Legendre, P. 1987. Constrained clustering. Pp. 289-307 *in*: P. Legendre & L. Legendre [eds.] *Developments in numerical ecology*. NATO ASI series, Vol. G-14. Springer-Verlag, Berlin. xi + 585 pages.

CONSTRAINED CLUSTERING

Pierre Legendre
Département de Sciences biologiques
Université de Montréal
C.P. 6128, Succursale A
Montréal, Québec H3C 3J7, Canada

Abstract - Results of cluster analysis usually depend to a large extent on the choice of a clustering method. Clustering with constraint (time or space) is a way of restricting the set of possible solutions to those that make sense in terms of these constraints. Time and space contiguity are so important in ecological theory that their imposition as an *a priori* model during clustering is reasonable. This paper reviews various methods that have been proposed for clustering with constraint, first in one dimension (space or time), then in two or more dimensions (space). It is shown, using autocorrelated simulated data series, that if patches do exist, constrained clustering always recovers a larger fraction of the information than the unconstrained equivalent. The comparison of autocorrelated to uncorrelated data series also shows that one can tell, from the results of agglomerative constrained clustering, whether the patches delineated by constrained clustering are real. Finally, it is shown how constrained clustering can be extended to domains other than space or time.

INTRODUCTION

Constrained clustering is part of a family of methods whose purpose is to delimit homogeneous regions on a univariate or multivariate surface, by forming blocks of pieces that are also adjacent in space or in time. As an alternative to clustering, this same problem of "regional analysis" can be addressed by ordination methods, as is the case with most other problems of descriptive data analysis. Various methods of "regional analysis" have been reviewed by Wartenberg (manuscript) who divided them into three basic classes: (1) *a posteriori* testing of nongeographic solutions; (2) clustering or ordering with absolute contiguity constraint; and, (3) geographic scaling of phenetic information.

Clustering with constraint is one way of imposing a model onto the data analysis process, whose end result otherwise would depend greatly on the clustering algorithm used. The model consists of a set of relationships that we wish the clustering results to preserve, in addition to the information contained in the resemblance matrix (or, for some clustering methods, in the raw data: Lefkovitch 1987). These relationships may consist of geographic information, placement along a time series, or may be of other types, as we will see. In any case, imposing a constraint or a set of constraints onto a data-analytic method is a way of restricting the set of possible solutions to those that are meaningful in terms of this additional information.

NATO ASI Series, Vol. G 14
Developments in Numerical Ecology
Edited by P. and L. Legendre

In this paper, we will first describe various forms of constrained clustering. Then we will examine the questions of whether constrained clustering is necessary to get meaningful results, and how to determine if the patches found by constrained clustering are real. Finally, we will suggest that the concept of constrained clustering can be extended to models other than space or time.

Ecologists are primarily interested in two types of natural constraints: space and time. Ecological sampling programs are usually designed along these physical axes, so that information about the position of ecological samples in space and in time is almost always known. Furthermore, various parts of ecological theory tell us that elements of an ecosystem that are closer in space or in time are more likely to be under the influence of the same generating process (competition theory, predator-prey interactions, succession theory), while other parts of ecological theory tell us that the discontinuities between such patches in space or in time are important for the structure (succession, species-environment relations) or for the dynamics of ecosystems (ergoclines).

These reasons are so compelling as to legitimize a clustering approach where the clusters will be considered valid only if they are made of contiguous elements. From this point of view, clusters of noncontiguous elements, such as can be obtained from the usual unconstrained clustering algorithms, are seen as an artifact resulting from the artificial aggregation of effects from different but converging generating processes. We will come back to this point later on.

ONE-DIMENSIONAL CONSTRAINT

In many ecological problems, the *a priori* information to be taken into account is one-dimensional. This is the case when the sampling takes place through time or along a transect, or else when studying sediment cores (that may represent either space or time series). The methods for dividing such data series into segments, using a constrained approach, go back to W. D. Fisher (1958), an economist, who suggested an algorithm for univariate data based on minimizing the weighted sum of within-group sums of squared distances to the group centroids. The user must also decide how many groups he/she wishes to obtain. Fisher's method was valid in both the constrained and the unconstrained situation. It was later generalized to multivariate data by Ward (1963), who considered only the unconstrained case, and proposed the well-known minimum-variance hierarchical clustering method.

Several other proposals have been reviewed by Wartenberg (manuscript). Among these, let us mention the method of Webster (1973), a soil scientist who needed to partition multivariate sequences corresponding to a space transect or to a core. Moving a window along the series, Webster compared the two halves of the segment covered by the window, either with Student's t or Mahalanobis' D^2, and he placed boundaries at points of maximum value of the statistic. While the results obtained depend in part on the window length, Webster's method is interesting in that it looks for points of maximal changes between regions.

The dual approach to this problem is to look for maximal homogeneity within segments. This was the point of view adopted by Hawkins and Merriam who proposed a method for segmenting a univariate (1973) or a multivariate (1974) data series into homogeneous units, using a dynamic programming algorithm. This method was advocated by Ibanez (1984) for the study of successional steps in ecosystems.

Although it represents a methodological improvement over previous ways of studying succession, this method is still problematic. First, the user must determine the number of segments she/he wishes to obtain, using as an indicator the increase in explained variation relative to the increase in the number of segments. A second problem with ecological data is that strings of multiple zeroes, which are very often found in species abundance data series, are likely to cause the formation of segments based on species absences. Actually, the method assumes each group to be drawn from a multivariate normal distribution and it is sensitive to departures from this condition, which is rarely met by ecological data. Finally, as the user increases the number of groups, group breaks that appear at one grouping level may change position at the next level (Legendre *et al.* 1985: 274).

Using the hierarchical clustering approach, Gordon and Birks (1972, 1974) and Gordon (1973) included the time constraint in a variety of algorithms to study pollen stratigraphy. They used constrained single linkage, constrained average linkage, and a constrained binary division algorithm. Their purpose was to define zones of pollen and spores that are homogeneous within zones and different between zones. They compared their various techniques, which led by and large to the same result. As we will see below, this was probably due to the predominant influence of the constraint on the results.

Legendre *et al.* (1985) used a very similar approach to study ecological successions through time. The basis of their method, called "chronological clustering", is proportional-link linkage hierarchical clustering with a constraint of time contiguity. This means that only time-adjacent groups are considered contiguous and are assessed for clustering. There is one

important addition to the ideas of Gordon and his co-workers, however: this algorithm is supplemented with a statistical test of cluster fusion whose hypotheses correspond to the ecological model of a succession evolving by steps.

Prior to this analysis, a distance matrix among samples has been computed, using a dissimilarity function appropriate to the problem at hand (ecological succession, or other). Considering two groups (1) that are contiguous and (2) that are proposed for fusion by the clustering algorithm, a one-tailed test is made of the null hypothesis that the "large distances" in the submatrix are distributed at random within and among these two groups. The test is performed by randomization; this test could actually be re-formulated as a special form of the Mantel test (1967). The above-mentioned paper shows the true probability of a type I error to be equal to the nominal significance level of the test. When the null hypothesis is accepted at the given confidence level, the two groups are fused. The computer program also allows for the elimination of aberrant samples that can form singletons and prevent the fusion of their neighboring groups, and it offers complementary tests of the similarity of non-adjacent groups. The end result is a nonhierarchical partition of the samples into a set of internally contiguous groups, the number of which has not been coined by the user.

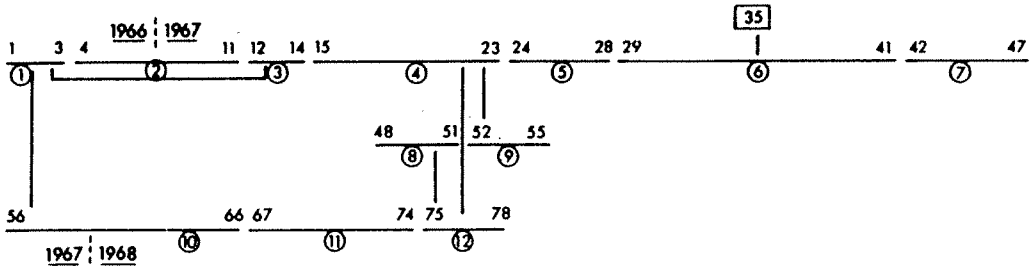

Fig. 1. Schematic representation of the chronological clustering of 78 samples of Mediterranean chaetognaths. Cluster numbers are circled. Between-group pairwise relationships are represented by vertical lines. The boxed sample is a singleton. Connectedness = 25%, α = 0.25. From Legendre *et al.* (1985), Figure 4.

I shall illustrate time-constrained clustering with this method. The example consists of a series of 78 samples of Mediterranean zooplankton (chaetognaths) obtained from 1966 to 1968 and analyzed by Legendre *et al.* (1985). In Figure 1, the series is folded to allow represention of the relationships among clusters; these relationships have been computed by *a posteriori* testing, using the test of cluster fusion described above. The ecological significance of the group breaks is discussed in the above-mentioned paper.

This data set was also subjected to chronological clustering using several values of connectedness during the proportional-link linkage agglomeration. Without the constraint, low values of connectedness have a space-contracting effect while high values cause an effect equivalent to an expansion of the reference space (Lance and Williams 1967). As shown in Figure 2, the results are quite stable through a range of connectedness values. This illustrates the predominant effect of the constraint during the clustering process, as previously noted by Gordon and Birks (*op. cit.*). Clustering the same data set by unconstrained proportional-link linkage produced scrambled, uninterpretable results (Legendre *et al.* 1985).

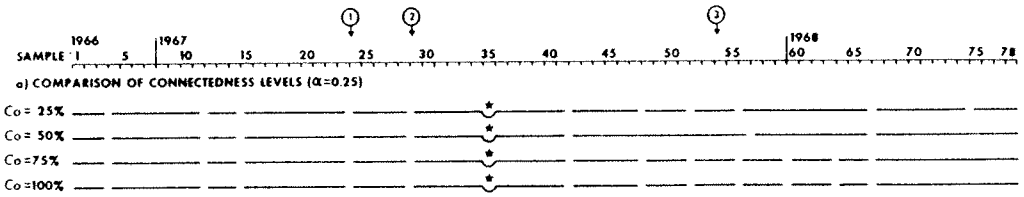

Fig. 2. Comparison of four connectedness levels (Co), keeping α fixed at 0.25. Same data as in Figure 1. Full horizontal lines: clusters of contiguous samples, with blanks representing significant breaks in the series. Stars: singletons. From Legendre *et al.* (1985), Figure 3.

Chronological clustering, which was developed with reference to the problem of species succession in ecosystems, could be applied to other problems where one hypothesizes sharp breaks within the data series. Besides the examples in Legendre *et al.* (1985), the method has been applied to a variety of other problems, that include the successional dynamics of bacteria through time in sewage treatment lagoons (Legendre *et al.*1984), the study of fish communities in a coral reef transect (Galzin and Legendre 1987) and of a stratigraphic sequence of fossil fish (Bell and Legendre 1987).

TWO-DIMENSIONAL AND HIGHER SPATIAL RELATIONSHIPS

Often, the spatially distributed data of interest to the ecologist are not sampled from a transect, but are spread across a surface or, in some instances, a volume. If the spatial relationships among samples are to be taken into account during the clustering process, it is important to define clearly what is meant by "contiguous samples".

If the data represent sub-units of the area under study, with these smaller surfaces touching one another, then a simple and natural way is to define as contiguous two surfaces that share a common border.

On the contrary, if the data can be seen as attached to points in space that are distant from one another, then there are various ways of defining the connection network among these points.

(a) The easiest way is to use the minimum spanning tree among points in geographic space. This method is also the least efficient in that it uses only a small fraction of the geographic information.

(b) Among the various types of connection networks, one that is often used is the Gabriel graph (Gabriel and Sokal 1969). In this graph, two points A and B are connected if no other point is found inside the circle whose diameter is the line joining A and B; in other words, connect A and B when $D^2_{AB} < D^2_{AC} + D^2_{BC}$ for every triplet of points A, B, C under study.

(c) Another commonly used type of connection network is the Delaunay triangulation. This is a way of dividing the whole plane into triangles without crossing edges. The algorithm proposed by Green and Sibson (1978) also allows the user to remove those long edges that form along the perimeter of the surface as "border effects". A Gabriel graph is a subset of a Delaunay triangulation (Matula and Sokal 1980).

(d) When the points form a regular grid (or when the surface is divided into squares or rectangles), it is a simple matter to connect them in 4 directions if they form a square lattice, or in 8 directions by adding diagonal edges. They could also be connected in 6 directions if they are positioned in staggered rows.

These connecting schemes can be extended to three dimensions if the points come from a volume of space, or if the volume is divided into regular or irregular blocks.

Using one or another of these connecting schemes, authors have constrained many of the usual clustering algorithms: linkage clustering, UPGMA, minimum-variance method, hierarchical binary division, and so on. Others used the geographic information *a posteriori*, selecting among the set of possible partitions those that are consistent with the spatial constraints. Wartenberg (manuscript) has reviewed these developments, that go back to Ray and Berry (1966).

Tests of various kinds have been developed, either as a part of constrained clustering algorithms, or to assess the interest of the results.

(a) Howe (1979) used a test of the difference between the means of adjacent groups, during pairwise agglomeration. In the same line of thought, Gabriel and Sokal (1969) developed a significance test of the homogeneity of a whole partition based on the sum of squares criterion. Given what we know now about the influence of spatial autocorrelation on statistical tests, and

especially on analysis of variance (e.g., Cliff and Ord 1981, ch. 7), these tests should be used with caution.

(b) Ray and Berry (1966) evaluated the various agglomeration levels by plotting the changes of the within-group and the among-group variances as a function of the number of groups. Changes in the slope of these curves indicate the best partition.

(c) Okabe (1981) developed an index for the difference between the constrained and the unconstrained solution, that he tested for significance by randomization. His index is based on the number of point displacements that are necessary to transform one solution into the other, but the Jaccard or the Rand index (described below), or information measures such as Rajski's metric (1961), could be used for the same purpose.

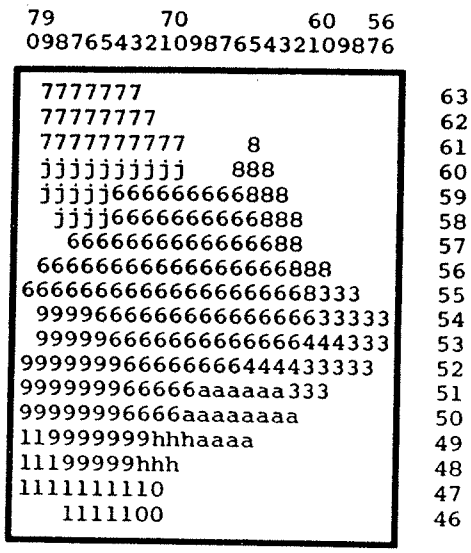

Fig. 3. One of the maps from the constrained clustering study of Legendre and Legendre (1984). This map represents clustering level S = 0.70 of the proportional-link linkage agglomeration, with connectedness of 50%. Each group of quadrats formed at this level is represented by a different letter or number. Longitude (W) and latitude (N) are shown outside the frame.

I will illustrate constrained clustering on a surface using results from our program (BIOGEO), which is a constrained proportional-link linkage agglomerative algorithm that can handle large data sets; this property comes from the fact that, in a constrained situation, the search for the next pair to join is limited to adjacent groups only, as previously noted by Openshaw (1974) and by Lebart (1978). The program can use either (a) points in a regular grid, or (b) a list of connections obtained for instance from a Delaunay triangulation. It presents the advantage of producing directly a series of maps, each corresponding to a clustering level, instead of the usual dendrogram. These maps are drawn either for the regular grid, or using the X and Y coordinates of the points. Figure 3 shows one such map, from a biogeographic study of freshwater fishes in the Québec peninsula (Legendre and Legendre 1984), based upon the presence/absence of 109 species in 289 units of territory. Figure 8 shows a pair of such maps for points positioned by their X and Y coordinates.

When constrained clustering has been completed, distant groups could be tested *a posteriori* to determine if recurrent group structures exist through space. See Cliff and Ord (1981) for tests of the difference among means in the presence of spatial autocorrelation.

IS CONSTRAINED CLUSTERING NECESSARY ?

The question has been raised, whether constrained clustering represents a methodological advance. Could the same results be obtained without the constraint ? A constraint is after all difficult to imbed into computer programs. I would like to argue that if one assumes the existence of an ecological process generating autocorrelation along the sampling axes (space or time), then one is more likely to miss uncovering the corresponding ecological structure if the clustering is carried out without constraint. This property of clustering algorithms will be demonstrated for agglomerative methods; divisive or nonhierarchical methods would likely lead to the same result.

For the sake of clarity, let us limit this discussion to spatially autocorrelated phenomena, although the results apply as well to autocorrelation along the time axis. In community ecology, one can often hypothesize generating processes related either to the abiotic environment, or to some form of contagious biological growth. If, for the scale of sampling under consideration, the generating process has produced a gradient, the existence of such a gradient can be demonstrated by spatial autocorrelation analysis (univariate autocorrelation analysis: Cliff and Ord 1981; multivariate Mantel correlogram: Sokal *et al.* 1987), while the gradient itself can be described adequately by ordination analysis (scaling). On the other hand, if the generating process has produced locally homogeneous community structures within some larger area subjected to sampling, then the description of these structures becomes a clustering problem. Since one is then interested in forming connected clusters of objects, there is no question as to the appropriateness of constrained clustering, since this is exactly what this family of methods does: it produces clusters of spatially connected points. On the contrary, clustering without constraint would open the door to clusters possibly formed by grouping objects whose apparent similarity is the result of different mechanisms that converged to produce somewhat similar effects on the community structure; these clusters would present a blurred picture, as noted by Monestiez (1978). Wartenberg (manuscript) gives a similar example from the health sciences, where lung ailments may be due to a variety of causes: occupational (i.e., from coal mining), ambient (such as near industrial areas), or personal habits (tobacco consumption), all of which can lead to light or severe lung conditions; unconstrained clustering would group the samples by severity of cases while spatially constrained clustering is more likely to delineate areas with similar types of causes.

The same rule applies to community ecology, where it is better to form the regional clusters first, and to find the relationship among clusters in a second step.

To demonstrate that constrained clustering is not only appropriate, but also necessary, we will rely on Monte Carlo simulations. Analyzing known conditions will show that one is less likely to get a meaningful answer after unconstrained clustering than if a constraint has been used, in cases where a generating process has produced patchiness.

Five groups of equal size (30 objects each) have been generated by an autocorrelated process with random components. To make them easier to picture, the groups are made to form for the moment a one-dimensional array of 150 objects. Within each group, one of the objects is selected at random to become the nucleus of the generating process giving rise to the group. A value is given to each of these nuclei; this value is drawn from a normal random distribution with mean 0 and variance VAR. The rest of each group is made to grow out of its nucleus by a contagious process, that consists of giving to a point located at distance n from the nucleus, the value of the point located at distance (n-1), plus a N(0,1) random normal deviate. Such autocorrelated Monte Carlo series have been generated with group nuclei variances VAR = 1, 5, 10, 15, 20, 25 and 30, as well as for the intermediate integer values of VAR between 1 and 10; the amount of variance added at random to the contagious within-group growth process is kept constant. The data sets, 150 objects long, are univariate; this should not affect the generality of the conclusions. Spatial autocorrelation analysis was performed on these series to verify that the data are indeed autocorrelated; significant positive autocorrelation extended to about distance 20 in each of these data series. Five of them are shown in Figure 4; the seed of the random number generator was the same for all runs.

After computing a (150 x 150) Euclidean distance matrix among objects, ag- glomerative clustering is performed using constrained as well as unconstrained clustering. Both of the algorithms used are based upon proportional-link linkage clustering, and a connectedeness value of 50% was used throughout for the sake of uniformity.

Since the "truth" is known from the generating process (five equal groups of 30 objects each), it can be used to assess the efficiency of each clustering model. To achieve this, a (150 x 150) half-matrix is first computed for any given partition level of the hierarchical classification, containing a "1" to describe two objects that are members of the same group at the said level, and "0" otherwise. Another such half-matrix is built for the reference classification of the objects into five groups. Milligan (1983) recommends using both the Jaccard and the Rand index to compare the two partitions:

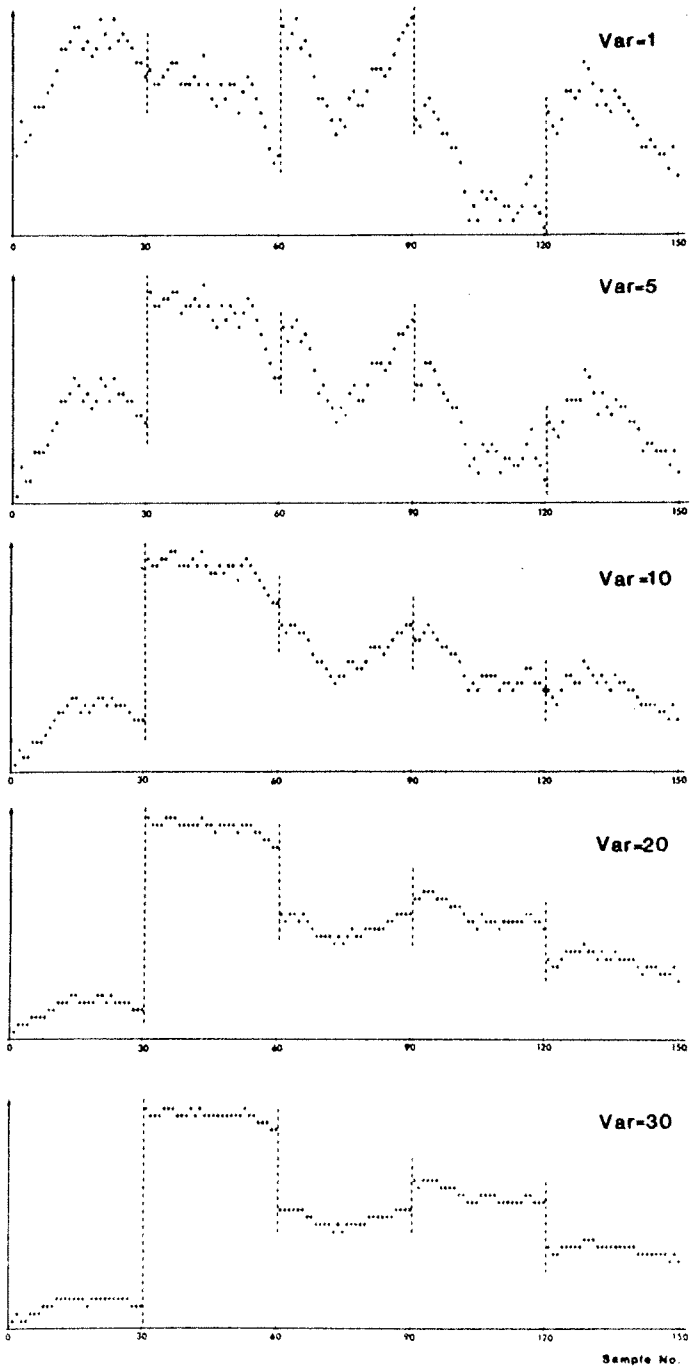

Fig. 4. Five autocorrelated Monte Carlo series, generated with different values (VAR) of group nuclei variance. Ordinate: the value attributed to each sample along the series. The seed of the random number generator was the same for these five runs. Group breaks are materialized by dashes.

$$\textbf{Jaccard} = a / (a + b + c)$$

$$\textbf{Rand} = (a + d) / (a + b + c + d)$$

where a, b, c, and d are the frequencies of the (2 x 2) contingency table comparing the two half-matrices. Since we are dealing with hierarchical clustering, there are many levels of partition; the partition was selected that maximizes the relationship between the computed classification and the "truth", for each of the indices (Jaccard and Rand).

First matrix \ Second matrix	1	0
1	*a*	*b*
0	*c*	*d*

The results obtained with the Jaccard index are clear (Fig. 5). For any amount of variance among group nuclei, constrained clustering recovers more of the original classification's information than does unconstrained clustering.

The results obtained with the Rand index are the same, although the Rand criterion, at low VAR values (VAR $\leq$ 5) and only in the unconstrained case, regularly picked out as optimal partition levels where very few points had been clustered, all the others being treated as singletons (one-member clusters). The Rand index could pick out these partition levels because the quantity to be maximized involves d, the number of pairs pertaining to unlike groups in both classifications.

These simulations lead to the conclusion that one should always use constrained clustering, when working under the assumption that the phenomenon under study is spatially (or temporally) autocorrelated.

VERIFYING THE ASSUMPTION OF PATCHINESS

What if one uses constrained clustering while there is no spatial structure, despite the assumptions to that effect ? Of course, one could have ascertained first that there is a patchy spatial structure, by spatial autocorrelation analysis (Sokal and Thomson 1987). Spatial correlograms, however, can only recognize patchiness when it is somewhat regular; they may fail to give a significant answer if the patches are greatly variable in size. So, constrained clustering may be needed even if spatial autocorrelation analysis has not demonstrated the existence of regular patches. Can we use the results of the clustering itself to tell us whether the patches obtained with constraint are real entities ?

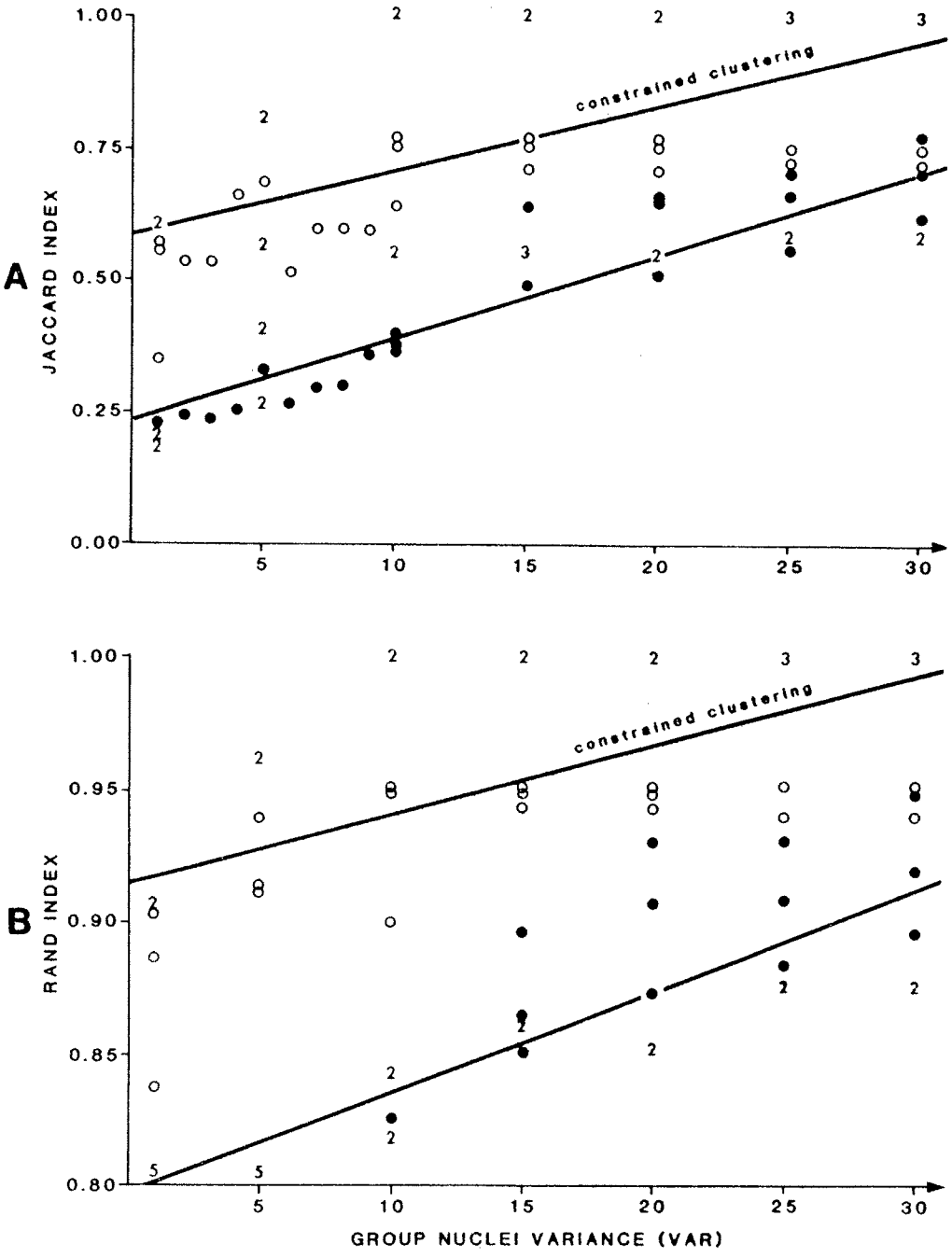

Fig. 5. Fraction of the group structure information recovered using constrained (open circles) and unconstrained (closed circles) clustering, according to (A) the Jaccard index, and (B) the Rand index, for groups generated with various amounts of variance among group nuclei (abscissa).

This can indeed be done. Let us compare what should happen during constrained clustering, in the absence or in the presence of patchiness. Let us consider first an example where the values to be clustered are the result of a strictly random process. In that case, the probability that two neighbors (groups, or single objects) will be the next most similar pair is equal among pairs of neighbors, and its value is (1/number of possible pairs) in ideal cases. It varies with group size in the case of space-contracting (like single linkage) or space-dilating (like complete

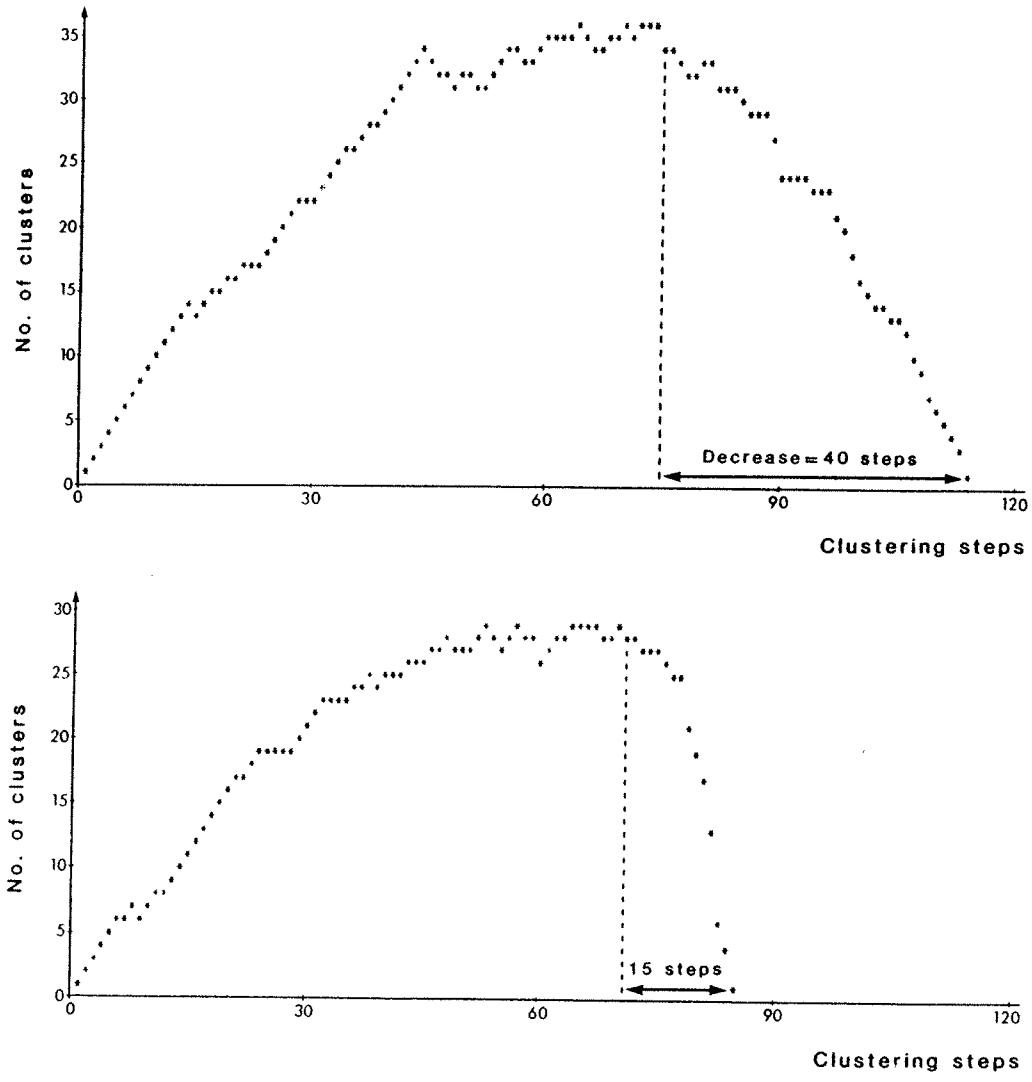

Fig. 6. Spatially autocorrelated data, from Figure 4 (VAR = 10), produce a longer zone of decrease (top) than 150 random points (bottom panel). The ordinate of each graph represents the number of groups, other than single-object clusters, that are present at the corresponding level (abscissa).

linkage) clustering methods (Lance and Williams 1967); this point deserves further investigation. In any case, one expects the random agglomeration mechanism to produce at first a large number of small patches, that grow according to some random model, while near the end of the clustering process, we can expect the quick formation of very large patches (within a few clustering steps), before the final formation of a single group. If there is a spatially autocorrelated structure, the beginning of the agglomeration should follow essentially the same pattern, since the points that cluster correspond at first to random within-group variations; near the end of the agglomerative process, the differences among groups should translate into extra steps in the larger distance classes, contrary to the no-structure case.

Actual experiments show that this is indeed what happens (Fig. 6 and 7). When the data series is one-dimensional (circles in Fig. 7), the difference in length of the zone of decline is very large at all values of connectedness, from 1% to 100%, used in the proportional-link linkage agglomeration. When the series are made to form a two-dimensional grid of 5 lines and 30

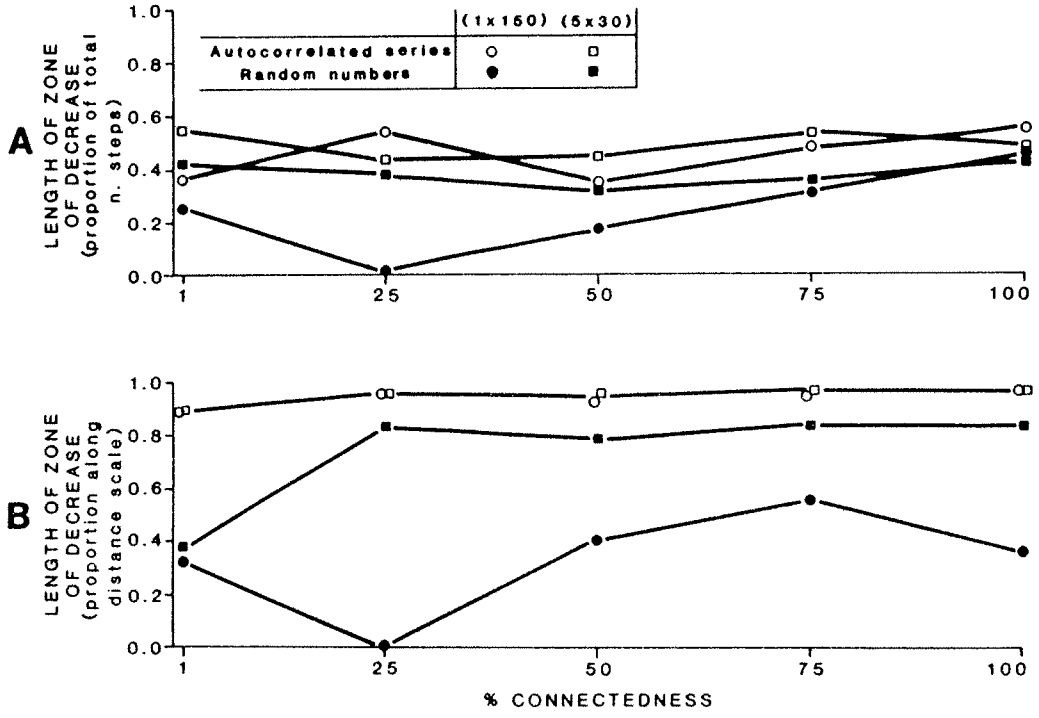

Fig. 7. Length of the zone of decrease, as a function of the connectedness (Co) used during linkage agglomeration, for autocorrelated series (150 points) and for random numbers (150 points). The zone of decrease is measured (A) as a proportion of the total number of steps, or (B) as a fraction of the range of distances where the number of groups decreases, over the total range of distances where agglomeration took place.

columns (chosen to agree with the autocorrelated group structure that we created), the difference is not as large, but it is still significant (sign test). In the lower panel of Figure 7, the ordinate value 0.85 seems to form a line separting the two processes; further statistical investigation of this property is obviously needed, either by Monte Carlo methods, or by studying the theoretical distribution of these statistics for constrained group formation.

ECOLOGICAL CLUSTERING WITH CONSTRAINT OTHER THAN SPACE OR TIME

One step further up the scale of abstraction consists of using constrained clustering to test the hypothesis that a variable or a set of multivariate data forms clusters that are autocorrelated in some other space than geography or time.

An example is given by a set of 99 forest sampling stations studied by P. Drapeau (CREM, Université de Montréal) in the "Municipalité Régionale de Comté du Haut Saint-Laurent", a few kilometers north of the Canada - U.S.A. border in western Québec. A relationship is sought between vegetation composition and edaphic conditions. The hypothesis to be tested is that vegetation is similar under related edaphic conditions. (1) Since the edaphic variables are of mixed types (geomorphology: qualitative; stoniness, soil texture, drainage, topography: semi-quantitative; slope: quantitative; orientation: quantitative circular), a similarity matrix among samples was first computed using the Estabrook and Rogers (1966) coefficient, that can combine data of these various types in a single measure of resemblance. (2) Principal coordinates were computed from this matrix and the first two principal coordinates were taken as an approximation of the edaphic space. (3) From the coordinates of the samples in that space, sample points were interconnected using a Delaunay triangulation (see above). The list of edges from this triangulation, derived only from edaphic data, provided the constraints fed into the space-constrained clustering program. (4) A Steinhaus (Motyka 1947; or Bray and Curtis 1957) similarity matrix was computed from another set of data consisting of the abundance of 28 species of trees in each quadrat; edaphic data were not used in these computations. (5) Proportional-link linkage clustering was performed with 50% connectedness from the vegetation similarity matrix, using as constraints the list of edges obtained above from the edaphic space. As a consequence, only sampling stations that are related in edaphic space were allowed to cluster, if the vegetation data allowed. Two of the clustering steps are represented in Figure 8, mapped onto the first two coordinates of the edaphic space. Botanists could then go back to the raw data and determine what group of species corresponds to each cluster in edaphic space.

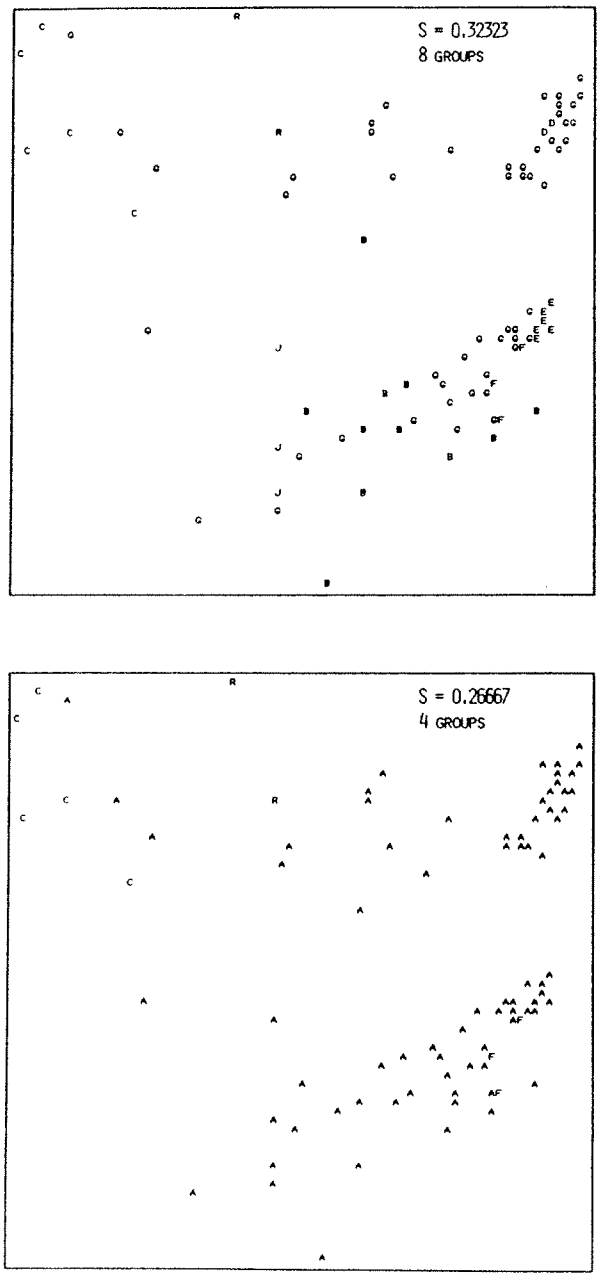

Fig. 8. Two of the steps during constrained agglomerative clustering of the forest vegetation data. Each step is represented by a map whose abscissa is principal coordinate I and ordinate is principal coordinate II of the edaphic space. The clustering similarity level is shown on each map. Each group of sampling stations is represented by a different letter (without order).

One should wonder first if the relationship is real between community structure and the edaphic space that we have constructed by principal coordinate analysis; studying the length of the zone of decrease of the number of clusters shows that the decrease occupies 0.390 of the total number of steps, and 0.442 along the distance scale; these figures fall in the "random numbers" zone of Figure 7, for a connectedness of 50%. So, instead of pursuing the interpretation of these results, one should conclude that the tree community structure data do not lead to significant clusters in the edaphic space, given the way it was created with the data and by the method described above.

CONCLUSION

Our experience with clustering methods that impose a constraint of contiguity through space or time is that the results obtained through a wide range of clustering methods -- linkage clustering, from single to complete linkage -- are much more similar to one another than without the constraint. This is because constraining the clustering process also constrains the set of solutions, eliminating a number of solutions that are compatible with the resemblance matrix, but that do not make much sense in view of the spatial or temporal relationships existing among the samples under study.

From the descriptive point of view, constrained clustering is one of the few ways available for synthetically representing multivariate data onto a map. With many ecological problems, this type of mapping is far more interesting than separate maps of the variables forming the multivariate data set. On the other hand, theories about the importance of dispersal routes for individual species or for whole biotic communities could be tested by comparing the unconstrained to the space-constrained classification of sites; many other hypotheses of contagiousness of ecological processes through space or time could be tested in the same way.

A number of constrained clustering programs have been written and are available to other users. This is the case at least with the present author's programs used in the examples presented above, as well as the program of Lebart (1978, for two-dimensional constraint), whose paper includes the program listing. De Soete *et al.* (1987) present algorithms for deriving constrained classifications in a more general context than that of the present paper, and they review the psychometric literature on the subject.

Geographic information can also be used with unconstrained clustering programs. If **A** is the ecological distance matrix among objects, build a matrix **B** ("penalty matrix") containing spatial information (either geographic distances, or 0 = connected and 1 = unconnected objects).

Compute $\mathbf{C} = \mathbf{A} + w\,\mathbf{B}$, where w is a scalar weight. Cluster **C** for different values of w and pick the result with the smallest w where all clusters are internally contiguous. This method can also be used to obtain constrained ordinations.

In the future, constrained clustering programs, if they are agglomerative, should be made to include some measure of the information content of the various clustering levels, and also perhaps a measure of "patchiness" such as the one developed in one of the previous sections. Since clustering with constraint includes, in the data analysis process, some *a priori* knowledge that is pertinent to many of the theories the ecologists are dealing with, it may be viewed by these same ecologists as an interesting method both for descriptive purposes and for hypothesis testing.

REFERENCES

Bell, M. A., and P. Legendre. 1987. Multicharacter chronological clustering in a sequence of fossil sticklebacks. Syst. Zool. 36: (in press).

Bray, R. J., and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27: 325-349.

Cliff, A. D., and J. K. Ord. 1981. Spatial processes: models and applications. Pion Limited, London. 266 p.

De Soete, G., J.D. Carroll, and W.S. DeSarbo. 1987. Least squares algorithms for constructing constrained ultrametric and additive tree representations of symmetric proximity data. J. Class. (in press).

Estabrook, G. F., and D. J. Rogers. 1966. A general method of taxonomic description for a computed similarity measure. BioScience 16: 789-793.

Fisher, W. D. 1958. On grouping for maximum homogeneity. J. Amer. Stat. Ass. 53: 789-798.

Gabriel, K. R., and R. R. Sokal. 1969. A new statistical approach to geographic variation analysis. Syst. Zool. 18: 259-278.

Galzin, R., and P. Legendre. 1987. The fish communities of a coral reef transect. Pac. Sci. (in press).

Gordon, A. D. 1973. Classification in the presence of constraints. Biometrics 29: 821-827.

Gordon, A. D., and H. J. B. Birks. 1972. Numerical methods in Quaternary palaeoecology. I. Zonation of pollen diagrams. New Phytol. 71: 961-979.

Gordon, A. D., and H. J. B. Birks. 1974. Numerical methods in Quaternary palaeoecology. II. Comparison of pollen diagrams. New Phytol. 73: 221-249.

Green, P. J., and R. Sibson. 1978. Computing Dirichlet tessellations in the plane. Computer J. 21: 168-173.

Hawkins, D. M., and D. F. Merriam. 1973. Optimal zonation of digitized sequential data. J. Int. Assoc. Math. Geology 5: 389-395.

Hawkins, D. M., and D. F. Merriam. 1974. Zonation of multivariate sequences of digitized geologic data. J. Int. Assoc. Math. Geology 6: 263-269.

Howe, S. E. 1979. Estimating regions and clustering spatial data: analysis and implementation of methods using Voronoi diagrams. Ph. D. Thesis, Department of Mathematics, Brown University.

Ibanez, F. 1984. Sur la segmentation des séries chronologiques planctoniques multivariables. Oceanol. Acta 7: 481-491.

Lance, G. N., and W. T. Williams. 1967. A general theory of classificatory sorting strategies. I. Hierarchical systems. Computer J. 9: 373-380.

Lebart, L. 1978. Programme d'agrégation avec contraintes (C. A. H. contiguïté). Cah. Anal. Données 3: 275-287.

Lefkovitch, L. P. 1987. Species associations and conditional clustering: clustering with or without pairwise resemblances. This volume.

Legendre, P., B. Baleux, and M. Troussellier. 1984. Dynamics of pollution-indicator and heterotrophic bacteria in sewage treatment lagoons. Appl. Environ. Microbiol. 48: 586-593.

Legendre, P., S. Dallot, and L. Legendre. 1985. Succession of species within a community: chronological clustering, with applications to marine and freshwater zooplankton. Am. Nat. 125: 257-288.

Legendre, P., and V. Legendre. 1984. Postglacial dispersal of freshwater fishes in the Québec peninsula. Can. J. Fish. Aquat. Sci. 41: 1781-1802.

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

Matula, D. W., and R. R. Sokal. 1980. Properties of Gabriel graphs relevant to geographic variation research and the clustering of points in the plane. Geogr. Anal. 12: 205-222.

Milligan, G. W. 1983. Characteristics of four external criterion measures, p. 167-173. *In* J. Felsenstein [ed.] Numerical taxonomy. NATO Advanced Study Institute Series G (Ecological Sciences), No. 1. Springer-Verlag, Berlin.

Monestiez, P. 1978. Méthodes de classification automatique sous contraintes spatiales, p. 367-379. *In* J. M. Legay and R. Tomassone [ed.] Biométrie et écologie. Société française de Biométrie, Paris.

Motyka, J. 1947. O zadaniach i metodach badan geobotanicznych. Sur les buts et les méthodes des recherches géobotaniques. Ann. Univ. Mariae Curie-Sklodowska Sect C, Suppl. I. viii + 168 p.

Okabe, A. 1981. Statistical analysis of the pattern similarity between 2 sets of regional clusters. Environment and Planning A 13: 547-562.

Openshaw, S. 1974. A regionalisation program for large data sets. Computer Appl. 3-4: 136-160.

Rajski, C. 1961. Entropy and metric space, p. 44-45. *In* C. Cherry [ed.] Information theory. Butterworths, London.

Ray, D. M., and B. J. L. Berry. 1966. Multivariate socioeconomic regionalization: a pilot study in central Canada, p. 75-130. *In* S. Ostry and T. Rymes [ed.] Papers on regional statistical studies. Univ. of Toronto Press, Toronto.

Sokal, R. R., N. L. Oden, and J. S. F. Barker. 1987. Spatial structure in *Drosophila buzzatii* populations: simple and directional spatial autocorrelation. Am. Nat. 129: 122-142.

Sokal, R. R., and J. D. Thomson. 1987. Applications of spatial autocorrelation in ecology. This volume.

Ward, J. H. Jr. 1963. Hierarchical grouping to optimize an objective function. J. Amer. Stat. Assoc. 58: 236-244.

Wartenberg, D. E. Regional analysis: describing multivariate data distributions using geographic information. Manuscript (cited with permission of the author).

Webster, R. 1973. Automatic soil-boundary location from transect data. J. Int. Assoc. Math. Geology 5: 27-37.