 1) available environmental and biological information is stored, 2) working territorial units are defined, 3) previously compiled information is allocated to these units, 4) units with reliable inventories are identified, and 5) biodiversity attribute (in this case, species richness) scores are obtained from the latter areas by studying the relationship between sampling effort and prediction success. b) Model development, where environmental and spatial variables for each well-inventoried territorial unit are related to species richness scores through a stepwise regression procedure to obtain a richness-score prediction function for the poorly-inventoried areas. c) Model validation, where model results are explored to identify possible outliers, and the predictive power of the model and the spatial distribution of its errors are identified.

Examples of the use of this protocol to predict the spatial distribution of Scarabaeinae dung beetles in different west Mediterranean areas can be found in Hortal and Lobo (2001, 2002), Hortal et al. (2001), Lobo and Martín-Piera (2002) and Lobo et al. (2002).

Compilation and analysis of information

Sources of data

The 51 50×50 UTM grid squares within the Portuguese mainland territory with a land surface of $>85\%$ were selected as working Territorial Units (hereafter, TUs) (see Hortal et al. 2001; Fig. 1).

Distribution data about Portugese butterflies (records of the presence of butterfly species) are mostly from Garcia-Pereira (2003). This information was complemented with similar data from ATLAMAR (García-Barros et al. unpubl.), a database on Iberian butterflies, for squares along the Portuguese/Spanish border (see Fig. 1). The data represent an exhaustive compilation on butterfly faunistics across the study area, including both literature records and unpublished information from public and private collections. Information was extracted

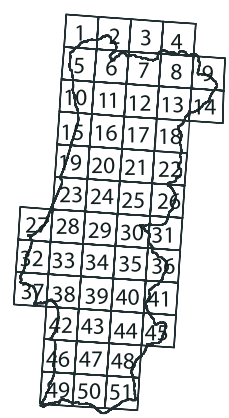

Fig. 1. Map showing the 51 TUs (50 km width UTM squares) in which we divided the Portuguese territory for this analysis (see Hortal et al. 2001). TU numbers as in Table 1.

for all individuals recorded in each TU. The data about dung beetles referred to by Hortal et al. (2001) and used for comparative purposes, came from BANDASCA, a distributional database on the Iberian species of Scarabaeine (see structure in Lobo and Martín-Piera 1991).

For each TU, predictors were taken to be the scores of 25 environmental and spatial variables (Table 2; see Hortal et al. 2001 for details): eight climate variables (minimum, maximum and mean annual temperature, annual range of temperatures, number of days of sun per year, mean annual rainfall, summer rainfall and annual range of monthly rainfall); four topographic variables (minimum, maximum and mean altitude, and altitude range); three bedrock geology variables (amount of surface covered by acid, limestone and clay-mineral bedrocks); four land-use variables (amount of surface covered by grassland, scrub, forest, and agricultural or urban area); two environmental diversity variables (bedrock-geology diversity and land-use diversity); two geographic variables (distance to Pyrenees and to each TU land surface); and two spatial variables (latitude and longitude of each TU centroid).

All variables were standardized to 0 means and unit variances to avoid the effect of different measurement scales, except for the case of latitude and longitude, which were standardized to the mean as recommended by Legendre and Legendre (1998).

Selection of Territorial Units with reliable information and estimation of local species richness

We used collector's curves to identify those units with inventories complete enough so as to produce reliable richness scores. These curves reflect the sampling effort carried out in a given area, related to the rate at which new species are added to the inventory (Soberón and Llorente 1993, Colwell and Coddington 1994, León-Cortés et al. 1998, Moreno and Halffter 2000, Gotelli and Colwell 2001). Our butterfly distribution data comes from heterogeneous sources (standardized samplings, non-professional collectors, etc.), not amenable to a single sampling effort unit; hence, it requires a surrogate measure. We have taken the number of individuals recorded in each TU as a surrogate of sampling effort, because: 1) the number of individuals collected is directly related to the sampling effort carried out (i.e. the greater the effort, the more individuals captured), although this relationship is not necessarily linear, and 2) it can be applied to all records in our database.

The slope of the collector's curve determines the rate of species accumulation at a given level of sampling effort. This slope diminishes as more sampling effort is developed and new species are found, reaching 0 when all species have been found. To determine the quality of the inventory in each TU, we fitted the relationship between the number of species inventoried and the number of individuals collected. As the shape of this relationship depends on the order in which the individuals were recorded, this order was randomized 100 times to obtain a smoothed "ideal" accumulation curve (EstimateS 6.0 software package; Colwell 2000). We compared the fit of the negative exponential function and the Clench function to the smoothed data of all TUs (Soberón and Llorente 1993, Colwell and Coddington 1994, Moreno and Halffter 2000). The Clench function, which produced better results, was selected to fit the curves of the 51 TUs:

$$S_{(ef)} = (a \times ef)/(1 + (b \times ef))$$

where $S_{(ef)}$ is the number of the species found per sampling effort unit (ef); a and b, the parameters of the function, were adjusted to the data of each curve by means of a Simplex and Quasi-Newton method (Anon. 1999). The slope of this curve at each point is defined as:

$$r_{(ef)} = a/(1 + (b \times ef)^2)$$

where $r_{(ef)}$ is the slope of the curve for ef sampling effort. When ef is the number of individuals collected up to date in a given TU, the slope score is the present rate of addition of new species to its inventory. A slope of 0.05 (that is, at present, collecting 20 individuals more would result in 1 new species added) was selected as the cut-off point below which an inventory should be considered reliable enough that its collector's curve can be used to estimate the total number of species in the Territorial Unit. Although a significant number of new species are expected to be added to the inventory of TUs with such a slope, after this point the estimates of total richness obtained with the species accumulation curves as new sampling effort is carried out become relatively stable and independent from the new effort invested (Hortal and Lobo unpubl.). Thus, we assume that, with slopes lower than this, our function would adequately describe the real, as well as the observed, relationship between sampling effort and species inventorying.

Although Hortal et al. (2001) used the number of species recorded in each well-sampled TU as the dependent variable, this score does not necessarily represent the total number of species that may constitute its inventory. Thus, the asymptote of the species accumulation curve was used as a local estimate of the species richness in each well-sampled TU. That is, the point where the slope of the curve reaches 0 (asymptotic species richness; $S_{TU} = a/b$; Soberón and Llorente 1993), which corresponds to the total number of species that would be recorded in the TU with a hypothetical infinite sampling effort. Hereafter, the dependent variable S_{TU} is defined as the butterfly species richness scores predicted by the collector's curve asymptote in each TU previously identified as well-sampled. To assess if modelling point estimates performs better that observed richness scores, we have also modelled the latter (S_{obs}), and compared

model building and model results from both dependent variables.

Model development

Our aim is to obtain a function, based on the environmental and spatial variables, able to predict accurately richness scores throughout Portugal. To find such a function, we used the regression techniques of General Linear Modelling (GLM; McCullagh and Nelder 1989, Dobson 1999; for applications in ecology see Nicholls 1989, 1991, and Crawley 1993). We followed three steps (Guisan and Zimmermann 2000, Hortal and Lobo 2002).

1) Model formulation. We assumed a Poisson distribution for species richness, and a logarithmic relationship between the dependent variable and the predictor (independent) variables, as recommended for biodiversity-related variables (Crawley 1993).

2) Selection of predictor variables. To account for environmental factors effecting species richness, we use the available environmental variables that could be related to species richness on our working scale (2500 km^2). We also included latitude and longitude (i.e., the spatial location of the TUs) as predictors in the model fitting procedure, to include effects due either to historic events or uncodified variables, as they may cause patterns in the spatial distribution of species richness that differ from the purely environmental. Collinearity among predictors is not a concern, to the extent that our aims were primarily to predict a map of the region represented by our data (by maximizing explained variance without inferring causal relationships from the model; Legendre and Legendre 1998, Hortal et al. 2001). Even though environmental variables are likely to be spatially autocorrelated (which invalidates the assumption of their independence), their deletion would lead to ignoring the effect of the processes responsible for the spatial distribution of the variable (Legendre 1993). They were not deleted, but the residuals from the final model were examined to see if they were spatially autocorrelated (see Model validation section).

3) Model fitting. We have used the procedure based on iterative stepwise GLM (Anon. 1999) described elsewhere (Hortal et al. 2001, Lobo and Martín-Piera 2002). The change in Deviance from that of a null model (i.e. the percentage of variation in S_{TU} explained) was used to evaluate the fit of the models, its significance being determined with a classic F test (McCullagh and Nelder 1989, Dobson 1999). This process accounts for: a) possible non-linear relationships between species richness and the environmental predictors, by including linear, quadratic and cubic functions (Austin et al. 1996); b) the effect of synergic interactions between pairs of variables on species richness, including interaction terms (multiplication terms) between all pairs of explanatory variables (Margules et al. 1987); and c) the spatial structure of S_{TU} unexplained by the environmental predictors, including the nine terms of the third-degree polynomial of latitude and longitude (Trend Surface Analysis, TSA; Legendre 1993, Legendre and Legendre 1998).

Model validation

When our purpose is to obtain accurate predictions, increasing the explained variation in species richness is a secondary task. The goodness-of-fit of the model only reflects its relationship with the data used in the process of the model construction, but not its validity and predictive usefulness in the rest of the territory studied. Thus, its strengths and weaknesses should be evaluated to determine if and how it can be improved, and if it is useful to better understand the broad patterns of butterfly species richness in Portugal (Hortal and Lobo 2002).

Residual analysis

We explored model residuals to identify potential outliers (TUs with raw residual scores higher than the standard deviation of predicted values), and Territorial Units with high scores of Potential Leverage (PLV; a measure of the distance of each observation to the centroid of the multidimensional space defined by the variables included in the model; Nicholls 1989). Each residual value was examined to determine whether they were due to erroneous data, or to the environmental uniqueness of TUs. Whilst the former were eliminated before the model fitting procedure, the latter were included with the rest of the observations in the final parameter estimation process.

To test whether final model residuals were spatially autocorrelated or not, we used Moran's I test (Legendre and Legendre 1998) over regular distance lags of 50 km width, that of TU size. Whenever, as a result of this analysis, any spatial structure could be seen to remain in the residuals, such autocorrelation was taken to indicate the existence of at least one further variable not included in the analysis, with a spatially structured effect on species richness. The inclusion of the third degree polynomial of latitude and longitude is presumed to compensate for this influence, so any detectable structure is likely to be due to processes occurring at a spatial scale smaller than used in our analysis (50×50 km), and beyond prediction.

Model reliability and goodness-of-fit

We estimated the predictive power of the model with a Jackknife test; model parameters were estimated as

many times as the number of well-sampled TUs (n), deleting each TU singly once, and comparing the $n-1$ function result with the asymptotic species richness observed in the deleted TU. Scores obtained this way are prediction errors (E_i), that is, the relative distance between the score predicted for each i observation when it is excluded from the model estimation process (P_i), and the observed value (O_i) (Pascual and Iribarne 1993). Percentage error for the ith case is:

$$E_i = \frac{|O_i - P_i|}{O_i} \times 100$$

We used Mean Prediction Error (EPM; the mean of all E_i error estimations) to measure the real error associated with the model. Its inverse ($EPM^{-1} = 100 - EPM$) can be used to estimate its Predictive Power (Pascual and Iribarne 1993).

Previous works use a similar way to assess model reliability (e.g. Wohlgemuth 1998). In these studies, a part of the observations (e.g., a third or a half) is extracted from model building procedure, becoming the "test dataset", and the performance of the model built from the "training dataset" is tested there. The jackknife method used here presents three advantages to test model reliability outside the observations (Hortal et al. 2001, Hortal and Lobo 2002): 1) the rationale of the method is the same, that is, test how a model built from a part of the full dataset performs in the other part, but here we do it one time per observation, instead of only one, increasing the power of this analysis; 2) possible bias due to the selection of the test dataset is avoided, as all observations are part of it in its own error estimation; and 3) it can be used with small datasets, such as the one presented here (only 23 observations), without losing a high number of degrees of freedom for the training dataset. A better assessment can be obtained testing model fitting against a set of different replicates of training/test datasets (see Wohlgemuth 1998 for an example). However, given the small size of our dataset, extracting a significant part of the observations may prevent it from embracing all species richness variation.

The functions for S_{TU} and S_{obs} obtained after deleting erroneous data and estimating final model parameters were applied to the predictor scores in each Portuguese TU, and the former was used to obtain a predicted map of butterfly species richness distribution in mainland Portugal.

Results

Sampling effort assessment

The number of individuals recorded in each TU is shown in Fig. 2a. As for Scarabaeinae (see Fig. 1 in Hortal et al. 2001), the spatial distribution of sampling effort in Portugal is highly heterogeneous. Some well-studied areas coincide for the two groups. Examples included areas near Lisbon, and the Algarve (southern coastal zone). On the other hand, eastern areas and Central System mountains (Serra da Estrela) were well-sampled for scarabeids, whilst northern mountainous were well sampled for butterflies.

The collector's curves for the 35 TUs with >100 recorded individuals were randomized 100 times, and then adjusted to a Clench equation (Fig. 3). Of these, the slopes of 23 fell below the 0.05 cut-off in the last observation, and were selected as well-sampled (Table 1; Fig. 2b); the scores of S_{TU} and S_{obs} for these TUs were correlated (Pearson $r = 0.93$; $n = 23$; $p < 0.001$). It is remarkable that, in the 23 well sampled TUs, whilst observed species richness scores were positively correlated with the number of recorded individuals ($r = 0.43$; $n = 23$; $p = 0.042$), S_{TU} scores showed no significant relationship ($r = 0.27$; $n = 23$; $p = 0.206$). Faunistic knowledge for butterflies seems to be more complete than it is for scarabeids, for which only 16 TUs could be selected using a similar criterion (Hortal et al. 2001). Obviously, however, some supplementary sampling effort directed mainly towards the Alentejo (the interior southern region) is still required.

Model building and validation

We related asymptotic species richness scores of well-sampled TUs (S_{TU}; see Table 1), with the third-degree polynomials of the environmental variables one-by-one in a backward stepwise analysis (Table 2). In successive steps, the model included the third degree polynomial of annual days of sun ($Dsun + Dsun^2 + Dsun^3$), the third-degree polynomial of minimum temperature ($Tmn + Tmn^2 + Tmn^3$), altitude range (Arn), and the interaction term of maximum temperature and limestone bedrock surface ($Tmx \times Gbs$), with their corresponding deletions (Table 3). Inclusion of the third-degree polynomial of latitude and longitude was not significant, so the final model selected was:

$$S_{TU} = EXP[c + f \times Dsun + g \times Tmn^2 + h \times Arn + i \times (Tmx \times Gbs)]$$

where S_{TU} is estimated butterfly species richness in each TU, c is the intercept, $f - i$ are model parameters for each term, and the rest of abbreviations are given in Table 2.

Model residuals Moran's I scores were not significant over any one of the seven distance classes. That is, they were not autocorrelated and, thus, there was no missing large-scale spatial structure due to a non-considered variable. After the Jackknife procedure, model Mean Prediction Error was 6.64% (± 5.18).

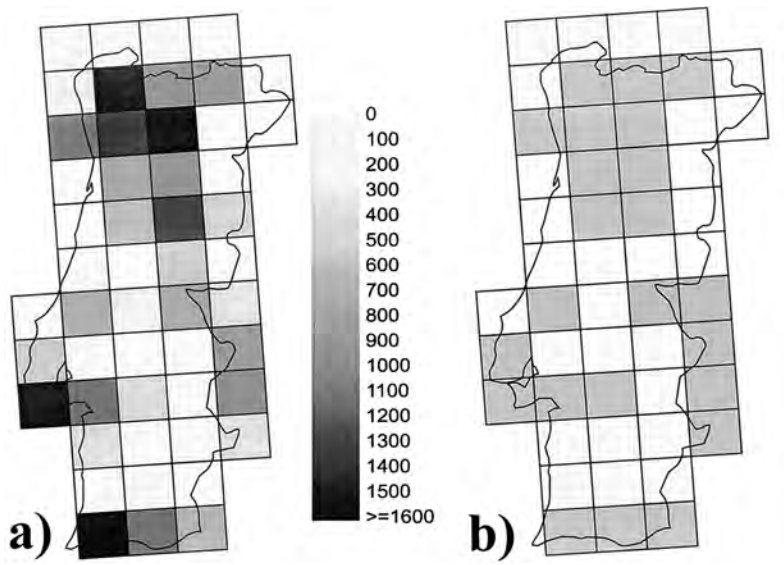

Fig. 2. a) Number of butterfly individuals recorded in each Portuguese TU (data from Garcia-Pereira 2003, and ATLAMAR; see text). b) TUs selected as well-sampled enough (see Table 1).

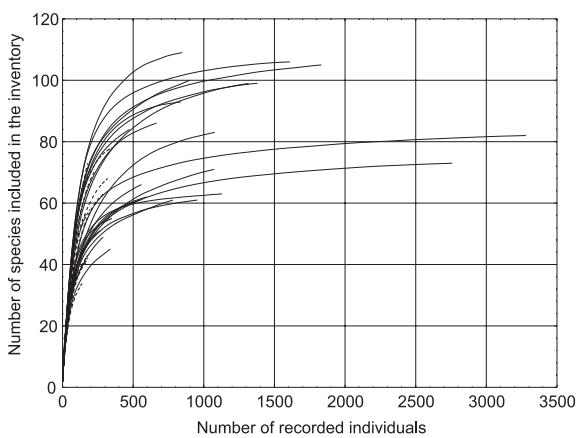

Fig. 3. Collector's curves from the 35 TUs with >100 recorded individuals (see Table 1). The order by which individuals are entered in the curve has been randomized 100 times. The 23 TUs selected as well-inventoried enough are shown in unbroken lines, and the rest in broken lines.

However, after an exploration of model residuals, six outliers were identified (Fig. 4). Of them, TU number 51 PLV was high. As it constitutes an environmentally unique TU, it was kept for model parameter estimation. Of the rest of the outliers, the ones with residuals higher than twice the standard deviation (TUs 20 and 39), were deleted.

After eliminating these two observations, we estimated model parameters again:

$$S_{TU} = EXP[4.358 - 0.163 \times Dsun + 0.062 \times Tmn^2 + 0.069 \times Arn + 0.055 \times (Tmx \times Gbs)]$$

All of them were significant at a 95% confidence level. Again, neither the model residuals nor the prediction errors from the Jackknife analysis displayed evidence of spatial autocorrelation. Both were also normally distributed. MPE decreased to 5.48% ($\pm$4.78), which implies a remarkably high Predictive Power (94.52%; ranging from 89.74 to 99.30%).

Figure 5a shows the extrapolation of this function to the 51 Portuguese Territorial Units. The north of the country appears as the area with higher species richness per TU. From there, S_{TU} diminishes southwards, and, less strikingly, from the centre to both sides, reaching minimum scores in the inner Guadiana basin, in the southeast. Finally, a slightly richer row is located on the southern coast, from the Tejo to Guadiana river mouths.

We followed the same steps to develop a model to predict S_{obs}, which resulted in a less explanatory function (see Table 4). Neither the inclusion of any interaction term, nor of the third-degree polynomial of latitude and longitude, were significant, so the final model selected was:

$$S_{obs} = EXP[q + j \times Dsun + k \times Gbas + l \times Gbas^2 + m \times Ush^2 + n \times Tmn^2]$$

where S_{obs} is observed butterfly species richness in each TU, q is the intercept, j – n are model parameters for each term, and variable codes follow Table 2. Residuals from this function were negatively autocorrelated in the first two distance lags (0–50 km lag, Moran's $I = -0.249$, $p < 0.05$; 50–100 km lag, $I = -0.323$, $p < 0.01$). That is, residual scores were spatially structured, but high scores in a given TU indicate low ones in the surrounding ones, and vice versa. This negative autocorrelation is only a minor problem, as indicates that,

Table 1. Species accumulation curves of the 35 TUs with >100 recorded individuals (see Fig. 3). TU numbers correspond with those of Fig. 1; name lists their UTM codes; S_{obs} is the number of observed species, and ind. The number of recorded individuals; a and b are the parameters of the Clench function adjusted to each curve; V. exp. is the percentage of Variance explained by each function; r is the slope score in the last recorded individual; S_{TU} is the asymptote value predicted by the Clench equation, and %S is the percentage of the predicted species that has been recorded already in that TU. The 23 TUs considered as well-sampled enough ($r<0.05$) are marked.

TU	Name	S_{obs}	Ind	a	b	V. exp.	r	S_{TU}	% S
1	29TNG1	43	178	0.860	0.0145	99.87	0.112	59.1	72.7
5	29TNG2	71	233	1.006	0.0098	99.99	0.161	102.2	69.5
6	**29TNG4**	**105**	**1833**	**1.080**	**0.0098**	**99.97**	**0.003**	**110.2**	**95.3**
7	**29TPG2**	**100**	**901**	**1.111**	**0.0101**	**99.88**	**0.013**	**109.5**	**91.4**
8	**29TPG4**	**109**	**848**	**1.114**	**0.0089**	**99.88**	**0.019**	**125.1**	**87.1**
10	**29TNF1**	**83**	**1077**	**0.639**	**0.0068**	**99.74**	**0.012**	**94.0**	**88.3**
11	**29TNF3**	**99**	**1380**	**1.072**	**0.0101**	**99.95**	**0.005**	**105.6**	**93.8**
12	**29TPF1**	**106**	**1609**	**1.268**	**0.0113**	**99.81**	**0.004**	**112.4**	**94.3**
13	29TPF3	73	178	0.976	0.0079	99.98	0.329	123.9	58.9
16	**29TNF4**	**86**	**665**	**1.041**	**0.0106**	**99.99**	**0.021**	**98.6**	**87.2**
17	**29TPF2**	**93**	**838**	**1.140**	**0.0109**	**99.89**	**0.013**	**104.6**	**89.0**
18	29TPF4	57	169	1.013	0.0119	99.98	0.200	85.0	67.1
20	**29TNE3**	**84**	**474**	**1.145**	**0.0115**	**99.93**	**0.038**	**99.9**	**84.1**
21	**29TPE1**	**99**	**1317**	**0.838**	**0.0078**	**99.77**	**0.008**	**107.6**	**92.0**
22	29TPE3	68	320	1.056	0.0124	99.99	0.063	85.3	79.7
23	29TNE2	42	180	1.090	0.0206	99.98	0.074	52.9	79.4
25	29TPE2	78	350	1.166	0.0118	99.78	0.065	99.1	78.7
26	29TPE4	40	138	1.066	0.0192	99.99	0.133	55.5	72.1
28	**29SND1**	**66**	**559**	**0.817**	**0.0108**	**99.78**	**0.022**	**75.9**	**86.9**
29	29SND3	58	195	1.103	0.0138	99.96	0.135	80.2	72.3
30	**29SPD1**	**62**	**597**	**0.886**	**0.0127**	**99.93**	**0.015**	**69.9**	**88.7**
31	**29SPD3**	**51**	**268**	**1.094**	**0.0177**	**99.99**	**0.047**	**61.9**	**82.5**
32	**29SMD4**	**55**	**350**	**1.139**	**0.0177**	**99.93**	**0.029**	**64.5**	**85.3**
33	29SND2	41	170	0.990	0.0184	99.98	0.092	53.9	76.1
36	**29SPD4**	**61**	**781**	**0.846**	**0.0130**	**99.39**	**0.008**	**65.2**	**93.6**
37	**29SMC3**	**82**	**3280**	**0.882**	**0.0106**	**98.95**	**0.001**	**82.9**	**98.9**
38	**29SNC1**	**63**	**1129**	**1.140**	**0.0171**	**99.91**	**0.003**	**66.5**	**94.8**
39	**29SNC3**	**45**	**338**	**0.884**	**0.0171**	**99.72**	**0.026**	**51.6**	**87.2**
40	29SPC1	34	138	0.957	0.0214	99.88	0.098	44.6	76.2
41	**29SPC3**	**61**	**955**	**0.968**	**0.0150**	**99.78**	**0.005**	**64.5**	**94.5**
42	29SNC2	63	291	0.960	0.0117	99.97	0.076	81.8	77.0
45	**29SPC4**	**49**	**287**	**0.943**	**0.0162**	**99.71**	**0.042**	**58.1**	**84.3**
49	**29SNB2**	**73**	**2759**	**0.697**	**0.0093**	**98.90**	**0.001**	**74.6**	**97.8**
50	**29SNB4**	**71**	**1074**	**0.693**	**0.0091**	**98.75**	**0.007**	**76.0**	**93.4**
51	**29SPB2**	**61**	**553**	**0.892**	**0.0129**	**99.74**	**0.017**	**69.1**	**88.3**

although richness scores from neighbouring localities are related, there are no clusters of nearby TUs with residuals higher or lower than their surroundings. After exploring model residuals, which presented higher standard deviation than the ones from the S_{TU} model, seven outliers were identified (Fig. 4). As for the former model, TUs 20 and 39, with residual scores higher than double standard deviation were deleted, and the model parameters were estimated again. The model obtained was:

$$S_{obs} = EXP[4.361 - 0.214 \times Dsun + 0.186 \times Gbas - 0.106 \times Gbas^2 - 0.027 \times Ush^2 + 0.072 \times Tmn^2]$$

from which Ush^2 parameter was significant only at a 90% level. Although the outcome of both predictive models was highly correlated in the 23 well-sampled Territorial Units (Pearson $r=0.962$, $n=23$, $p<0.001$), MPE of this model was higher, although this estimate was more uncertain (12.83% $\pm$16.30), with a Predictive Power (87.17%) ranging from 70.81 to 100%. Residuals from both models (using S_{TU} or S_{obs}) were highly correlated ($r=0.708$, $n=23$, $p<0.001$), presenting similar spatial distributions (Fig. 6). However, prediction errors of both models showed different distributions (Wilcoxon matched pairs test; $n=23$, $Z=2.2507$, $p=0.024$), evidencing the existence of different error sources in the two datasets (estimated and observed richness scores). Here, it is advisable that prediction errors of the S_{obs} model showed a negative but significant relationship with raw richness scores (that is, the higher the richness score, the lower the error; Pearson $r=-0.585$, $n=23$, $p=0.003$).

Discussion

Effect of sampling effort on observed species richness

Biodiversity scores such as species richness extracted from atlas data are inaccurate due to bias in sampling effort. They do not contain information about taxonomic and spatial bias, real absences of species, or sampling intensity (Rich 1998). Sampling effort has a considerable impact on both species richness and species

Table 2. Explanatory variables included in the analysis, with their respective codes. Deviance and Change in Deviance were calculated through the comparison with a null model (with no explanatory variables) of butterfly species richness. The linear, quadratic or cubic functions of each variable were selected when they produced a Change in Deviance significant at a 5% level. Dev: Deviance; Ch. Dev: Change in Deviance; * $p < 0.05$.

Variable	Code	Selected terms	DF	Dev.	Ch. Dev.	F
	Null model		22	119.01		
Topographic variables						
Minimum altitude	Amn	Amn	21	92.7	26.31	5.96*
Maximum altitude	Amx	Amx	21	47.6	71.33	31.42*
		$Amx+Amx^2$	20	47.2	71.81	30.42
		$Amx+Amx^2+Amx^3$	19	34.5	12.67	6.97*
Mean altitude	Amd	Amd	21	50.7	68.32	28.30*
Altitude range	Arn	Arn	21	58.3	60.64	21.81*
Climate variables						
Minimum temperature	Tmn	Tmn	21	53.6	65.40	25.62*
		$Tmn+Tmn^2$	20	49.2	69.78	28.35
		$Tmn+Tmn^2+Tmn^3$	19	36.3	82.73	43.31*
Maximum temperature	Tmx	Tmx	21	58.7	60.65	21.82*
Mean annual temperature	Tmd	Tmd	21	28.9	90.13	65.53*
		$Tmd+Tmd^2$	20	28.5	90.53	63.55
		$Tmd+Tmd^2+Tmd^3$	19	24.8	94.25	72.31*
Annual temperature range	Trn	Trn	21	118.8	0.23	0.04
		$Trn+Trn^2$	20	102.0	17.04	3.34
Annual days of sun	Dsun	Dsun	21	33.6	85.46	53.48*
		$Dsun+Dsun^2$	20	33.4	85.57	51.18
		$Dsun+Dsun^2+Dsun^3$	19	22.6	96.38	80.88*
Mean annual precipitation	Pmd	Pm	21	64.3	54.75	17.89
		$Pm+Pm^2$	20	44.0	75.00	34.08*
Summer precipitation	Psm	Psm	21	57.8	61.24	22.26
		$Psm+Psm^2$	20	42.8	76.23	35.64
Annual precipitation range	Prn	Prn	21	68.7	50.27	15.35
		$Prn+Prn^2$	20	55.3	63.69	23.02
Land use variables						
Cultivated and urban area	Ucu	Ucu	21	118.2	0.80	0.14
		$Ucu+Ucu^2$	20	82.9	36.07	8.70*
Forest area	Ufr	Ufr	21	111.2	7.83	1.48
Shrub area	Ush	Ush	21	116.3	2.76	0.50
Grassland area	Ugs	Ugs	21	118.9	0.17	0.03
		$Ugs+Ugs^2$	20	93.7	25.28	5.39
Bedrock geology variables						
Basic bedrock area	Gbs	Gbs	21	109.2	9.83	1.89
Clay-mineral bedrock area	Gcl	Gcl	21	98.9	20.11	4.27
		$Gcl+Gcl^2$	20	90.6	28.40	6.27
		$Gcl+Gcl^2+Gcl^3$	19	77.7	41.27	10.09*
Acid bedrock area	Gac	Gac	21	97.1	21.95	4.75
Environmental diversity variables						
Land use diversity	Dlu	Dlu	21	86.1	32.94	8.04
Bedrock geology diversity	Dbg	Dbg	21	98.6	20.42	4.35
Geographic and spatial variables						
Distance to Pyrenees	Dpr	Dpr	21	118.0	1.02	0.18
		$Dpr+Dpr^2$	20	104.1	14.91	2.86
Land surface	Lsr	Lsr	21	116.6	2.40	0.43
Latitude	Lat	Lat	21	37.3	81.70	45.98*
		$Lat+Lat^2$	20	26.3	92.72	70.51*
		$Lat+Lat^2+Lat^3$	19	18.4	100.64	104.02*
Longitude	Long	Long	21	117.9	1.14	0.20

incidence estimates of different areas (Dennis et al. 1999). Although collector visits are biased toward richest areas, they are also biased toward areas near home (Dennis and Thomas 2000). Thus, it is likely that the bias detected in our sampling effort assessment may be explained by entomologists' habits, or by the distribution of entomologists themselves. Our results show that sampling effort has indeed been directed to areas with higher observed richness. However, scores estimated from species accumulation curves were not correlated to the number of individuals, suggesting that the biases identified above do not explain all the variation in butterfly species richness, a fact that may be true for many of the distribution Atlas available at present in European countries.

As pointed out by our results, even with such biased data, a survey effort assessment, such as the one developed here, can be used to produce predictive models that fill in gaps in knowledge (Dennis and Hardy 1999, Lobo 2000, Hortal et al. 2001, Hortal and Lobo 2002, Lobo and Martín-Piera 2002). Although the model developed from observed richness scores could

Table 3. Summary of the stepwise variable selection procedure used to build asymptotic butterfly species-richness (S_{TU}) models in Portuguese well-sampled TUs. The Change in Deviance after the inclusion of each term in the model was evaluated with an F test with 0.05 confidence level. Variable codes follow Table 2. Marked percentage of explained deviance is the score of the model selected at the end of the process. Dev: Deviance; Ch. Dev: Change in Deviance; % Dev: percentage of total deviance explained.

Model	Dev	DF	Ch. Dev.	F	p	% Dev
Null	119.0162	22				
Step 1						
+Dsun	33.5574	21	85.46	53.48		71.80
+Dsun^2	33.4426	20	0.11	0.07	0.7940	
	33.4426	20	85.57	51.18		71.90
+Dsun^3	22.6399	19	10.80	9.07	0.0072	
	22.6399	19	96.38	80.88		80.98
− Dsun^2	23.9174	20	− 1.28	− 1.07	0.3133	
	23.9174	20	95.10	79.52		79.90
Step 2						
+Tmn+Tmn^2+Tmn^3	12.1203	17	11.80	16.55	0.0008	
	12.1203	17	106.90	149.93		89.82
− Dsun^3	13.8040	18	− 1.68	− 2.20	0.1553	
	13.8040	18	105.21	137.19		88.40
− Tmn^3	15.7390	19	− 1.93	− 2.34	0.1426	
	15.7390	19	103.28	124.67		86.78
− Tmn	16.1546	20	− 0.42	− 0.51	0.4834	
	16.1546	20	102.86	127.35		86.43
Step 3						
+Arn	11.3327	19	4.82	8.08	0.0104	
	11.3327	19	107.68	180.54		90.48
Step 4						
+Tmx × Gbs	7.0212	18	4.31	11.05	0.0038	
	7.0212	18	111.99	287.11		**94.10**

explain more variation, the one produced from asymptotic estimates was more consistent, with less standard deviation, and lower residual scores throughout all Portugal. Function errors were higher than 4 species only in a few localized areas: the Algarve (SW corner), Beira Litoral (western coast) and Trás-os-Montes (NE). Thus, this latter model is likely to be more explanatory and reliable, evidencing that using local estimates of richness before the modelling procedure produces a better picture of biodiversity patterns, or, at least, homogenizes the variable, diminishing the noise present in the dependent variable, in a way that improves how it

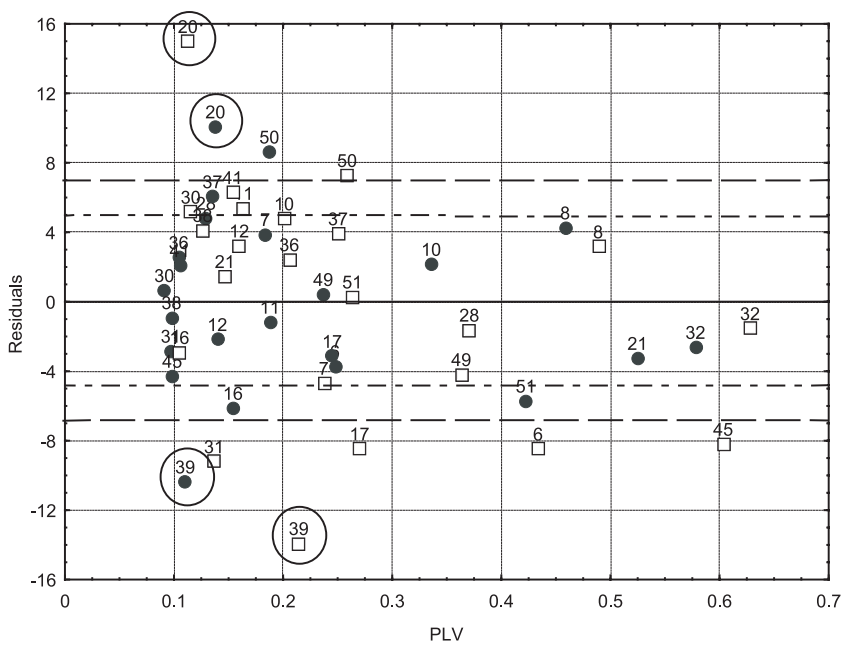

Fig. 4. Residuals and Potential Leverage (PLV) from both models in well-sampled TUs; full dots correspond to S_{TU} scores, and empty squares S_{obs}. Standard deviation of predicted scores is shown as dots and dashes for the former, and dashes for the latter. Residuals with scores higher than double standard deviation (circles) have been deleted in the final estimation of model parameters.

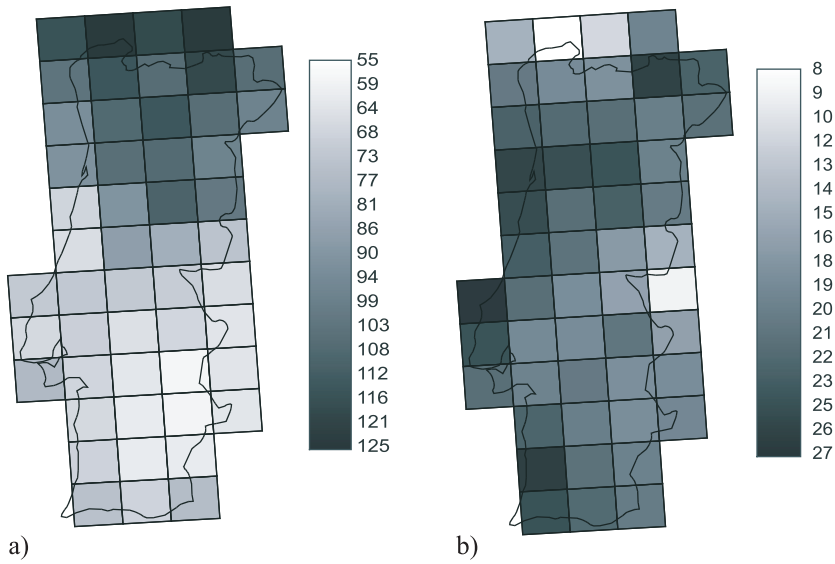

Fig. 5. Species-richness scores for the 51 Portuguese TUs (abbreviations as in Table 2): a) Number of butterfly species estimated by the final model ($S_{TU} = EXP[4.358 - 0.163 \times Dsun + 0.062 \times Tmn^2 + 0.069 \times Arn + 0.055 \times (Tmx \times Gbs)]$; b) Number of Scarabaeinae species estimated by the model $S = EXP[3166 - 0.464 \times Pmd^2 + 0.120 \times Pmd^3 + 0.234 \times Pmd \times Ufr]$ (Hortal et al. 2001).

Table 4. Summary of the stepwise variable selection procedure used to build observed butterfly species-richness (S_{obs}) models in Portuguese well-sampled TUs. Analyses as in Table 3; abbreviations and variable codes as in Tables 2 and 3.

Model	Dev	DF	Ch. Dev.	F	p	% Dev
Null	119.0162	22				
Step 1						
+Dsun	59.2891	21	85.20	30.18		58.97
+Dsun^2	58.9675	20	0.32	0.11	0.1207	
	58.9675	21	85.52	29.00		59.19
+Dsun^3	48.0275	19	10.94	4.33	0.9717	
	48.0275	19	96.46	38.16		66.76
Step 2						
+Gbas+Gbas^2+Gbas^3	25.4968	16	22.53	14.14	0.9997	
	25.4968	16	118.99	74.67		82.35
− Dsun^2	25.7255	17	−0.23	−0.15	0.1381	
	25.7255	17	118.76	78.48		82.19
− Gbas^3	26.7944	18	−1.07	−0.72	0.4997	
	26.7944	18	117.69	79.06		81.46
Step 3						
+Ush^2	18.5338	17	8.26	7.58	0.9955	
	18.5338	17	125.95	115.53		87.17
Step 4						
+Tmn+Tmn^2+Tmn^3	12.6429	14	5.89	6.52	0.9900	
	12.6429	14	131.84	145.99		91.25
− Tmn^3	12.7875	15	−0.14	−0.17	0.1547	
	12.7875	15	131.70	154.48		91.15
− Tmn	13.3006	16	−0.51	−0.62	0.4496	
	13.3006	16	131.18	157.81		90.79

can be fitted with the used predictors. However, to carry out this previous assessment, a measurement of survey evenness and intensity is needed (Dennis et al. 1999). Using the incomplete and sparse sampling information available in the scientific literature, and private and institutional collections, both database-records for Scarabaeinae (Hortal et al. 2001, Lobo and Martín-Piera 2002) and recorded individuals for butterflies (this work), have proved to be useful surrogates of sampling effort. The minimum effort of recording only a little reliable information associated with each capture, such as that found in good entomological labels, makes virtually all data useful in assessing or predicting biodiversity.

Predictive models and causal explanations

Modelling procedures have been used in ecology to: 1) identify causal factors that best explain the dependent

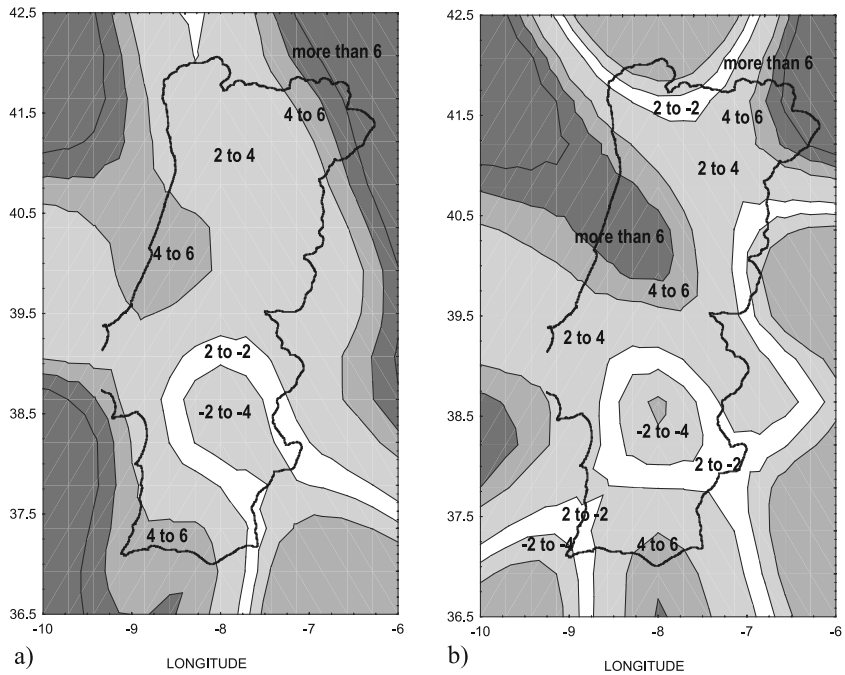

Fig. 6. Geographic distribution of the residuals produced by a) the S_{TU} model, and b) the S_{obs} model. Longitude and Latitude are expressed in decimal Greenwich degrees, and the shadowed areas represent the raw residual scores. The residual score ranges are marked in the figure.

variable (e.g. McCarthy et al. 2001); and 2) find a set of factors to predict changes in the modelled variable, without assuming any causal relationship. Despite the fact that the former task needs additional experimental information to certify causality (Lawton 1999, Mac Nally 2000), these two models are frequently confounded (see Guisan and Zimmermann 2000). The main problems of ascertaining causes from modelling come from two sources: usefulness of the procedure to identify causal relationships, and the existence of important factors not included in the set of predictors used.

Many papers have used predictive models based on regression methodologies to make hypotheses about causal relationships between biodiversity variables, such as species richness, and several environmental and land use variables (e.g. Margules et al. 1987, Austin et al. 1990, 1996, Heikkinen and Neuvonen 1997). However, regression procedures are not able to identify causes, but instead specify covariation between the scores of the variables (Mac Nally 2000, 2002). Thus, using these methods to infer biodiversity responses to the environment may be misleading, as we are estimating only patterns of spatial co-occurrence, and not causation. In this framework, predictive maps may allow us to identify the main spatial patterns of many biodiversity attributes (Levin 1992), but, as no causal relationships are identified, its outcome should not be extrapolated outside of the spatial and numeric range of the explanatory variables used (Legendre and Legendre 1998), nor to the future scenarios that may appear after climate or land-use changes.

Present distribution of biodiversity is conditioned by the responses of the species to the environment, as well as by several unique contingent factors. For many taxa, geographic patterns in the composition of West Palaearctic assemblages have been conditioned by the extreme climatic fluctuations occurring since the Pleistocene (e.g. de Jong 1998, Hewitt 1999). Post-glacial dispersal and redistribution of European butterfly faunas has been related with two of their key characteristics (Dennis et al. 1991, Dennis 1993): 1) a close relationship to floristic composition (due to the host-specificity of most butterfly species to a given plant species), and 2) great sensibility to climatic alterations (due to physiological restrictions). The effect of floral relationships in structuring Portuguese assemblages can not be explored with our analysis, while climatic and other environmental effects are implicitly included. However, environmental variables are not independent, and are also spatially structured. Thus, no direct causal relationships can be demonstrated with the regression methodology used (Austin et al. 1996, Lobo et al. 2001).

Furthermore, the spatial scale used limits the outcome of our analysis to the broad spatial resolution of biogeography (see Blackburn and Gaston 2002 for a discussion). Thus, our results differ from that of Mac Nally et al.'s (2003) meso-scale analysis, which explained butterfly species richness in terms of several topographic variables. These variables probably act as surrogates of

microclimatic variation on small scales, but the effects operating at this scale are unlikely to work at broader ones. Nonetheless, their predictive function was useful only on the mountain range where their data came from, failing for an adjacent region. Differences between faunal assemblages of different areas make the extrapolation of such functions impossible to other, even close, regions.

In a previous ordination analysis, Garcia-Pereira (2003) found a relationship between butterfly species composition and many environmental factors, of which precipitation, altitude and, less strikingly, temperature stand out. The positive relationship found between the number of rainy days and richest areas (composed of European elements, and located in the north of the country), was confirmed by the strong negative relationship (ca 80%) found between richness and the number of sunny days, its inverse measure (see Tables 2 and 3). Especially in relation to southern faunistic replacement, temperature variables also had a significant effect which stands out in both analyses. Other variables had subsidiary effects on the composition of assemblages; for example, increases in species richness with increasing altitude may be related to higher environmental heterogeneity. The association of the Algarve fauna (composed of Mediterranean thermophylous species), on the southern coast, with calcareous bedrock and high temperatures (see Table 3), may be related to the soil specificity of many host plants.

All the three areas that presented high residuals, Beira Litoral, the Algarve and Trás-os-Montes, showed particular features which may be difficult to predict by a general function such as the one developed here. The first constitutes an altitude gradient, from the hillsides of the Serra da Estrela, towards the western coast. The second and third, apart from being the only hotspots shared by butterflies and scarabeids (see below), are inhabited by "singular" butterfly assemblages, which include many endemics, and species rare or inexistent in the surrounding areas (Garcia-Pereira 2003).

Conservation assessment through predicting biogeographic patterns

Given the spatial scale studied, it is not possible to delimit exactly the sites most important for butterfly conservation in Portugal. However, those regions where some of these sites are located can be identified. Once conservation goals are well defined for a specific region, work on a local scale should be done to define the shape, extent and legal regulation of the protected sites (see Bassett and Edwards 2003). Trás-os-Montes and the Algarve are the only hotspots shared by both butterflies and dung beetles (see Fig. 5). These two areas host "singular" butterfly faunas, as well as mid-to-high endemicity scores for scarabeids (Lumaret and Lobo 1996, Verdú and Galante 2002). This is specially so for the south, where the appearance of Ibero-Maghrebian endemic species produces high latitudinal β-diversity scores (Hortal-Muñoz et al. 2000). The Portuguese conservation network should include places from both regions. Similar conclusions were reached independently by means of complementarity analyses on the basis of data from other diverse animal and plant taxa (Araújo 1999). The northernmost assemblages are protected by the Montesinho Natural Park; only a few new sites would be necessary to maximize protection success. In the Algarve, however, many important conservation areas are still unprotected.

Due to the southwards and peripheral impoverishment in Trans-European butterfly richness described, and the absence of Alpine species, the Serra da Estrela, the western part of the Iberian Central System mountain range, constitutes a hotspot only for dung beetles. Using butterflies alone as indicators for conservation would leave out this area, important at least for scarabeids, as Eurosiberian and Mediterranean faunas overlap (Hortal et al. 2001). Again, protected-area networks should include sites from this region.

Using only predicted maps of species richness is not a good strategy to identify networks of areas for conservation, unless they are used in combination with information about the compositional variability in the studied area. Once enough knowledge on this issue is available, the richest TUs in each compositionally different region seem to be a good criterion to place areas for conservation. Thus, it is advisable to include sites from the "shadow zones", poor in species, for both butterflies (Alentejo), and scarabeids (Tejo river basin). The Alentejo, a transition area between Mediterranean and Eurosiberian faunas, where many range margins occur, is home to impoverished Scarabaeinae and Rhopalocera assemblages (Hortal-Muñoz et al. 2000, Garcia-Pereira 2003). However, biogeographic crossroads are important for conservation policies (Spector 2002), and this region has been identified as an endemicity area for diverse, independent groups of plants and animals (Martín et al. 2000, García-Barros et al. 2002). Using species richness as a single criterion in conservation planning would have ignored this area.

Hotspots are frequently found in areas of ecological transition and may host both core and marginal populations (Araújo 2002), the latter which may be subject to important evolutionary processes (Thomas et al. 2001). However, as singular assemblages may often be absent from such areas, they may be undervalued in conservation policies. The Serra da São Mamede, the only mountain range exceeding 1000 m in altitude south of the river Tejo (TU number 35; see Fig. 1), may be a good example of the situation described above, where Eurosiberian and Mediterranean intergradation condi-

tions favour both the sympatric occurrence of sister taxa from both zones and a rich diversity of species (Amphibians and Reptiles; Sá-Sousa 2000); it also hosts a local Scarabaeinae hotspot (Hortal et al. 2001).

Final considerations

Biodiversity assessment from Atlas data

Atlas data alone are far from accurate in covering the spatial distribution of biodiversity measures, due to sampling bias (Dennis et al. 1999, Dennis and Thomas 2000). As new atlases are developed, and more sampling is carried out, the diversity measures obtained vary greatly on different scales and extents, and the new estimates of diversity may be uncorrelated with those derived from old atlases (e.g. butterflies, Dennis 1997, Dennis and Shreeve 2003; birds, Lobo pers. comm.). New sampling designs are needed to obtain reliable information on biodiversity patterns (Dennis and Hardy 1999, Dennis et al. 1999, Dennis and Thomas 2000, Hortal and Lobo unpubl.). A measure of sampling effort or, at least, a description of the surveys, included in the databases used to build these atlases (Dennis et al. 1999, Dennis and Thomas 2000) is necessary for the production of maps of predicted biodiversity components such as species richness, rarity or endemism (Carroll and Pearson 1998a, Pearson and Carroll 1999, Lobo 2000, Hortal et al. 2001, Hortal and Lobo 2002, Lobo and Martín-Piera 2002). With the procedure described here and atlas databases, sound predictions of these surrogates can be obtained for a given set of areas, and then interpolated for the rest of the territory studied.

Which biodiversity surrogates predict, and how?

Unfortunately, a map of a single biodiversity surrogate is not enough to describe all the spatial variation in the assemblages from a given group. To decide where and how to locate protected areas, other biodiversity components (e.g. species turnover, faunal composition, rarity or endemism) must also be estimated, and used jointly with biogeographic and ecological information (see Lobo et al. 2001, 2002 for some examples). The scale of the work presented here is too large to design concrete boundaries for the new protected sites that may be integrated in the conservation network. However, the main patterns of variation may be obscured by small-scale processes at smaller spatial resolutions. Thus, the goodness-of-fit and utility of predicted maps increase on larger scales (Carroll and Pearson 1998b, Pearson and Carroll 1999). In this framework, regional conservation goals should be stated on broad scales. Then, site networks should be designed in each important area on a smaller scale (see above), taking into account local parameters operating at a landscape level, such as land use, land owners, productive and urban areas, local impacts, etc. This flexible approach connects broad and landscape scale conservation planning in a reasonable way: interesting areas are identified at a coarse-grain scale, and reserves are designed in situ.

The resolution used here (50 km grid cells) seems useful for conservation planning. Many Atlas data can be referred to this scale, or to a similar one, as the units of the 0.5° geographic grid have approximately the same surface area throughout the world, except at extreme north and south latitudes. Moreover, a large amount of world-wide GIS environmental information is easily available at this spatial resolution, or even at smaller ones (García Hernández and Bosque Sendra 2001; see also the United Nations Environmental Program geographic information database; < http://www.grid.unep.ch/ >).

Utility of indicator taxa

Carroll and Pearson (1998a, Pearson and Carroll 1999) successfully used a mixed procedure, joining together an indicator taxon and geostatistical analyses to predict species richness of another group. However, species richness of different taxa are not necessarily correlated (as seen in data about butterflies and dung beetles presented here). This study shows that when the coverage of spatio-temporal variation is complete enough, (see Hortal and Lobo unpubl.), additional predictor taxa may not be necessary. Moreover, current results indicate that geostatistics alone can also help interpolate such maps when good geographic coverage of well-sampled TUs is available (Lobo and Hortal unpubl.).

Thus, the use of any single group as an indicator for all biodiversity can be misleading; it is necessary to use as many groups as possible. As stated above, good Atlas data are available only for a few groups, mainly plants and vertebrates, although using raw scores from them introduces inaccuracies. These maps can be obtained rapidly and accurately with good databases, and the spatial prediction of biodiversity surrogates. With the information so-obtained, present-day arbitrary or inaccurate-data-based area selections can be improved from a by far wider spectrum of all biodiversity.

Acknowledgements – We are indebted to M. Munguira and J. Martín Cano for providing us with some Spanish data, and to J. Cerne for the English review. The comments of Jorge Lobo and Miguel B. Araújo improved substantially previous drafts of the manuscript. This paper was supported by the Spanish D.G.I. projects REN2001-1136/GLO and REN2000-0466/GLO (ATLAMAR). J.H. was supported by a PhD MNCN/C.S.I.C./CM grant, and P.G.P. by a PhD grant from the Praxis XXI program (BD 9101/96; Ministério da Ciência, Portugal).

References

Anon. 1999. STATISTICA for Windows, computer program manual. – StatSoft, Tulsa, OK.

Araújo, M. B. 1999. Distribution patterns of biodiversity and the design of a representative reserve network in Portugal. – Div. Distrib. 5: 151–163.
Araújo, M. B. 2002. Biodiversity hotspots and zones of ecological transition. – Conserv. Biol. 16: 1662–1663.
Austin, M. P., Nicholls, A. O. and Margules, C. R. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. – Ecol. Monogr. 60: 161–177.
Austin, M. P., Pausas, G. J. and Nicholls, A. O. 1996. Patterns of tree species richness in relation to environment in south-eastern New South Wales, Australia. – Aust. J. Ecol. 21: 154–164.
Bassett, S. D. and Edwards Jr, T. C. 2003. Effects of different sampling schemes on the spatial placement of conservation reserves in Utah, USA. – Biol. Conserv. 113: 141–151.
Blackburn, T. M. and Gaston, K. J. 2002. Scale in macroecology. – Global Ecol. Biogeogr. 11: 185–189.
Carroll, S. S. and Pearson, D. L. 1998a. Spatial modeling of butterfly species diversity using tiger beetles as a bioindicator taxon. – Ecol. Appl. 8: 531–543.
Carroll, S. S. and Pearson, D. L. 1998b. The effects of scale and sample size on the accuracy of spatial predictions of tiger beetle (Cicindelidae) species richness. – Ecography 21: 401–414.
Colwell, R. K. 2000. EstimateS, statistical estimation of species richness and shared species from samples. Ver. 6. – User's guide and computer program available at <http://viceroy.eeb.uconn.edu/estimates>.
Colwell, R. K. and Coddington, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. – Philos. Trans. R. Soc. Lond. B 345: 101–118.
Crawley, M. J. 1993. GLIM for ecologists. – Blackwell.
de Jong, H. 1998. In search of historical biogeographic patterns in the western Mediterranean fauna. – Biol. J. Linn. Soc. 65: 99–164.
Dennis, R. L. H. 1993. Butterflies and climate change. – Univ. Press, Manchester.
Dennis, R. L. H. 1997. An inflated conservation load for European butterflies: increases in rarity and endemism accompany increases in species richness. – J. Insect Conserv. 1: 43–62.
Dennis, R. L. H. and Williams, W. R. 1995. Implications of biogeographical structures for the conservation of European butterflies. – In: Pullin, A. S. (ed.), Ecology and conservation of butterflies. Chapman and Hall, pp. 213–229.
Dennis, R. L. H. and Hardy, P. B. 1999. Targeting squares for survey: predicting species richness and incidence of species for a butterfly atlas. – Global Ecol. Biogeogr. 8: 443–454.
Dennis, R. L. H. and Thomas, C. D. 2000. Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. – J. Insect Conserv. 4: 73–77.
Dennis, R. L. H. and Shreeve, T. G. 2003. Gains and losses of French butterflies: test of predictions under-recording and regional extinction from data in a new atlas. – Biol. Conserv. 110: 131–139.
Dennis, R. L. H., Williams, W. R. and Shreeve, T. G. 1991. A multivariate approach to the determination of faunal structures among European butterfly species (Lepidoptera: Rhopalocera). – Zool. J. Linn. Soc. 101: 1–49.
Dennis, R. L. H., Sparks, T. H. and Hardy, P. B. 1999. Bias in butterfly distribution maps: the effects of sampling effort. – J. Insect Conserv. 3: 33–42.
Dobson, A. 1999. An introduction to Generalized Linear Models. – Chapman and Hall/CRC.
Edwards, J. L., Lane, M. A. and Nielsen, E. S. 2000. Interoperability of biodiversity databases: biodiversity information on every desktop. – Science 289: 2312–2314.
García-Barros, E. et al. 2002. Parsimony analysis of endemicity and its application to animal and plant geographical distributions in the Ibero-Balearic region (Western Mediterranean). – J. Biogeogr. 29: 109–124.
García Hernández, E. and Bosque Sendra, J. 2001. Bases de datos cartográficas de cobertura global accesibles on-line. – Geofocus (Recursos) 1: 5–10. Available at <http://geofocus.rediris.com>.
Garcia-Pereira, P. N. C. 2003. Mariposas Diurnas de Portugal Continental: Faunística, Biogeografía y Conservación. – Ph.D. thesis, Facultad de Ciencias, Dept de Biología, Univ. Autónoma de Madrid.
Gotelli, N. J. and Colwell, R. K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. – Ecol. Lett. 4: 379–391.
Griffiths, G. H., Eversham, B. C. and Roy, D. B. 1999. Integrating species and habitat data for nature conservation in Great Britain: data sources and methods. – Global Ecol. Biogeogr. 8: 329–345.
Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Modell. 135: 147–186.
Heikkinen, R. K. and Neuvonen, S. 1997. Species richness of vascular plants in the subarctic landscape of northern Finland: modelling relationships to the environment. – Biodiv. Conserv. 6: 1181–1201.
Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. – Biol. J. Linn. Soc. 68: 87–112.
Hortal, J. and Lobo, J. M. 2001. A preliminary methodological approach to model the spatial distribution of biodiversity attributes. – In: Mateu, J. and Montes, F. (eds), Spatio-temporal modelling of environmental processes. Proc. of the 1st Spanish Workshop of Spatio-temporal Modelling of Environmental Processes. Publicacions de la Univ. Jaume I, Col·lecció "Treballs d'Informàtica I Tecnologia" 10, Castelló de la Plana, pp. 211–239.
Hortal, J. and Lobo, J. M. 2002. Una metodología para predecir la distribución espacial de la diversidad biológica. – Ecología (n.s.) 16: 151–178+14 Fig.
Hortal, J., Lobo, J. M. and Martín-Piera, F. 2001. Forecasting insect species richness scores in poorly surveyed territories: the case of the Portuguese dung beetles (Col. Scarabaeinae). – Biodiv. Conserv. 10: 1343–1367.
Hortal-Muñoz, J., Martín-Piera, F. and Lobo, J. M. 2000. Dung beetle geographic diversity variation along a Western Iberian latitudinal transect (Coleoptera: Scarabaeidae). – Ann. Entomol. Soc. Am. 93: 235–243.
Johnston, C. A. 1998. Geographic Information Systems in ecology. – Blackwell.
Lawton, J. H. 1999. Are there general laws in ecology? – Oikos 84: 177–192.
Lawton, J. H., Prendergast, J. R. and Eversham, B. C. 1994. The numbers and spatial distributions of species: analyses of British data. – In: Forey, P. L., Humphries, C. J. and Vane-Wright, R. I. (eds), Systematics and conservation evaluation. Syst. Assoc. Spec. Vol. No. 50. Clarendon Press, Oxford, pp. 177–195.
Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? – Ecology 74: 1659–1673.
Legendre, P. and Legendre, L. 1998. Numerical ecology, 2nd ed. – Elsevier.
León Cortés, J. L., Soberón-Mainero, J. and Llorente-Bousquests, J. 1998. Assesing completeness of Mexican sphinx moth inventories through species accumulation functions. – Div. Distrib. 4: 37–44.
Levin, S. A. 1992. The problem of pattern and scale in ecology. – Ecology 73: 1943–1967.
Lobo, J. M. 2000. ¿Es posible predecir la distribución geográfica de las especies basándonos en variables ambientales? – In: Martín-Piera, F., Morrone, J. J. and Melic, A. (eds), Hacia un proyecto CYTED para el inventario y estimación de la diversidad entomológica en Iberoamérica: PrIBES 2000. m3m-Monografías Tercer Milenio. Vol. 1. Sociedad Entomológica Aragonesa (SEA), Zaragoza, pp. 55–68.
Lobo, J. M. and 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, J. M. and Martín-Piera, F. 2002. Searching for a predictive model for Iberian dung beetle species richness based on spatial and environmental variables. – Conserv. Biol. 16: 158–173.
Lobo, J. M., Castro, I. and Moreno, J. C. 2001. Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. – Biol. J. Linn. Soc. 73: 233–253.
Lobo, J. M., Lumaret, J. P. and Jay-Robert, P. 2002. Modelling the species richness distribution of French dung beetles and delimiting the predictive capacity of different groups of explanatory variables (Coleoptera, Scarabaeidae). – Global Ecol. Biogeogr. 11: 265–277.
Lumaret, J. P. and Lobo, J. M. 1996. Geographic distribution of endemic dung beetles (Coleoptera Scarabaeoidea) in the Western Palaearctic region. – Biodiv. Lett. 3: 192–199.
Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: the disctintion between – and reconciliation of – "predictive" and "explanatory models. – Biodiv. Conserv. 9: 655–671.
Mac Nally, R. 2002. Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. – Biodiv. Conserv. 11: 1397–1401.
Mac Nally, R. et al. 2003. Modelling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the Great Basin of western North America. – Biol. Conserv. 110: 21–31.
Margules, C. R., Nicholls, A. O. and Austin, M. P. 1987. Diversity of *Eucalyptus* species predicted by a multi-variable environment gradient. – Oecologia 71: 229–232.
Martín, J. et al. 2000. High endemism areas in the Iberian Peninsula. – Belg. J. Entomol. 2: 47–57.
McCarthy, M. A., Lindenmayer, D. B. and Possingham, H. P. 2001. Assessing spatial PVA models of arboreal marsupials using significance test and Bayesian statistics. – Biol. Conserv. 98: 191–200.
McCullagh, P. and Nelder, J. A. 1989. Generalized Linear Models, 2nd ed. – Chapman and Hall.
Miller, R. I. 1994. Mapping the diversity of nature. – Chapman and Hall.
Moreno, C. E. and Halffter, G. 2000. Assesing the completeness of bat biodiversity inventories using species accumulation curves. – J. Appl. Ecol. 37: 149–158.
Nicholls, A. O. 1989. How to make biological surveys go further with generalised linear models. – Biol. Conserv. 50: 51–75.
Nicholls, A. O. 1991. Examples of the use of Generalised Linear Models in analysis of survey data for conservation evaluation. – In: Margules, C. R. and Austin, M. P. (eds), Nature conservation: cost-effective biological surveys and data analysis. CSIRO, pp. 54–63.
Pascual, M. A. and Iribarne, O. O. 1993. How good are empirical predictions of natural mortality? – Fish. Res. 16: 17–24.
Pearson, D. L. and Carroll, S. S. 1999. The influence of spatial scale on cross-taxon congruence patterns and prediction accuracy of species richness. – J. Biogeogr. 26: 1079–1090.
Purvis, A. and Hector, A. 2000. Getting the measure of biodiversity. – Nature 405: 212–219.
Ramos, M. A., Lobo, J. M. and Esteban, M. 2001. Ten years inventorying the Iberian fauna: results and perspectives. – Biodiv. Conserv. 10: 19–28.
Rich, T. 1998. Squaring the circles – bias in distribution maps. – Brit. Wildl. 9: 213–219.
Sá-Sousa, P. 2000. A predictive distribution model for the Iberian wall lizard (*Podarcis hispanicus*) in Portugal. – Herpetol. J. 10: 1–11.
Scott, J. M. 1998. Large-area mapping of biodiversity. – Ann. Missouri Bot. Gard. 85: 34–47.
Soberón, J. and Llorente, J. 1993. The use of species accumulation functions for the prediction of species richness. – Conserv. Biol. 7: 480–488.
Spector, S. 2002. Biogeographic crossroads as priority areas for conservation. – Conserv. Biol. 16: 1480–1487.
Thomas, C. D. et al. 2001. Ecological and evolutionary processes at expanding range margins. – Nature 411: 577–581.
Verdú, J. R. and Galante, E. 2002. Climatic stress, food availability and human activity as determinants of endemism patterns in the Mediterranean region: the case of dung beetles (Coleoptera, Scarabaeoidea) in the Iberian Peninsula. – Div. Distrib. 8: 259–274.
Wohlgemuth, T. 1998. Modelling floristic species richness on a regional scale: a case study in Switzerland. – Biodiv. Conserv. 7: 159–177.