

Designing a survey protocol to overcome the Wallacean shortfall: a working guide using bryophyte distribution data on Terceira Island (Azores)

Silvia C. Aranda^{1,2,3}, Rosalina Gabriel¹, Paulo A. V. Borges¹,
Eduardo Brito de Azevedo⁴, and Jorge M. Lobo²

¹ Azorean Biodiversity Group – CITA-A, Dep. Ciências Agrárias, Universidade dos Açores, 9700-042 Angra do Heroísmo, Portugal; ² Dep. Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC). C/ José Gutiérrez Abascal, 2. 28006, Madrid, Spain; ⁴ CMMG (Centro de Estudos do Clima, Meteorologia e Mudanças Globais), Universidade dos Açores, Dep. de Ciências Agrárias, 9700-042 Angra do Heroísmo, Portugal

ABSTRACT. The increasing availability of open access data on species occurrences is leading researchers to generate more hypotheses about patterns of species distributions. However, when all of this information is mapped onto a particular geographical scale, gaps usually appear due to lack of knowledge and sampling spatial bias (the so-called *Wallacean shortfall*). To overcome these problems as efficiently as possible, field surveys should be designed after distinguishing well-surveyed places from those with incomplete inventories in order to carry out the extra survey effort in those areas not represented environmentally and spatially by the well-surveyed places. This procedure requires (1) gathering, cleaning and standardizing data; (2) selecting environmental variables that are important for the group considered according to field experience and the literature; and (3) making statistical decisions about the number and location of areas that should be surveyed according to the available resources. Here, we summarize most concepts and procedures devoted to the evaluation of biodiversity data, offering some general recommendations on how to use them for optimizing new survey designs. As a practical guide for potential users, we provide an example describing its application to a comprehensive database on bryophyte distribution on Terceira Island (Azores, Portugal). More than 8,000 bryophyte records were gathered, but (i) less than half of the island area has been surveyed at least once and (ii) less than 1% of these have reliable inventories (placed on the few remnants of laurel forests that have been traditionally better surveyed). Nevertheless, surveying just 15 additional localities evenly distributed across the major environmental regions and habitats on Terceira Island seems to represent the existing environmental diversity. We believe that the survey protocol presented here for bryophytes of Terceira Island could be flexibly applied to other taxa or areas.

KEYWORDS. Environmental diversity, inventory completeness, liverworts, mosses, species richness estimators, stratified sampling.

³ Corresponding author's e-mail:

mcnsc850@mncn.csic.es

DOI: 10.1639/0007-2745-114.3.611



Species distribution is one of the key macroecological variables (Gaston 1994, 2003; Gaston & Blackburn 2000; Lawton 2000). The structure of a species' geographic range and the spatial scale on which species distributions are measured are fundamental features for the understanding of processes explaining current patterns of local and regional species richness (Gaston 2003). However, the available biological information may not necessarily represent real patterns on species distributions. Traditionally, collectors choose to conduct their surveys at the most diverse sites, in the most accessible places, or at historically recorded sites (e.g., Kadmon et al. 2004; Reddy & Davalos 2003). Therefore, biological data generally possess taxonomic incompleteness and spatial biases and gaps in geographic knowledge, the so-called *Linnaean* and *Wallacean shortfalls* (Brown & Lomolino 1998). To overcome the *Wallacean shortfall* as efficiently as possible, new field survey protocols are needed.

For maximum effectiveness, a survey protocol should assess the minimum number of sites or minimum total area necessary to represent all species of a region. Because both fieldwork and species identification require a large investment in terms of time, money and personnel training, two helpful procedures have been proposed to accelerate the study of species distribution patterns: (i) the use of environmental variables as proxies for species diversity to select survey localities (Belbin 1993; Faith & Walker 2004) and/or (ii) distinguishing well-surveyed localities from those with incomplete inventories to optimize the survey process (e.g., Hortal et al. 2004). Ideally, the next best collection localities should be placed where the environmental conditions are less well-represented by the network of existing localities. The best surveyed localities can be differentiated by analyzing their inventory completeness, i.e., the relationship between the observed and the maximum expected number of species for each locality (Colwell & Coddington 1994; Soberón et al. 2000).

Presently, such exploratory analyses may be carried out using the impressive amount of information already available on the Internet, including freely downloadable databases for both biota (e.g., GBIF at www.gbif.org) and the environment (e.g., WorldClimate at www.worldclimate.com). The most recent informatic tools allow not only to access and import databases but also to share the users' material. Although this material opens a wide range of possibilities for data treatment, the errors, gaps and biases associated with those sources of biological information require some caution in their use (Chapman 2005; Soberón & Peterson 2004).

The first part of this paper presents an overview of the main concepts and procedures devoted to the evaluation of the existing species distribution data as well as some general recommendations on how to use them for building a database from which well-surveyed localities may be identified. We stress the importance of accounting for the environmental heterogeneity in field designs when aiming to unveil the "real" species distribution patterns and we explain how to optimize the survey planning by taking into account the information on well-surveyed localities previously identified. As an illustration of the above-mentioned methodology, the second part of this work describes its application to a bryophyte database from Terceira Island (Azores, Portugal). Our results will be discussed to provide guidelines applicable to any taxonomic group and spatial scale for which species occurrence data and environmental variables can be measured.

Importance of planning a field survey design.

The design of a survey is shaped primarily by the purpose of the study, which must be clearly defined in advance. Estimating species richness, composition and abundance at a regional scale or species distributions along a gradient, or defining conservation areas for specific taxa, for example, require different survey designs. Moreover, planning a field survey may be important to optimize the taxonomical resources when funding or time is limited.

The design of a survey scheme must take into account that most species in the communities are rare, i.e., the species range size and species abundance distributions are right skewed (Gaston 1994, 1999, 2003; Lawton 2000). As a result, a geographically extensive scheme will sample spatially rare species but will miss some local rarities, whereas a locally intensive program will detect short lived species but will miss species that only occur in other places. An adequate survey design should aim at obtaining a reliable representation of localities across the whole territory but also to detect as many species as possible in each locality. This involves a trade-off in allocating the available work resources between and within localities.

Geographically extensive surveys have frequently been designed using different statistical strategies. Those that select a representative sample of points at random or regularly at spacing intervals have been shown to be logistically demanding and costly, while also ignoring the underlying non-random relationship between biota and their physical environment (Southwood & Henderson 2000). Alternatively, the gradsect and stratified survey methods are considered powerful methods capturing the overall environmental conditions of the analyzed territory while minimizing the survey costs (e.g., Austin & Heyligers 1991; Wessels et al. 1998). The gradsect method tries to find the steepest gradient of environmental variation in the study region along which a transect is oriented, while the stratified method identifies more or less “discrete” patches (or *strata*) that are environmentally homogeneous. Obviously, choosing the most important variables for the targeted taxonomic group is a key decision when dividing the region to be surveyed, for which expert knowledge and bibliographic documentation will be essential. Currently, this task can be facilitated using common statistical packages. The main environmental gradients in the study area can be identified with a preliminary ordination analysis and the so-generated results used to obtain the ‘environmental domains’ or ‘ecoregions’ by means of clustering algorithms (e.g., Leathwick et al. 2003). Using geographic information system tools (GIS), the clusters or regions so derived can be mapped and overlaid with other digital cartographic layers, such

as land-use maps, to account for present-day modifications of natural environmental conditions (i.e., identifying the main habitat types within each environmental area). Each region could be considered a stratum to be surveyed, and the number of survey localities within each stratum may be selected by assigning equal numbers of localities across all strata or, even better, by selecting a number of localities proportional to the area of each stratum and its internal variability (the Neyman allocation procedure; Neyman 1934). Both the number of localities to be surveyed and their spatial placement should be determined *a priori* considering the aims of the study and the maximum possible survey effort, related with the availability of financial and time resources, and with the accessibility limitations of the area. Moreover, we also stress that the stratified sampling design may be further optimized by analyzing previously collected information and estimating the proportion of environmental variation covered by well-surveyed localities to avoid oversampling while focusing on lesser visited or previously ignored places.

Estimating the completeness of species inventories. Most of the information on species distribution currently available in digital databases comes from taxonomic catalogs of the surveyed localities. Comparison of these inventories requires verifying firstly the significance of differences between the observed species richness and the maximum number of species that could be expected with exhaustive survey efforts. Different methods have been proposed to calculate the expected maximum number of species in a locality (Colwell & Coddington 1994; Gotelli & Colwell 2001; Longino et al. 2002), which can be divided into three main groups: i) fitting a statistical distribution to rank abundance data, ii) producing species accumulation curves and iii) using non-parametric estimators. Only the last two methods can, however, be applied when species occurrences are the only available data, as is the case in most bryophyte’s studies.

Accumulation curves first appeared in the paleoecology and entomology literature and are derived from the Michaelis-Menten enzymatic kinetics function. Such curves are based on the assumption that the higher the survey effort, the

greater the number of species found with the relationship being, usually, asymptotic (Soberón & Llorente 1993). This asymptote can be calculated using different types of fitting curves (Flather 1996), but in all cases it may serve as an estimate of the species pool or the total number of species. Accumulation curves have the advantage of being visually intuitive and useful for extrapolation purposes. In fact, they have been applied to the prediction of the expected number of discovered species over time (e.g., Aranda et al. 2010; Medellín & Soberón 1999). However, when all the available biological information is coming from heterogeneous sources using different methodologies, it is difficult to obtain a reliable and standard measure of survey effort. An acceptable solution in this case consists of using the number of database records as a surrogate of survey effort (Lobo 2008). Criteria for building a biological database and setting a 'record' are user defined, but it is useful to include all available taxonomic and distributional data relating to date, collector and locality for each recorded species (see Hortal et al. 2006).

Non-parametric methods are currently more accepted because they seem to generate more reliable estimates of expected species richness values (e.g., Walther & Martin 2001). Non-parametric equations were originally developed to estimate population sizes in capture-recapture data, and they rely on the assumption that rarity strongly influences richness estimates. Hence, in contrast with accumulation curves, they establish different probabilities of occurrence for each species, which are: for presence-absence data, 'uniques' or 'duplicates' (referring to species that occur in only one or two localities, respectively) and, for abundance data, 'singletons' or 'doubletons' (referring to species with only one or two individuals, respectively) (Colwell & Coddington 1994). As the survey effort increases, it also increases the likelihood of finding the rarest species, so there is a good chance that the maximum species richness will be reached.

Although a wealth of literature about richness estimators exists, a universal consensus procedure has not yet been reached. Among the most accepted, Díaz-Francés & Soberón (2005) recommend the Clench and exponential models for species

accumulation curves, while Hortal et al. (2006) argue for the use of Chao2 and Jackknife2 non-parametric estimators. The choice of the method will ultimately depend on the biology of the taxonomic group, the considered spatial scale, and the type of available dataset (see Walther & Moore 2005 and references therein). In either case, the ratio between the number of observed species (which is usually an underestimation of reality) and the estimated theoretical maximum number of species in the survey locality will be indicative of the 'completeness' of the inventory (Soberón et al. 2000), thereby allowing a comparison among localities.

A CASE STUDY: BRYOPHYTE DIVERSITY ON TERCEIRA ISLAND

Azorean bryophytes constitute a good example to illustrate the above-mentioned procedures, as the Azores are one of the Macaronesian archipelagos most affected by land-use changes involving predominant man-made habitats. The level of change is so dramatic that the native subtropical broadleaf evergreen forests or *laurisilva* has disappeared from some islands and is very scarce and highly fragmented in others (Cardoso et al. 2009; Borges et al. 2010a). Herewith we focused on bryophytes' distribution records from Terceira Island, which is one of the most species-rich and better surveyed islands (see Gabriel & Bates 2005) but for which the taxonomic inventory is still incomplete (Aranda et al. 2010). In this study, we aimed to estimate the variation in species diversity among site plots within each survey locality, placed along the whole environmental gradient of the island. To do this, (1) we explored pre-existing information to identify geographical gaps, assessing the inventory completeness of all 500×500 m UTM-cells covering the island, and (2) we designed a survey protocol by stratifying the island into environmental regions and land uses, and selecting the cells to be surveyed among those previously not identified as well-surveyed capable of maximizing the environmental variability.

Analyzing pre-existing information. As mentioned above, assessing the degree of completeness of species inventories for a given locality may be achieved by following three basic

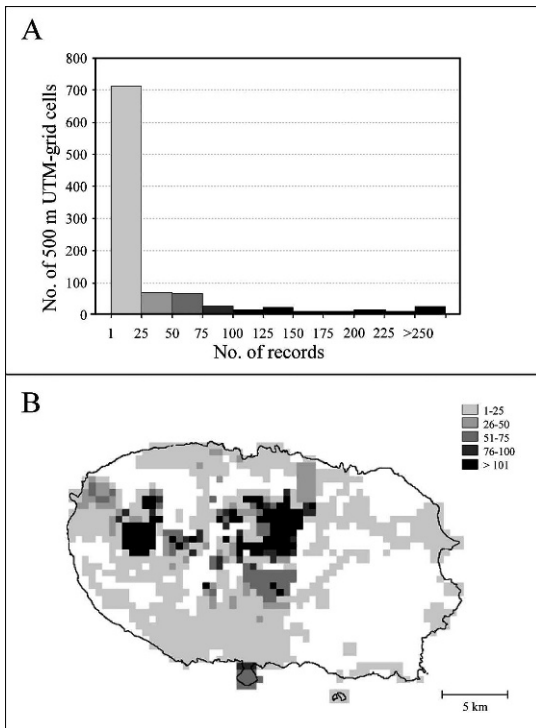


Figure 1. Survey effort carried out for bryophytes on Terceira Island, measured as the number of database records compiled for all species present in each 500×500 m UTM-cell that was visited at least once ($n = 962$). **A.** Number of database records per grid cell. **B.** Geographic distribution of database record density per grid cell. Categories include the upper bounds.

steps: (i) measuring the survey effort, (ii) calculating the maximum estimated richness, and (iii) obtaining the ratio between observed (S_{obs}) and estimated richness (S_{est}).

We used a database built from a thorough review of the literature and herbaria that includes 8,140 bryophyte distribution records for Terceira Island. Based on 374 surveyed localities, this dataset includes information on 348 species, belonging to 68 families of mosses, liverworts and hornworts. We employed the number of database records for each survey locality (defined as the number of times that a species has been recorded by different collectors on different dates, at varied altitudes and/or on distinct substrates) as a surrogate of the survey effort carried out on the island. This information was georeferenced by experts using 500×500 m UTM-grid cells, which is an appropriate resolution for the island area (c. 400 km², around 1,700 cells); a public version of the database may be consulted at the

Azorean Biodiversity Portal website (www.azoresbioportal.angra.uac.pt). Among the surveyed localities, 128 (from old studies) could not be accurately placed at this fine resolution; therefore, we assigned them a lower precision, oscillating from four neighboring grid cells to a maximum of 50 neighboring grid cells (two cases only). This strategy may overestimate the local species richness in some cells and the amount of cells finally considered as well-surveyed. In our case, and due to the high resolution of the used data (cells of 0.25 km²), we decided to maintain this dataset of lower precision (around 2,247 database records, c. 28% of the total used), but authors should reach a decision balancing the advantages of using a greater quantity of data with reliable but imprecise information at the expense of more geographically accurate data.

Although nearly half of the island's area was visited at least once, the histogram of survey effort is highly right-skewed (**Fig. 1a**), meaning that most surveyed cells contain less than 25 database records, and only around 10% of the cells have more than 100 database records. This biased pattern is related to the distribution of land uses as collectors have been noticeably attracted towards the few remnants of native habitats (e.g., *laurisilva*, bogs, natural grasslands), mainly located above 500 m in the middle of the island, whereas most of the less-surveyed cells correspond to highly modified habitats (mainly intensive pastureland, urban and agriculture areas) along the margin (**Fig. 1b**). Hence, taking into account this uneven distribution of survey effort, it is not possible to discern if those localities with low values of species richness were effectively poor in species or simply less surveyed. However, what happens in the cells that are more intensively surveyed? Are their species inventories truly reliable?

Once the distribution of the number of database records for each grid cell was found, we estimated the maximum expected number of species in each grid cell according to its survey effort. First, we selected cells in the upper quartile in terms of number of database records provided that the number of database records for the cell exceeded the number of observed species. In total, 155 such intensively surveyed (hereafter IS) grid cells were identified, representing around 15% of all visited cells (see

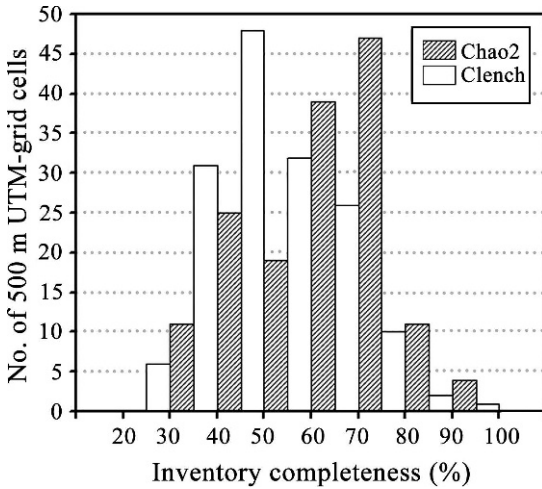


Figure 2. Variation of completeness values for the intensively surveyed (IS) grid cells ($n = 155$). Inventory completeness was defined as the ratio of the currently known species to the maximum expected richness that was obtained according to both the non-parametric Chao2 estimator and the asymptote of the Clench curve. Categories include the upper bounds.

Fig. 1) and covering less than 1% of the island’s total area.

Next, we estimated the species richness (S_{est}) for these cells using two different methods: the species accumulation curve based on the Clench function and the Chao2 non-parametric estimator. The Clench estimator is appropriate for either small areas or well-known taxa, and it assumes that the probability of adding new species (S) to the inventory decreases with the number of species already found but increases with the time spent in the field. We used the STATISTICA package (StatSoft 2007) to calculate the Clench equation (Eq. 1) by using the Simplex and quasi-Newton method to estimate the parameters (cf. Jiménez-Valverde & Hortal 2003). Thus, the Clench curve, $S(r)$, was fitted to the observed species richness (S_{obs}) as a function of the number of database records (r) and model parameters (a and b):

$$S(r) = \frac{a \cdot r}{1 + b \cdot r} \quad (\text{Eq.1})$$

Because the shape of the species accumulation curve may vary as a consequence of the order in which samples are added, the entry of database records was randomized 100 times (Colwell & Coddington 1994). Finally, we calculated the

asymptote of the curve using the obtained model parameters as $S_{est} = a/b$ (Soberón & Llorente 1993). In contrast, the S_{est} for the Chao2 estimator is given by the equation:

$$S_{est} = S_{obs} + \left(\frac{r-1}{r}\right) \left(\frac{U(U-1)}{2(D+1)}\right) \quad (\text{Eq.2})$$

where U and D are the frequencies of unique and duplicate species, respectively. According to Hortal et al. (2006), Chao2 is recommended for comparable and/or small cell sizes, and presently, it is among the most widely applied non-parametric estimators (e.g., Nakamura & Soberón 2009). We obtained Chao2 directly from the user-friendly *EstimateS* software (Colwell 2005) following the settings options explained in the user’s guide.

Finally, we calculated the inventory completeness for all of the IS grid cells. Completeness values estimated by the Clench function were highly variable, being lower than 70% in most cases (**Fig. 2**), whereas estimations based on Chao2 yielded higher values, as supported by other studies (e.g., Peterson & Slade 1998). As decisions about the most reliable richness estimator are highly dependent on the available dataset (Walther & Moore 2005), we selected the S_{obs} values of native forests (the best surveyed habitat in Terceira) as a “guesstimate” with which to compare or validate the results obtained with the Clench and Chao2 estimators. Correlations between S_{obs} and S_{est} for the IS cells belonging to native forest were higher for Chao2 ($r = 0.68, n = 68, p < 0.05$) than for Clench ($r = 0.59, n = 68, p < 0.05$). Thus, based on the Chao2 results, we selected the 15 grid cells with completeness values higher than 80% (**Fig. 2, Supplementary Fig. S1**). Most of these well-surveyed cells are located within the two main patches of relatively unmodified native forests (**Supplementary Fig. S1**): Serra de Santa Bárbara (localities 1–6) and Terra Brava (localities 11–14).

Designing the survey protocol. How much environmental variation do well-surveyed cells recover, and which habitats are better represented? Where should future surveys take place, taking into account existing data on well-surveyed cells? Answering these questions requires following three basic steps: (i) choosing the most appropriate environmental and land-use information, (ii) stratifying the area of study according to this

information, and (iii) determining the new cells to be surveyed considering those previously identified with reliable inventories.

The selection of the environmental and habitat information needed to stratify the survey was first based on a review of the environmental determinants of bryophyte diversity. The available literature suggests that variation in macroclimatic factors like rainfall, temperature, and humidity usually controls bryophyte cover and diversity at broad spatial scales (Bates et al. 2005; Frahm & Klaus 2001). To some extent, light intensity (Gabriel & Bates 2003; Tinya et al. 2009) and site conditions such as topography (Ah-Peng et al. 2007; Callaghan & Ashton 2008) and substrate characteristics (Gabriel & Bates 2005; Spitale et al. 2009) may contribute to explain the patterns of bryophyte diversity. In addition, and although bryophytes are able to disperse over long-distances, it seems that landscape management and habitat fragmentation also influence the naturally occurring species assemblages (Vanderpoorten et al. 2004; Vellak & Ingerpuu 2005).

On this basis, we created the environmental matrix by gathering digital information about climate (temperature, rainfall, relative humidity and solar radiation), topography (altitude, slope and aspect) and geology for all grid cells of Terceira Island (see **Supplementary Table S1** for further details on data sources). All these variables were managed by using the IDRISI Kilimanjaro GIS software (Clark Labs 2003). We considered both mean and extreme values to obtain a total of 41 variables (**Supplementary Table S1**) and subsequently we performed a principal components analysis (PCA) to reduce the number of collinear variables to a new subset of orthogonal environmental factors that characterize the main trends in the environmental variation of the island. PCA results showed that around 95% of the environmental variability of Terceira may be explained by four principal environmental factors (from PC1 to PC4) that were negatively correlated with the maximum relative humidity and positively with diversity of slopes, annual solar radiation and winter rainfall, respectively. With regard to habitat variables, we used a digital map of the five major land uses currently present on the island (**Supplementary**

Table S1): (i) native forests of *laurisilva* species and *Ericaceae*-type vegetation, (ii) exotic forests (*Eucalyptus* spp., *Pittosporum undulatum* and *Cryptomeria japonica*), (iii) semi-natural pastures, (iv) intensively managed pastures, and (v) social areas (including mostly cities, towns and villages, but also some industrial sites and other human constructions).

Once the relevant variables were selected, we used them to define the zones (or strata) to be surveyed in Terceira Island. To do this, we applied an agglomerative *k*-means partitioning technique using the four principal environmental factors as input variables. The *k*-means analysis begins with a user-specified number of *k* clusters into which the objects (i.e., map grid-cells) are to be grouped. All grid cells are examined iteratively to find the most widely environmentally separated sets of cells until the process converges on a particular grouping scheme (Clarke & Warwick 2001). To choose the *k* groups for data partitioning ($1 < k < N$), it is recommended to take into account the maximum survey effort that can be accomplished (cf. Jiménez-Valverde & Lobo 2004). As no pilot study was possible, we performed a preliminary cluster analysis to find environmentally homogenous groups (not shown). According to the resulting dendrogram and based on field experience, we decided to divide the island into four environmental regions. An analysis of similarities (ANOSIM) carried out with the Primer software (Clarke & Gorley 2001) allowed us to confirm that the four defined regions were significantly different in terms of the environmental variables used ($R = 0.469$, $n = 1706$, $p < 0.001$). Briefly, region one (R1) occupies the lowland periphery of the island and is characterized by higher temperatures and low precipitation. By contrast, the highest annual rainfall values and the lowest temperatures occur in R4 because this region is near the highest point of Santa Bárbara, the most abrupt mountain chain on the island. R2 and R3 have similar environmental conditions and correspond mainly to pastureland zones that occupy the majority of the island area. However, R2 is located at a slightly lower altitude than R3, occupying the flattest zone of the island. Finally, we overlaid the generated map of environmental regions with the land-use map of the

Table 1. Selection of survey zones (SZ) in the planned field design. In bold, zones that were considered sufficiently surveyed because contained at least one 500 m grid-cell with reliable inventories (ID as **Fig. S1**). Asterisks show grid-cells that are not truly representative of the semi-natural habitat as they include some bogs (see text for details). Habitat categories as **Fig. 3**.

SZ	Land-uses	Environmental regions	Area (%)	Number of grid cells	ID well-surveyed grid cells
1	Social	R1	9.19	157	15
2	ExoFor	R1	1.79	31	–
3	NatFor	R1	0.15	3	–
4	IntPast	R1	18.81	321	–
5	SemiPast	R1	0.09	1	–
6	Social	R2	2.36	40	–
7	ExoFor	R2	4.75	81	–
8	NatFor	R2	0.59	10	9
9	IntPast	R2	22.84	390	–
10	SemiPast	R2	3.26	56	10*
11	Social	R3	0.81	14	–
12	ExoFor	R3	4.46	76	–
13	NatFor	R3	3.39	58	11–14
14	IntPast	R3	7.85	134	–
15	SemiPast	R3	6.96	119	7
16	Social	R4	0.18	3	–
17	ExoFor	R4	2.56	44	–
18	NatFor	R4	4.81	82	1–6
19	IntPast	R4	0.09	1	–
20	SemiPast	R4	5.08	87	8*

five categories previously established to define 20 main survey zones (hereafter SZ). These SZ delimit areas with similar landscapes across the environmental spectrum of the island (**Table 1**).

Based on pre-existing information, our knowledge about bryophyte diversity in Terceira is obviously conditioned by the species occurring in native habitats, as most cells with reliable inventories are located in the natural evergreen laurel forests (**Figs. 3, Supplementary Fig. S1, Table 1**). Three well-surveyed cells occur in semi-natural pastures, although two of them (Lagoa do Negro and Turfeira do Pico da Bagacina) are not truly representative of this habitat because they correspond to bogs. The high completeness for Canado do Porto (locality 15) is explained by its easy accessibility, since it is located near one of the two largest urban cores of Terceira. By contrast, no well-surveyed grid cells exist for either exotic forests — where, in fact, a great diversity has been demonstrated (Bates 2000) — nor for intensive pastures. Although the latter habitat is not especially species rich, it should be better surveyed because it accounts for nearly half of the island's area

and because it probably harbors different species (e.g. Sérgio & Gabriel 1995). Also, pastures are surrounded by fences made of volcanic basaltic rocks that are an excellent substrate for bryophytes, but few data are available about their species richness and composition on this substrate. Regarding the environmental coverage, R3 and R4 are overrepresented because they correspond mostly to native forests located in the uplands (**Fig. 3**). Overall, only five of the 20 SZ previously defined have at least one cell with a reliable inventory (**Table 1**), which represents around 21% of the island area. So, how may future surveys be improved? How to choose the following cells to be surveyed within the 15 zones that had not been sufficiently inventoried?

Due to survey effort limitations, we decided to select only one locality of around 1 ha from every SZ, but the number and size of the plots surveyed within each locality varied as a function of the type of habitat and the dominant substrates found there (ground, rocks and/or tree trunks). On average, we surveyed between 15 and 21, 100×100 cm plots in pastures, 16 5×5 cm plots in social areas and

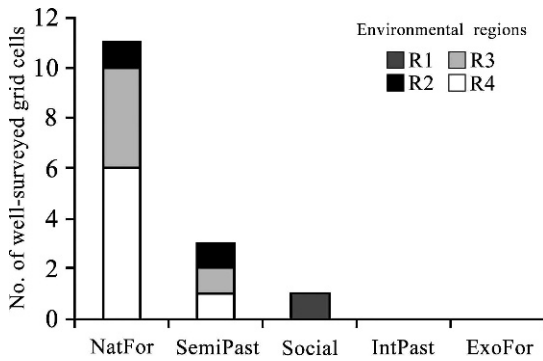


Figure 3. Land-use coverage and environmental variation explained by the well-surveyed grid cells on Terceira Island ($n = 15$). Land use categories are: ‘NatFor’ (native forests), ‘ExoFor’ (exotic forests), ‘SemiPast’ (semi-natural pastures), ‘IntPast’ (intensive pastures), and ‘Social’ (including mainly cities, towns and villages, but also other human constructions). Environmental regions were obtained according to a k -means clustering analysis (see methods).

between 39 and 44 30×30 cm standard plots in forests. In order to place the 15 survey localities within each SZ, we developed a ‘decision rule’ applying the following constraints: the patch with the largest area, smallest perimeter/area relationship (i.e., more globular) and located farthest from the most different habitat (in terms of disturbance) was selected. Finally, we placed the survey localities ($n = 15$), taking into account their accessibility and proximity to the mean environmental conditions (PC1 to PC4 values) of each SZ (Fig. 4).

DRAWBACKS AND UNCERTAINTIES

The current inventory of living species is unsatisfactory in terms of completeness and accuracy for both the majority of taxa and geographical areas; these have been colloquially termed the *Linnaean* and *Wallacean shortfalls*, respectively (Brown & Lomolino 1998). Even in those cases where an apparently large amount of data is available, a map of recording intensity *per se* shows only the distribution of collecting effort without indicating anything about how complete the survey has been. This is the case for the bryophytes of Terceira, where, despite the existence of more than 8,000 records for all species, only half of the island’s area has been visited at least once, and of these sites, less than 1% has reliable inventories. For instance, we found that one

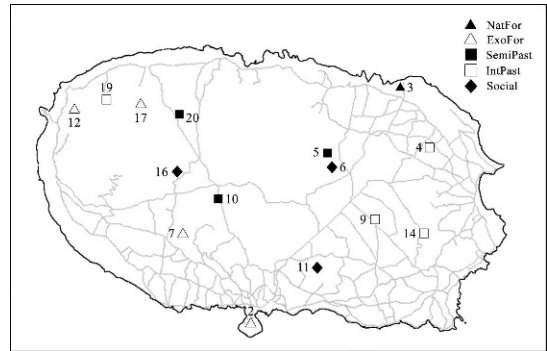


Figure 4. Placement of survey localities resulted from the stratified survey design using the main environmental gradients and the different land uses recognized on the island (see **Supplementary Table S1**). Point numbers correspond to the survey zones indicated in **Table 1**. Land-use categories correspond to **Fig. 3**.

500×500 m grid cell in Serra de Santa Bárbara (one of the most surveyed in the island) has 326 database records, and yet 60% of the 89 recorded species had only one or two records. This implies that the completeness of this inventory does not reach 70%. However, completeness values should be considered cautiously and not treated as gold standards. In the case of Serra de Santa Bárbara, the low completeness may be due to the existence of ‘unique’ and ‘duplicate’ species that are truly rare and hard to find (e.g., *Acrobolbus wilsonii*, *Breutelia azorica*; see Homem & Gabriel 2008) or, conversely, due to the occurrence of relatively common species that were intentionally less collected by classical taxonomists (e.g., *Rhytidialdelphus squarrosus*, *Conocephalum conicum*).

Another critical decision regarding the completeness is the selection of a threshold to discriminate well-surveyed localities. The issue of “stopping rules”, specifying when to optimally stop the survey, is a trade-off between obtaining a new species for the inventory and the survey costs. Although decision rules offer great advantages in terms of avoiding under-surveyed sites and making efficient efforts, they are typically applied only to species accumulation curves (Chirsten & Nakamura 2003; Soberón & Llorente 1993), and there is still no general recommendation about which is the best completeness cut-off for any situation. Cardoso (2009) suggested three subjective levels of

completeness for arthropod inventories: (i) ‘reasonable’ (about 50% of the species ought to be sampled), (ii) ‘comprehensive’ (reaching 70–80% of existing species), and (iii) ‘exhaustive’ (90% of all known species). This author considers that collecting 100% of species would demand too much effort and be unfeasible. Chao et al. (2009) agree that the effort required to detect all estimated species will often be “prohibitively large”, and they propose empirical formulas using Chao estimators for more realistic objectives, showing that substantially less effort is needed to detect 90% or even 95% of the species. In the present study, we chose the grid cells with completeness inventories up to 80% as a trade-off between selecting a small number of well-surveyed cells or choosing a greater number of less inventoried cells. Although decisions on the completeness threshold may change the obtained results, we followed a standardized protocol that may be repeatable and validated in the future with the addition of new field collection data.

Regarding the survey design, we used a stratified sampling protocol because, among all alternatives, it seems to perform quite well (Hirzel & Guisan 2002). For bryophytes, however, preferential sampling has traditionally been used (e.g., Gradstein et al. 1996; Kürschner 2003) instead of systematic or stratified designs (Hill & Lozano 1994; McCune et al. 2000; Newmaster et al. 2005). The issue of how to stratify the survey and locate survey localities is inherently difficult, but the environmental regionalization approach is among the most accepted (Mackey et al. 2008). However, the available classification techniques are not perfectly objective because, as we have seen, the stratified sampling procedure is ultimately determined by four choices that are somewhat subjective: (1) the variables used, (2) the clustering algorithm, (3) the number of regions to be selected, and (4) the number and placement of survey localities within each region.

Based on literature and personal experience, the environmental and habitat variables should be selected according to well-known causal relationships with the abundance or occurrence of a species. However, such an approach is often impracticable due to a lack of knowledge of the true factors determining these patterns. Although climate plays a

major role in delimiting the distribution of vegetation, other important variables such as soil, habitat type, slope or orientation may interact with climate or simply be correlated with it. To what extent large-scale environmental variations influence the local distributions of species remains debatable and unstudied in the case of bryophytes. Lastly, in the present study we analyzed only the five most dominant land-use categories in the island, but other habitats important for bryophyte distribution, like peatlands and coastal zones, should be further investigated in the future.

Another source of uncertainty is related with the statistical technique used to discriminate the environmental strata. Unlike agglomerative hierarchical algorithms, non-hierarchical partitioning such as the *k*-means is sometimes preferred for this purpose because: (i) it strongly minimizes heterogeneity within groups while maximizing distinctiveness between groups and (ii) it is computationally less demanding (Legendre & Legendre 1998; Trakhtenbrot & Kadmon 2006). Some authors argue, however, that cluster analysis obscures useful information for predicting the relative amounts of biodiversity represented by different sets of areas and they propose the ‘environmental diversity’ or ED criterion based on heuristic decision rules (Faith & Walker 1996; Hortal & Lobo 2005). In this study, we opted for agglomerative *k*-means partitioning technique because it is simple and easy to use with standard statistical packages.

The level of division at which to stratify the sampling is a critical choice because it will essentially define the sample size for latter comparisons. Such a decision must be partially made on a statistical basis while also following common sense and experience. Thus, coarse divisions accurately capture regional environmental differences, whereas fine divisions highlight local gradients, ecotones and clines (Bailey 2005). In our case, we observed that the choice of divisions was relatively unimportant beyond *k* = 4 partitions, because we obtained environmental regions that were similar although, obviously, progressively more fragmented.

Finally, we selected only one locality per survey zone for logistical reasons and because we aimed at

estimating the variations of species richness and composition among site plots (our sampling unit) regularly located in each survey locality placed along the whole environmental gradient of the island. Lastly, we chose the geographical positions of survey localities not randomly, but rather based on criteria that are known to affect general diversity patterns, such as edge effects and the impact of regional conditions on local diversity (Cardoso *et al.* 2009), which have proven to also be relevant for bryophytes (Moen & Jonsson 2003; Pharo *et al.* 2004; Zartman & Shaw 2006).

CONCLUSIONS

It could be argued that primary biodiversity data available in natural history collections and the taxonomic literature are not useful for estimating diversity patterns due to the different initial purposes of species collecting and the heterogeneous methodologies used. However, we show that the biological information collected by taxonomists over time is highly valuable because it may be useful in designing field surveys directed specifically at unveiling biodiversity distribution patterns (Aranda *et al.* 2010; Solow & Roberts 2006). This often neglected information may provide an overall indicator of where future surveys should take place to optimize the search for additional species or citations that would otherwise demand an impracticable amount of effort. For a group as diverse as bryophytes in a region as rich as Macaronesia, the remarkable prior work carried out to gather taxonomical information—which has recently been published online through the Azorean Biodiversity Portal (www.azoresbioportal.angra.uac.pt; see Borges *et al.* 2010b)—allowed us to optimize the survey effort using a well-planned stratified survey design. Because error propagation and biases may start early in the process of field data compilation, we also offer some practical recommendations about how to standardize the available information before processing and interpretation. We believe that the survey protocol presented here for bryophytes of Terceira Island could be flexibly applied to other taxa or areas.

ACKNOWLEDGMENTS

The authors are especially grateful to two anonymous referees for their critical suggestions on the manuscript. Climate data was obtained from Project CLIMAAT, PIC - INTERREG_3B (MAC 2.3/A3). SA is a PhD candidate from University of Azores and was supported by a grant from Direcção Regional da Ciência e Tecnologia dos Açores (M311/I009A/2005). PAVB and RG worked on this project under the DRCT projects M2.1.2/1/017/2007, M.2.1.2/1/003/2008 and the EU projects INTERREGIII B “ATLÂNTICO” (2004–2006) and BIONATURA (2006–2008).

LITERATURE CITED

- Ah-Peng, C., M. Chuah-Petiot, B. Descamps-Julien, J. Bardat, P. Stamenoff & D. Strasberg. 2007. Bryophyte diversity and distribution along an altitudinal gradient on a lava flow in La Reunion. *Diversity and Distributions* 13: 654–662.
- Aranda, S. C., R. Gabriel, P. A. V. Borges & J. M. Lobo. 2010. Assessing the completeness of bryophytes inventories: an oceanic island as a case study (Terceira, Azorean archipelago). *Biodiversity and Conservation* 19: 2469–2484.
- Austin, M. P. & P. C. Heyligers. 1991. New approach to vegetation survey design: Gradsect sampling. Pages 31–51. *In* C. R. Margules & M. P. Austin (eds.), *Nature Conservation: Cost effective biological surveys and data analysis*. Csiro, Melbourne, Australia.
- Azevedo, E. B., L. S. Pereira & B. Itier. 1999. Modelling the local climate in island environments: water balance applications. *Agricultural Water Management* 40: 393–403.
- . 2003. Projecto CLIMAAT - Clima e Meteorologia dos Arquipélagos Atlânticos. PIC Interreg_III B - Mac 2.3/A3.
- . 2008. Projecto MacSIMAR - FEDER PIC MAC/1/A089.
- Bailey, R. G. 2005. Identifying ecoregion boundaries. *Environmental Management* 34: S14–S26.
- Bates, J. W. 2000. Introduction to the Azores and its bryophytes. *Bulletin of the British Bryological Society* 76: 21–23.
- , K. Thompson & J. P. Grime. 2005. Effects of simulated long-term climatic change on the bryophytes of a limestone grassland community. *Global Change Biology* 11: 757–769.
- Belbin, L. 1993. Environmental representativeness: regional partitioning and reserve selection. *Biological Conservation* 66: 223–230.
- Borges, P. A. V., E. B. Azevedo, A. Borba, F. O. Dinis, R. Gabriel & E. Silva. 2010a. Ilhas Oceánicas. Pages 461–508. *In* H. M. Pereira, T. Domingos & L. Vicente (eds.), *Portugal Millenium Ecosystem Assessment*. Escolar Editora, Lisboa.
- , R. Gabriel, A. Arroiz, A. Costa, R. Cunha, L. Silva, E. Mendonça, A. F. Martins, F. Reis & P. Cardoso. 2010b. The Azorean Biodiversity Portal: a regional model internet database on biodiversity data sharing. *Systematics and Biodiversity* 8: 423–434.

- Brown, J. H. & M. V. Lomolino. 1998. *Biogeography*, 2nd ed. Sinauer Press, Sunderland, Massachusetts.
- Callaghan, D. A. & P. A. Ashton. 2008. Bryophyte distribution and environment across and oceanic temperate landscape. *Journal of Bryology* 30: 23–35.
- Cardoso, P. 2009. Standardization and optimization of arthropod inventories—the case of Iberian spiders. *Biodiversity and Conservation* 18: 3949–3962.
- , S. C. Aranda, J. M. Lobo, F. Dinis, C. Gaspar & P. A. V. Borges. 2009. A spatial scale assessment of habitat effects on arthropod communities of an oceanic island. *Acta Oecologica-International Journal of Ecology* 35: 590–597.
- Chao, A., R. K. Colwell, C.-W. Lin & N. J. Gotelli. 2009. Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90: 1125–1133.
- Chapman, A. D. 2005. Principles and methods of data cleaning – Primary species and species-occurrence data, version 1.0. Report for the Global Biodiversity Information Facility, GBIF, Copenhagen.
- Christen, J. A. & M. Nakamura. 2003. Sequential stopping rules for species accumulation. *Journal of Agricultural Biological and Environmental Statistics* 8: 184–195.
- Clark Labs. 2003. IDRISI Kilimanjaro. GIS software package. Clark Labs, Worcester, MA, USA.
- Clarke, K. R. & R. N. Gorley. 2001. PRIMER v5: user manual/tutorial. PRIMER-E, Plymouth, UK.
- & R. M. Warwick. 2001. Change in Marine Communities: an Approach to Statistical Analysis and Interpretation. PRIMER-E, Plymouth, UK.
- Colwell, R. K. & J. A. Coddington. 1994. Estimating Terrestrial Biodiversity through Extrapolation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 345: 101–118.
- . 2005. EstimateS: Statistical estimation of species richness and shared species from samples. Version 7.5.
- Díaz-Francés, E. & J. Soberón. 2005. Statistical estimation and model selection of species-accumulation functions. *Conservation Biology* 19: 569–573.
- DROTRH. 2008. Carta de ocupação do solo da região Autónoma dos Açores - Projecto SUEMAC. Secretaria Regional do Ambiente, Direcção Regional do Ordenamento do território e dos Recursos Hídricos.
- Faith, D. P. & P. A. Walker. 1996. Environmental diversity: On the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodiversity and Conservation* 5: 399–415.
- & ———. 2004. The ED strategy: how species-level surrogates indicate general biodiversity patterns through an ‘environmental diversity’ perspective. *Journal of Biogeography* 31: 1207–1217.
- Flatler, C. H. 1996. Fitting species-accumulation functions and assessing regional land use impacts on avian diversity. *Journal of Biogeography* 23: 155–168.
- Frahm, J.-P. & D. Klaus. 2001. Bryophytes as indicators of recent climate fluctuations in Central Europe. *Lindbergia* 26: 97–104.
- Gabriel, R. & J. W. Bates. 2003. Response of photosynthesis to irradiance in seven bryophytes from the Azorean forests. *Journal of Bryology* 25: 101–105.
- & ———. 2005. Bryophyte community composition and habitat specificity in the natural forests of Terceira, Azores. *Plant Ecology* 177: 125–144.
- Gaston, K. J. 1994. *Rarity*. Chapman & Hall, London.
- . 1999. Implications of interspecific and intraspecific abundance-occupancy relationships. *Oikos* 86: 195–207.
- . 2003. *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.
- & T. M. Blackburn. 2000. *Pattern and process in macroecology*. Blackwell Science, Oxford, UK.
- Gotelli, N. J. & R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4: 379–391.
- Gradstein, S. R., P. Hietz, R. Lücking, A. Lücking, H. J. M. Sipman, H. F. M. Vester, J. Wolf & E. Gardette. 1996. How to sample the epiphytic diversity of tropical rainforests. *Ecotropica* 2: 59–72.
- Hill, M. O. & F. D. Lozano. 1994. A numerical analysis of the distribution of liverworts in Britain. Pages 11–20. *In* M. O. Hill, C. D. Preston & A. J. E. Smith (eds.), *Atlas of the Bryophytes of Britain and Ireland*, vol. 3. Harley Books, Colchester, United Kingdom.
- Hirzel, A. & A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling* 157: 331–341.
- Homem, N. & R. Gabriel. 2008. Briófitos raros dos Açores / Azorean Rare Bryophytes. *Princípios*, Lisboa, Portugal.
- Hortal, J., P. Garcia-Pereira & E. Garcia-Barros. 2004. Butterfly species richness in mainland Portugal: Predictive models of geographic distribution patterns. *Ecography* 27: 68–82.
- & J. M. Lobo. 2005. An ED-based protocol for optimal sampling of biodiversity. *Biodiversity and Conservation* 14: 2913–2947.
- , P. A. V. Borges & C. Gaspar. 2006. Evaluating the performance of species richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology* 75: 274–287.
- Jiménez-Valverde, A. & J. Hortal. 2003. Las curvas de acumulación de especies y la necesidad de evaluar la calidad de los inventarios biológicos. *Revista Ibérica de Aracnología* 8: 151–161.
- & J. M. Lobo. 2004. Un método sencillo para seleccionar puntos de muestreo con el objeto de inventariar taxones hiperdiversos: el caso práctico de las familias Araneidae y Thomisidae (Araneae) en la Comunidad de Madrid, España. *Ecología* 18: 297–308.
- Kadmon, R., O. Farber & A. Danin. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications* 14: 401–413.

- Kürschner, H. 2003. Epiphytic bryophyte communities of southwestern Arabia: phytosociology, ecology and life strategies. *Nova Hedwigia* 77: 55–71.
- Lawton, J. H. 2000. Community ecology in a changing world. International Ecology Institute, Oldendorf/Luhe, Germany.
- Leathwick, J. R., J. M. Overton & M. McLeod. 2003. An environmental domain classification of New Zealand and its use as a tool for biodiversity management. *Conservation Biology* 17: 1612–1623.
- Legendre, P. & L. Legendre. 1998. Numerical ecology. 2nd ed. Elsevier Science BV, Amsterdam, The Netherlands.
- Lobo, J. M. 2008. Database records as a surrogate for sampling effort provide higher species richness estimations. *Biodiversity and Conservation* 17: 1572–9710.
- Longino, J. T., J. Coddington & R. K. Colwell. 2003. The ant fauna of a tropical rain forest: estimating species richness three different ways. *Ecology* 83: 689–702.
- Mackey, B. G., S. L. Berry & T. Brown. 2008. Reconciling approaches to biogeographical regionalization: a systematic and generic framework examined with a case study of the Australian continent. *Journal of Biogeography* 35: 213–229.
- McCune, B., R. Rosentreter, J. M. Ponzetti & D. C. Shaw. 2000. Epiphyte habitats in an old conifer forest in Western Washington, USA. *The Bryologist* 103: 417–427.
- Medellín, R. A. & J. Soberón. 1999. Predictions of mammal diversity on four land masses. *Conservation Biology* 13: 143–149.
- Moen, J. & B. G. Jonsson. 2003. Edge effects on liverworts and lichens in forest patches in a mosaic of boreal forest and wetland. *Conservation Biology* 17: 380–388.
- Nakamura, M. & J. Soberón. 2009. Use of approximate inference in an index of completeness of biological inventories. *Conservation Biology* 23: 469–474.
- Newmaster, S. G., R. J. Belland, A. Arsenaault, D. H. Vitt & T. R. Stephens. 2005. The ones we left behind: Comparing plot sampling and floristic habitat sampling for estimating bryophyte diversity. *Diversity and Distributions* 11: 57–72.
- Neyman, J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97: 558–606.
- Peterson, A. T. & N. A. Slade. 1998. Extrapolating inventory results into biodiversity estimates and the importance of stopping rules. *Diversity and Distributions* 5: 95–105.
- Pharo, E. J., D. B. Lindenmayer & N. Taws. 2004. The effects of large-scale fragmentation on bryophytes in temperate forests. *Journal of Applied Ecology* 41: 910–921.
- Pinheiro, J., J. Madruga, J. Sampaio & A. R. Pereira. 2010. Carta Pedológica da Ilha Terceira. Universidade dos Açores, Angra do Heroísmo, Azores, Portugal.
- Reddy, S. & L. M. Davalos. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography* 30: 1719–1727.
- Sérgio, C. & R. Gabriel. 1995. Novos dados para os Açores sobre o género *Riccia*. *Revista de Biologia (Lisboa)* 15: 184–185.
- Solow, A. R. & D. L. Roberts. 2006. Museum collections, species distributions, and rarefaction. *Diversity and Distributions* 12: 423–424.
- Soberón, J. & J. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. *Conservation Biology* 7: 480–488.
- , ——— & L. Oñate. 2000. The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. *Diversity and Conservation* 9: 1441–1466.
- & A. T. Peterson. 2004. Diversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London B* 359: 689–698.
- Southwood, T. R. E. & P. A. Henderson. 2000. Ecological methods. 3rd ed. Wiley-Blackwell, Oxford.
- StatSoft Inc. 2007. STATISTICA (data analysis software system), version 8.0. www.statsoft.com.
- Spitale, D., A. Petraglia & M. Tomaselli. 2009. Structural equation modelling detects unexpected differences between bryophyte and vascular plant richness along multiple environmental gradients. *Journal of Biogeography* 36: 745–755.
- Tinya, F., S. Marialigeti, I. Kiraly, B. Nemeth & P. Odor. 2009. The effect of light conditions on herbs, bryophytes and seedlings of temperate mixed forests in Ars.g, Western Hungary. *Plant Ecology* 204: 69–81.
- Trakhtenbrot, A. & R. Kadmon. 2006. Effectiveness of environmental cluster analysis in representing regional species diversity. *Conservation Biology* 20: 1087–1098.
- Vanderpoorten, A., P. Engels & A. Sotiaux. 2004. Trends in diversity and abundance of obligate epiphytic bryophytes in a highly managed landscape. *Ecography* 27: 567–576.
- Vellak, K. & N. Ingerpuu. 2005. Management effects on bryophytes in Estonian forests. *Biodiversity and Conservation* 14: 3255–3263.
- Walther, B. A. & J.-L. Martin. 2001. Species richness estimation of bird communities: how to control for sampling effort. *Ibis* 143: 413–419.
- & J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28: 815–829.
- Wessels, K. J., A. S. Van Jaarsveld, J. D. Grimbeek & M. J. Van der Linde. 1998. An evaluation of the gradsect biological survey method. *Biodiversity and Conservation* 7: 1093–1121.
- Zartman, C. E. & A. J. Shaw. 2006. Metapopulation extinction thresholds in rain forest remnants. *The American Naturalist* 167: 177–189.

Supplementary documents on-line:

Table S1: Topographic, bio-climatic, geologic and habitat variables used to stratify the island. The values of each variable were given for all the 500×500 m UTM-cells covering the island.

Fig. S1: Locations of well-surveyed 500×500 m UTM-grid cells defined on the basis of completeness values higher than 80% of estimated species richness according to the Chao2 estimator. The legend on the right indicates the identifier (ID) and the corresponding name of the survey locality for each selected grid cell.

Supplementary Table for Aranda, S. C., R. Gabriel, P. A. V. Borges, E. Brito de Azevedo & J. M. Lobo. 2011. Designing a survey protocol to overcome the Wallacean shortfall: a working guide using bryophyte distribution data on Terceira Island (Azores). *The Bryologist* 114(3): 611–624.

Table S1. Topographic, bio-climatic, geologic and habitat variables used to stratify the island. The values of each variable were given for all the 500×500 m UTM-cells covering the island.

Topographic. ALT: Average altitude (m); SLO: Average slope (degree, 0–90); SLODIV and ASPDIV: Slope and aspect diversity were both calculated using the Shannon-Wiener index of the relative frequency of each class in all the twenty-five pixels of 100×100 m contained in every 500 m UTM-square (dimensionless).

Bio-climatic. ETP: Potential evapotranspiration (mm); Monthly RAIN (12 variables): Monthly rainfall (mm); RAIN: Annual rainfall (mm); RAINSUM: Summer rainfall (mm); RAINWIN: Winter rainfall (mm); RHMAX: Maximum relative humidity (percentage); RHMIN: Minimum relative humidity (percentage); RNMAX: Maximum net solar radiation ($\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}$); RNMIN: Minimum net solar radiation ($\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}$); RN: Annual average net solar radiation ($\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}$); Monthly TMED (12 variables): Monthly average temperature ($^{\circ}\text{C}$); TMAX: Annual maximum temperature ($^{\circ}\text{C}$); TMIN: Annual minimum temperature ($^{\circ}\text{C}$); TMED: Annual average temperature ($^{\circ}\text{C}$).

Geologic. Multinomial variable with five categories: basalts, ignimbrites, peralkalines, trachybasalts, trachytes.

Land-use. NatFor: Native forests of laurisilva species and Ericaceae-type vegetation; ExoFor: Exotic forests of *Eucalyptus* spp., *Pittosporum undulatum* and *Cryptomeria*

japonica; IntPast: Pastures under intensive management and sowed from one to ten years intervals, characterised also by a poor vascular flora of five or fewer contributing species; SemiPast: Pastures under low intensive management for more than 30 years, with a higher diversity of grasses and forbs and grazing primarily concentrated in the spring-summer; Social: Mostly urban zones, but also some industrial, crop and farming areas

Data sources. Topographic variables were derived from a digital elevation model (DEM) at a 100 m resolution provided by the Cartographic Service of the Portuguese Army (available at <http://snig.igeo.pt/>). Bioclimatic variables were obtained from the CIELO Model (acronym of “Clima Insular à Escala Local”) also at a 100 m resolution (Azevedo *et al.* 1999). This model was developed to dynamically simulate local climate in island environments. The CIELO is a physically based model that simulates the climatic variables in an island using data from a single reference meteorological station. The reference station “knows” its position in the orographic and dynamic regime context. The domain of computation is a GIS raster grid parameterized with a DEM. The model consists of two main sub-models. One, relative to the advective component simulation, that assumes the Föhn effect to reproduce the dynamic and thermodynamic processes occurring when an air mass moves through the orographic obstacle that justify the observed values at reference station. This makes possible to reproduce the spatial evolution of the air pressure, air temperature, air humidity, orographic clouds and the reinforcement of the precipitation as influenced by the orography along the air displacement. The second concerns the spatial evolution of the radiative component taking in account the effect of the clouds of orographic origin and the shadow produced by the topography. The CIELO model has been calibrated and validated to most Azorean Islands, and is available through CLIMAAT and MacSIMAR projects (FEDER

PIC - INTERREG_3B -MAC 2.3/A3: MAC/1/A089; Azevedo 2003, 2008; see <http://www.climaat.angra.uac.pt/>). Geologic variables come from Pinheiro et al. (2010). Land-Use data was extracted from DROTRH (2008), based on aerial photography and fieldwork data (F. Dinis, unpubl.).

Supplementary Figure for Aranda, S. C., R. Gabriel, P. A. V. Borges, E. Brito de Azevedo & J. M. Lobo. 2011. Designing a survey protocol to overcome the Wallacean shortfall: a working guide using bryophyte distribution data on Terceira Island (Azores). *The Bryologist* 114(3): 611–624.

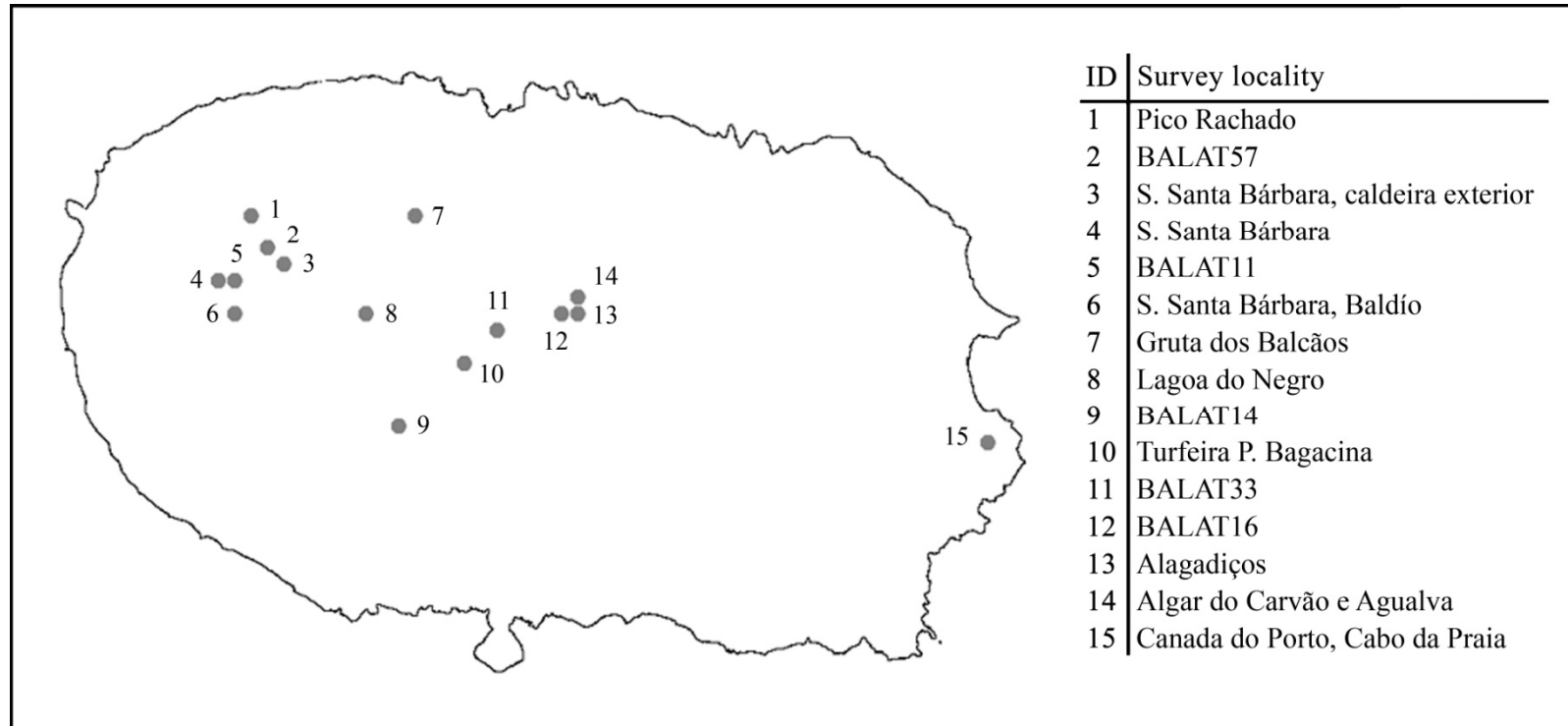


Figure S1. Locations of well-surveyed 500×500 m UTM-grid cells defined on the basis of completeness values higher than 80% of estimated species richness according to the Chao2 estimator. The legend on the right indicates the identifier (ID) and the corresponding name of the survey locality for each selected grid cell.