COENOSES 11: 29-36, 1996
©C.E.T.A., Gorizia, Italy

A RANDOMIZATION-BASED SOLUTION FOR VEGETATION CLASSIFICATION AND HOMOGENEITY TESTING

Pillar, Valério De Patta

Departamento de Botânica, Universidade Federal do Rio Grande do Sul, Campus Central, Porto Alegre, RS, Brazil 90046-900.
Fax: +55 51 2272295, E-mail: pillar@vortex.ufrgs.br

Keywords: Campos, Cluster, Community, Hypothesis, Probability, Resemblance.

Abstract. The paper offers a probabilistic solution to classification and homogeneity testing in vegetation studies. The method described overcomes the problem of having to make subjective choices about the number of groups or the acceptable level of termination in cluster analysis. It also has advantages over other probabilistic solutions in the way that probabilities are determined. A randomization procedure generates probabilities in pairwise relevé comparisons for input into cluster analysis to control the homogeneity of groups, or in homogeneity testing within given groups of an existing partition. The methods are illustrated on data from Campos grasslands.

Introduction

Cluster analysis is a commonly used technique in plant community classifications. Many clustering algorithms have been offered for in conjunction with a plethora of resemblance indices (Anderberg 1973, Orlóci 1978, Legendre and Legendre 1983, Digby and Kempton 1987, Pillar and Orlóci 1993a, Podani 1994). Despite the objectiveness of the algorithms, one still has to make subjective choices, such as the number of groups or an acceptable level of within-group dissimilarity. Dale (1988) discusses comprehensively the approaches adopted to deal with this problem. As one alternative, I believe that probabilistic indices of resemblance are useful in the evaluation of group homogeneity. Orlóci (1978:221) discusses earlier techniques along this line, pointing out the requirement of large samples and uncorrelated variables when computing probabilities, such as in the case of the Goodall (1966) and Feoli and Lagonegro (1983) indices.

The technical problem in probabilistic indices is to find a suitable method for defining probabilities. Past work suggests three options: axiomatic distributions, observed distributions in the actual sample, and distributions generated in randomization and related methods. The technique in Goodall (1966) is an example of the first two and in Strauss (1982), Jaksic and Medel (1990) and Pillar and Orlóci (1993a:69) of the third. Techniques of randomization are reviewed in general terms in Edgington (1987), Manly (1991) and Crowley (1992). Precedents for the use of randomization and other Monte Carlo methods are many in related fields. Examples include Rohlf (1965) and Orlóci and Beshir (1976) in population biology, and Orlóci and Kenkel (1985:80), Orlóci *et al.* (1986), Legendre and Fortin (1989), Podani (1991), Cornelius and Reynolds (1991), ter Braak and Wiertz (1994) and Pillar and Orlóci (1995 mscr.) in community ecology.

In this paper I build on the idea that the probability associated with a resemblance quantity can itself be interpreted as a measure of resemblance. Goodall (1966) offered resemblance measures of this kind, but did not use randomization. The randomization alternative relies on a larger reference base, which has a good chance of improving accuracy. I apply a randomization method with the objective to construct community groups with maximum homogeneity in one case, and to evaluate the homogeneity within *a priori* defined groups in another case. The probabilities generated in randomization are direct input for cluster analysis.

The randomization algorithm

The probabilities generated for pairwise relevé resemblances express how common the observed values would be if the null hypothesis of *random composition* was in fact true. Random composition implies a random assortment of taxa among the communities described by relevés. This would be expected if the populations were both environmentally neutral and non-interacting. Different algorithms can be used to generate random data sets according to this null hypothesis. The problem is related to testing randomnesss of species co-occurrences in islands, which is addressed in Connor and Simberloff (1979), Wright and Biehl (1982), Gilpin and Diamond (1987), Wilson (1987, 1988) and reviewed in Manly (1991:233), who points out the difficulties and ambiguities in defining a random allocation algorithm for this purpose. It is also related to algorithms discussed in Strauss (1982) and Jacsic and Medel (1990) for probabilistic species clustering. Indeed, various algorithms are possible, depending on whether all taxa are equally like to occur in a community or the probability of each taxon is proportional to its frequency in the data, whether the number of taxa in a given community is constant or not, and how quantitative data should be treated.

I adopt here the algorithm described in Pillar and Orlóci (1993a:69). The input data set comprises n relevés, each containing the records of s_i populations. Each population belongs to a taxon and the same taxon may appear in different relevés. The complete data set forms a pool list of $\pi = s_1 + s_2 + ... + s_n$ plant populations and corresponding taxon-identities. The taxon-identities are likely to be repeated in the pool. This list is used in the design of the reference set under the null hypothesis of random composition. The reference set (Hope 1968, Edgington 1987:305) contains all p! data sets that result by permuting the p taxon-identities among the populations. Actually, because of repeating taxon-identities in the list, the reference set will contain $\prod_{h=1}^{s} f_h!$ identical copies, where s is the number of non-repeated taxon-identities listed in the data set and f_h is the frequency of occurrence of taxon h in the n relevés. However, these repetitions will not affect the derived probabilities. As the reference set may be too large for direct computation, a sample with a limited number of random permutations is taken (Hope 1968, Edgington 1987:43). For this, the list in the ith relevé is formed by s_i randomly allocated taxon-identities from the pool, without replacement.

In quantitative data a performance value, such as density, cover, or frequency observed in the community is attached to each population. The observed performance value is taken as being inherent in the plant population and not subject to randomization. The assumption that the community has an internal structure in which some populations are dominant is implicit in this. Performance values of populations that end up assigned to the same taxon-identities in the same random relevé are pooled. Likewise, repeated taxon-identities are pooled in qualitative data, but the value attached to the pooled population remains one. Because of pooling populations, the number of taxa in a random relevé, its richness, may be smaller than the observed but will not be larger. It is assumed that the total performance value in a relevé and secondarily its richness define the carrying capacity of the site, its environmental limit. Figure 1 illustrates the generation of random data sets.

It has to be assumed that the measuring scale of performance is linear to allow simple additivity of the measurements in the randomization procedure, for which Fig. 1 is example. In other words, the sum of the measured performances of two populations taken individually should equal what would have been measured if they were not considered distinct. I should note that transformations may not establish additivity.

Having an element of the reference set, either randomly generated or the observed, the analysis continues with the computation of a resemblance matrix between relevés. Any resemblance function that can reveal relevant structures in the data is admissible. In a resemblance matrix of n relevés, there are n(n-1)/2 distinct pairwise comparisons that are potentially informative. For each pairwise comparison of relevés i and j the test involves an observed resemblance value d_{ij} and an associated distribution under the null hypothesis. This distribution defines the probability $P(d_{ij}^{o} \geq d_{ij})$ that any randomly chosen value d_{ij}^{o} from among those that materialize in the reference set will be at least as large as dij.

The randomization process is performed by an iterative algorithm. The n(n-1)/2 pairwise resemblance values of the matrix computed at each iteration are compared with the corresponding values in the resemblance matrix computed with the observed data set. After a large number of iterations, each probability $P(d_{ij}^{o} \geq d_{ij})$ is obtained as the proportion of iterations in which randomly generated d_{ij}^{o} was at least as large as the observed dij. The minimum $P(d_{ij}^{o} \geq d_{ij})$ is 1/total number of iterations) . The result is a matrix of probabilities. Since the observed data set is counted as one iteration, in at least one iteration d_{ij}^{o} is itself the observed value d_{ij}. When d_{ij} measures dissimilarity, $1 - P(d_{ij}^{o} \geq d_{ij}) = P(d_{ij}^{o} < d_{ij})$ is used, so that large $P(d_{ij}^{o} < d_{ij})$ values will correspond to large dissimilarities. The example in Figure 1 explains the *modus operandi*. The number of iterations must be large, so that the probabilities will be close to the exact ones that would be obtained in complete evaluation of the reference set (Hope 1968). As suggested by Crowley (1992), the number of iterations must be at least 1000 for a 5% significance level.

Assessing relevé group homogeneity

The probabilities generated by the randomization method are used to test the commonness of the observed resemblances. The one complement $1 - P(d_{ij}^{o} \geq d_{ij})$ is a measure of the extent to which the observed dissimilarity dij departs from its expected value when the relevés are constructed by random assignment of taxon-identities under a random composition null hypothesis. On the basis of these probabilities one can decide whether the dissimilarity of two relevés can be considered extreme (very small or very large). An extreme value is indicated if $P(d_{ij}^{o} \geq d_{ij})$ is small or $1 - P(d_{ij}^{o} \geq d_{ij})$ is small. Any threshold value α can serve as the limit to what is to be regarded as a sufficiently small probability. Usually α is 0.05 or 0.01. A small $P(d_{ij}^{o} \geq d_{ij})$ indicates greater dissimilarity than would be expected by chance and a small $1 - P(d_{ij}^{o} \geq d_{ij})$ indicates that the dissimilarity is smaller than could be expected by chance. The latter is the more likely case in actual samples. If it is true, the relevé pair is considered indistinguishable. In other words, the idea of indistiguishability is broadened beyond the case of the least dissimilar pair, allowing one to declare that they belong to the same community type. I apply this idea to set limits in the evaluation of homogeneity within relevés groups.

The matrix of observed resemblances will have a corresponding matrix of probability values that can be input into cluster analysis as can any other resemblance matrix. The ob-

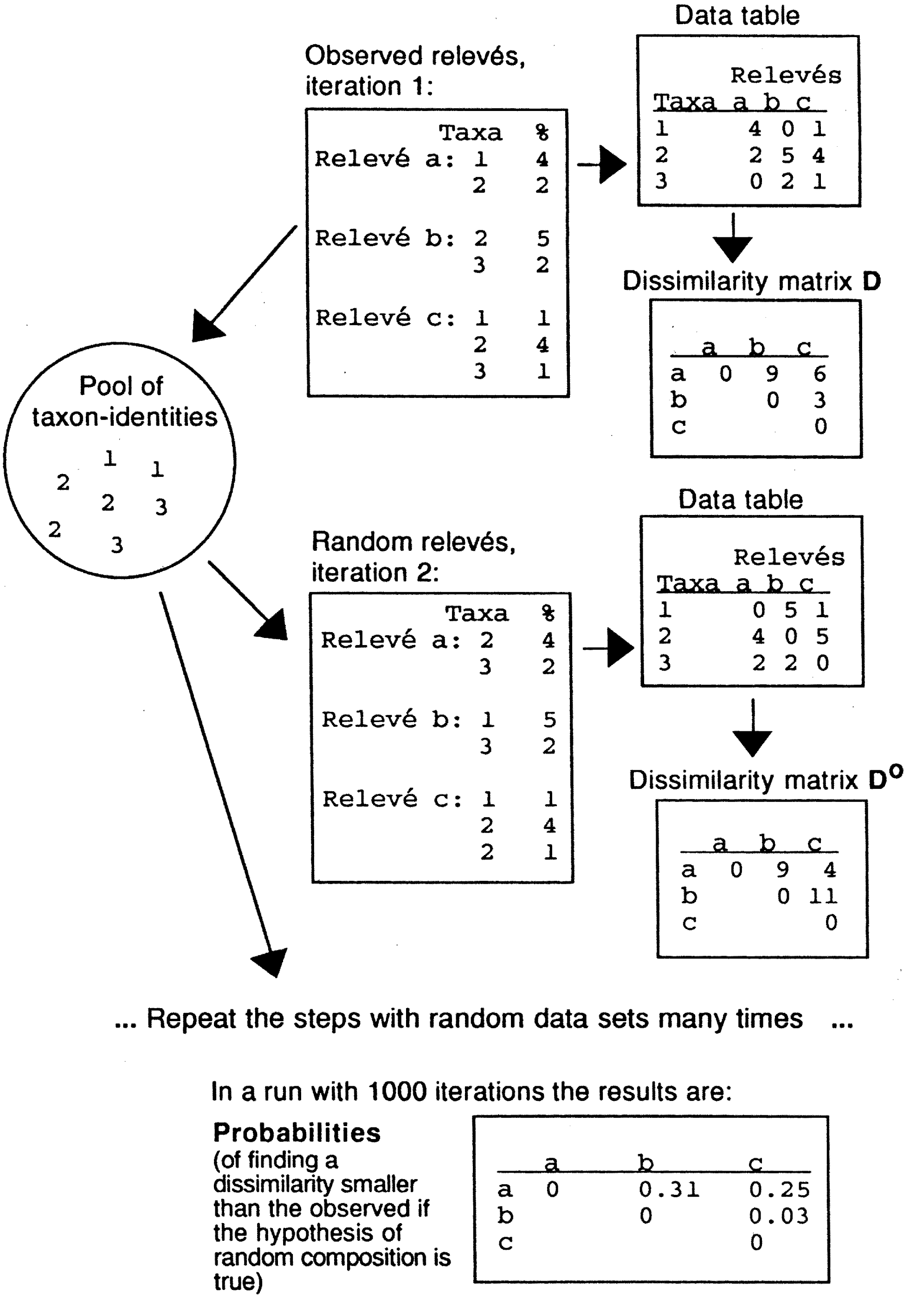

Figure 1. Randomization algorithm applied to artificial data under the random composition hypothesis. The main text describes concepts and technique. Three vegetation relevés (a, b, c) are involved. The number of taxa is three. Performance values are cover percentage estimates divided by 10. Each taxon-identity in the pool is repeated as many times as the number of relevés in which it occurs in the data. The cover values remain unchanged in the expanded matrix, only taxon-identities are permuted. Population quantities that are identified by the same taxon in a random relevé are pooled. Resemblance matrices in this case were computed based on the absolute value function. For example, in the observed data set $d_{ab} = |4-0| + |2-5| + |0-2| = 9$. Note that the observed data is iteration 1. In a run of 1000 iterations the $1 - P(d^{o}_{ij} \geq d_{ij})$ probability of obtaining a $d^{o}_{ab} < 9$ was 0.31, included as the second element in the probability matrix. When interpreting the diagram, the reader should follow pathways indicated by arrows starting with the observed relevés. Note that d_{ij} is an element of D and d^{o}_{ij} is an element of D^{o}.

Table 1. Resemblance values for 15 Campos relevés (data from Pillar 1992). Matrix A contains in the upper half squared Euclidean distances and in the lower half the one complements ($1-P(d^0_{ij} \geq d_{ij})$) of the corresponding probabilities generated in randomization under random composition hypothesis. The number of iterations is 1000. Rows and columns correspond to relevés a to o. The $1-P(d^0_{ij} \geq d_{ij})$ value indicates how rare is a resemblance value. For instance in A, for relevés a and b the probability of obtaining a distance smaller than 120 is 0.041, which is sufficiently low to regard the relevés as not being significantly different. The opposite is indicated for relevés a and d where the probability corresponding to 197 is 0.424. Compare these values to the frequency distributions in Fig. 2. Matrix B contains Goodall type probabilistic dissimilarity indices (Goodall 1966) computed by program PINDEX (Orlóci 1990).

A

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
a	0	**120**	92	**197**	91	85	100	76	128	188	193	85	87	60	**274**
b	**.041**	0	204	213	211	201	206	178	198	118	187	195	113	190	268
c	.005	.189	0	251	59	101	66	56	180	158	261	61	107	142	298
d	**.424**	.402	.498	0	186	242	221	207	119	171	160	234	146	211	191
e	.015	.345	.0	.311	0	70	69	63	171	215	232	88	136	99	263
f	.006	.225	.008	.506	.003	0	65	57	143	223	256	98	128	51	291
g	.009	.225	.0	.345	.001	.0	0	56	158	206	193	87	129	94	250
h	.0	.18	.0	.385	.003	.0	.0	0	156	170	233	91	103	70	250
i	.061	.258	.14	.05	.208	.06	.072	.118	0	136	243	143	91	120	274
j	.108	.015	.046	.088	.209	.193	.141	.074	.031	0	287	181	63	226	324
k	.287	.188	.503	.137	.514	.517	.172	.48	.569	.558	0	242	210	257	63
l	.005	.211	.0	.507	.01	.01	.002	.01	.076	.089	.52	0	80	107	261
m	.015	.047	.013	.22	.124	.051	.045	.033	.016	.0	.607	.004	0	103	247
n	.0	.161	.042	.403	.022	.0	.003	.001	.035	.218	.572	.019	.02	0	288
o	**.575**	.46	.53	.2	.561	.546	.318	.429	.563	.575	.001	.432	.597	.567	0

B

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
a	0														
b	.9932	0													
c	1	.9999	0												
d	1	.9998	1	0											
e	.9997	1	.9973	.9995	0										
f	.9994	.9995	.9998	1	.9920	0									
g	1	.9868	.8747	1	.9996	.9933	0								
h	.9989	.9977	.9353	1	.9959	.9999	.8497	0							
i	.9907	.9999	1	.9801	1	.9997	1	1	0						
j	1	.9593	.9592	.9983	1	.9999	.8647	.9932	.9978	0					
k	.9998	.9982	1	.7855	1	1	1	1	.9999	1	0				
l	.9982	1	.9949	1	.9996	.9894	.9999	.9947	.9978	.9999	1	0			
m	.9977	.9997	.9996	1	1	1	.9987	1	.9942	.9723	1	.9895	0		
n	.9214	1	1	.9995	.9818	.9998	.9995	.9533	.9978	1	1	1	.9999	0	
o	1	1	1	.9611	.9999	1	1	1	.9999	1	1	1	1	1	0

jective is to generate probabilistically justified homogeneous groups. For this the whole matrix of probabilities, treated as a dissimilarity matrix, is the input in complete linkage cluster analysis (Sneath and Sokal 1973:222). This clustering algorithm minimizes the dissimilarity between the most dissimilar relevés within the groups formed in the agglomerative process. Complete linkage clustering is advantageous in this case since the level of within group dissimilarity at each node corresponds to the largest pairwise $1 - P(d^0_{ij} \geq d_{ij})$ of the relevés in the group joined at the node and, therefore, will indicate directly the probabilistic level of group homogeneity. If this level is not larger than the specified threshold value α, the group is considered a homogeneous vegetation type.

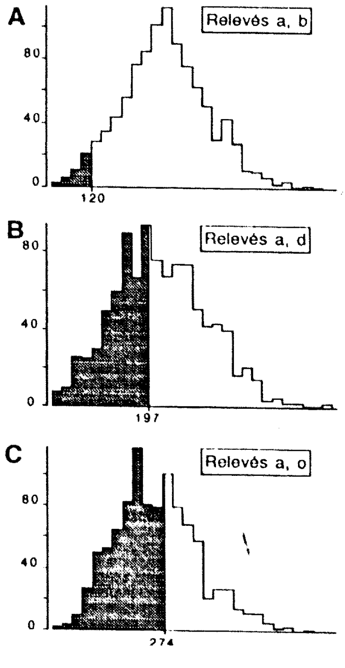

Figure 2. Frequency distributions of 1000 randomly generated distance values for selected Campos relevé pairs. Shaded areas under histograms correspond to the proportion of random distances that are smaller than the observed. The numbers 120, 197, 274 indicate the observed squared Euclidean distance taken from the upper half of matrix A in Table 1.

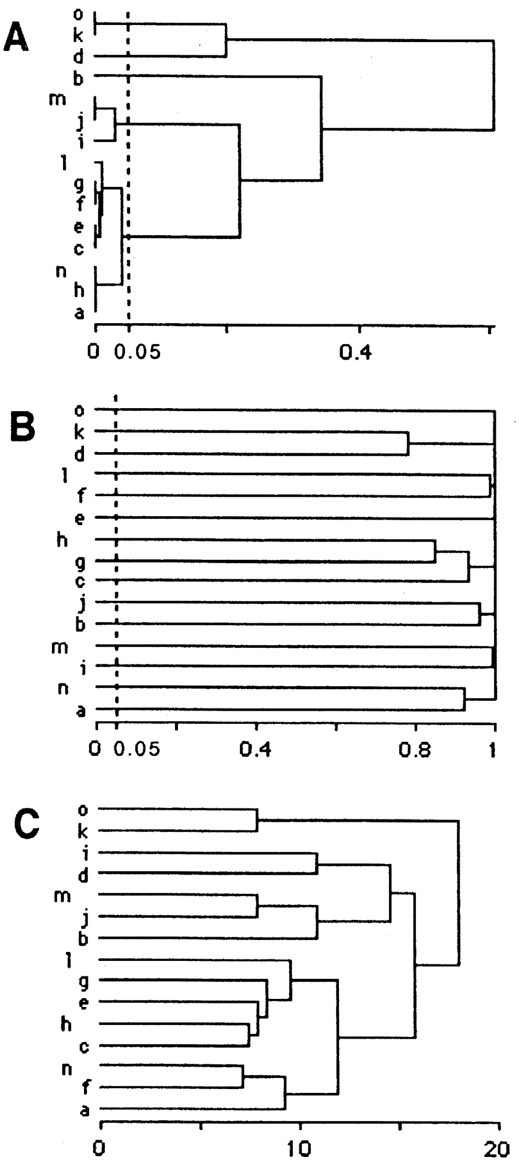

Figure 3. Complete linkage clustering of Campos relevés (labels a to o) based on (A) $1\text{-}P(d^{o}_{ij} \geq d_{ij})$ probabilities generated in randomization (lower half of matrix A in Table 1); (B) Goodall's (1966) index (Table 1B) and (C) Euclidean distance (upper half of matrix A in Table 1). The methods are described in the main text. Horizontal scales indicate the maximum distance within the group merged at the level in tree C, and the probability of the group merged being heterogeneous in trees A and B. At the probability threshold 0.05, 5 homogeneous groups are recognized in tree A [(a, e, f, g, h, c, l, n), (b), (d), (i, j, m) and (k, o)]. This is not much different from tree C. In tree B, the group structure is completely different.

In another case, the objective is to evaluate homogeneity probabilistically within relevé groups in the context of an existing partition. For this, for each group a submatrix of probabilities is extracted from the original probability matrix, containing the comparisons of relevés belonging to the group within which homogeneity is to be assessed. For the group being homogeneous, the largest pairwise probabilistic dissimilarity $1 - P(d^{o}_{ij} \geq d_{ij})$ of the relevés in the group must not be larger than the specified threshold value α.

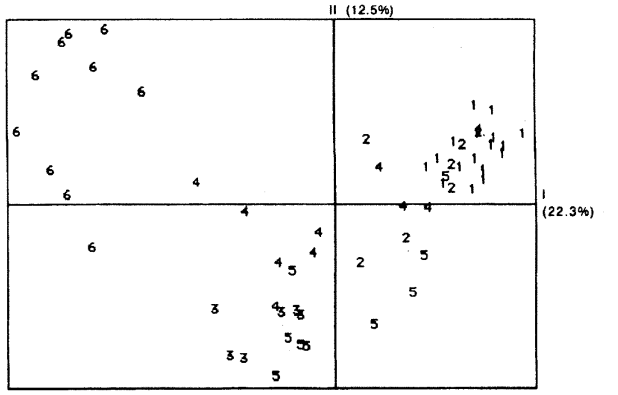

Figure 4. Ordination of 60 Campos relevés. The labels 1 to 6 identify vegetation types in Pillar et al. (1992): 1. *Facelis-Paspalum*, 2. *Aristida-Borreria-Paspalum*, 3. *Baccharis-Andropogon*, 4. *Desmodium-Axonopus-Paspalum*, 5. *Eryngium*, 6. *Eleocharis-Centella*.. . The ordination is Principal Coordinates on a matrix of Euclidean chord distances. The data set is in Pillar (1988).

Finding homogeneous groups in real data

The data set contains 15 relevés from Campos grasslands described by 27 character-set types (CSTs in Orlóci and Orlóci 1985, Orlóci 1991, Pillar 1992, Pillar and Orlóci 1993a,b). The CSTs are defined by the states of 5 characters: growth-form 1 (stoloniferous, rhizomatous, else), growth-form 2 (solitary, rosette, caespitose), leaf cross section (straight, folded, rolled, else), leaf width (ordered classes) and stem (culm) tissue type (herbaceous, woody, no stem). The complete data set is in Pillar and Orlóci (1993a). A character-based description is a peculiarity of the example, not strictly required for the test. Table 1 contains the resemblance values and corresponding probabilities obtained by randomization and using one of Goodall's techniques (Goodall 1966). Figure 2 gives the frequency distributions of pairwise resemblances generated through randomization.

The dendrogram in Fig. 3A indicates five homogeneous groups of relevés at probability threshold 0.05. The dendrogram in Fig. 3B is based on the Goodall index. Fig. 3C displays a non-probabilistic cluster analysis based on Euclidean distances. It is evident that the group structure derived with the Goodall index has little connection to the group structures revealed by the other methods. Goodall et al. (1987) point out that the Goodall index is problematic for quantitative data in the presence of many zeros. As an alternative they recommend another algorithm (AFFINC) which involves data manipulations beyond what I would consider

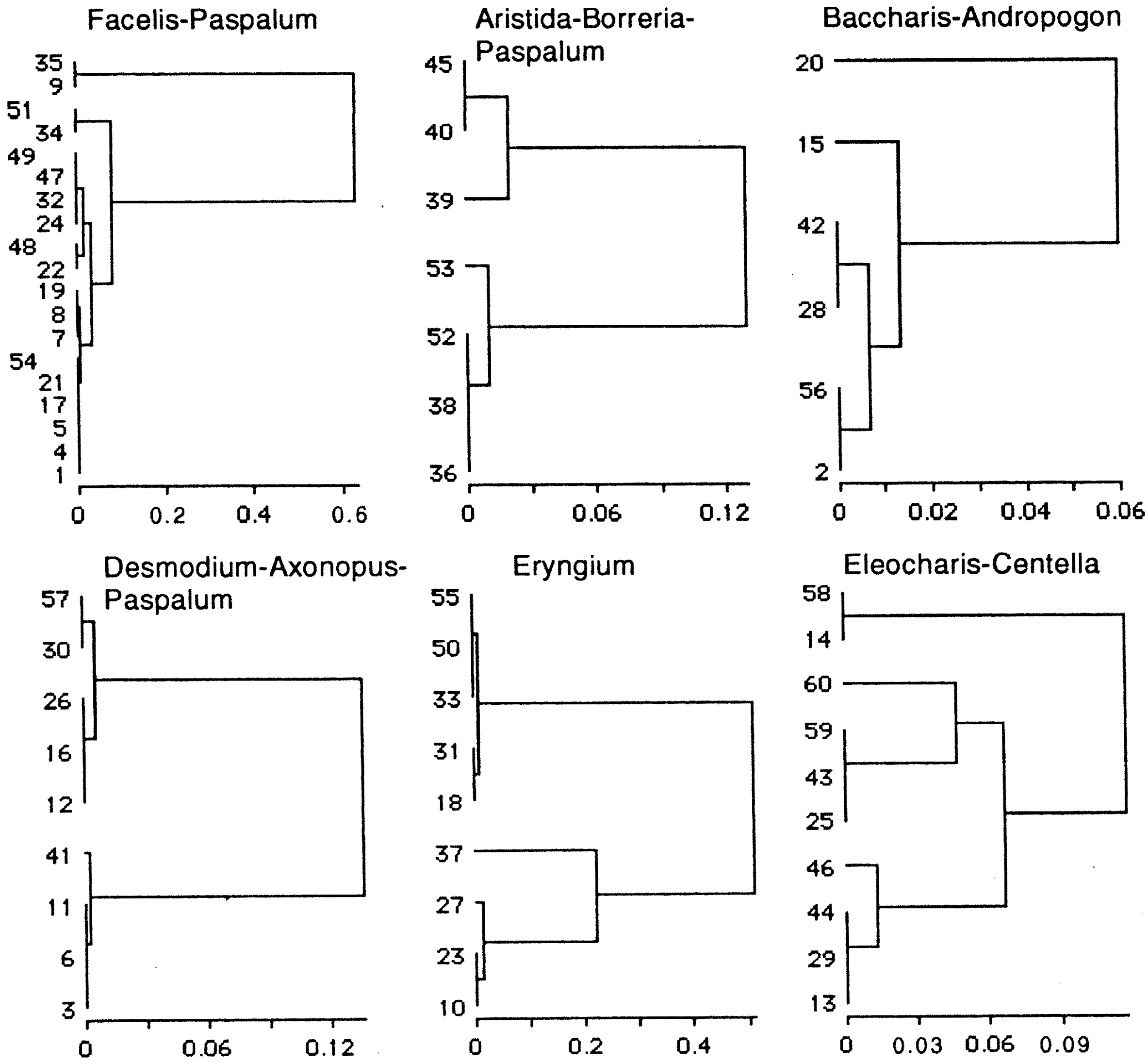

Figure 5. Homogeneity tests in 6 community types identified in the caption of Fig. 4. The clustering algorithm is complete linkage applied to probabilistic relevé resemblances within the type. The probabilities were generated in randomization in the set of 60 relevés. The horizontal scales indicate $1 - P(d^{o}_{ij} \geq d_{ij})$ probabilities. Numbers on the vertical scales are relevé labels. The most homogeneous type is *Baccharis-Andropogon* with $1 - P(d^{o}_{ij} \geq d_{ij}) = 0.06$ and the least *Facelis-Paspalum* ($1 - P(d^{o}_{ij} \geq d_{ij}) = 0.63$).

appropriate under the circumstances (see also Goodall and Feoli 1988).

Assessing community type homogeneity in real data

This example involves 6 vegetation types described in Fig. 4. The data set contains 60 grassland relevés described by 60 species near Porto Alegre in Campos vegetation (Rambo 1956, Cabrera 1971). The numbers 60/60 are coincidental. Randomization is applied to the entire data set to generate probabilities for 60(60-1)/2 pairwise relevé chord distances (Orlóci 1967). Six matrices of probabilities for the test are extracted from the master probability matrix, relevant to pairwise comparisons of relevés within community types. Complete linkage clustering is performed on the probabilities within the types. The result (Fig. 5) shows that only the *Baccharis-Andropogon* type is homogeneous at a reasonably low threshold level (0.06). The other types would have to be further divided to attain this level of group homogeneity.

Discussion

It is seen from the foregoing text that relatively simple (albeit computationally tedious) manipulations can provide a randomization solution to classification and homogeneity testing. The advantage of this is that a well-justified criterion determines the termination of cluster analysis or the interception of a dendrogram to define community types. This is not so in the other methods where probabilities come from theoretical distributions or from the observed distribution within the sample. The problem with these is that the assumptions of the theoretical distribution are not met or the sample is too small. The randomization and Monte Carlo methods in Strauss (1982) and Jaksic and Medel (1990) could be adapted to vegetation data, but they generate a common distribution instead of a distribution for each pairwise comparison, which is problematic in defining the significance of cluster nodes. Furthermore, Jaksic and Medel's algorithm, instead of random reordering, applies random sampling with replacement in the re-assignment of descriptors to units and, therefore, is appropriate only to data sets in which sampling units are selected by random sampling, since this must be assumed in such a Monte Carlo method (see Crowley 1992). This requirement may be very restricting in vegetation studies.

The randomization techniques designed for group comparisons (see, e.g., Edgington 1987, Manly 1991, Pillar & Orlóci 1995 mscr.) are not appropriate for testing group homogeneity as defined here. There would be a circular approach because the tests would use the same compositional data that were used in cluster analysis to maximize the differences among groups; it is likely that significance will be declared. These methods, defined under a random partition null hypothesis, are more appropriate to test whether relevé groups defined by one set of variables (factors) are significantly different when described by another set of variables.

Nevertheless, some comments are needed on the randomization algorithm used here. It may be argued that pooling populations belonging to the same taxon within random relevés determines that more common taxa will tend to be pooled more frequently than less common ones, and thus the ratio between the frequency in the list and the probability of being present in a relevé will be larger for the less common taxa. In spite of this, taxa observed in more relevés will be proportionately more common in the reference set. By extension, an observed coincidence of rare taxa will also be rare in the reference set (Goodall 1966). This is important, since otherwise any taxon would have the same probability irrespective of its frequency in the observed data set. This is not desired, because (for example) two relevés sharing a rare taxon in the data set would not appear any different from the point of the test than if they were sharing a common taxon. The alternative in Strauss (1982) and Wilson (1987) is not appropriate because their constraint, that the number of taxa on each site and the number of sites each taxon occurs in a random data set are the same as those in the observed data set, would not fit under the null hypothesis defined here. A typical example is when a species occurs in all relevés; following such a constraint there would not be any random allocation of this species because it would have to be allocated *a priori* to all relevés.

I should note also that the probabilities and the perception of homogeneity are local and may only be valid within the same data set. A different probability may arise for the same relevé pair, or homogeneity may be differently perceived within the same group, if new relevés are added to or relevés are removed from the set (Orlóci 1978:64). A possible solution to this problem is to use process sampling, in which the sample stability measured by a specific stress function indicates optimal sample size (Orlóci and Pillar 1989) and implies "stable" probabilities. Similarly, the use of a different resemblance measure may produce different probabilities.

Acknowledgments. I am grateful to L. Orlóci for discussions and suggestions. Thanks are extended to M. E. Biondini for helpful suggestions on an early version of the manuscript. Financial support for this project from CNPq (Brazil) is gratefully acknowledged.

For specifications and availability of the SYNCSA package, inquiries should be directed to V. Pillar at the given address.

References

Anderberg, M. R. 1973. Cluster Analysis for Applications. Academic Press, New York.

Cabrera, A. L. 1971. Fitogeografia de la Republica Argentina. Sociedad Argentina de Botanica, Buenos Aires.

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

Cornelius, J. M. and J. F. Reynolds. 1991. On determining the statistical significance of discontinuities within ordered ecological data. Ecology 72:2057-207.

Crowley, P. H. 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. Annual Review in Ecology and Systematics 23:405-47.

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

Digby, P. G. N. and R. A. Kempton. 1987. Multivariate Analysis of Ecological Communities. Chapman & Hall, London.

Edgington, E. S. 1987. Randomization Tests. Marcel Dekker, New York.

Feoli, E. and M. Lagonegro. 1983. A resemblance function based on probability: applications to field and simulated data. Vegetatio 53:3-9.

Gilpin, M. E. and J. M. Diamond. 1987. Comments on Wilson's Null Model. Oecologia (Berlin) 74: 159-160.

Goodall, D. W. 1966. A new similarity index based on probability. Biometrics 22: 882-907.

Goodall, D. W., P. Ganis and E. Feoli. 1987. Probabilistic Methods in Classification: A Manual for Seven Computer Programs. Università degli Studi di Trieste, Trieste.

Goodall, D. W. and E. Feoli. 1988. Application of probabilistic methods in the analysis of phytosociological data. Coenoses 3:1-10.

Hope, A. C. A. 1968. A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society 30: 582-598.

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

Legendre, P. and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107-138.

Legendre, L. and P. Legendre. 1983. Numerical Ecology. Elsevier, Amsterdam.

Maarel, E. van der 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. Vegetatio 39: 97-114.

Manly, B. F. J. 1991. Randomization and Monte Carlo Methods in Biology. Chapman and Hall, London.

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.

Orlóci, L. 1990. Ecological Programs for Instructional Computing on the Macintosh. SPB Academic Publishing, The Hague.

Orlóci, L. 1991. On character-based plant community analysis: choice, arrangement, comparison. Coenoses 6:103-107.

Orlóci, L. and E. Beshir. 1976. A heuristic test for homogeneity in species populations. Vegetatio 31: 141-145.

Orlóci, L., E. Feoli, D. Lausi and P. Nimis. 1986. Estimation of character structure convergence (divergence) in plant communities; a nested hierarchical model. Coenoses 1: 11-2.

Orlóci, L. and N. C. Kenkel. 1985. Introduction to Data Analysis; with examples from population and community ecology. International Co-operative Publishing House, Fairland, Maryland.

Orlóci, L. and M. Orlóci. 1985. Comparison of communities without the use of species: model and example. Annali di Botanica 43: 275-285.

Orlóci, L. and V. De Patta Pillar. 1989. On sample size optimality in ecosystem survey. Biometrie-Praximetrie 29: 173-184.

Pillar, V. De Patta. 1988. Fatores de Ambiente Relacionados à Variaç_o da Vegetaç_o de um Campo Natural. M.Sc. Dissertation, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Pillar, V. De Patta. 1992. The Theory of Character-Based Community Analysis. Ph.D. Thesis, University of Western Ontario, London, Canada.

Pillar, V. De Patta, , A. V. A. Jacques and I. I. Boldrini. 1992. Fatores de ambiente relacionados à variaç_o da vegetaç_o de um campo natural. Pesquisa Agropecuária Brasileira 27:1089-1101.

Pillar, V. De Patta and L. Orlóci. 1993a. Character-Based Community Analysis; The Theory and an Application Program. SPB Academic Publishing, The Hague.

Pillar, V. De Patta and L. Orlóci. 1993b. Taxonomy and perception in vegetation analysis. Coenoses 8: 53-66.

Podani, J. 1991. On the standardization of Procrustes statistics for the comparison of ordinations. Abstracta Botanica 15: 43-46.

Podani, J. 1994. Multivariate Data Analysis in Ecology and Systematics. SPB Academic Publishing, The Hague.

Rambo, B. 1956. A Fisionomia do Rio Grande do Sul. 2nd. ed. Selbach, Porto Alegre.

Rohlf, F. J. 1965. A randomization test of the nonspecificity hypothesis in numerical taxonomy. Taxon 14: 262-267.

Sneath, P. H. A. and R. R. Sokal. 1973. Numerical Taxonomy; the principles and practice of numerical classification. Freeman, San Francisco.

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

ter Braak, C. J. F. and J. Wiertz. 1994. On the statistical analysis of vegetation change: a wetland affected by water extraction and soil acidification. Journal of Vegetation Science 5: 361-372.

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

Wilson, J. B. 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.

Received November 11, 1995