An ED-based protocol for optimal sampling of biodiversity

JOAQUÌN HORTAL[*] and JORGE M. LOBO
*Departamento de Biodiversidad y Biología Evolutiva, MuseoNacional de Ciencias Naturales, C/ José Gutiérrez Abascal, 2, 28006 Madrid, Spain; *Author for correspondence (e-mail: mcnjh521@ mncn.csic.es; phone: 34 + 91 + 4111328; fax: 34 + 91 + 5645078)*

Abstract. While conservation planning requires good biodiversity data, our knowledge of most living groups is scarce and patchy even in well-sampled regions. Therefore, we need methodologies for rapid assessments for particular groups and regions. Maps of any biodiversity surrogate can be interpolated from even a few well-known sites, but such places are usually lacking. We therefore propose a protocol for designing field surveys to obtain good coverage of pattern variations of biodiversity in a given region. To represent biodiversity patterns comprehensively, we use a rule step site-allocation procedure, partially based on Faith and Walker's ED criterion that takes environmental and spatial variation into account, together with other criteria such as survey costs. A preliminary assessment of the adequacy of this site sampling strategy is made. Then a set of complementary sites is selected for further sampling. Using the ED criterion, during the stepwise process a *p-median* analysis is applied both to an environmental distance matrix and to a spatial distance matrix, to maximize the amount of variation covered by our survey planning. This rule-set allocation procedure is integrated into a continuous sampling design protocol directed to ensure we can sample all biodiversity of a region. This protocol requires the gathering of both biological and environmental information, an assessment of previously available information, the choice of sampling methods and dates, and a continuous assessment of the success of the survey being carried out. An example of the application of this protocol to the survey design of dung beetle (Coleoptera, Scarabaeoidea) diversity in the Comunidad de Madrid (Spain) is included.

Introduction

One of the main factors bearing on the current biodiversity crisis is the incompleteness of our knowledge of the patterns of the variety of life in most regions of the planet. In a rapidly changing world, knowledge of species

distribution patterns is necessary to improve the effectiveness of conservation policies, both for the selection of reserves and for their subsequent management (Austin and Heyligers 1989; Neldner et al. 1995). Usually, information about inventories is incomplete or lacking for most regions and living groups. However, with only a few reliable well-distributed local inventories (i.e. maximizing the representation of the assemblage variation in the region), it is possible to extrapolate maps of their estimated distribution over all the region (Margules et al. 1987; Bojorquez-Tapia et al. 1996; Iverson and Prasad 1998; Zimmermann and Kienast 1999; Hortal and Lobo 2001; Hortal et al. 2001; Lobo and Martín-Piera 2002; Lobo et al. 2002). So, a biodiversity assessment system, comprising surveys, extrapolation and reserve selection techniques is needed for conservation purposes (Austin 1998; Margules and Pressey 2000; Ferrier 2002; Ferrier et al. 2002a, b). A necessary component of any useful strategy to explore the main patterns of biodiversity within a region (i.e. species richness and composition of assemblages) is the accomplishing of systematic field surveys (Haila and Margules 1996; Ferrier 2002). However, funds directed to biodiversity sampling are often insufficient to maintain large-scale survey campaigns. So, new methods are needed to design rapid and accurate surveys that minimize resource consumption and obtain reliable information about distribution of unknown living organisms (Neldner et al. 1995; Barnett and Stohlgren 2003). Sampling methodologies specifically designed to obtain a comprehensive picture of biodiversity in a region are needed. Methods of allocating sampling points in a given territory have been directed either to select points randomly in space, to cover variation regularly in space (Southwood and Henderson 2000; Hirzel and Guisan 2002), to obtain truly statistically-independent samples (Davis and Goetz 1990; Pereira and Itami 1991), or to survey throughout all the environmental spectrum (stratified and gradsect methods; see Bunce et al. 1996 and Hirzel and Guisan 2002 for the former, and Austin and Heyligers 1989, 1991 for the latter).

If environmental diversity and biological diversity were positively related, the use of stratified and gradsect methods would be expected to recover most of the biodiversity of a region (Belbin 1993; Faith and Walker 1996; Wessels et al. 1998). However, both methods are based on partial representations of overall environmental variability (a classification, and the most important gradients, respectively). Further improvements on survey selection methods should try to use all environmental variability in the studied region. This requirement is accomplished by ED (Faith and Walker 1994, 1996), a site-allocation criterion to develop territorial selections minimizing the variation uncovered by the selected sites in the matrix of distances between all available locations. This matrix accounts for regional complementarity (that is, biodiversity variation throughout the region), either from species data or biodiversity variation surrogates (e.g., between-sites environmental and/or spatial distances). While data about the spatial distribution of biodiversity is usually scarce and biased, nowadays is quite easy to obtain and manage high-quality GIS environmental data. Thus, ED has been commonly used based on a description of environmental variability

between sites as a surrogate (although this description has also been built up from spatial distances, see below). Here, the *p-median* procedure was considered the best-possible alternative to allocate sites in a matrix of distances developed from a given surrogate of biodiversity. Although ED has been commonly used for reserve-selection procedures, it is also commonly applied in Australia to design field surveys using spatial information (Ferrier 2002; Faith 2003). As it tries to cover all the variability of the used surrogate in the studied area, ED constitutes a better approach to describe environmental variations in a region than stratified of gradsect methods.

A recent study showed that the ED criterion proposed by Faith and Walker (1996) to select areas for conservation to cover environmental diversity variation in a given region was not able to recover more species than expected by chance in some of the studied living groups at European extent (Araújo et al. 2001; see debate in Faith 2003 and Araújo et al. 2003), probably due to the existence of a geographic pattern different from a purely environmental one. This inconsistency may be due, first, to the fact that many different groups are analyzed jointly even through each one reacts to variations in environmental diversity in a different way; thus the matrix selected to describe the environmental variations in Europe at the spatial scale considered may not adequately describe these complex environmental requirements (see Faith 2003). Second, as Araújo and collaborators point out, many contingent factors, such as historical events and population-related processes, can modify the spatial distribution of biodiversity. These contingent factors are difficult to express as explanatory variables, but, if their effect varies over space, they generate a spatial structure in biodiversity that can be identified and modelized using the spatial location of each site (Legendre and Legendre 1998). So, in order to increase the probability of capturing all the variability in the biological response variable across the region, it will be convenient to maximize both the spatial and environmental coverage of our survey. This issue has been pointed out by Ferrier (2002), and has been applied to recent field survey campaigns in New South Wales, where the ED method applied to environmental and spatial distances (geo-sampling; Faith and Walker 1994) was used to allocate more than 4000 sampling points (see Ferrier 2002, Ferrier et al. 2002a, b and Faith 2003).

While surveying the environmental diversity of a region would allow us to record a great proportion of species present, such an approach can be improved in two main ways:

(i) ED has been commonly intended as a surrogate for overall biodiversity. However, as different environmental factors influence taxonomic groups differently, the selected explanatory variables and their relative importance should be specific to each taxonomic group.
(ii) Current species distributions may not be in equilibrium with current environmental conditions, due to historical and geographical contingency (Ricklefs and Schluter 1993). So, places with similar environmental conditions may not hold the same assemblages. Thus, when identifying

samples we must try to cover this non-environmental variability (see discussion in Ferrier 2002).

In this paper, we describe a hierarchical, stepwise and iterative allocation protocol to optimize the balance between the spatio-environmental spectrum sampled and the funds available for survey campaigns. Its rationale is partially based on heuristic reserve-selection techniques (Pressey and Nicholls 1989a), as well as and on the ED criterion (Faith and Walker 1994, 1996). The performance of this *p-median* approach in maximizing biodiversity coverage using environmental diversity as a surrogate has received moderate support in former studies (e.g. Ferrier and Watson 1997), being unable to obtain solutions that performed better than chance with several vertebrate groups (Araújo et al. 2001, 2003). However, it did covered better than previously used biodiversity surrogates (Faith 2003), so, although it does not produce statistically significant selections all times, it remains as the best-available approach to design a survey. Although biodiversity distribution may not be effected only by variations in environmental diversity, it is likely that environment influences the composition of species assemblages in some way, so ED applied to environmental distances still remains as a sound option for covering biodiversity patterns in absence of good biological data. The site allocation protocol described here must be considered as a practical application of the ED concept, in an attempt to answer the question raised by Faith (2003): How does ED could help us to make the best possible use of surrogate data? We use Faith and Walker (1994, 1996) algorithm to locate sampling localities, improving its implementation by using a weighted matrix of environmental distances specific to each taxonomic group, and also by maximizing the amount of spatial diversity covered, in addition to environmental diversity covered. The process of designing a survey of dung beetle distribution in a Spanish region is used to illustrate the method.

Environmental and spatial distance matrices

As our objective is to select sites to maximize the environmental diversity covered, taking into account spatial distribution, we apply the ED method using two different matrices of distance between sites (see Ferrier 2002), including: (i) an environmental matrix where the distances are weighted according to the environmental requirements of the taxonomic group, and (ii) a spatial matrix which includes the spatial distances between sites. Whilst the former accounts for between-site environmental differences, the latter does so for the spatial variation between sites with similar environmental conditions.

Building the environmental matrix

Former approaches to build environmental diversity matrices have used common distance indices, such as Euclidean distance, to obtain a metric

environmental space. However, although it presents some advantages, a metric Euclidean space is not necessary to work directly with the distance matrix. Moreover, to be useful in describing environmental preferences of each taxonomic group, the proximity measure used must address two main issues; (i) weighting the environmental variables used to build it (see discussion below), and (ii) including qualitative environmental descriptors (e.g. soil types or land cover categories), which are difficult to codify as continuous variables. Thus, we propose to use the Gower similarity index to build the environmental matrix. This distance measure allows to weight variables entered individually (Legendre and Legendre 1998), resulting in a weighted non-Euclidean space able to define similarity between sites, in which different weights are given to different variables (Gower 1971; Casgrain and Legendre 2001). This approach is alternative (and even complementary) to that of Faith and Walker (1996; Faith 2003), where ED includes the option of giving differential weights to concrete areas or sites (not to variables).

Environmental variables are often highly correlated. So, to minimize the collinearity among environmental variables, it is useful to ordinate each one of the main groups of environmental descriptors and reduce them into few new independent factors. These groups of environmental variables can be climate (precipitation, temperature, cloudiness, etc.), geomorphologic (digital elevation models and related structural variables), energy or resource availability variables (remote-sensing indices, food availability, feeding success, etc.), or substrate data (geology, soil, etc.), but any other requirement of the taxonomical group can be included. After defining the groups of environmental variables, each one is reduced to a new manageable set of independent, orthogonal, variables or factors by means of an ordination method, which can retain the gradient of variation present in the original variables (Faith and Walker 1996). For continuous variables, such as temperature, elevation or slope, a Principal Components Analysis (PCA) can be successfully used. However, as pointed out by Legendre and Legendre (1998), when the variables used are qualitative, or have been derived from qualitative information (e.g. proportions of each soil class in each area), PCA is not an appropriate ordination technique. In this case, a Principal Coordinates Analysis (PCoA; Gower 1966) is recommended. This analysis is similar to PCA but the resemblance matrix is constructed from a similarity index applicable to binary, qualitative, semiquantitative and quantitative descriptors together, such as the above-mentioned Gower's similarity index (Gower 1971). In both ordination analyses (PCA and PCoA), meaningless factors can be detected by a broken-stick method (Frontier 1976; Legendre and Legendre 1998). All meaningful factors for each group of variables are retained to build the environmental matrix. But there remains the question of which weights would be applied to these new variables or factors.

Former works have considered ED as a surrogate for overall biodiversity. However, when studying groups of environmentally or evolutionary related species, this approach can be refined. We thus argue that weights must come from previous hypotheses of the relative influence of each factor on the distribution of

the taxonomic (or functional) group that we are going to sample. These hypotheses can be formulated starting from: (i) the available biological and environmental data for the studied region at the scale to be considered; (ii) the available biological and environmental data for other regions and scales; (iii) the literature, or our basic empirical knowledge of the group.

If reliable faunistic information is available for a reduced set of well-sampled sites in the studied region, each factor can be explored separately to ascertain its effect on the variability of the taxonomic group at the scale and extent under consideration. Here, the best option is to use faunistic complementarity (e.g., variation in species composition, see Hortal et al. 2003) as a surrogate measure of variations in biodiversity. However, most times biological data is too scarce to obtain reliable descriptions of species composition patterns, so other biodiversity attributes could be used, such as species richness, rarity, endemicity or species richness. From them, species richness (i.e. number of species; Gaston 1996) seems to be a good option to determine the relative influence of environmental factors on overall biodiversity, due to both practical and theoretical considerations. On the one hand, it is easy to extract, and the less sensible measure to differences in sampling effort among the well-sampled territorial units. On the other hand, recent theoretical and experimental work supports species richness as a key property of ecosystems, supporting and/or being the outcome of processes driving to productivity, structure and resistance to alteration of biological communities (see Tilman et al. 1996, 1997; Yachi and Loreau 1999; Loreau et al. 2003; Jordano et al. 2003; Bascompte et al. 2003), being its spatial patters also influenced by historical processes (see Ricklefs 2004). Both issues stress the utility of species richness as a surrogate to study the spatial influence of biodiversity determinants.

The scores of this dependent variable can be related with each factor by means of General Linear Models (GLM; McCullagh and Nelder 1989; Dobson 1999), a broad group of analytic tools that can be used with a wide range of non-normal distributions for the random component, which include common linear regression analysis, as well as logistic regressions and many other similar analyses. To account for nonlinear relationships, we also suggest including the cubic function of the factor in a backward selection stepwise analysis, where non-significant terms in the equation are deleted. The variability in the dependent variable explained by significant terms is measured as their change in deviance from a null model (Dobson 1999). The percentage of deviance explained will be the relative weight of this factor in the process of building the environmental matrix. An example of the use of GLM modelling with variations in species composition can be found in Hortal et al. (2003). This process must be carried out for the factors extracted from all the groups of environmental variables. Recently developed dissimilarity modelling (GDM; see Ferrier 2002, Ferrier et al. 2002b; Faith and Ferrier 2002), which analyzes directly the distance matrices, may produce a better description of the relationships between biodiversity and environment, even with scarce data, and thus may be a better option to obtain the relative group weights.

When insufficient sites are well sampled, we can estimate the relative importance of each factor using the biological and environmental data from other regions, other scales, or other geographical extents ('external data'). Here, the main drawback is the assumption that the relative influences are similar on our working scale or region, which may not be true. If no empirical data are available, the relative weights of each environmental group can be hypothesized taking into account existing literature or field knowledge of the ecological requirements of the taxonomic group. If any *a priori* hypothesis can be formulated, the weight applied to all groups must be equal to one. With the information provided by the first sampling campaigns, these weightings can be adjusted. In later sampling evaluations (see Stage 6 below), factor weights can be recalculated when enough sites become well-sampled, yielding a new and more accurate matrix of environmental distances.

Building the spatial distance matrix

Environmentally similar localities may hold different species assemblages due to unique contingent events (e.g. historical or population-related). To incorporate this spatial variability we use the 'geo-sampling' proposed by Faith and Walker (1994), that is, ED applied to geographic space (see discussion in Ferrier 2002), by means of between-site Euclidean distances (spatial matrix). However, this spatial variability would be better described by a spatial distance matrix in which connectivity (i.e. differential costs of moving through spatial patches) is considered (Ferrier 2002). Thus, as the cost of movement differs among taxonomic groups, as well as its degree of connectivity or isolation, these differences can be expressed as cost surfaces, which characterize the difficulty in moving across each given patch in the region for the species of the studied group (see Ferrier et al. 2002a). This cost surface can be developed using GIS variables such as land cover, elevation, etc.

Multi-criteria allocation of sample sites

Once environmental and spatial matrices have been defined, the survey of biodiversity patterns requires maximizing together the amount of environmental and spatial coverage sampled. This should be expected to improve the accuracy of total biodiversity estimates and inventories of a given region, and also give reliable information about its distribution in the spatio-environmental multivariate spectrum of the region, using the smallest number of samples possible.

The p-median allocation procedure

As pointed out by Faith and Walker (1996), the *p-median* procedures (Church and Sorensen 1994; Church 2002) in which the ED method is based seem to be

the best-available method to maximize gains in environmental and spatial coverage. They involve determining the spatial location of a given new feature (such as a hospital) in such a way as to minimize the distance from each target point of the defined space ('demand points', e.g. the streets of a city) to the nearest feature location ('locality sites', e.g. any hospital in the city). That is, the spatial location where the sum of distances from all 'demand points' to the next 'locality site' is lowest. These distances can be calculated in many ways, depending on the purpose of our feature allocation. In the hospital example, this distance can be calculated as the amount of time an ambulance spends to reach the hospital from each city street using the street network and taking into account the lengths of the streets and the speed of traffic in each one.

Two main approaches have been used to solve the *p-median* problem for selecting areas to maximize the environmental variability recovered: (i) greedy algorithms (Faith and Walker 1996; Faith 2003), where new features are added iteratively to the former set of selected areas, and (ii) optimal algorithms (Araújo et al. 2001, 2003), where a unique optimal solution is found, and all desired features are added in a single step. The latter may produce a better coverage of the overall variability present in the region, as greedy algorithms (that is, take the best-choice at each step) may produce sub-optimal solutions (Araújo et al. 2003). However, the greedy approach presents two main advantages for the systematic selection of areas for survey in real-world bio-diversity assessment, where most times both funds and time are limited. On the one hand, these methods allow to maximize the biodiversity coverage as more funds or time to sample more areas are being available. Site-selection can be easily carried out by means of a stepwise procedure, where the new sampling sites produce direct improvements in the gains in biodiversity coverage obtained in the previous sampling campaigns. On the other, in such a stepwise approach, it is easy to determine the moment when the cost (in time and/or funds) of surveying more areas would not improve our biodiversity coverage enough to be worthy. Depending on which other areas have been previously selected, a given area's complementarity value can be more or less than its cost (Faith et al. 1996). Thus, we recommend the greedy *p-median* approaches for stepwise survey planning processes.

Another issue when dealing with *p-median* algorithms is choosing between continuous and discrete forms of describing the environmental space (Faith 2003). Briefly, whilst the former treats the environmental space as a continuum between the extremes present in the region, the latter restricts this space only to the environmental conditions present in the sites (spatial locations) used as 'demand points' (in our case, all territorial units in the region; see Faith et al. 1996, Faith 2003; Araújo et al. 2003 for a discussion on both *p-median* forms). The continuous *p-median* is tied to the unimodal response of each species to environment model (Austin 1987), which may not be true (see Ferrier et al. 2002a). Assuming the environmental space as a continuum would bias the selection of areas to the representation of inexistent conditions, preventing us of maximizing the coverage of actual biodiversity patterns

present in the region (Araújo *et al.* 2003). On the contrary, discrete *p-median* accounts for the actual environmental (or spatial) variability of the studied area. This approach may produce the effect of clustering the selected areas (Faith 2003), which could potentially reduce species representation in a theoretical scenario where the spatial responses of the species are uniformly distributed along geographic gradients. However, geographic variations in the distribution of biodiversity are the effect of these spatial responses to local habitat conditions and historical and population processes (realised niches), which modify their potential response to a full environmental spectrum (Araújo et al. 2003). The assumption of a unimodal response model is not needed in discrete *p-median* applications. Most applications of this allocation criterion aim to obtain a selection based in distances obtained from real spatial information of all demand points in the studied area, independently of the shape of the relationship between distances and so-obtained (see review in Church 2002). Thus, an approach based on actual conditions present in the studied localities, rather than on the full variation in the environmental spectrum, should be a better descriptor of the patterns of biodiversity variation throughout the studied region.

To maximize the cost/effectiveness of a survey in describing biodiversity patterns, the selection of areas to be sampled in a given region must try to reflect the main compositional variations present in that region. Thus, at least for a systematic survey planning directed to identify and describe biodiversity patterns in a given region, it seems better to use a discrete *p-median* form to select the TUs that are going to be surveyed, and we use such an approach (see description below). Studies aimed to characterize potential ecological responses of the species to environmental variations may obtain better gains using the continuous *p-median* form, although the spatially-dependent variation present in the data must be taken into account to separate the effect of non-environmental spatially-structured factors (see Borcard et al. 1992, Legendre and Legendre 1998, Lobo et al. 2001, 2002, and Legendre et al. 2002).

Integrating different criteria in a rule-step selection procedure

Since each single step of the greedy algorithm involves the use of different criteria (e.g. maximizing both environmental and spatial coverage), a procedure able to take all of them into account is needed. A complete allocation procedure can be expressed as a set of iterative rules, each pertaining to different criteria, based on heuristic reserve-selection techniques (see Pressey and Nicholls 1989a and references therein). ED has been commonly used as part of such multi-criteria analyses (see Faith et al. 1996). Here, our target points are poorly-sampled sites, and the starting features are those considered well-sampled, a set iteratively enlarged to include sites to be sampled one by one. Explicit rules make the method easily repeatable and understandable (Neldner et al. 1995). These rules may be chosen from the following, with the order altered at will,

depending on the purposes of the study or our previous knowledge of the distribution patterns of the group in the region, and the previous survey effort carried out there.

Rule 1. Maximize the environmental variation recorded. This rule accounts for the environmental influence on the distribution of the group. An ED procedure is applied to the environmental distance matrix. First, the environmental distance from each i site (demand point) to the set of s previously selected and well-sampled ones ($ed_{i,s}$) is obtained as the distance of site i to the nearest one of the s group. Uncovered environmental variation is calculated as the sum of all these distances in the n unselected sites ($\sum_{i=1}^{i=n} ed_{i,s}$). Then, the site h that minimizes the sum that remains when incorporated in the previous s set ($\sum_{i=1}^{i=n-h} ed_{i,s+h}$) is selected. If there are more than one site falling within the confidence interval given by this minimum $\sum_{i=1}^{i=n-h} ed_{i,s+h}$ and the standard deviation of all $\sum_{i=1}^{i=n-h} ed_{i,s+h}$, they are also selected and enter the next rule.

Rule 2. Capture extreme environments. Since biodiversity is highly sensitive to extreme environments, where species rare to the region are expected to appear, the environmentally most-distant-site (the one that accounts for the maximum $ed_{i,s}$) is chosen. The confidence interval is defined here as the range between this maximum $ed_{i,s}$ and the standard deviation of all $ed_{i,s}$. All sites in this interval pass to the next rule.

Rule 3. Maximize the spatial spectrum recorded. Non-environmental processes, such as historic or population-related events, modify the response of the group to the environment. Such response may be different among localities with similar environmental conditions in the region. To consider such possible differences, the space covered by the sampling method is maximized again with a *p-median* method, where we choose the site that minimizes the sum of distances to the set of selected sites ($\sum_{i=1}^{i=n} sd_{i,s}$) with its inclusion. Again, all those sites included in the confidence interval of the minimum $\sum_{i=1}^{i=n-h} sd_{i,s+h}$ are retained to be tested in the next rule.

Rule 4. Minimize sampling costs. To take into account one of the more pressing issues of planning biodiversity sampling, the reduction of survey costs (Austin and Heyligers 1991; Missios 1998), different rules can be used. First, it is useful to concentrate survey effort on those sites already, though poorly, sampled. So, we will select the ones that account for greater previous sampling effort. A second rule can be the choosing of those sites where, with similar potential gains in biodiversity coverage, the costs of access, due to road density or distance to the research institutions, are lowest.

Additional Rules. Depending on the purposes of the survey, other rules can be implemented. For example, if we want to take into account human habitat alteration, the mean alteration of each site can be calculated easily as a mean score, using a GIS land-use coverage, after assigning a degree of alteration to each land-use class. Then, the matrix of metric distances between the degree of alteration of each site may be used for the *p-median* selection procedure. Finally, if there is more than one site selected after applying the last rule, one of them may be chosen at random.

The remaining non-selected sites enter again into the first rule of the set. This process is repeated until a stopping rule identifies the distribution of sampling as satisfactory. This stopping rule must take into account two main issues: availability of funds for surveying, and completeness of spatio-environmental variation of the region covered (see Faith et al. 1996 and Faith 2001b for a discussion on the use of ED and costs). That is, we must stop if the maximum number of sites that we have funds for sampling has been selected or if the $\Sigma_n ED_i$ and the $\Sigma_n SD_i$ are small enough. The latter condition can be assessed by studying the relationships between the number of places selected and the uncovered $\Sigma_n ED_i$ and $\Sigma_n SD_i$ that remain after each inclusion (trade-off curves). As the slope of these curves becomes smaller, it becomes progressively more expensive to increase our spatio-environmental coverage. Thus, we stop selecting sites when the potential benefits of increasing our sampling are low, that is, when the slopes of both curves are small. A more complete description of trade-off curves and the application of a more developed methodology to the selection of areas when funds are limited is given by Faith et al. (1996).

Sampling design

As mentioned above, to recover biodiversity distribution of a region, sampling design must cover as much as possible of both environmental and spatial variability as they are perceived by the group studied throughout the region. In addition, this sampling design must include a continuous sampling-level evaluation procedure, to determine which parts of the region have been correctly inventoried, and where correct inventories can be easily obtained from previous information. This evaluation procedure should be used after each sampling campaign to assess effectiveness, and identify those areas which need extra sampling effort. To account for these issues, we propose a six-stage procedure:

Stage 1: Compiling environmental and biological information

First, it is necessary to define exactly the extent (*sensu* Whittakker et al. 2001) and taxonomic group we are going to study. Then, all the relevant information available for this area and taxon must be compiled in a biological database and a spatially explicit environmental database stored in a GIS.

Nowadays, the creation of trans-national extensive data banks containing up-to-date information on biota distribution is becoming a reality, by means of the GBIF (Global Biodiversity Information Facility; see http://www.gbif.org and Edwards et al. 2000a). However, this is a complex and slow process, and for most territories this information must be still compiled in a standardized format (Bisby 2000a, b; Edwards et al. 2000a, 2000b; Smith et al. 2000). When studying a given region and taxon, it may be necessary to create the biological

database from scratch, so we must first develop a database to include all taxonomic and distributional information available. This information is often incomplete and heterogeneous, scattered over sources such as scientific publications with many different objectives, or Natural History Museums and private collections. The structure of such databases is often complex (see for example Biotica, in http://www.conabio.gob.mx/biotica/, or Biota, in http://www.gobcan.es/medioambiente/biodiversidad/ceplam/bancodatos/biota.html), but it is useful to include, at least, the following data fields: date of capture or observation; place (including spatial location coordinates in a common reference system, such as Geographic (lat/long) or Universal Transverse Mercator (UTM)); relevant ecologic data (habitat type, feeding, altitude, host species, etc.); number and sex, if possible, of specimens; capture or observation method, collector and identification responsible; place of storage (for Natural History collection specimens) or bibliographic reference; and other available useful data, such as genetic sequences or morphotype. Although aggregating the information about the variation in one of this fields (e.g., different dates in a single year) would result in saving disk-space and funds during the compilation process, reducing the information stored would prevent us to make the best-possible use of such an information.

Conversely, the development of environmental databases is usually an easier task. There is an increasing amount of geographically referenced environmental information worldwide. Many countries and trans-national institutions own GIS data banks, which centralize the environmental information gathered by different projects, as well as its origin and development methodologies (see García Hernández and Bosque Sendra 2001 or http://www.grid.unep.ch/data/grid). If this information is available, although scattered, a previous compilation and treatment may be necessary to develop the environmental database of the studied area. To characterize the environmental spectrum of the region, the environmental database must include, at least, the following data at the best-possible resolution: climate data, including common variables such as mean annual temperature or mean annual precipitation, as well as others which may be important for the studied group; geomorphologic data, including a digital elevation model and derived variables (slopes, aspects, etc.); substrate data, such as geology, soil type and composition or hydrology; land use/land cover data; and other information about the concrete taxon environmental requirements. All this information must be digitized, and fitted onto the same cartographic base, that is, same extent (map borders), spatial resolution (pixel width), and reference system (e.g. UTM or Geographic), so the studied region is described homogeneously.

When the biological database and the environmental database are available, it is necessary to define the Territorial Units (TUs) used to study the region, i.e. the spatial resolution of our survey planning. These units must be the smallest possible, taking into account: the accuracy of the spatial locations in the biological database; the spatial resolution of the environmental database; the extent of the studied region; the sampling objectives; and the investment

available to carry out the sampling. It is also advisable to choose spatially homogeneous TUs, kept to a similar shape and area over all the region. The best option is to use common reference grids, such as those based on geographic coordinates (e.g. 0.5-degree grid) or those derived from the UTM system (e.g. 5 km UTM grid). However, irregular TUs, such as vegetation patches or administrative divisions, can be used when necessary, taking care of the consequences due to differences in size and shape among the selected TUs (see Fortin 1999).

Stage 2: Assessing sampling-effort level

Once all available distributional information has been compiled, it is necessary to determine the accuracy of the inventories for each TU contained in the biological database. Collector's curves can provide a good assessment of the level of knowledge reached with present sampling-effort (see Figure 1). In these curves, also called species accumulation curves, the cumulative sampling effort carried out in an area is related to the number of species that are being sequentially added to its inventory. This relationship is fitted to a function, which relates the species accumulation with the sampling effort carried out (Soberón and Llorente 1993; Colwell and Coddington 1994; León-Cortés et al. 1998; Moreno and Halffter 2000; Gotelli and Colwell 2001). These curves have been used as a tool to predict total species richness by extrapolating the function found to its asymptote, but its utility is still being debated (see for example Colwell and Coddington 1994; Gotelli and Colwell 2001; Willott 2001; Moreno and Halffter 2001; Hortal et al. 2004). However, it is well known that they are very useful in determining the present rate of finding new species for the inventory. The slope of the curve at each point determines the species accumulation rate at that sampling level. As our inventory becomes more complete, it becomes more infrequent to register the presence of an unknown species, so this rate (curve slope) decreases. The higher the slope of the curve, the greater the sampling effort necessary to obtain a good inventory. However, when this rate is small great sampling effort cannot add significant numbers of species to the inventory of a given TU. The missing species are those locally rare or vagrants (see Moreno and Halffter 2000), so the inventory of this TU can be considered reliable enough, although incomplete.

Collector's curves require a spatially and temporally homogeneous and comparable sampling effort unit. In the case of standardized surveys, this unit is easily defined as hours/person, traps/day, and so on. Unfortunately, distributional information contained in a database such as the biological database described, which comes from heterogeneous sources, does not lead to a unit of this kind. As usually there is no information on the sampling effort carried out when no specimen capture or observation was made (i.e. the survey effort carried out with no results), it is necessary to examine the biological database to define a measure of sampling effort, possibly as the number of days spent
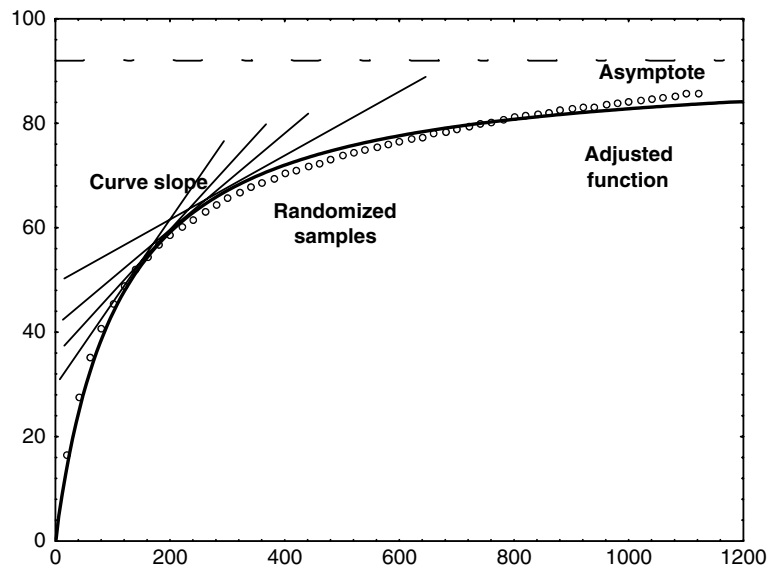
Figure 1. Species accumulation curve for the historic dung beetle inventory (Col., Scarabaeoidea) collected in the UTM 10 km^2 grid square 30TVL11 (Cercedilla, Madrid, TU number 18; J. Lobo, F. Martín-Piera and J. Hortal, unpublished data). *X* axis shows the sampling effort carried out (*n*; in this case, number of records in SCAMAD 1.0 database). *Y* axis shows the number of species found in the grid cell for each level of sampling effort (S_n). The order the samples enter in the curve has been randomized 500 times using the EstimateS 6 program (Colwell 2000). Circles: randomized curve (only 1 of each 20 records is shown). Continuous line: mathematical function fitted to the curve (Clench equation; $S_n = [0.828 \cdot n/1 + (0.009 \cdot n)]$). Dotted lines are the successive lines tangent to this function as the sampling effort carried out raises, that is, the slope of the curve in each sampling effort level. Finally, the horizontal straight line (dots and lines) shows the asymptote predicted by the adjusted function. Jiménez-Valverde and Hortal (2003).

sampling per person and TU, the number of specimens captured, the number of database-records, and so on.

To assess survey quality on each TU we must fit a function that describes the increase in the number of species added to the inventory as the sampling effort increases. As the shape of this curve may vary depending on the order of sample accumulation, an 'ideal' smoothed curve may be obtained by randomizing this order (Colwell 2000). Although many functions have been proposed to describe this relationship, the negative exponential function, the Clench equation, and the Weibull function have often been used for this task (Soberón and Llorente 1993; Colwell and Coddington 1994; Flather 1996; Fagan and Kareiva 1997; Moreno and Halffter 2000).

Once the function has been fitted, it is necessary to decide when this inventory is complete enough to consider a TU as well-sampled. Although there are many promising approaches for this task (Gotelli and Colwell 2001;

Christen and Nakamura 2003), they are still preliminary or too complex to be easily used. A simple and promising measure of inventory completeness is the slope of the curve at present sampling-effort level. As no studies have explored this issue, it is practical to identify as well-sampled all TUs slopes lower than an arbitrary cut-off value, such as 0.05 (i.e. one species each 20 units of sampling effort).

Stage 3: Selecting units to sample

As explained above, in this stage the key issue is to sample as much of the spatio-environmental variability as possible by using the iterative ED assignation procedure (Faith and Walker 1994, 1996). To maximize our knowledge of the biodiversity in the region, the goal is to complete the set of well-sampled TUs, maximizing the coverage of the spatio-environmental spectrum perceived by the taxonomic group, by means of weighted matrices of environmental and spatial distances. To select each TU, the iterative rule-set multi-criteria decision method proposed before is applied. The best TU to be sampled is selected at each step by following the previously defined sequential set of rules, which may try to maximize the effectiveness of surveys both economizing funds and covering potential biodiversity. This process is repeated until a stopping rule indicates that the distribution of sampling is complete enough, either when no more funds are available or when the non-sampled spatial and environmental spaces fall below a previously defined cut-off point.

Stage 4: Selecting sampling points on each territorial unit

Within each TU, a variable number of sampling points must be selected to cover its spatio-environmental variability. Here, a balance between optimum TU size and the dispersal capacity of the group must be struck. If the TUs are large, a procedure similar to that described in the previous stage could be used to allocate sampling points inside them. Otherwise, if TUs are small enough, simple land cover or physioclimatic-regions maps can be used to place sampling points covering all the different habitats present. A road and/or path network can be overlaid on these maps, helping to place each sampling point in easy-access sites.

Stage 5: Choosing field sampling techniques

At each sampling point we must try to collect as many species of the group as possible. Using a single survey method usually restricts the proportion of the species of a given group that can be captured. Then, it is advisable to use as many different kinds of traps, baits or survey methods as possible, taking into

account the costs of increasing sampling effort. We can define a sample unit as the combined effort made with methods that have shown to be successful (in the literature or in a previous study; see an example in Jiménez-Valverde and Lobo 2004). Furthermore, the number of sampling units used also determines survey success. Usually, literature and field knowledge about the group provide a criterion to decide how many samples must be used to balance increasing effort with costs. When this knowledge is not available, it is possible to carry out preliminary tests to determine the number of samples that maximizes the benefits and minimizes the costs. As discussed before for each TU (see Stage 2), the relationship between sampling effort (here, number of sampling units per sampling point) and the increase of the captured number of species can be represented by a species accumulation curve, from which the point of diminishing returns can be identified. Lastly, the samples must be distributed over time to account for the seasonal variation of the group.

Stage 6: Sampling success evaluation

To increase survey success, it is necessary to evaluate and improve sampling design process. This step allows us to use the knowledge we have acquired to maximize the cost/benefit ratio of the subsequent survey campaigns.

After carrying out each campaign, results must be evaluated, determining the degree of success. When these results are included in the biological database, a new species accumulation curve should be plotted for each sampled TU, to assess the completeness of the new inventory (see Stage 2). If the TU is considered sufficiently well sampled, we can assume that this inventory is complete. However, a new survey campaign must be carried out of those TUs where knowledge remains incomplete, until each selected TU can be defined as well sampled. The sampling effort in the TUs that remain incompletely sampled must be examined to determine improvements for future campaigns, by either increasing the number of sampling sites, survey methods, and samples in each site or sampling dates.

As mentioned before, if the factor weights used for the environmental diversity matrix calculation were not extracted from data at the same scale and extent, this matrix should be recalculated from newly well-sampled TUs. Of course, all TUs that had been raised to well-sampled status, or even all those that had been sampled in the first campaign, may then be included in the initial set of the ED selection procedure (see Stage 3).

A practical example

To illustrate this methodology we present the process of designing a survey to capture distribution patterns of dung beetles (Coleoptera; Scarabaeoidea) in the Comunidad de Madrid, an Iberian region.

Step 1: Information compilation

The Comunidad de Madrid (CM) is a Spanish Autonomous Community, a triangle-shaped territory, nearly 8000 km^2 in extent located in Central Iberia (Figure 2). Its northern angle lies at 41°8′ N, whilst its southernmost point is located 140 km away, at 39°52′ N in the Tajo valley. The triangle base is about 125 km long, between the longitudes 4°31′ W and 3°6′ W. Although the mean altitude is around 800 m.a.s.l., it ranges from 2430 m.a.s.l. in the "Sistema Central" mountain range to 434 m.a.s.l. in the Alberche valley. Moreover, its geological history is highly heterogeneous, leading to the present heterogeneous lithology, that includes acid-rock mountains (granite and gneiss), a ramp of acidic and coarse-grained sands, many alluvial, fine-grained soils at lowlands, and a clay, limestone and gypsum soil plateau. Thus, in spite of its small extent, CM is a highly heterogeneous region; its habitats represent almost all of those of the Iberian Central Basin. A preliminary survey design is necessary to capture all Scarabaeoidea diversity patterns that may have been produced by this environmental heterogeneity and the complex history of Central Iberia.

Scamad is a biological database that compiles all the available information for CM and its surrounding territories from Museum, University and private collections, as well as the distributional information published in the literature. This information is geo-referenced when possible, with at least a 10 × 10 km resolution. Iberian Scarabaeoidea dung beetles are a taxonomically well-known group (Martín-Piera 2000; López-Colón 2000; Veiga 2000), which comprises the species of three families: Scarabaeidae, Geotrupidae and Aphodiidae.

Figure 2. Geographic location and shape of the Comunidad de Madrid region.

The information used to construct the environmental database comes from heterogeneous sources. The CM digital elevation model was extracted from a Global DEM with 1 km spatial resolution (Clark Labs 2000a). Slopes and Aspects for each 1 km cell where calculated using Idrisi 32 GIS software (Clark Labs 2000b). Thirty-year mean monthly values of precipitation and maximum and minimum temperatures for 41 stations of Central Iberia where obtained from an agroclimatic atlas (Ministerio de Agricultura, Pesca y Alimentación 1986). This data provided mean annual and summer precipitation, mean, maximum and minimum annual temperatures, and temperature range, and then interpolated onto 1 km spatial resolution maps using a distance weighted-average method (INTERPOL module of Idrisi; Clark Labs 2000b). An 11-category lithology map was also digitized from a CM Atlas (ITGE 1988), also at a 1 km spatial resolution. A land cover map at a 1 km spatial resolution was obtained by reclassifying and enlarging the 250 m European Land Use/Land Cover map provided by the CORINE programme (European Environment Agency 1996). A 1:200.000 phytoclimatic regionalization of Spain was digitized for the CM territory (Rivas-Martínez 1987). Finally, we obtained additional cartography, such as roads, rivers, or administrative limits from the digital version of the CM 1:200.000 map (Servicio Cartográfico de la Comunidad de Madrid 1996). All digital maps where reprojected to a UTM 30N reference system when necessary.

The environmental database spatial resolution was 1 km, that is, 100 pixels for each territorial unit of 10 km^2. On the other hand, although many biological records were referred to 1 km UTM coordinates, nearly all the biological records in SCAMAD could be referred to a 10 km resolution grid, so choosing a 1 km scale may have led to the lose of some useful biological information. Moreover, many dung beetle species can fly several kilometres a day, so at a smaller scale this high vagility would obscure biodiversity patterns. Thus, it makes sense to use the 10 km UTM grid to define the TUs we are going to use. The 108 UTM 10 × 10 km grid squares that had more than 5% of their territory in CM where chosen as TUs (Figure 3a).

Step 2: Sampling-effort assessment

In this step, we analyze the initial version of SCAMAD (v. 1.0) to determine which TUs can be considered well-sampled enough to obtain a good picture of its dung beetle fauna. First, this database was examined to ascertain which sampling-measure unit could be extracted for further analyses. Although many records contained date of capture, others lack this information. In addition, the origin of records was highly heterogeneous (standardized samplings, non-professional collections, etc.), so different dates also result in differences in sampling effort. Thus, to homogenize this information, using nearly all records, we define a sampling unit as each record in the biological database (i.e. each time a species is recorded by a different method or collector, regardless of the number or sex of specimens). To give rise to a new sampling unit, these records must differ at least
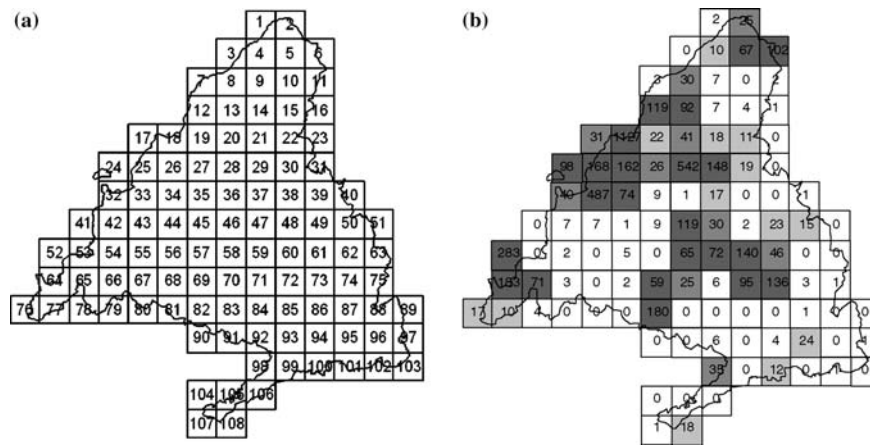
Figure 3. (a) Ten km² UTM cells of the Comunidad de Madrid selected as territorial units (TUs) for dung beetle inventory assessment. (b) Number of database records in SCAMAD (a database of the Scarabaeoidea dung beetles of the Comunidad de Madrid) for each 10 km² UTM grid square. White: less than 10 database records; Light Grey: 10–24 database records; Medium Grey: 25–49 records; Dark Grey: more than 50 records. The species accumulation curves have been carried out only for this latter group (see Table 1).

in one of the following SCAMAD database fields: capture date, place of capture, habitat type, feeding, capture or observation method and collector.

Five thousand three hundred and sixty four (5364) database records for 146 species were available at the beginning of the sampling design (on SCAMAD 1.0), although only 5237 could be referred to any one of the 108 TUs (Figure 3b). We estimated the 'ideal' species accumulation curves only for those TUs with more than 50 database records, randomizing the order of entrance of each record in the curve 500 times, and calculating the mean values of richness at each cumulative-effort-step using the EstimateS program (Colwell 2000). Then, we tried to fit these curves both to the negative exponential function and to the Clench equation. The latter gave better results, so we used it to fit the randomized curves, calculating the final slope of the fitted function for each TU (Table 1). Preliminary results showed that the asymptote values predicted for the curves as records increase become regular at a slope value of 0.1, so we used this score as a cut-off point. The TUs with final slopes equal to or lower than 0.1 where selected as well-sampled (Table 1, Figure 4), constituting the initial set of the following selection procedure.

Step 3: Selecting territorial units to sample

To define our selection rules we assumed, as a preliminary hypothesis, that dung beetle diversity in CM is influenced by three main groups of environmental variables (climate, topographic and geologic), and by many contingent
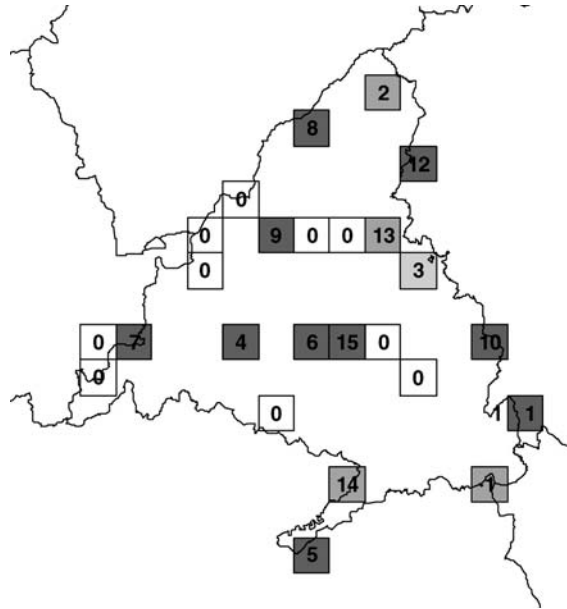
Figure 4. Territorial units of the Comunidad de Madrid chosen by the iterative rule set selection procedure taking into account both environmental and spatial information (see text and Table 5). Numbers are the step at which each TU was selected, and colours the final rule (criteria) used for each selection. White: initially well-sampled (rule 0); Dark Grey: environmental *p-median* (ED applied to environmental distances; rule 1); Medium Grey: number of records (rule 2); Light Grey: spatial *p-median* (ED applied to geographic distances; rule 4).

factors that may have produced a spatial pattern. In addition, we must take into account funds availability. So we defined six iterative rules to select the TUs that may be sampled (see the multi-criteria allocation of samples section). If more than one TU is selected by a rule, they enter into the next one:

0. Select as initial set all the well sampled TUs (i.e., with final slopes lower than 0.1).
1. Select the TU that minimizes the non-recorded environmental spectrum, by means of applying an ED allocation procedure to the matrix of environmental distances. This rule aims to increase biodiversity coverage through sampling the full environmental spectrum (see *Rule 1* in the *Multi-criteria allocation of sample sites* section).
2. Select the TU with most database-records. This way, field costs are reduced, as less survey effort (which will result in new database records) is needed to raise a good inventory of a given TU when previous information is available (see *Rule 4*).
3. Select the TU environmentally more distant from the previously selected set, i.e., the one at a maximum distance from the selected set in the matrix of environmental distances. Here, the aim of the rule is to sample extreme

Table 1. Species accumulation curves fitted for all TUs (10 km² UTM cells of the Comunidad de Madrid) with more than 50 database records (see Figure 4).

UTM Cell		SCAMAD		Curve Fitting			Results	
N°	Name	S_r	r	A	B	VE (%)	Asymptote	Final slope
5	30TVL54	34	67	1.082	0.0165	99.96	65.6	0.2442
6	30TVL64	41	102	1.122	0.0170	99.82	66.0	0.1502
12	30TVL22	43	119	1.021	0.0153	100	66.7	0.1283
13	30TVL32	37	92	1.078	0.0182	99.97	59.3	0.1511
18	**30TVL11**	**86**	**1127**	**0.828**	**0.0090**	**99.28**	**92.0**	**0.0067**
24	30TUL90	38	98	1.080	0.0182	99.99	59.4	0.1397
25	**30TVL00**	**50**	**168**	**0.988**	**0.0140**	**99.99**	**70.8**	**0.0884**
26	30TVL10	56	162	1.040	0.0123	99.99	84.5	0.1160
28	**30TVL30**	**43**	**542**	**1.118**	**0.0246**	**99.86**	**45.4**	**0.0054**
29	**30TVL40**	**29**	**148**	**0.960**	**0.0264**	**99.96**	**36.3**	**0.0398**
33	**30TVK09**	**89**	**487**	**0.861**	**0.0079**	**99.56**	**108.5**	**0.0364**
34	30TVK19	36	74	1.013	0.0150	99.95	67.6	0.2278
46	30TVK38	45	119	0.798	0.0096	99.89	83.4	0.1746
52	**30TUK77**	**41**	**283**	**0.779**	**0.0155**	**99.73**	**50.2**	**0.0268**
58	30TVK37	41	65	0.982	0.0087	99.97	112.6	0.4001
59	30TVK47	39	72	0.940	0.0110	99.98	85.1	0.2918
60	**30TVK57**	**36**	**140**	**1.034**	**0.0215**	**99.99**	**48.1**	**0.0643**
64	**30TUK76**	**50**	**183**	**0.954**	**0.0137**	**99.95**	**69.4**	**0.0772**
65	30TUK86	39	71	0.905	0.0094	99.95	96.6	0.3264
69	30TVK26	27	59	0.923	0.0174	99.98	53.1	0.2248
72	30TVK56	41	95	0.861	0.0109	99.84	79.0	0.2078
73	**30TVK66**	**30**	**136**	**0.979**	**0.0259**	**99.87**	**37.8**	**0.0480**
82	**30TVK25**	**51**	**180**	**0.964**	**0.0134**	**100**	**72.2**	**0.0832**

S_r and r are, respectively, the number of species observed and the number of database records extracted from SCAMAD 1.0 (the database of Scarabaeoidea dung beetles of the Comunidad de Madrid). A and B are fitted scores for the parameters of the Clench Equation ($S_r = A \cdot r / 1 + B \cdot r$; Soberón and Llorente 1993). *VE* is the percentage of variance explained by the fitted equation. The predicted asymptote was calculated as A/B, following Soberón and Llorente (1993). TUs identified as well sampled are highlighted. TU numbers as in Figure 3.

environments, which may host species rare in the region due to their selection of marginal habitats (see *Rule 2*).

4. Select the TU that maximizes the spatial spectrum recorded, that is, apply an ED allocation procedure to the matrix of spatial distances. This rule and the following one aim to cover faunistic differences due to the non-environmental spatial replacement (see *Rule 3*).

5. Select the TU at a maximum geographic distance from the selected set using the matrix of spatial distances.

6. Select just one TU randomly from all that have reached this rule.

To define the environmental distance matrix, we first extracted from the environmental database the scores of the environmental variables for each TU, assigning each one to the three environmental groups of variables previously defined (see Table 2). All environmental variables were first mean centered. Then, a PCA was applied to the climate and topographic group of variables.

Geologic variables were derived as proportions from a categorical map, so a PCoA was carried out for this group of variables. Significant factors are chosen using the broken-stick method (Table 3). PCA and PCoA calculations were made using STATISTICA (StatSoft 1999) and R-Package (Casgrain and Legendre 2001), respectively.

As there were only 10 well-sampled TUs, and they were unevenly distributed (see Figure 4), the relative weights of each group of variables could not be extracted from the Madrid dataset. No preliminary hypothesis was available on the magnitudes of these weights, so the environmental and biological data were extracted from a previous study of the Iberian Scarabaeinae species on a spatial scale of 50 km (see Lobo and Martín-Piera 2002). We modeled species richness with the variables that belong to each environmental group of variables using a GLM analysis, testing the deviance from a null model (Table 4). The explained deviance for each group of variables was taken as its respective weight.

The environmental distance matrix was calculated using the selected significant factors of each group of variables, and the Gower's similarity index as the measure of resemblance. The weight applied in the Gower's similarity index

Table 2. Environmental variables included in each one of the three environmental groups of variables defined.

Groups of Variables	Variable name	Definition
Climate		
	Summer precipitation	Average value in the TU
	Annual precipitation	"
	Maximum annual temperature	"
	Mean annual temperature	"
	Minimum annual temperature	"
	Annual temperature variation	maximum minus minimum temperature
Topographic		
	Mean aspect	Average value in the TU
	Mean slope	"
	Slope variation	Maximum minus and minimum slopes
	Aspect diversity	Shannon-Wiener index of the frequencies of each
	Slope diversity	aspect or slope class in the TU
	Minimum elevation	Minimum value in the TU
	Maximum elevation	Maximum value in the TU
	Mean elevation	Average value in the TU
	Altitude range	Maximum minus minimum elevation
Geologic		
	Soil geology diversity	Shannon-Wiener Index of the frequencies of each soil class in the TU
	Acid bedrocks	Surface covered
	Basic bedrocks	"
	Acid sediments	"
	Basic sediments	"

Table 3. Significant factors extracted in the ordination analysis carried out with the environmental variables of each variable group (Table 2).

Variables group	Factors	Eigenvalue	Explained variation (%)	Cumulative variation (%)
Climate				
	Factor 1	4.408	73.47	73.47
Topographic				
	Factor 1	6.490	72.12	72.12
Geologic				
	Factor 1	7.681	45.01	45.01
	Factor 2	2.736	16.03	61.04
	Factor 3	1.535	8.99	70.03
	Factor 4	1.336	7.83	77.87

For climate and topographic variables PCA factors were selected using the Broken-Stick criterion (see Legendre and Legendre 1998) whilst, in the case of the PcoA (Geologic variables), all factors with eigenvalues higher than one were selected (see Casgrain and Legendre 2001). In the case of PcoA a negative eigenvalue correction by Lingoes (1971) has been made (see also Casgrain and Legendre 2001).

was the variation explained by each factor multiplied by the weights of each group of variables previously calculated using all the 50 km UTM Iberian data. The spatial distance matrix was calculated as the Euclidean distances between the centroids of all TUs.

All 108 TUs entered in the rule-set selection procedure. Ten were identified as well-sampled, and the rest were tested for inclusion, being selected one by one (see process in Figure 3). After each inclusion in the survey group, we examined the curves that describe the amount of environmental and spatial distance uncovered as we add new TUs to this group (Figure 5). After 15 inclusions, the slope of both curves started to fit the horizontal thus reducing potential benefits in spatio-environmental coverage provided by the addition of new sites to the group of selected areas. In addition, according to the funds and staff available, it would have prevented us from carrying out and evaluating all surveys in the following two years. So we decided to sample only these 15 TUs (Table 5 and Figure 4). In all cases but one, the selection was tied to the first

Table 4. Goodness of fit of the GLM models accomplished by using the species richness of Scarabaeidae species in the Iberian UTM 50 km grid squares as dependent variable, and different groups of environmental variables in these squares (see Lobo and Martín-Piera 2002).

	df	dev	dev/df	Δdev	F	VE (%)
Null model	81	153.48	1.895			
Climate variables	74	107.78	1.456	45.70	31.37	29.77
Topographic variables	74	121.87	1.647	31.61	19.19	20.59
Geologic variables	77	136.47	1.772	17.01	9.60	11.09

df, degrees of freedom; dev, deviance; Δdev, change in deviance; VE is the percentage of variation explained by the models.
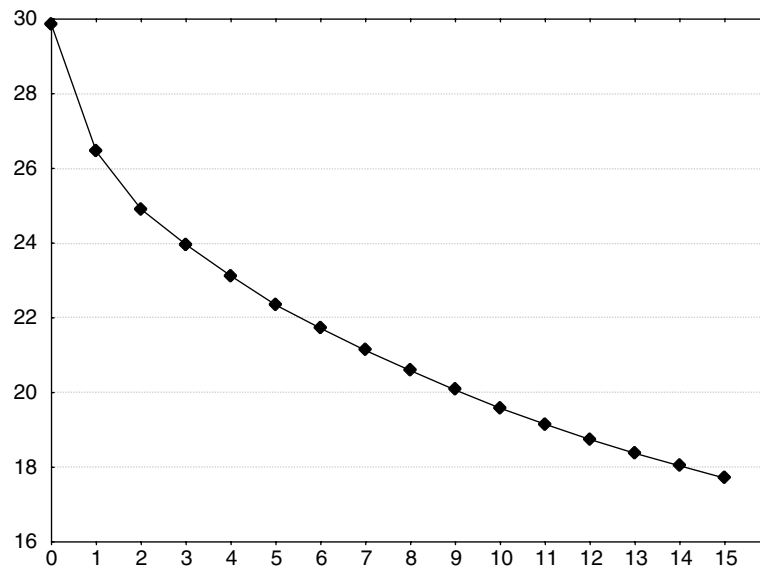
Figure 5. Decrease of the amount of uncovered environmental variability of the Comunidad de Madrid region after each inclusion of a new TU to be sampled. According to the ED criterion applied to environmental distances, uncovered environmental variability is measured as the sum of pairwise distances between each non-selected TU and the previously selected set. Step numbers (*X* axis) are as in text, Figure 4 and Table 5.

two rules, so the here-obtained selection was mostly based on environmental variation (Table 5). That is, the times when several TUs providing equivalent gains in environmental coverage were selected, sampling effort differences promoted the selection of one of them. In the only case were a geographic coverage was needed (Step 3), two contiguous Territorial Units (TU numbers 31 and 39, see Figure 3) passed the first three rules. TU 39 was selected with the fourth one (maximize spatial coverage), as TU 31 was closer to TU 29, which is part of the initial set (see Figure 4). However, both TUs were environmentally similar, so environmental variation also ruled out for this selection.

Step 4: Selecting sampling points within each territorial unit

TU resolution (10 km^2) was small enough to determine sampling points directly with the land cover data (European Environment Agency 1996) and phytoclimatic classification maps (Rivas-Martínez 1987). We also overlaid roads on both maps and determined three 1 km cells in each selected TU. Each cell was also placed as far as possible from the other two in order to account for non-ecological small-scale spatial patterns in species composition. As dung beetles feed and nest mainly on livestock droppings, they are likely to appear

near this food source. On the other hand, in the Mediterranean area, open biomes support much greater dung beetle abundance and richness than do closed biomes, which do not have exclusive species (Lumaret 1980; Lumaret and Kirk 1987). These issues were taken into account during field work, when sampling points were placed in the most suitable places inside those grid cells.

Step 5: Choosing sampling methodologies

As Iberian dung beetle populations present two main seasonal peaks, in spring and autumn (Martín-Piera et al. 1992; Lobo and Martín-Piera 1993; Lobo et al. 1997), the sampling program was designed to collect both spring and autumn inventories: each sampling point was sampled in both seasons. The baited pitfall traps recommended by Lobo et al. (1988) and Veiga et al. (1989) were used to sample dung beetle communities. In Mediterranean conditions very few baited traps are needed to give a good picture of the structure and composition of a dung beetle assemblage (Lobo et al. 1998). To collect the main assemblage composition, five traps placed 10 m apart and baited with fresh cow dung were set in each sampling site for a period of 48 h each season.

Step 6: Sampling success evaluation

After carrying out the field work in the selected TUs, it is necessary to evaluate both the reliability of the inventories obtained in each territorial unit, and the consistency of the regional inventory estimated with all the selected TUs. The first task can be accomplished incorporating the new biological information for each TU and recalculating the species accumulation curves; the second one using available parametric and non-parametric methods to estimate species richness (Colwell and Coddington 1994), comparing with the estimated and the recovered number of species.

In the case of CM dung beetles, two preliminary survey campaigns carried out in four southeastern TUs (numbers 89, 98, 102 and 108) located in an unsampled region (see Figure 3) have allowed us to find abundant populations of some species previously considered rare in Central Iberia, and the presence of others unknown in this area. The accumulation curves for each TU indicate that a high proportion of the local dung beetle inventory was captured. However, when the final slopes of these TUs were calculated they were much higher than the 0.1 cut-off point (Figure 6). This could mean that we are using a low number of (i) sampling places, (ii) sampling methodologies, or (iii) sampling dates, so we cannot capture reliable inventories. To overcome this drawback, in the following campaigns, a fourth sampling point was placed on each selected TU. Moreover, additional sampling was intended to be done throughout the year, by means of light traps or transects on foot capturing dung beetles by hand. Due to time and funds limitations, surveying the 15

Table 5. Summary of the iterative procedure to select those territorial units (TUs) to be sampled.

Step	TU	UTM designation	Selection rule	Value	r
0	28	30TVL30	0	0.0054	542
0	18	30TVL11	0	0.0067	1127
0	52	30TUK77	0	0.0268	283
0	33	30TVK09	0	0.0364	487
0	29	30TVL40	0	0.0398	148
0	73	30TVK66	0	0.0480	136
0	60	30TVK57	0	0.0643	140
0	64	30TUK76	0	0.0772	183
0	82	30TVK25	0	0.0832	180
0	25	30TVL00	0	0.0884	168
1	102	30TVK83	2	1	1
2	5	30TVL54	2	67	67
3	39	30TVK69	4	1526.65	0
4	56	30TVK17	1	23.12	5
5	108	30SVK31	1	22.34	18
6	58	30TVK37	1	21.72	65
7	53	30TUK87	1	21.12	0
8	8	30TVL33	1	20.59	30
9	27	30TVL20	1	20.07	26
10	63	30TVK87	1	19.58	0
11	89	30TVK95	1	19.15	0
12	16	30TVL62	1	18.74	1
13	30	30TVL50	2	19	19
14	98	30TVK43	2	35	35
15	59	30TVK47	2	72	72

The initial set of well-sampled TUs is listed at step 0. TU number as given in Figure 3. Selection rule refers to the criterion that finally selected the TU, and the numbers correspond to those described in the text (see *Selecting territorial units to sample*). Value is the score of the TU in this selection rule: rule 0 – final slope in the species accumulation curve, 1 – Sum of environmental distances to the selected set after this inclusion, 2 – Number of records in SCAMAD 1.0, and 4 – Sum of spatial distances to the selected set after this inclusion, in km. *r* is the number of database records compiled in SCAMAD 1.0 for each TU.

selected TUs required 2 years of field work. Although this work has been almost totally accomplished, it is still necessary to conclude the taxonomic determination of the collected material to include survey results in a new version of SCAMAD, and analyze current sampling-effort level in the sampled TUs as made with version 1.0 of this database (see *Step 2*). The so-obtained data is needed to evaluate the successfulness of the here presented sampling design protocol.

Discussion

Funds availability is always an important issue in recent survey design literature (Austin and Heyligers 1989, 1991; Neldner et al. 1995; Neave et al. 1997;
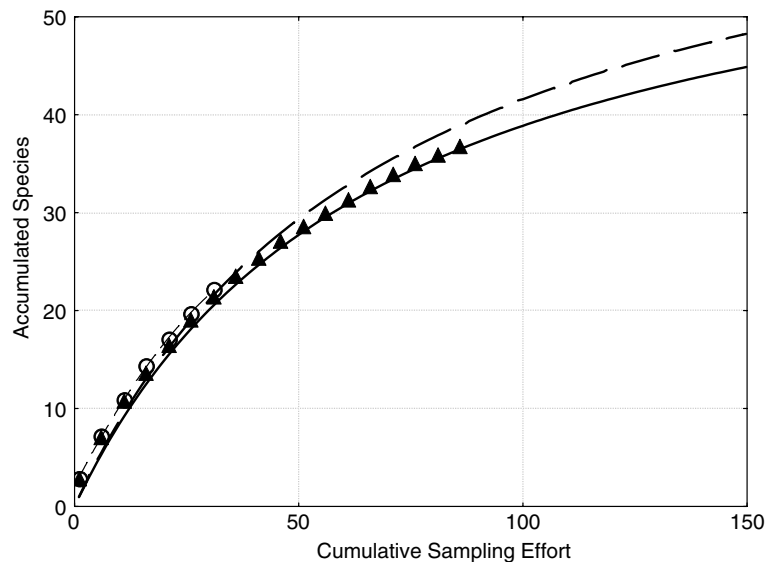
Figure 6. Collector's curves of the TU 98 (UTM code 30TVK43) before and after the first sampling campaign. Cumulative sampling effort is measured as the number of records stored in the SCAMAD 1.0 database. Both curves have been randomized 100 times (Colwell 2000; see text). Empty circles and discontinuous line correspond, respectively to the observed data (only each fifth record is shown) and the estimated curve before the sampling. Full triangles and continuous line do so for the observed data and estimated curve after the sampling was made.

Wessels et al. 1998; Austin 1998; Barnett and Stohlgren 2003). Fortunately, nowadays it is possible to use many methodologies to extend the information from some localities to the full extent of the region. Sampling the full range of environmental and spatial variation ensures that predictive models truly interpolate biodiversity variables in other environmentally similar localities, rather than extrapolate the scores of the predictive variable to localities with different environmental conditions (Austin and Heyligers 1989; Ferrier 2002; Ferrier et al. 2002a, b). Although it is not necessary to sample the whole territory, survey costs must be balanced against the completeness of the sampled variation of biodiversity. Our method aims to obtain better cost-effective surveys. It is possible to diminish the amount of sampling effort necessary to obtain a good picture of the diverse assemblages present there by: (i) using the previous knowledge about the inventories of a region; (ii) establishing obtained results feedback to the survey design; (iii) optimizing our survey design to both cover a greater range of variation of the studied group with fewer sampling points; (iv) locating them in easily accessible places.

In a recent simulation, Hirzel and Guisan (2002) found that data obtained by regular and equal-stratified (the same number of sampling plots per environmental region) sampling designs improved robustness and accuracy of

single-species predictive models. However, although in their simulation the 'real' distribution of a virtual species in a restricted studied area was a direct function of its environmental requirements, they found a higher effect of sample size in that estimates. When sampling natural patterns, unconsidered effects on species distributions (i.e. historic events, meta-population dynamics and stochasticity) also appear, an even higher sampling effort may be needed. This becomes even more important when surveying a group of species instead of just one. Although still untested, including both spatial location and environmental variation in site selection may produce better coverage of distribution patterns that random, regular or stratified samplings, diminishing the resource and time-consuming task of increasing sampling effort.

It has been widely accepted that covering all the environmental diversity in a region leads to the detection (or protection in the case of reserve selection procedures) of the full range of variation of biological diversity (but see debate in Araújo et al. 2001, 2003 and Faith 2003). From this picture, gradsect (Gillison and Brewer 1985; Austin and Heyligers 1989, 1991) and stratified (Bunce et al. 1996) methods seem to be a correct approach for biological diversity sampling. In general, environmental stratification provides a more efficient strategy than random or systematic sampling (Cochran 1977), because the latter would not lead to different types of landscape to being surveyed, resulting in a limitation of the heterogeneity sampled (Bunce et al. 1996). Kish (1965) pointed out that gradsect and stratified methods produce gains that are only small or moderate, an opinion supported by the results of Neave et al. (1997), who find no benefit in using these approaches instead of random sampling in estimating bird species richness in SE Australia. However, the results of Wessels et al. (1998) show clearly that both stratified and gradsect methods produced better coverage of the diversity of dung beetles and birds in a South-African reserve. These results are consistent with the intuitive idea that randomization is largely irrelevant where the principle survey objective is the detection of the maximum diversity possible (Gillison and Brewer 1985; Wessels et al. 1998; Faith 2003). Applying the ED criterion to environmental diversity may improve the results of these two former methods, as it takes into account all the variability among sites, not only land classes (stratified methods) or just a given environmental gradient (gradsect method).

However, environmental diversity alone may not be a good predictor of biodiversity (Araújo et al. 2001, but see Faith 2003; Araújo et al. 2003). As pointed out by Austin and Heyligers (1991), gradsect selection does not ensure adequate representation of diversity variation due to the absence of factors not included in the selection criteria, an effect that increases with increasing extents and spatial resolutions. Historic and population-dynamics events modify the pure environmental component of biodiversity distribution, leading to spatially modified patterns (see Ferrier 2002 for a discussion). So distributing our sampling points over both environmental and spatial spectrums would theoretically improve the range of diversity variation covered by the survey, thus giving a better representation of the assemblages present in the region. Ferrier

accounts for this using both environmental and spatial data in the generation of a single distance matrix through GDM, which is then used for selection processes.

The development of our selection procedure, however, relied largely on the results of the ED-environmental criterion used in the first rule. Apart from being the first per order of application, two factors may have promoted the importance of this rule: extent of analysis and spatial autocorrelation of environmental variables. On the one hand, Madrid is a small region (with maximum spatial distances of around 150 km), and only 108 TUs were used for the analysis. On the other, it constitutes an environmentally heterogeneous territory, where the coupled variation in altitude, climate, geology and land use has produced a few marked environmental gradients in a small geographic space, as well as the absence of similar habitats in separate areas. Thus, whilst spatial distances are not of much importance in Madrid, environmental conditions vary strongly between spatially close TUs. Analyses carried out at larger extents (e.g. the Iberian Peninsula or Europe) and/or environmentally homogeneous areas (e.g. Patagonian steppes) would take a better approach of geographic distance rules.

Although still incomplete, preliminary results stress the usefulness of the here-presented protocol. From the 108 defined TUs, 10 proved to be sufficiently well-sampled, and sampling only an additional 15 could provide a good spatio-environmental coverage. It seems that with only 60 sampling points (15 TUs $\times$ 4 sampling points in each TU), comprising 600 baited pitfall traps (5 traps per site $\times$ 2 season campaigns), a good knowledge of the distributions of dung beetle species and assemblages in Madrid can be obtained. A preliminary survey carried out in the selected TUs located farther from the well-sampled ones has detected both species and assemblages previously unknown or considered rare in Madrid (J. Hortal, J.M. Lobo, F.J. Cabrero-Sañudo and F. Martín-Piera, unpublished), one of the best-known regions in Iberia. However, once the complete results of this survey are available, an evaluation similar to the one carried out by Wessels et al. (1998) is needed to determine by how much this method improves on the results of prior ones.

We have used a heuristic algorithm for TU selection due to the need for a multi-criteria assessment. The main advances in this field are referred to reserve selection procedures, where the aim is to protect as much variation as possible of the attributes selected. As is pointed out by Pressey and Nicholls (1989b), including the different criteria iteratively and separately (i.e. the heuristic approach) gives better results than scoring them in order to obtain single values for each TU. Also, this approach involves simple computations, once all the criteria have been well defined and are available for calculation. Moreover, it allows indirect weighting of the importance of each criteria by means of their order of entrance in the analysis. Optimization algorithms, where all criteria are optimized at the same time instead of one by one, are a promising tool for this assessment, as nowadays they can be easily implemented using linear programming techniques, and computing time is relatively low with the

desktop computers available (Rodrigues and Gaston 2002). However, greedy selection methods, such as the one used here, provide an useful tool to develop an ordered list of sites, where each addition seeks for the best-possible improvement in the previous set. This step-by-step-gains rationale is most similar to the way that conservation policies (or sampling campaigns) are carried out: unfortunately, in many cases funds availability imposes that new sites are added in this fashion, once funds are being available.

This method may be useful for elucidating many of the present conservation insights focused on biodiversity. Once sampling campaigns have been improved, their final design can be repeated in the future in order to also sample temporal changes in biodiversity due to climate or land use changes. On the other hand, reserve selection procedures provide better results the more reliable the biological data. Furthermore, our method has been designed following a similar rationale to that of many reserve selection procedures. If it succeeds in allowing us to sample all diversity variation, including similar criteria in reserve selection protocols, such as BioRap (see Faith 2001a, b and references therein), may allow us to protect all diversity variation when both good biological data and funds for surveying are lacking. To explore this possibility, an evaluation of its potential as a reserve selection procedure, such as the Araújo et al. (2001) one, must be made.

Acknowledgements

References

Araújo M.B., Humphries C.J., Densham P.J., Lampinen R., Hagemeijer W.J.M., Mitchell-Jones A.J. and Gascet J.P. 2001. Would environmental diversity be a good surrogate for species diversity? Ecography 24: 103–110.

Araújo M.B., Densham P.J. and Humphries C.J. 2003. Predicting species diversity with ED: the quest for evidence. Ecography 26: 380–383.

Austin M.P. 1987. Models for the analysis of species response to environmental gradients. Vegetation 69: 35–45.

Austin M.P. 1998. An ecological perspective on biodiversity investigations: examples from Australian eucalypt forests. Ann. Missouri Bot. Garden 85: 2–17.

Austin M.P. and Heyligers P.C. 1989. Vegetation survey design for conservation: gradsect sampling of forests in north-eastern New South Wales. Biol. Conserv. 50: 13–32.

Austin M.P. and Heyligers P.C. 1991. New approach to vegetation survey design: gradsect sampling. In: Margules C.R. and Austin M.P. (eds), Nature Conservation: Cost Effective Biological Surveys and Data Analysis CSIRO, Australia pp. 31–36.

Barnett D.T. and Stohlgren T.J. 2003. A nested-intensity design for surveying plant diversity. Biodiver. Conserv. 12: 255–278.

Bascompte J., Jordano P., Melián C.J. and Olesen J.M. 2003. The nested assembly of plant–animal mutualistic networks. Proceedings of the National Academy of Sciences of the United States of America 100: 9383–9387.

Belbin L. 1993. Environmental representativeness: regional partitioning and reserve selection. Biol. Conserv. 66: 223–230.

Bisby F.A. 2000a. The quiet revolution: biodiversity informatics and the internet. Science 289: 2309–2312.

Bisby F.A. 2000b. Response to Databases tailored for biodiversity conservation. Science 290: 2074.

Bojorquez-Tapia L.A., Azuara I., Escurra E. and Flores-Villela O. 1996. Identifying conservation priorities in México through geographic information systems and modelling. Ecol. Appl. 5: 215–231.

Borcard D., Legendre P. and Drapeau P. 1992. Partialling out the spatial component of ecological variation. Ecology 73: 1045–1055.

Bunce R.G.H., Barr C.J., Clarke R.T., Howard D.C. and Lane A.M.J. 1996. Land classification for strategic ecological survey. J. Environ. Manag. 47: 37–60.

Casgrain P. and Legendre P. 2001. The R Package for Multivariate and Spatial Analysis, version 4.0d3.. Département de Sciences Biologiques, Université de Montréal Available at http://www.fas.umontreal.ca/BIOL/legendre/.

Christen J.A. and Nakamura M. 2003. Sequential stopping rules for species accumulation. J. Agric. Biol. Environ. Stat. 8: 184–195.

Church R.L. 2002. Geographical information systems and location science. Comput. Operation Res. 29: 541–562.

Church R.L. and Sorensen P. 1994. Integrating Normative Location Models into GIS: Problems and Prospects with the p-median Model. Technical Report, NGCIA, Available at http://www.ncgia.ucsb.edu/Publications/.

Clark Labs 2000a. Global Change Data Archive Vol. 3. 1 km Global Elevation Model. CD-Rom. Clark University.

Clark Labs 2000b. Idrisi 32.2. GIS software package, Clark University, CD-Rom.

Cochran W.G. 1977. Sampling Techniques, 3rd edn. Wiley and Sons, New York.

Colwell R.K. and Coddington J.A. 1994. Estimating terrestrial biodiversity through extrapolation. Philos. Trans. R. Soc. Lond. B 345: 101–118.

Colwell R.K. 2000. EstimateS 6.0b1.. Computer program and manual available at http://viceroy.eeb.conn.edu/Estimates6/.

Davis F.W. and Goetz S. 1990. Modeling vegetation–environment association using digital satellite and terrain data. Landscape Ecol. 4: 69–80.

Dobson A. 1999. An Introduction to Generalized Linear Models. Chapman and Hall/CRC, London.

Edwards J.L., Lane M.A. and Nielsen E.S. 2000a. Interoperability of biodiversity databases: biodiversity information on every desktop. Science 289: 2312–2314.

Edwards J.L., Lane M.A. and Nielsen E.S. 2000b. Response to Databases tailored for biodiversity conservation. Science 290: 2073–2074.

European Environment Agency 1996. Natural Resources CD-Rom. European Environment Agency.

Fagan W.F. and Kareiva P.M. 1997. Using compiled species list to make biodiversity comparisons among regions: a test case using Oregon butterflies. Biol. Conserv. 80: 249–259.

Faith D.P. 2001a. Overlap of species richness and development-opportunity does not imply conflict. Science Online 293: 1591a, Available at http://www.sciencemag.org/cgi/eletters/293/5535/1591.

Faith D.P. 2001b. Cost-effective biodiversity planning. Science Online 293: 2207a, Available at http://www.sciencemag.org/cgi/eletters/293/5538/2207.

Faith D.P. 2003. Environmental diversity (ED) as a surrogate information for species-level biodiversity. Ecography 26: 374–379.

Faith D.P. and Walker P.A. 1994. DIVERSITY: A Software Package for Sampling Phylogenetical and Environmental Diversity. Reference and User's Guide. v. 2.1.. CSIRO Division of Wildlife and Ecology, Canberra, Available at http://www.amonline.net.au/systematics/staff_faith.htm.

Faith D.P. and Walker P.A. 1996. Environmental diversity: on the best-possible use of surrogate data for assessing the relative biodiversity set of areas. Biodiver. Conserv. 5: 399–415.

Faith D.P. and Ferrier S. 2002. Linking beta-diversity, environmental variation, and biodiversity assessment. Science Online 295, Available at http://www.sciencemag.org/cgi/eletters/295/5555/636#504.

Faith D.P., Walker P.A., Ive J.R. and Belbin L. 1996. Integrating conservation and forestry production: exploring trade-offs between biodiversity and production in regional land-use assessment. Forest Ecol. Manag. 85: 251–260.

Ferrier S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? Syst. Biol. 51: 331–363.

Ferrier S. and Watson G. 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. Environment Australia.

Ferrier S., Watson G., Pearce J. and Drielsma M. 2002a. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodiver. Conserv. 11: 2275–2307.

Ferrier S., Drielsma M., Manion G. and Watson G. 2002b. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. Biodiver. Conserv. 11: 2309–2338.

Flather C.H. 1996. Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. J. Biogeogr. 23: 155–168.

Fortin M.J. 1999. Effects of sampling unit resolution on the estimation of spatial autocorrelation. Écoscience 6(4): 636–641.

Frontier S. 1976. Étude de la décroissance des valeurs propres dans un analyse en composantes principales: comparaison avec le modèle du bâton brisé. J. Exp. Mar. Biol. Ecol. 14: 217–224.

García Hernández E. and Bosque Sendra J. 2001. Bases de datos cartográficas de cobertura global accesibles 'on-line'. Geofocus (Recursos) 1: 5–10.

Gaston K.J. 1996. Species richness: measure and measurement. In: Gaston K.J. (ed), Biodiversity. A Biology of Numbers and Difference. Blackwell Science, Oxford, pp. 77–113.

Gillison A.N. and Brewer K.R.W. 1985. The use of gradient directed transects or gradsects in natural resource surveys. J. Environ. Manag. 20: 103–127.

Gotelli N.J. and Colwell R.K. 2001. Quantifying biodiversity: procedures and pitfalls in measurement and comparison of species richness. Ecol. Lett. 4: 379–391.

Gower J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325–338.

Gower J.C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27: 857–871.

Haila Y. and Margules C.R. 1996. Survey research in conservation biology. Ecography 19: 323–331.

Hirzel A. and Guisan A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecol. Model. 157: 331–341.

Hortal J. and Lobo J.M. 2001. A preliminary methodological approach to model the spatial distribution of biodiversity attributes. In: Mateu J. and Montes F. (eds), Spatio-temporal

Modelling of Environmental Processes: Proceedings of the 1st Spanish Workshop of Spatio-temporal Modelling of Environmental Processes. Publicacions de la Universitat Jaume I, Collecció "Treballs d'Informàtica I Tecnologia" 10, Castelló de la Plana, pp. 211–239.

Hortal J., Lobo J.M. and Martín-Piera F. 2001. Forecasting insect species richness scores in poorly surveyed territories: the case of the Portuguese dung beetles (Col. Scarabaeinae). Biodiver. Conserv. 10: 1343–1367.

Hortal J., Lobo J.M. and Martín-Piera F. 2003. Una estrategia para obtener regionalizaciones bióticas fiables a partir de datos incompletos: el caso de los Escarabeidos (Coleoptera) Ibérico-Baleares. Graellsia 59: 331–343.

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

ITGE 1988. Atlas Geocientífico y del Medio Natural de la Comunidad de Madrid. Instituto Tecnológico GeoMinero de España, Serie Medio Ambiente, Madrid.

Iverson L.R. and Prasad A.M. 1998. Estimating regional plant biodiversity with GIS modelling. Diversity and Distributions 4: 49–61.

Jiménez-Valverde A. and Hortal J. 2003. Las curvas de acumulación de especies y la necesidad de evaluar la calidad de los inventarios biológicos. Rev. Iber. Aracnol. 8: 151–161.

Jiménez-Valverde A. and Lobo J.M. 2004. Determining a combined sampling procedure for a reliable estimation of Araneidae and Thomisidae assemblages (Arachnida: Aranaeae). J. Arachnol., in press.

Jordano P., Bascompte J. and Olesen J.M. 2003. Invariant properties in coevolutionary networks of plant-animal interactions. Ecol. Lett. 6: 69–81.

Kish L. 1965. Survey Sampling. Wiley and Sons, New York.

Legendre P. and Legendre L. 1998. Numerical Ecology, 2nd edn. Elsevier, Amsterdam.

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.

León Cortés J.L., Soberón-Mainero J. and Llorente-Bousquets J. 1998. Assesing completeness of Mexican sphinx moth inventories through species accumulation functions. Divers. Distribut. 4: 37–44.

Lingoes J.C. 1971. Some boundary conditions for a monotone analysis of symmetric matrices. Psychometrika 36: 195–203.

Lobo J.M. and Martín-Piera F. 1993. Análisis comparado de las comunidades primaverales de escarabeidos coprófagos (Col., Scarabaeoidea) del archipiélago balear. Ecologia Mediterranea 3/4: 29–41.

Lobo J.M. and Martín-Piera F. 2002. Searching for a predictive model for Iberian dung beetle species richness based on spatial and environmental variables. Conserv. Biol. 16: 158–173.

Lobo J.M., Martín-Piera F. and Veiga C.M. 1988. Las trampas pitfall con cebo, sus posibilidades en el estudio de las comunidades coprófagas de Scarabaeoidea (Col.) I. Características determinantes de su capacidad de captura. Revue d'Ecologie et de Biologie du Sol 25: 77–100.

Lobo J.M., Sanmartín I. and Martín-Piera F. 1997. Diversity and Spatial Turnover of dung beetle (Coleoptera: Scarabaeoidea) communities in a protected area of South Europe (Doñana National Park, Huelva, Spain). Elytron 11: 71–88.

Lobo J.M., Lumaret J.P. and Jay-Robert P. 1998. Sampling dung beetles in the French Mediterranean area: effects of abiotic factors and farm practices. Pedobiologia 42: 252–266.

Lobo J.M., Castro I. and Moreno J.C. 2001. Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. Biol. J. Linn. Soc. 73: 233–253.

Lobo J.M., Lumaret J.P. and Jay-Robert P. 2002. Modelling the species richness distribution of French dung beetles (Coleoptera, Scarabaeidae) and delimiting the predictive capacity of different groups of explanatory variables. Global Ecol. Biogeogr. 11(4): 265–277.

López-Colón J.I. 2000. Familia Geotrupidae. In: Martín-Piera F. and López-Colón J.I. (eds), Fauna Ibérica 14. Scarabaeoidea I. Coleoptera, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid, pp. 105–183.

Loreau M., Mouquet N. and Gonzalez A. 2003. Biodiversity as spatial insurance in heterogeneous landscapes. Proceedings of the National Academy of Sciences of the United States of America 100: 12765–12770.

Lumaret J.P. 1980. Analyse des communautés de scarabéides coprophages dans le maquis Corse et étude de leur role dans l'utilisation des excréments. Ecologia Mediterranea 5: 51–58.

Lumaret J.P. and Kirk A.A. 1987. Ecology of dung beetles in the French Mediterranean region (Coleoptera, Scarabaeidae). Acta Zool. Mexicana 24: 1–55.

Margules C.R. and Pressey R.L. 2000. Systematic conservation planning. Nature 405: 243–253.

Margules C.R., Nicholls A.O. and Austin M.P. 1987. Diversity of Eucalyptus species predicted by a multi-variable environment gradient. Oecologia 71: 229–232.

Martín-Piera F. 2000. Familia Scarabaeidae. In: Martín-Piera F. and López-Colón J.I. (eds), Fauna Ibérica 14. Coleoptera, Scarabaeoidea I. Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid, pp. 205–432.

Martín-Piera F., Veiga C.M. and Lobo J.M. 1992. Ecology and biogeography of dung-beetle communities (Coleoptera, Scarabaeoidea) in an Iberian mountain range. J. Biogeogr. 19: 677–691.

McCullagh P. and Nelder J.A. 1989. Generalized Linear Models, 2nd edn.. Chapman and Hall, London.

Ministerio de Agricultura and Pesca y Alimentación 1986. Atlas Agroclimático Nacional de España. 2ª Ed. Dirección General de la Producción Agraria, Subdirección General de la Producción Vegetal, Madrid.

Missios P.C. 1998. Optimal sampling intensity in biodiversity prospecting and the financing of conservation. Annual Conference of the Canadian Economics Association, Available at http://www.feem.it/gnee/papers.html.

Moreno C.E. and Halffter G. 2000. Assessing the completeness of bat biodiversity inventories using species accumulation curves. J. Appl. Ecol. 37: 149–158.

Moreno C.E. and Halffter G. 2001. On the measure of sampling effort used in species accumulation curves. J. Appl. Ecol. 38: 487–490.

Neave H.M., Cunningham R.B., Norton T.W. and Nix H.A. 1997. Preliminary evaluation of sampling strategies to estimate the species richness of diurnal, terrestrial birds using Monte Carlo simulation. Ecol. Model. 95: 17–27.

Neldner V.J., Crossley D.C. and Cofinas M. 1995. Using Geographic Information Systems (GIS) to determine the adequacy of sampling in vegetation surveys. Biol. Conserv. 73: 1–17.

Pereira J.M.C. and Itami R.M. 1991. GIS-based habitat modelling using logistic multiple regression: a study of the Mt. Graham Red Squirrel. Photogramm. Eng. Rem. Sens. 57: 1475–1486.

Pressey R.L. and Nicholls A.O. 1989a. Application of a numerical algorithm to the selection of reserves in semi-arid New South Wales. Biol. Conserv. 50: 263–278.

Pressey R.L. and Nicholls A.O. 1989b. Efficiency in conservation evaluation: scoring versus iterative approaches. Biol. Conserv. 50: 199–218.

Ricklefs R.E. 2004. A crompehensive framework for global patterns in biodiversity. Ecol. Lett. 7: 1–15.

Ricklefs R.E. and Schluter D. 1993. Species diversity: regional and historical influences. In: Ricklefs R.E. and Schluter D. (eds), Species Diversity in Ecological Communities. The University of Chicago Press, Chicago, pp. 350–364.

Rivas-Martínez S. 1987. Memoria del Mapa de las Series de Vegetación de España. ICONA, Madrid.

Rodrigues A.S.L. and Gaston K.J. 2002. Optimisation in reserve selection procedures - why not? Biol. Conserv. 107: 123–129.

Servicio Cartográfico de la Comunidad de Madrid 1996. Mapa de la Comunidad de Madrid. Escala 1:200.000. Consejería de Ordenación Territorial, Comunidad de Madrid, Madrid.

Smith A.T., Boitani L., Bibby C., Brackett D., Corsi F., da Fonseca G.A.B., Gascon C., Gimenez Dixon M., Hilton-Taylor C., Mace G., Mittermeier R.A., Rabinovich J., Richardson B.J., Rylands A., Stein B., Stuart S., Thomsen J. and Wilson C. 2000. Databases tailored for biodiversity conservation. Science 290: 2073.

Soberón J. and Llorente J. 1993. The use of species accumulation functions for the prediction of species richness. Conserv. Biol. 7: 480–488.

Southwood T.R.E. and Henderson P.A. 2000. Ecological Methods, 3rd edn. Blackwell Science, Oxford.

StatSoft Inc. 1999. STATISTICA for Windows. Computer program manual. StatSoft Inc., Tulsa, OK.

Tilman D., Wedin D. and Knops J. 1996. Productivity and sustainability influenced by biodiversity in grassland ecosystems. Nature 379: 718–720.

Tilman D., Naeem S., Knops J., Reich P., Siemann E., Wedin D., Ritchie M. and Lawton J. 1997. Biodiversity and ecosystem properties. Science 278: 1865.

Veiga C.M. 2000. Los Aphodiinae (Coleoptera, Aphodiidae) Ibéricos. Unpublished PhD thesis, Universidad Complutense de Madrid, Facultad de Ciencias Biológicas, Dpto. de Biología Animal I.

Veiga C.M., Lobo J.M. and Martín-Piera F. 1989. Las trampas pitfall con cebo, sus posibilidades en el estudio de las comunidades coprófagas de Scarabaeoidea (Col.) II. Análisis de efectividad. Revue d'Ecologie et de Biologie du Sol 26: 91–109.

Wessels K.J., van Jaarsveld A.S., Grimbeek J.D. and van der Linde M.J. 1998. An evaluation of the gradsect biological survey method. Biodiver. Conserv. 7: 1093–1121.

Whittaker R.J., Willis K.J. and Field R. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. J. Biogeogr. 28: 453–470.

Willott S.J. 2001. Species accumulation curves and the measure of sampling effort. J. Appl. Ecol. 38: 484–486.

Yachi S. and Loreau M. 1999. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. Proc. Nat. Acad. Sci. USA 96: 1463–1468.

Zimmermann N.E. and Kienast F. 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. J. Veg. Sci. 10: 469–482.