

Sampling sufficiency in ecological surveys

V. DePatta Pillar

Department of Ecology, Federal University of Rio Grande do Sul, Porto Alegre, RS, 91540-000, Brazil. Fax: +55 51 3191568, e-mail: vpillar@ecologia.ufrgs.br

Keywords: Bootstrap, Cluster analysis, Confidence intervals, Estimation, Ordination, Precision, Preferential, Quadrat, Sample size, Sampling.

Abstract. The paper investigates the problem of evaluating sampling sufficiency in ecological surveys that aim at estimating parameters or detecting complex patterns. It offers a general approach and new methods based upon bootstrap resampling. The methods are useful in pilot studies to anticipate sufficiency or to find out a posteriori if the sample is adequate for a given type of analysis. Regarding the basic idea, the bootstrap algorithm generates frequency distributions of the parameter of interest in samples with increasing size. The parameter may be simple such as means, or describe cluster structure or ordination configuration. By definition the sample is considered sufficient if the parameter reaches stability or the required level of precision within the range of sample sizes evaluated. Examples with artificial and ecological data sets demonstrate consistency and utility. An application program performs the analyses and is available from the author.

Introduction

In any sampling survey, we have to choose an adequate number of sampling units for the study intended. We may fall back for orientation on classical sampling theory (Sampford 1962, Cochran 1977, Green 1979, Krishnaiah & Rao 1988), but in the case of ecological surveys the classical solutions I believe are not optimal. The problem is that these assume a well-behaved medium and not one as complex as in nature owing to inherent high diversity, nonlinear responses, complicated interactions and most important from the sampling point of view, non-random arrangements in geographical and resource space terms. The discussion of the medium and sampling problem under such conditions by Orlóci (1993) and Kenkel et al. (1989) is relevant.

The objectives of sampling in ecology are mainly from two categories: estimation of a quantity or the recognition of pattern. Indeed, ecological surveys often aim at describing species richness and diversity (Pielou 1975, Palmer 1990, 1991), or more complex attributes as in classifications, ordinations and other multivariate problems (Orlóci 1978, Pielou 1984, Podani 1994). Often the attributes do not have an explicit numerical

value, for they are reflections of the sample structure. Because of the limited value of theoretical probability distributions to evaluate precision under these circumstances (Patil et al. 1988, Orlóci 1993), other means are needed to derive probability distributions. These are computer intensive techniques, especially the bootstrap methods, which generate empirical distributions (Efron 1979, Efron & Tibshirani 1993). Resampling techniques can generate confidence intervals, which are used to evaluate sampling sufficiency (Bros & Cowel 1987, Rao 1988, Rao & Wu 1988, Patil et al. 1988, Manly 1992, 1993, Chaudhuri & Stenger 1992:201). Although the usual application of these techniques concerns simple parameters, such as mean values and variances, some authors applied bootstrapping to determine confidence intervals of eigenvalues (Beran & Srivastava 1985, Efron & Tibshirani 1993:63), to examine significance of components (Jackson 1993) or sampling effects in ordinations (Stauffer et al. 1985, Knox & Peet 1989, Knox 1989).

A complication for the use of conventional sampling theoretical rules in many ecological surveys is that the sampling units have to be delimited as areal units or volumes. These entities are aggregates representing subdivisions in a continuous medium. The lack of their

uniqueness creates a different sampling medium than assumed in most statistical texts (Kenkel et al. 1989, Orlóci 1993). The ecologist has to choose sampling unit size and shape, that is to say select the 'grain' of the investigation (O'Neill et al. 1986, Wiens 1989). As a consequence, the sampling universe embodies an infinite number of potential sampling units. This raises also the problem of the parameters being scale dependent (Juhász-Nagy 1967, Greig-Smith 1983, Juhász-Nagy & Podani 1983, Kenkel et al. 1989, Camiz & Gergely 1990, Camiz 1993, Podani et al. 1993). A case of this is manifested by, for example, the dependence of variance on sampling unit size. Larger units result in smaller sampling variance as a rule compared to small sampling units. As a consequence, a smaller number of units is needed for the estimation of the mean with the same precision. The mean is scale independent, but considerations of the feasibility and cost of using a smaller number of large units or many small units are involved. Parameters such as diversity, covariances, Euclidean distances, and other similar quantities are scale dependent. In these, different trends will likely be detected by changing the unit size. Owing to a nonrandom assortment of species, the differences of sampling units are enhanced by small, square-shaped area units, but other scales may also reveal interesting aspects of pattern.

Another decision always involved in sampling concerns the sampling design, that is the method used to select or arrange sampling units. Sampling designs in ecological surveys are often systematic, stratified or even preferential, but rarely unrestricted random (see, e.g., Orlóci 1978, Gauch 1989, Jongman et al. 1995). In unrestricted random sampling all sampling units are equally probable in the sample, but the design may not be practical in the field. In systematic sampling, the sampling units are evenly spaced in geographical space, while in stratified sampling they may be evenly spaced in resource space (Goedickemeier et al. 1997). The sampling design may affect sampling sufficiency. Preferential sampling is known to reduce sampling effort in pattern detection (Orlóci 1993). Stratified random sampling may force the inclusion of rare community types in the sample with less sampling effort than in unrestricted random or systematic sampling (Goedickemeier et al. 1997).

In this paper, I investigate the utility of computer intensive techniques for evaluating sample size sufficiency. I give emphasis to cases in which the main objective is detecting patterns by use of multivariate analysis. My methods follow the idea of process sampling, reminiscent of Poore's (1955, 1956) successive approximation as defined in Orlóci and Pillar (1989,

1991). The idea is to conceive sampling as a process, involving step-by-step expansions of the sample, with each next step intricately tied to the evolution of sample structures, monitored in concurrent data analysis based on which their stability is judged. How can we judge stability in the sample? This question has not yet been addressed beyond the examination of the parameter curves as functions of increasing sample size (Pillar & Orlóci 1989, 1991). I expand this approach by applying bootstrap resampling and judging sampling stability probabilistically.

Methods

I start with n sampling units and p variables. The data set on these may represent a pilot sample at one point in process sampling. This set is expanded as required in the field according to the results of sampling sufficiency calculations. An existing data set may also be a starting point. The sampling units may have meaning as an area or a volume as already explained. The variables may be species presence/absence or quantities, or similar types but for site descriptors. The bootstrap algorithm resamples the data set to generate samples with increasing size, and calculates for each resampling step k , with sample of size n_k the parameter θ_k^* . The exact resampling method, the type of parameters and the sampling design problem will be described in the sequel. The methods are implemented in C++ in the application program **SAMPLER** for Macintosh and Windows systems available from the author.

The basic question I ask is whether the parameter of interest reaches stability or the chosen probability level of precision. If it does I conclude that the sample is sufficient within the range of sample sizes evaluated. The actual indicator of precision is the magnitude of the confidence interval or probability for a given sample size.

Process sampling and sampling stability

Precision and sample stability are linked concepts. The linkage is through the fact that a narrow range of variation of the sample parameter at a given sample size also indicates that the parameter will tend to be more stable across larger sample sizes. This is the principle at the core of process sampling. Sample stability may be perceived in relative terms as a complement of the magnitude of change in the parameter between sampling steps with increasing sample sizes $n_1, n_2, \dots, n_k, \dots, n$, where n_1 is an initial sample size (at sampling step 1). The constant increment of sample size is s , the number of sampling units added at each following sampling step. The total number of sampling steps is $t = 1 + \text{INT}((n - n_1)/s)$, plus 1 more step if n_k at the last step

does not coincide with n . Symbol INT signifies the integer portion of the quotient. Selecting a small s will produce a larger number of sampling steps and a smoother graph of θ_k^* . A typical result is shown in Fig. 1A (see also Greig-Smith 1983:32, Orlóci & Pillar 1989, 1991). However, if the random resampling process is repeated, different sampling units will form each sample of size n_k , and θ_k^* at given n_k may not be the same between runs (see Fig. 1B). Thus, stability cannot be easily evaluated based on a single run. If we perform several runs, the resulting value of θ_k will fluctuate between limits set by the information in the data (Fig. 1C). The definition of these limits is explained in the sequel.

Bootstrap resampling

The bootstrap is a resampling method devised by Efron (1979, see also Efron & Tibshirani 1993) based on the principle that, not having better information, the distribution of observational values in a sample is the best indicator of the distribution in the sampling universe from which the sample was taken. If this is true, resampling the sample with replacement will mimic resampling the sampling universe. Each sample obtained by resampling the sample is a 'bootstrap sample'. The sample being resampled simulates a 'pseudo sampling universe' (PSU).

In the following bootstrap methods, the PSU is simulated by resampling the data set of n sampling units. The algorithm at each resampling takes a random sample with replacement of size n_k and computes the sample parameter θ_k . After performing a large number B of resampling iterations, the ordered values θ_{k1}^* , θ_{k2}^* , ..., θ_{kB}^* define an empirical distribution, from which we can set a confidence interval. For instance, with 1000 iterations and an α probability of 0.05, the lower confidence limit at a given sample size will be θ_{k25}^* , that is, the highest of the lowest 25 values found in 1000 iterations, and the upper limit will be θ_{k976}^* , that is, the lowest of the highest 25 values found in the same 1000 iterations. The number of iterations may be adjusted such that $B\alpha/2$ is an integer to avoid the need for interpolations.

We could instead define PSU's with sizes n_k when bootstrapping sample sizes $n_k < n$. A PSU in this case would have to be a random subsample taken without replacement at each iteration from the large set of n units. This approach would simulate intermediate steps in the process of obtaining a sample of size n . I prefer, however, to define the PSU with n sampling

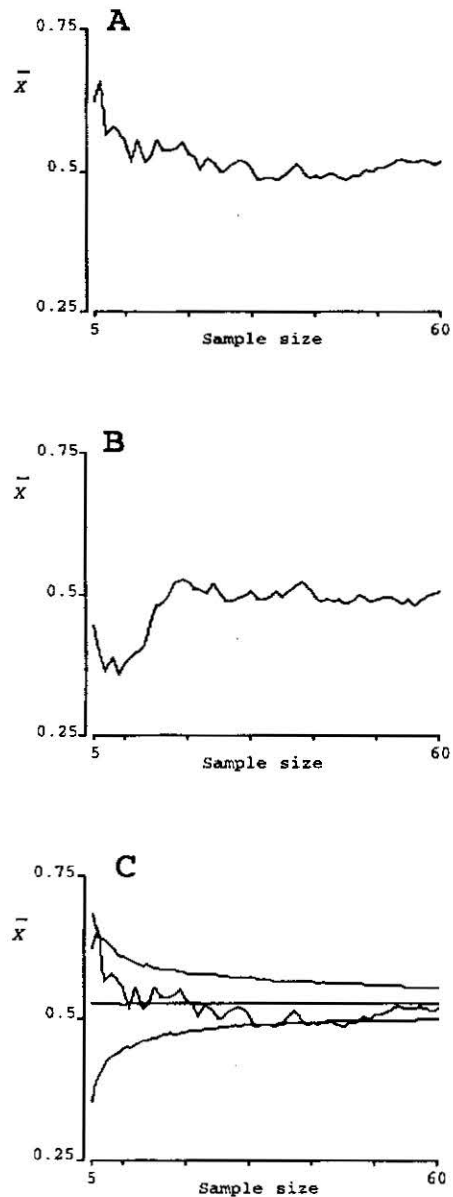


Figure 1. Mean values of a variable X obtained by resampling with replacement a data set of 60 sampling units. The sample sizes vary from 5 to 60. Two of the many sequences of possible sample means are shown in A-B. In C 90% confidence intervals are set based on 1000 resampling iterations at each sample size. For illustration, the case in A and the mean of the 1000 means at each sample size (almost a straight line and at a value identical to the mean of variable X) are drawn with the limits in C. For estimation purposes, by using, e.g., a sample with 5 sampling units the mean is expected in 90% of the cases to lie between 0.36 and 0.67, that is, means that differ as much as 0.31 may not come from different populations (assuming the populations have the same distributions). Much smaller differences may be detectable with 30 sampling units: the mean will lie between 0.42 and 0.60. There is little advantage in taking 60 sampling units; the confidence interval is between 0.45 and 0.58.

units for it is the best representation we have from the real sampling universe.

By comparing successive sampling steps with size n_k , sampling sufficiency may be declared at the sample size at which the confidence intervals reach stability, for beyond this point, increasing the sample size does not add new information to the sample so to affect the variation of the parameter of interest. However, this criterion is strict. A smaller sample size may still give results with the required precision level. We take this consideration, for instance, when we select sample sizes for estimation of the mean of a variable (see Fig. 1C). If the mean of θ_k^* over B iterations coincides with θ obtained in the PSU, the parameter estimated with sample size n_k is considered unbiased (see Fig. 1C). However, unbiasedness would be important only for estimation purposes, such as when means and variances are sought.

The ordered values $\theta_{k1}^*, \theta_{k2}^*, \dots, \theta_{kB}^*$ may instead define a probability according to a given null hypothesis. This probability is used to judge if the null hypothesis should be rejected or not. Actually, the θ_k^* values do not have to be stored and ordered. At each bootstrap resampling and sample size n_k the algorithm compares θ_k^* to its expectation θ_k^0 under the null hypothesis. If $\theta_k^0 \geq \theta_k^*$, the algorithm adds one to the cumulative frequency $F(\theta_k^0 \geq \theta_k^*)$ for sample size n_k . After B bootstrap iterations, the probability $P(\theta_k^0 \geq \theta_k^*)$ is the proportion $F(\theta_k^0 \geq \theta_k^*) / B$. A probability $P(\theta_k^0 \geq \theta_k^*)$ is similarly defined. Regarding the number of resampling iterations for each sample size n_k , the larger is B , the closer the probabilities will be to the true value. I will define the null hypothesis and the computation of θ_k^0 when describing specific parameters.

Parameters

The sample state at any point in the sampling process is described by parameters. The sampling may aim at estimating the mean of an environmental variable. In this case the parameter is simple and the researcher may follow traditional methods (e.g., Cochran 1977) to examine sampling precision by the sample variance and confidence intervals. But theoretical assumptions (normal distribution) implicit in the methods may not be met. Very often sampling is needed to estimate quantities of each species in a community (see Mueller-Dombois & Ellenberg 1974), or the association between variables, in which case these assumptions are even more inappropriate. It is beyond my objective to examine parameters of this sort here, yet

they are suitable for the evaluation of sampling sufficiency by bootstrap resampling, similarly as in the example of Fig. 1C.

Ecological surveys often aim at detecting and explaining patterns by means of ordination and cluster analysis techniques. To examine sampling sufficiency in this context, we have to find sample parameters that are indicators of the state of the sample patterns. If the data are to be analyzed by ordination, the parameter should directly reflect patterns in ordination. Similarly, if the analysis involves a classification, the parameter should indicate the state of partitions generated by the classification technique. I will give further details when describing each parameter.

Ordination structures

The state of data structures that metric ordination reveals is frequently characterized by the ratio between each of the first few eigenvalues computed in the bootstrap sample with size n_k and the total number of non-zero eigenvalues (Stauffer et al. 1985 and Efron & Tibshirani 1993:61). This ratio, however, in addition of being restricted to metric ordination, is not a direct measure of the state and stability of structures in any ordination space (Wilson 1981, Knox & Peet 1989). Similar conclusions may also apply to methods that Jackson (1993) evaluated to determine the number of non-trivial components in ordination.

I characterize ordination structures by a matrix correlation of the bootstrap sample and the PSU,

$$\theta_k^* = \rho(D_k^*; D_n^*)$$

In this, D_k^* is a matrix of pairwise distances computed for n_k sampling units based on their scores on chosen ordination dimensions of a bootstrap sample. Matrix D_n^* is similarly defined for the same sampling units, but based on the ordination of the complete sample with n sampling units. The ordination scores may come from any metric or non-metric method (Orlóci 1978, Pielou 1984, Podani 1994). The computation of D_k^* and D_n^* may use one or more ordination axes. As a possibility, distances may be computed from ranks within each ordination dimension rather than the scores directly. The quantity ρ is the Pearson product moment correlation, but other measures could be contemplated, such as rank correlation (Kendall & Gibbons 1990) or a Procrustes statistic (Schönemann & Carroll 1970).

Probabilities $P(\theta_k^0 \geq \theta_k^*)$ are generated as already explained. I define θ_k^0 as the correlation $\rho(D_k^*; D_n^*)$ based on a sample identical to the PSU but with the values permuted within variables. Each bootstrap

iteration uses a new permutation of the PSU to obtain bootstrap samples and compute θ_k^0 values for sample sizes n_k . The probability $P(\theta_k^0 \geq \theta_k^*)$ will then indicate where the ordination at sample size n_k stands with reference to a random data set generated under a null hypothesis of no correlation (association) structure. By setting an α probability threshold to help the interpretation of $P(\theta_k^0 \geq \theta_k^*)$, a small $P(\theta_k^0 \geq \theta_k^*)$, that is, not larger than α , will indicate that the ordination subspace in consideration is significantly more stable than that would be expected for the same subspace in the ordination of a random data set. In this case we can conclude, with a probability $P(\theta_k^0 \geq \theta_k^*)$ of being wrong, that the given ordination dimension (or dimensions) is non-trivial and worthy of interpretation. Otherwise, the ordination dimension should be taken as unstable (at least in isolation) and indistinguishable from a random data ordination. This method has proven reliable in finding the correct dimensionality in artificial data sets (Pillar, manuscript).

How do we translate these conclusions in terms of sampling sufficiency? When the given ordination dimension is significant at a sample size n_k we should consider that sample size n_k is sufficient to reveal consistent ordination patterns. Why is the significance of the axis equated to sampling sufficiency? The answer relies on my observation that in general the probabilities, if not yet flattened, tend to monotonically decrease or increase with increasing sample size. Therefore, once the axis is taken as significant because $P(\theta_k^0 \geq \theta_k^*)$ is low, if the probabilities are decreasing, it is likely that they will continue decreasing until they flatten at larger sample sizes, which will not change the conclusion of significance. For the same reason, large probabilities that are increasing will continue increasing or flatten, and this will not change the conclusion of non-significance. However, if the probabilities are large and are at some point decreasing with sample size, the only way to interpret sampling sufficiency is based on their stability.

Group structures

When the survey objective is to reveal groups by cluster analysis, it is relevant to ask whether groups with similar make-up would appear again if the survey were repeated in the same sampling universe and with the same sampling intensity and design. The technique includes the computation of the state and stability of the group structure at each sample size n_k and resampling iteration. Any clustering algorithm may be used. We could use the sum of squared dissimilarity between groups (Ward 1963, Orlóci 1967, Pillar & Orlóci

1996) as the parameter to indicate the state of the group structure, but it is not a direct measure of the state of group partitions. Instead, I defined the measuring parameter (Pillar, in press) as the similarity between the group partition in the bootstrap sample and the partition with the same number of groups in the PSU. Both partitions are defined by the same clustering method. The n sampling units in the PSU and the n_k sampling units in the bootstrap sample are viewed as if they were points in the same space of p variables. In this, we compare g groups in the PSU with the g groups in the bootstrap sample by

$$\theta_k^* = G_k^* = 1 - \frac{S}{T}$$

a relative measure of agreement between PSU and bootstrap sample. In this equation, T is a total sum of squares, involving $(n + n_k)(n + n_k - 1)/2$ pair-wise squared dissimilarities of $n + n_k$ sampling units. The matrix containing these dissimilarities must have Euclidean properties (Orlóci 1978, Gower & Legendre 1986). S is a sum of squares for pair-wise contrasts between the groups of the bootstrap sample and their nearest neighbor groups mapped one-to-one in the PSU. S is found by computation of sum of squares for all g^2 pair-wise contrasts between the g groups in the bootstrap sample and the g groups in the PSU. The computation of S from the distance matrix follows similarly defined computations of contrasts (Pillar & Orlóci 1996). I will give details in Pillar (in press). For any pair-wise contrast between group i in the bootstrap sample and group j in the PSU we compute Q_{ij} , the between groups sum of squares for the contrast. The Q_{ij} of g^2 pair-wise contrasts are arranged in a g by g matrix, with rows identifying the groups found in the bootstrap sample and the columns the groups in the PSU. The $g!$ permutations of the columns of this matrix are examined such that a minimum trace is found. The minimum trace is the value of S we are seeking. The g pair-wise contrasts in the main diagonal represent the best approximation of a one-to-one correspondence of the groups in the bootstrap sample with their nearest groups in the PSU.

G_k^* is compared to G_k^0 generated at each bootstrap iteration under the null hypothesis that the groups are sharp. If the null hypothesis is true, each group found by cluster analysis in the bootstrap sample is a random sample of the corresponding nearest neighbor group in the reference sample. For each group j with size n_j in the bootstrap sample, a random sample with n_j units is taken with replacement from its nearest group in the PSU. This is a null sample. The $n + n_k$ sampling units in the PSU and the null sample are put together. The

distance matrix used for the computation of G_k^* will provide the distances to make up the distance matrix between these units. With this matrix, the total sum of squares T^0 and contrast sum of squares Q_k^0 between already known nearest groups are computed. Their sum over g group contrasts will give

$$G_k^0 = 1 - \frac{\sum_{j=1}^g Q_j^0}{T^0}$$

The probability $P(G_k^0 \leq G_k^*)$ is defined as the proportion of iterations in which $G_k^0 \leq G_k^*$.

Due to the nature of the null hypothesis, the interpretation of $P(G_k^0 \leq G_k^*)$ differs from previous cases. The same considerations I made for the interpretation of probabilities in ordination evaluation are applicable, but reversed concerning significance. Likewise, the rejection of the null hypothesis based on $P(G_k^0 \leq G_k^*)$ is equated to sampling sufficiency under similar circumstances. Thus, if $P(G_k^0 \leq G_k^*)$ is not larger than a specified threshold α , we reject the null hypothesis and sample size n_k is deemed sufficient. This is equivalent to stating, with a probability $P(G_k^0 \leq G_k^*)$ of being wrong, that the groups are fuzzy and unstable. A $P(G_k^0 \leq G_k^*)$ larger than a specified threshold α may indicate the g groups in the partition are sharp, i.e., consistently reappear in resampling, or that the sample size n_k is too small. Whether sample size n_k is sufficient or not is in this case judged by inspection of the probabilities for stability across increasing sample sizes.

Sampling designs

The bootstrap method was originally described assuming unrestricted random sampling with replacement in data collection (Efron and Tibshirani 1993). Nevertheless, random sampling with replacement is not the rule in survey sampling, even more in ecological surveys. This problem would require modifications in the bootstrap method to accommodate stratified and systematic sampling (Rao & Wu 1988, Efron & Tibshirani 1993 p. 396), if not hindering the application of bootstrap in preferential sampling. I argue that this limitation is real only when estimation (e.g., means, variances) is the objective of the survey. Random sampling in the field should not be a requirement for the bootstrap when the objective is pattern recognition. In this case it seems any sampling design is valid. We may not be interested in making inferences about the real sampling universe, but to make inferences about

patterns in the sampling universe that is in the sample (PSU), no matter how the real sampling universe is represented. This being accepted, sampling design is not a limitation for bootstrap resampling in pattern recognition.

The PSU simulated by resampling with replacement the data set with n sampling units is actually infinitely larger than n (though in essence with the same properties). If the real sampling universe is infinitely large, there is no inconsistency between applying resampling with replacement in a data set acquired by sampling without replacement, the norm in ecological surveys. It is irrelevant replacing or not replacing sampling units in a real survey of an infinitely large sampling universe. As explained, sampling units in ecological surveys are often arbitrary dissections of a continuum and, therefore, the sampling universe is infinitely large. The problem with bootstrap without replacement is that the number of different sample combinations is a function of sample size, being equal to n at $n_k = 1$, maximal around $n_k = n/2$, and just one at $n_k = n$. That is, the confidence intervals will invariably converge when $n_k = n$, which precludes unbiased judgment on sample sufficiency at sizes larger than $n/2$ (Cadenazzi 1996, Bros & Cowell 1987). Alternative methods in Sitter (1992) can partially avoid this limitation.

In systematic sampling, consider a transect of length L measured on a continuous scale and consider quadrats with length $m < L$ located on the transect. In unrestricted random sampling there is no assurance that the quadrats will not overlap and there will be an infinitely large number of possible locations to lay the quadrats. In systematic sampling, if L/m is an integer and the sampling interval i is equal to m , there will be contiguity and no overlap of quadrats in any sample with L/i quadrats. Notice that bootstrap results may be affected by lack of independence of sampling units (Efron & Tibshirani 1993 p. 396). This may occur when the units are close in space, especially when quadrats are contiguous. If the sampling interval is such that $i > m$, there will be no contiguity; if $i < m$ there will be overlap. The pivot of systematic sampling, if randomly sited, is a location given by a distance from one of the transect ends, a real number in the interval $[0, i]$, which by convention may indicate where the quadrat side closest to the transect end is placed. Obviously, under these circumstances there is an infinitely large number of possible pivots and, thus, an infinitely large number of possible samples with size L/i . The problem of fractional units is always present in quadrat sampling. To avoid this, the transect ends may be viewed as if they were connected. Truncated

quadrats located at one end of the transect may continue on the other end, but this is cumbersome and may not be worth the effort. In fact, the effect of eliminating these quadrats from the sample is inversely proportional to the L/m ratio.

Preferential sampling in vegetation ecology is a tradition, in which quadrats are placed on sites that are perceived as typical of given conditions indicated by homogeneity (Braun-Blanquet 1979, Mueller-Dombois and Ellenberg 1974). Preferential sampling has been questioned as a valid sampling technique when the objective is conventional statistical analysis, on the grounds that it does not supply a representative sample (Orlóci 1978, Greig-Smith 1983, Kenkel et al. 1989) or lacks explicit and repeatable procedures (Goedickemeier et al. 1997). However, adopting a less orthodox view, Orlóci (1993) considers preferential sampling a shortcut to pattern recognition and points out that much of biological knowledge is rooted in non-statistical sampling. I go further and view the sample as one possible outcome of the repeated application by the researcher of his/her criteria in selecting sampling units in the same area. If this view is accepted, there is a sampling universe defined by all these possibilities and thus it is valid to examine sampling sufficiency by bootstrap resampling, as in the other sampling methods. The sampling universe is infinitely large considering that the researcher lays the units on a continuum. Upon repeating the sampling with 'typical quadrats' selected there will be variation from one trial to another in the exact quadrats' locations. Therefore, when the sampling objective is pattern recognition, there is no clear reason to believe that bootstrap resampling with replacement (and sampling sufficiency evaluation) is unjustified in data collected by preferential sampling. Whether random and preferential samples, taken by different researchers on the same area, are representative of the same sampling universe, however, is a different question that is not answered by evaluation of sampling sufficiency.

Examples

Data sets

I use two data sets, both from grasslands in the South of Brazil. There are 60 quadrats in one set, each 0.5 x 0.5 m selected on a 30 ha natural area at UFRGS experimental station (EEA) near Porto Alegre, Rio Grande do Sul. Quadrat siting was preferential, aimed at finding homogeneous patches. The community variables are visual estimates of cover-abundance (1-9 scale) for 60 species. Pillar et al. (1992) describe the survey procedure and the variables.

The other data set is from a large scale survey (Tcacenco and Pillar 1996). This set includes 37 quadrats, each 30 x 90 m in anthropogenic vegetation. The percentage cover of 12 species was evaluated by a point quadrat method. The survey covered a 4000 km² area in Santa Catarina state. Quadrat siting was systematic on the regional scale and preferential on the local scale.

As a reference, the methods are also tested on a random 'community' data set with 60 sampling units and 60 variables in the interval [0,1]. The example for sampling sufficiency determination in group partitions also includes an artificial data set containing 60 sampling units described by 60 random variables, divided in three well defined groups: In units 1-20 the observations in the variables are random numbers in the interval [0,1]; in units 21-40 they are in the interval [10,11]; and in units 41-60 they are in the interval [15-16].

Sampling sufficiency for ordination

The analysis used principal coordinates analysis of Euclidean distance matrices defined for sampling units. As expected, the ordination axes generated by bootstrapping the random data set were not significant, as indicated by the $P(\theta_k^0 \geq \theta_k^*)$ values in Fig. 2A-B. By inspection, the probability graphs for the first and second axes are steadily near $P(\theta_k^0 \geq \theta_k^*) = 0.5$. The uniformity suggests that this pattern should continue with sample sizes even larger than 60. We can conclude that sample sizes larger than 5 are sufficient to determine the non-significance of the ordination axis in this data set.

The analysis with the EEA grassland community data set indicated significance ($\alpha = 0.05$) and sampling sufficiency (sample sizes larger than 45) for the first ordination axis (Fig. 2C). There was no significance and no sampling sufficiency (up to 60 sampling units) concerning the interpretation of the probabilities of the second axis considered in isolation (Fig. 2D). However, the second and third axes taken jointly were significant and the probabilities flatten at sample sizes larger than 45. This suggests that in the EEA data set interpretations of the vegetation variation on the first three ordination axes are stable when based on samples sizes larger than 45 (Fig. 2E). The results from the Santa Catarina grassland community data set indicated sampling sufficiency (probabilities flatten at sample sizes larger than 25) but no significance ($\alpha = 0.05$) for the first three ordination axes (Fig. 2F-H). By setting a less strict $\alpha = 0.1$, however, the first axis would be significant. In the same data set, the first and second axes considered jointly were found not significant (graph not shown).

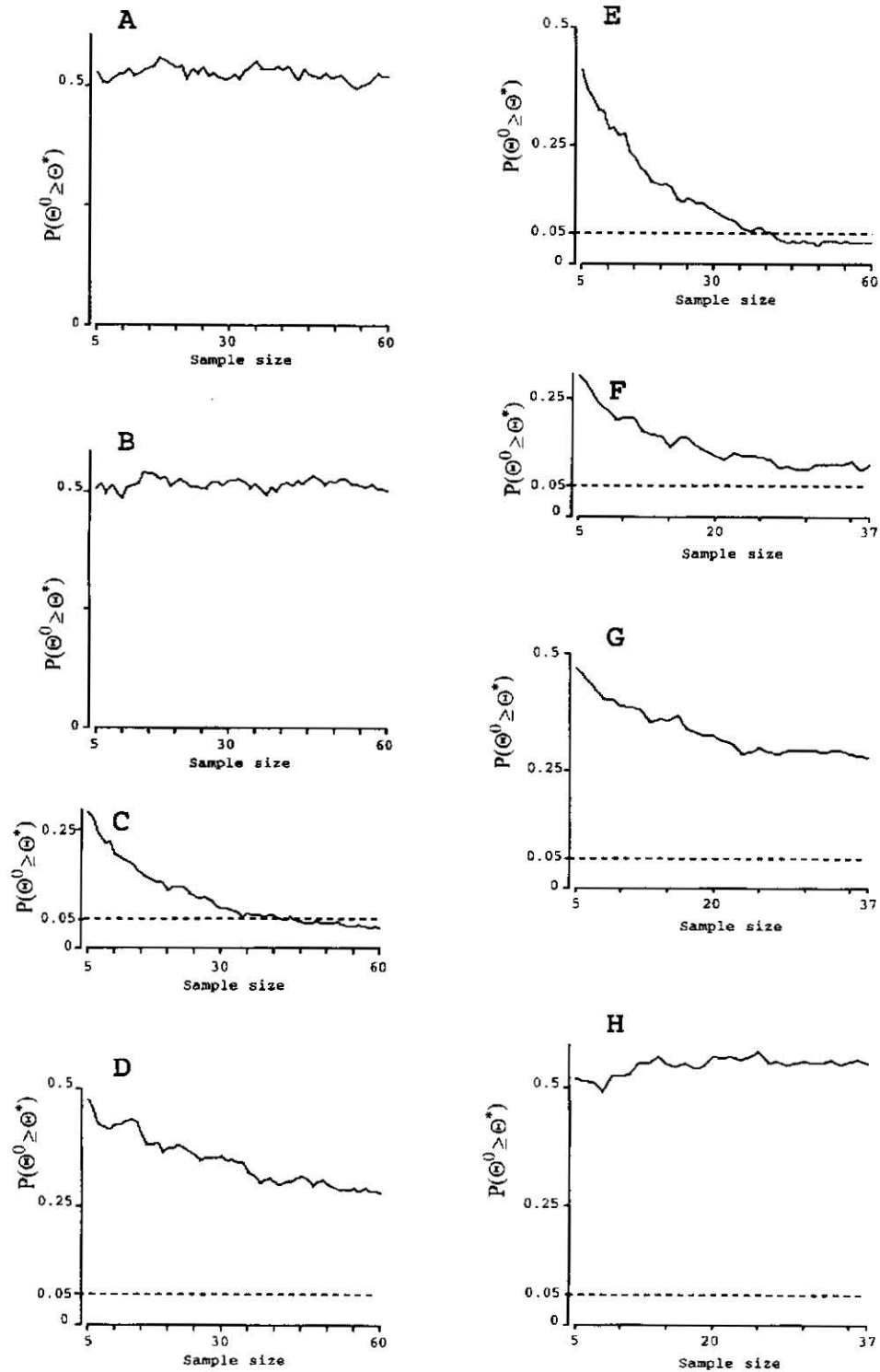


Figure 2. Evaluation of sampling sufficiency and significance for ordination structures in different data sets by probabilities $P(\theta_k^0 \geq \theta_k^*)$ for $\theta_k^* = \rho(D_k; D_n)$. The probabilities were generated in 1000 iterations of bootstrap resampling at each sample size. The ordination method is principal coordinates analysis. Data sets and ordination axes are the following: Data set with 60 units described by random variables, axes 1 (A) and 2 (B) are considered; EEA grassland data set with 60 sampling units and 60 species (Pillar et al. 1992), using axes 1 (C), 2 (D) and 2-3 (E); Santa Catarina grassland data set (Tcenceno & Pillar 1986) with 37 sampling units and 12 species, using axes 1 (F), 2 (G) and 3 (H).

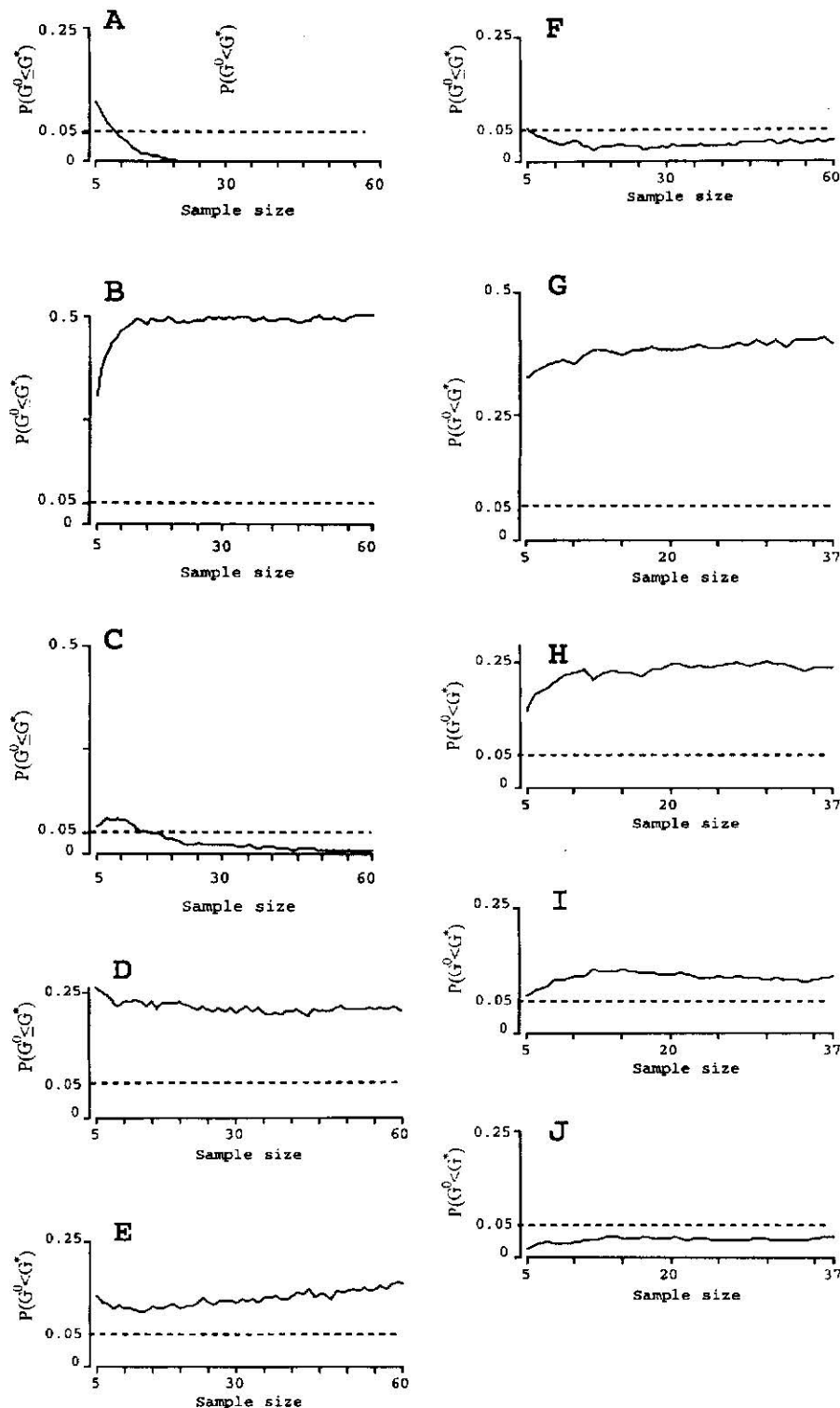


Figure 3. Evaluation of sampling sufficiency and significance for group partition levels in different data sets by probabilities $P(G_k^0 \leq G_k)$. The probabilities were generated in 10000 iterations of bootstrap resampling at each sample size. Data sets and partition levels are: (A) Artificial data of 60 units described by random variables, partition level 3 groups; (B-C) Artificial data set of 3 well defined groups, partition levels 3 and 4 groups; (D-F) EEA grassland data set (Pillar et al. 1992), partition levels 2, 3 and 4 groups; and (G-J) Santa Catarina grassland data set (Tcacenco & Pillar 1986), partition levels 2, 3, 4 and 5 groups. The groups were defined by sum of squares clustering.

Sampling sufficiency for cluster analysis

The examples used sum of squares clustering to generate group partitions at different levels. The results obtained with the random data show that probabilities $P(G_k^0 \leq G_k^*)$ are low and, therefore, that the groups are fuzzy and unstable at the three-groups partition level with sample sizes larger than 10 (Fig. 3A). The sample is sufficient, as shown by the decreasing and flat probability graph. The null hypothesis has to be rejected ($\alpha = 0.05$). The analyses of the artificial data set with three well-defined groups detected sharp partitions with three groups (Fig. 3B). Sampling sufficiency is indicated by the stability of the probabilities with sample sizes larger than 15. The groups are fuzzy and unstable, and sample size around 20 is sufficient, at a higher partition level (Fig. 3C).

The partitions generated by cluster analysis of the EEA sample are sharp when the number of groups is two or three, but not when the number of groups is four (Fig. 3D-F). Sample size sufficiency is reached near 25 sampling units for 2-group partitions (Fig. 3D). For 3-group partitions the probabilities are still increasing at sample size 60 (Fig. 3E), but this will not change the conclusion that the groups are sharp. Samples with more than 5-10 sampling units are sufficient to find that 4-group partitions are fuzzy (Fig. 3F). In the Santa Catarina grassland data set, the groups are sharp and sample sizes around 20 are sufficient when the number of groups is not larger than four (Fig. 3G-I). At partition level five the groups are not sharp (Fig. 3J) but the sample is sufficient since the probabilities level off around sample size 15.

Conclusions

Clearly, it is more productive to approach the sampling sufficiency problem through new flexible methods rather than trying to fit the problem into the Procrustean bed of classical statistics. Methods relying on bootstrap resampling are needed in view of the limitations of classical methods when applied in ecological contexts. The new approach I offer fulfills this need and is of general application. The examples demonstrate the utility of combining the idea of process sampling and the bootstrap method to support an objective appraisal of pattern stability and sampling sufficiency in the samples. Reliability is thus enhanced by this new approach despite the exploratory nature of ordination and cluster analysis and the use of unorthodox sampling designs.

The examples indicated that sampling sufficiency may be reached at different sample sizes for the same data set depending on the parameter considered in each test.

A sample size may be sufficient, for instance, to reveal groups at a certain partition level but insufficient to interpret the first two axes in ordination. Therefore, sampling sufficiency is obviously objective dependent.

Though the concepts are linked, a clear distinction must be made between sampling sufficiency and significance. The test may indicate that a sample is sufficient to evaluate the first ordination axis, to give an example, but also that the axis is not stable (significant) in revealing similar patterns in repeated samples with the same size. Similarly, a sample may be found sufficient to evaluate group structure sharpness at a given partition level, but the group structure may not necessarily be sharp.

Bootstrap methods are computationally demanding, but not beyond the capabilities of present day microcomputers. For instance, computation time with program **SAMPLER**, on a Macintosh equipped with a PowerPC 603e 200 MHz CPU, to run the total number of iterations at each sample size, varied from 13 to 87 seconds in the examples in Fig. 2 and from 23 to 83 seconds in Fig. 3.

Acknowledgments: This work was supported by grants and fellowships I received from CNPq and CAPES (Brazil) and was in part executed at the Department of Plant Sciences, University of Western Ontario, London, Canada. I thank L. Orlóci for hospitality and gratefully acknowledge his contribution to ideas here expressed. I also thank S. Camiz for comments.

References

- Beran, R. J. & M. Srivastava. 1985. Bootstrap test and confidence regions for functions of a covariance matrix. *Annals of Statistics* 13: 95-115.
- Braun-Blanquet, J. 1979. *Fitosociologia; bases para el estudio de las comunidades vegetales*. Blume, Madrid.
- Bros, W. E. & B. C. Cowell. 1987. A technique for optimizing sample size (replication). *Journal of Experimental Marine Biology and Ecology* 114: 63-71.
- Cadenazzi, M. 1996. *Otimização de amostragem para estudos de padrões em vegetação*. M.Sc. thesis, Faculdade de Agronomia, UFRGS, Porto Alegre, Brazil.
- Camiz, S. 1993. STATIS ordinations vs. the Juhász-Nagy models: The predictability of an exploratory tool. *Abstracta Botanica* 17: 29-36.
- Camiz, S. & A. Gergely. 1990. An exploratory method for determining optimal plot size in plant community studies. *Abstracta Botanica* 14: 83-108.
- Chaudhuri, A. & H. Stenger. 1992. *Survey sampling: theory and methods*. Marcel Dekker, New York, USA.
- Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. Wiley, New York, USA.

- Efron, B. 1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7: 1-25.
- Efron, B. & R. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, London.
- Gauch, H. G. 1989. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge.
- Goedickemeier, I., O. Wildi & F. Kienast. 1997. Sampling for vegetation survey: Some properties of a GIS-based stratification compared to other statistical sampling methods. *Coenoses* 12: 43-50.
- Gower, J. C. & P. Legendre. 1986. Metric and euclidean properties of dissimilarity coefficients. *Journal of Classification* 3: 5-48.
- Green, R. H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. Wiley, New York.
- Greig-Smith, P. 1983. *Quantitative Plant Ecology* 3rd ed. Blackwell, Oxford.
- Jackson, D. A. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* 74: 2204-2214.
- Jongman, R. H. G., C. J. F. ter Braak & O. F. R. van Tongeren (eds.). 1995. *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge.
- Juhász-Nagy, P. 1967. On some 'characteristic areas' of plant community stands. In: A. Rényi (ed.), *Proceedings of the Colloquium on Information Theory*, Bolyai Mathematical Society, Akadémiai Kiadó, Budapest, pp. 269-282.
- Juhász-Nagy, P. & J. Podani. 1983. Information theory methods for the study of spatial processes and succession. *Vegetatio* 51: 129-140.
- Kendall, M. & J. D. Gibbons. 1990. *Rank Correlation Methods* 5th ed. Edward Arnold, London.
- Kenkel, N. C., P. Juhász-Nagy and J. Podani. 1989. On sampling procedures in population and community ecology. *Vegetatio* 83: 195-207.
- Knox, R. G. 1989. Effects of detrending and rescaling on correspondence analysis: solutions on stability and accuracy. *Vegetatio* 83: 129-136.
- Knox, R. G., and R. K. Peet. 1989. Bootstrapped ordination: a method for estimating sampling effects in indirect gradient analysis. *Vegetatio* 80: 153-165.
- Krishnaiah, P. R. & C. R. Rao (eds.). 1988. *Sampling*. North-Holland, Amsterdam.
- Manly, B. F. J. 1992. Bootstrapping for determining sample sizes in biological studies. *Journal of Experimental Marine Biology and Ecology* 158: 189-196.
- Manly, B. F. J. 1993. A review of computer intensive multivariate methods in ecology. In: G. P. Patil & C. R. Rao, (eds.), *Multivariate Environmental Statistics*. North-Holland, Amsterdam, The Netherlands, pp. 307-346.
- Mueller-Dombois, D. & H. Ellenberg. 1974. *Aims and Methods of Vegetation Ecology*. Wiley, New York.
- O'Neill, R. V., D. L. DeAngelis, J. B. Waide & T. F. H. Allen. 1986. *A Hierarchical Concept of Ecosystems*. Princeton University Press, Princeton, N. J., USA.
- Orlóci, L. 1967. An agglomerative method for classification of plant communities. *Journal of Ecology* 55: 193-206.
- Orlóci, L. 1978. *Multivariate Analysis in Vegetation Research*. 2nd ed. Junk, The Hague, The Netherlands.
- Orlóci, L. 1993. The complexities and scenarios of ecosystem analysis. In: G. P. Patil & C. R. Rao (eds.), *Multivariate Environmental Statistics*. North-Holland, Amsterdam, The Netherlands, pp. 423-432.
- Orlóci, L. & V. D. Pillar. 1989. On sample size optimality in ecosystem survey. *Biométrie-Praximétrie* 29: 173-184.
- Orlóci, L. & V. D. Pillar. 1991. On sample size optimality in ecosystem survey. In: E. Feoli & L. Orlóci (eds.), *Computer Assisted Vegetation Analysis*. Kluwer, Dordrecht, The Netherlands, pp. 41-46.
- Palmer, M. W. 1990. The estimation of species richness by extrapolation. *Ecology* 71: 1195-1198.
- Palmer, M. W. 1991. Estimating species richness: the second-order jackknife reconsidered. *Ecology* 72: 1512-1513.
- Patil, G. P., G. J. Babu, R. C. Hennemuth, W. L. Myers, M. B. Rajarshi, and C. Taillie. 1988. Data-based sampling and model-based estimation for environmental resources. Pages 489-513 in P. R. Krishnaiah, and C. R. Rao, editors. *Sampling*. North-Holland, Amsterdam, The Netherlands.
- Pielou, E. C. 1975. *Ecological Diversity*. Wiley, New York.
- Pielou, E. C. 1984. *The interpretation of Ecological Data. A primer on Classification and Ordination*. Wiley-Interscience, New York, USA.
- Pillar, V. D. (manuscript) The bootstrapped ordination reexamined.
- Pillar, V. D. (in press) How sharp are classifications? *Ecology*.
- Pillar, V. D., A. V. A. Jacques & I. I. Boldrini. 1992. Fatores de ambiente relacionados à variação da vegetação de um campo natural. *Pesquisa Agropecuária Brasileira* 27: 1089-1101.
- Pillar, V. D. & Orlóci, L. 1996. On randomization testing in vegetation science: multifactor comparisons of relevé groups. *Journal of Vegetation Science* 7: 585-592.
- Podani, J. 1994. *Multivariate data analysis in ecology and systematics*. SPB, The Hague, The Netherlands.
- Podani, J., T. Czárán & S. Bartha. 1993. Pattern, area and diversity: the importance of spatial scale in species assemblages. *Abstracta Botanica* 17: 37-51.
- Poore, M. E. D. 1955. The use of phytosociological methods in ecological investigations. *Journal of Ecology* 44: 28-50.
- Poore, M. E. D. 1956. The use of phytosociological methods in ecological investigations. IV General discussion of phytosociological problems. *Journal of Ecology* 44: 28-50.
- Rao, J. N. K. 1988. Variance estimation in sample surveys. In: P. R. Krishnaiah & C. R. Rao (eds.), *Sampling*. North-Holland, Amsterdam, The Netherlands, pp. 427-447.
- Rao, J. N. K. & C. F. J. Wu 1988. Resampling inference with complex survey data. *Journal of the American Statistical Association* 83: 231-241.
- Sampford, M. R. 1962. *An Introduction to Sampling Theory*. Oliver and Boyd, Edinburgh, UK.

- Schönemann, P. H. & R. M. Carroll. 1970. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika* 35: 245-256.
- Sitter, R. V. 1992. Bootstrap methods for survey data. *Canadian Journal of Statistics* 20: 135-154.
- Stauffer, D. F., E. O. Garton & R. K. Steinhorst. 1985. A comparison of principal components from real and random data. *Ecology* 66: 1693-1698.
- Teacenco, F. A. & V. D. Pillar. 1996. Vegetation-environment relations in anthropogenic grasslands of northeastern Santa Catarina, Brazil. *Coenoses* 11: 103-108.
- Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58: 236-244.
- Wiens, J. A. 1989. Spatial scaling in ecology. *Functional Ecology* 3: 385-397.
- Wilson, M. V. 1981. A statistical test of the accuracy and consistency of ordinations. *Ecology* 62: 8-12.