

Mapping species richness of endemic macroinvertebrates by overlapping distribution maps

Miquel PALMER, Guillem X. PONS, Raimundo OUTERELO, Eduard VIVES, Xavier BELLÉS, Lluç GARCIA, Marta LINDE, Lluís GÓMEZ-PUJOL and David MARCH

SHNB



SOCIETAT D'HISTÒRIA
NATURAL DE LES BALEARS

Palmer, M., Pons, G.X., Outerelo, R., Vives, E., Bellés, X., Garcia, L., Linde, M., Gómez-Pujol, L. and March, D. 2012. Mapping species richness of endemic macroinvertebrates by overlapping distribution maps. *Boll. Soc. Hist. Nat. Balears*, 55: 99-122. ISSN 0212-260X. Palma de Mallorca.

Inventory-based approaches (i.e., those that directly relate species richness to explanatory variables) do not work on the medium and local scale that was studied in this paper; even when ecologically-meaningful environmental predictors were used. The use of an alternative approach (*Taxon-based diversity mapping*) is proposed. This approach starts by modelling the probability of occurrence of twelve target endemic species using environmental variables as predictors. Next, it projects the probability of occurrence from 48 sampled sites to a grid of 532 1-km² units. Finally, the approach estimates endemic species richness at these 532 1-km² units by overlapping the twelve maps inferred. We also propose a new approach whose purpose is to detect possible false absences and unstable presences. These doubtful observations were excluded from ecological niche modelling. In addition to species richness, the patterns experienced by species composition were analysed. Species composition experienced strong changes (i.e., large turnover), while species richness remained constant. This stasis in species richness is neither related to low environmental variability (both the biotic and abiotic scenarios are diverse) nor to the lack of species-environment relationships (not only species composition but also species-specific responses are correlated with environment). The area studied is environmentally diverse and species turnover of the macroinvertebrate community was moderate to large. Species composition was significantly correlated with environment (minimum temperature). The predicted number of target endemic species per cell in 92% of the studied area fell between 3 and 5. Therefore *Taxon-based diversity mapping* has been confirmed as a valid alternative to conventional inventory-based diversity mapping. Additionally, the new procedure proposed here for dealing with noisy presence/absence data produces more accurate distributional maps of individual species.

Keywords: *species richness, distribution maps, occurrence patterns, spatial structure, GLZ modelling.*

CARTOGRAFIA DE LA RIQUESA D'ESPÈCIES DE MACROINVERTEBRATS ENDÈMICS A PARTIR DE LA SUPERPOSICIÓ DE MAPES. Els resultats de la cartografia biogeogràfica que es recolcen en predictors de significat ecològic derivats d'inventaris faunístics (i.e. aquells que relacionen directament la riquesa d'espècies amb variables explicatives) no acostumen a oferir resultats òptims a les escales espacials d'ordre mig i local, com les que aborda aquest treball en el marc de la Serra de Tramuntana (Mallorca). A mode d'alternativa es proposa l'ús d'un mètode de cartografia centrat en la diversitat dels tàxons. Aquesta aproximació modelitza la probabilitat d'aparició de dotze grups d'espècies endèmiques tot emprant les variables ambientals com a predictors. Així les coses, es representa la probabilitat d'ocurrència a partir de 48 punts de mostreig d'una xarxa de 532 cel·les d'1 km². A partir de dites probabilitats es superposen els 12 mapes d'inferència per tal d'estimar

la riquesa d'espècies a cada una de les 532 unitats d'anàlisi. Adicionalment es proposa un mètode per detectar les falses absències així com les presències fortuïtes. Aquestes observacions dubtoses són excloses del model de nínxol ecològic. En paral·lel a la representació de la riquesa d'espècies, s'aborden els patrons d'agrupació d'espècies. La composició dels grups experimenta variacions importants, tot i que la riquesa d'espècies roman constant. Aquesta invariabilitat en el nombre d'espècies no sembla estar relacionada amb l'escassa variabilitat ambiental, ja que tant el marc biòtic com l'abiòtic són diversos, així com tampoc amb la manca de correlació entre espècies i medi (que es dona tant pel que fa a la composició d'espècies, així com també amb la resposta específica d'algunes espècies). L'àrea d'estudi és diversa des d'un punt de vista ambiental i el volum d'espècies de macroinvertebrats és de moderat a important. La composició dels grups d'espècies presenta una correlació significativament amb el medi (temperatura mínima). El nombre de grups d'espècies endèmiques per a un 92% de les unitats d'anàlisi oscil·la entre 3 i 5. Així doncs, la cartografia de la biodiversitat a partir de tàxons s'ha confirmat com una alternativa vàlida als mètodes convencionals fonamentats en inventaris. De més a més, la proposta per tal de detectar les falses absències/presències resulta en una cartografia de distribució específica de les espècies força més acurats.

Keywords: *riquesa d'espècies, mapes de distribució, patrons de presència, estructura espacial, modelització GLZ.*

Miquel PALMER, Marta LINDE i David MARCH, Institut Mediterrani d'Estudis Avançats, IMEDEA (CSIC-UIB), Miquel Marqués 21, 07190 Esporles. Raimundo OUTERELO, Departamento de Zoología y Antropología Física, Universidad Complutense de Madrid, José Antonio Novais 2, 28040 Madrid. Guillem X. PONS, Dept. Ciències de la Terra, Universitat de les Illes Balears, Cra. Valldemossa, km 7.5. Eduard VIVES, San Antoni 73, 08221 Terrassa. Xavier BELLÉS, Institut de Biologia Evolutiva (CSIC-UPF), Passeig Marítim 37, 08003 Barcelona. Lluc GARCIA, Museu Balear de Ciències Naturals, Cra. Palma-Port de Sóller km 30, 07100 Sóller. Lluís GÓMEZ-PUJOL, ICTS SOCIB, Sistema d'Observació i Predicció Costaner de les Illes Balears, Parc Bit, Cra. Valldemossa km 7.4, 07121 Palma. E-mail: lgomez-pujol@uib.cat.

Recepció del manuscrit: 03-mai-12; revisió acceptada: 22-nov-12

Introduction

The main goal of this contribution is to improve description and analysis of the patterns depicted by species richness, and to compare them with those suggested by species composition. Different types of causal hypotheses have been proposed to explain the spatial patterns of diversity at large scales (Ruggiero and Kitzberger, 2004). Energy-related hypotheses (e.g., Hawkins *et al.*, 2003) predict that an increase in productivity promotes an increase in both biomass and number of individuals, which, in turn, are related to higher diversity. The mechanistic link

between biomass and diversity has been proposed by the Unified Neutral theory (Hubbell, 2001). Alternatively, different versions of the environmental variability hypothesis suggest that less variable environments enable a greater number of species to coexist because species are able to evolve and occupy narrower ecological niches (Brown and Lomolino, 1998). The potential effects of historical contingency have also been proposed to explain the occurrence of high diversity patches (Hawkins *et al.*, 2003). Finally, peaks of diversity occurring at the geographic centre of latitudinal or altitudinal gradients have been related to the effect of the mere

geometric limits. These limits constrain the species ranges and claim that these ranges increasingly overlap toward the centre (mid-domain effect, Colwell *et al.*, 2004). At the mesoscale, empirical models successfully estimate species richness from a variety of environmental variables (e.g., Stefanescu *et al.*, 2004; Bhattarai *et al.*, 2004; Pei-Fen *et al.*, 2004; Nogués-Bravo and Martínez-Rica, 2004; Mac Nally *et al.*, 2003). However, uncertainties on the causal processes involved usually remain. The problem is exacerbated in case studies at local scale because the variability in both the response (species richness) and/or the explanatory variables (environment) is usually too small to reveal any significant pattern. Therefore, the lack of theoretical and empirical models that precludes mapping diversity at local geographical scale is too great to be useful for environmental managers.

Here we explore an alternative approach for mapping diversity and species distributions at the local scale. *Taxon-based diversity mapping* derives diversity maps by overlaying the distribution maps of a number of target species (Müller *et al.*, 2003). The *a priori* selection of the specific set of target species is an important advantage for environmental managers because high diversity hot-spots, in terms of endangered or endemic species, are not confounded with other sites that display a large number of broadly distributed species. The analytical strategy adopted here is summarized in Figure 1. The method is grid-based and the diversity at each geographical unit could ideally be estimated by the number of target species occurring in this unit. However, the current level of knowledge on the spatial distribution of most species remains inadequate to be useful for this purpose. Therefore, the procedure for taxon-based mapping is to

infer the most plausible distribution for each species considered.

Such an inferential procedure masks a number of theoretical and applied caveats.

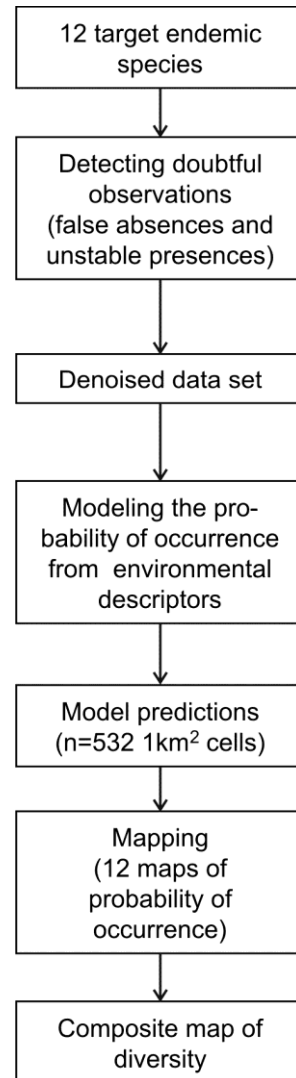


Fig. 1. General analytical strategy for taxon-based diversity mapping.

Fig. 1. Estratègia d'anàlisi per a la cartografia de biodiversitat basada en tàxons.

From a theoretical perspective, the conventional procedure for ecological niche modeling (the *correlative approach*; Soberón and Peterson, 2005) tries to find the correlation patterns between presence and/or absence of a species and environmental variables. It uses these patterns to extrapolate the areas of (predicted) occurrence from a relatively small number of surveyed sites to the entire landscape of interest (provided that the environmental data sets cover it). However, strictly speaking, this approach only finds regions that environmentally resemble those where occurrences are currently located (Soberón and Peterson, 2005). Interpretation of this correlation pattern, in terms of ecological niche (i.e., set of conditions under which a species is able to maintain populations without immigrations), should be done with caution because niche modeling is usually based only on abiotic data (e.g., climate, physical environment, soil conditions, and other variables typically implemented in geographic information systems). Species' occurrence also depends on: (i) the dispersal abilities of the species considered that can introduce a state of non-equilibrium between appropriate environment and species distribution; and (ii) biotic factors (mutualisms, competition, predation, etc.) (Soberón and Peterson, 2005).

Concerning the first point, it is expected that the importance of non-equilibrium due to dispersal is considerable at a large spatial scale but ameliorate at a small scale. All the points in the landscape should be considered reachable by a species when (i) the spatial scale is small, (ii) the biology of the species allows for efficient dispersal and (iii) the landscape does not include obvious barriers (Soberón and Peterson, 2005). All these three conditions are plausibly attained in the case considered here.

Concerning the second point, the subspace of sites with favorable abiotic conditions for a specific target species (the "fundamental ecological niche") does not necessarily coincide with the subspace of sites where appropriate species are present and inappropriate species (including man) are absent (Soberón and Peterson, 2005). Assuming that all the points in the landscape can be reached by a species, the intersection of these two subspaces constitutes the geographical expression of the *realized niche* (Soberón and Peterson, 2005). Consequently, any correlation inference based in the observed pattern of occurrence is restricted to the currently realized niche (as shown by Guisan and Zimmermann, 2000). However, even the inference of the geographical expression of the realized niche is problematic when: (i) using noisy presence/absence data-sets (data with false absences and unstable presences) and (ii) dealing with spatially-structured data-sets. Here, we propose a new procedure for dealing with noisy presence/absence data. We use a spatially-explicit modeling procedure for estimating the probability of occurrence of each of the target species in the entire scenario.

Material and methods

Sampling and data

The island of Mallorca (Balearic Islands, Western Mediterranean) covers an area of 3640 km² and is located in the middle of the Western Mediterranean. It has a typical Mediterranean climate with hot dry summers and mild wet winters. The mean annual temperature is approximately 17°C, with mean winter and summer values of 10 and 25°C, respectively; the mean annual precipitation is about 500 mm and is mostly concentrated in autumn (Guijarro,

1986). The vegetation is typically Mediterranean with two clear community types: holm oaks, *Cyclamini-Quercetum ilicis*, with boreal characteristics abundant at the lowest altitudes and macchia and garrigue bushes, *Oleo-Ceratonion*, *Hypericion balearici*, *Rosmarino-Ericion* mainly in the drier lowlands (Bolòs, 1996).

The study area is a mountain range located at the NW of Mallorca Island, rising from the seashore to 1445 m a.s.l. The Serra de Tramuntana forms the northwest side of the island and comprises several tectonic zones (Fallot, 1922; Gelabert, 1998), mainly built of Mesozoic sediments, cut by longitudinal valleys, karstic canyons and poljes, and with a spectacular cliffed coast (Gelabert *et al.*, 2003). The study area

comprises a grid of 532 UTM 1-km² cells. Coastal, urban, peri-urban, and intensive agricultural cells were not considered in the analysis. Forty-eight cells were selected and evenly distributed along the main range axis (100 km long) for sampling (Fig. 2). Five pitfall traps (10 cm wide; detergent and salt used as preservative) were set 2 m apart at each site. The position of the sites considered was GPS-determined. The traps remained in the field for two months (mid May –mid July). The sampling schedule was complemented with one hour of direct search (e.g. under stones) at each site on a 1000 m² plot around the pitfall traps. The lists of species obtained from the five pitfalls and from the direct search were pooled as a single sample per cell. We

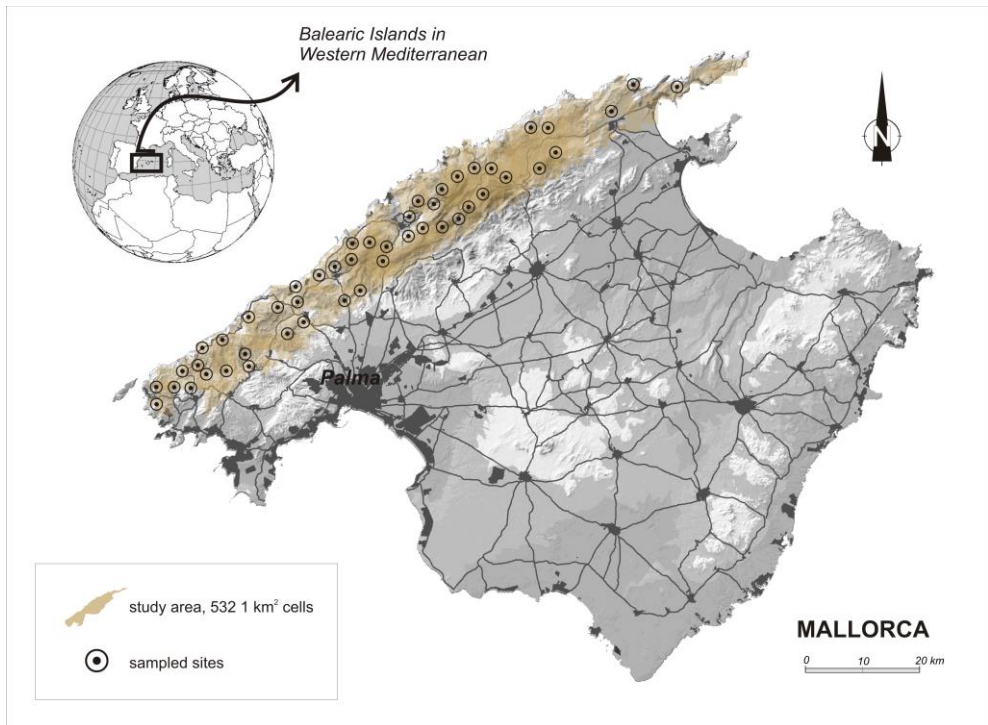


Fig. 2. Location of the area studied and sampled sites.

Fig. 2. Àrea d'estudi i localització dels punts de mostratge.

focused on four groups of invertebrates: Coleoptera, Arachnida, Isopoda and Gastropoda. In regard to Gastropoda, species were also considered to be present when empty shells were found. Some of the species (23 of 196; 11%) were not readily determinable and were identified as morphospecies (Oliver and Beattie, 1996).

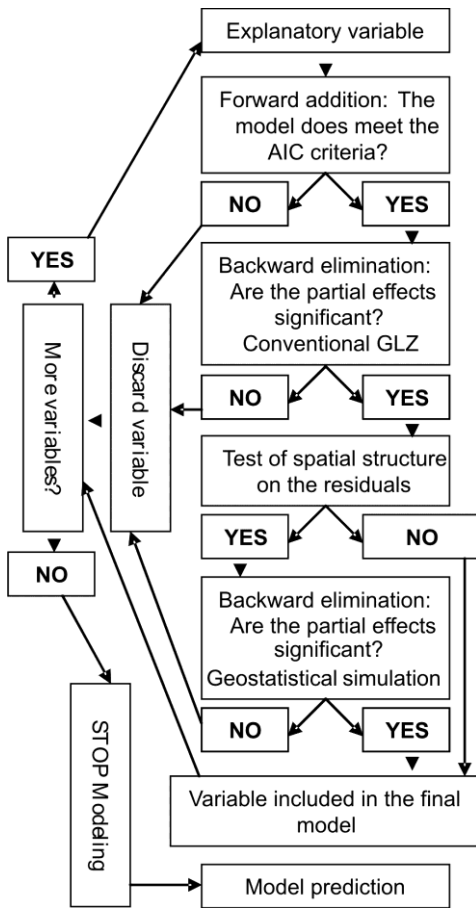


Fig. 3. Analytical strategy for selecting the explanatory variables included in the species-specific final models.

Fig. 3. Estratègia analítica per a la selecció de variables explicatives incloses en els models finals específics per a espècies.

For each of the 48 cells sampled, the values of 12 explanatory variables were recorded. These variables included: altitude, aspect (transformed following Moisen and Frescino, 2002), slope, distance to the closest house (from the centre of the cell; ln-transformed), NVDI (a non-linear combination of reflectance at specific lambdas extracted from a LANDSAT satellite image; Kerr and Ostrovsky, 2003, Turner *et al.*, 2003). Climatic data including: annual average temperature, minimal and maximal average monthly temperature, minimum absolute monthly temperature, total precipitation, and actual and potential evapotranspiration, were estimated using a regression-based interpolation procedure (Guijarro, 1986). All these variables were also recorded also for each of the 532 UTM 1-km² cells of the entire landscape of interest. The methods and sources for estimating these values are detailed elsewhere (Palmer *et al.*, 2002).

Species richness conventional analysis

A generalized linear model (GLZ) was used to model the relationships between species richness and environmental variables. Environmental variables were included in the model following a forward step-by-step sequence until the maximum explanatory power was reached. Then, some variables can be backward eliminated, looking for generality and significant partial effects of all the variables. The process of variable selection is summarized in Fig. 3. In order to avoid overfitting, the measure of explanatory power (deviance) was penalized by the number of variables included in the model. This was achieved using the AIC criterion. Some highly correlated environmental variables were *a priori* left out. The patterns of correlation between explanatory variables were depicted using Principal Components Analysis.

Some of the sampling sites are in close proximity to each other. At the local scale population dynamics, interspecific interactions and other biotic processes that operate only in the immediate neighborhood produce spatial autocorrelation, which in turn might mask environmental cues or suggest spurious relationships. Alternatively, nearby observations of species richness could be similar due to a spatially structured environment.

The statistical and conceptual problems posed by spatial structure have been considered elsewhere (e.g., Lennon, 2000; Diniz-Filho *et al.*, 2003; Lichstein *et al.*, 2002; Liebhold and Sharov, 1998; Fortin and Payette, 2005; Legendre *et al.*, 2002; Keitt *et al.*, 2002). Hypothesis testing tends to be too liberal when the assumption of independence of the residuals is violated (e.g., Lennon, 2000). In these cases, the null hypothesis of no correlation between species richness and environment is rejected more frequently than expected by chance.

The existence of spatial structure on the residuals (or on the raw data) can be explored using some methods. Here we use semivariograms (Legendre and Legendre, 1998). A semivariogram is a graph displaying a convenient measure of the variability of the between-site differences (semivariance; Legendre and Legendre, 1998) at progressively increasing distance intervals. The fitted variograms are used to predict the distance-related decay of a variable at any point in the landscape. This information is used by most of the methods capable of managing data that exhibit spatial structure.

Among these methods, we have selected geostatistical simulation (Liebhold and Sharov, 1998). This method simulates a random landscape with the same spatial structure exhibited by the observed

landscape. The strength of the relationship (measured, for example, by the F-statistic) between the response variable (species number) and the data extracted from the simulated landscape was measured. This process was iterated a large number of times and the F-value was recorded at each iteration. The distribution of F-values under the null-hypothesis was then constructed for each specific dataset and compared to the observed F-value.

Model building, semivariogram analysis and geostatistical simulation were completed with the libraries GeoR (Ribeiro and Diggle, 2001) and the tool RandomFields (Schlather, 2001) of the R package. The version 2.0.1 of the R package was used (<http://www.r-project.org/>).

Taxon-based diversity mapping

This second approach of diversity analysis involves building up the occurrence maps for a number of target species and inferring the underlying spatial pattern of diversity by summing these individual maps. The analytical strategy adopted here is summarized in Fig. 1.

Twelve target endemic species were chosen from the 24 endemic species detected in the studied area, with the constraint that their prevalence should fall between 20 and 80% in order to avoid considering very rare or very conspicuous species in the analysis (Table 1).

The statistical problems of inferring the unsuitable sites in conspicuous species are the same as inferring suitable sites for rare species. Moreover, very conspicuous species do not contribute to enhance between-site differences in biodiversity. In our case, the number of target endemics in a site was a surrogate of the total number of endemics ($n = 48$; $r = 0.85$; Prob. < 0.001) and the total number of species ($n = 48$; $r = 0.61$; Prob. < 0.001).

Class/Order	Species	Number of cells
Gastropoda	<i>Xerocrassa frater</i> (Dohrn and Heynemann, 1862)	46
Gastropoda*	<i>Iberellus balearicus</i> (Ziegler, 1853)	38
Gastropoda*	<i>Tudorella ferruginea</i> Lamarck, 1823	36
Coleoptera*	<i>Percus plicatus</i> Dejean, 1828	27
Arachnida*	<i>Euscorpius balearicus</i> Caporiacco, 1950	26
Gastropoda*	<i>Oxychilus lentiformis</i> Kobelt, 1882	24
Coleoptera*	<i>Phylan semicostatus</i> (Mulsant and Rey, 1854)	19
Arachnida*	<i>Malthonica balearica</i> Brignoli, 1978	18
Arachnida*	<i>Harpactea dufouri</i> Thorell, 1873	15
Coleoptera*	<i>Asida planipennis</i> Schaufuss, 1869	14
Coleoptera*	<i>Asida barceloi</i> Pérez Arcas, 1868	14
Arachnida*	<i>Nemesia brauni</i> Koch, 1882	14
Coleoptera*	<i>Nesotes viridicollis</i> (Schaufuss, 1869)	10
Coleoptera	<i>Alphasida depressa</i> (Solier, 1836)	6
Coleoptera	<i>Jekelius balearicus</i> (López-Colón, 1984)	5
Gastropoda	<i>Iberellus (Allognathus) graellsianus</i> (Pfeiffer, 1848)	4
Coleoptera	<i>Orthomus balearicus</i> (Piochard, 1867)	4
Arachnida	<i>Dasylobus ferrugineus</i> Thorell, 1876	4
Coleoptera	<i>Dendarus depressus</i> Reitter, 1915	3
Coleoptera	<i>Timarcha balearica</i> Gory, 1829	2
Coleoptera	<i>Platyderus majoricus</i> Jeanne, 1988	1
Isopoda	<i>Ballodillium pilosum</i> Vandel, 1961	1
Arachnida	<i>Hahnia hauseri</i> Brignoli, 1978	1
Coleoptera	<i>Xantholinus balearicus</i> Coiffait, 1962	1

Table 1. List of the endemisms found. The twelve target endemic species are indicated with an *.
Taula 1. Relació dels endemismes capturats. * Espècies que formen part del grup diagnòstic.

Identifying and removing doubtful observations

The essential procedure of taxon-based diversity mapping is to extrapolate the distribution of the target species from a relatively small number of sampled cells to all the cells of the landscape of interest (Müller *et al.*, 2003). This goal was achieved by using ecological niche modeling (Soberón and Peterson, 2005). However, theoretical background suggests that false absences (absences in sites where the species is currently present) and unstable presences (temporary presences in sites with unsuitable conditions for maintaining a stable population) have a significant impact in niche modeling and in projecting niche models onto maps. A

simplified example of the consequences of noise on the modeling procedure is shown in Fig. 4, where only two doubtful data points are introduced in an otherwise “clean” data set. The ecological tolerance is estimated to be larger in the noisy data set. Consequently, the geographical projection of these two models will be very different. Here, we propose that biases in ecological modeling related to data set noise could be ameliorated after the removal of these doubtful observations from the raw data. Specifically, the presence/absence data set for each target species was denoised using the information gathered from the presence/absence data of all other species in the community (Palmer *et al.*, 2003). The proposed vetting procedure was based on

the continuum theory, which implies a gradual change in species composition along continuous environmental gradients (Gauch and Whitaker, 1972; Putman and Wratten, 1984, ter Braak and Smilauer, 2002). In spite of the ongoing debate on the shape of the response curves of species to environment (Oksanen and Minchin, 2002), or on the rules controlling species turnover and species packing (Chave and Levin, 2003, Buttel and Durret, 2002), it is reasonable to assume that some expectation on the presence/absence of an individual species could be deduced from the presence/absence of all other species in the community.

The analytical strategy for detecting doubtful observations is summarized in Fig. 5 and is comprised of three steps: (i) to

reduce the dimensionality of the raw data (presence/absence of 107 species in 48 sites), (ii) to model probability of occurrence and (iii) to use the confidence interval of the model predictions to detect doubtful observations.

The problem in the first step is that the number of species is usually too large in relation to the number of samples. This fact imposes numerical instability on any attempt to parameterize a model aimed at using all species as potential explanatory variables. Therefore, as a preliminary step, matrix dimensionality of the original presence/absence data set (107 species after the removal of 89 species that were detected in less than 5 sites) was reduced using Correspondence Analysis (CA, ter Braak and Smilauer, 2002). A specific site

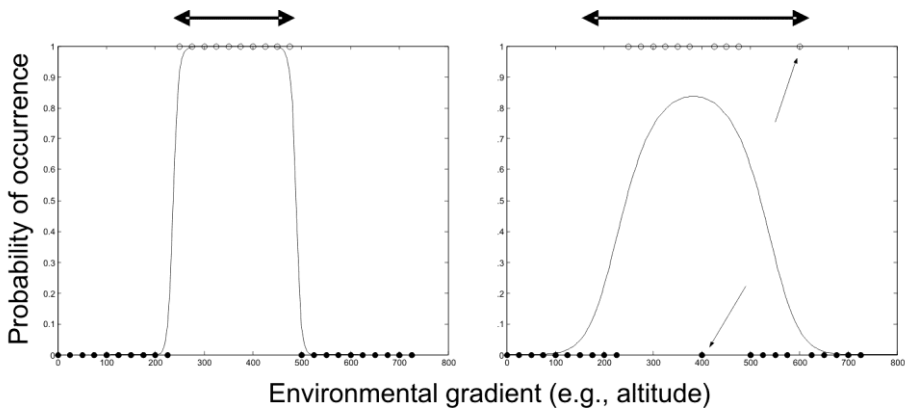


Fig. 4. Simplified example of modelling two sets of presence/absence datasets. Occurrences are indicated by open circles and absences by closed circles. The lines are obtained by adjusting the data to a 2-degree logit model. A clear-cut relationship is shown in the left panel (the species occur at “middle altitude” only). The left panel shows the results after simulating one false absence (the species is not detected in one site, just in the middle) and one unstable presence. These two doubtful observations are denoted by arrows. The inferred ecological niche and the predicted spatial distribution will be broader in the second case.

Fig. 4. Exemples simplificats de dues de les aproximacions a la modelització de la presència/absència. Les presències estan representades per punts blancs i les absències per punts negres. El model de la gràfica de l'esquerra presenta uns límits clars entre les absències i les presències, mentre que el de la dreta, més inestables, presenten espècies no detectades o falses presències (indicats amb les fletxes). La inferència del nínxol ecològic i la predicció espacial de la distribució de l'espècie serà menys precisa en el segon dels casos.

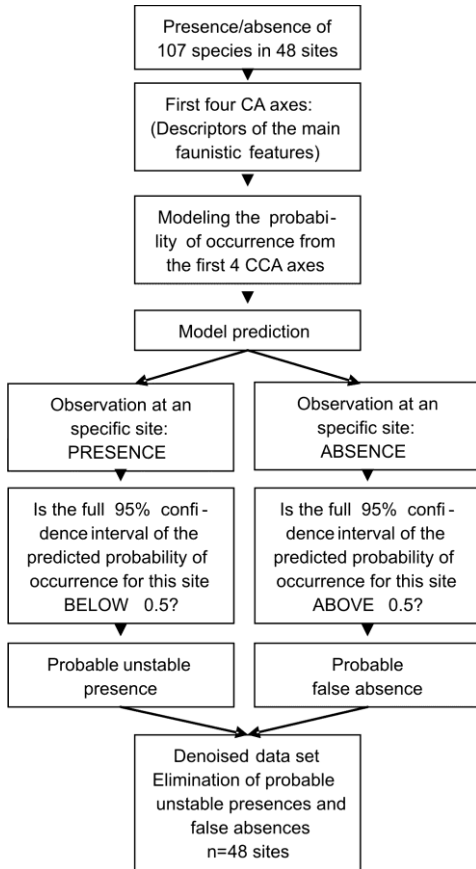


Fig. 5. Summary of the procedure for detecting false absences and unstable presences.

Fig. 5. *Resum del procediment per detectar falses absències o presències inestables.*

site can be viewed as a point in a space of 107 dimensions. The “scores” of each site on these 107 dimensions were the presence (1) or absence (0) of the 107 species. CA reduces the dimensionality of the original matrix similarly to Principal Components Analysis (PCA). All these techniques define a new space. The faunistic differences (determined using a convenient measure; here the chi-squared distance; Legendre and Legendre, 1998) between

sites are the same in the original space and in the space defined by the CA axes. However, in the new (CA) space, the first few axes summarize almost all the variability displayed by the 107 dimensions of the original space. Therefore, faunistic composition of each site can be approximately defined using the scores of the first few CA axes only without losing relevant information.

In the second step, presence/absence of the target species (response variable) was modeled using Generalized Linear modeling (GLM; 2-degree polynomial logistic model with binomial distribution of errors) on the scores of the first four CA axes (explanatory variables). This model predicted the probability of the occurrence of a target species on a specific set of values of CA axes. The latter are surrogates of a specific faunistic combination of presences of appropriate species and absences of inappropriate species. The selection of the best subset of the explanatory variables is the same as what was summarized in Fig. 3. It was fully explained in the previous section (Conventional analysis of species richness).

The third step consisted of detecting doubtful observations using the confidence interval of the model predictions (i.e., the probability of occurrence). A false absence was assumed to occur when the full 95% confidence interval of the expected probability of presence of the target species on a site was above 0.5. This occurred when a site was faunistically very similar to all other sites where the target species occurred (except for the absence of the target species). Similarly, a false presence was declared when the confidence interval fell below 0.5. This meant that the target species occurred in a site together with a very unusual species set, and an example of computer-simulated data is shown in Fig. 6.

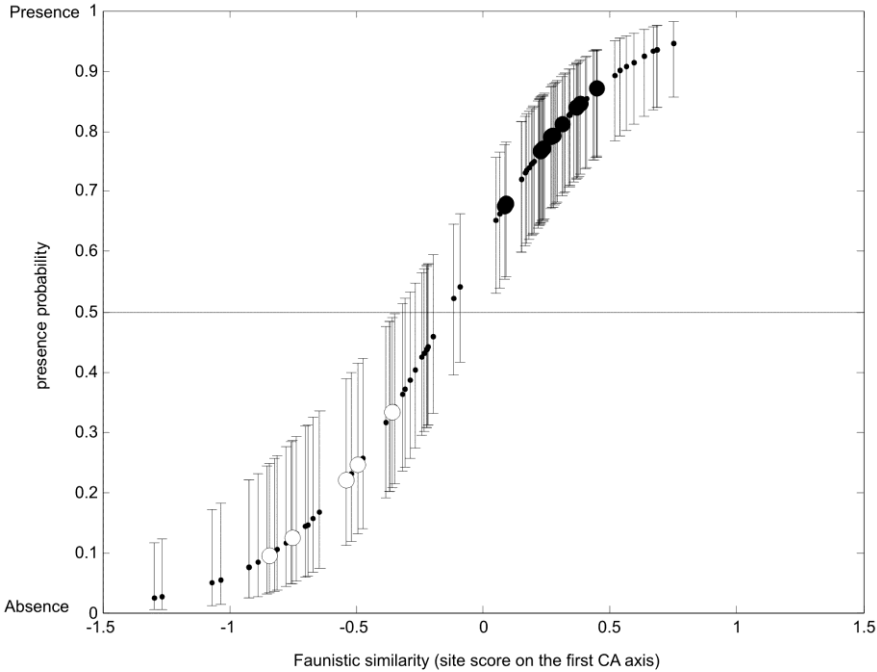


Fig. 6. Example of the vetting procedure using computer simulated data. A community was simulated with the constraint of being correlated with a single environmental gradient. Therefore, the first CA axis accounts for almost all the faunistic variability. Presence/absence data of one target species (response variable) was modeled (logistic model) using the scores on the first CA axis as an explanatory variable. 95% confidence intervals of all samples are shown. Doubtful observations are indicated with a larger circle. Filled circles correspond to false absences (actual absences in sites with high probability of presence of the target species) and open circles to unstable presences (actual presences in sites with high probability of absence).

Fig. 6. Exemple del procés d'avaluació-selecció a partir de simulacions d'ordinador. Una comunitat és simulada amb les restriccions de la correlació amb un únic gradient ambiental. Així doncs, el primer eix de l'AC explica bona part de la variabilitat faunística. La presència/absència del grup diagnòstic d'espècies (variable depenent) fou modelitzada (model logístic) emprant les puntuacions del primer eix de l'AC com a variable explicativa. Es presenta l'interval del 95% de confiança per a totes les mostres, mentre que el casos dubtosos es marquen amb un cercle de majors dimensions. Els cercles negres corresponen a falses absències (absències a localitat amb una alta probabilitat de presència del grup diagnòstic) i els cercles blancs les falses presències (presències actual a localitats amb una alta probabilitat d'absència).

Ecological niche modeling

Doubtful observations were removed and the denoised data sets were used for the ecological niche modeling. We used a

Generalized Linear model (GLZ; 2-degree polynomial logistic model with binomial distribution of errors). The best subset among the 12 putative environmental

predictors was selected using the same method described below (Fig. 3; the same combination of step-by-step forward procedure and backward testing of the partial effects of the variables already included in the model that has been detailed below). Up to 2-degree polynomial extensions of all the environmental variables was also considered in order to accommodate possible unimodal responses.

Mapping diversity and species distribution

The next step consisted of predicting the probability of occurrence of the 12 target species at the 532 1 km²-units (i.e., the entire landscape) using the best subset of explanatory variables as predictors. The best specific subset of explanatory variables was species-specific. The values of all the explanatory variables were known at each of the landscape cells.

The final step consisted of combining the 12 maps of probability of occurrence (one per species) into a single map that depicted the expected number of target species per cell.

The accuracy of the predictions of the ecological niche models corresponding to the full ($n=48$) and the denoised data set ($n \leq 48$) were compared using a threshold-independent method (AUC) because the model's output are probabilities of occurrence (Fielding and Bell, 2003). In addition to these empirical results, a computer simulation experiment also indicated that the maps produced by denoised data are more accurate than the corresponding maps inferred directly from the raw (observed) presence/absence data set.

Model building, ordination analyses (Correspondence analyses), semivariograms and geostatistical simulations were completed using the libraries Stat (basic module) GeoR (Ribeiro and Diggle, 2001),

RandomFields (Schlather, 2001) and Vegan (Oksanen, 2005) of the R package. The version 2.0.1 of the R package was used (<http://www.r-project.org/>). The detection of doubtful observations and the combination of the 12 maps were completed using MATLAB routines.

Species composition patterns

Univariate generalized linear models were used to evaluate the existence of significant correlation patterns between the responses of a single variable (e.g., species richness) to environment. However, species composition (presence/absence of 107 species) has a multivariate nature. Therefore, we use a multivariate direct method that relates a response matrix (presence /absence of 107 species in 48 sites) with an explanatory matrix (12 environmental variables). Specifically, we used Canonical Correspondence Analysis (CCA, Legendre and Legendre, 1998; ter Braak and Smilauer, 2002) to estimate how much variation in species composition (response variable) can be attributed to the environmental variables considered. Therefore, we focused on the capability of multivariate methods for testing the existence of a relationship between the main (in the sense of the most common for the species) faunistic gradients and their putative explanatory environmental variables. The ratio between the variability (inertia in the multivariate jargon) explained by the model and the residual inertia was used to test the model's significance (by using Monte Carlo permutation methods). The strategy adopted here for model building partially emulated the one shown in Fig. 3, but extrapolated from univariate to multivariate data. The selection of the best subset of explanatory variables was completed following a step-by-step forward addition procedure (AIC-based), and testing the significance of

partial effects (i.e., partial CCAs) of the variable added at each model building step.

Concerning the spatial structure, we applied recent developments aimed at integrating multivariate ordination techniques (e.g., canonical correspondence analysis) and geostatistics. We used diagnostic tools for testing autocorrelation of residuals and depicted any eventual scale-independent correlation structure (Wagner, 2003, Wagner, 2005, Couteron and Ollier, 2005). Emulating the univariate semivariograms described above, the raw multivariate data and the multivariate residuals (residuals from canonical correspondence analyses) can be partitioned by distance lags and integrated with geostatistics using multivariate variograms (Wagner, 2005). These variograms were used as diagnostic tools for: (i) looking for stationary (i.e., the spatial variability tends to stabilize at large distances), (ii) testing autocorrelation of the residuals (spatially-related variability in species composition not explained by the environmental variables included in CCA), and (iii) for exploring scale-dependent correlation structure (Wagner, 2005). This latter capability is very interesting because it compares the spatial structure of the overall variability (i.e., in the raw data) with the sum of the spatial variability of explained and unexplained variances. These quantities cancel out when the entire landscape is considered, but the spatial partitioning can evidence, for example, the existence of subareas displaying different relationships between species composition and environment (i.e., regionalization).

Multivariate analyses were completed using the R library *Vegan* (Oksanen, 2005). Calculations, significance tests and plotting of the multivariate variograms were completed using the R function *mso* (Wagner, 2005). The version 2.0.1 of the R

package was used (<http://www.r-project.org/>).

Results

Species richness conventional analysis

Sampling of the 48 sites rendered 196 species in the four macroinvertebrate groups considered (Coleoptera, Arachnida, Gastropoda and Isopoda). A GLZ analysis was completed using species richness as the response variable. None of the 12 environmental variables considered explained a significant amount of the variability found in species richness or in the number of endemic species. The best predictor of species richness was the log-transformed distance to the nearest house (*Houses*; GLZ with Gaussian-distributed residuals: Slope of the regression line=4.6; Probability that the slope were different from zero=0.14; AIC of the null model=343.0; AIC of the model=342.7; variance explained by the model 4.5%; Fig. 7). Log-transformation of the response variables rendered qualitatively the same results. Observations should be considered fully independent because the residuals of this best model did not display spatial structure (Fig. 7). The results obtained that replaced species richness with the number of endemic species were qualitatively comparable.

Therefore, it should be concluded that inventory-based approaches (i.e., those that directly relate species richness to explanatory variables) do not work at the scale studied even when using ecologically-meaningful environmental predictors.

Taxon-based diversity mapping

The twelve target endemic species that were selected are listed in Table 1. For each species, the raw data (presence/absence in 48 cells) were first submitted to the vetting

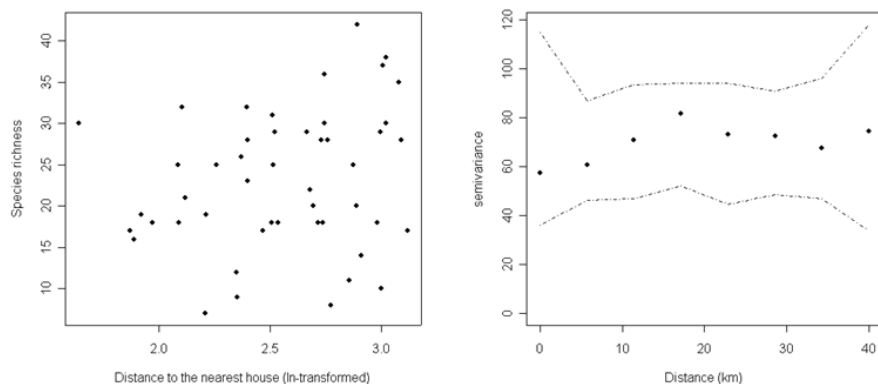


Fig. 7. Left panel: Scatter plot of the relationship between species richness and the environmental variable with the best explanatory power. Right panel: semivariogram of the residuals of this model (there is no evidence of spatial structure because the observed semivariations are within the 95% confidence interval envelope for the 8 lags considered).

Fig. 7. Gràfica esquerra: Gràfic de dispersió de la relació entre riquesa d'espècies i les variables ambientals de major poder explicatiu. Gràfica dreta: semivariograma dels residuals del model anterior (no hi ha evidències de cap estructura espacial perquè les semivariàncies observades s'ubiquen dins el marge de l'interval de confiança del 95% que inclou 8 dels lags considerats).

Species	Doubtful Observations	Variables selected	Explanatory power
<i>Iberellus balearicus</i> (Ziegler, 1853)	5	–	–
<i>Tudorella ferruginea</i> (Lamarck, 1923)	6	Tm	22.3%
<i>Percus plicatus</i> (Dejean, 1828)	3	Nvdi+Alt+PET ²	29.3%
<i>Euscorpius balearicus</i> (Caporiacco, 1950)	0	Tma+TM ² +Aspect ²	37.2%
<i>Oxychilus lentiformis</i> (Kobelt, 1882)	2	T+Aspect	18.0%
<i>Phylan semicostatus</i> (Mulsant and Rey, 1854)	3	T+Alt ²	23.1%
<i>Malthonica balearica</i> (Brignoli, 1978)	4	Prec+Slope	24.5%
<i>Harpactea dufouri</i> (Thorell, 1873)	1	Prec+T ² +Aspect ²	39.5%
<i>Asida planipennis</i> (Schaufuss, 1869)	4	Houses+slope	26.7%
<i>Asida barceloi</i> (Pérez Arcas, 1868)	4	T	7.3%
<i>Nemesia brauni</i> (Koch, 182)	4	Aspect+Slope ²	30.0%
<i>Nesotes viridicollis</i> (Schaufuss, 1869)	2	Alt+Alt ²	30.2%

Table 2. List of the 12 target species indicating the number of doubtful observations, the variables included in the predictive model and the explanatory power measures as $100 \cdot r^2$. Acronyms for variables are: Averaged minimum temperature (Tm), Altitude (Alt), Potential evapotranspiration (PET), Absolute minimum temperature (Tma), Averaged maximum temperature (TM), Mean temperature (T), Slope, Aspect, NVDI and Distance to the nearest house (Houses).

Taula 2. Llista de les 12 espècies de control i el nombre de presències dubtoses associades. Les variables incloses en el model predictiu i el poder explicatiu associat en la forma de $100 \cdot r^2$. Els acrònims són: Mitjana de les temperatures mínimes (Tm), altitud (Alt), potencial d'evapotranspiració (PET), temperatura mínima absoluta (Tma), Mitjana de les temperatures màximes (TM), temperatura mitjana (T), pendent, aspecte (orientació vessant), NVDI i distància a la casa més propera (Houses).

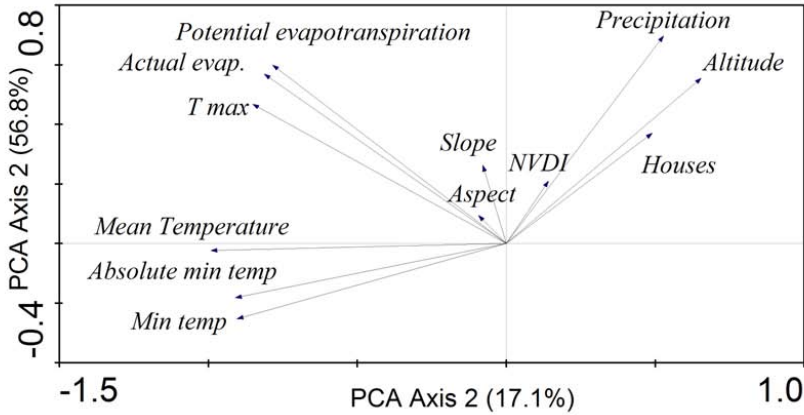


Fig. 8. Plot of principal component analysis of the 12 explanatory variables. The angle between two vectors is proportional to the correlation degree (e.g., Actual and potential evapotranspiration are highly correlated). Note, however, that variable vectors pointing in the same direction and opposite sense indicate that they are highly but inversely correlated (e.g., Min temp and Altitude).

Fig. 8. Gràfic de components principals de les 12 variables explicatives. L'angle entre els dos vectors és proporcional al grau de correlació (e.g. l'evapotranspiració actuals i la potencial estan altament correlacionades). Observeu que variables amb vectors amb la mateixa direcció però de sentit oposat indiquen que les variables tenen una alta correlació, tot i que inversa. (e.g. Min temp i Altitude).

procedure described in Fig. 5, in order to detect false absences and unstable presences. The number of doubtful observations detected ranged between zero and six (Table 2). The environmental variables selected as the best predictors, and the variance explained by such models, depended on the species (Table 2). The principal component analysis of the explanatory variables showed that some variables were highly correlated (Fig. 8). For example, the human impact-related variable (distance to the nearest house) was significantly correlated with altitude ($r=0.66$; $\text{prob.}<0.0001$). However, in the cases where more than one variable was included, they were not severely correlated.

The results of the vetting procedure are exemplified in the case of *Phylan semicostatus*. The denoised data ($n=45$) showed better accuracy ($\text{AUC}=0.80$) in

comparison to the raw data ($n=48$; $\text{AUC}=0.69$; AUC varies from 1 [perfect model] to 0.5 [random model]). Concerning the other species, AUC either remained equal (no doubtful observations were detected in one case) or improved after the removal of doubtful observations. The predicted landscape of probabilities of occurrence of *Phylan semicostatus* for the raw data and the recoded data are compared in Fig 9. Occurrence was assumed when the probability for a cell was above 0.5; the predicted prevalence decreased from 28% (raw data) to 19% (recoded data).

No significant model was obtained for *Iberellus balearicus*. This species was problematic because the recoding procedure detected 5 false absences. Therefore, the observed prevalence increased from 79% to 88%. This example seems to reinforce the need to limit the use of logistic models to

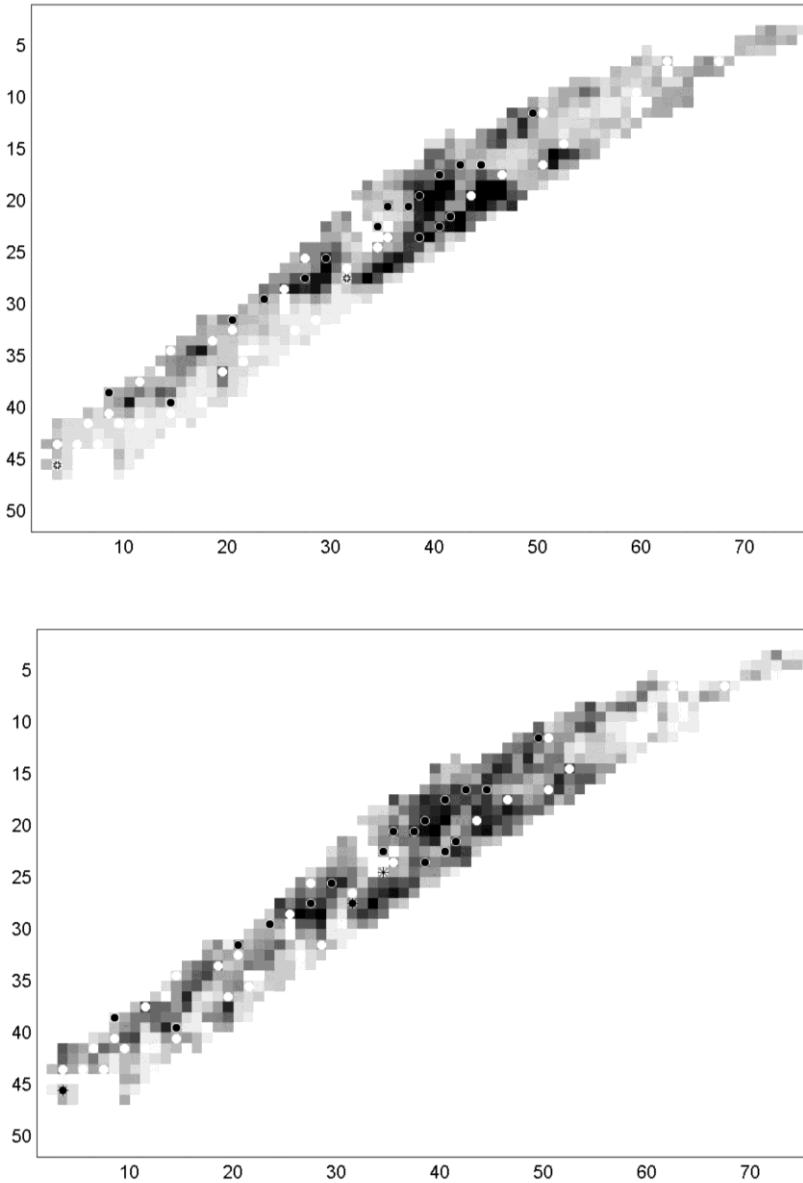


Fig. 9. Maps of predicted probabilities of occurrence of *Phylan semicostatus* for the denoised data (upper panel) and the raw data (lower panel). Observed presences are denoted by a black point and observed absences by a white point. Doubtful observations are marked with an asterisk.

Fig. 9. Cartografia de la probabilitat de presència de *Phylan semicostatus* per a les dades filtrades (mapa superior) i les dades brutes (mapa inferior). Les presències s'indiquen amb un punt negre i les absències amb un punt blanc. Les presències dubtoses es marquen amb un asterisc.

species showing medium prevalence because the few absences observed do not provide enough information for characterizing unsuitable environmental conditions. The remaining models were all significant (Prob<0.01). The variance explained by these models (estimated following the method proposed by (Lichstein, Simons, et al. 2002) varied from 7.2% (*Asida barceloi*) to 39.5% (*Harpactea dufouri*; Table 2). The predictive power of the model corresponding to *Asida barceloi* was considered to be too low. The average predictive power of the other 10 models (26.2%) was acceptable, especially when the stochasticity related to presence/absence data is taken into account. The cases of *Iberellus balearicus* and *Asida barceloi* were not considered in the next analytical steps.

Finally, these ten maps were combined into a single map that summarized the expected number of endemic species per cell (Fig. 10). The stasis of the predicted number of target species per cell in the entire mountain range is noticeable. This figure fell between 3.0 and 5.0 target endemics per cell in 92% of the area studied. It is not plausible that this stasis resulted from an artifact related to the complex methodological approach adopted here because the method correctly reproduced computer simulated landscapes with areas displaying well-differentiated hot-spots of species richness.

The predicted data of the number of target species per cell did not suggest any clear relationship between diversity and environment. However, some groups of species can be recognized in regard to their spatial and ecological preferences. These species groups are: (i) conspicuous species (e.g., *Iberellus balearicus*), (ii) lowland species (with *Harpactea dufouri* being the clearest case), (iii) species that gradually

decrease in abundance toward the mountain tops and/or toward colder sites (e.g., *Euscorpis balearicus*, *Oxychilus lentiformis* and *Tudorella ferruginea*), (iv) mid-slope species (e.g., *Nesotes viridicollis*), and (v) highland species, typical of open habitats (*Phylan semicostatus* and *Asida planipennis*).

The four remaining target species displayed intermediate distribution patterns and can be included in more than one of the aforementioned categories. *Percus plicatus* deserves special attention because it shows two separate clusters of environmentally suitable cells. Interestingly, the morphological variability of this species (e.g., sculpture pattern of elytra) is noticeable, suggesting the possibility that differentiated populations share the studied area.

Correlation patterns in species composition

Concerning the spatial structure of species composition, the variogram of CA scores indicated a significant spatial correlation between cells placed at first lag interval (i.e., cells located 0 to 5 km apart; Fig. 11).

However, spatial autocorrelation in this lag interval was not significant in the variogram corresponding to the CCA residuals (i.e., the variability unexplained by the environment; Fig. 11). This suggests that spatial structure in species composition is related to a spatially structured environment (Wagner, 2005). This CCA included minimum temperature as an explanatory variable (the proprieties of this model are detailed below). Additionally, the variance explained by the environment plus the residual variance always fell within the expected 95% confidence interval. This suggests that the relationship between environment and species composition is

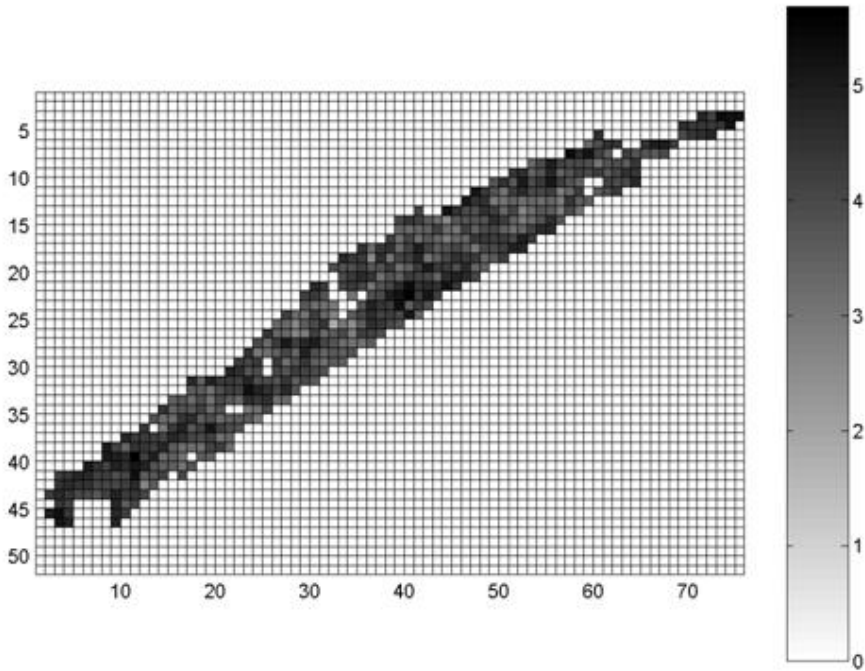


Fig. 10. Diversity (expected number of target species per cell) map combining the probability of the presence of ten target species.

Fig. 10. *Cartografia de diversitat combinant els mapes de probabilitat de les diferents espècies diagnòstic (nombre probable d'espècies diagnòstic per cel·la).*

scale-independent or, in other words, there is no evidence of regionalization (i.e., subareas with different relationships between species composition and environment; Wagner, 2005). Moreover, the variances explained tend to stabilize, which agrees that any (spatially structured) hidden environmental variable was left out of the model (Wagner, 2005). In summary, the analyses of the spatial structure suggest that patterns and significance tests of standard CCA are fully meaningful.

In contrast to the small variance found in species richness, variability in species composition was large (the total inertia of the raw data was 3.9). This suggests the occurrence of an important species turnover along the environmental gradient.

Concerning the relationship between species composition and the environment, a CCA with minimum temperature as the explanatory variable explained 7.6% of the total variability. The magnitude of the percentage of variance explained is apparently small. However, this figure is related to the response of 107 species, some of them slightly plausibly correlated with minimum temperature. In addition, presence/absence data does not allow a modulated response (as in the case of biomass, number of individuals or other abundance estimates). Therefore, the presence/absence data of a large number of species is prone to decrease the percentage of explained variance. Figures of variability explained below 10% are usual in this type

of study. However, the key point is that this figure is highly significant ($n=48$; trace=0.292; F-ratio = 1.775; Prob. value < 0.0002) and points to a substantial impact of the environment on species composition (ter Braak and Smilauer, 2002).

Discussion and concluding remarks

It has been claimed that ecological niche modeling (the correlative approach) only finds regions that environmentally resemble those where occurrences are currently

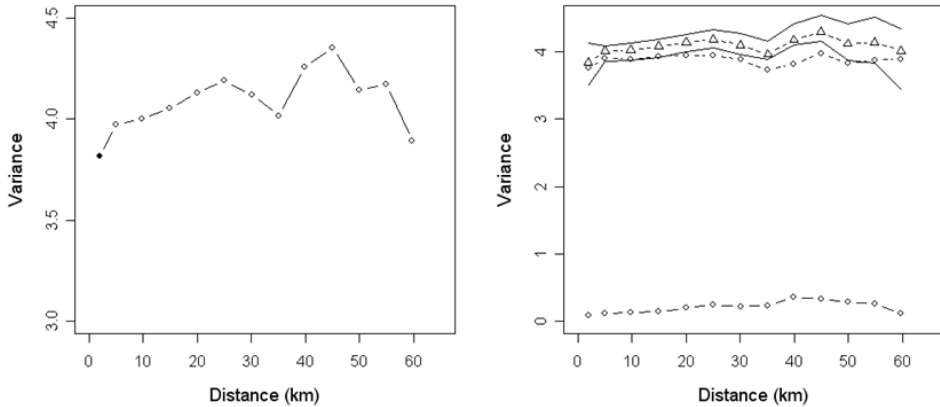


Fig. 11. Left panel: multivariate variogram of the raw data (i.e., results from correspondence analysis). Species composition is more similar than expected by chance at the first distance lag only (filled black point; probability <0.05 after applying the Bonferroni sequential correction). Right panel: multivariate variogram of the residuals extracted by a CCA model (the single explanatory variable included in the model was minimum temperature). The dashed line of points corresponds to the spatial variability in unexplained variance (the residuals of the CCA model). There is no spatial structure in the residuals (i.e., all lags show within-lag residual variance smaller than the expected by chance). The solid line of points (at the bottom) represents the spatial variability experienced by the variance explained by the CCA model. The envelope delimited by the solid lines defines the 95% confidence interval of the spatial variability of total variance (i.e., the CI for the variable displayed at the left panel but note the different vertical scale). The line of triangles describes the spatial variability of the sum of explained and unexplained variance.

Fig. 11. Gràfic esquerra: variograma multivariant de les dades brutes (i.e., resultats de l'anàlisi de correspondències). La composició de les espècies és només més semblant del que s'espera per atzar al primer interval de distàncies (punt negre; probabilitat <0.05 després d'aplicar la correcció seqüencial de Bonferroni). Gràfic dreta: variograma multivariant dels residuals producte del model CCA (l'única variable explicativa inclosa en el model és la temperatura mínima). La línia discontinua de punts correspon a la variabilitat espacial en la variància que no ha pogut ser explicada (els residuals del model CCA). No hi ha estructura espacial en els residuals. La línia sòlida de punts (a la base del panell) representa la variabilitat espacial explicada pel model CCA. Els límits que marquen les línies contínues delimiten l'interval de confiança del 95% de la variabilitat espacial de la variància total. La línia de triangles descriu la suma d'ambdues variàncies, l'explicada i la no explicada.

located (Soberón and Peterson, 2005). This approach is usually based on abiotic data only, when the occurrence pattern of a species also depends on biotic factors (i.e., presence/absence of other species). We propose that the vetting procedure described here (i.e., detection of false absences and unstable presences) represents not only a technical improvement (i.e., it increases the predictive power of statistical models) but also a theoretical one. The procedure is a way of introducing between-species interactions into the correlative approach by increasing the resemblance between the fundamental and the realized niche (after assuming that the entire geographical scenario can be reached by all the species). Note that we are not claiming the existence (and detailed description) of deterministic relationships (in a biologically explicit sense; e.g., host-parasite), but the existence of an empirical pattern resulting from the diffuse relationships of any individual species with a large number of other species in the community.

A second point is that Taxon-based diversity mapping is confirmed to be a valid alternative to conventional inventory-based diversity mapping. Inventory-based diversity mapping is a three-step process involving: 1) estimating species richness in a relatively small number of sites, 2) elaborating and evaluating a model relating species richness and environment, and 3) predicting species richness within a wider spatial area. In regard to the second step, no significant relationship has been found between species richness and environmental variables in our study, although these variables have been used successfully to predict species richness in other cases (e.g., Stefanescu *et al.*, 2004; Bhattarai *et al.*, 2004; Pei-Fen *et al.*, 2004; Nogués-Bravo and Martínez-Rica, 2004;

Mac Nally *et al.*, 2003). We argue that at relatively small spatial scales, the variability in species richness tends to be small. For example, the relationship between species richness and altitude is frequently unimodal (McCain, 2004), but this pattern most likely becomes undetectable when narrow altitude ranges are considered. Indeed, finding negative results when trying to demonstrate correlations between species richness and environmental variables is possibly the most frequent result of this type of analyses, although these negative results usually remain unreported.

A third interesting point is more specific to the case study presented here. The canonical view of a faunistic unit, as the one considered here, is that species composition experiences strong changes (i.e., large turnover), while species richness remains constant (or very stable). Consequently, conventional approaches fail to correlate species richness and environment. However, this stasis in species richness is neither related to low environmental variability (both the biotic and abiotic scenarios are diverse) nor to the lack of species-environment relationships (not only species composition but also species-specific responses are correlated with environment).

The diversity map obtained using Taxon-based diversity mapping showed that the number of endemic species remains stable along the entire mountain range considered, with two minor deviations. The first deviation concerns the decreased number of endemic species detected at some, but not all, lowland cells. Since, altitude and human-impact are correlated in the area studied (Fig. 8), it is difficult to attribute this decrease to human impact because effects of altitude and human

impact overlap (as in the case of Pei-Fen *et al.*, 2004). The surroundings of Pollença bay, a highly populated tourist center located in the NW of the area studied, is an example of lowland and extensively urbanized areas where this deviation occurs. The second deviation concerns the moderate decrease in diversity observed at the mountaintops. This decrease can be attributed to the reduced area located above 1000 m (the widely-accepted species-area relationship predicts a reduced species number in such a domain). It is also possible that the mountaintops were occupied by specialized species. These species could be locally abundant, but rare at the scale of the entire mountain range. Examples of rare species found only at the mountaintops are *Cyrtonus majoricensis* Breit and *Cryptocephalus tramuntanae* Petitpierre (Chrysomelidae, Coleoptera). In our study, all target species displayed a minimum prevalence of 20%. Consequently, it is possible that our selection was inadvertently against these specialized species.

These two minor deviations from the main trend do not invalidate the general statement that species richness remains insensitive to strong environmental gradients. Altitude in our case ranged from 100 to 1000 m. Vegetation cover changed from dense Oak tree and Pine tree woods in low to mid altitude, to open vegetation at the mountain tops. Human impact was diverse also, in spite of the exclusion of urban areas from the analysis. The fact that the number of endemic species remains constant along this environmental gradient invalidates the use of this crude number (as a surrogate of species richness) to identify biodiversity hotspots (geographical units of relatively small area and relatively large biodiversity) within the spatial scale considered (a 100 km mountain range).

Instead, the macroinvertebrate community studied showed important changes in species composition. This seems to be a general trend: species richness and other synthetic biodiversity indexes can experience small changes when under disturbance or other environmental constraints. For example, the number of species in some taxonomic groups remains insensitive to substantial habitat modification and maximum richness is not always reached in undisturbed habitats (Lawton *et al.*, 1998). The number of endemic species does not necessarily reflect pristine habitats because some endemic species can be found close to human settlements (Palmer *et al.*, 2004 and present study). Consequently, management decisions based on naive interpretations of biodiversity can be misleading. In contrast, the analysis of invertebrate species composition, especially if a broad taxonomical approach is used (Mattoni *et al.*, 2000), emerges as a suitable alternative to the correct interpretation of spatial and environmental gradients.

A final interesting point is that the area studied formed a unit from a faunistic point of view because it did not show any evidence of regionalization. Multivariate methods show that species composition remains correlated with the same set of environmental variables along the entire area. Spatial structure in species composition results from spatial structure in environmental variables. There is no evidence suggesting the existence of spatial autocorrelation in species composition due to population dynamics, species interactions or other biotic processes.

In summary, in cases like those presented here, managing decisions should not be founded upon species richness but on species composition. Additionally management decisions should include the entire distribution area because ranking of the

importance of some subsets of grid units in terms of diversity and species composition is not possible.

Acknowledgements

The authors thank Conselleria de Medi Ambient del Govern de les Illes Balears and Caixa de Balears "Sa Nostra" for their financial support, and the land owners of the sampling sites for allowing us to install pitfall traps on their properties. J. Monrde and M. Mascaró invaluablely helped us during the fieldwork.

References

- Bolòs, O. 1996. *La vegetació de les Illes Balears. Comunitats de plantes*. Institut d'Estudis Catalans. Barcelona.
- Fallot, P. 1922. *Etude géologique de la Sierra de Majorque*. Thèse d'état, Libr. Polytechnique Ch. Béranger, Paris and Liège. 480 pp.
- Gelabert, B. 1998. *La estructura geològica de la mitad occidental de la isla de Mallorca*. ITGME. Madrid.
- Gelabert, B., Fornós, J.J. and Gómez-Pujol, L. 2003. Geomorphological characteristics and slope processes associated with different basins: Mallorca (Western Mediterranean). *Geomorphology*, 52: 253-267.
- Guijarro, J.A. 1986. *Contribución a la bioclimatología de Baleares*. PhD Thesis. Universitat de les Illes Balears.
- Bhattarai, K.R., Vetaas, O.R. and Grytnes, J.A. 2004. Fern species richness along a central Himalayan elevation gradient, Nepal. *Journal of Biogeography*, 31: 389-400.
- Brown, J.H. and Lomolino, M.V. 1998. *Biogeography*. Sinauer Assoc., Sunderland.
- Buttel, L. A. and Durret, R. 2002. Competition and species packing in patchy environments. *Theoretical Population Biology*, 61: 265-276.
- Chave, J. and Levin, S. 2003. Scale and scaling in ecological and economic systems. *Environmental and Resource Economics*, 26: 527-557.
- Colwell, R.K., Rahbek, C. and Gotelli, N.J. 2004. The Mid-Domain effect and species richness patterns: What have we learned so far?. *The American Naturalist*, 163: 1-23.
- Couteron, P. and Ollier, B. 2005. A generalized, variogram-based framework for multi-scale ordination. *Ecology*, 86: 828-834.
- Diniz-Filho, J.A.F., Bini, L. M. and Hawkins, B. A. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, 12: 53-64.
- Fielding, A.H. and Bell, A.F. 2003. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24: 38-49.
- Fortin, M.J. and Payette, S. 2005. How to test the significance of the relation between spatially autocorrelated data at the landscape scale: A case study using fire and forest maps. *Ecoscience*, 9: 213-218.
- Gauch, H. G. and Whitaker, R. H. 1972. Coenocline simulation. *Ecology*, 53: 446-451.
- Guisan, A. and Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135: 147-186.
- Hawkins, B., Porter, E.E. and Diniz-Filho, J.A.F. 2003. Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology*, 84: 1608-1623.
- Hubbell, S.P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.
- Keitt, T H., Bjørnstad, O., Dixon, P. M. and Citron-Pousty, S. 2002. Accounting for spatial pattern when modelling organism-environment interactions. *Ecography*, 25: 616-625.
- Kerr, J.T. and Ostrovsky, M. 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology and Evolution*, 18: 299-305.
- Lawton, J.H., Bignell, D.E., Bolton, B., Bloemers, G.F., Eggleton, P., Hammond, P.M., Hodda, M., Holt, R.D., Larsen, T.B., Mawdsley, N.A., Stork, N.E., Srivastava,

- D.S. and Watt, A.D. 1998. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, 391: 72-76.
- Legendre, P., Dale, M.R.T., Fortin, M.J., Gurevitch, J., Hohn, M. and Myers, D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*, 25: 601-615.
- Legendre, P. and Legendre, L. 1998. *Numerical Ecology*. Elsevier Science B.V., Amsterdam.
- Lennon, J.J. 2000. Red-shifts and red herrings in geographical ecology. *Ecography*, 23: 101-113.
- Lichstein, J.W., Simons, T.R., Shiner, S.A. and Franzreb, K. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 73: 445-463.
- Liebholt, A.M. and Sharov, A.A. 1998. Testing for correlation in the presence of spatial autocorrelation in insect count data. In Baumgartner, J., Brandmayr, P. and Manly, B.F.J. (ed.), *Population and community ecology for insect management and conservation*. Bakelma, Rotterdam.
- Mac Nally, R., Fleishman, E., Fay, J.P. and Murphy, D.D. 2003. Modelling butterflies species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the Great Basin of western North America. *Biological Conservation*, 110: 21-31.
- Mattoni, R., Longcore, T. and Novotny, V. 2000. Arthropod monitoring for fine-scale habitat analysis: a case study of the El Segundo sand dunes. *Environmental Management*, 25: 445-452.
- McCain, C.M. 2004. The mid-domain effect applied to elevational gradients: species richness of small mammals in Costa Rica. *Journal of Biogeography*, 31: 19-31.
- Moisen, G.G. and Frescino, T.S. 2002. Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, 157: 209-225.
- Müller, R., Nowicki, C., Barthlott, W. and Ibsch, P.L. 2003. Biodiversity and endemism mapping as tool for regional conservation planning - case study of the Pleurothallidinae (Orchidaceae) of the Andean rain forests in Bolivia. *Biodiversity and Conservation*, 12: 2005-2024.
- Nogués-Bravo, D. and Martínez-Rica, J.P. 2004. Factors controlling the spatial species richness pattern of four groups of terrestrial vertebrates in an area between two different geographic regions in the northern Spain. *Journal of Biogeography*, 31: 629-640.
- Oksanen, J. 2005. *Multivariate Analysis of Ecological Communities in R: vegan tutorial*. R package at: <http://cran.r-project.org/>.
- Oksanen, J. and Minchin, P.R. 2002. Continuum theory revisited: what shape are species responses along ecological gradients?. *Ecological Modelling*, 157: 119-129.
- Oliver, I. and Beattie, A.J. 1996. Invertebrate morphospecies as surrogates for species: a case study. *Conservation Biology*, 10: 99-109.
- Palmer, M., Gómez-Pujol, L., Pons, G. X., Mateu, J., Meminn, M. And Rodríguez, A. 2002. *Cartografia de la distribució d'espècies endèmiques i bioindicadores a la Serra de Tramuntana: una aproximació des de la teledetecció i la geoestadística*. Conselleria de Medi Ambient, Govern Balear, Palma de Mallorca. Informe inèdit.
- Palmer, M., Gómez-Pujol, L., Pons, G. X., Mateu, J. and Linde, M. 2003. Noisy data and distribution maps: The example of *Phylan semicostatus* Mulasant and Rey, 1984 (Coleoptera, Tenebrionidae) from Serra de Tramuntana (Mallorca, Western Mediterranean). *Graellsia*, 59: 389-398.
- Palmer, M., Linde, M. and Pons, G. X. 2004. Correlational patterns between invertebrate species composition and the presence of an invasive plant. *Acta Oecologica*, 26: 219-226.
- Pei-Fen, L., Tzung-Su, D., Fu-Hsiung, H., Shu, G. 2004. Breeding bird species richness in Taiwan: distribution on gradients of elevation, primary productivity and urbanization. *Journal of Biogeography*, 31: 307-314.
- Putman, R.J. and Wratten, S.D. 1984. *Principles of Ecology*. University of California Press, Berkeley.

- Ribeiro, P.J. and Diggle, P.J. 2001. *GeoR: A package for geostatistical analysis*. Available for download at: <http://cran.r-project.org>. *R-News*, 1, 15-18.
- Ruggiero, A. and Kitzberger, T. 2004. Environmental correlates of mammal species richness in South America: effects of spatial structure, taxonomy and geographical range. *Ecography*, 27: 401- 416.
- Schlather, M., 2001. Simulation of stationary and isotropic random fields. *R-News*, 1: 18-20.
- Soberón, J. and Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2: 1-10.
- Stefanescu, C., Herrando, S. and Páramo, F. 2004. Butterfly species richness in the north-west Mediterranean Basin: The role of natural and human-induced factors. *Journal Biogeography*, 31: 905-915.
- ter Braak, C.J.F. and Smilauer, P. 2002. *CANOCO reference manual and user's guide to Canoco for Windows: Software for Canonical Community ordination*. Microcomputer Power, Ithaca, NY, USA.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. and Steininger, M. 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, 18: 306-314.
- Wagner, H.H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. *Ecology*, 84: 1045-1057.
- Wagner, H.H. 2005. Direct multiscale ordination with canonical correspondence analysis. *Ecology*, 85: 342-351.