



**Succession of Species within a Community: Chronological Clustering, with Applications to Marine and Freshwater Zooplankton**

P. Legendre; S. Dallot; L. Legendre

*American Naturalist*, Vol. 125, No. 2 (Feb., 1985), 257-288.

Stable URL:

<http://links.jstor.org/sici?sici=0003-0147%28198502%29125%3A2%3C257%3ASOSWAC%3E2.0.CO%3B2-J>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*American Naturalist* is published by The University of Chicago Press. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

---

*American Naturalist*

©1985 The University of Chicago Press

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

## SUCCESSION OF SPECIES WITHIN A COMMUNITY: CHRONOLOGICAL CLUSTERING, WITH APPLICATIONS TO MARINE AND FRESHWATER ZOOPLANKTON

P. LEGENDRE,\* S. DALLOT,† AND L. LEGENDRE‡

\*Département de sciences biologiques, Université de Montréal, C.P. 6128, Succursale A, Montreal, Quebec H3C 3J7; †Station zoologique, Université Pierre et Marie Curie, 06230 Villefranche-sur-Mer, France; ‡GIROQ, Département de biologie, Université Laval, Quebec, Quebec G1K 7P4

*Submitted April 7, 1983; Revised May 7, 1984; Accepted June 7, 1984*

According to Margalef (1968), the theory of species succession within an ecosystem plays the same role in ecology as does the evolutionary theory in general biology. Several phenomena are associated with the succession of species, such as increases in biomass, diversity, and homeostasis. These phenomena have been summarized by Margalef (1968, 1974), Odum (1971), Frontier (1977), and Gutierrez and Fey (1980). They are measurable and lead to an increase in homeostasis and in information content of the ecosystem.

There is as yet no general model of succession integrating all these phenomena. Therefore, it would be of help to ecologists if it were at least possible to describe accurately and unambiguously the steps through which an ecosystem goes during the process of ecological succession. It could also help in the modeling of successional events for management purposes.

The very idea that an ecological succession evolves in steps, instead of smoothly, is well known to students of successional phenomena. It has been clearly summarized in the ecological model described by Allen et al. (1977, pp. 1076, 1082) who state: "There are structural forces intrinsic to the community, and the community composition moves from one center of community structure to another only by the intervention of intermittent shifts in environmental structure. . . . The model suggested . . . resembles a gently wobbling, scalloped surface with a steel ball held in craters. Intermittently, a randomly directed magnetic force moves the ball, sometimes into a new hollow." (See also fig. 7.) Allen et al. (1977) analyzed the first-difference vectors of weekly species occurrences, through principal component analysis, to show this process operating in a real community, while in this paper we will attempt to reveal it using another type of analysis, based on the original data matrix.

The contribution of this paper lies in the development of a method to describe disjunctions occurring during an ecological succession. Models emphasizing disjunction are useful for many kinds of analyses, while admittedly, succession may

be viewed under its dual aspect of continuity, for other purposes. In this latter case, another multivariate method, which can be considered to be complementary to the one developed below, has been proposed by Orlóci (1981) for uncovering successional trends and separating them into monotonic and cyclic components. On the other hand, the disjunction model may well not be applicable to all kinds of successions. It has been found, however, to be adequate in many types of communities and environments, and among them phytoplankton (Allen et al. 1977, cited above); zooplankton (Dessier and Laurec 1978; Sprules 1980); plant seral stages with distinctive breeding bird species (Johnston and Odum 1956); and transitional discontinuity between pioneer stage and building stage in a subtropical rain forest (Williams et al. 1969).

The ecological model described above becomes useful only if one can observe disjunctions in real series of ecological samples. Empirical methods (like the usual agglomerative clustering techniques, for instance) for doing so exist, but they have the drawback that they always partition the samples into groups, without providing indications as to the reality of these groups. Here we proceed the other way around. A mathematical model is found which corresponds to the ecological model of discrete successional steps, and this statistical model is imbedded in a computer program capable of handling real and relatively large data sets. This procedure is applied to two real data sets representing, respectively, seasonal (within-year) and long-term successions. It is then compared with other methods proposed in the literature.

More precisely, our objectives are the following: (1) to identify the discontinuities in a time series of multispecies biological samples, collected at a single station; (2) to take into account the time sequence of the sampling, this method thus being different from the usual clustering techniques; (3) to make it possible to eliminate singletons.

This last preoccupation with singletons, which might be thought of as "aberrant samples," comes from the very nature of ecological data series. (1) In most ecological data series, random events may temporarily perturb the sequence of the succession. These random events include external forcings, such as perturbed or missing strata in sediment cores, or temporary shifts of water masses at a fixed station in an aquatic ecosystem, where it is far from certain that an ecologist will always sample the same water mass. Also included under this heading are emigration and immigration phenomena, pertaining to another scale of perception than the maturation of the succession under study. They can perturb the data series in an unpredictable, temporary, and nonsignificant manner. (2) A singleton in a data series may also be the result of an extreme stochastic variation within the range of the given successional step. Indeed, the statistical criterion proposed in Appendix C causes the rejection of the null hypothesis that says "this observation should be included in the successional step," with a probability of error  $\alpha$ , accounting for these extreme cases. (3) Finally, bizarre samples may be the result of improper sampling or inadequate preservation of the samples.

These events may happen in any ecological data series, except when one is sampling fixed organisms, long-lived relative to the sampling interval, from permanent and well-defined quadrats. These are found for instance in studies of trees,

or of shallow-water sessile organisms that are observed without being removed. Consequently, because our second objective imposes a time constraint on the grouping of samples, we decided on the following compromise as ecologically meaningful: a single sample, significantly different from those found immediately before and after itself, should be considered temporarily aberrant by the algorithm, in order to prevent it from interrupting a sequence of ecologically homogeneous samples. This singleton can later be tried and fitted with other sequences of the time series (in which case an external forcing may be hypothesized). For this purpose, a switch has been included in our algorithm (fig. 2) allowing the ecologist to have these singletons identified and removed from the data series, if he so wishes. However, any group of samples longer than one will be kept as representing a step in the succession.

#### METHODS

The bases of the chronological clustering method follow naturally from the principles discussed above. One begins by computing a matrix of samples  $\times$  samples distance. The choice of a measure of resemblance appropriate to the data and to the problem to be solved is discussed in Appendix B. This preliminary step is illustrated in figure 1, and it is followed by the chronological clustering itself, explained in some detail below with reference to figure 2.

After reading in the distance matrix (step 1 in fig. 2), clustering the samples is carried out from the distance matrix using any method of agglomerative clustering, with the time contiguity constraint explained in the next paragraph. Proportional-link linkage (Sneath 1966), used hereinafter (steps 2 to 5 in fig. 2), is an agglomerative and hierarchical algorithm of intermediate linkage clustering. In this clustering, the user specifies, out of the maximum number of links that can form between any two clusters, the proportion of links that are necessary for fusion of these two clusters into one. This proportion will be called  $Co$  (connect-*edness*) and it is set by the user (step 1 in fig. 2).

Clustering with a constraint of contiguity has been proposed by several authors in the past few years (time contiguity: Gordon and Birks 1972; Gordon 1973; North 1977; space contiguity: Lefkovitch 1978, 1980; Monestiez 1978; Lebart 1978; Roche 1978). Time contiguity, in particular, means that only those samples that are contiguous along the time series are allowed to cluster. This principle is applied to the study of successions in our method (during step 2 in fig. 2). It is modified by the identification and removal of singletons, as defined in the introduction. This is a fundamental problem as far as ecological successions are concerned, one that has yet to be tackled by any of the authors mentioned above.

A hierarchical clustering algorithm naturally produces a hierarchy. Since this hierarchy has little bearing on the problem of succession (it must actually be considered as a computational artifact), we will seek instead a single nonhierarchical final solution, corresponding somewhat to a broken vertical line through the horizontal hierarchical tree. To achieve this, a statistical criterion of cluster fusion must be developed (step 4 in fig. 2). This topic will occupy the next subsection.

After presenting the criterion of cluster fusion, the last subsections of the

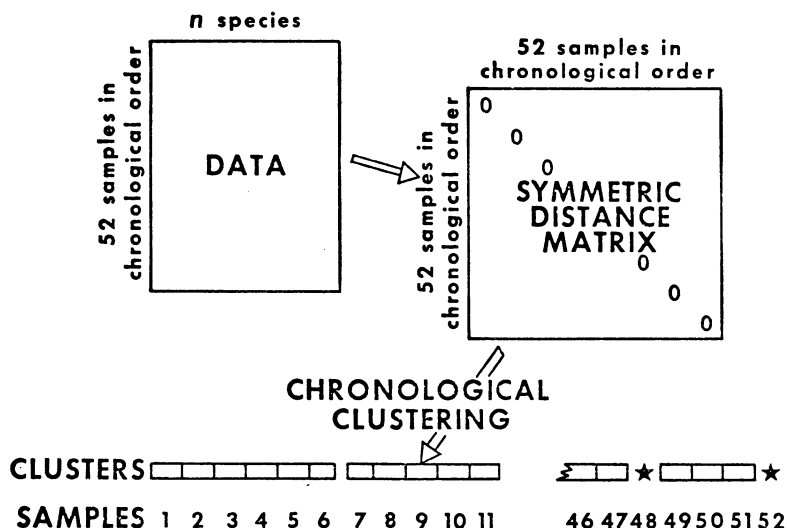


FIG. 1.—Summary of the chronological clustering procedure, for a series of 52 samples  $\times$   $n$  species. The nonhierarchical clusters are symbolized by connected boxes. Stars are singletons.

METHODS will be devoted, respectively, to the procedure for removing aberrant samples (step 6 in fig. 2), the group expansion tests (step 7), the a posteriori testing (step 8), and finally a summary of the chronological clustering method.

*A test of cluster fusion* (step 4 in fig. 2).—As mentioned in the introduction, the ecological model upon which this chronological clustering is built says that succession evolves in steps, rather than smoothly. This implies that within a step, the variation is small compared to the changes occurring between steps. This statement, in turn, can be translated into the following terms: within a step, all the time-ordered samples behave as if they resulted from randomly sampling the same biological community. This formulation corresponds to the null hypothesis of the statistical test of cluster fusion, described in more detail in Appendix C.

*Removal of singletons* (step 6 in fig. 2).—It has already been mentioned that we wish to obtain a set of clusters that is not affected by singleton perturbations in the sample series, as are frequently found in ecological data series (see the introduction). A procedure for identifying singletons naturally follows from the test of significance described above. The intermediate linkage algorithm agglomerates hierarchically from lower to higher distance levels. A singleton occurs when the procedure comes to the point where a given sample is considered for clustering, and it is found that this sample can cluster neither at its left nor at its right (except, of course, for the first and last samples).

When a sample has been identified as a singleton, the procedure is immediately interrupted. The singleton is marked (by a star, in our computer program) and the clustering procedure is started again from the beginning after removing the singleton. This is made necessary by the reason illustrated in table 1: by its very presence in the series, a singleton may prevent its neighbors from clustering with one another, even though they are more alike than any other pair at that linkage

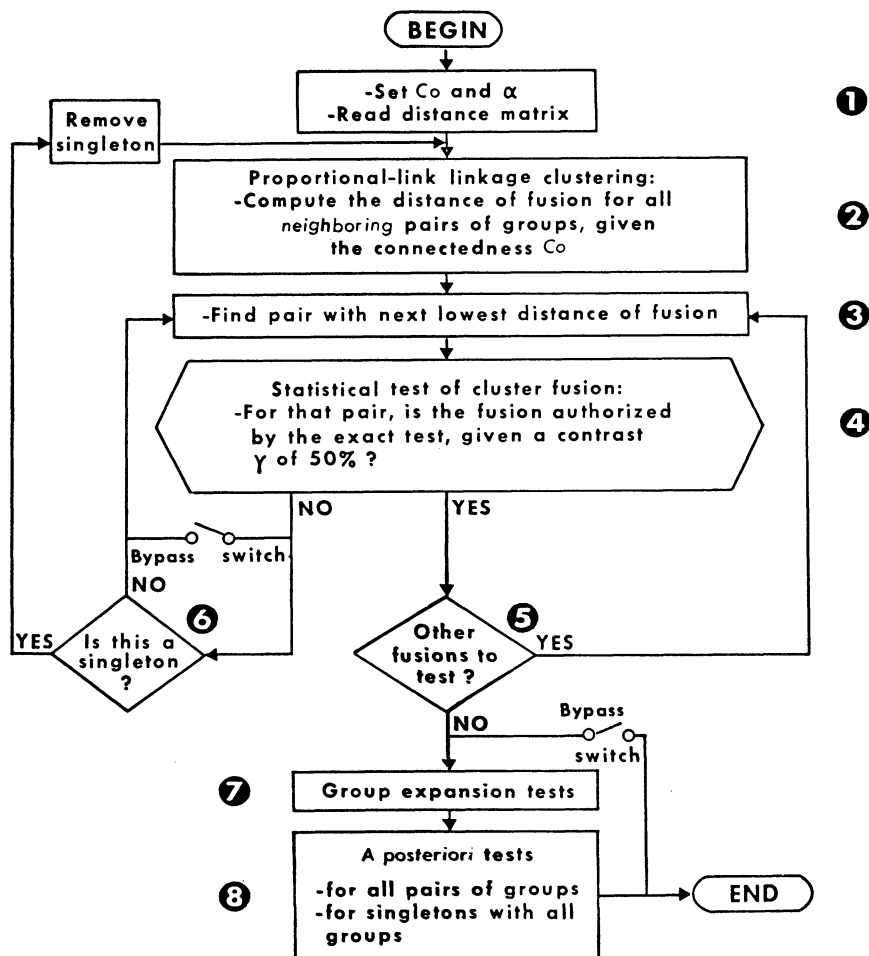


Fig. 2.—Flow diagram of the chronological clustering algorithm. Circled numbers are referred to in the text.

level (compare level  $D = 0.05$  before and after removal of sample 5, in the table). This phenomenon may well irreversibly alter the clustering topology right to the end of the procedure, so that it is safer to start again from the beginning.

*Group expansion tests and interval graphs* (step 7 in fig. 2).—The clustering algorithm was designed to produce a single set of nonoverlapping clusters. Ecologically speaking, it is important to determine whether these clusters represent completely different and independent phenomena resulting, for instance, from an external forcing or, rather, relatively stable subseries of events within a continuous ecological succession. In the first case (independent phenomena), we would expect, if we destroyed the clustering structure except for one cluster, that this cluster will remain the same. In the second case (continuum), however, this cluster could still “grow” a little, by incorporating some of the samples located immediately at its left or right; in the chronological clustering procedure, these

TABLE 1  
EXAMPLE OF CLUSTER FORMATION BEFORE AND AFTER REMOVAL OF A SINGLETON (★)  
(Synthetic distance matrix with eight samples)

LEVELS	SAMPLES							
	1	2	3	4	5	6	7	8
D = .02	...	...	A	A	...	...	...	...
D = .04	...	A	A	A	...	...	...	...
D = .05	...	A	A	A	...	B	B	...
D = .06	...	A	A	A	...	B	B	B
D = .09	A	A	A	A	...	B	B	B
The singleton (sample 5) is removed†								
D = .02	...	...	A	A	★	...	...	...
D = .02	...	...	A	A	★	A	...	...
D = .05	A	A	B	B	★	B	...	...
D = .06	A	A	B	B	★	B	C	C
D = .08	A	A	A	A	★	A	B	B

NOTE.—Co = 50%, α = .25. Letters identify groups, from left to right.  
† At this point, the probability of fusion of sample 5 was .20 on its left and .25 on its right.

samples have been prevented from joining the said cluster by the fact that they were already members of other clusters. Taking alone, in turn, each of the previously formed clusters and allowing it to expand along the time series (singletons removed) leads to the identification of overlapping clusters that can be represented by the interval graphs of figures 3 and 5. It should be obvious that, as a cluster expands, the test becomes more permissive and can accept more and more heterogeneous samples, which is a desirable property in this case.

*A posteriori tests* (step 8 in fig. 2).—Recurrent biological events may be identified from the results of the chronological clustering by testing a posteriori if there are differences between clusters, independently from the time sequence. In the same way, one can try to relate the singletons to distant clusters in the time series, in which case they can be called *satellites* of these distant clusters. In figure 4, for instance, it seems that samples [56 to 66] represent a recurrence of the cluster [1 to 3]. In the same way, singleton 35★ is a satellite of cluster [29 to 41]. See also the last section of Appendix C for more comments on a posteriori testing.

Summary of the Chronological Clustering Algorithm

To summarize (fig. 2), a proportional-link linkage algorithm is used to agglomerate neighboring samples hierarchically. It generates different pairs of samples or groups of samples, in turn, to be candidates for clustering. The test for accepting the fusion of two groups is a randomization test involving a contrast parameter γ (set at 50%) to classify the distances into high and low distances. The hypothesis that the ratio

$$\left[ \frac{\text{no. high } D}{\text{total no. } D} \right]_{\text{between-group matrix}} = \left[ \frac{\text{no. high } D}{\text{total no. } D} \right]_{\text{triangular matrix}}$$

is tested, and if the probability of its being true is larger than a predetermined significance level  $\alpha$ , then the two groups under consideration are clustered.

Singletons are identified using the same test. When one is found, it is removed from the series until the a posteriori tests, and the whole clustering procedure is started again from the beginning.

At the end of the clustering proper, each group is allowed to expand at the expense of its neighbors, in order to determine whether the successional structure found represents relatively stable subseries of events within a continuum, or different and independent phenomena succeeding one another. Furthermore, a posteriori tests are carried out between nonneighboring clusters, as well as between singletons and each of the clusters, in order to determine their relationships.

#### RESULTS USING NATURAL COMMUNITIES

##### *Example 1: Mediterranean Zooplankton*

As a first example of the behavior of chronological clustering, we used a series of marine zooplankton samples collected at the entrance of Villefranche Bay (Mediterranean, France: 7°15' E, 43°41' N) by vertical bottom-to-surface (75 m) hauls of a Juday-Bogorov net (330  $\mu$ m mesh) between 15 November 1966 and 14 May 1968. Data are counts of 13 groups of chaetognaths (7 species, 3 sexual maturity levels being considered for each of the three most abundant species). The first year of this data series has been analyzed for other purposes by Ibanez and Dallot (1969) and by Ibanez (1972, 1974). In order to remove small-scale variability, individual counts (3–11 samples per wk) were combined into 78 weekly averages. This series is actually the beginning of a 6-yr series which will be analyzed elsewhere. This segment was chosen to be short enough to be represented in an uninterrupted graph (fig. 3), yet long enough to illustrate the recurrence of seasonal zooplankton events (1½ yr).

The data were subjected to the usual  $\log(y + 1)$  transformation. For all the variables so transformed, normality was not rejected at the 1% level (class zero removed) by the Kolmogorov-Smirnov test of goodness-of-fit, as modified by Lilliefors (1967).

The Mediterranean environment is known to be more mature and stable than that of example 2 (below), for instance. Its maturity was confirmed by the shape of the rank-frequency diagrams of the 78 samples. Since a type-I coefficient of distance is appropriate in this situation (see Appendix B), Whittaker's index of association was computed between all pairs of samples, and the chronological clustering was run on this distance matrix.

In figure 3, each cluster of contiguous samples is represented by a horizontal line segment, while discontinuities are represented by blanks. Four values of connectedness are compared in figure 3a. The breakpoints seem fairly stable throughout the various connectedness levels. Connectedness 25% is chosen for the remainder of this example, although other levels could have been used and will be mentioned later.



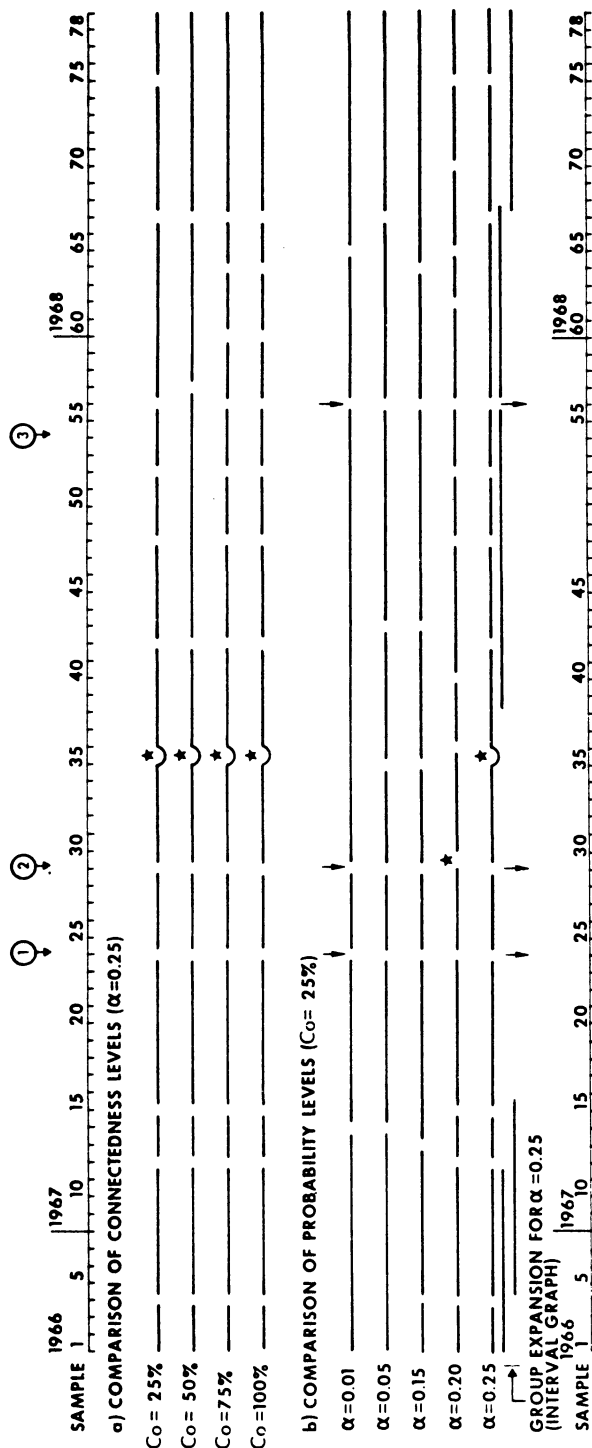


FIG. 3.—Chronological clustering of the Villefranche chaetognaths series from the Whittaker index-of-association matrix. Full horizontal lines: clusters of contiguous samples. Stars: singletons. *a*, Comparison of four connectedness levels, keeping  $\alpha$  fixed at 0.25. *b*, Comparison of five  $\alpha$  probability levels, keeping  $C_0 = 25\%$ . The group expansion tests are shown for  $\alpha = 0.25$ . Arrows indicate the most important breakpoints in the data series. Circled numbers at the top of the graph identify the most important hydrometeorological events: 1, establishment of thermal stratification; 2, breakdown of thermal stratification; 3, mistral storm.

As could be expected, the smallest  $\alpha$  values in figure 3b produce fewer, more encompassing groups. Indeed, looking at the nature of the test for cluster fusion in Appendix C, one realizes that the test allows more fusions when  $\alpha$  is small, which leads to larger groups. Thus, varying  $\alpha$  from 0.01 to 0.25 is equivalent to probing the data series with increasing intensity, the larger  $\alpha$  values allowing finer details to emerge. In this respect, it is noticeable that singletons show up only at  $\alpha = 0.20$  and above. Indeed, well-detached singletons can be expected to be less "obvious" than the major breakpoints structuring the series, which can be found through the whole scale of  $\alpha$  values.

One may also wonder what is the interest of using a probability level of 20% or larger. Actually, the test is used here mostly as a clustering intensity parameter, although it is shown in Appendix C that  $\alpha$  is equal to the probability of a Type I error during the test. In the present case, it is more interesting to break the series into pieces smaller than those obtained with  $\alpha = 0.01$ . At  $\alpha = 0.30$ , however, too many small groups (17) are obtained, so that  $\alpha = 0.20$  or 0.25 produces more interesting results. The following paragraphs are based on the 12 groups obtained at  $\alpha = 0.25$ .

It is, however, worth considering a whole set of levels. Indeed, looking at figure 3b, the most important breakpoints in the data series become obvious. They are indicated by arrows on the figure and are found between samples 23 and 24, 28 and 29, and 55 and 56. The importance of these breakpoints is confirmed by the fact that no group expansion covers them. Larger  $\alpha$ -values allow the identification of smaller steps within the main ones, which may give a truer picture of the ecological reality (see DISCUSSION AND CONCLUSIONS).

The main breakpoints in the data series, marked by arrows in figure 3b, correspond to the major hydrological events positioned at the top of the figure by circled numbers. The 23–24 breakpoint (end of April 1967) corresponds to the formation of thermal stratification of the water mass, the vertical stratification (measured by  $\Delta\sigma_t$  = the difference in water density between bottom and surface water) changing suddenly from 0.27 (week 23) to 0.70 (week 24). The 28–29 breakpoint is the result of several days of strong mistral wind. This blew surface waters away from the coast, and these were replaced by an upwelling of subsurface water (Bougis 1968), with the consequence that the concentrations of epipelagic species, usually found near the shore, were lowered. This perturbation was followed by a typical summer community dominated by *Sagitta setosa* and high numbers of young *S. enflata* in weeks 29 to 41, very different from the preceding springtime community, poor in species. This transition is so abrupt that sample 29\* behaved as a singleton in one of the chronological clustering runs. The last major breakpoint, between weeks 55 and 56 (December 1967), follows closely the breakdown of thermal stratification ( $\Delta\sigma_t$  was 0.04 on week 54). The 2-wk delay in the reaction of the zooplankton community may be attributed to the homeostatic properties of the community.

Sample 35\* also behaved as a singleton in all clustering runs at  $\alpha = 0.25$ . Singletons may be appreciated correctly only by going back to the data, since many causes may lead to the production of singletons (see introduction). Sample 35\* is singled out because of a general increase in population densities that was

irregular from one group to another. The peculiarities of this sample become visible only at high  $\alpha$ -values, which are known to bring up finer details. The a posteriori tests show, however, that the probability of sample 35\* belonging to the group [29 to 41] is 1.00, after this group has completed its membership. Some of the secondary breakpoints can also be attributed to hydrometeorological causes.

The groups formed by expansion follow in a striking way the main phases of the hydrological cycle, mainly characterized by the vertical water stability variable ( $\Delta\sigma_t$ ) and the temperature. The two expanded groups [1 to 11] and [4 to 15] correspond to the hydrological winter period, with low vertical stability and low temperature; the corresponding period in 1967–1968 forms a single cluster, [56 to 66]. On the other hand, the expanded group [38 to 55] corresponds to the fall period, beginning at maximum annual temperature and ending shortly after the vertical stratification has disappeared; unfortunately, the 1968 data series in this example does not extend far enough to cover this same period.

The a posteriori tests between groups, described in the METHODS, can be used to investigate the assumption of recurrence of successional events from year to year. The between-group tests with a probability higher than  $\alpha$  are represented in figure 4, which is drawn in such a way as to preserve the temporal sequence on successive lines while showing the between-group relationships as simply as possible. As expected, this figure shows a major folding of the series corresponding to the second year of sampling, with group 1 related to group 10 (wintertimes) and group 4 related to group 12 (springtimes). In computer runs with a connectedness of 75% and 100% (not illustrated), another a posteriori link joined group [12 to 14] with a group [64 to 66], thus strengthening the between-years relationship. During the year 1967, a secondary fold is present in the data series (links between groups 4 and 9, and between 8 and 12) relating the two springtime communities with the fall community, even though these communities pertain to different phases of the hydrological cycle. Interestingly, the same spring-fall relationship will be seen in the freshwater community of a northern reservoir, in the next example.

#### *Example 2: Zooplankton in Freshwater Reservoir*

In May 1977, the *Société d'énergie de la baie James* impounded a small reservoir (7 km<sup>2</sup> approximate area) called Desaulniers, in northwestern Quebec (77°32' W, 53°36' N). Several sampling stations were established and visited both before and after the flooding. The deepest of these stations, G2-129 (maximum depth 13 m), will be studied hereinafter for the evolution of its zooplankton community (crustaceans and rotifers). The counts for the 33 species are in numbers of individuals per volume of water. Before the flooding, this station was located in the riverbed, and only zooplankton animals drifting from lakes located upstream could be found; thus, this is an example of primary succession. Its evolution has been described in more detail elsewhere (Pinel-Alloul et al. 1982).

An important difference between this example and the Villefranche chaetognath data is that the time lag between samples is anything but constant. This is due in part to the difficulties of winter sampling in the northern regions. It also reflects

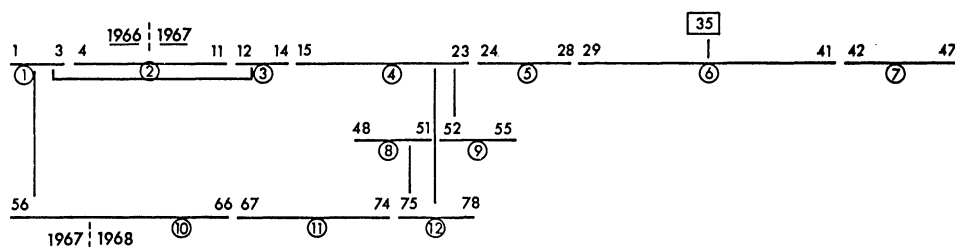


FIG. 4.—Schematic representation of the chronological clustering of example 1, showing the between-group relationships ( $p > \alpha$ ) indicated by the a posteriori tests. The boxed sample is a singleton.

the a priori knowledge of biologists who expect the zooplankton communities to evolve much more slowly during winter. This example will then serve to illustrate that, contrary to time-series analysis, the chronological clustering method does not require the data to be sampled at fixed intervals. Of course, differences in sampling density along time will have to be given due attention when interpreting the clustering results.

Prior to analysis, the 33 species were normalized by the log ( $y + 1$ ) transformation, as in example 1. Only one of the variables so transformed was found not to be normal at the 1% probability level (class zero removed), by the Kolmogorov-Smirnov goodness-of-fit test, as modified by Lilliefors (1967). Since the beginning of a primary succession forms by definition an immature community that is likely to remain so for some time, especially in a northern environment, the Canberra metric (a type-2 coefficient of distance, see Appendix B) was used to compute the relationship between all pairs of samples.

At 75% and 100% connectedness, the chronological clustering of the Canberra distance matrix gave exactly the same results as with  $Co = 50\%$ . At 25% some groups are fused (samples 20 to 28, with 25 as an in-group singleton; samples 39 to 44) without any change in the position of the breakpoints. Therefore, the connectedness will be fixed at 50% for the remainder of this example.

Figure 5 shows that, here again, smaller  $\alpha$  values produce fewer but larger clusters. The most important breakpoints in the series can be described at the smallest  $\alpha$  levels. The flooding of the reservoir (marked 1 at the top of the figure) at the end of May 1977 makes a clear departure from the zooplankton population that was present in the river in 1976, drifting from the lakes located upstream. This event corresponds to an increase in the number of species from 2 to 8. In mid-July 1977, after rising above 139 m altitude, the water level began to drop and the number of species increased from 13 to 19. In mid-August 1977, after a drop to the 137.5-m mark, the water level rose again, and the number of species increased from 14 to 20. Between the October and December samplings, the water temperature fell below  $4^{\circ}\text{C}$ , the fall overturn took place, the ice cover was established, and the number of species fell from 17 to 5 (more samples would have been desirable during this transition period). Finally, at the end of May 1978, the rising water temperature crossed above  $4^{\circ}\text{C}$  (spring overturn) and the number of species

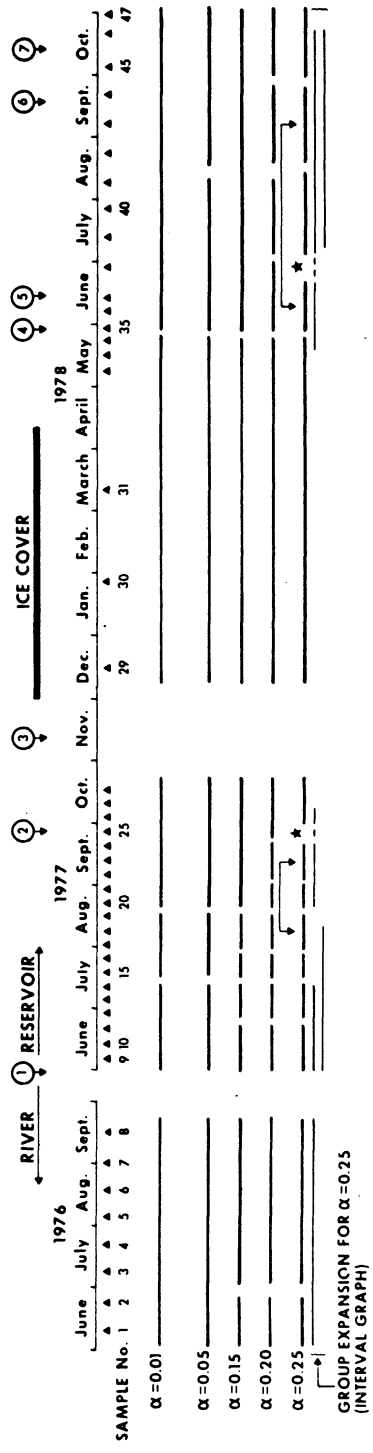


FIG. 5.—Effect of the probability level  $\alpha$  of the exact test, on the clustering structure of the Desaulniers zooplankton series, using the Canberra distance function. Connectedness is 50%. Group expansion tests are shown for  $\alpha = 0.25$ . Full horizontal lines: clusters of contiguous samples. Stars: singletons. Double arrows: a posteriori tests with probability of fusion higher than  $\alpha$ . Circled numbers at the top of the graph identify the most important hydrological events: 1, creation of the reservoir; 2, cool-off reaches 10° C; 3, cool-off reaches 4° C; 4, warming-up reaches 4° C; 5, warming-up reaches 10° C; 6, cool-off reaches 10° C; 7, cool-off reaches 4° C.

increased from 9 to 17. It is noticeable that no other break is produced by the  $\alpha = 0.01$  clustering. At  $\alpha = 0.05$ , another break appears in August 1978, corresponding to the beginning of a drop in the water level, after it had almost reached 142 m.

Clustering at  $\alpha = 0.25$  is also worth considering, since it results in smaller groups which correspond better to the reduced-space ordination (fig. 6) obtained by the nonmetric multidimensional scaling method, applied to the Canberra distance matrix. The most interesting feature is the presence of two singletons. The first one (sample 25) is clearly isolated in figure 6 and corresponds to the point where the water temperature drops below the  $10^{\circ}\text{C}$  mark, while the second one (sample 38) corresponds to a major drop in species evenness which may indicate a restructuring of the zooplankton community interactions, after the water temperature had passed the  $10^{\circ}\text{C}$  mark in its spring warm-up (hydrological event 5 in fig. 5). The  $10^{\circ}\text{C}$  transition seems relevant in freshwater zooplankton community development (Jacobs and Bouwhuis 1979). Another feature of the  $\alpha = 0.25$  clustering is the presence of a break in the series in the fall of 1978, some time after the  $10^{\circ}\text{C}$  transition, accompanied by a drop from 16 to 8 in the number of species, between samples 44 and 46. This delay in the reaction of the zooplankton community to the beginning of the fall hydrological period was also observed in the Mediterranean zooplankton series (example 1).

The interval graph resulting from the group expansion tests computed for the  $\alpha = 0.25$  clustering is plotted in figure 5. It confirms that the most important breakpoints are those identified by clustering at the lower  $\alpha$  values. The a posteriori tests for the two singletons show, above the level of  $\alpha = 0.25$ , the existence of possible relationships only with the three two-sample groups, with a probability of 0.66667. As indicated in table C1, this value is the lowest probability that can be found between a singleton and a two-sample group, so no meaning is attributed to these findings.

Finally, the a posteriori tests between all pairs of groups are computed. Very few show a probability larger than  $\alpha$ . In a first group, the three small two-sample groups [1–2], [15–16], and [20–21] show among them a probability of 0.33333. Table C1 shows that, since  $\gamma = 0.50$ , this value is also the smallest possible probability of fusion obtainable from the exact test. Consequently, no attention is to be paid to these results. Besides these, only two between-group relationships are left, between groups [17–19] and [22–24], and between groups [35–37] and [42–44], with a probability of fusion equal to 0.40000. They are drawn in figure 5 by double arrows.

From figures 5 and 6, we can draw the following description of the evolution of zooplankton communities related to the impoundment of Reservoir Desaulniers. In 1976, zooplankton were drifting at random from the small lakes located upstream; this is evidenced by low species numbers and highly fluctuating evenness, which indicate that no stable community was present at station G2-129. After the flooding of the reservoir, the community departs rapidly from the river status (fig. 6) and forms a fairly well developed assemblage with 13 to 20 species in the summer of 1977, despite large chemical and water-level fluctuations (Pinel-Alloul et al. 1982). The August [17–19] and September [22–24] communities are suffi-

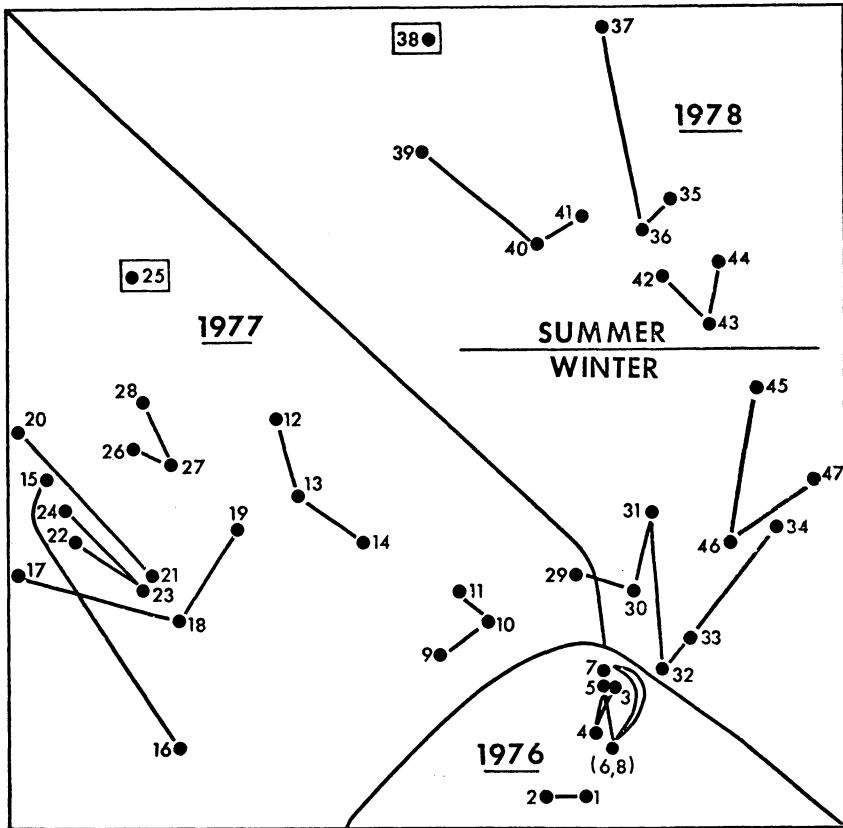


FIG. 6.—Nonmetric multidimensional scaling of the Desaulniers zooplankton series, after computation of the Canberra distance between samples. Line segments show the results of the chronological clustering with  $Co = 50\%$  and  $\alpha = 0.25$ . Sectors of the graph correspond to sampling years. Boxed samples are singletons.

ciently similar to be interconnected by an a posteriori link. After the fall overturn and during the 1977–1978 winter period, the community moves away from the previous summer's status. When spring comes (sample 35), the community has reached a zone of the multidimensional scaling graph (fig. 6) quite distinct from the one it occupied in summer 1977. The faunal composition is now completely dominated by rotifers, which increased from 70% to 87% in numbers and from 18% to 23% in biomass between 1977 and 1978, with a corresponding decrease for the crustaceans, while the physical and chemical conditions are stabilizing (Pinel-Alloul et al. 1982). It is noticeable that the early [35–37] and late [42–44] steps in the zooplankton succession of the summer 1978, identified by clustering with  $\alpha = 0.25$ , are interconnected by an a posteriori link. When the succession is interrupted by the overturn, in the fall of 1978, the last group [45–47] is found in figure 6 near the position of the previous winter's samples [29–34], thus indicating that the following year's cycle might resemble the 1978 succession.

This was an example of primary succession (the area was not previously occupied by a lake) with large between-year variations, which is very different from that of example 1. In Reservoir Desaulniers, the community was evolving rapidly from a river to a lake condition, with none of the repetitions of previous community structures that can be seen in the folds of figure 4 (chaetognaths).

## DISCUSSION AND CONCLUSIONS

### *A. Comparison of Methods*

The methods specifically designed for the identification of cycles in time or space data series, such as periodograms, correlograms, or the various types of spectral analysis, are not relevant to the study of succession, since we do not hypothesize any constant periodicity, other than trivial (year), in a succession of species. Other methods have been used in the past for this same kind of study, however. They can be grouped under three headings.

1. Chernoff (1973) has proposed to use drawings of faces to delimit homogeneous time intervals in sediment cores. The core is subsampled at equal depth intervals, thus forming a time series of discrete samples. Each sample is analyzed for a given set of descriptors, and a face is drawn to represent it. In this face, each characteristic (nose, mouth, and so on) corresponds to one of the descriptors and varies in size or position according to the value of that descriptor in succeeding samples. Our experience is that although faces may be an excellent method to summarize and represent the results of a chronological clustering, faces alone do not lead to the same breakpoints for different observers. More important, divergences are produced when a given descriptor is coded by two ecologists into different face characteristics.

2. Ordinations in reduced space have been used by several authors to represent multispecies time series. In particular, various methods of factor analysis have been used by Williams et al. (1969: vegetation, principal coordinates), by Levings (1975: benthos, principal coordinates); by Allen et al. (1977: phytoplankton, principal components of the first difference of species' weekly occurrences); and by Dessier and Laurec (1978: zooplankton, principal components and correspondence analysis); Sprules (1980) used nonmetric multidimensional scaling on zooplankton data. Reduced-space ordinations assume the independence of the samples, in order to produce an ellipsoid-shaped dispersion of points. Consequently, these authors expect their samples to form some sort of a circle in the plane of the first two axes, since successive samples are more likely to fall close to one another in the reduced-space ordination (Appendix D). Departures from a regular distribution in a circular pattern are then interpreted as evidences of subsets in the data series. Such subsets may indeed be observed in these projections (although not necessarily). A more or less circular pattern is found in the ordinations published by Levings (1975), Allen et al. (1977), and Dessier and Laurec (1978), mentioned above, who investigated 1-yr successional cycles. The "circles" are not closed in Sprules' 7-mo study, nor in the pioneering succession study of Williams et al. (12 samples gathered over 7 yr).



When used alone, these methods are left with two major drawbacks. (*a*) The distances to those objects which are, in some particular way, different from the main between-group variability are likely to be expressed by some minor principal axis, orthogonal to the main projection plane. Consequently, these objects may well be projected, in the reduced-space ordination, within a group from which they are quite distant. Moreover, we have noticed that the "circle" of samples may be deformed in a spoon-shape, so that in one part of the two-dimensional ordination graph, groups that are distinct in a third or higher dimension may well be packed together. These problems are common to all methods of ordination when used alone for a purpose of clustering, but much less so in the case of nonmetric multidimensional scaling. This is the reason why in numerical taxonomy, several authors including Gower and Ross (1969), Rohlf (1970), Schnell (1970), and Jackson and Crovello (1971) have proposed, independently, to take advantage of both the clustering and the ordination methods, associating them on the same graph, as did Allen et al. (1977) (with hierarchical complete linkage clustering) for the study of phytoplankton succession in Lake Wingra. See also our figure 6. (*b*) The second drawback is the lack of criteria for assigning samples to groups from an ordination diagram alone, so that the published diagrams often seem arbitrary. Rather than relying on unwritten and changing rules, one may prefer to state the rules clearly. Such a set of rules of sample assignment to groups is called a clustering algorithm.

With our data, the ordination diagram of figure 6 (nonmetric multidimensional scaling) is helpful as a complement to the clustering results, because we deal with a primary succession, with very little recurrence of community structure in the segment under study. Such was not the case, however, with the data of example 1. Although several ordination methods (not published here) were tried on these data, the results were always vague and not useful for the interpretation. The reason becomes clear when looking at figure 4: with a series longer than 1 yr in a stabilized temperate-climate environment, and when the series shows such complex between-group relationships, it is unlikely that the time sequence of sampling would help to recognize homogeneous groups of samples in the reduced space, without knowing beforehand the clustering structure and the between-group links.

These remarks lead to the following recommendations: an ordination alone (preferably by nonmetric multidimensional scaling of a carefully chosen resemblance matrix) may be sufficient to uncover the structure of a small series of samples, representing either a short time period or a pioneering community loosely sampled over several years. Ordinations are also a helpful support for clustering results in 1-yr studies (under temperate-climate latitudes, where annual cycles are present), as well as in longer-lasting studies of pioneering communities, as long as the successional trend occupies more variance than the within-year differences. Finally, climax communities showing year-to-year recurrence (possibly identifiable as species associations) are best represented by chronological clustering, as in figure 4.

3. Hawkins and Merriam (1974) have proposed a method for segmenting a multivariate data series into homogeneous units, by minimizing variability within segments. This method was advocated by F. Ibanez (1984) for the study of

successional steps. A first problem with this method is that the user must determine the number of segments, using as an indication the increase in explained variation relative to the increase in the number of segments. In ecology, a second and more important problem is that strings of multiple zeros in the multivariate series (often the case with species abundance data) are likely to cause the formation of segments based on species absences. A third drawback comes from the fact that 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. There are, however, two similarities to our method: the multidimensional series is partitioned into homogeneous groups using an objective criterion, and this partitioning is made with a constraint of contiguity within the series, as was explained in the METHODS.

The results obtained by Hawkins and Merriam's partitioning algorithm are compared in table 2 with those obtained from chronological clustering, for the log-transformed Desaulniers zooplankton data. Comparisons of these 10 partitions using Rajski's metric (1961) or the symmetric uncertainty coefficient (Nie et al. 1975) show that in most instances a chronological clustering resembles most the Hawkins and Merriam's partition into the same number of groups, thus showing the correspondence between the two partitioning methods. Similar results were obtained with the Mediterranean zooplankton data. These symmetric uncertainty coefficients have values around 80%, which indicates some degree of difference between the two methods—but not larger, perhaps, than between two chronological clustering partitions obtained from different distance functions. Hawkins and Merriam's method does not, however, provide the ecologist with indications of when to stop the partitioning process. Furthermore, it is likely to produce cuts where singletons are identified by chronological clustering, as is the case in the results of Ibanez (1984); this author used Hawkins & Merriam's partitioning method on the same Mediterranean zooplankton data as in our example 1, and he obtained a break in the series before sample 35.

4. Unconstrained clustering has also been applied to our example data sets, with varying degrees of success. In order to make the results as closely comparable as possible, proportional-link linkage clustering was applied to the same distance matrices as chronological clustering, with the same value of connectedness. For the Desaulniers zooplankton data, which form a series in rapid primary succession showing few bridges between distant groups, the results are the following (representing the hierarchical structure by nested sets of parentheses, and basic groups in square brackets): (((([1–8, 32] [29–33])[9–11])[34, 45–47])[35–44]) ([12–16] [17–28])). The main groupings obtained by unconstrained hierarchical clustering do correspond to the three main regions of the multidimensional scaling graph (fig. 6), in this case, although some mistakes are produced in the time sequence. At finer levels, relationships become blurred, including the relationships [17–19]–[21–24] and [35–37]–[42–44] that had been identified by the a posteriori tests of the chronological clustering method (fig. 5).

The results of unconstrained clustering obtained from the Mediterranean zooplankton data are not as easy to represent, because the sampling sequence is often completely broken down. The longest sequences still recognizable are made of

TABLE 2

COMPARISON OF RESULTS FROM CHRONOLOGICAL CLUSTERING AND FROM HAWKINS AND MERRIAM'S SEGMENTATION, FOR THE LOG-TRANSFORMED DESAULNIERS ZOOPLANKTON ABUNDANCE DATA

SAMPLE No.	CHRONOLOGICAL CLUSTERING					HAWKINS & MERRIAM				
	$\alpha$ -LEVEL									
	.01	.05	.15	.20	.25					
	GROUP NUMBER					GROUP NUMBER				
	6	7	9	14	16	6	7	9	14	16
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	2	2	2	1	1	1	1	1
4	1	1	2	2	2	1	1	1	1	1
5	1	1	2	2	2	1	1	1	1	1
6	1	1	2	2	2	1	1	1	1	1
7	1	1	2	2	2	1	1	1	2	2
8	1	1	2	2	2	1	1	1	2	2
9	2	2	3	3	3	1	1	2	2	2
10	2	2	3	3	3	2	2	2	3	3
11	2	2	3	3	3	2	2	2	3	3
12	2	2	4	4	4	2	2	2	3	4
13	2	2	4	4	4	2	2	3	4	4
14	2	2	4	4	4	2	2	3	4	4
15	3	3	5	5	5	2	2	3	4	5
16	3	3	5	5	5	2	2	3	4	5
17	3	3	6	6	6	3	3	4	5	6
18	3	3	6	6	6	3	3	4	5	6
19	3	3	6	6	6	3	3	4	5	6
20	4	4	7	7	7	3	3	4	6	7
21	4	4	7	7	7	3	3	4	6	7
22	4	4	7	8	8	3	3	5	6	7
23	4	4	7	8	8	3	3	5	7	8
24	4	4	7	8	8	3	3	5	7	8
25	4	4	7	9	9	3	3	5	7	8
26	4	4	7	9	10	3	3	5	7	8
27	4	4	7	9	10	4	4	6	8	9
28	4	4	7	9	10	4	4	6	8	9
29	5	5	8	10	11	4	4	6	9	10
30	5	5	8	10	11	4	4	6	9	10
31	5	5	8	10	11	4	5	7	10	11
32	5	5	8	10	11	5	5	7	10	11
33	5	5	8	10	11	5	5	7	10	11
34	5	5	8	10	11	5	5	7	11	12
35	6	6	9	11	12	5	5	7	11	12
36	6	6	9	11	12	5	6	8	12	13
37	6	6	9	11	12	6	6	8	12	13
38	6	6	9	11	13	6	6	8	12	14
39	6	6	9	12	14	6	6	8	13	14
40	6	6	9	12	14	6	6	8	13	15
41	6	6	9	12	14	6	6	8	13	15
42	6	7	9	13	15	6	6	8	13	15
43	6	7	9	13	15	6	7	9	14	16
44	6	7	9	13	15	6	7	9	14	16
45	6	7	9	14	16	6	7	9	14	16
46	6	7	9	14	16	6	7	9	14	16
47	6	7	9	14	16	6	7	9	14	16

NOTE.—Left: group numbers obtained in fig. 5 for various  $\alpha$ -levels. Right: partitions obtained with the same numbers of groups. Horizontal lines represent group breaks.

five samples: they are [7–11] and [24–28]. This is due in part to the between-group relationships illustrated in figure 6, and also to the fact that single samples from a group often manage to cluster with neighboring groups. The high degree of sampling variability found in ecological data makes it desirable to cluster with a constraint of time contiguity, especially when recurrence in community structure is present, as can be identified by between-group *a posteriori* testing.

### *B. Ecological Significance*

“The analysis of succession in complex communities typically takes the form of a bulky statistical or verbal description” (Lewis 1978, p. 401). The method expounded here is a contribution to the simplification of graphical results, as well as to the mathematical modeling of successional phenomena. It is based on the assumption that succession is discontinuous, at least in many types of communities. This concept is well documented in the literature (Allen et al. 1977; Lewis 1978, and others). On the other hand, the method takes advantage of the major tools available for the analysis of multidimensional data, namely clustering and ordination, which had been tried with varying success in the study of successions. The addition of a time constraint and of an exact statistical test of cluster fusion produces an algorithm that models (mimics) efficiently the thought processes of ecologists who are looking for successional steps and breaks. The algorithm itself is summarized in the last section of the METHODS and in figure 2. It does not require the data in the series to be equidistant in time.

The type of results to be obtained by this method depends to a large extent on the choice of an appropriate measure of the affinity or resemblance between samples. This choice may be based on the nature of the data (binary or quantitative) as well as on known mathematical properties of the coefficients available in the literature. A final choice between results obtained from seemingly equivalent coefficients could be made using independent criteria (such as the hydrological events in the examples above), or other dependent indicators of successional events, such as the Jassby-Goldman succession-rate index (Jassby and Goldman 1974), the summed-difference succession-rate index (Lewis 1978), or Goodall's (1966) deviant index. Obviously, the use of different but equally desirable indices may bring complementary information about the data series.

The algorithm proposed for chronological clustering requires the user to set two parameters at the start of the computer run: the type of agglomerative clustering (set by the value of connectedness of the intermediate linkage clustering, in our own program), and the probability level ( $\alpha$ ) of the cluster-fusion criterion. It was found that the connectedness parameter did not have much influence on the results. One should simply stay away from single linkage which seems undesirable in succession studies. On the other hand, complementary information was derived from various  $\alpha$  levels of clustering. The smallest  $\alpha$  values show where the most important breakpoints in the data series lie, while larger  $\alpha$  levels bring out the finer structure of the successional events.

Increasing  $\alpha$  acts as a probe into the finer details of the series, so that the simultaneous use of several partitions obtained with various  $\alpha$  values is recom-

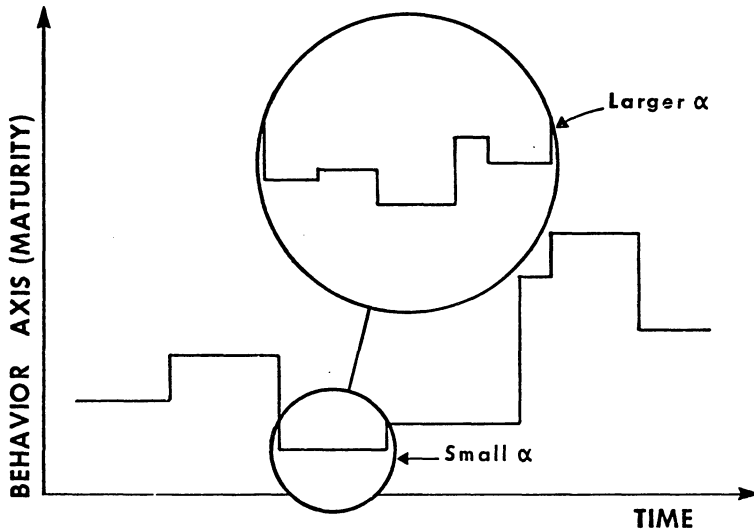


FIG. 7.—Ecological model: biological community succession proceeds in steps, which in turn may contain a series of smaller steps. Larger  $\alpha$ -values of the chronological clustering method make it possible to bring out these finer steps.

mended. The implicit model is that of figure 7, where main succession steps can be divided into smaller steps when the appropriate investigation tool is applied (larger  $\alpha$ ), up to the limit set by the sampling interval. This model can be applied to series of events in any system where periods of stability are followed by periods of instability (it is equivalent to Gould and Eldredge's [1977] model of *punctuated equilibria*, applicable to paleontological series), when the sampling interval and the length of the data series make it possible to identify two or more of these periods of relative stability. When extended in time, this model has the advantage of showing succession as an ever-evolving process, a given series leading to a climax that is itself only a step in a succession operating on a much longer time scale. A classical example in this respect is a lake filling with silt, that changes gradually from a deep to a shallow lake, then to a marsh, and beyond this to a dry-land forest (Whittaker 1970), with many stable states being reached and broken along this path.

We have shown furthermore that the proposed algorithm makes it possible to apply this model to real communities exhibiting very different types of successional phenomena. In the first example, the series exhibited within-year succession coupled with recurrence between years, as can be expected from the Mediterranean marine environment. On the other hand, the data from Reservoir Desaulniers show both within-year and between-year succession after impoundment, with very little recurrence of community structure in the first 2 yr of its existence. The method of chronological clustering performed equally well in these two situations, although the data imposed the choice of different distance mea-

tures. Chronological clustering has recently been applied to study the successional dynamics of bacteria in sewage treatment lagoons (Legendre et al. 1984). Work now in progress includes applications of the method to sediment cores, and to horizontal transects of community samples across a discontinuous environment, both representing spatial series.

#### SUMMARY

A clustering method is presented to describe the discontinuities in a multivariate (multispecies) series of biological samples, obtained from a single station at successive times. The method takes into account the sequence of sampling (time contiguity constraint) and makes it possible to eliminate singletons. Such singletons can be found in most ecological series, due to random components or to external forcings such as a temporary shift of water masses or immigration and emigration at a fixed station.

The clustering proceeds from a sample  $\times$  sample association matrix, built with an appropriately chosen similarity or distance coefficient. Agglomerative clustering is applied with the time constraint, and a randomization test is performed to verify whether the fusion is valid. This test compares the number of "high" distances in the between-group matrix to that in the fusion matrix of the two groups tested. When a singleton is discovered, with this same test, it is temporarily removed from the study and the clustering is started again from the beginning. This procedure is used because a singleton can well disturb the whole clustering geometry, as a result of the contiguity constraint.

The end result is a nonhierarchical partition of the series into nonoverlapping homogeneous groups, which are the steps of the ecological succession. Interval graph tests and a posteriori tests help to understand the relationships between the groups, as well as the origin of the singletons.

Examples are presented of a 78-wk Mediterranean zooplankton (chaetognaths) series and also of a 28-mo series of the zooplankton community in a freshwater reservoir. The results are compared to those obtained with several other methods of analyzing succession data. In the Appendixes we discuss various mathematical aspects of the work.

#### ACKNOWLEDGMENTS

We wish to express our thanks to J.-P. Croteau et A. Vaudor who participated in the birth of the statistical test of the between-group distance matrix, during an informal seminar on chronological clustering held near Quebec City in October 1980. The presence of S. Dallot in Quebec at this time had been sponsored by the France-Quebec exchange program between GIROQ and Université Pierre et Marie Curie (Paris VI). The criticism of the test was led by Y. Lepage, statistician at the Université de Montréal, with the help of U. Maag and of C. Tremblay. We are indebted also to J. Wong, University of Toronto, who suggested the expansion tests for the interval graphs. All the programming aspects of the development of the chronological clustering method have been handled by A. Vaudor. The com-

puter time was provided by the Service de l'informatique de l'Université du Québec à Montréal (former affiliation of P. Legendre and A. Vaudor) and by the Centre de calcul de l'Université de Montréal. This work was also supported by NSERCC grants to P. Legendre and to L. Legendre. The data of example 2 were made available by B. Pinel-Alloul, Université de Montréal, and by the Société d'énergie de la Baie James, Montreal. Our thanks also to Philippe Ross for his critical review of the manuscript.

## APPENDIX A

### COMPUTER PROGRAM

An efficient computer program for the chronological clustering method has been written in the PASCAL language, for a CDC CYBER computer, by Alain Vaudor, computer analyst, Département de sciences biologiques, Université de Montréal. Program listings and documentation are available free of charge from P. Legendre. On our Control Data CYBER 173 operating under NOS/BE, it took 228 CP seconds to run the 78 samples of example 1, and 28 CP seconds to run the 47 samples of example 2. Only 10% to 20% of this time was taken by the clustering procedure itself, the rest being used by the group expansion and the a posteriori tests.

## APPENDIX B

### THE CHOICE OF A DISTANCE MEASURE

Choosing among the many indices of similarity or distance available in the literature is a critical task, since different measures of resemblance may lead to different clustering results. Elements of a rationale to guide this choice have been published recently by Legendre and Legendre (1983, 1984), Bloom (1981), Hajdu (1981), and Wolda (1981), while measures of resemblance have recently been reviewed by Gower (1984) and by Legendre and Legendre (1983, 1984). Another aspect of the decision process is developed here.

Let us assume that the raw species-abundance distributions have been made normal by some appropriate transformation. The chronological clustering method itself does not require normality of the data. Since normalized variables are not skewed, however (since skewness is the result of populations occasionally showing exponential growth within the community), differences in normalized abundances are more likely to have the same ecological meaning throughout the range of the abundance scale. Furthermore, normality gives access to a larger number of distance functions that are well adapted to the problem under study.

Among the coefficients of resemblance that are adapted to species-abundance data (coefficients for quantitative measurements, excluding double absences from the comparison), three main types may be recognized.

*Type 1.*—All resemblance coefficients compare the abundances found in two samples, for each of the species present in either one. Let us assume that a given difference is found, between the 2 samples, for an abundant species and also for a rare species. In the coefficients of the first type, a difference found between 2 samples for an abundant species contributes the same to the distance as the same difference found between these samples for rarer species. Whittaker's *index of association* (1952) pertains to this group. It is a sum of differences of percentages between 2 samples, 1 and 2:

$$D_w = \frac{1}{2} \sum_{i=1}^n \left| \frac{y_{i1}}{\sum_{i=1}^n y_{i1}} - \frac{y_{i2}}{\sum_{i=1}^n y_{i2}} \right|$$

where  $n$  is the number of species. From the following simple example:

Species:	1	2
sample 1	90	10
sample 2	80	20

it is easy to verify that the difference of 10 for the abundant species 1 contributes 0.1 to the sum, while the same difference for the rare species 2 also contributes 0.1. A similar result would be obtained with unequal sample sizes. Steinhaus' coefficient, whose 1-complement ( $1 - \text{coefficient}$ ) is also known as Odum's (1950) *percentage difference* or as the coefficient of Bray and Curtis (1957), pertains to this same type.

*Type 2a.*—In coefficients of this type, a difference found between 2 samples for an abundant species contributes less to the distance (more to the similarity) than the same difference found between these samples for rarer species. The *Canberra metric* (Lance and Williams 1967) between two samples, 1 and 2, belongs to this group:

$$D_c = \sum_{i=1}^n \frac{|y_{i1} - y_{i2}|}{(y_{i1} + y_{i2})}$$

where  $n$  is the number of species. Related forms, such as Clark's coefficient of divergence (1952), are found in the literature. Computing the simple example above gives  $(10/170 + 10/30)$ : the abundant species 1 contributes  $10/170$  to the distance, while the rare species 2 contributes  $10/30$ .

*Type 2b.*—In this type are found the distances for which an equal difference receives a weight inversely proportional to the variability of the species in the whole set of samples under study. This is the case with the 1-complement of Gower's (1971) similarity coefficient, where the measure of variability is the range of variation of each species in all samples. The 1-complement of the similarity of Legendre and Chodorowski (1977) also pertains to this group since the spread parameter  $k$  of the partial similarity function imbedded in this coefficient is usually made proportional to each species' variability. In this respect, coefficients of type 2b behave much as those of type 2a. The difference resides in the fact that with type-2a coefficients, the weight of a species varies depending upon the pair of samples being compared, while with type 2b, the variability of a species (and its weight) is the same for all pairs of sample comparisons.

A decision about the choice of the type of coefficient most appropriate to a particular case can be made by adding ecological criteria to these mathematical considerations. For a biological sample, a vector of species abundances can be written, from which one may plot a species rank-frequency diagram for this sample. The shape of this diagram indicates the degree of maturity of the community, as explained by Frontier (1976) and by Legendre (1983). The argument is based on the fact that rare species are generally not well sampled, and consequently the comparison of dominant or intermediate-density species is more meaningful. In immature communities, most of the species are rare while only a few are well sampled. In such communities, one may prefer to give more weight to the few well-sampled species in assessing the similarity, which can be done by using type-2 coefficients. On the other hand, more of the species in mature ecosystems are of high or intermediate density, so that type-1 coefficients may be adequate to describe the resemblance relationships.

## APPENDIX C

### THE RANDOMIZATION TEST OF CLUSTER FUSION

Following the METHODS section, let us assume that a symmetric distance matrix has been computed between all pairs of samples, ordered according to the time of sampling (as in figs. 1, C1, C2). *Distance* is used here as synonymous with *dissimilarity*. Let us suppose



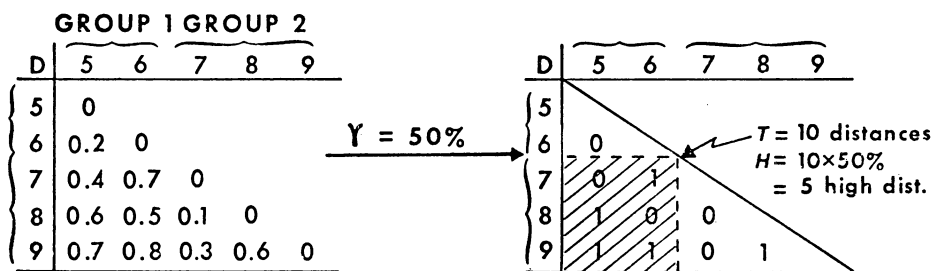


FIG. C1.—Application of a contrast value  $\gamma = 50\%$  to a hypothetical distance matrix, made of 2 contiguous groups of chronological samples. The 2 groups are candidates for fusion.

further that the intermediate linkage algorithm (fig. 2, step 3) points to the two contiguous groups of figure C1 as the next candidates for clustering. The problem to be tested may be formulated as follows. If the 2 groups represent in fact a single cluster, then the presence of 2 groups is merely an artifact of the hierarchical agglomerative algorithm of intermediate linkage. In this case, the highest distances found in this distance matrix should be distributed at random in the joint triangular distance matrix of these two groups (with  $T = 10$  distances in fig. C1) and not only in the between-group matrix (shaded, with  $B = 6$  distances, in fig. C1). This is the null hypothesis ( $H_0$ ). The alternative hypothesis ( $H_1$ ) is that, if these two groups do pertain to different successional steps, significantly more high distances will be found in the between-group matrix than in the fusion matrix of these two groups. Is the actual distribution of the highest distances far enough from random to reject this null hypothesis ( $H_0$ ) of random distribution?

It would be possible to formulate the fusion criterion (below) in terms of the distribution of actual distances. We felt, however, that it would be more appropriate to the present problem to look at distances as being simply high or low. Indeed, given the importance of random components in the sampling itself (for instance, 1 liter from a lake, or a sea . . .), we cannot attach undue importance to small variations in the distance between 2 samples, so that the result of the chronological clustering may be expected to be more robust if the statistical criterion is very contrasting. On the other hand, if actual distances were used, the fusion criterion would be formulated in terms of "sums of distances" rather than "numbers of high distances," with 2 drawbacks. First, 2 distance coefficients that behave monotonically but not linearly could lead to different results, which renders even more delicate the choice of the resemblance coefficient to be used. Second, if a distance coefficient does not have the same efficiency at measuring differences over its whole range (if the difference between 0.1 and 0.2 does not mean the same thing as the difference between distance 0.9 and distance 1.0) then the criterion should include a weighting by the magnitude of the distances to be tested. This phenomenon is extremely difficult to evaluate, for lack of a unique reference, so that we feel that there would be no way of preserving the same power for the criterion throughout the full range of distances.

For these reasons, we looked for a way of dividing the  $T$  distances of the triangular matrix joining 2 groups into 2 subsets. These 2 subsets are called the high and the low distances, so defined that the statistical test would make sense for between-group matrices of any size, or shape (square or elongated). The criterion we chose is a contrast parameter, called  $\gamma$  (gamma), determining what proportion of the  $T$  distances are to be called high distances. The number of high distances among the  $T$  distances is called  $H$ , computed by:

$$H = \lfloor \gamma T \rfloor$$

where  $\lfloor \gamma T \rfloor$  means the largest integer contained in  $\gamma T$ . In figure C1, for instance, where there are  $T = 10$  distances, setting  $\gamma$  to 50% means that  $H = 10 \times 50\% = 5$  high distances. Then we identify the 5 highest distances, in the triangular matrix, and replace them by

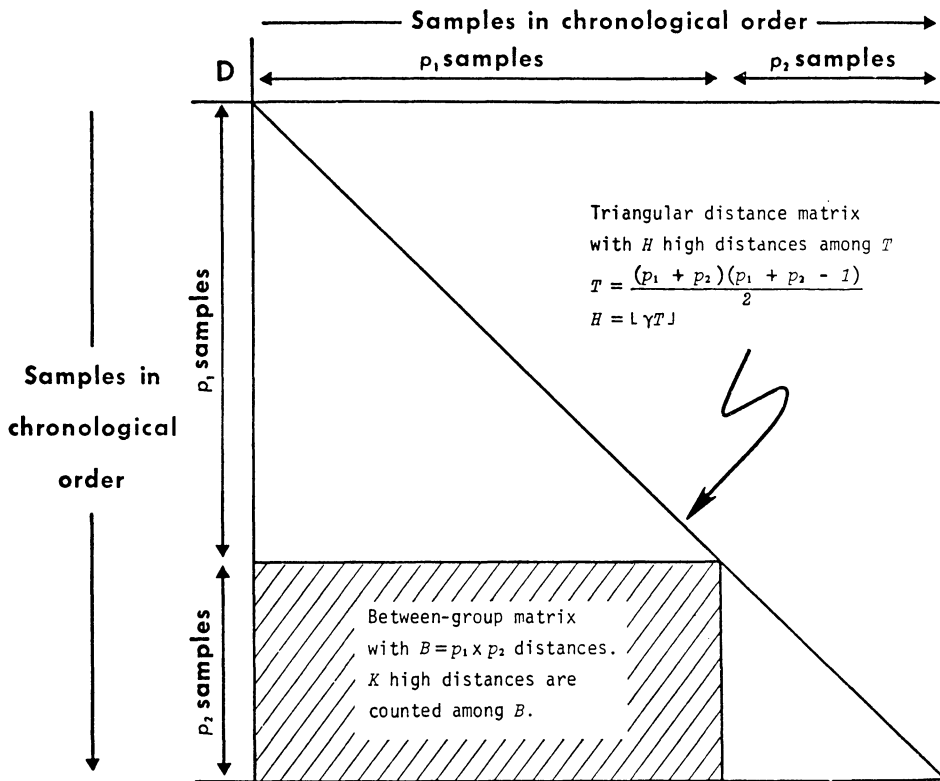


FIG. C2.—Definition of the parameters  $T$ ,  $H$ ,  $B$ , and  $K$  used in the test of fusion of 2 contiguous groups.

ones. All the  $T-H$  other (low) distances are replaced by zeros. Finally, we look at the between-group matrix containing  $B$  distances, and count how many ( $K$ ) among them are high distances, or ones (fig. C2).  $K$  is then a random variable, which opens the way to a test of statistical significance. For the example in figure C1, there are  $K = 4$  high distances among the  $B = 6$ . The fusion of pairs of samples at the beginning of the agglomerative procedure is performed automatically, even though this is the limit case where the test collapses,  $H$  being equal to 0.

The random variable  $K$ , computed after setting  $\gamma$  to a given value (see below), is then tested for statistical significance using an exact test.

#### Computing Procedure

A step-by-step exposition of the procedure will make it clear.

1. The null hypothesis  $H_0$  to be tested is the following: the between-group matrix contains the same proportion of high distances as the triangular matrix, allowing for statistical variation. The alternative hypothesis ( $H_1$ ) is that significantly more high distances are found in the between-group matrix than in the fusion matrix of these two groups. In other words,

$$H_0: (K/B) = (H/T) \quad H_1: (K/B) > (H/T).$$

It should be noted that the test is one-tailed. Indeed, if  $(K/B) < (H/T)$ , this is no reason to reject the hypothesis of clustering homogeneity.

2. The test is based upon the following question: for the given triangular matrix containing  $T$  distances, what is the probability of finding a between-group matrix of  $B$  distances containing at least the same number  $K$  of high distances?

3. Set a probability level  $\alpha$  for rejecting  $H_0$  (ex.,  $\alpha = .05, .10, .25, \dots$ ).

4. Look into the given triangular matrix of  $T$  distances, among which  $H = [\gamma T]$  are high. An exact reference distribution may be obtained by randomization. This method, from R. A. Fisher, is expounded in Box et al. (1978) and in Sokal and Rohlf (1981), among others. The procedure in the present case goes as follows. (a)  $H_0$ : there is no difference between the 2 groups. (b) Accordingly, any one observation could have been a member of either of the two groups. (c) Change the labels of the objects in a random manner; what value of the test statistic ( $K$ ) is obtained? (d) Make all the possible groupings of these objects into 2 groups of the same sizes as the groups tested, in order to obtain the sampling distribution of the statistic  $K$  for this particular pair of matrices. There are  $N$  possible rearrangements of these objects into 2 groups.

5. Count how many ( $M$ ) of these ( $N$ ) have a number of high distances larger than or equal to  $K$ .

6. If  $(M/N) > \alpha$ ,  $H_0$  is accepted and so is the fusion of the 2 groups under consideration. On the other hand, if  $(M/N) \leq \alpha$ ,  $H_0$  must be rejected at significance level  $\alpha$  and we turn to the alternative hypothesis that these 2 groups should not cluster for reason of a significant difference in the distribution of the high distances. The meaning of the  $\alpha$ -level is further discussed in a separate section below. In both cases, the algorithm turns to the intermediate linkage clustering procedure in order to find out which pair of groups is to be tested next.

Looking at all the possible between-group matrices is relatively easy with small matrices (for instance, for 2 groups of 2 objects each, there are only 6 possible between-group matrices of size  $2 \times 2$ ). This is not the case with larger matrices, however. For instance, with 2 groups of 10 objects each, there are 184,756 possible between-group matrices of size  $10 \times 10$ . In such cases, it becomes impractical to examine them all, so that our computer program (see Appendix A) proceeds by a sampled randomization test, as follows: if there are no more than 10,000 possible between-group matrices, then we examine all of them. If there are more, then we sample at random and examine 1,000 of them. A comparison between the complete and the sampling methods, in cases ranging from 18,564 to 352,716 possible between-group matrices, showed that the probability obtained was the same, at least up to the third decimal place.

#### *Selection of an Optimal Value of Contrast $\gamma$*

It is obvious that the test will gain in power if there are enough, but not too many, high distances. Indeed, if there are too many or too few, it might be difficult to decide whether the between-group matrix is different from the triangular matrix, as illustrated in figure C3. The impression left by this figure was formalized by computing the probability of  $H_0$  under a whole range of values for the contrast parameter  $\gamma$ , using some real matrices where  $H_0$  was obviously to be rejected. These computations showed that  $H_0$  is more efficiently rejected (small probability) around the middle of the range of  $\gamma$ . For this reason,  $\gamma$  was set at 50%. This places the breakpoint between high and low distances at the median of the array of distances.

#### *Meaning of the $\alpha$ -Level*

One might argue that testing only between groups that have already been formed by clustering increases the chances of finding a significant difference (biased test). This problem can be answered both theoretically, and practically through simulation.

The theoretical justification is 3-fold:

1. Since the two groups have been obtained after clustering with the contiguity con-

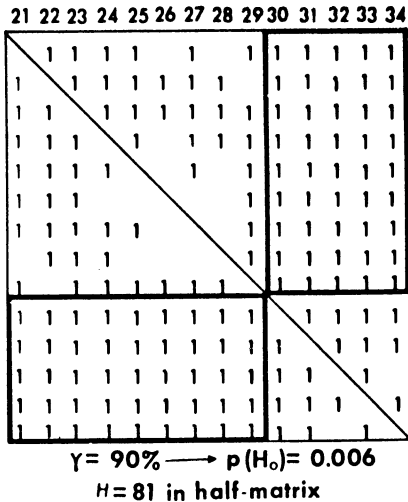
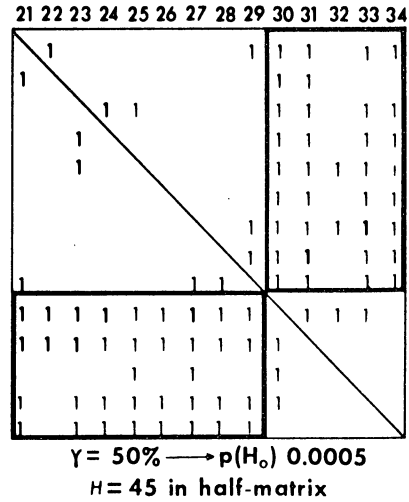
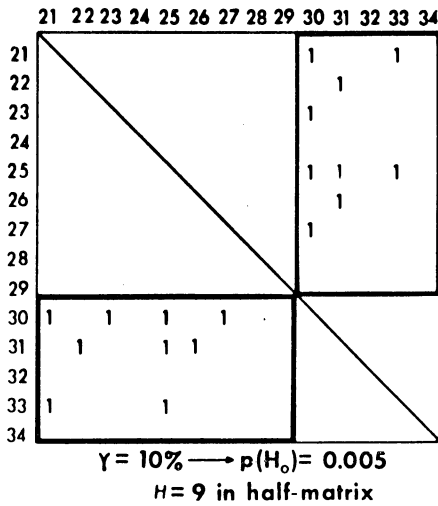


FIG. C3.—A real case where the fusion of some groups [21 to 29] and [30 to 34] is to be tested. The 1's represent the high distances for  $\gamma$  values of 10%, 50%, and 90%, respectively. When  $\gamma$  is set at 50%, it is easier to make the obvious decision that these 2 groups should not cluster, most of the high distances being located in the between-group matrix. The exact test indicates that the probability of  $H_0$  is also smaller when  $\gamma = 50\%$ .  $T = 91$  in the half-matrix.

straint, they are unlikely to be optimal. This consideration lessens the problem without eliminating it totally.

2. Actually, one would like to test all the breakpoints of all possible pairs of groups. It is only because this cannot be done (for computing time reasons), on real data sets, that only the breakpoints singled out by the agglomerative procedure are tested. Even though the agglomerative structure is, in and of itself, only a computational artifact, one may suppose that if the breakpoint that has been singled out by the agglomerative clustering algorithm (among all the possible breakpoints that should be tested) is nonsignificant, then the null hypothesis of ecological stability for the 2 groups combined should verify itself at any of the other possible breakpoints within the combination of these 2 groups.

3. We only seek a fusion criterion, even though this criterion does not have all the properties of a statistical test. As Jambu (1978) puts it, tests by simulation are helpful to

provide stopping criteria during clustering, since statistical tests are ill adapted to the task. Perruchet (1983) has reviewed the various types of significance tests for clustering.

A Monte Carlo simulation was run in order to assess the true probability of a Type I error for various  $\alpha$  values used in the computation algorithm (fig. 2). In order to simulate the situation described above as the null hypothesis, 30 objects were positioned at random in a 30-dimensional unit cube, for each simulation. The "chronological order of sampling" was set to be the order of the objects in the random list. The Euclidean distance was computed for all pairs of objects, and the chronological clustering program was run after setting  $\alpha$  to a given value. Since there is no structure other than random in the objects, the probability of finding breaks in the data set should be equal to  $\alpha$ . The following statistic was calculated for each simulation run:

$$\text{RATIO} = \frac{\text{How many times } H_0 \text{ is rejected}}{\text{How many tests are performed}}.$$

Thirty such data sets were generated, and each one was run with an  $\alpha$ -value (called ALPHA hereafter) of 2%, 25%, and 40%. A regression of RATIO as a function of ALPHA, for these 90 experimental points, gave the following linear equation:

$$\text{RATIO} = -0.00924 + 1.01481 \text{ ALPHA}.$$

The 95% confidence interval was  $[-0.0603, 0.0418]$  for the intercept and  $[0.8299, 1.1997]$  for the slope. This result confirms the expectation that ALPHA equals RATIO in situations corresponding to the null hypothesis and supports the theoretical arguments of the previous paragraph.

#### *The $\alpha$ -Level in A Posteriori Tests*

One should be careful when performing a posteriori tests on singletons. One should not simply take a probability value higher than  $\alpha$  as an indication of relationship, without due

TABLE C1  
LOWEST POSSIBLE PROBABILITY OF FUSION ( $p_{\min}$ ) OF TWO GROUPS  
WITH  $p_1$  AND  $p_2$  OBJECTS, RESPECTIVELY

$p_1$	$p_2$				
	1	2	3	4	5
2 .....	.66667	.33333			
3 .....	.25000	.10000	.10000		
4 .....	.20000	.06667	.02857	.02857	
5 .....	.16667	.04762	.01786	.00794	.00794
6 .....	.14286	.03571	.01190	.00476	.00217
7 .....	.12500	.02778	.00833	.00303	.00126
8 .....	.11111	.02222	.00666	.00202	.00078
9 .....	.10000	.01818	.00455	.00140	.00050
10 .....	.09091	.01515	.00350	.00100	.00033
11 .....	.08333	.01282	.00275	.00073	.00023
12 .....	.07692	.01099	.00220	.00055	.00016
13 .....	.07143	.00952	.00179	.00042	.00012
14 .....	.06667	.00833	.00147	.00033	.00009
15 .....	.06250	.00735	.00123	.00026	.00006
16 .....	.05882	.00654	.00103	.00021	.00005
17 .....	.05556	.00585	.00088	.00017	.00004
18 .....	.05263	.00526	.00075	.00014	.00003
19 .....	.05000	.00476	.00065	.00011	.00002
20 .....	.04762	.00433	.00056	.00009	.00002

consideration of the number of objects in these groups. Indeed, simple computations with the randomization test show that the probability of a singleton fusing, for example, with a three-sample group can never be less than 0.25, when  $\gamma$  is fixed at 0.50. In general, it can be shown that the lowest probability that can be taken by the randomization test, for the fusion of two groups with  $p_1$  and  $p_2$  objects, respectively, is (when  $\gamma = 0.50$ )

$$p_{min} = \frac{p_1! p_2!}{(p_1 + p_2)!}$$

except when  $p_1$  equals  $p_2$ , or when  $p_1 = 2$  and  $p_2 = 1$ , in which cases this formula must be doubled. When  $p_1 = p_2 = 1$ , the test cannot be performed because there is no high distance. Some values are presented in table C1. Thus, one could base a correct interpretation of the singletons on a value  $\Delta p$ , taken as the difference between the computed probability of fusion, minus the lowest possible probability, shown in column " $p_2 = 1$ " of table C1. This table also shows that this effect becomes less important as the number of objects in the groups increases.

## APPENDIX D

### A NULL-HYPOTHESIS MODEL FOR SUCCESSION BY STEPS

Although the authors (mentioned in the main text) who used reduced-space ordination methods to study annual successions did assume that they would get a more or less circular distribution of sample points in reduced space, this assumption has never been shown to be

TABLE D1  
DATA TABLE REPRESENTING A SMOOTH SPECIES SUCCESSION (10 species  $\times$  10 samples)

SPECIES	SAMPLES				
	1	2	3	4	5
1	sin(36°)	sin(72°)	sin(108°)	sin(144°)	sin(180°)
2	sin(72°)	sin(108°)	sin(144°)	sin(180°)	sin(216°)
3	sin(108°)	sin(144°)	sin(180°)	sin(216°)	sin(252°)
4	sin(144°)	sin(180°)	sin(216°)	sin(252°)	sin(288°)
5	sin(180°)	sin(216°)	sin(252°)	sin(288°)	sin(324°)
6	sin(216°)	sin(252°)	sin(288°)	sin(324°)	sin(360°)
7	sin(252°)	sin(288°)	sin(324°)	sin(360°)	sin(36°)
8	sin(288°)	sin(324°)	sin(360°)	sin(36°)	sin(72°)
9	sin(324°)	sin(360°)	sin(36°)	sin(72°)	sin(108°)
10	sin(360°)	sin(36°)	sin(72°)	sin(108°)	sin(144°)
SPECIES	6	7	8	9	10
	6	7	8	9	10
1	sin(216°)	sin(252°)	sin(288°)	sin(324°)	sin(360°)
2	sin(252°)	sin(288°)	sin(324°)	sin(360°)	sin(36°)
3	sin(288°)	sin(324°)	sin(360°)	sin(36°)	sin(72°)
4	sin(324°)	sin(360°)	sin(36°)	sin(72°)	sin(108°)
5	sin(360°)	sin(36°)	sin(72°)	sin(108°)	sin(144°)
6	sin(36°)	sin(72°)	sin(108°)	sin(144°)	sin(180°)
7	sin(72°)	sin(108°)	sin(144°)	sin(180°)	sin(216°)
8	sin(108°)	sin(144°)	sin(180°)	sin(216°)	sin(252°)
9	sin(144°)	sin(180°)	sin(216°)	sin(252°)	sin(288°)
10	sin(180°)	sin(216°)	sin(252°)	sin(288°)	sin(324°)

legitimate. We will show here that the underlying ecological model does indeed lead to such an arrangement of the points in reduced space.

The ecological model sought to represent a community evolving smoothly should have the following characteristics: (1) it should be multidimensional, in order to represent the interaction of many species; (2) each species should vary through time; (3) in order to form a structure without steps, each species' variation should be smooth, while the sum of all species abundances should be constant; (4) for the series of points to form a closed shape, each species should come back, at the end of the cycle, to a value close to its starting value; (5) the samples should be equi-spaced through time.

Table D1 represents one of the possible data sets obeying these conditions. It could be expanded at will to as many species or samples as necessary. Each species is represented by a sine function. The species are lagged by one-tenth of a circle, relative to the previous one, and they all have the same amplitude, in order to make the column sums equal.

A principal component analysis of table D1 produces only 2 eigenvalues, each accounting for 50% of the total variance. In the reduced space, the samples are positioned at equal distances along a perfect circle. This result was expected, since the data table is built in such a way that the Euclidean distances between successive samples are equal. This is to say that other data sets could be built to represent a smooth species succession. The conditions are: same distance between all neighboring points, this distance being smaller than that between points farther apart; same weight to all points, provided by equal column sums.

Accordingly, the circle can be used as a null hypothesis for the study of sharp transitions during succession, as was done by the authors mentioned in the *Comparison of Methods* section.

#### LITERATURE CITED

- Allen, T. F. H., S. M. Bartell, and J. F. Koonce. 1977. Multiple stable configurations in ordination of phytoplankton community change rates. *Ecology* 58:1076–1084.
- Bloom, S. A. 1981. Similarity indices in community studies: potential pitfalls. *Mar. Ecol. Prog. Ser.* 5:125–128.
- Bougis, P. 1968. Le problème des remontées d'eaux profondes à Villefranche-sur-Mer. *Cah. Océanogr.* 20:597–603.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. *Statistics for experimenters—an introduction to design, data analysis, and model building*. Wiley, New York.
- Bray, R. J., and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27:325–349.
- Chernoff, H. 1973. The use of faces to represent points in  $k$ -dimensional space graphically. *J. Am. Stat. Assoc.* 68:361–368.
- Clark, P. J. 1952. An extension of the coefficient of divergence for use with multiple characters. *Copeia* 1952:61–64.
- Dessier, A., and A. Laurec. 1978. Le cycle annuel du zooplancton à Pointe-Noire (RP Congo). Description mathématique. *Oceanol. Acta* 1:285–304.
- Frontier, S. 1976. Utilisation des diagrammes rang-fréquence pour l'analyse des écosystèmes. *J. Rech. Océanogr.* 1:35–48.
- . 1977. Réflexion pour une théorie des écosystèmes. *Bull. Ecol.* 8:445–464.
- Goodall, D. W. 1966. Deviant index: a new tool for numerical taxonomy. *Nature (Lond.)* 210:216.
- 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.
- Gould, S. J., and N. Eldredge. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3:115–151.
- Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27:857–871.

- . 1984. Measures of similarity, dissimilarity and distance. In S. Kotz, N. L. Johnson, and C. B. Read, eds. *Encyclopaedia of statistical sciences*. Vol. 5. Wiley, New York.
- Gower, J. C., and G. J. S. Ross. 1969. Minimum spanning trees and single linkage cluster analysis. *Appl. Stat.* 18:54–64.
- Gutierrez, L. T., and W. R. Fey. 1980. Ecosystem succession—a general hypothesis and a test model of a grassland. M.I.T. Press, Cambridge, Mass.
- Hajdu, L. J. 1981. Graphical comparison of resemblance measures in phytosociology. *Vegetatio* 48:47–59.
- Hawkins, D. M., and D. F. Merriam. 1974. Zonation of multivariate sequences of digitized geologic data. *J. Int. Assoc. Math. Geol.* 6:263–269.
- Ibanez, F. 1972. Un essai d'interprétation écologique par l'analyse spectrale conjuguée avec l'analyse des composantes principales. *Bull. Union Océanogr. France* 9:46–53.
- . 1974. Une cotation d'abondance réduite à trois classes: justification de son emploi en analyse des composantes principales. Mise en oeuvre et intérêt pratique en planctologie. *Ann. Inst. Océanogr.* 50:185–198.
- . 1984. Sur la segmentation des séries chronologiques planctoniques multivariées. *Oceanol. Acta* (in press).
- Ibanez, F., and S. Dallot. 1969. Étude du cycle annuel des chaetognathes planctoniques de la rade de Villefranche par la méthode d'analyse des composantes principales. *Mar. Biol.* 3:11–17.
- Jackson, R. C., and T. J. Crovello. 1971. A comparison of numerical and biosystematic studies in *Haplopappus*. *Brittonia* 23:54–70.
- Jacobs, R. P. W. M., and A. M. J. Bouwhuis. 1979. The year cycle of *Eudiaptomus vulgaris* (Schmeil, 1896) (Copepoda, Calanoida) in a small, acid water body during 1973. Development in the natural habitat and relationships between temperature and duration of development stages. *Hydrobiologia* 64:17–36.
- Jambu, M. 1978. Classification automatique pour l'analyse des données. Tome I. Méthodes et algorithmes. Dunod, Paris.
- Jassby, A. D., and C. R. Goldman. 1974. A quantitative measure of succession rate and its applications to the phytoplankton of lakes. *Am. Nat.* 108:688–693.
- Johnston, D. W., and E. P. Odum. 1956. Breeding bird populations in relation to plant succession on the Piedmont of Georgia. *Ecology* 37:50–62.
- Lance, G. N., and W. T. Williams. 1967. Mixed-data classificatory programs. I. Agglomerative systems. *Aust. Comput. J.* 1:15–20.
- Lebart, L. 1978. Programme d'agrégation avec contraintes (C.A.H. contiguïté). *Cah. Anal. Données* 3:275–287.
- Lefkovich, L. P. 1978. Cluster generation and grouping using mathematical programming. *Math. Biosci.* 41:91–110.
- . 1980. Conditional clustering. *Biometrics* 36:43–58.
- Legendre, L., and P. Legendre. 1983. Numerical ecology. Developments in environmental modelling. No. 3. Elsevier, The Hague.
- . 1984. *Ecologie numérique*. 2d ed. Tome 2. La structure des données écologiques. Collection d'écologie. No. 13. Masson, Paris, et les Presses de l'Université du Québec.
- Legendre, P. 1983. Numerical ecology: developments and recent trends. Pages 505–523 in J. Felsenstein, ed. *Numerical taxonomy: Proceedings of a NATO Advanced Study Institute*. NATO Advanced Study Institute Series G (Ecological Sciences). No. 1. Springer-Verlag, Berlin.
- 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., and A. Chodorowski. 1977. A generalization of Jaccard's association coefficient for Q analysis of multi-state ecological data matrices. *Ecol. Pol.* 25:297–308.
- Levings, C. D. 1975. Analyses of temporal variation in the structure of a shallow-water benthic community in Nova Scotia. *Int. Rev. Ges. Hydrobiol.* 60:449–470.
- Lewis, W. M., Jr. 1978. Analysis of succession in a tropical phytoplankton community and a new measure of succession rate. *Am. Nat.* 112:401–414.
- Lilliefors, H. W. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* 62:399–402.



- Margalef, R. 1968. Perspectives in ecological theory. University of Chicago Press, Chicago.
- . 1974. *Ecologia*. Ediciones Omega, Barcelona.
- Monestiez, P. 1978. Méthodes de classification automatique sous contraintes spatiales. Pages 367–379 in J. M. Legay and R. Tomassone. *Biométrie et écologie*. Société française de Biométrie. Paris.
- Nie, N. H., C. H. Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Bent. 1975. SPSS—statistical package for the social sciences. 2d ed. McGraw-Hill, New York.
- North, P. M. 1977. A novel clustering method for estimating numbers of bird territories. *Appl. Stat.* 26:149–155.
- Odum, E. P. 1950. Bird populations of the Highlands (North Carolina) plateau in relation to plant succession and avian invasion. *Ecology* 31:587–605.
- . 1971. *Fundamentals of ecology*. 3d ed. Saunders, Philadelphia.
- Orlóci, L. 1981. Probing time series vegetation data for evidence of succession. *Vegetatio* 46:31–35.
- Perruchet, C. 1983. Significance tests for clusters: overview and comments. Pages 199–208 in J. Felsenstein, ed. *Numerical taxonomy: Proceedings of a NATO Advanced Study Institute. NATO Advanced Study Institute Series G (Ecological Sciences)*. No. 1. Springer-Verlag, Berlin.
- Pinel-Aloul, B., E. Magnin, and G. Codin-Blumer. 1982. Effets de la mise en eau du réservoir Desaulniers (Territoire de la Baie de James) sur le zooplancton d'une rivière et d'une tourbière réticulée. *Hydrobiologia* 86:271–296.
- Rajski, C. 1961. Entropy and metric spaces. Pages 44–45 in C. Cherry, ed. *Information theory*. Butterworths, London.
- Roche, C. 1978. Exemple de classification hiérarchique avec contrainte de contiguïté. Le partage d'Aix-en-Provence en quartiers homogènes. *Cah. Anal. Données* 3:289–305.
- Rohlf, F. J. 1970. Adaptive hierarchical clustering schemes. *Syst. Zool.* 19:58–82.
- Schnell, G. D. 1970. A phenetic study of the suborder Lari (Aves). I. Methods and results of principal components analyses. *Syst. Zool.* 19:35–57.
- Sneath, P. H. A. 1966. A comparison of different clustering methods as applied to randomly-spaced points. *Classif. Soc. Bull.* 1:2–18.
- Sokal, R. R., and F. J. Rohlf. 1981. *Biometry—the principles and practice of statistics in biological research*. 2d ed. Freeman, San Francisco.
- Sprules, W. G. 1980. Nonmetric multidimensional scaling analyses of temporal variation in the structure of limnetic zooplankton communities. *Hydrobiologia* 69:139–146.
- Whittaker, R. H. 1952. A study of summer foliage insect communities in the Great Smoky Mountains. *Ecol. Monogr.* 22:1–44.
- . 1970. *Communities and ecosystems*. Macmillan, London.
- Williams, W. T., G. N. Lance, L. J. Webb, J. G. Tracey, and M. B. Dale. 1969. Studies in the numerical analysis of complex rain-forest communities. III. The analysis of successional data. *J. Ecol.* 57:515–535.
- Wolda, H. 1981. Similarity indices, sample size and diversity. *Oecologia (Berl.)* 50:296–302.