



Forecasting insect species richness scores in poorly surveyed territories: the case of the Portuguese dung beetles (Col. Scarabaeinae)

JOAQUÍN HORTAL*, JORGE M. LOBO and FERMÍN MARTÍN-PIERA

*Departamento de Biodiversidad y Biología Evolutiva (Entomología). Museo Nacional de Ciencias Naturales (C.S.I.C.), C/ José Gutiérrez Abascal 2, 28006 Madrid, Spain; *Author for correspondence (e-mail: jhortal@mcn.csic.es; fax: +34 +91 5645078)*

Received 31 May 2000; accepted in revised form 28 September 2000

Abstract. Large-scale biodiversity assessment of faunal distribution is needed in poorly sampled areas. In this paper, Scarabaeinae dung beetle species richness in Portugal is forecasted from a model built with a data set from areas identified as well sampled. Generalized linear models are used to relate the number of Scarabaeinae species in each Portuguese UTM 50 × 50 grid square with a set of 25 predictor variables (geographic, topographic, climatic and land cover) extracted from a geographic information system (GIS). Between-squares sampling effort unevenness, spatial autocorrelation of environmental data, non-linear relationships between variables and an assessment of the models' predictive power, the main shortcomings in geographic species richness modelling, are addressed. This methodological approach has proved to be reliable and accurate enough in estimating species richness distribution, thus providing a tool to identify areas as potential targets for conservation policies in poorly inventoried countries.

Key words: dung beetle, generalized linear models, geographic information system (GIS), insect conservation, Portugal, sampling effort, Scarabaeinae, spatial autocorrelation, species richness forecast

Abbreviations: ALTMAX – maximum altitude; ALTMED – mean altitude; ALTMIN – minimum altitude; ALTRNG – altitude range; BANDASCA – database of Iberian dung beetles (Col. Scarabaeinae) which contains all the available information for this group geographically referenced; DEM – digital elevation model; DOSUN – days of sun; DPYR – distance from Pyrenees; GACID – acid rock area; GCAL – calcareous rock area; GCLAY – clay area; GDIV – geologic diversity; GIS – geographic information system; GLM – generalized linear models; LAT – latitude; LON – longitude; LUDIV – land use diversity; LUFOR – forest area; LUGRS – grassland area; LUSHRB – scrub area; LUURB – cultivated and urban area; MPE – mean prediction error; PLV – potential leverage; PMED – mean precipitation; PRNG – precipitation range; PSUMM – summer precipitation; TERRAR – terrestrial area; TMAX – maximum temperature; TMED – mean temperature; TMIN – minimum temperature; TRNG – temperature range; UTM – universal transverse mercator geographic reference system; UTM50 – UTM 50 × 50 km grid square

Introduction

Quantitative determination of biodiversity provides tools to rank areas by species richness to optimise conservation effort (see Araújo 1999; Wessels et al. 2000 and references therein). Only in countries such as UK there is a long-standing tradition of natural history and sufficient resources to produce good distribution maps based on

adequate sampling of a range of taxonomic groups (Lawton et al. 1994; Griffiths et al. 1999). As well-known groups and exhaustively inventoried areas are not the general rule for most biota, only a few local inventories (e.g. Melic and Blasco-Zumeta 1999) and the results of sporadic uncoordinated captures exist. This is especially true for hyperdiverse organisms such as insects inhabiting severely threatened-rich areas such as the Mediterranean basin (Balleto and Cassale 1991). With currently available resources it would be impossible to sample Mediterranean fauna adequately before diversity decreased below critical levels.

As occurs in all orders of insects (Samways 1993, 1994; Martín-Piera 1997) concern has been increasing over dung beetle conservation in the last decade (Klein 1989; Lumaret 1994; Martín-Piera and Lobo 1995; Barbero et al. 1999; Van Rensburg et al. 1999). Agro pastoral civilization decline in the last few decades – the massive extension of monocultures, the disappearance of domestic livestock from highly mechanized agricultural landscapes (Barbero et al. 1999) and anti-parasitic treatments (Lumaret 1986; Lumaret et al. 1993) probably led many species that were abundant at the beginning of this century, to become very scarce or even extinct (Johnson 1962; Leclerc et al. 1980; Lumaret 1990; Väisänen and Rassi 1990; Biström et al. 1991; Lumaret and Kirk 1991; Miessen 1997; Lobo 2001). The effect of such decline on dung beetle diversity is particularly striking in the Mediterranean region, as this region contains numerous large species, particularly Scarabaeinae, which are absent at northern latitudes (Lumaret and Kirk 1991).

Scarabaeinae dung beetle feed and breed currently on large herbivorous feces. They dig a vertical tunnel below the dung pat, carrying dung into the bottom of the burrow or roll away a dung ball, digging a tunnel outside the dung pat. Dung beetle play a major ecological role in pastures and grassland biomes removing vertebrate feces (mainly domestic and wild ungulates), draining soils, and aiding recycling of organic matter and nutrients, thus improving pasture quality and productivity (Fincher 1981; Rougon et al. 1988). Hence, dung beetles can be viewed as a cornerstone of natural grassland biomes and Mediterranean livestock grazing maintained by traditional grazing coupled with low-intensity agricultural systems (Martín-Piera and Lobo 1995).

The recovering of a picture of species richness distribution of highly diverse taxa in poorly surveyed areas, becomes a colossal task because of the resources constraints both in manpower and financial support. Obviously, an important first step is to compile and map all available biological information. Once this information is mapped large spatial gaps appear. As the key question is whether areas of low species richness are either truly poor or bad-sampled areas is necessary to discriminate poorly from well-sampled territories in order to develop new sampling programs to assess biodiversity rapidly in the poorly sampled areas. However, as investments are limited, spatial distribution forecast from environmental information has been proposed as a surrogate method for species richness estimation (Gaston 1996).

Commonly, the richest areas are supposed to be those most intensely sampled (Gaston 1996). Methodologies to estimate species richness that compensate for un-

equal sampling effort have been proposed (Prendergast et al. 1993; Colwell and Coddington 1995; Colwell 1997). However, in regions where the sampling effort has been close to zero or null, species richness scores cannot be corrected for unequal sampling effort. Hence, species richness predictive functions from environmental data must be recovered from adequately sampled areas. These areas can be identified through the number of species added to the inventory as an asymptotic function of sampling effort (Soberón and Llorente 1993; Fagan and Kareiva 1997).

In order to model relationships between species richness and environmental variables, recent papers (Austin et al. 1996; Heikkinen and Neuvonen 1997) have explored the use of generalized linear models (GLM; McCullagh and Nelder 1989). Herein, these models have been used to forecast (*sensu* Legendre and Legendre 1998) Scarabaeinae dung beetle species richness from environmental variables in poorly sampled areas (Nicholls 1989).

Taxonomic knowledge of this group in the Iberian Peninsula is fairly good (see Martín-Piera 2000). However, Portuguese scarabeid dung beetles (Coleoptera, Scarabaeinae) have been poorly sampled. The only available faunistic data for Portugal comes from old catalogues (Preudhomme de Borre 1886; Oliveira 1894; Seabra 1907, 1909; Ladeiro 1950; Luna de Carvalho 1950) and a recent study (Hortal-Muñoz et al. 2000). A generalized sampling effort of Portuguese Scarabaeinae over all the country, especially at intermediate latitudes, is sorely needed.

The general aim of this paper is to demonstrate that reliable species richness estimations can be obtained from a small dataset within a poorly sampled zone. The number of Scarabaeinae species from the well sampled 50 km UTM grid squares (UTM50) was modelled to forecast feasible species richness scores for this group over Iberian Portugal. GLM were applied to the biological information compiled by a database of the Iberian dung beetles (*BANDASCA*; Lobo and Martín-Piera 1991), using 25 environmental predictor variables to find a dung beetle species richness model able to produce reliable results. This model was used to forecast the richness scores for the total UTM50 comprising the entire Portuguese territory. Emphasis has been made on the shortcomings derived from the building of species richness predictive models (collinearity and spatial autocorrelation of predictor variables, interaction between explanatory variables, and non-linear relationships between dependent variable and independent variables).

Methods

Selecting reliable biological information

Origin of biological data

An adequate description of species richness distribution must include a measure of the associated sampling effort (Gaston 1996). But data from heterogeneous sources

(bibliographic, museum, etc.), collected according to varying goals and techniques lack such a measure. This is the case for the distribution data of the Portuguese dung beetles in *BANDASCA*, a database which compiles all the biological and geographical information from literature, museum and private collections, doctoral thesis, as well as other unpublished data available for the entire Iberian Peninsula (see structure in Lobo and Martín-Piera 1991). At present, it contains 15 023 database-records of 96 981 individuals of the 55 Iberian Scarabaeinae species.

The number of database-records in each UTM50 grid square was chosen as a surrogate for the sampling effort undertaken within it, on the assumption that probability of species occurrence in each square correlates positively with the number of database-records. A *BANDASCA* database-record was defined as a pool of specimens of a single species with identical database field values (locality, UTM coordinates, altitude, date of capture (day/month/year), type of habitat and food resource; among others) regardless of the number of specimens; so any difference in any database field value gave rise to a new database-record. Table 1 shows the number of database-records and species found in each grid square.

The 51 UTM50 that cover Portuguese territory and have 15% or more area not covered by water (Table 1, Figure 1) were used. A sampling-effort map (Figure 1) indicates that almost a half of those squares have fewer than 10 database-records (25 squares, 49.02%; Table 1). Hence, correcting species richness for sampling bias (Prendergast et al. 1993; Colwell and Coddington 1995; Colwell 1997) in Portugal is impossible.

Identifying the adequately-sampled squares

To ensure forecasting model reliability, adequately-sampled squares were identified through the asymptotic relationship between the number of database-records and species richness in the 51 Portuguese UTM50. This relationship was calculated for each Portuguese physioclimatic subregion, because environmental conditions may determine different values for this asymptote. To identify Portuguese physioclimatic subregions, all grid squares were classified by cluster analysis according to 15 environmental variables using Ward's method as linkage rule and the squared Euclidean distance as a measure of the similarity between each pair of UTM50 squares. The squared Euclidean distance was chosen because it maximizes the distance for groups further apart (STATISTICA 1999). The environmental variables used to classify the Portuguese UTM50 squares were: eight climatic variables (minimum, maximum and mean annual temperature, annual temperature range and days of sun, mean annual and summer precipitation, and annual precipitation range); four topographic (maximum, minimum and mean altitude, and altitude range); and three bedrock geologic (calcareous rock, acid rock and clay area). All variables were standardised. Finally, using the groups of squares previously defined by cluster analysis, a discriminant function analysis was used to verify to which group a particular square belongs; the misclassified squares were reclassified according to the classification functions.

Table 1. UTM 50 × 50 km grid cells (UTM50) used. Data for number of database-records (DB-records), species and individuals have been extracted from BANDASCA database (Lobo and Martín-Piera 1991). UTM50 identified as well sampled are given in bold.

UTM grid cell number	UTM designation	No. of DB-records	Species found	Collected individuals	UTM cell number	UTM designation	No. of DB-records	Species found	Collected individuals
1	29TNG1	53	15	148	27	29SMD3	0	0	0
2	29TNG3	14	7	21	28	29SND1	7	5	7
3	29TPG1	3	3	3	29	29SND3	8	8	140
4	29TPG3	14	6	26	30	29SPD1	7	7	44
5	29TNG2	5	5	5	31	29SPD3	38	9	97
6	29TNG4	42	16	45	32	29SMD4	5	5	5
7	29TPG2	3	3	3	33	29SND2	42	18	46
8	29TPG4	14	8	23	34	29SND4	14	12	72
9	29TQG2	3	3	3	35	29SPD2	6	6	6
10	29TNE1	12	5	12	36	29SPD4	100	18	130
11	29TNE3	8	8	153	37	29SMC3	39	24	39
12	29TPE1	5	5	5	38	29SNC1	36	14	70
13	29TPE3	7	6	7	39	29SNC3	11	9	11
14	29TQF1	110	17	566	40	29SPC1	0	0	0
15	29TNE2	0	0	0	41	29SPC3	35	21	37
16	29TNE4	14	9	850	42	29SNC2	2	2	2
17	29TPE2	2	2	9	43	29SNC4	16	17	456
18	29TPE4	95	19	466	44	29SPC2	0	1	0
19	29TNE1	3	2	4	45	29SPC4	12	9	26
20	29TNE3	48	27	530	46	29SNB1	6	5	6
21	29TPE1	57	22	179	47	29SNB3	5	5	5
22	29TPE3	153	24	618	48	29SPB1	3	3	3
23	29TNE2	35	24	99	49	29SNB2	62	23	82
24	29TNE4	2	2	2	50	29SNB4	22	9	27
25	29TPE2	5	5	10	51	29SPB2	6	5	7
26	29TPE4	39	14	111					

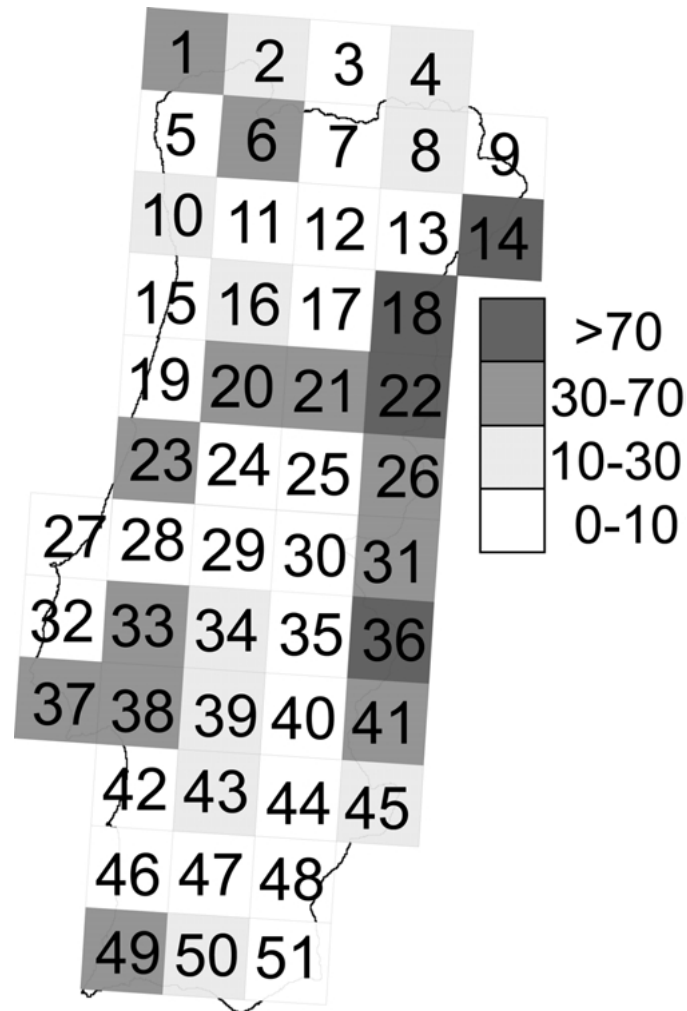


Figure 1. UTM 50×50 grid cells (UTM50) used in this paper, showing the identification number used for each one (see Table 1), and sampling effort distribution in Portugal. Four sampling effort categories were established from the number of database-records (see right scale).

Climatic data for each UTM50 were provided by W. Cramer (CLIMATE database version 2; <http://www.pik-potsdam.de/~cramer/climate.html>). Annual temperature and precipitation ranges as the difference between the extreme monthly data for each case were calculated. The topographic data from a digital elevation model (DEM) of the entire Iberian Peninsula, spatial resolution (pixel width) of 1 km, was extracted by overlaying the DEM with the polygons of the UTM50 in a geographic information system (GIS; Idrisi 2.0 1998). To obtain bedrock data, a three category (calcareous, acid and clay soils) map of Iberia (Instituto Geográfico Nacional 1995), spatial resolution 1 km, was digitised and the same overlaying technique applied.

For well-known taxa (ideally, all the species can be collected and identified), the adequacy of sampling in each square can be determined by a negative exponential function relating the number of species (S_r) to the number of database-records (r). According to Soberón and Llorente (1993) and Colwell and Coddington (1995) this relationship is:

$$S_r = S_{\max}[1 - \exp(-br)]$$

where S_{\max} , the asymptote, is the estimated total number of species per square, and b is a fitted constant that controls the shape of the curve. The function was curvilinearly fitted by the Quasi-Newton method.

Theoretically, reaching 100% of total richness would require an infinite number of database-records. In order to identify those squares adequately-sampled enough to be included in richness estimations, the number of database-records necessary to reach a rate of species increment ≤ 0.01 ($r_{0.01}$; one added species for each 100 additional database-records) was calculated for each one of the physioclimatic regions previously defined. According to Soberón and Llorente (1993), the number of database-records necessary is:

$$r_{0.01} = 1/b \cdot \ln(1 + b/0.01)$$

where $r_{0.01}$ was chosen as the number of database-records necessary to qualify a UTM50 as adequately-sampled rather than a more restrictive level (e.g. $r_{0.001}$; one added species each 1000 additional database-records) because of the small number of cases (UTM50; 51) and the maximum number of database-records per square (153; Table 1).

Forecasting species richness

Origin of the explanatory variables

In order to model and forecast Scarabaeinae spatial patterns of species richness in Portugal, biological and environmental data from the adequately-sampled UTM50 identified by the above-mentioned negative exponential function were used.

The 25 continuous predictor variables chosen were the same as those used to define the physioclimatic subregions, plus two spatial variables (UTM square centroid longitude and latitude; LON and LAT respectively), two geographic (distance from Pyrenees and land area in each UTM50), four land use variables (grassland area, cultivated and urban area, scrub area and forest area), and two environmental diversity variables (bedrock diversity and land use diversity).

The geographic and spatial variables were extracted from the Iberian DEM included in the GIS by the same overlay technique used to obtain the topographic data. To extract the land use data from the geographic raster CORINE Land Use/Land

Cover database (CORINE Programme 1985–1990; spatial resolution of 282 m) provided by the European Environment Agency (1996) the same overlay technique was applied. The 44 categories for Portugal and Spain were previously reclassified in the GIS (Idrisi 2.0 1998) to group all forest areas (any kind of forest), all scrub areas, all grassland areas (either natural and artificial) and all areas with intensive human occupation (urban, industrial and cultivated areas).

Land use diversity was estimated with the Shannon Diversity Index (Magurran 1988):

$$H' = -\sum p_i \cdot \ln p_i$$

where p_i is the relative frequency of each one of the 44 land use categories recovered from the CORINE Land Use/Land Cover database. Bedrock diversity was estimated by applying the same index to the relative frequencies of the three bedrock classes.

Predictor variables were standardised to zero means and unit variances to eliminate the effect of differences in the measurement scale for the different independent variables. The algorithm used for the standardization was:

$$\text{Std. value} = (\text{raw value} - \text{mean})/\text{SD}$$

except for the case of LAT and LON, that were standardized as recommended by Legendre and Legendre (1998):

$$\text{Std. value} = \text{raw value} - \text{mean}$$

Model building

There are four main problems with the use of environmental variables to build species richness predictive models: (1) the collinearity, and thus interdependence, of predictor variables used; (2) the spatial autocorrelation of variables; (3) the usual non-linear relationship between the dependent and independent variables; and (4) the frequently complex interactions among explanatory variables.

Environmental variables are usually correlated so that the problem of multicollinearity among independent variables becomes an important shortcoming in the model building procedures when the target is to find those variables that likely influence the variation in the dependent variable (Mac Nally 2000). But, if the aim is to forecast (to find a regression equation which describes the changes in a variable in response to changes in others for future use to make predictions) maximizing the explained variance of the data, and not to make ecological inferences, collinearity of the explanatory variables is not a concern (Legendre and Legendre 1998, pp. 518–519).

As spatial heterogeneity in nature is the result of non-random processes, spatial autocorrelation is also an intrinsic propriety of biological and environmental variables (Legendre and Legendre 1998). Autocorrelation in one variable implies the spatial dependence of observations, invalidating the assumption on which classical statistical tools are based, since the observed values of variables at any given locality are

influenced by those of the neighboring localities. What can be done in this case? The removal of spatial autocorrelation, a consequence of the processes that lead to species richness spatial patterns, would diminish the impact of spatially structured factors (i.e. neither randomly nor uniformly distributed), thus reducing the forecasting ability of the model (Smith 1994; Legendre and Legendre 1998). The key criterion when developing regression models with spatially autocorrelated data is to check if function errors (residual scores) from the final model are spatially autocorrelated. In this case, at least one spatially structured variable was not included in the analysis (Cliff and Ord 1981; Odland 1988). Spatial variables have been included in the modelling procedure (see below) in order to include these hypothetically ignored variables. To check that function errors are not spatially structured Moran's *I* and Geary's *C* autocorrelation tests were used (Legendre and Vaudor 1991), dividing all the possible UTM50 pairwise comparisons in eight distance classes.

A GLM procedure was used to model variation in Scarabaeinae species richness as a function of the most significant environmental and spatial explanatory variables (McCullagh and Nelder 1989; Dobson 1999; see Austin 1980; Nicholls 1989, 1991; Tonteri 1994; Austin et al. 1996 and Heikkinen and Neuvonen 1997 for a description of the method and some examples). A Poisson error distribution for the number of Scarabaeinae species was assumed, and was linked to the set of predictor variables by means of a logarithmic link function (Crawley 1993).

The adequacy of the models developed was tested by means of the change from a null model in which the number of parameters is equal to the total number of observations ($n = 16$) and the species richness is modelled alone (with no explanatory variables; see Dobson 1999). The goodness-of-fit of the models was measured by the deviance statistic and the change in deviance *F*-ratio tested (McCullagh and Nelder 1989; Dobson 1999). As the low number of cases could influence the reliability of the model built, in order to choose the most reliable, two different models were developed at 0.01 and 0.05 change in deviance significance levels.

A forward stepwise procedure was used to enter the variables into the model (see Austin et al. 1996 and Heikkinen and Neuvonen 1997). In order to account for non-linear relationships, in a first step the total number of Scarabaeinae species registered on each adequately-sampled UTM50 was related, one-by-one, with each predictor variable's linear, quadratic and cubic function (including LAT and LON). The function that accounted for the highest reduction in deviance from that of the null model was selected (Austin 1980; Margules et al. 1987; Austin et al. 1996). Next, from the functions selected before, the one that accounted for the most important change in deviance was chosen. Then all the remaining functions were added one-by-one to the model and tested again for significance in the change of deviance. The one which accounted for the most significant change was included in the model. After each significant inclusion, the new model was submitted to a backward selection procedure, in order to eliminate those terms that had become non-significant. This procedure was repeated iteratively until no more statistically significant changes remained.

As interactions between variables are often highly predictive (Margules et al. 1987), subsequently the importance of all the interaction terms between explanatory variables (including spatial) was tested by adding them sequentially one by one to the previously obtained model. Again, a backward procedure was used after each forward inclusion.

Finally, spatial variables were added to the model. As commented before, this would include in our analysis the influence of ignored spatially structured variables, also diminishing the probability of occurrence of spatially autocorrelated residuals. The nine terms of the third-degree polynomial equation of latitude and longitude (trend surface analysis; $b_1\text{LAT} + b_2\text{LON} + b_3\text{LAT}^2 + b_4\text{LAT} \times \text{LON} + b_5\text{LON}^2 + b_6\text{LAT}^3 + b_7\text{LAT}^2 \times \text{LON} + b_8\text{LAT} \times \text{LON}^2 + b_9\text{LON}^3$) were added to the former model and submitted to the backward stepwise selection procedure in order to remove the non-significant terms (see Legendre 1993; Legendre and Legendre 1998).

Stepwise multiple selection procedures now are regarded as highly flawed techniques in which the resulting final model can be biased (Daniel and Wood 1980; Derksen and Keselman 1992; Mac Nally 2000). The critics claim an exhaustive search for computing all 2^k possible models in order to select just the 'best' model. We are not engaged in such as an exhaustive search because of the few considered observations and high quantity of independent variables (2^{25}) plus the interaction terms and complex spatial terms. We are mostly interested in the finding of a model with a high predictive power and the inclusion of non-linear relationships, the exhaustive exploration of all interaction terms among independent variables, and the consideration of the spatial variables warranty the building of a high predictive power model.

Goodness-of-fit and predictive power of the models

The best method to test model reliability is empirical: an inventory taken in the poorly sampled zones to check if predicted and real scores are similar. However, as this method exceeds the scope of this work, statistical techniques must be applied to the dataset to test model reliability (Mac Nally 2000). To check the final model, a Jackknife test was carried out. With a data set of sixteen UTM50 squares the model was recalculated sixteen times, leaving out one square in turn. Each one of the regression models based on the $n - 1$ grid squares was then applied to that excluded square, to predict species richness score in each UTM50. Then observed and estimated values were checked for correlation using the Pearson correlation coefficient.

The percentage of explained deviance for each model was calculated to obtain an estimation of the total variability of the data explained by each model (see Dobson 1999). Moreover, to estimate the predictive power of the model, the relative distance between the predicted value for case i when excluded in the model building (P_i) and the observed score (O_i) is used as a prediction error (E_i) for that observation (Pascual and Iribarne 1993). The percentage error for case i is:

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

The mean of all the error estimations (mean prediction error; MPE) was used as a measure of the prediction error associated with each one of the models, and the inverse of this measure ($MPE^{-1} = 100 - MPE$) as an estimation of the predictive power of each one of them. These results were used to assess the selection of the most reliable model between the two found at 0.05 and 0.01 *P*-levels.

Residual analysis

Once the best model was chosen, the residual analysis recommended by Nicholls (1989) was carried out to identify outliers, those grid squares in which the residual absolute value is greater than the standard deviation of the predicted values. Points with high scores of potential leverage (PLV) were also selected. The PLV is a measure of the distance of each observation from the centroid of the multi-dimensional space defined by the variables included in the model. Each of the outliers and observations with high PLV was explored in order to ascertain if it were due to erroneous data, or if it included environmental variability not found in the rest of the observations. While the former should be deleted, the latter kind of observations may remain in the model in order to include as much environmental heterogeneity as possible. The final model parameters were then estimated after the deletion of the real outliers.

The STATISTICA package (1999) was used for all statistical computations, except the autocorrelation tests, made by means of the *R* Package (Legendre and Vaudor 1991).

Results

Identifying adequately-sampled squares

Cluster analysis of environmental data shows two well-defined groups of Portuguese UTM squares: the Portuguese Eurosiberian and Mediterranean physioclimatic subregions (Figure 2a). Discriminant analyses of the two regions shows that 96.1% of the UTM squares were well classified by the Cluster analysis. The two misclassified Eurosiberian squares (UTM50 numbers 25 and 35; see Figure 2b) were reassigned to the Mediterranean region. A Mann–Whitney *U*-test ratified that the explanatory variables of the squares assigned to the two regions differ significantly, except on annual temperature range (ATR) and calcareous rock area (CRA) (Table 2).

Figure 3 shows the species accumulation or ‘collector’s’ curves for both regions. The number of database-records necessary to reach a rate of species increment of one each 100 additional database-records was 40.9 for the Eurosiberian, and 33.7 for the Mediterranean region, identifying 7 Eurosiberian UTM squares (numbers 1, 6, 14, 18, 20, 21, 22), and 9 Mediterranean UTM squares (numbers 23, 26, 31, 33, 36, 37, 38, 41 and 49) as adequately sampled (see Figure 1 and Table 1).

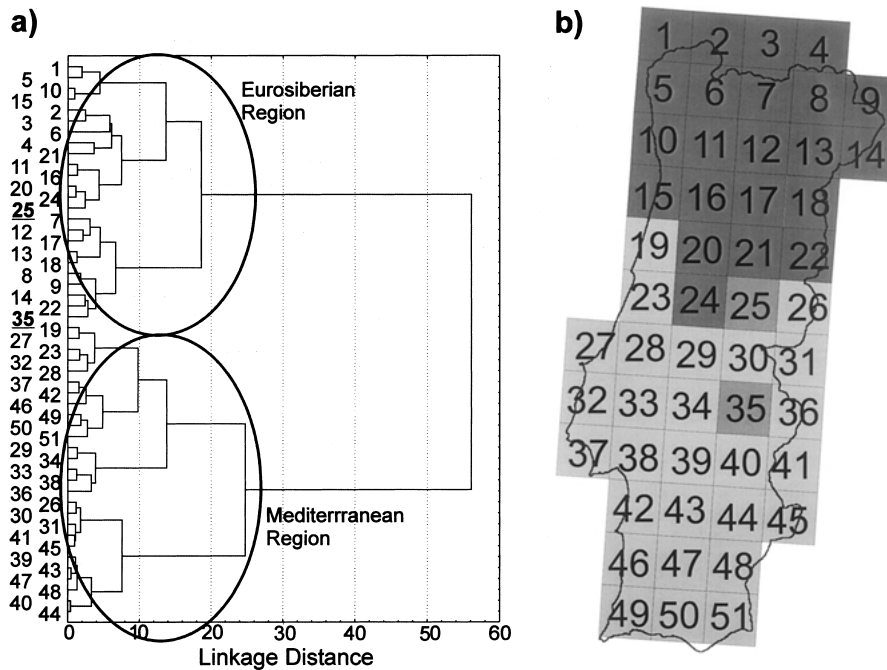


Figure 2. (a) Dendrogram from the cluster analysis for the 51 Portuguese UTM50 cells. Linkage rule was Ward's method and similarity measure between UTM squares was square Euclidean distance. Numbers on the left are UTM50 identifiers (see Table 1). Main groups identified have been rounded, and the squares that were identified as misclassified by the discriminant function analysis are highlighted. (b) Physioclimatic regions identified. Deep grey: Eurosiberian region; light grey: mediterranean region; medium grey: squares identified as Eurosiberian by the cluster analysis and reassigned to the Mediterranean region by the discriminant function analysis.

Forecast model building

Only 5 of the 23 explanatory variables accounted for a significant change in deviance as either a linear, quadratic or cubic function (Table 3). As the cubic function of mean annual precipitation ($PMED + PMED^2 + PMED^3$; change in deviance = 9.78) accounted for the maximum reduction in deviance, it was selected in the first step of the stepwise procedure (Table 4).

None of the most explanatory functions of the other 24 explanatory variables accounted for a significant change in deviance, neither at 0.01 nor at 0.05 significance levels, when added to the former function. Next, all the explanatory variables interaction terms were tested one by one. The interaction term between mean annual precipitation and the forested area ($PMED \times LUFOR$) accounts for the greatest significance. Adding this interaction term removed the linear term of the annual mean precipitation from the model. No more interaction terms nor spatial third degree poly-

Table 2. Results from the Mann–Whitney U -test ($n_1 = 22$; $n_2 = 29$) with the explanatory variable scores between the UTM50 from each physioclimatic region identified (see Figure 2).

	Z
Minimum altitude	2.273*
Maximum altitude	4.945***
Mean altitude	4.545***
Altitude range	4.831***
Minimum temperature	-4.574***
Maximum temperature	-2.967**
Mean temperature	-5.183***
Temperature range	1.626
Days of sun	-5.839***
Mean precipitation	4.907***
Summer precipitation	6.019***
Precipitation range	3.956***
Calcareous rock area	-1.312
Clay area	-3.433***
Acid rock area	3.119**

*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.

nomial terms accounted for a reduction in deviance at the 0.01 significance level. However, at a 0.05 significance level another two interaction terms (maximum temperature \times cultivated and urban area, and grassland area \times bedrock diversity) enter in the model. The model developed at a 0.01 significance level explains 76.4% of total deviance, while the 0.05 significance level model explains 92.8% (Table 4).

Choice of the most reliable model

The correlation between observed versus Jackknife-predicted scores suggest that the 0.01 significance level model gives higher predictive power (0.05 model: Pearson $r = 0.257$, $P = 0.338$; 0.01 model: Pearson $r = 0.6085$, $P = 0.012$). The MPE scores confirm this, because this model's mean prediction error is smaller with less standard deviation (0.05 model: MPE = 25.28; SD = 57.38; 0.01 model: MPE = 20.07; SD = 23.45). Thus, because of its greater reliability, the model developed at a 0.01 significance level was chosen.

Residual analysis

The residuals from this model were explored to identify possible outliers. Figure 4 shows that the residual absolute value of UTM square 14 is higher than the SD score (4.35). Data from this square indicate that this is a real outlier rather than an environmentally singular square. Figure 2a shows that it is environmentally quite similar

Table 3. Considered explanatory variables included in the analysis with respective code. Deviance and change in the deviance from a null model for total dung beetle species number. The linear, quadratic or cubic functions of each variable have been selected when accounted as statistically significant at 5% change in the deviance. The sign column shows the sign of each term of the significant functions.

Variable	Code used	Selected terms	df	Deviance	Change in deviance	F	Sign
Topographic variables							
		Null model	15	19.24			
Minimum altitude	ALTMIN	ALTMIN	14	19.11	0.13	0.094	
Maximum altitude	ALTMAX	ALTMAX	14	18.27	0.97	0.745	
Mean altitude	ALTMED	ALTMED	14	19.24	0.00	0.001	
Altitude range	ALTRNG	ALTRNG	14	17.67	1.56	1.239	
Climatic variables							
Minimum temperature	TMIN	TMIN	14	19.20	0.04	0.032	
Maximum temperature	TMAX	TMAX	14	11.81	7.43	8.806*	-
Mean annual temperature	TMED	TMED	14	17.50	1.74	1.389	
Annual temperature range	TRNG	TRNG	14	16.91	2.33	1.927	
Annual days of sun	DOSUN	DOSUN	14	19.13	0.10	0.076	
Mean annual precipitation	PMED	PMED	14	19.20	0.04	0.029	
		PMED + PMED ²	13	13.66	5.58	5.305*	+ -
		PMED + PMED ² + PMED ³	12	9.46	9.78	12.401**	+ - +
Summer precipitation	PSUMM	PSUMM	14	19.09	0.15	0.111	
Annual precipitation range	PRNG	PRNG	14	18.97	0.27	0.200	
		PRNG+ PRNG ²	13	13.63	5.61	5.356*	+ -
Land use variables							
Cultivated and urban area	LUURB	LUURB	14	19.11	0.12	0.091	
Forest area	LUFOR	LUFOR	14	18.20	1.04	0.797	
Scrub area	LUSHRB	LUSHRB	14	18.24	0.99	0.763	
Grassland area	LUGRS	LUGRS	14	14.81	4.43	4.190	
		LUGRS + LUGRS ²	13	12.97	6.27	6.288*	- -
		LUGRS + LUGRS ² + LUGRS ³	12	10.93	8.31	9.129**	- + -

Geologic variables						
Calcareous rock area	GCAL	14	16.60	2.64	2.223	
Clay area	GCLAY	14	18.37	0.87	0.664	
Acid rock area	GACID	14	19.16	0.08	0.061	
Environmental diversity variables						
Land use diversity	LUDIV	14	18.52	0.72	0.546	
Bedrock diversity	GDIV	14	18.56	0.68	0.512	
	GDIV + GDIV ²	13	11.49	7.75	8.774*	- +
Geographic variables						
Distance from pyrenees	DPYR	14	17.95	1.29	1.007	
Land area	TERRAR	14	18.15	1.09	0.838	

** $P < 0.01$; * $P < 0.05$.

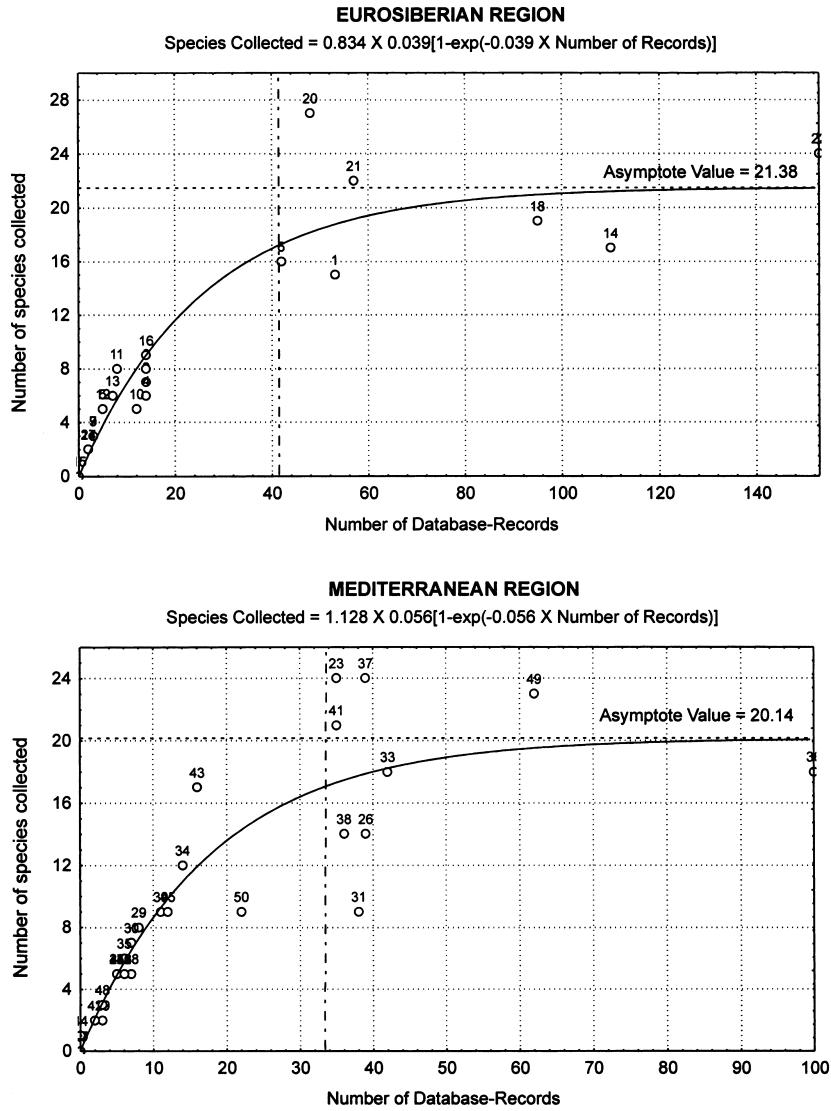


Figure 3. ‘Collector’s’ curves for each of the two Portuguese physioclimatic regions (see Figure 2b). The asymptotic function for each region is shown, as well as the asymptote value. Curve asymptotes are presented as discontinuous lines, and the number of database-records needed to identify a UTM grid cell as well sampled is shown as dots and lines. A database-record is defined as a single observation (the set of information common to one or more specimens of the same sex belonging to a single species in the BANDASCA database; see Methods section). Database-records differ by at least one database-field.

to another two adequately-sampled squares (18 and 22). The PLV of this square is also very small, showing its redundancy, so this case was deleted from the modelling process.

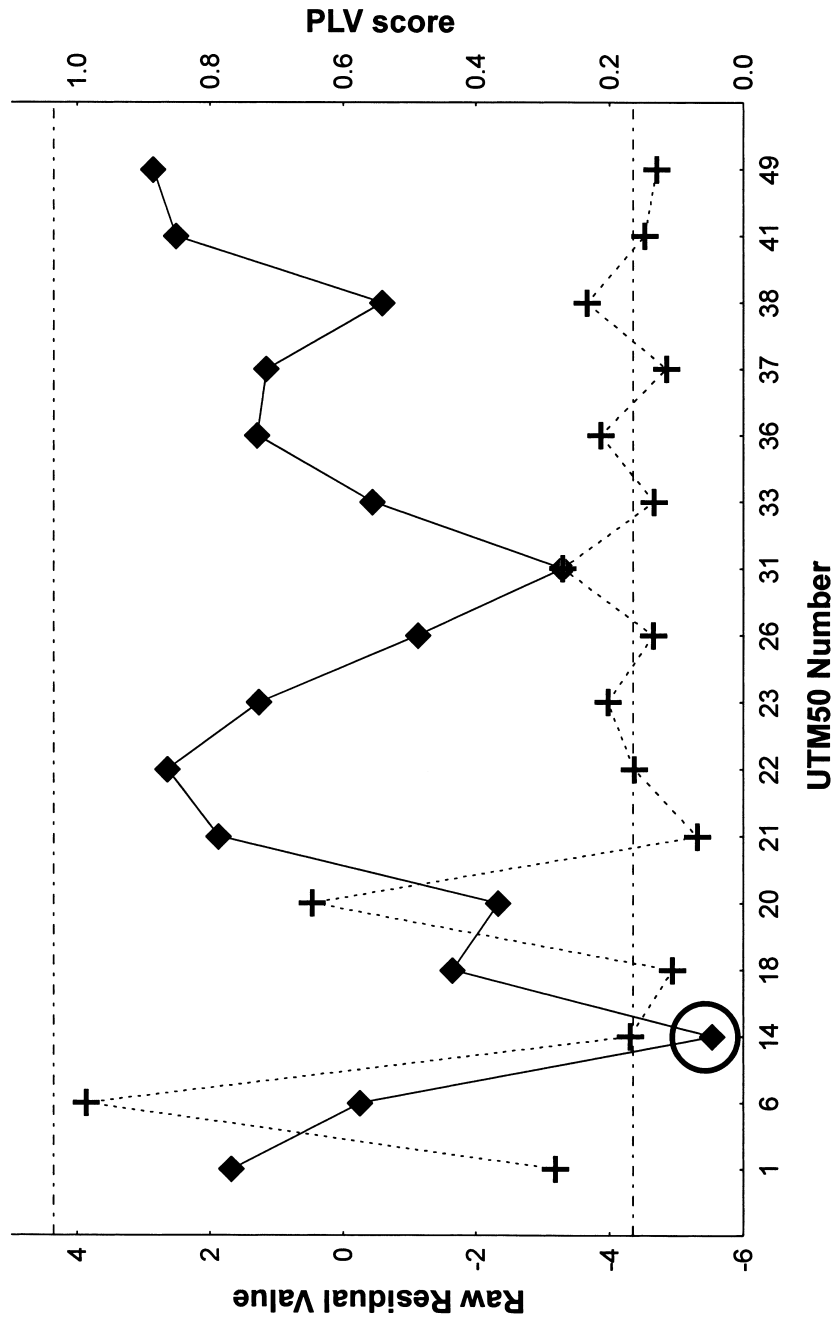


Figure 4. Residuals from the 0.01 significance level model and potential leverage (PLV) in this model of the well-sampled UTM50. Raw residual values are presented as rhombus and continuous lines, and PLV scores as crosses and discontinuous lines. Standard deviation of the predicted values is shown as points and lines, and residuals with scores higher than that are circled.

Table 4. Summary of the stepwise forward selection of variables to build the models for dung beetle species richness in the Portuguese well-sampled UTM50. The change in deviance after the inclusion of each term in the model has been tested by an *F*-ratio test with probability less than 0.05. The variable codes are the same as in Table 2.

Model	Deviance	df	Change in deviance	<i>F</i>	<i>P</i>	Total deviance (%)
Null	19.24	15				
Model at 0.01 significance level						
Step 1						
+PMED	19.20	14	0.04	0.03	NS	0.21
+PMED ²	13.66	13	5.54	5.27	0.039	28.98
+PMED ³	9.46	12	4.20	5.33	0.040	50.82
PMED + PMED ² + PMED ³	9.46	12	9.78	12.40	0.004	
Step 2						
+PMED*LUFOR	4.41	11	5.05	12.58	0.005	77.06
-PMED	4.54	12	-0.12	0.33	NS	76.42*
Model at 0.05 significance level						
Step 3						
+TMAX*LUURB	3.11	11	1.43	5.07	0.046	83.86
Step 4						
+LUGRS*GDIV	1.39	10	1.71	12.26	0.006	92.75*

* Final values for each model.

NS = not significant.

Final model parameter estimation and validation

The modelling (at a 0.01 significance level) after the deletion of the outlier did not change the variables in the model. The final model, which explains 85.4% of total deviance, was:

$$S = \text{EXP}[3.166 - 0.464 \text{PMED}^2 + 0.120 \text{PMED}^3 + 0.234(\text{PMED} \times \text{LUFOR})]$$

This model is more reliable than the former. The Pearson correlation coefficient between observed *versus* Jackknife-predicted scores changed from 0.61 to 0.93. The MPE diminished from 20.07 to 9.17%, its standard deviation being 7.85 instead of the former 23.45. The predictive power of the model (MPE^{-1}) increased from 79.97 to 90.83%.

Neither Moran's *I* nor Geary's *C* autocorrelation test for residuals show significant positive spatial autocorrelation scores at any distance class. The lack of positive spatial autocorrelation in the residuals indicates that the majority of the spatially structured variation in the data is retained by the model. In fact, when the residuals from the model are mapped (Figure 5), almost all function errors are small (up to 2

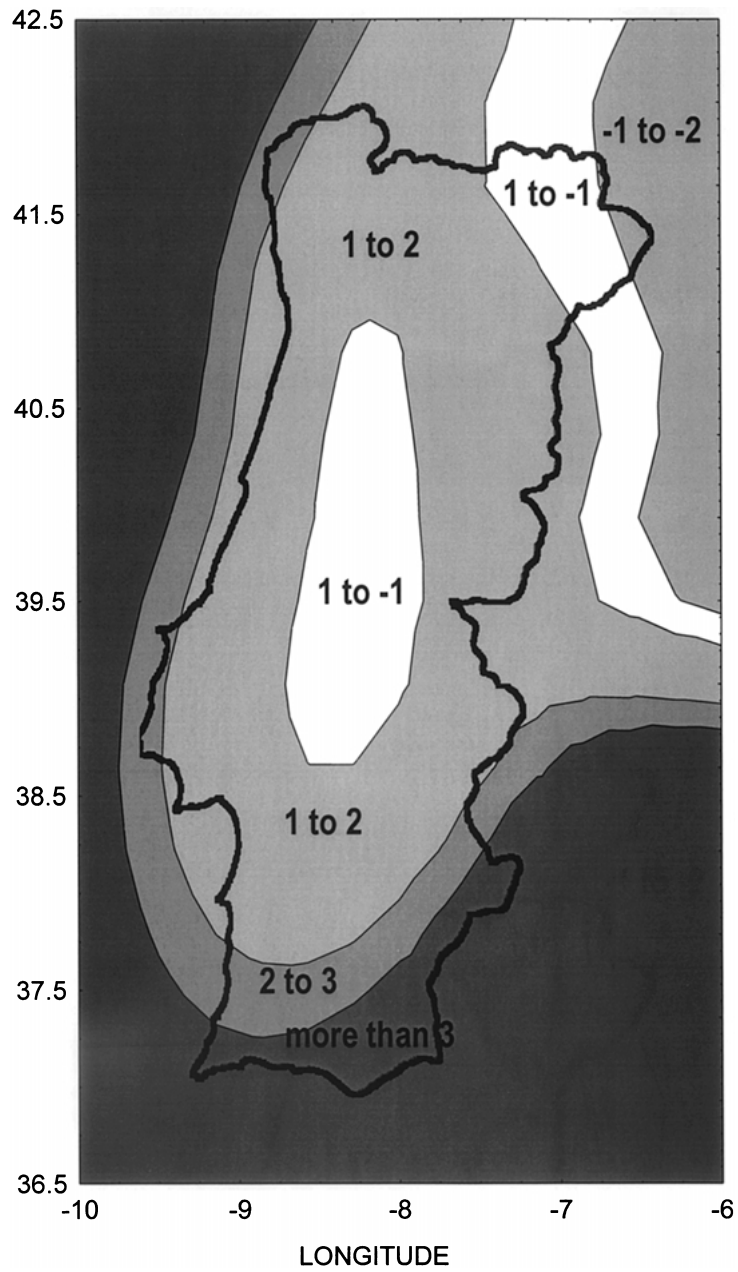


Figure 5. Geographical distribution of the residuals produced by the final model $S = \text{EXP}[3.166 - 0.464 \text{PMED}^2 + 0.120 \text{PMED}^3 + 0.234(\text{PMED} \times \text{LUFOR})]$. Longitude and latitude are expressed in decimal Greenwich degrees, and the shadowed areas represent the raw residual scores. The residual score ranges are marked in the figure.

residual absolute values). Only the residuals in the South-eastern zone are higher than 2 (the estimated species number is lower than the real scores).

Applying the final model to the data from all the Portuguese UTM grid squares, a potential map of species richness distribution was obtained (Figure 6). In this map Scarabaeinae species richness can be seen to be higher in central Portugal, around the Iberian central system western spurs (squares 20, 24 and surroundings). Two species-poor zones, with predicted richness scores lower than 17 species also appear, one in the Tejo Valley (specially squares 31, 34, 38 and 42), and another in the north-west (squares 1 and 5 and neighbouring ones).

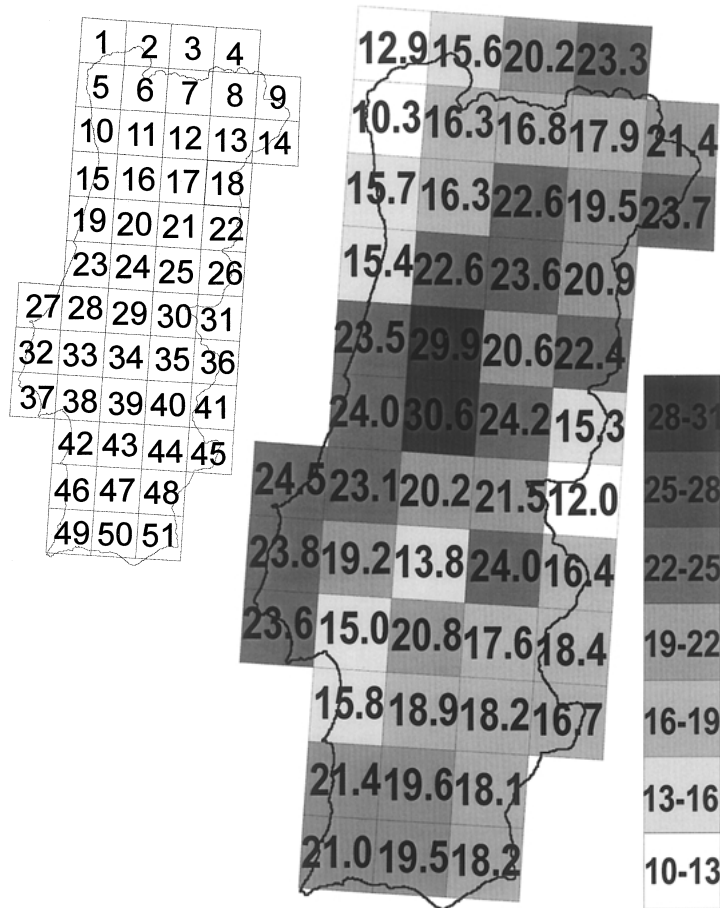


Figure 6. Potential Scarabaeinae richness scores forecast by the final model $S = \text{EXP}[3.166 - 0.464 \text{PMED}^2 + 0.120 \text{PMED}^3 + 0.234(\text{PMED} \times \text{LUFOR})]$ for all the Portuguese UTM50. The numbers on the right grey scale represent the Richness intervals used for the shadowing of the map. For UTM correspondences see Table 1.

Discussion

Highly heterogeneous biologic information sources (independent inventories kept in museums and private collections, unrelated studies, etc.) were used, with only 16 adequately-sampled Portuguese UTM 50×50 grid squares (31.4% of the total) and a broad spatial scale (UTM squares of 2500 km²). In spite of these drawbacks, the stepwise GLM-based procedure with environmental variables produced a forecasting model with a predictor capability (MPE^{-1}) greater than 90%. Obviously, better models should be developed on more and better biological information on a smaller spatial scale related more directly to the processes that determine the spatial patterns of species richness distributions. Hence, additional sampling effort must be carried out in most Portuguese areas, to empirically validate results, develop more accurate predictive models and, lastly, map species richness distribution as accurately as possible. This additional collecting effort must cover those patches of the environmental variability not included in the adequately-sampled areas.

The present Scarabaeinae richness map (Figure 6) is consistent with the patterns found by Martín-Piera et al. (1992), Lobo et al. (1997) and Lobo and Martín-Piera (1999) for all the Iberian Peninsula, in which the Iberian central system hosts most of the diversity hotspots for this family. The remarkable hotspot of Central Portugal (squares 20 and 24) corresponds to the foothills of this mountain range, and it is also the geographic area where the residuals from the model are smaller (Figure 5), so it is likely to be the area where greater Scarabaeinae species richness occurs. Eurosiberian and Mediterranean fauna co-occurrence could feasibly explain this hotspot.

A similar faunistic co-occurrence pattern has been detected for amphibians and reptiles in the Serra da Sao Mamede (Sá-Sousa 2000), which corresponds to a localized Scarabaeinae hotspot (UTM50 number 35; see Figure 6). Eurosiberian and Mediterranean intergradation conditions in this mountain range, the only one south of the river Tejo that exceeds 1000 m altitude, favour both the sympatric occurrence of sister taxa from both zones and a rich diversity of species (Sá-Sousa 2000). Almost all the rest of the flat southern basin of the Tejo (around 300 m altitude) is characterised by low richness scores and low residuals. This lowland species poverty also appears in north-western Portugal, the only Eurosiberian low altitude region. However, the expected richness scores from the south and south-eastern squares might be higher than that predicted because of their high residuals.

There is a pressing need to investigate the distribution of species richness, rarity, endemism and phylogenetic diversity in order to maximize the biodiversity preserved with the funds available (Solbrig 1991; Gaston 1994; Williams and Humphries 1994). As the investment both in time and manpower necessary to obtain complete inventories is still too huge in most Mediterranean areas, forecast maps could be quite useful in conservation policies. The study presented herein points out that, by means of geographically referenced environmental and land use variables, as well as an adequate sampling effort directed toward small areas of each physioclimatic

subregion, it should be possible to recover reliable species richness scores for broad territories which remain unexplored.

Acknowledgements

We are grateful to The European Environment Agency, and Dr W. Cramer for making the Land Use/Land Cover and climatic data, respectively available; to Miguel Araújo, whose comments improved the manuscript, and to James Cerne, who helped us with the English. This paper was supported by the project 'Patrones de diversidad geográfica en insectos: una aproximación a la evaluación de áreas prioritarias de conservación en España Central' (Spanish D.G.I.C.Y.T.; grant: PB97-1149), and also by a PhD Museo Nacional de Ciencias Naturales/C.S.I.C./Comunidad de Madrid and a postdoc Comunidad de Madrid grants.

References

- Araújo MB (1999) Distribution patterns of biodiversity and the design of a representative reserve network in Portugal. *Diversity and Distributions* 5: 151–163
- Austin MP (1980) Searching for a model for use in vegetation analysis. *Vegetatio* 42: 11–21
- Austin MP, Pausas JG and Nicholls AO (1996) Patterns of tree species richness in relation to environment in southern–eastern New South Wales, Australia. *Australian Journal of Ecology* 21: 154–164
- Balleto E and Cassalle A (1991) Mediterranean Insect Conservation. In: Collins NM and Thomas JA (eds) *The Conservation of Insects and Its Habitats*, pp 121–140. Academic Press, London
- Barbero E, Palestrini C and Rolando A (1999) Dung beetle conservation: effects of habitat and resource selection (Coleoptera: Scarabaeoidea). *Journal of Insect Conservation*. 3: 75–84
- Biström O, Silfverberg H and Rütanen I (1991) Abundance and distribution of coprophilous Histerini (Histeridae) and *Onthophagus* and *Aphodius* (Scarabaeidae) in Finland (Coleoptera). *Entomologica Fennica* 2: 53–66
- Cliff AD and Ord JK (1981) *Spatial Processes. Models and Applications*. Pion Limited, London
- Colwell RK (1997) EstimateS, Statistical estimation of species richness and shared species from samples. Version 5. User's guide and application published at <http://viceroy.eeb.uconn.edu/estimates>
- Colwell RK and Coddington JA (1995) Estimating terrestrial biodiversity through extrapolation. In: Hawksworth DL (ed) *Biodiversity, Measurement and Estimation*, pp 101–118. Chapman & Hall, London
- Crawley MJ (1993) *GLIM for Ecologists*. Blackwell Scientific Publications, Oxford
- Daniel C and Wood FS (1980) *Fitting Equations to Data. Computer Analysis of Multifactor Data*. John Wiley & Sons, New York
- Derksen S and Keselman HJ (1992) Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology* 45: 265–282
- Dobson A (1999) *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC, London
- European Environment Agency (1996) *Natural Resources*. CD-Rom, European Environment Agency, Copenhagen, Denmark
- Fagan WF and Kareiva PM (1997) Using compiled species list to make biodiversity comparisons among regions: a test case using Oregon butterflies. *Biological Conservation* 80: 249–259
- Fincher GT (1981) The potential value of dung beetles in pasture ecosystems. *Journal of the Georgia Entomological Society* 16: 316–333

- Gaston KJ (1994) *Rarity*. Chapman & Hall, London
- Gaston KJ (1996) Species richness: measure and measurement. In: Gaston KJ (ed) *Biodiversity. A Biology of Numbers and Difference*, pp 77–113. Blackwell Science, Oxford
- Griffiths GH, Eversham BC and Roy DB (1999) Integrating species and habitat data for nature conservation in Great Britain: data sources and methods. *Global Ecology and Biogeography* 8: 329–345
- Heikkinen RK and Neuvonen S (1997) Species richness of vascular plants in the subarctic landscape of northern Finland: modelling relationships to the environment. *Biodiversity and Conservation* 6: 1181–1201
- Hortal-Muñoz J, Martín-Piera F and Lobo JM (2000) Dung beetle geographic diversity variation along a western Iberian latitudinal transect (Coleoptera: Scarabaeidae). *Annals of the Entomological Society of America* 93(2): 235–243
- Idrisi 2.0 (1998) *Geographic Information System*. Clark Labs, Clark University, Worcester, Massachusetts
- Instituto Geográfico nacional (1995) *Atlas nacional de España*. Tomos I y II. Centro Nacional de Información Geográfica, Madrid, Spain
- Johnson C (1962) The scarabaeoid (Coleoptera) fauna of Lancashire and Cheshire and its apparent changes over the last 100 years. *The Entomologist* 95: 153–165
- Klein BC (1989) Effects of forest fragmentation on dung and carrion beetle communities in Central Amazonia. *Ecology* 70: 1715–1725
- Ladeiro JM (1950) Os lamellicórnios Portugueses do Museu Zoológico de Universidade da Coimbra. *Memórias dos Estudos del Museu Zoológico da Universidade de Coimbra* 196: 1–23
- Lawton JH, Prendergast JR and Eversham BC (1994) The numbers and spatial distributions of species: analyses of British data. In: Forey PL, Humphries CJ and Vane-Wright RI (eds) *Systematics and Conservation Evaluation, Systematics Association Special Volume No. 50*, pp 177–195. Clarendon Press, Oxford
- Leclerc J, Gaspar C, Marchal JL, Verstraeten C and Wonville C (1980) Analyse des 1600 premières cartes de l'Atlas provisoire des insectes de Belgique, et première liste rouge d'insectes menacés dans la faune belge. *Notes Fauniques de Gembloux* 4: 1–104
- Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 74: 1659–1673
- Legendre P and Legendre L (1998) *Numerical Ecology*, 2nd edn. Elsevier, Amsterdam
- Legendre P and Vaudor P (1991) *The R Package: multidimensional analysis, spatial analysis*. Département de sciences biologiques, Université de Montréal, Montréal
- Lobo JM (2001) Decline of roller dung beetle (Scarabaeinae) populations in the Iberian Peninsula during the 20th century. *Biological Conservation* 97(1): 43–50
- Lobo JM 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: 31–37
- Lobo JM and Martín-Piera F (1999) Between-group differences in the Iberian dung beetle species-area relationship (Coleoptera: Scarabaeidae). *Acta Oecologica* 20(6): 587–597
- Lobo JM, Sanmartín I and Martín-Piera F (1997) Diversity and spatial turnover of dung beetle (Coleoptera: Scarabaeoidea) communities in a protected area of South Europe (Doñana National Park, Huelva, Spain). *Elytron* 11: 71–88
- Lumaret JP (1986) Toxicité de certains helminthocides vis-a-vis des insectes coprophages et conséquences sur la disparition des excréments de la surface du sol. *Acta Oecologica, Oecología Applicata* 7: 313–324
- Lumaret JP (1990) *Atlas des Coléoptères Scarabéides Laparosticti de France*. Collection inventaires de faune et de flore, fasc. 1. Secrétariat Faune-Flore/Museum National d'Histoire Naturelle, Paris
- Lumaret JP (1994) La conservation de l'entomofaune dans les aires naturelles protégées. In: Jiménez-Peydró R and Marcos-García MA (eds) *Environmental Management and Arthropod Conservation*, pp 57–65. Asociación Española de Entomología, Valencia
- Lumaret JP and Kirk AA (1991) South temperate dung beetles. In: Hanski I and Cambefort Y (eds) *Dung Beetle Ecology*, pp 97–115. Princeton University Press, Princeton, New Jersey
- Lumaret JP, Galante E, Lumbreras C, Mena J, Bertrand M, Bernal JL, Cooper JL, Kadiri N and Crowe D (1993) Field effects of ivermectin residues on dung beetles. *Journal of Applied Ecology* 30: 428–436
- Luna de Carvalho E (1950) Contribuições para o inventario da fauna lusitânica. *Insecta. Aditamento ao inventario dos coleópteros do dr. A. F. de Seabra. Memórias e Estudos do Museu Zoológico da Universidade de Coimbra* 203: 1–24

- Mac Nally R (2000) Regression and model building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – ‘predictive’ and ‘explanatory’ models. *Biodiversity and Conservation* 9: 655–671
- Magurran AE (1988) *Ecological Diversity and Its Measurement*. Princeton University Press, Princeton, New Jersey
- Margules CR, Nicholls AO and Austin MP (1987) Diversity of *Eucalyptus* species predicted by a multi-variable environment gradient. *Oecologia* 71: 229–232
- Martín-Piera F (1997) Apuntes sobre biodiversidad y conservación de insectos: dilemas, ficciones y ‘soluciones’. *Boletín de la Sociedad entomológica Aragonesa* 20: 1–31
- Martín-Piera F (2000) Familia Scarabaeidae. In: Martín-Piera F and López-Colón JI (eds) *Coleoptera, Scarabaeoidea I. Fauna Ibérica Vol. 14* (Ramos MA et al. (eds)). Museo Nacional de Ciencias Naturales, CSIC, Madrid
- Martín-Piera F and Lobo JM (1995) Diversity and ecological role of dung beetles in Iberian grassland biomes. In: McCracken DI, Bignal EM and Wenlock SE (eds) *Farming on the Edge: The Nature of Traditional Farmland in Europe*, pp 147–153. Joint Nature Conservation Committee, Peterborough
- Martín-Piera F, Veiga CM and Lobo JM (1992) Ecology and biogeography of dung beetle communities (Coleoptera: Scarabaeoidea) in an Iberian mountain range. *Journal of Biogeography* 19: 677–691
- McCullagh P and Nelder JA (1989) *Generalized Linear Models*, 2nd edn. Chapman & Hall, London
- Melic A and Blasco-Zumeta J (eds) (1999) *Manifiesto Científico por los Monegros*. *Boletín de la Sociedad Entomológica Aragonesa* 24. Volumen Monográfico, 266 pp
- Miessen G (1997) Contribution à l’étude du genre *Onthophagus* en Belgique (Coleoptera: Scarabaeidae). *Bulletin des Annales de la Société royal belge d’Entomologie* 133: 45–70
- Nicholls AO (1989) How to make biological surveys go further with generalised linear models. *Biological Conservation* 50: 51–75
- Nicholls AO (1991) Examples of the use of Generalised Linear Models in analysis of survey data for conservation evaluation. In: Margules CR and Austin MP (eds) *Nature Conservation: Cost-Effective Biological Surveys and Data Analysis*, pp 54–63. CSIRO, Canberra
- Odland J (1988) *Spatial Autocorrelation*. Sage Publications, Newbury Park
- Oliveira MP (1894) *Catalogue des insectes de Portugal*. *Coleoptères*. Imprensa da Universidade de Coimbra, Coimbra
- Pascual MA and Iribarne OO (1993) How good are empirical predictions of natural mortality? *Fisheries Research* 16: 17–24
- Prendergast JR, Wood SN, Lawton JH and Eversham BC (1993) Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiversity Letters* 1: 39–53
- Preudhomme de Borre A (1886) Liste des Lamellicornes Laparoscitiques recueillies par feu Camille Van Volxem. *Annales de la Société Entomologique de Belgique* 30: 98–102
- Rougon D, Rougon C, Trichet J and Levieux J (1988) Enrichissement en matière organique d’un sol sahélien au Niger par les insectes coprophages (Coléoptères: Scarabaeidae): implications agronomiques. *Revue d’Ecologie et de Biologie du Sol* 25: 413–434
- Samways MJ (1993) A spatial and process sub-regional framework for insect and biodiversity conservation research and management. In: Gaston KJ, New TR and Samways MJ (eds) *Perspectives on Insect Conservation*, pp 1–28. Intercept, Andover, UK
- Samways MJ (1994) *Insect Conservation Biology*. Chapman & Hall, London
- Sá-Sousa P (2000) A predictive distribution model for the Iberian wall lizard (*Podarcis hispanicus*) in Portugal. *Herpetological Journal* 10: 1–11
- Seabra AF (1907) Estudos sobre os animaes uteis e nocivos á agricultura. IV. Esboço monographico sobre os Scarabeidos de Portugal (Coprini). Imprensa Nacional, Lisboa, Portugal
- Seabra AF (1909) Estudos sobre os animaes uteis e nocivos á agricultura. VI. Esboço monographico sobre os Scarabeidos de Portugal (Aphodiini e Hybosorini). Imprensa Nacional, Lisboa, Portugal
- Smith PA (1994) Autocorrelation in logistic regression modelling of species’ distributions. *Global Ecology and Biogeography Letters* 4: 47–61
- Soberón J and Llorente BJ (1993) The use of species accumulation functions for the prediction of species richness. *Conservation Biology* 7: 480–488

- Solbrig OT (1991) Biodiversity. A review of the scientific issues and a proposal for a collaborative program of research. MAB Digest 9, UNESCO
- STATISTICA for Windows (1999) Computer Program Manual. StatSoft, Inc. Tulsa, Oklahoma
- Tonteri T (1994) Species richness of boreal understorey forest vegetation in relation to site type and successional factors. *Annali Zoologici Fennici* 31: 53–60
- Väisänen R and Rassi P (1990) Abundance and distribution of *Geotrupes stercorarius* in Finland (Coleoptera: Scarabaeidae). *Entomologica Fennica* 1: 107–111
- Van Rensburg B, McGeoch MA, Chown SL and Van Jaarsveld AS (1999) Conservation of heterogeneity among dung beetles in the Maputaland centre of endemism, South Africa. *Biological Conservation* 88: 145–153
- Wessels KJ, Reyers B and Van Jaarsveld AS (2000) Incorporating land cover information into regional biodiversity assessments in South Africa. *Animal Conservation* 3: 67–79
- Williams PH and Humphries CJ (1994) Biodiversity, taxonomic relatedness and endemism in conservation. In: Forey PL, Humphries CJ and Vane-Wright RI (eds) *Systematics and Conservation Evaluation*, Systematics Association Special Volume No. 50, pp 177–195. Clarendon Press, Oxford

