



Modelling the species richness distribution of French dung beetles (Coleoptera, Scarabaeidae) and delimiting the predictive capacity of different groups of explanatory variables

JORGE M. LOBO*, JEAN-PIERRE LUMARET† and PIERRE JAY-ROBERT†

*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), C/José Gutiérrez Abascal 2, 28006 Madrid, Spain, E-mail: mcnj117@mncn.csic.es, and †Laboratoire de Zoogéographie, Université Paul Valéry Montpellier III, Route de Mende, F-34199 Montpellier Cedex 5, France

ABSTRACT

Aim To predict French Scarabaeidae dung beetle species richness distribution, and to determine the possible underlying causal factors.

Location The entire French territory has been studied by dividing it into 301 grid cells of 0.72×0.36 degrees.

Method Species richness distribution was predicted using generalized linear models to relate the number of species with spatial, topographic and climate variables in grid squares previously identified as well sampled ($n = 66$). The predictive function includes the curvilinear relationship between variables, interaction terms and the significant third-degree polynomial terms of latitude and longitude. The final model was validated by a jack-knife procedure. The underlying causal factors were investigated by partial regression analysis, decomposing the variation in species richness among spatial, topographic and climate type variables.

Results The final model accounts for 86.2% of total deviance, with a mean jack-knife predictive error of 17.7%. The species richness map obtained highlights the Mediterranean

as the region richest in species, and the less well-explored south-western region as also being species-rich. The largest fraction of variability (38%) in the number of species is accounted for by the combined effect of the three groups of explanatory variables. The spatially structured climate component explains 21% of variation, while the pure climate and pure spatial components explain 14% and 11%, respectively. The effect of topography was negligible.

Conclusions Delimiting the adequately inventoried areas and elaborating forecasting models using simple environmental variables can rapidly produce an estimate of the species richness distribution. Scarabaeidae species richness distribution seems to be mainly influenced by temperature. Minimum mean temperature is the most influential variable on a local scale, while maximum and mean temperature are the most important spatially structured variables. We suggest that species richness variation is mainly conditioned by the failure of many species to go beyond determined temperature range limits.

Key words dung beetles, France, modelling distributions, species richness determinants, species richness distribution, variance partitioning.

INTRODUCTION

As there tends to be a correlation between the number of species and many other biodiversity measures (Gaston, 1996), an accurate description of the geographical distribution of

species richness is critical for biologists (Miller, 1994) and can aid in the elaboration of precise conservation strategies (Margules & Usher, 1981; Haila & Kouki, 1994). Unfortunately, even in countries with a tradition of naturalist studies, taxonomic knowledge is not uniform across regions, and the pattern of species richness is influenced frequently by sampling effort (Prendergast *et al.*, 1993; Gaston, 1996).

In order to quantify the species richness distribution of a taxonomic group one could identify well-inventoried regions and increase the recording effort in poorly inventoried

Correspondence: Jorge M. Lobo, Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), C/José Gutiérrez Abascal 2, 28006 Madrid, Spain. E-mail: mcnj117@mncn.csic.es

regions, both time-consuming and slow. However, if there are enough biological data, geographical gaps in information can be filled through the use of environmental surrogate measures, with predictive species richness models based on biological survey data and mapped environmental variables (Margules *et al.*, 1987; Nicholls, 1989; Walker, 1990; Buckland & Elston, 1993; Carpenter *et al.*, 1993; Iverson & Prasad, 1998; Hortal *et al.*, 2001; Pearce & Ferrier, 2001; Lobo & Martín-Piera, 2002). This strategy would yield results more quickly, while additional fieldwork sampling could empirically validate the proposed model of species richness distribution.

Explanatory environmental variables are usually correlated, and thus collinear, which may bias model parameter estimation. If the aim is not to make ecological inferences, but rather to forecast, maximizing the explained variation of the data, collinearity of the explanatory variables is not a concern (Legendre & Legendre, 1998). However, when the influence of these same explanatory variables on the variation in species number is being investigated, collinearity, a source of ambiguity, can hamper the identification of the effects of different groups of variables. Caution is then necessary in the estimation of causal relationships underlying multivariate environmental and spatial correlations. To delimit the influence of different variables, one must first identify the factors correlated with species richness and, subsequently, establish probable causal influences by performing experiments or by assessing the correspondence between the correlations and a causal statistical model (Levin, 1992; Legendre & Legendre, 1998; Shipley, 1999). As manipulative experiments and causal modelling are almost impossible approximations in the case of coarser spatial scale studies, factors correlated with the number of species provide the starting point for the identification of factors that might possibly explain geographical species richness patterns (see Richerson & Lum, 1980; Austin *et al.*, 1984; Owen, 1989; Tonteri, 1994; Birks, 1996; Heikkinen & Neuvonen, 1997; Fraser, 1998; Qian, 1998; Wohlgemuth, 1998; Kerr & Packer, 1999; Lobo *et al.*, 2001). Explanatory variables have been analysed jointly, but rarely have combined effects or the independent influence of each variable been included. Partial regression analysis facilitates the examination of species richness dependence on different variable groups. Although decomposing the total variation of the dependent variable into environmental and spatial components would not necessarily identify causal factors, it would help to identify the component overlap, and facilitate the formulation of hypotheses about the processes that may have generated the observed patterns (Borcard *et al.*, 1992; Legendre, 1993; Pinel-Alloul *et al.*, 1995; Qinghong & Bråkenhielm, 1995; Heikkinen & Birks, 1996; Anderson & Gribble, 1998; Legendre & Legendre, 1998; Lobo *et al.*, 2001).

Thanks to a long-standing naturalist tradition, the species richness distribution of some insect groups is relatively well

known in France. Scarabaeidae dung beetles (*sensu* Scholtz, 1990), which have been studied for a long time (Paulian & Baraud, 1982; Baraud, 1992), constitute one such group (Lumaret, 1990; Lumaret *et al.*, 1996). Using the French dung beetle database (37 300 records), the asymptotic relationship between the number of species and the number of records in 0.72×0.36 degree grid squares (about 2230 km² each) was analysed for each one of the five biogeographic subprovinces. These asymptotic curves allow one to obtain an estimate of the number of species present in the squares of each subregion. Subsequently, the number of database records required to obtain at least 70% of the estimated number of species was calculated for each subregion; squares with such a number of database records (66 squares, 22% of 301) were considered well inventoried (Lobo *et al.*, 1997). As no large inventory is planned at present, it would now be particularly useful to obtain a predictive map of dung beetle species richness distribution. The purpose of this paper is to elaborate such a predictive model for Scarabaeidae dung beetle species richness in France using environmental surrogates and a modelling technique. The species richness distribution obtained is mapped and discussed. In addition, the most important determinants of French dung beetle species richness were estimated by decomposing its variation into spatial, topographic and climate components in order to provide a preliminary understanding of the possible underlying causal factors that have led to the current species richness distribution.

MATERIALS AND METHODS

Selecting environmental variables

The French territory (France mainland and Corsica island) was divided into 301 0.72×0.36 degree grid squares (Fig. 1). For each square, six climate variables were utilized: mean annual temperature (TM), temperature variation (TV; difference between the highest and the lowest mean annual temperatures within the square), maximum mean temperature (T; temperature in July), minimum mean temperature (t; temperature in January), mean annual precipitation (P) and precipitation variation (PV; difference between the highest and the lowest mean annual precipitation within the square). Two topographic and two spatial variables were also used: mean elevation (E) and altitude range (AR; difference between the highest and the lowest elevations within the square), central latitude (LAT) and central longitude (LON). All variables are continuous. TV, PV and AR are measures of the spatial heterogeneity of squares. Both climate and topographic variables were extracted from digitized versions of maps from published Atlases (Serryn & Blasselle, 1970; Bessemoulin, 1974). Maps representing temperature (equidistance between isotherms = 1 °C), precipitation (equidistance = 200 mm) and altitude (equidistance = 300 m) were

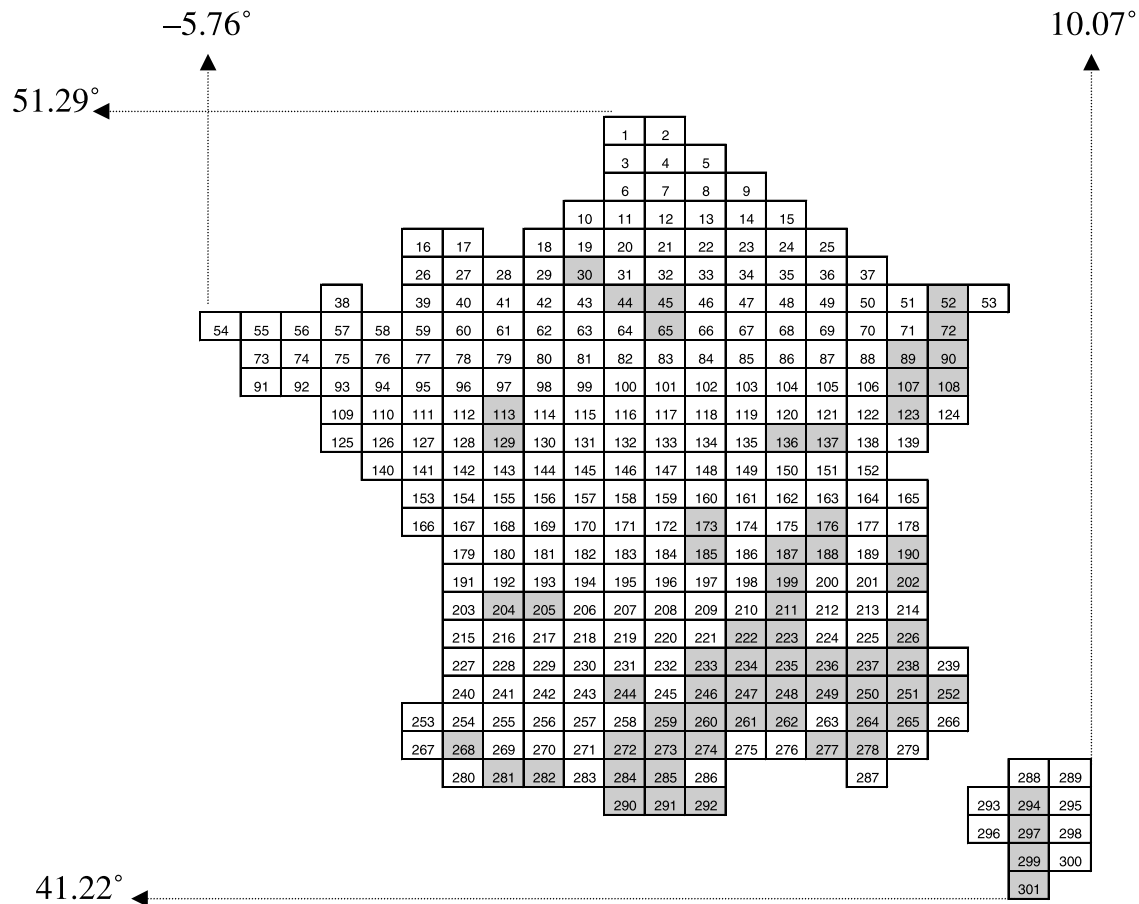


Fig. 1 Continental France and Corsica Island divided into 301 squares of 0.72×0.36 degrees. Broken lines represent maximum and minimum latitude and longitude. Shaded squares are the 66 well-sampled grid squares (Lobo *et al.*, 1997).

superimposed on the grid square polygons by means of a geographical information system (GIS). For each square the mean values were calculated taking into account the area covered by each temperature, precipitation or altitude variable. Heterogeneity variables have been estimated considering the difference between the most extreme values in each square. We assumed that on the broad spatial scale analysed the environmental data for each square obtained by this method are a fairly good approximation to the real data.

The forecasting models

Generalized linear models (GLM; McCullagh & Nelder, 1989) have been used to summarize the relationship between Scarabaeidae dung beetle species number and the above-mentioned explanatory variables in order to obtain an appropriate forecasting model (e.g. Nicholls, 1989; Austin *et al.*, 1990, 1996; Heikkinen & Neuvonen, 1997). GLM allow a range of model specification distributions other than the normal distribution for the random component,

also avoiding the constraint imposed by the assumption of linearity between the dependent and independent variables. In the present case, a Poisson error distribution for the number of dung beetle species was assumed. It was linked to the set of predictor variables via a logarithmic link function (see Crawley, 1993).

As a first step, the total number of Scarabaeinae in the 66 well-surveyed 0.72×0.36 degree grid squares (Lobo *et al.*, 1997) was related separately to each environmental variable. Environmental variables were standardized to eliminate the effect of differences in measurement scale. As the relationship between the species number and environmental variables is often curvilinear (Austin, 1980), either a linear, a quadratic or a cubic function of each environmental variable was selected, comparing their reduction in deviance with a full model in which the number of parameters was equal to the total number of observations (Dobson, 1990). The goodness-of-fit for the competing functions was measured by the deviance statistic and the change in deviance tested by a *F*-ratio test (McCullagh & Nelder, 1989; Dobson, 1990) with a 5%

significance level. The percentage of explained deviance was calculated for each model (see Dobson, 1990).

Next, a forward stepwise procedure was used to enter the variables into the model (see Nicholls, 1989; Austin *et al.*, 1996; Heikkinen & Neuvonen, 1997). First, the linear, quadratic or cubic function of the variable that accounts for the most important change in deviance was selected for inclusion. Next, all the remaining variables were tested for significance, adding them sequentially one by one. The procedure was repeated iteratively until no more statistically significant explanatory variables remained ($P \leq 0.05$). After each inclusion, the significance of the terms previously selected was also tested, submitting the new model to a backward selection procedure in order to exclude the terms that had become non-significant.

Margules *et al.* (1987) demonstrated that interactions between variables often provide a higher predictive power than the same variables separately. Therefore, after testing all the environmental variables for inclusion, the significance of each pairwise interaction term (between explanatory variables) was tested (45 interaction terms). All interaction terms were added sequentially one by one to the model and the interaction term that accounts for the most important change in deviance was selected for inclusion. The procedure was repeated iteratively until no more significant interaction terms remained.

Lastly, the spatial variables were included in the model because spatial structures can incorporate the effects caused by other historical, biotic or environmental variables not taken into consideration (Legendre & Legendre, 1998). To ensure that complex spatial structures were recovered, the third-degree polynomial equation of the central latitude (LAT) and longitude (LON) of each square was used as an independent variable: $b_1\text{LAT} + b_2\text{LON} + b_3\text{LAT}^2 + b_4\text{LAT} \times \text{LON} + b_5\text{LON}^2 + b_6\text{LAT}^3 + b_7\text{LAT}^2 \times \text{LON} + b_8\text{LAT} \times \text{LON}^2 + b_9\text{LON}^3$ (trend surface analysis; see Legendre, 1993). Latitude and longitude were centred on their respective means. We performed a backward stepwise regression with the nine terms of the equation as predictor variables and the number of dung beetle species as dependent variable in order to remove the non-significant spatial terms (Legendre & Legendre, 1998). To test if they account for a significant change in deviance, the significant spatial terms ($P \leq 0.05$) were retained and included in the previous model. The STATISTICA (1995) package was used for all statistical computations.

Statistical considerations, residuals and model validation

Covariance between explanatory variables and their spatial autocorrelation are currently considered to be the main problems when modelling species richness (Margules *et al.*, 1987; Austin *et al.*, 1996; Gaston, 1996). When the goal is to

forecast the value of a response variable, collinearity in explanatory variables is not a concern because the aim is to maximize the predictive power of the function and not to correctly estimate model parameters (Legendre & Legendre, 1998). On the other hand, the autocorrelation of variables considered in a regression analysis implies the spatial dependence of observations, as the observed values of variables at any given locality are influenced by the values at neighbouring localities. As most environmental variables are autocorrelated on a geographical scale, assessing the relationships between variables is complicated by the ineffectiveness of classical statistical tools. The broad-scale spatial structure of spatially autocorrelated variables could be removed, but consequently the impact of those spatially structured factors would be reduced along with the forecasting ability of the model, so autocorrelated data, an important source of information, must be taken into account to improve the predictive power of models (Smith, 1994; Legendre & Legendre, 1998). When autocorrelated variables are used, it is necessary to examine the spatial independence of errors (Cliff & Ord, 1981; Odland, 1988). To check if residuals are autocorrelated over space, the autocorrelation for the residuals from the final model assuming independence was calculated by Moran's I spatial autocorrelation statistic, testing for significance with the Bonferroni-corrected significance level (Cliff & Ord, 1981; Sawada, 1999). Spatial autocorrelation in the residuals usually indicates that the model must be enlarged in order to accommodate some spatially structured variables not taken into account by the model (Cliff & Ord, 1981; Odland, 1988). Hence, the final addition of complex spatial terms (the third degree polynomial of latitude and longitude) to the environmental variable and interaction term model could be expected to take into account those spatially structured variables not previously considered.

The adequacy of the final model was improved by detecting outliers that disproportionately influence the fit of the model. Therefore, the standard errors of coefficients were examined, the residuals vs. fitted values were plotted and the residuals were plotted in a normal probability plot (Nicholls, 1989; Dobson, 1990). Also examined was the observation's leverage, 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 high leverage value, from outlier or other observation, was examined to ascertain if it was due to erroneous data, or to noteworthy environmental variability. In the former case, the observation must be deleted; in the latter case the observation may remain in the model in order to include as much environmental heterogeneity as possible (Nicholls, 1989).

To validate the final model, a jack-knife procedure was applied. Using the dataset of the 66 well-sampled grid squares, the model was recalculated 66 times, leaving out one square in turn. Each one of the regression models

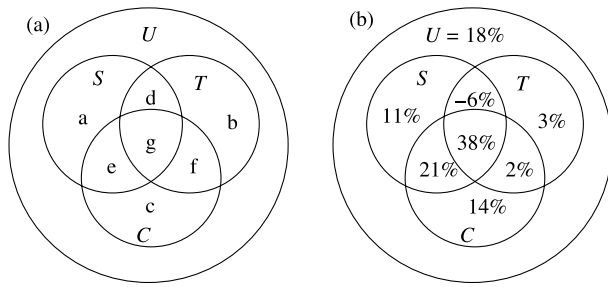


Fig. 2 (a) Diagram showing the variation in the dependent variable (number of species) due to three groups of explanatory variables: S (space), T (topography), and C (climate), and U is the unexplained variation; a, b and c are the pure effects of spatial variation, topography, and climate, respectively; while d, e and f are the combined variation due to the joint effect of space and climate, the joint effect of space and topography, and the joint effect of topography and climate, respectively; g is the combined variation due to the joint effect of the three variables. (b) Variation partitioning in the number of French Scarabaeidae species between the three set of explanatory variables. The decomposition of the variation has been carried out by partial regression analysis (Legendre & Legendre, 1998) using generalized linear models (Crawley, 1993) and the percentage of total explained deviance as the measure of explained variation.

based on the $n-1$ grid squares was then applied to that excluded square in order to predict species richness scores for each one of the squares. The predictive power of the model was checked both by examining the correlation between the recorded and the jack-knife-predicted species richness, and by estimating the percentage error for each observation:

$$E_i = (|y_i - \hat{y}_i|/y_i) \times 100$$

where E_i is the predictive error for observation i ; y_i is the observed value; and \hat{y}_i the predicted value for y_i when this observation is excluded. The mean of the 66 estimates of error provides a measure of the prediction error associated with the model and data used (Pascual & Iribarne, 1993).

Variation partitioning

The significant terms of the explanatory variables selected in the first step of the regression analysis were analysed further to determine the comparative influence of spatial (S), topographic (T) and climate (C) variables in Scarabaeidae species richness variation. The total variation within the number of species was decomposed among the three types of explanatory variable (i.e. S, T and C) and the percentage of explained deviance calculated for eight different components (Fig. 2a): pure effect of spatial variation alone (a), pure effect of topography alone (b), pure effect of climate alone (c), combined variation due to the joint effect of spatial and

topographic components (d), combined variation due to the joint effect of spatial and climate components (e), combined variation due to the joint effect of topographic and climate components (f), combined variation due to the joint effect of the three components (g) and variation not explained by the independent variables included in the analysis (U). The decomposition of the variation in the species richness into the three sets of explanatory variables was carried out by means of a partial regression analysis (see Legendre & Legendre, 1998). As explanatory variables were not mutually independent, this approach can help determine the amount of variation overlap and clarify the influence of the variables considered (Borcard *et al.*, 1992; Qinghong & Bråkenhielm, 1995; Anderson & Gribble, 1998). In the process of variation decomposition the species richness (y) was regressed with the three types of variables together, which represent the total explained variation in the dataset ($a + b + c + d + e + f + g$ in Fig. 2a). Regressing y with each one of the explanatory variables yields the variation attributable separately to S ($a + d + e + g$), T ($b + d + f + g$), and C ($c + e + f + g$). Subsequently, residuals of the regression of S with T + C variables were calculated, and y regressed with these residuals in order to estimate the spatial variation in the species richness that is independent of any considered climate or topographic variables (pure spatial variation; a). Fractions b and c were estimated in the same way after computing the regression residuals of T with S + C, and the regression residuals of C with S + T, respectively. The remaining variation fractions were computed according to two sets of equations (Qinghong & Bråkenhielm, 1995; Anderson & Gribble, 1998), where:

$$\begin{aligned} d + e + g &= S - a \\ d + f + g &= T - b \\ e + f + g &= C - c \end{aligned}$$

and

$$\begin{aligned} d &= (S + T) - (e + f + g) - (a + b) \\ e &= (S + C) - (d + f + g) - (a + c) \\ f &= (T + C) - (d + e + g) - (b + c) \\ g &= (d + e + g) - d - e = (d + f + g) - d - f = (e + f + g) - e - f \end{aligned}$$

RESULTS

Using all the well-surveyed squares ($n = 66$), a first model was run. Six environmental variables were significant when tested separately as either linear, quadratic or cubic functions: mean elevation, altitude range, annual mean temperature, temperature variation, maximum mean temperature and minimum mean temperature (Table 1). The cubic function of annual mean temperature and the linear term of maximum mean temperature were the most important variables in the model (respectively, 38.6% and 38.2% of the total deviance). With regard to the spatial variables, the backward stepwise

Table 1 Deviance and change in the deviance from a full model for Scarabaeidae species number, considering explanatory variables in the 0.73×0.36 degrees grid squares of France with their respective code. The linear, quadratic or cubic functions of each variable have been selected if they account for a significant change in the deviance ($P < 0.05$). The sign column indicates the sign for each selected term of each function. Spatial variables were chosen after removing the non-significant terms of a third degree polynomial equation of latitude and longitude by a backward stepwise selection

Variables	Code	Selected terms	Deviance	d.f.	Change in deviance	F	Sign.
Full			226.37	65			
Longitude	LON	LAT + LAT ² + LAT ³ + LON					
Latitude	LAT	+ LON ³ + LAT ² * LON	83.24	59	143.03	101.44***	- + + - +
Mean elevation	E	E	218.37	64	8.00	2.34	
		E + E ²	184.79	63	33.59	11.45**	+ -
Altitude range	AR	AR	225.56	64	0.81	0.23	
		AR + AR ²	196.86	63	28.70	9.18**	+ -
Annual mean temperature	TM	TM	152.27	64	74.10	31.15***	+
		TM + TM ²	151.50	63	74.87	31.13***	+ -
		TM + TM ² + TM ³	138.97	62	87.40	38.99***	+ - +
Temperature variation	TV	TV	188.33	64	38.03	12.92***	+
Maximum mean temperature	T	T	139.93	64	86.43	39.53***	+
Minimum mean temperature	t	t	170.27	64	56.10	21.09***	+
Mean annual precipitation	P	P	213.42	64	12.95	3.88	
Precipitation variation	PV	PV	225.98	64	0.39	0.11	

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

regression of the third-degree polynomial terms of latitude and longitude selected the cubic function of latitude, the linear and cubic terms of longitude, and one interaction term (see Table 1).

As the cubic function of annual mean temperature accounted for the most important change in deviance, this variable was the first included in the model (Table 2a). In the second step, the quadratic function of mean elevation was selected but the backward procedure eliminated the linear term of annual mean temperature function (TM). Next, the linear terms of maximum mean temperature (T) and temperature variation (TV) were selected. The inclusion of this last term excluded mean elevation (E). None of the remaining environmental variables was significant when added to the model. After fitting the significant environmental variables, the importance of interaction terms added one by one was tested. The interaction between temperature variation (TV) and maximum mean temperature (T) accounted for the most significant change in deviance. The product of altitude range (AR) and mean annual precipitation (P) was the other significant interaction term that entered in the model, although its inclusion removed the quadratic term of mean elevation (Table 2a). Finally, the significant spatial terms (Table 1) were added to the model, but without a significant increase in the percentage of explained deviance.

Thus, model 1 was:

$$S = \exp(c + TM + TM^3 + T + TV + TV * T + AR * P)$$

where S is the total number of predicted dung beetle species and c is the intercept. This model explains 71.1% of the total deviance (see coefficients and standard errors in Table 2a).

A jack-knife test procedure on this model showed that the correlation between observed and jack-knife-predicted values was significant and positive (Pearson correlation coefficient: $r = 0.808$, $P < 0.001$; Spearman's rank correlation: $r_s = 0.824$, $P < 0.0001$). However, the mean jack-knife predictive error of the observations was relatively high (24.7%; 95% confidence interval from 18.1% to 31.4%). Moreover, six squares showed positive standard residuals higher than +1.5 (fitted values lower than observed) and two squares showed negative standard residuals lower than -1.5 (fitted values higher than observed, Fig. 3). As the leverage of these squares was small (between 0.06 and 0.16), these eight observations were deleted and the modelling performed again. The four first steps of the new model ($n = 58$) selected the same terms and variables as in the former model (Table 2b), but the linear term of mean elevation (E) was not deleted from the model in this case. The interaction between temperature variation (TV) and maximum mean temperature (T) was the only interaction term that accounted for a significant change in deviance. The addition of spatial terms to the model did not increase significantly the percentage of explained deviance (Table 2b).

Model 2:

$$S = \exp(2.81 - 0.14TM^2 + 0.06TM^3 + 0.15E - 0.07E^2 + 0.19T + 0.13TV - 0.06TV * T)$$

Table 2 Summary of the stepwise forward selection of variables to build the models for the Scarabaeidae species number in the 0.73×0.36 degrees grid squares. (a) With the 66 squares considered as well surveyed (Lobo *et al.*, 1997). (b) Eliminating the eight grid squares as outliers in the former model ($n = 58$). The change in deviance after including a term in the model has been tested by an F -ratio test with a probability < 0.05 . Variable codes as in Table 1

	d.f.	Deviance	Change in deviance	F	% of explained deviance	Coefficients	Standard error
(a) Full model	65	226.37					
Intercept						2.773	0.058
Step 1							
TM	64	152.27	74.10	31.15***	32.74		
+ TM ²	63	151.50	0.77	0.32	33.07	-0.084	0.031
+ TM ³	62	138.97	12.53	5.59*	38.61	0.042	0.012
Step 2							
+ E	61	138.85	0.12	0.05	38.66		
+ E ²	60	103.65	35.20	20.37***	54.21		
- TM	61	105.57	1.92	1.11	53.36		
Step 3							
+ T	60	84.27	21.30	15.17***	62.77	0.218	0.055
Step 4							
+ TV	59	74.27	10.00	7.94**	67.19	0.181	0.026
- E	60	76.18	1.91	1.50	66.34		
Step 5							
+ TV \times T	59	68.86	7.32	6.27*	69.58	-0.086	0.025
Step 6							
+ AR \times P	58	64.10	4.76	4.31*	71.68	-0.064	0.022
- E ²	59	65.45	1.35	1.21	71.09		
(b) Full model	57	172.56					
Intercept						2.805	0.067
Step 1							
TM	56	108.30	64.26	33.23***	0.5216		
+ TM ²	55	106.49	1.81	0.94	0.5296	-0.141	0.037
+ TM ³	54	92.17	14.32	8.39**	0.5928	0.065	0.015
Step 2							
+ E	53	87.59	4.57	2.77	0.6130	0.154	0.079
+ E ²	52	54.85	32.74	31.04***	0.7577	-0.074	0.032
- TM	53	56.84	1.99	1.85	0.7489		
Step 3							
+ T	52	45.92	10.92	12.36***	0.7971	0.190	0.063
Step 4							
+ TV	51	36.52	9.40	13.13***	0.8387	0.127	0.034
Step 5							
+ TV \times T	50	31.30	5.22	8.35**	0.8617	-0.063	0.027

explained a higher percentage of the total deviance than model 1 (86.2% and 71.1%, respectively) (see full coefficients and standard errors in Table 2b).

The jack-knife test for model 2 showed a higher correlation than for model 1 between observed and jack-knife-predicted values (Pearson's correlation coefficient: $r = 0.868$, $P < 0.0001$; Spearman's rank correlation: $r_s = 0.875$, $P < 0.0001$). The mean jack-knife predictive error of observations was reduced to 17.7% (95% confidence interval, ranging between 13.6% and 21.7%) and the examination of residuals did not show especially aberrant features or spatial autocorrelation.

None of the Moran's I -values in the different lag-classes were significant at a 0.05 significance level with the Bonferroni correction. These results show that the model is able to predict reasonably well the dung beetle species richness in the cells not yet well-enough sampled.

When applied to the entire territory of continental France and Corsica, model 2 showed a clear geographical distribution of species richness (Fig. 4). The richest area (≥ 20 species per square) corresponded to the Mediterranean bioclimatic region and its surroundings and included Corsica (south-eastern France). Roughly, the remaining part of continental

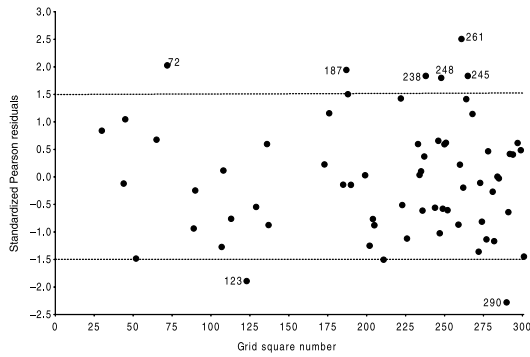


Fig. 3 Standardized Pearson residuals scores of the 66 well-sampled grid squares (see Fig. 1) used to build the forecasting model described in Table 2a. Square numbers are from Fig. 1. Discontinuous lines delimit eight squares with positive residuals higher than 1.5 or with negative residuals less than -1.5.

France could be divided into two main areas: a southern half, characterized by intermediate richness (15–19 species per square), and a poorer northern half (8–14 species per square). Inside the southern zone, the large mountain ranges of the Alps and Massif Central showed a characteristically low diversity (4–9 species per square), while the Pyrenean region (south-western boundary) was intermediate (10–11 species per square).

The amount of deviance explained by all the statistically significant environmental and spatial variables (Table 1) is 82%. The decomposition of the variation shows that the largest fraction of the variability (38%) in the number of species is accounted for by the combined effect of the three groups of explanatory variables (spatial, climate and topographic), while the spatially structured climate component explains 21% of total variation in species richness. The variation that might be attributed both to the climate and the spatial components alone are similar and moderate (14% and 11%, respectively). However, the pure effect from topography was negligible. The joint influence of topography + climate and topography + space is also unimportant (Fig. 2b). The

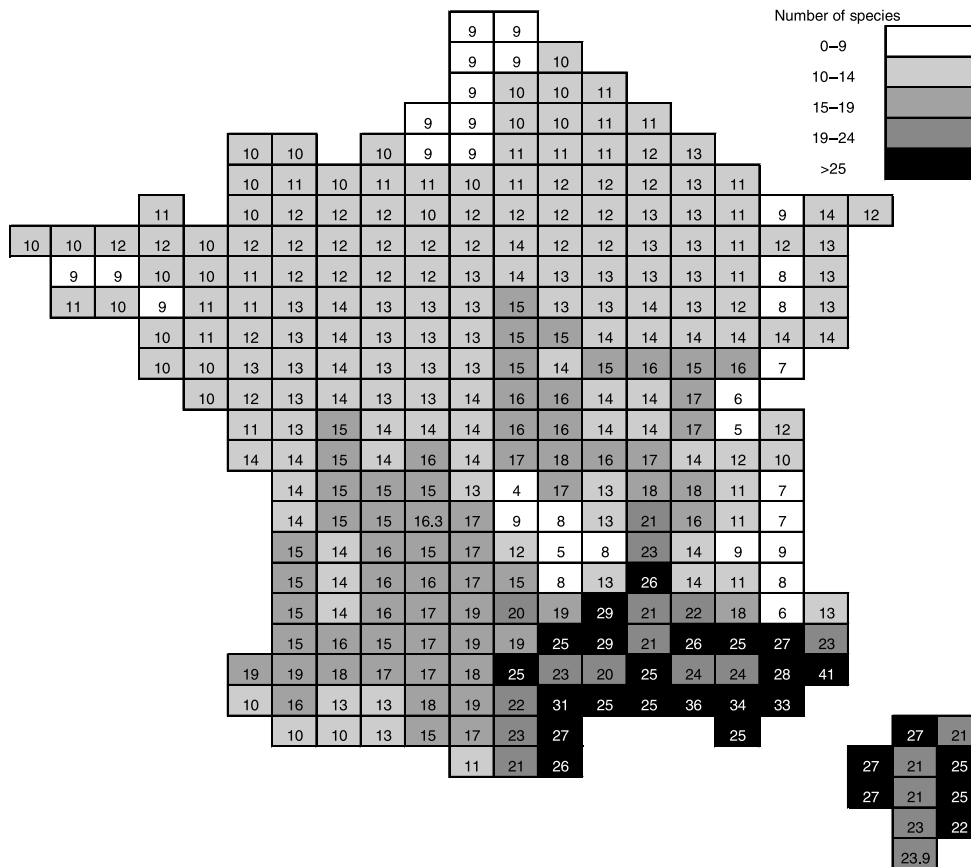


Fig. 4 Predicted Scarabaeidae species richness in France and Corsica. The species richness scores were forecasted by applying the final multiple regression model (see Table 2b) where the data of eight outliers were omitted.

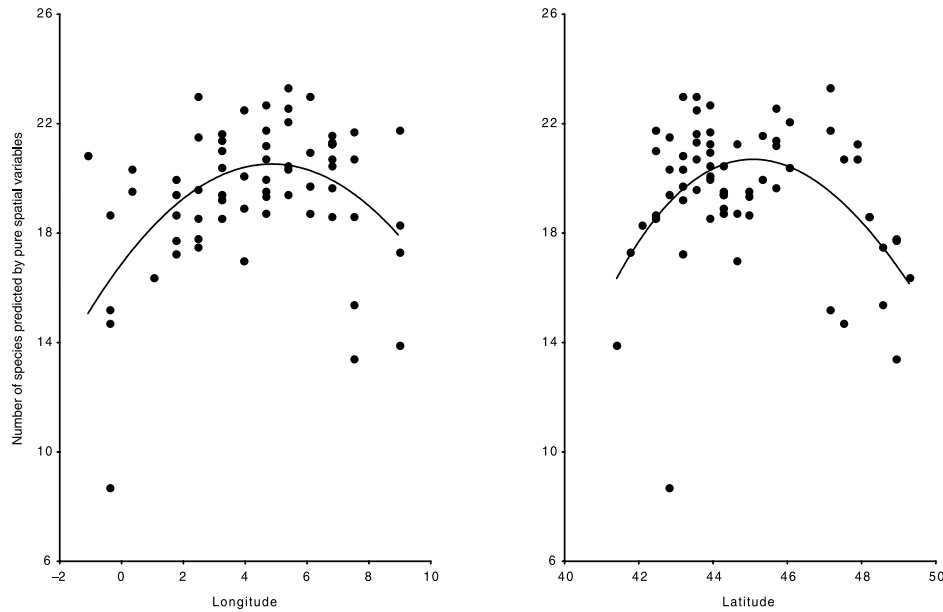


Fig. 5 Relationships between the number of dung beetle species predicted by 'pure' spatial variables and longitude and latitude. Both latitude and longitude were submitted to a backward stepwise regression with significant topographic and climate variables as independent variables, retaining residuals as a measure of the purely spatial contribution. The number of species was regressed with these residuals calculating the species richness predicted by these pure spatial variables.

negative value of the last combination of variables is due probably to their having opposite effects on species richness (Legendre & Legendre, 1998). To ascertain the contribution of temperature variables to the purely climate component, the large-scale spatial structure of these temperature-related variables was subtracted by performing a backward stepwise regression for each variable with the third-degree polynomial of the latitude and longitude of each grid cell. The retained residuals of these regressions are a good measure of the predictive power of purely temperature-related variables. This analysis yielded minimum mean temperature as the only significant variable ($P < 0.05$), with a positive sign. Similarly, to examine the contribution of purely spatial variables, both latitude and longitude were submitted to a backward stepwise regression with significant topographic and climate variables as independent variables. Significant in this analysis were the linear term of longitude, the quadratic term of latitude, plus two interaction terms ($\text{longitude}^2 \times \text{latitude}$ and $\text{latitude}^2 \times \text{longitude}$). The relationship between the number of species predicted by these pure spatial variables and latitude or longitude shows that the maximum number of species occurs at intermediate latitudes and longitudes (Fig. 5).

DISCUSSION

Little accurate geographical information on species richness is available, except for some groups of vertebrates, higher plants

and very few groups of invertebrates (Williams & Gaston, 1994). The spatial pattern of species richness of French dung beetles is not described adequately by the available data. Biogeographic diversity patterns for the majority of insect groups reflect the distribution of the areas investigated by entomologists (Dennis & Hardy, 1999). With so much sampling still to be carried out for a supposedly 'well-known' insect group, how can the spatial distribution of biodiversity be described within a reasonable period of time? Sampling all poorly explored or unexplored territories would be expensive and impractical. Our results demonstrate that delimiting the adequately sampled areas and elaborating predictive models using simple environmental variables is quicker and cheaper. By such means we can rapidly produce an estimate of the spatial distribution of species richness obtaining testable geographical patterns. Moreover, the predictive figures could be validated and improved continually by adding new empirical data.

The inventory of the French Scarabaeidae species was completed in 1990 (Lumaret, 1990). This work gave rise to new studies and regional surveys (Gangloff, 1991), but no large-scale survey is planned at present. Consequently, the present model of species richness distribution provides a good substitute for information that could not be provided in the coming years. The model also helps to focus sampling efforts on areas that were doubtless inadequately inventoried.

The model showed that the richest area was also the best prospected (south-eastern France). For more than a century, French and European entomologists have been collecting beetles in that Mediterranean area, known for its rich entomofauna. It is now very unlikely that a new Scarabaeidae species could be encountered in this area. However, the model showed that south-western France, between the Atlantic and the Massif Central, although less well explored, is also rich (Fig. 1). This region should be targeted for further sampling to gain greater knowledge of the French Scarabaeidae fauna. The geographical Scarabaeidae species richness pattern provided by the model is relatively simple, and reflects the latitude gradient of the bioclimate transition zones. Southward, the species richness is similar to the richness observed in the Iberian Peninsula (Lobo & Martín-Piera, 1999). The French Mediterranean region is home to 11 genera (seven tribes) found in two subfamilies: Scarabaeinae (three) and Coprinae (eight) (Lumaret, 1990). Conversely, northward, Scarabaeinae are represented only by one species (*Sisyphus schaefferi* (L.)) and Coprinae are mainly represented by the genus *Onthophagus* Latreille (Lumaret, 1990). Within the latitude gradient, the influence of the Alps is reflected in the model by a marked decrease in species richness in these mountains. No similar drop in richness was observed in the Pyrenees mountains (south-western boundary).

Detecting diversity patterns and identifying probable causal variables and processes is important for management (Levin, 1992). Unfortunately, there are many candidate factors (Hengeveld, 1990; Huston, 1994; Whittaker *et al.*, 2001). The forecasting method used maximizes the predictive power of the model but its dependence on established correlations should not be taken to imply causal relationships (Legendre & Legendre, 1998; MacNally, 2000). The variation-partitioning method selects factors of greatest probable influence from among those analysed. The largest fraction of variability is accounted for by the combined effect of spatial and climate or topographic variables. The majority of the variation in French dung beetle species richness can be attributed indifferently to these three groups of explanatory variables. However, the signs of the significant explanatory variables (Table 1) show that higher species richness occurs in the southern and eastern cells with lower altitudes and higher temperatures, results in agreement with the recognized environmental adaptations of Scarabaeidae dung beetles. In fact, unlike the other dung beetle families (Aphodiidae and Geotrupidae), Scarabaeidae species predominate in tropical, lowland and southern European dung beetle assemblages (Hanski, 1986; Cambefort, 1991; Lumaret & Kirk, 1991; Lobo & Halfpeter, 2000).

Topography, alone or combined with other variables, contributes little to, and probably does not directly influence, species richness. Thus, we suggest that the recognized

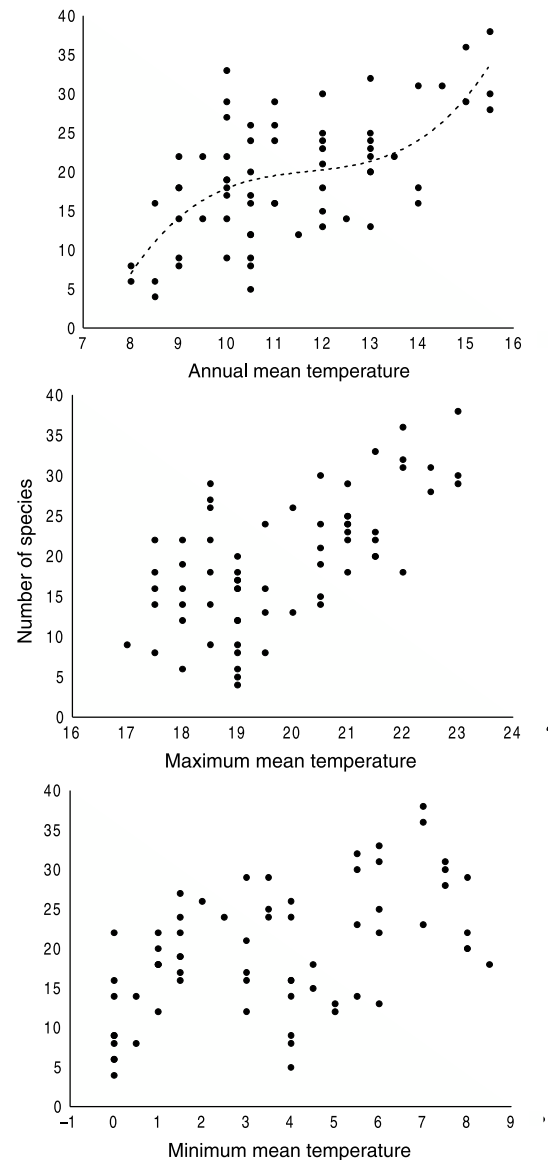


Fig. 6 Relationship between the number of French Scarabaeidae species in the 66 well-sampled grid squares and annual, maximum, and minimum mean temperature in these squares. The line is the third degree polynomial fit.

negative effect of altitude on French Scarabaeidae dung beetle richness (Jay-Robert *et al.*, 1997) is due mainly to the associated change in climate variables. The influence of purely spatial variables can be regarded as the consequence of the role played by other environmental variables, biotic interactions or past contingent events (Legendre & Legendre, 1998). The scores of the number of species predicted by spatial variables alone are greater in the region of the Massif Central and the Alps (between 3° and 7° in

longitude°, and 43° and 47° in latitude). Although these mountain squares are species-poor, the topographic and environmental heterogeneity of the adjacent squares would have enhanced their capacity to harbour a combination of species from the neighbouring Mediterranean and Eurosiberian regions.

As Cambefort (1991) suggests, climate variables are the most relevant explanatory factors for species richness of dung beetles on both local (as purely climate variables) and geographical scales (as spatially structured climate variables). Of the climate variables, precipitation seems to be of negligible effect. On the spatial scale analysed herein the species richness distribution of Scarabaeidae dung beetles in France seems to be most closely related with temperature variables. Maximum, minimum and annual mean temperatures are each related positively to the number of species, forming linear, linear and curvilinear relationships, respectively. Below an annual mean temperature of 10 °C, a marked decay in species richness occurs (Fig. 6). The influence of the minimum mean temperature, seemingly the most important climate variable by itself (regardless of spatial structure) on species richness, does not depend on geographical location, i.e. should be felt primarily on a local scale (Borcard *et al.*, 1992). The present results suggest that higher minimum temperatures may favour the local co-existence of species, probably because Scarabaeidae dung beetles may not survive cold climate conditions during winter months in great numbers. Maximum and mean temperatures, the most influential spatially structured climate variables, are both significantly and negatively correlated with latitude (Spearman's rank correlation coefficient: $r_s = -0.68$, $P < 0.001$; and $r_s = -0.63$, $P < 0.001$, respectively) but not with longitude ($r_s = 0.16$ and -0.17 , respectively); their geographical influence is probably responsible for the gradient with latitude of Scarabaeidae dung beetles in France.

Accordingly, over the geographical extent analysed, and on this broad spatial scale, Scarabaeidae species richness variation seems to be influenced primarily by climate factors. If the evolution of Scarabaeidae species are similarly constrained by temperature and few species can tolerate cold environments, we then argue that species richness variation across France is due mainly to the failure of many species to go beyond determined temperature range limits. Thus, the regions richest in species are those with a temperature compatible with the maintenance of populations.

ACKNOWLEDGMENTS

We are grateful to James Cerne, who helped us with the English. This paper was supported by the project 'Faunística predictiva: Análisis comparado de la efectividad de distintas metodologías y su aplicación para la selección de reservas naturales' (Spanish D.G.I.; grant: REN2001-1136/GLO).

REFERENCES

- Anderson, M.J. & Gribble, N.A. (1998) Partitioning the variation among spatial, temporal and environmental components in a multivariate dataset. *Australian Journal of Ecology*, **23**, 158–167.
- Austin, M.P. (1980) Searching for a model for use in vegetation analysis. *Vegetatio*, **42**, 11–21.
- Austin, M.P., Cunningham, R.B. & Fleming, P.M. (1984) New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio*, **55**, 11–27.
- Austin, M.P., Nicholls, A.O. & Margules, C.R. (1990) Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs*, **60**, 161–177.
- Austin, M.P., Pausas, J.G. & Nicholls, A.O. (1996) Patterns of tree species richness in relation to environment in south-eastern New South Wales, Australia. *Australian Journal of Ecology*, **21**, 154–164.
- Baraud, J. (1992) *Coléoptères Scarabaeoidea d'Europe, Faune de France 78*. Fédération Française des Sociétés de Sciences Naturelles, Paris and Société Linnéenne de Lyon, Lyon.
- Bessemoulin, J. (1974) *Atlas Climatique de la France Edition Réduite*. Ministère des Transports, Secrétariat Général de l'Aviation Civile et Direction de la Météorologie Nationale, Paris.
- Birks, H.J.B. (1996) Statistical approaches to interpreting diversity patterns in the Norwegian mountain flora. *Ecography*, **19**, 332–340.
- Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.
- Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution wildlife. *Journal of Applied Ecology*, **30**, 478–495.
- Cambefort, Y. (1991) Biogeography and evolution. In: *Dung beetle ecology* (ed. by I. Hanski and Y. Cambefort), pp. 51–67. Princeton University Press, Princeton, N.J.
- Carpenter, G., Gillison, A.N. & Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, **2**, 667–680.
- Cliff, A.D. & Ord, J.K. (1981) *Spatial processes: models and applications*. Pion, London.
- Crawley, M.J. (1993) *GLIM for ecologists*. Blackwell Scientific Publications, Oxford.
- Dennis, R.L.H. & Hardy, P.B. (1999) Targeting squares for survey: predicting species richness and incidence of species for a butterfly atlas. *Global Ecology and Biogeography*, **8**, 443–454.
- Dobson, A.J. (1990) *An introduction to generalized linear models*. Chapman & Hall, Boca Raton, Florida.
- Fraser, R.H. (1998) Vertebrate species richness at the mesoscale: relative roles of energy and heterogeneity. *Global Ecology and Biogeography*, **7**, 215–220.
- Gangloff, L. (1991) *Catalogue et atlas des Coléoptères d'Alsace*. Vol. 4: *Lamellicornia, Scarabaeidae, Lucanidae*. Société alsacienne d'entomologie, Musée zoologique de l'université et de la ville de Strasbourg, Strasbourg.
- Gaston, K.J. (1996) Species richness: measure and measurement. In: *Biodiversity. A biology of numbers and difference* (ed. by K.J. Gaston), pp. 77–113. Blackwell Science, Oxford.

- Haila, Y. & Kouki, J. (1994) The phenomenon of biodiversity in conservation biology. *Annales Zoologici Fennici*, **31**, 5–18.
- Hanski, I. (1986) Individual behaviour, population dynamics and community structure of *Aphodius* in Europe. *Acta Oecologica, Oecologia Generalis*, **7**, 171–187.
- Heikkinen, R.K. & Birks, H.J.B. (1996) Spatial and environmental components of variation in the distribution patterns of subarctic plant species at Kevo, N Finland — a case study at the meso-scale level. *Ecography*, **19**, 341–351.
- Heikkinen, R.K. & 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.
- Hengeveld, R. (1990) *Dynamic biogeography*. Cambridge University Press, Cambridge.
- Hortal, J., Lobo, J.M. & Martín-Piera, F. (2001) Forecasting insect species richness scores in poorly surveyed territories: the case of the Portuguese dung beetles (Col. Scarabaeinae). *Biodiversity and Conservation*, **10**, 1343–1367.
- Huston, M.A. (1994) *Biological diversity: the coexistence of species in changing landscapes*. Cambridge University Press, Cambridge.
- Iverson, L.R. & Prasad, A. (1998) Estimating regional plant biodiversity with GIS modelling. *Diversity and Distributions*, **4**, 49–61.
- Jay-Robert, P., Lobo, J.M. & Lumaret, J.P. (1997) Elevational turnover and species richness variation in European mountain dung beetle assemblages. *Arctic, Antarctic, and Alpine Research*, **29**, 196–205.
- Kerr, J.T. & Packer, L. (1999) The environmental basis of North American species richness patterns among *Epicauta* (Coleoptera: Meloidae). *Biodiversity and Conservation*, **8**, 617–628.
- Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659–1673.
- Legendre, P. & Legendre, L. (1998) *Numerical ecology*. Elsevier, Amsterdam.
- Levin, S.A. (1992) The problem of pattern and scale in ecology. *Ecology*, **73**, 1943–1967.
- Lobo, J.M., Castro, I. & Moreno, J.C. (2001) Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic islands. *Biological Journal of the Linnean Society*, **72**, 233–253.
- Lobo, J.M. & Halffter, G. (2000) Biogeographical and ecological factors affecting the altitudinal variation of mountainous communities of coprophagous beetles (Coleoptera: Scarabaeoidea): a comparative study. *Annals of the Entomological Society of America*, **93**, 115–126.
- Lobo, J.M., Lumaret, J.P. & Jay-Robert, P. (1997) Les atlas faunistiques comme outils d'analyse spatiale de la biodiversité. *Annales de la Société Entomologique de France (N.S.)*, **33**, 129–138.
- Lobo, J.M. & Martín-Piera, F. (1999) Between-group differences in the Iberian dung beetle species–area relationship (Coleoptera: Scarabaeidae). *Acta Oecologica*, **20**, 587–597.
- Lobo, J.M. & Martín-Piera, F. (2002) Searching for a predictive model for Iberian dung beetle species richness (Col., Scarabaeinae) using spatial and environmental variables. *Conservation Biology*, **16**, 158–173.
- Lumaret, J.P. (1990) *Atlas Des Scarabéides Laparosticti de France*. Série Inventaires de Faune et de Flore, fascicule 1, Secrétariat Faune-Flore/MNHN, Paris.
- Lumaret, J.P. & Kirk, A.A. (1991) South temperate dung beetles. In: *Dung beetle ecology* (ed. by I. Hanski and Y. Cambefort), pp. 97–115. Princeton University Press, Princeton, N.J.
- Lumaret, J.P., Lobo, J.M. & Jay-Robert, P. (1996) *Catalogue et Atlas Des Scarabéides Laparosticti Endémiques de France*. Collection Patrimoines Naturels, Vol. 26. Service du Patrimoine Naturel/IEGB/MNHN–OPIE, Paris.
- MacNally, 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.
- Margules, C.R., Nicholls, A.O. & Austin, M.P. (1987) Diversity of *Eucalyptus* species predicted by a multi-variable environment gradient. *Oecologia*, **71**, 229–232.
- Margules, C. & Usher, M.B. (1981) Criteria used in assessing wildlife conservation potential: a review. *Biological Conservation*, **21**, 79–109.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*. Chapman & Hall, London.
- Miller, R.I. (1994) *Mapping the diversity of nature*. Chapman & Hall, London.
- Nicholls, A.O. (1989) How to make biological surveys go further with generalised linear models. *Biological Conservation*, **50**, 51–75.
- Odland, J. (1988) *Spatial autocorrelation*. Sage Publications, California.
- Owen, J.G. (1989) Patterns of herpetofaunal species richness: relation to temperature, precipitation, and variance in elevation. *Journal of Biogeography*, **16**, 141–150.
- Pascual, M.A. & Iribarne, O.O. (1993) How good are empirical predictions of natural mortality? *Fisheries Research*, **16**, 17–24.
- Paulian, R. & Baraud, J. (1982) *Faune Des Coléoptères de France*. II. *Lucanoidea et Scarabaeoidea*. Lechevalier, Paris.
- Pearce, J. & Ferrier, S. (2001) The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation*, **98**, 33–43.
- Pinel-Alloul, B., Niyonsenga, T. & Legendre, P. (1995) Spatial and environmental components of freshwater zooplankton structure. *Ecoscience*, **2**, 1–19.
- Prendergast, J.R., Wood, S.N., Lawton, J.H. & Eversham, B.C. (1993) Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiversity Letters*, **1**, 39–53.
- Qian, H. (1998) Large-scale biogeographic patterns of vascular plant richness in North America: an analysis at the generic level. *Journal of Biogeography*, **25**, 829–836.
- Qinghong, L. & Bråkenhielm, S. (1995) A statistical approach to decompose ecological variation. *Water, Air and Soil Pollution*, **85**, 1587–1592.
- Richerson, P.J. & Lum, K.L. (1980) Patterns of plant species diversity in California: relation to weather and topography. *American Naturalist*, **116**, 504–536.
- Sawada, M. (1999) ROOKCASE: a1n Excel 97/2000 Visual Basic (VB) add-in for exploring global and local spatial autocorrelation. *Bulletin of the Ecological Society of America*, **80**, 231–234.
- Scholtz, C.H. (1990) Phylogenetic trends in the Scarabaeoidea (Coleoptera). *Journal of Natural History*, **24**, 1027–1066.
- Serryn, P. & Blasselle, R. (1970) *Atlas Général Bordas. La France, le Monde*. Bordas, Paris.

- Shipley, B. (1999) Testing causal explanations in organismal biology: causation, correlation and structural equation modelling. *Oikos*, **86**, 374–382.
- Smith, P.A. (1994) Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters*, **4**, 47–61.
- STATISTICA (1995) *Computer program manual for Windows*. StatSoft, Inc., Tulsa, OK.
- Tonteri, T. (1994) Species richness of boreal understorey forest vegetation in relation to site type and successional factors. *Annales Zoologici Fennici*, **31**, 53–60.
- Walker, P.A. (1990) Modelling wildlife distribution using a geographic information system: kangaroos in relation to climate. *Journal of Biogeography*, **17**, 279–289.
- Whittaker, R.J., Willis, K.J. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453–470.
- Williams, P. & Gaston, K.J. (1994) Measuring more 'diversity': can higher taxon richness predict wholesale species richness? *Biological Conservation*, **67**, 211–217.
- Wohlgemuth, T. (1998) Modelling floristic species richness on a regional scale: a case study in Switzerland. *Biodiversity and Conservation*, **7**, 159–177.

BIOSKETCHES

Jorge M. Lobo is a specialist in the biogeography and ecology of dung beetles. He conducts studies on topics such as microsuccession, variation in species richness, changes in composition with altitude, and trophic preferences of species. At present, he is interested in the elaboration of predictive models for biodiversity attributes using spatial and environmental variables, and in the delimitation of historical and ecological processes that could explain species richness distributions.

Jean-Pierre Lumaret's main research interests focus on the biogeography and ecology of dung beetles in the Western Palaearctic region; larval taxonomy and phylogeny of Scarabaeoidea; ecology and biodiversity of pastures; pesticides and environmental risks.

Pierre Jay-Robert's research interests focus also on western Palaearctic dung beetles, mainly on community structure and the spatial distribution of diversity.