Ecology, 80(8), 1999, pp. 2508–2516

HOW SHARP ARE CLASSIFICATIONS?

VALÉRIO DEPATTA PILLAR[1]

Departamento de Ecologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 91540-000, Brazil

Abstract. Ecologists often use cluster analysis as a tool in the classification and mapping of entities such as communities or landscapes. The problem is that the researcher has to choose an adequate group partition level. In addition, cluster analysis techniques will always reveal groups, even if the data set does not have a clear group structure. This paper offers a method to test statistically for fuzziness of the partitions in cluster analysis of sampling units that can be used with a wide range of data types and clustering methods. The method applies bootstrap resampling. In this, partitions found in bootstrap samples are compared to the observed partition by the similarity of the sampling units that form the groups. The method tests the null hypothesis that the clusters in the bootstrap samples are random samples of their most similar corresponding clusters mapped one-to-one into the observed data. The resulting probability indicates whether the groups in the partition are sharp enough to reappear consistently in resampling. Examples with artificial and vegetational field data show that the test gives consistent and useful results. Though the method is computationally demanding, its implementation in a C++ program can run very fast on microcomputers.

Key words: bootstrap resampling; classification methods; cluster analysis; fuzzy vs. sharp groups; group structure; mapping; partition, appropriate level for cluster analysis; resampling; sampling; similarity, statistical analysis; vegetation, cluster analysis.

INTRODUCTION

Cluster analysis is a useful tool in classification when the description of the objects is multivariate. Ecologists often apply cluster analysis for the classification and mapping of entities, such as communities or landscapes. For this purpose many clustering algorithms are available, combined with many resemblance indices (Orlóci 1978, Pielou 1984, Podani 1994). Although the algorithms are objective, to obtain a classification the researcher has to make a choice, which is often subjective, of an adequate group partition level. Furthermore, cluster analysis techniques will always reveal groups, even if the data set does not have a clear group structure.

Milligan and Cooper (1985) evaluated many procedures available at that time to find the number of clusters; they used simulated data sets with different numbers of known clusters, dimensions, and group sizes. The best criterion they found was the use of the Calinski and Harabasz (1974) index, which is a ratio of pooled sum of squares between and within clusters; in the tests, this method determined the correct number of clusters in 90% of the cases evaluated. Other criteria also had good overall performance, but did not perform well in data sets with two known clusters. The authors, however, warn that the performances may be data dependent. Also, since all simulated data sets had well-defined clusters, there was no evaluation of the criterion ''resolving power.'' Dale (1988) discusses at length possible alternatives to finding an optimal partition level. According to him, there is no unique answer, for different objectives are sought, and most of the methods are specific in their requirements.

The use of probabilistic resemblance is one of the solutions I offered earlier (Pillar 1996) to evaluate group homogeneity in plant community classification. In this, the probabilities are generated by randomization under the null hypothesis of random community composition. A similar null hypothesis is involved in testing randomness of species co-occurrences in islands (Connor and Simberloff 1979, Wright and Biehl 1982, Gilpin and Diamond 1987, Wilson 1987, 1988, Manly 1993), or in testing the significance of species clusters (Strauss 1982, Jaksic and Medel 1990). However, there is no agreement on exactly how the algorithm should implement the null hypothesis of random composition. All taxa may or may not have the same probability to occur in a community. The richness of each site may or may not be constant. Also, there are different ways to treat quantitative compositional data. Again, even if we agree on a reasonable implementation, the solution is specific, in that random composition has to be a relevant null hypothesis. It is not relevant if the variables describing the sampling units are not compositional, such as climate, soil, or land use attributes.

The bootstrap is a resampling method devised by Efron (1979, Efron and Tibshirani 1993) that adopts

Manuscript received 1 December 1997; revised 27 October 1998; accepted 13 November 1998; final version received 14 December 1998.

[1] E-mail: vpillar@ecologia.ufrgs.br

the principle that, in the absence of better information, the distribution of observations in a sample is the best indicator of the distribution in the sampling universe. If this principle is tenable, resampling a sample with replacement will mimic resampling the sampling universe. Each sample, obtained by resampling the sample, is a ''bootstrap'' sample. The distribution of a given attribute in the sampling universe can be estimated by the distribution of the same attribute computed from bootstrap samples. Confidence intervals can then be defined and probabilities evaluated.

A few antecedents use bootstrap resampling to evaluate the significance of clusters. Peck et al. (1989) apply bootstrap to find an optimal number of clusters that maximize the parsimony of clustering and minimize the loss of information. Felsenstein (1985, 1988) applies bootstrap resampling to set confidence limits on phylogenies, yet Manly (1993) and Brown (1994) criticize the technique. In Felsenstein's method, the objects of resampling are not the species, but the vectors of character states that describe the species. This is questionable, for the method considers the set of vectors as a random sample from a pool of potential independent characters. Nemec and Brinkhurst (1988) use bootstrap to evaluate the significance of clustering steps, but their solution is specific to cases in which there are replicates of each object, such as communities or sites. In the same paper, they use bootstrap resampling to compare group partitions of the same set of objects. The comparison involves shared objects, and the test in one instance has an objective akin to the Mantel (1967) test.

Here, I offer a more general method, based on bootstrap resampling. It takes a data set from survey sampling, performs a cluster analysis of sampling units, and then examines the stability of the partition at a given level by simulated resampling of the same universe. Unstable partitions will indicate the data set has a fuzzy group structure at the level being considered. In the resampling, different samples will likely not contain exactly the same objects, but the cluster analysis may produce similar groups. This method differs from those described in Nemec and Brinkhurst (1988) by incorporating a broader notion of similarity: partitions are considered similar not only when they share the same sampling units, but also when they share more similar sampling units. For the same reason, my method does not use any of the suggested measures reviewed in Podani (1989) to compare partitions with the same objects.

Method

A data set of n sampling units and p variables is considered. The method supplies the probability needed to evaluate the stability of the partition. A partition with k groups is represented by a vector of n elements containing integer values from 1 to k. These values identify the group to which each sampling unit belongs.

An iterative algorithm uses the complete sample and its k-group partition as references to which other k-group partitions obtained in the sequel, by resampling the sample, are compared. Each iteration takes a bootstrap sample of n units with replacement and subjects these sampling units to cluster analysis. Then it computes G^*, the similarity of the k-group partition in the bootstrap sample, to the k-group partition in the reference sample, and computes G^0 under the null hypothesis that the partition is sharp. The comparison of G^* and G^0 concludes one iteration. Any clustering technique may be used. A numerical illustration of the method is given in Appendix A.

The G attribute*

The $2n$ sampling units in the reference sample and in the bootstrap sample are viewed as if they were points in the same space defined by p variables. Similarly, in the same space, we can compare the k groups in the reference sample with the k groups in the bootstrap sample. In this space, G^* is a relative measure of agreement between reference and bootstrap samples defined for the k-group partition level by

$$G^* = 1 - \frac{S}{T}. \tag{1}$$

In this equation, T is a total sum of squares, defined by

$$T = \frac{1}{2n} \sum_{h=1}^{2n-1} \sum_{i=h+1}^{2n} d_{hi}^2. \tag{2}$$

It is seen that T involves $2n(2n - 1)/2$ pair-wise squared dissimilarities (d_{hi}^2) between the sampling units in the pooled samples. S is the sum of squares between groups attributed to sampling, i.e., the 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 reference sample.

Finding the value of S involves the computation of sum of squares for all k^2 pair-wise contrasts between the k groups in the bootstrap sample and the k groups in the reference sample. The computation of contrasts based on the distance matrix follows Pillar and Orlóci (1996). For any pair-wise contrast j between one of the groups in the bootstrap sample and another group in the reference sample, we compute the following: T_j, the total sum of squares involving squared distances of the sampling units in the two groups; $W_{j\mathrm{B}}$, the sum of squares within the group in the bootstrap sample; and $W_{j\mathrm{R}}$, the sum of squares within the group in the reference sample. T_j is based on Eq. 2, but using only the squared distances of the $n_{c\mathrm{B}} + n_{c\mathrm{R}}$ sampling units in the groups involved in the contrast. $W_{j\mathrm{B}}$ is defined by

$$W_{j\mathrm{B}} = \frac{1}{n_{c\mathrm{B}}} \sum_{h=1}^{2n-1} \sum_{i=h+1}^{2n} d_{hi}^2 \delta(h, i, c_{\mathrm{B}}) \tag{3}$$

where $n_{c\mathrm{B}}$ is the size of group c in the bootstrap sample,

and the indicator variable $\delta(h, i, c_B) = 1$, if sampling unit h and i belong to group c_B, and takes on the value zero otherwise. W_{jR} is similarly defined. The between groups sum of squares for contrast j is then

$$Q_j = T_j - (W_{jB} + W_{jR}). \quad (4)$$

The sum of squares Q_j of k^2 pair-wise contrasts are arranged in a $k \times k$ matrix, with rows identifying the groups found in the bootstrap sample and the columns identifying the groups in the reference sample. The $k!$ 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 k pair-wise contrasts in the main diagonal indicate the best approximation of a one-to-one correspondence of the groups in the bootstrap sample with the groups in the reference sample. If k is large, too many permutations will have to be examined. However, the computations may use random permutations and an algorithm that seeks minimum values in the matrix, repeating the process until a stable minimum trace is found. The stable minimum trace is a good approximation of the true value of S.

It is important to note that the additivity of the terms in Eq. 4 and, as a consequence, the computation of G^* is valid only for a dissimilarity matrix with Euclidean metric properties (Orlóci 1978, Gower and Legendre 1986). As a simplification, computing a sum of squares from the distance matrix avoids the need for going back to the original data to calculate group centroids. This also enables the application of the method to data sets of any type (quantitative, presence–absence, qualitative, or mixed), provided an appropriate dissimilarity measure is used.

The G^0 attribute

The similarity index G^* may range from zero, a conceptual limit, to one, when bootstrap and reference partitions coincide. The sharper the partition, the more likely that the groups in the bootstrap sample will recover the group partition in the observed sample, and, thus, the larger will be G^*. However, G^* close to 1 does not by itself indicate that the partition is not fuzzy. G^* must therefore be compared to a value of G^0 generated under the null hypothesis that the groups are sharp. The null hypothesis states that 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 the same size n_j and with replacement is taken from its nearest group in the reference sample. This is the null bootstrap sample. The $2n$ units in the reference and the null bootstrap sample are put together. The distance matrix used in Eq. 2 will provide the distances to make up the distance matrix between these units. With this matrix, the total sum of squares T^0 (Eq. 2) and the contrast sum of squares Q_j^0 (Eq. 4) are computed. Their sum over k group contrasts between reference and null bootstrap sample will give the following:

$$G^0 = 1 - \frac{\sum_{j=1}^{k} Q_j^0}{T^0}. \quad (5)$$

After performing a large number of resampling iterations, the probability $P(G^0 \leq G^*)$ is defined. The larger the number of iterations B, the more stable and closer to its true value will be $P(G^0 \leq G^*)$. The choice of the number of iterations should balance precision, efficiency of the implemented algorithm, and available computing power.

If $P(G^0 \leq G^*)$ is not larger than a specified threshold α, we conclude, with a probability $P(G^0 \leq G^*)$ of being wrong, that the k groups in the partition are not sharp enough to consistently reappear in resampling. That is, we reject the null hypothesis and conclude that the groups are fuzzy. If, instead, we accept the null hypothesis, we conclude that there is not enough evidence to refute that the groups are sharp.

Sample size influences the power of the group partition stability test. A small sample size will likely increase the error type II, the probability of accepting the null hypothesis when it is actually false. That is, $P(G^0 \leq G^*) > \alpha$ may either indicate that the group structure is sharp, or that the sample size is too small. Sample size sufficiency can be evaluated by examining the stability of probabilities across a series of sample sizes. For this, probabilities $P(G_k^0 \leq G_k^*)$ are found by bootstrap resampling for each sample size $n_k \leq n$, taken from the observed sample with n sampling units (Pillar 1998). The sample is considered sufficient if the probabilities reach stability within the range of sample sizes evaluated. Typical cases are depicted in Fig. 1.

EXAMPLES

Artificial data sets

The first example uses simulated data sets with two equally sized clusters separated by expected difference d in each variable ranging from $d = 0$ (a single group) to $d = 0.32$ (very sharp groups). The data sets with 40 sampling units contain 40 random variables. The values range from zero to one in one cluster and from $0 + d$ to $1 + d$ in the other cluster; also the variables have equal variances v within each cluster, in one case with normal ($v \approx 0.004$) distribution, and in the other case with uniform random distribution. Ten data sets were generated for each difference d, type of distribution, and partition level. Results are given in Fig. 2a. As seen, the values of $P(G^0 \leq G^*)$ for partitions with two groups are low for $d \leq 0.02$ with normal distribution, and for $d \leq 0.05$ with uniform distribution (Fig. 2a). This indicates the rejection (threshold $\alpha = 0.1$) of the null hypothesis of group sharpness, since there is no evident group structure other than by chance. As d increases slightly, the values of $P(G^0 \leq G^*)$ with two

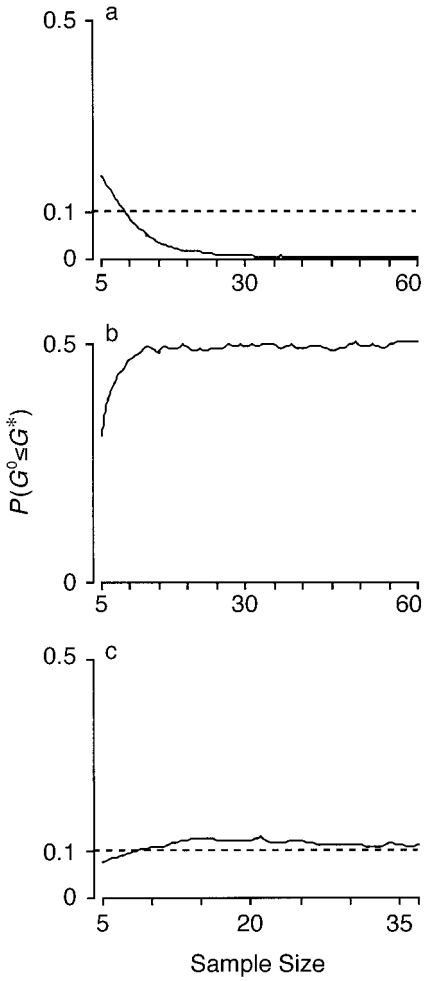

FIG. 1. Examples of the influence of sample size on the probabilities $P(G^0 \leq G^*)$. In (a), the data set is a 60 × 60 matrix with random numbers in the interval [0, 1]. The partition level is two groups, and the partition is found to be fuzzy with sample sizes >10. In (b), the dimensions are the same as in (a), but the observations are random numbers in the interval [0, 1] in sampling units 1–20, in the interval [10, 11] in sampling units 21–40, and in the interval [15, 16] in sampling units 41–60. The partition level is three groups, and the probabilities stabilize with sample sizes >15; thus the sample is sufficient to conclude that the groups are sharp. In (c), the partition level is four groups, and the data set is from Tcacenco and Pillar (1996), the same used in Table 2. Probabilities stabilize and are larger than $\alpha = 0.1$ with sample sizes >25.

groups steeply increase to ~0.5, correctly indicating that the groups are sharp at this partition level. The probability curve is monotonic within limits set by the standard deviations of the mean. With three groups (and higher partition levels not shown) the values of $P(G^0 \leq G^*)$ rightly indicate fuzziness for any d in both data types.

The second example uses similarly defined series of simulated data sets, but there is one sampling unit in one group and 39 in the other, with the variables normally distributed. The results are given in Fig. 2b. At partitions with two groups, the test indicates that the groups are fuzzy ($\alpha = 0.1$) for $d \leq 0.06$. Again, at this partition level there is a steep increase in the probabilities for larger values of d, correctly indicating two sharp groups. The values of $P(G^0 \leq G^*)$ at higher partition levels correctly indicate fuzziness (Fig. 2b for three-group partitions, others not shown).

The third example is similar to the first, but the data sets contain four variables instead of 40 (Fig. 2a). The reduced dimensionality does not affect the effectiveness of the test in detecting the correct number of groups (two), though the probabilities for three- and four-group partitions are close to the threshold $\alpha = 0.1$ (Fig. 2c). Partitions with two groups are fuzzy for $d \leq 0.04$.

In the fourth example, the simulated data sets have five equally sized groups. Values of d range 0–0.64, but expected differences within each data set have increasing values; that is, $0.25d$ between clusters 1 and 2, $0.5d$ between clusters 2 and 3, $0.75d$ between clusters 3 and 4, and d between clusters 4 and 5. The interpretation of the probabilities (Table 1), using $\alpha = 0.1$, indicates that partitions with the correct number of five groups are sharp (not fuzzy) at minimum $d \geq 0.04$, but partitions with two, three, or four groups, which are also real, are sharp at smaller minimum d. It is relevant to note that when five-group partitions are found to be sharp, lower level partitions should by consequence be sharp as well, unless ambiguities are present when groups are all equally separated, which is unlikely in nature. A pattern similar to the previous cases is observed: as d increases, probabilities $P(G^0 \leq G^*)$ increase monotonically and then stabilize. Partitions with six groups were found fuzzy for any d.

Real data sets

One of the examples uses a data set from Cadenazzi (1996) with 20 sampling units, which are vegetation quadrats described by the cover of 67 species. The 0.1 × 1 m quadrats are laid systematically on a 40 m transect on grassland in the Campos formation near Porto Alegre, Brazil. The data was checked for spatial autocorrelation, which was found nonsignificant. The results in Table 2 for this case indicate that the null hypothesis of sharp groups should be rejected, i.e., the groups are fuzzy at any partition level from two to nine groups ($\alpha = 0.1$).

Another example uses a data set with 37 sampling units. These are vegetation quadrats described by the percentage cover of 12 species (Tcacenco and Pillar 1996). The quadrats were 30 × 90 m in size and were laid in anthropogenic grasslands over 4000 km^2 in Santa Catarina state, Brazil, systematically on the regional scale and preferentially on the local scale. In this case, Table 2 indicates that the partitions with two, three, and four groups are not fuzzy, although the partition with four groups is very close to the threshold

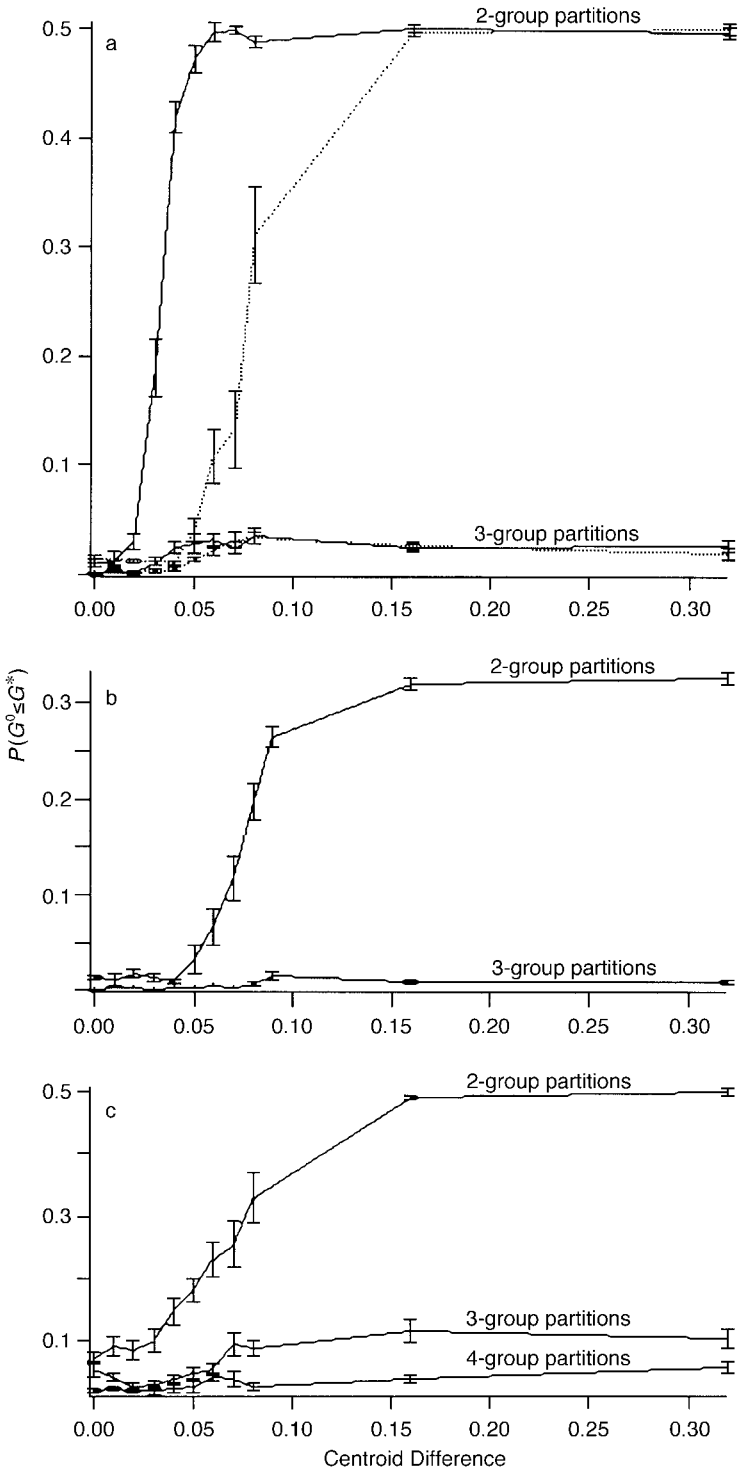

FIG. 2. Probability curve of $P(G^0 \leq G^*)$ for increasing separation between groups in simulated data. Random data sets were defined with two groups separated by expected difference d between centroids ranging from $d = 0$ (a single group) to $d = 0.32$ (clearly two groups). In (a) and (c), the groups have equal sizes (20 sampling units); in (b), one group has one sampling unit and the other group has 39 sampling units. The data contain 40 variables in (a) and (b) and four variables in (c), with normal (——) or uniform (⋯⋯ [only in (a)]) distribution within each group. Standard deviations of the means based on 10 data sets in each case are indicated. The partition level after cluster analysis is indicated on each line. The number of iterations is 1000 for each combination of centroid difference, partition level, distribution type, and data set replicate.

TABLE 1. Probabilities $P(G^0 \leq G^*)$ generated by bootstrap resampling on artificial data sets at different partition levels.

Centroid difference	Partition level				
	2	3	4	5	6
0	0.010 ± 0.002	0.002 ± 0.001	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
0.0025–0.01	0.010 ± 0.002	0.002 ± 0.001	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
0.005–0.02	0.159 ± 0.029	0.008 ± 0.002	0.004 ± 0.001	0.001 ± 0.000	0.000 ± 0.000
0.0075–0.03	0.259 ± 0.027	0.060 ± 0.017	0.016 ± 0.005	0.002 ± 0.001	0.000 ± 0.000
0.01–0.04	0.214 ± 0.025	0.172 ± 0.031	0.018 ± 0.004	0.004 ± 0.002	0.001 ± 0.000
0.0125–0.05	0.255 ± 0.043	0.239 ± 0.032	0.085 ± 0.018	0.017 ± 0.003	0.002 ± 0.001
0.015–0.06	0.360 ± 0.025	0.273 ± 0.022	0.237 ± 0.024	0.027 ± 0.005	0.003 ± 0.001
0.0175–0.07	0.355 ± 0.024	0.285 ± 0.033	0.313 ± 0.026	0.034 ± 0.003	0.008 ± 0.003
0.02–0.08	0.289 ± 0.034	0.312 ± 0.015	0.417 ± 0.015	0.035 ± 0.006	0.011 ± 0.002
0.04–0.16	0.352 ± 0.025	0.306 ± 0.012	0.499 ± 0.004	0.287 ± 0.032	0.041 ± 0.005
0.08–0.32	0.302 ± 0.017	0.324 ± 0.009	0.516 ± 0.008	0.507 ± 0.006	0.063 ± 0.008
0.16–0.64	0.333 ± 0.005	0.304 ± 0.007	0.503 ± 0.003	0.496 ± 0.004	0.059 ± 0.007

Notes: Table entries are means ± 1 SE, based on 10 data sets. The data sets are defined with five equally sized groups separated by expected difference d between centroids, ranging from $d = 0$ (a single group) to $d = 0.64$, which is variable within each data set: $0.25d$ between clusters 1 and 2; $0.5d$ between clusters 2 and 3; $0.75d$ between clusters 3 and 4; and d between clusters 4 and 5. The data sets contain 40 sampling units and 40 variables with normal distribution within each group, with random values ranging from 0 to 1 in the first group, from $0 + 0.25d$ to $1 + 0.25d$ in the second group, and so on. The number of iterations is 1000 in each data set.

$\alpha = 0.1$. Since the groups are not fuzzy, are they also sharp? Sample size reached sufficiency, as indicated by the fact that probabilities increase and then are consistently stable and $>\alpha$ at sample sizes >25, for two-, three-, and four-group partitions; the latter is depicted in Fig. 1c. Since sample size is sufficient, we conclude that the groups at these partition levels are indeed likely sharp. Interestingly, Tcacenco and Pillar (1996) detected four grassland types by inspection of the dendrogram in Fig. 3b, a decision now supported by the partition significance test. Indeed, this dendrogram shows a sharper increase in the dissimilarity level at the last few clustering steps than does the dendrogram in Fig. 3a. Similarly, a sharper group structure can be seen in the point configuration in Fig. 4b than in Fig. 4a.

DISCUSSION

The results with the simulated data sets indicated accurate numbers of known groups, for sufficiently large cluster separation, in all cases examined. The use of $\alpha = 0.1$ as a threshold seems adequate. A more strict (smaller) threshold, say $\alpha = 0.05$, would have indicated sharpness in several cases of partitions known to be fuzzy (three-group partitions in Fig. 2c and six-group partitions in Table 1). This would lead to increased error type II, the probability of accepting the null hypothesis when it is actually false.

The hypothesis test actually proves fuzziness of the group partition by rejecting the null hypothesis. If the null hypothesis is accepted, either the groups are really sharp or sample size is too small. It may be argued that the logic of proving sharpness by accepting a null hypothesis is faulty. It is clear that sample size may affect the results of the test. The decision of discriminating between sharpness and insufficient sample size is nonstatistical; nevertheless, it is helped by the examination of sample size sufficiency prior to making conclusions. Upon the null hypothesis being accepted and sample being sufficient, concluding that the groups are sharp is not less faulty than, for instance, concluding that there is no effect of treatments in an experiment.

It can be shown that the sampling universe, simulated in bootstrap resampling with replacement, is infinitely larger than the sample itself. If the real sampling universe is also infinitely large, there is no inconsistency of resampling with replacement in a data set acquired by sampling without replacement. This is so, because replacing or not replacing sampling units in a real survey of an infinitely large sampling universe is without consequence. Infinitely large sampling universes are defined, for instance, when the sampling units are ag-

TABLE 2. Probabilities $P(G^0 \leq G^*)$ generated by bootstrap resampling on two vegetation data sets at different partition levels (number of groups).

No. quadrats	No. species	Number of groups							
		2	3	4	5	6	7	8	9
20†	67	0.0446	0.0233	0.0203	0.0205	0.0154	0.008	0.0036	0.0029
37‡	12	0.4128	0.2364	0.0979	0.0394	0.0278	0.0193	0.0099	0.0061

Note: The number of iterations is 10 000.
† Cadenazzi (1996).
‡ Tcacenco and Pillar (1996).

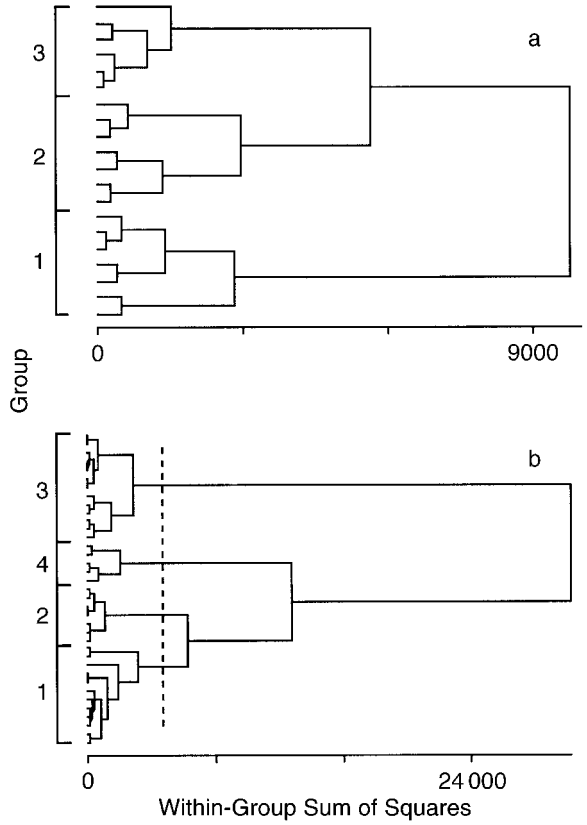

FIG. 3. Cluster analysis of (a) 20 grassland quadrats, 0.1 × 1 m (Cadenazzi 1996), and (b) 37 grassland quadrats, 30 × 90 m (Tcacenco and Pillar 1996). Partitions with any number of groups are fuzzy in (a) as seen in Table 2. A four-group partition is identified in (b), which is sharp, as concluded from Table 2. The clustering criterion is minimum variance, and the analysis uses Euclidean distances.

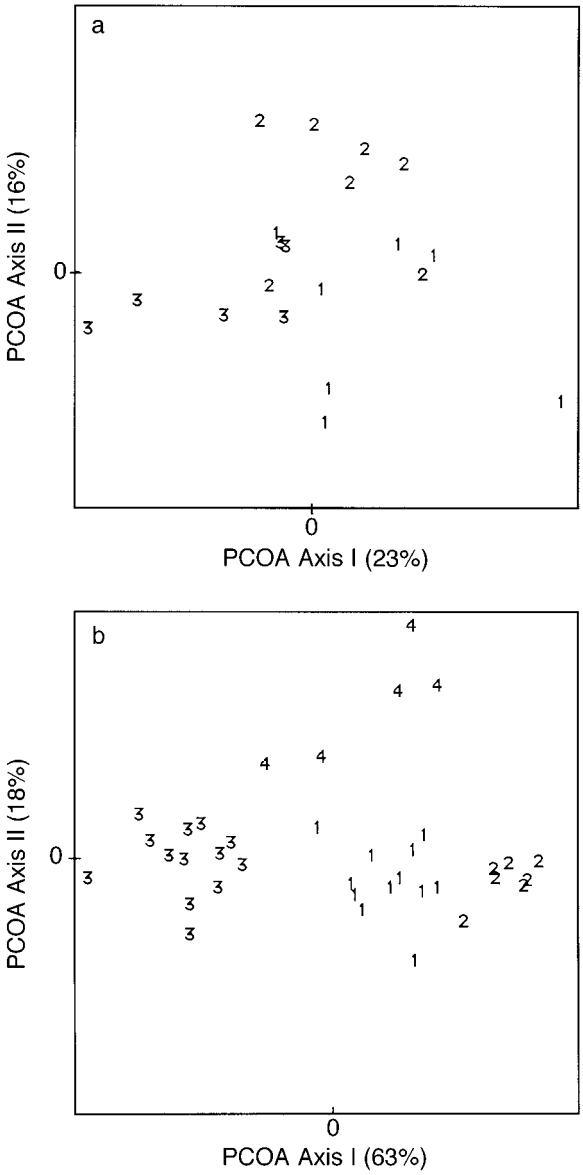

FIG. 4. Ordination of (a) 20 vegetation quadrats, 0.1 × 1 m (Cadenazzi 1996), and (b) 37 grassland quadrats, 30 × 90 m (Tcacenco and Pillar 1996). The method is principal coordinates analysis from Euclidean distances. The labels identify groups found in cluster analysis (see Fig. 3).

gregates with size and shape set by the investigator, as in quadrat sampling in an ecological study. The resampling with replacement must be understood as a way to simulate the pseudosampling universe and not the actual sampling process. Likewise, simulating an infinitely large sampling universe is compatible with sampling with replacement in a finite sampling universe.

The algorithm is computationally demanding, but its implementation in C++ is reasonably fast on recent microcomputers (see Appendix B for availability of the application program). On a Macintosh computer with a PowerPC 200 MHz 603e CPU, the run time for generating the probabilities in the first case in Table 2 took 11 s for two groups and 69 s for nine groups. It took 32–117 s to compute the probabilities for the second case in Table 2.

It is true that the results of the test will be influenced by insufficient sample size. In addition, the results may be affected by lack of independence of the sampling units (Efron and Tibshirani 1993:396), such as in contiguous quadrats. Nevertheless, the method I offer to test statistical significance of partitions in cluster analysis of sampling units possesses the following positive attributes: (1) it is generally applicable under a broader range of conditions than other available methods; (2) it produces consistent results in artificial data sets with known properties; (3) it has proven to be useful in real data sets; and (4) it is computationally feasible even on microcomputers.

ACKNOWLEDGMENTS

This research received financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Cien-

tífico e Tecnológico (CNPq; Brazil) and was in part executed during a leave at the Department of Plant Sciences, University of Western Ontario, Canada. I gratefully acknowledge M. Cadenazzi for allowing me to use her data and L. Orlóci, D. Jackson and P. Dixon for helpful suggestions.

Literature Cited

Brown, J. K. M. 1994. Bootstrap hypothesis tests for evolutionary trees and other dendrograms. Proceedings of the National Academy of Sciences of the USA **91**:12293–12297.

Cadenazzi, M. 1996. Otimização de amostragem para estudos de padrões em vegetação. Thesis. Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Calinski, R. B., and J. Harabasz. 1974. A dendrite method for cluster analysis. Communications in Statistics **3**:1–27.

Connor, E. F., and D. Simberloff. 1979. The assembly of species communities: chance or competition? Ecology **60**: 1132–1140.

Dale, M. B. 1988. Knowing when to stop: cluster concept—concept cluster. Coenoses **3**:11–32.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. The Annals of Statistics **7**:1–25.

Efron, B., and R. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, London, UK.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

———. 1988. Phylogenies from molecular sequences: inference and reliability. Annual Review in Genetics **22**:521–565.

Gilpin, M. E., and J. M. Diamond. 1987. Comments on Wilson's null model. Oecologia **74**:159–160.

Gower, J. C., and P. Legendre. 1986. Metric and Euclidean properties of dissimilarity coefficients. Journal of Classification **3**:5–48.

Jaksic, F. M., and R. G. Medel. 1990. Objective recognition of guilds: testing for statistically significant species clusters. Oecologia **82**:87–92.

Manly, B. F. J. 1993. A review of computer intensive multivariate methods in ecology. Pages 307–346 *in* G. P. Patil and C. R. Rao, editors, Multivariate environmental Statistics, Elsevier, Amsterdam, The Netherlands.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research **27**:209–220.

Milligan, G. W., and M. C. Cooper. 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika **50**:159–179.

Nemec, A. F. L., and R. O. Brinkhurst 1988. Using the bootstrap to assess statistical significance in the cluster analysis of species abundance data. Canadian Journal of Fisheries and Aquatic Sciences **45**:965–970.

Orlóci, L. 1978. Multivariate analysis in vegetation research. Second edition. Junk, The Hague, The Netherlands.

Peck, R., L. Fisher, and J. V. Ness 1989. Approximate confidence intervals for the number of clusters. Journal of the American Statistical Association **84**:184–191.

Pielou, E. C. 1984. The interpretation of ecological data. A primer on classification and ordination. Wiley Interscience, New York, New York, USA.

Pillar, V. D. 1996. A randomization-based solution for vegetation classification and homogeneity testing. Coenoses **11**:29–36.

———. 1998. Sampling sufficiency in ecological surveys. Abstracta Botanica **22**:37–48.

Pillar, V. D., and L. Orlóci. 1996. On randomization testing in vegetation science: multifactor comparisons of relevé groups. Journal of Vegetation Science **7**:585–592.

Podani, J. 1989. Comparison of ordinations and classifications of vegetation data. Vegetatio **83**:111–128.

———. 1994. Multivariate data analysis in ecology and systematics. SPB Academic Publishing, The Hague, The Netherlands.

Strauss, R. E. 1982. Statistical significance of species clusters in association analysis. Ecology **63**:634–639.

Tcacenco, F. A., and V. D. Pillar. 1996. Vegetation–environment relations in anthropogenic grasslands of northeastern Santa Catarina, Brazil. Coenoses **11**:103–108.

Wilson, J. B. 1987. Methods for detecting non-randomness in species co-occurrences: a contribution. Oecologia **73**: 579–582.

———. 1988. Community structure in the flora of islands in Lake Manapouri, New Zealand. Journal of Ecology **76**: 1030–1042.

Wright, S. J., and C. C. Biehl. 1982. Island biogeographic distributions: testing for random, regular, and aggregated patterns of species occurrence. The American Naturalist **119**:345–357.

APPENDIX A

In this Appendix, the bootstrap resampling algorithm for testing the significance of partitions in cluster analysis of sampling units is explained with a numerical illustration.

1) Complete data set (variables are rows, sampling units are columns):

17	14	27	21	16
5	9	8	5	0
5	8	0	0	10

2) Distance matrix (squared Euclidean) of sampling units:

0	34	134	41	51
	0	234	129	89
		0	45	285
			0	15
				0

3) Reference partition with two groups generated by cluster analysis:

Sampling units:	1	2	3	4	5
Groups:	1	1	2	2	1

4) Get a bootstrap sample:

Sampling units:	1	5	4	4	5

5) Distance matrix (squared Euclidean) of sampling units (reference plus bootstrap):

0	34	134	41	51	0	51	41	41	51
	0	234	129	89	34	89	129	129	89
		0	45	285	134	285	45	45	285
			0	150	41	150	0	0	150
				0	51	0	150	150	0
					0	51	41	41	51
						0	150	150	0
							0	0	150
								0	150
									0

6) Bootstrap sample partition with groups 3 and 4 generated by cluster analysis:

Sampling units:	1	5	4	4	5
Groups:	3	4	3	3	4

7) Sum of squares for contrasts between groups of sampling units in the reference (rows) and bootstrap sample (columns); matrix is rearranged:

	3	4		4	3
1	78.2	32.8	→	32.8	78.2
2	28.6	206		206	28.6

8) Total sum of squares computed from distance matrix of step 5:

$$T = (34 + \ldots + 51 + \ldots + 150)/10 = 411.6.$$

One-to-one nearest neighbor sum of squares between partitions:

$$S = 32.8 + 28.6 = 61.3667.$$

Nearest neighbor groups: 1, 4; 2, 3;

$$G^* = 1 - S/T = 0.8509.$$

9) Null bootstrap sample (the units in each group are taken at random from the nearest group in the reference sample):

Sampling units:	3	1	3	4	5
Groups:	3	4	3	3	4

10) Distance matrix of sampling units (reference plus null bootstrap sample):

0	34	134	41	51	134	0	134	41	51
	0	234	129	89	234	34	234	129	89
		0	45	285	0	134	0	45	285
			0	285	51	285	150	0	
				0	134	41	51		
					0	45	285		
						0	150		
							0		

11) Sum of squares for contrasts between nearest neighbor groups of sampling units in the reference and null bootstrap sample:

1, 4:	6.5
2, 3:	1.5

12) Total sum of squares computed from distance matrix of step 10:

$$T^0 = (34 + \ldots + 51 + \ldots + 285 + 150)/10 = 495.8.$$

Exclusive nearest neighbor sum of squares between partitions:

$$S^0 = 6.5 + 1.5 = 8 \qquad G^0 = 1 - S^0/T^0 = 0.9839.$$

Since $G^0 > G^*$, this iteration will add zero to the cumulative frequency $F(G^0 \leq G^*)$.

13) Repeat steps (4)–(12) up to B times. A run with $B = 10\,000$ iterations gave $P(G^0 \leq G^*) = 0.3839$ and a mean value of $G^* = 0.9068$.

APPENDIX B

The application program in C++ for Macintosh or Windows is available in ESA's Electronic Data Archive: *Ecological Archives* E080-014.