COENOSES 9(1):1-13, 1994

CLASSIFICATION WITH MULTIPLE DISSIMILARITY MATRICES

M.B. Dale[1] & P.T. Dale[2]

1 Present Address: 10/217 Hawken Dr., St. Lucia, Qld., 4067 Australia,
2 Dept. Computer Science, Univ. Queensland, St. Lucia, Qld., 4072 Australia.

Keywords: Constrained classification, Dissimilarity, Multiple-objective optimisation, Pareto optimality.

Abstract. Several ordination methods, including Principal Component analysis, Correspondence analysis and Multidimensional Scaling, are capable of analysing several matrices simultaneously, that is they can analyse three-mode data. In contrast, almost all classification methods which have been used in phytosociology have employed a single symmetric dissimilarity matrix. At first, this may not seem to pose any particular problem. However, on inspecting such single matrices more closely, we find that they are formed by uniting several matrices, often by summing contributions from several variables (species). Thus implicitly we have adopted a particular method of analysis without examining the alternatives. And there are other situations in which simultaneous analysis of several matrices might be regarded as potentially profitable.

There is one exceptional case in classification where multiple matrices have been recognised. Certain procedures utilise ancillary information to constrain the analysis, commonly by permitting only spatially adjacent stands to fuse, or by requiring that subgroups formed by division remain spatially coherent (e.g. Gordon & Birks 1972). Such constraint could equally be derived from any other variable. In such cases elements of one matrix dominate the elements of the other, so the analysis is markedly asymmetric in its treatment. Modifying the constraint matrix during the classification procedure gives a dynamic system. We may use several matrices in such constrained analyses, defining a logical function to combine the constraints, and modifying each matrix idiosyncratically. Furthermore, we could introduce weighting to replace complete prohibition, thus encouraging, say, adjacency without demanding it. However, asymmetry is not always desired, especially in cases where multiple matrices arise from reasons other than constraint satisfaction.

In this paper we first examine various situations where several matrices of dissimilarity might reasonably be generated simultaneously and where it does not seem entirely desirable that they be summed to form a single overall dissimilarity matrix. We shall then examine some approaches to the problem of classification using multiple dissimilarity matrices, with and without asymmetric treatment, and illustrate some of their potentialities and problems. Finally we indicate how our results reflect on some fundamentals of phytosociological investigation.

Introduction

Several ordination methods, such as Principal Component analysis and Multidimensional Scaling are capable of analysing several matrices simultaneously. In contrast, almost all classification methods which have been used in phytosociology have employed a single symmetric dissimilarity matrix. One major exception has been those programs which utilise adjacency information to constrain the analysis, usually by permitting only spatially (or temporally) adjacent stands to fuse, or by requiring that subgroups formed by division remain spatially coherent (see e.g. Gordon & Birks 1972, Lebart 1978, Margules, Faith & Belbin 1985). Yet such adjacency information is not the *only* situation in which a user might require a classification based on more than one set of dissimilarities, and in any case the treatment of the two matrices is markedly asymmetric. Are there alternative methods which might prove useful in such cases? If so then they are remarkable by their absence from phytosociological studies so far. Here we seek to partially redress this neglect.

Sources of Multiplicity

We can distinguish at least 9 different situations in which a user might expect 2 or more dissimilarity matrices to be pertinent to solving a phytosociological problem. We shall discuss these in turn.

1. Distinguishable Subsets of Attributes

This case considers the species to be divisible into subsets; for example, subsets based on life forms or functional groups. Note that this includes the case where each species forms its own dissimilarity matrix and these are then combined into a single matrix. So, when we calculate a dissimilarity matrix using almost any of the presently used coefficients, we do in fact operate first on each attribute independently, and then sum the components to make an overall dissimilarity measure.

Weir (1970) noted that this is not always what is required. In his NEWCLAS program he permitted the user to distinguish subsets of attributes. Each subset could be associated with its OWN dissimilarity measure, and the combination of these was accomplished by a weighted summation. Within each subset the attribute contributions were combined through the usual simple summation, that is with each attribute equally weighted. The TAXON package (Ross *et al.* 1984) provided the same functionality by defining a specific procedure (MATWAD; **MAT**rix **W**eighted **AD**dition) for combining dissimilarity matrices through weighted summation; the user simply generated a matrix for each set of attributes and equal weights, then combined the matrices as required.

Phytosociologically this could be an extremely attractive procedure. Users commonly distinguish subgroups of species in terms of life form, growth form, phenology, chorology or even syntaxonomic relevance. Given that these record different aspects of the vegetation, it might well seem sensible to allocate more weight to, say, the trees than to the bryoids. Yet if there are more species of bryoids than of trees the traditional simple summation procedure of combining their contributions will assign more weight to the bryoids whereas weights based on inverse floristic richness might be regarded as more appropriate.

Still further complications can be added. In the previous proposals, the summation of the matrices was made once at the start of the classification procedure. Belbin (1991) has carried this further by allowing the user to define a different sorting strategy for each dissimilarity matrix, and performing the summation at each stage in the fusion. This also permits him to use dissimilarity values to introduce several constraints rather than combining all the matrices into one.

Most important of all is the recognition by Lefkovitch (1978) that each attribute could be assessed independently and some form of consensus obtained. This proposal was made, in part, to avoid some of the difficult problems of polarising characters in taxonomic studies, but it is equally relevant to phytosociology.

2. *Spatial Scale of Sampling*

In this case we have the different matrices associated with different sizes of plot, as formally suggested by Podani (1984, see also Noy-Meir & Anderson, 1970). In many cases where the growth forms of the plants vary widely, investigators have employed different plot sizes for different classes of species. In such circumstances it may well seem sensible to differentiate the contributions to the measure of dissimilarity in a manner related to the plot size. This source is clearly related to the previous one, but differs in that it is not primarily a property of the plants themselves which indicates the use of several matrices.

3. *Differing Scales of Measure for Attributes*

Where the observations have been made in various scales of measures, such as nominal, ordinal, interval and/or angular, are each appropriately treated and weighted. Most floristic data are commensurate; they are all presence/absence or all cover-abundance measures, though there is an interesting question as to whether such measures *mean* the same for all species. If we attempt to incorporate environmental variables as well, then we may find several scales of measure simultaneously present. This immediately poses problems about how to combine the contributions - what can be done with mixed data? There are of course various suggestions in the literature - Goodall, 1966; Lance & Williams, 1967 - but even if we can define an equivalent dissimilarity measure for each different scale, some scales may be represented by many more attributes than others.

Consider for example a presence/absence floristic data set which is to be analysed together with some few ordinal, interval and angular environmental variables or even some associated quantitative information on the species themselves. Any investigator might well desire the combination of variables and, more important, the matrices derived from them, to be made in a controllable manner and not left solely to the default provided by some computer program. In such cases independent generation of dissimilarity matrices is sensible.

4. *Multiple measurements recorded for each species*

We can obtain multiple matrices simply by measuring several characteristics of the plants. Perhaps the most obvious would be the use of both structural and floristic information, or presence and abundance, but other choices are possible. For example consider the series "presence, density, valence, cover, cover-repetition, dry weight":

1. we start with the existence of the species on the plot, i.e. presence and record further the number of elements composing this incidence which gives us the density,
2. the pattern of occurrence within the plot, Greig-Smith's (1983) valence,
3. the size of these elements as projected cover,
4. the distribution of elements vertically as replicated cover and
5. the physical density of the materials composing the plant to give us the biomass at a particular height in the canopy.

Distinction of photosynthetic material and other such discriminations can be added to this list.

Each step along the list adds further information additional to that already available, probably increasingly noisy measurements!, and it would be interesting to examine how these differ in their organisation of the vegetation. It is possible to separate the component elements and generate dissimilarity matrices for each of them by using appropriate information statistics. It would also be interesting to examine the various dissimilarity matrices which can be generated at each stage, if only to determine where the major components of information reside.

5. *Replication in time*

In the previous section, the variables used to record vegetation performance formed a sequence in terms of information content. This can easily be extended to sequences in time; it just involves repeated sampling of the *same* stands at several different times. We can clearly calculate a dissimilarity matrix for the stands at each such period, and it would then seem reasonable to request a classification which treats each such matrix as a distinct entity.

There are of course other means of analysing such data, including regarding each stand as being described by a matrix of attribute x time values, which enlarges the set of descriptors but leads to a single dissimilarity matrix. There are also such methods as the transition matrices of Williams, Lance, Webb, Tracey & Dale (1969) or the strings used by Dale, Coutts & Dale (1988). In both these cases we convert the stand x species x time description into a single structure be it transition matrix or string of symbols, and then calculate a single dissimilarity matrix employing these structures as single attributes.

6. Replication in space

Here we have two alternative cases which might lead to multiple matrices. In one case we could use different matrices to represent different vertical layers, as in Lipmaa's (1939) unions. We could also record asymmetric matrices representing under/over relationships and then partition this into symmetric and skew-symmetric components.

In the other case, we could have subplots within the sampling area laid out in some systematic manner, for example as in Makirinta's (1978) sampling of Lake Kukkia where the subsamples surrounded a boat in a particular pattern. Since corresponding subplots can be identified in each sample, we could obtain replicate dissimilarity matrices by matching the subsamples. However, random subsampling is not a source which can be used, since there is no mechanism for determining which subsamples are to be regarded as analogous.

7. Alternative concepts of dissimilarity

Tversky's (1977) contrast measures distinguish components of "similarity"; that is similarity itself is regarded as a composite measure.

Tversky suggests a contrast measure as a better representation of human judgements. This is the basis of the Noest-van der Maarel (1989) suggestion of using both Bray-Curtis (1957) and Gower (1971) measures in weight-ed combination. Essentially this involves the weighted combination of 2 measures, *commonality* and *difference*. The former is concerned with the overlap (intersection) and the simplest form of this is the minimum of the 2 values, min(a,b). The latter involves 2 directional differences. If we define $(a,b)^+ = \max(0,a-b)$ then we can use $(a,b)+ (b,a)^+$. These are combined using weights to give $d_{ij} = f\{\min(a,b), a(a,b)^+, b(b,a)+\}$. We might obtain a,b by using the suggestions of Krumhansl (1978) or Ozawa (1983) that similarity is a function of local point density, although it is less obvious how to make the function f data dependent, and what the relationship between f and a,b should be.[1]

Austin & Belbin (1982) argued for a directional measure which is also present in Tversky's formulation. In fact any pair of such measures might be combined, for example Calhoun's topological distance, (Bartels *et al.* 1970) a difference measure, and Ulam's Levenshtein distance for ordered category data (Critchlow 1985), an intersection measure. It would also be possible to combine measures of the same kind, say several difference measures, but these often differ more in standardisation than in substance.

Again the solution adopted to make the combination was a weighted summation. Yet it is not clear that this is either the only procedure or the best procedure. What other methods might be used in such situations? And do these offer any particular advantages over the weighted summation method?

8. Asymmetric matrices

These can be directly observed, for example by recording vertical relationships of under/over. They can also be generated using the Austin-Belbin (1982) method for estimating similarity of species. Krumhansl (1978) and Ozawa's (1983) have both proposed local density modifications. In effect they chose to scale the observed pairwise dissimilarity by some local measure attributable to each point individually, primarily the local point density defined in terms of something like distance to k-th nearest neighbour. I have extended this to introduce local linearity and planarity measures and local conformance to some specific model.

The reason that asymmetric matrices lead to multiple matrices is because they can be partitioned into a symmetric part, a skew-symmetric part and an indicator matrix. However, there are appropriate methods of analysis specific to this case, so that it is not necessary to make the partition at all. We shall therefore ignore asymmetric matrices in the remainder of the paper.

9. Multiple Interpretative Viewpoints

The social sciences have long been concerned with the problem that the view a human observer takes of a situation can be modified by changing the role of the observed. In vegetation terms the view of a rainforest taken by a conservationist is unlikely to be the same as that taken by a indigenous inhabitant, a forester, a swidden practitioner, a politician or any of the other possible evaluators of the forest as a resource. I have worked on a project where a landscape artist attempted to provide a description of "forest experience" and the major problem was to relate his evaluation of emotional response to structural and floristic descriptions.

---

[1] There might be something to be obtained from the GUHA notion of *founded almost implication* of Havránek (1981) where founded almost implication(p,s) is *true*
iff $\#(f \cap j) * [\#(f \cap j) + \#(f \cap -j)]^{-1} \geq p$ and $\#(f \cap j) \geq s$

What is to be Done?

Given that multiple matrices are the foundation on which our single matrices are based, and that they can occur in several ways, what can we do to make them analysable. In essence each such matrix is a contribution of evidence and we are seeking to combine this evidence into some comprehensible, explanatory patterns.[2]

Again a variety of possibilities exist, although I do not propose to illustrate all of them here. Instead I shall discuss several alternatives and then concentrate on one of them.

1. Many to One

Given several matrices, from whatever sources they are derived, usually we seek to reduce them to a single matrix. The commonest method of doing this is to employ a linear combination of the elements, perhaps with weights. Thus Gower's (1971) distance measure sums differences for all species but weights each difference by the range for the attribute in question, the Canberra metric (Lance & Williams 1967) sums minima but scales by the sum, while Goodall's (1966) probabilistic measure combines the probabilities of similarity for each species, assuming independence. There may be a further normalisation of the resulting number often by the number of attributes but sometimes by other things. Thus the Bray-Curtis (1957) coefficient sums minima, but normalises by the sum of maxima.

Summation is not the only possibility, and indeed is less than satisfactory, if only because all the dissimilarities have to be initially scaled to some comparable levels (usually 0...1, or -1:1), if one is not to dominate entirely. To avoid this dominance we could better use the product. In essence we are replacing the arithmetic (weighted) mean by a geometric mean, and I suppose a case might be made for the harmonic mean in some circumstances. If enough different matrices were available then the mode might also be of interest, but mode estimation on small samples is not very profitable. (c.f. Taylor 1987, on fitting histograms optimally)

But relying on the numerical values may not be congenial, for it is easily disturbed by a few outlying values. Instead, it would be preferable perhaps to rely only on the ordering of the dissimilarities We could also choose to employ any other 'measure of central tendency', or indeed any defined quantile. Thus we could use the median, the midrange, the minimum, the maximum or even the 63rd percentile. Alternatively we might argue that we would essentially invariant estimates of dissimilarity because these indicate stable structure in the data. In such a case we would use a diversity or variance measure rather than any kind of average. Yet in making such a reduction we are losing information about the exact sources of similarity or difference.

Note too, that the general practice has been to combine the matrices statically before initiating the classification procedure, rather than dynamically before each fusion. Retaining the separate matrices throughout permits the user to select different sorting strategies for each matrix; a case where this was desirable was reported by Margules, Faith & Belbin (1985) .

2. Putting things in order

Rather than arithmetically combining the matrices, we can look for an ordinal methodology to attain the same end. In their two-step procedure Austin & Belbin (1982) proposed using the rows of the dissimilarity matrix itself as new attributes. If we replace the dissimilarity measures by their ranks, we can regard our problem as one of first estimating if the matrices are concordant in their values, and if so, further estimating an optimal ranking for the items using some non-null rank model.

This would be that pair with the lowest rank sum if dissimilarities were ranked (increasing) within each matrix. In this case we are effectively performing an analysis of variance and it should be possible to ask if there is evidence for this rank sum being different from that expected if the ranks are randomly permuted. Probably the correct statistic is Kendall's concordance coefficient (Kendall & Babington-Smith 1939)

However, we do not only seek to examine this statistic, since we are not interested solely in whether all the dissimilarities are concordant. Instead we employ a further test to determine if the sums of ranks can be distinguished; the process is analogous to post hoc tests in ANOVA. In fact the Lyerly (1952) and Page (1963) tests for ordered response is simply a simple measure of intergroup concordance. Meddes (1980) shows how contrasts might be evaluated in the general case. Note also Prentice (1978) and Ying (1990), who provide rank methods for censored and truncated data respectively; it seems likely that these approaches could be extended to estimation of correlation if this seemed desirable. Schemper (1984) permits arbitrary ties as well as censoring. If we follow Faith, Minchin & Belbin (1987) and Belbin (1991) in suppressing some of the larger dissimilarities, such procedures may be useful.

Here we would want to examine the smallest sum of ranks and see

> *a.* if it was abnormally small. This can be done by examining the deviations of the rank sums from their expected values (Whitfield, 1953). This provides an approximate test for termination of the classification.
>
> *b.* if it could be distinguished from other small values, which requires a multiple range test. Such a multiple comparison test is indeed possible (Whitfield 1954) based on the differences between the rank totals.

This second test permits the identification of several competing fusions to be followed further, using all pairs

---

[2] The philosophical literature provides a large number of alternative proposals for formally dealing with this problem - belief theories, possibility theory, modal logics, nonmonotonic logics but I am here concerned with more practical possibilities.

whose rank totals are indistinguishable from the smallest at the $\alpha\%$ significance level. To use such a test at least 3 dissimilarity matrices would be necessary and the test is not very powerful with few matrices and small numbers of items.

Further possibilities exist for using ordinal methods to reduce the multiple matrices to a single matrix. Schucany & Frawley (1973; see also Li & Schucany 1975) examined the problem of concordance between two groups of judges. They proposed a statistic based on the rank totals, essentially using average Spearman ρ_s values, so that concordance between groups can only be exhibited if there is concordance within. We can regard each matrix as providing a set of judges, the stands. Each of these ranks the other stands in dissimilarity, and this is repeated k, the number of matrices, times. Thus we can directly obtain a coefficient representing the similarity of the matrices themselves. However, a stand is most similar to itself so even its most alike neighbour can only attain rank 2, which prevents complete agreement ever being attained. Lewis (1971) examined how to avoid problems which arise from such self-ranking.

In contrast Ehrenberg (1952) looked at the change in concordance when 2 separate rankings are adjoined and the coefficient recalculated, so that he uses Total-within$_a$-within$_b$ (note the resemblance to homogeneity measures and information approaches). He bases his calculation on the average Kendall τ coefficient not the Spearman ρ_s, which is useful because the former is somewhat easier to interpret. Hays (1960), Linhart (1960), Quade (1962) and Feigin & Cohen (1978) have all examined somewhat similar methods; for other work see Snell (1983) and Fligner & Rust (1983).

Whichever specific method we choose, by comparing between matrices we can obtain a measure of the concordance between stands - in effect we generate a new (dis)similarity matrix based on the degree to which the various stands are concordant among themselves, with diagonal elements showing within-matrix concordance. We have then reduced our matrices to a single matrix which can be analysed normally. If there exist ambiguities in this matrix then we can follow several paths. Note finally that with concordance measures it is possible to test the significance of the coefficient as shown by Palachek & Schucany (1984).There are other possibilities which have phytosociological interest in the use of these ordinal models, especially with the so-called non-null models, but we hope to discuss these elsewhere (see Critchlow, Fligner & Verducci 1991, for a recent survey).

3. Subject to Permission

This is the archetypal constrained classification, where the smallest element of one matrix is selected, provided that the other dissimilarity values associated with this pair meet some specified logical predicate criterion (see Batagelj, 1984). Obviously one matrix is dominated by the others, hence, as noted earlier, this is an asymmetric analysis. It would be possible to utilise quite complex constraints, for example:

```
IF the maximal dissimilarity in the specified
matrix $\leq$ some specified value
AND the other corresonding dissimilarities are
ALL $\geq$ some critical value
OR u < some others < c
THEN the fusion will be permitted
ELSE it is prohibited.
```

Obviously the complication of constraints is limited only by the imagination of the user, and the ingenuity of the programmer! As noted in the Introduction, it would be possible to weight the choices rather than use complete prohibition, and I have previously developed a method for such weighting of spatial coherence, based on adjacency. In such a system, spatial coherence is not mandatory but it is emphasised.

Constraints can appear within the matrix itself. For example, the program TRANMAT in the TAXON package provides a sequential agglomerative hierarchical non-overlapping classification of transition matrices, with a similarity measure derived from information theory. However this particular measure cannot distinguish between completely *identical* matrices, where all the entries are numerically identical, and completely *disjoint* ones where every numeric entry in one matrix is matched by a zero entry in the other. We term this case *disassociation* and it is really telling us that if 2 sequences share no common transitions then you cannot say, on internal evidence alone, how different they are. Given a set of matrices the fusion sequence may, but need not, generate such cases, but if it does then the analysis must terminate before the hierarchy is complete.

The adjacency criterion most commonly used is spatial (Margules, Faith & Belbin 1985), or temporal, as with Dale & Walker's (1970, see also Gordon & Birks 1972) suggestion concerning pollen diagrams, but the user can define any such constraint derived from any other variables. Thus it may be desirable that some variable be retained within particular bounds, or that there is an upper limit which should not be passed. Here, for example, we could employ pH to order the stands and permit only stands with adjacent, or closely similar, pH values to fuse.

In general such constraints tend to be logically discrete, i.e. they are either met or not met. This is again not a necessary condition, for the constraint approach is obviously a special case of combining several matrices into one. In this case the assessment of adjacency is dynamic, that is it is not fixed at the start of the analysis, but is modified during the course of the processing. I have, for example, used Krishna-Iyer's (1949) method of assessing contiguity to provide a measure of adjacency to be combined with the floristic similarity (Dale, unpublished). The implementation used the product of the selected adjacency measure and the similarity measure to choose the pair for fusion. In practice this led to three different approaches based on different means of assessing adjacency.

a. The adjacency was assessed on the single group formed by the fusion - Individual or Black-Black adjacency

b. The adjacency was assessed on the 2 groups being fused - Local or Black-White

c. The adjacency was assessed over all groups present after fusion - Global or multi-hued.

In general these can be expected to give similar but not identical results. Interestingly the same three categories are pertinent in stopping rules for classifications which are based on internal group homogeneity criteria. A similar approach can be adopted in divisive classification, if desired. In neither case need the constraint be totally binding, so the analysis can be regarded as encouraging adjacency *inter pares*. Obviously there will in practice be some sort of trade-off between spatial adjacency and floristic similarity.

4. Now are we all agreed?

Another possibility is to first classify each matrix separately and then combine the results using a consensus technique. It is not difficult to define measures of similarity or dissimilarity between trees, or dendrograms, but what we are seeking here is equivalent more to a mean or median tree, for we are seeking to identify the common branching structure. The consensus approach has been championed by Podani (1989), although with more emphasis on its value as a means of *comparing* the results of several classifications than as a technique for treating multiple matrices. Lefkovitch (1978) also applies consensus techniques to ultrametric distances derived from classifications. However, he has further argued that consensus could be applied to *single* attributes, especially in cladistic studies where determining the directionality of relationship between character states is a major problem. Using the consensus delays the necessity for deciding directionality.

In fact there are several ways of attempting to obtain consensus. We can start with a strict consensus which demands that any structure in the consensus must be present in ALL other matrices, then relax the restriction so that only a majority of the matrices is necessary. There are several other definitions in the taxonomic literature - median trees, Steinbrickner' s-methods (1984a,b), and so on, each of which, in principle, can lead to a different final consensus. Most of the methods come down to choosing between various possible committee decision rules. For example, strict consensus relies on unanimity, majority rule on majority agreement. The work of Farebrother (1987) indicates that, at least in some cases, there is a relationship to L_1 approximation, i. e. minimising a sum of absolute differences, so that our choice would be based on the median dissimilarities.

There are many other voting systems which could be used; e.g. plurality, dictatorship, or power hierarchy. This last is useful if the matrices can be ordered in some manner to indicate their significance; in effect it is a kind of lazy dictatorship. We first order the dissimilarities in terms of the power assigned to the matrix (the weight) and then, for each in sequence, ask if the suggested fusion was acceptable, was unacceptable or indecisive. If the decision was indecision then we move to the NEXT dissimilarity measure in sequence to provide the decision. To order the matrices we might use Raunkiaer's (1934) life form categories and order the matrices

trees > shrubs > chamaephytes > hemicryptophytes > geophytes > therophytes

where > means "are more important than".

We first examine the tree matrix and have 3 possible outcomes:

a. a unique acceptable fusion can be identified. We proceed immediately to make the fusion

b. no acceptable fusion can be identified. Terminate the analysis

c. no decision is possible using this matrix. This means that ambiguities exist, so we proceed to the next lower-ranked matrix, the shrubs. Again we have the three choices, and again, in the case of indecision, we can move down the order.

d. If there is no lower ranked option remaining, ie the therophyte matrix was reached and is still indecisive, we again have a choice of actions. We may terminate, select a single fusion pair in some random or non-deterministic manner, or follow all remaining ambiguous pairs.

The same kind of sequential approach might be possible if the dissimilarities are given as intervals (e. g. confidence limits or some range of values) and not as point estimates. Again we have 3 cases

a. The interval with the minimal lower bound is disjoint from all others and can be immediately accepted.

b. the interval with the lower bound contains other intervals but these do not extend beyond its upper limit

c. overlapping intervals can be identified.

With multiple matrices it may be possible to resolve the ambiguities implied by cases b and c, so that a unique pair can still be identified. Note that in the constrained approach the constraints may well be defined as acceptable intervals.

5. A Fuzzy thought

It is possible to derive a classification methodology based on fuzzy logic. More specifically we search for that pair most approximately equal to the smallest values in each matrix. There have been proposals for fuzzy measures of 'approximately smaller then', 'approximately larger then' or 'approximately equal to'. A fuzzy approach to multiple dissimilarity matrix clustering can be based on the proposals of Naplatanoff & Lakov (1977). They define the following fuzzy functions where X is the degree of belonging. Thus:

$u1 \underset{\sim}{<} a \quad X1= \{=1|\ u1 \le a;\ \ =[1+(u1-a)4/a]-1\ |u1>a\}$
approximately smaller than

$u2 \underset{\sim}{=} b \quad X2= \{=1|\ u2=b;\ \ =[1+(u2-b)4/b]-1\ |u2<>b\}$
approximately equal to

$u3 \underset{\sim}{>} c \quad X3= \{=0|\ u31 \le c;\ \ =[c/1+(u3-c)4]-1\ |\ u3>c\}$
approximately bigger than

What we are seeking is that pair which belongs most to the class approximately smaller than all others. This is the intersection of the individual measures and hence is determined by the minimum "approximately smaller than" belonging. In fact any monotone decreasing function could be substituted for the fourth-power expressions. So we are in effect choosing that pair whose dissimilarities deviate least from the smallest for any matrix. If there is a pair which is uniquely smallest we shall obtain it. We could also seek that pair which was least approximately bigger than all others, that is the pair which all other pairs were approximately bigger than.

An alternative approach can be found in the use of linguistic quantifiers (see e.g., Boryslawski 1989) In essence though, this remains a committee approach, as is obvious from Kacprzyk (1985) who indicates that a variety of solutions are possible, in two major classes which he terms consensory and competitive.

6. Another way of looking

The use of transition matrices was mentioned earlier as one approach to analysing sequential data. In this case each stand is "described" by a matrix of the frequency of transitions between states which took place there. Each stand is therefore associated with a t x t matrix of state transitions. An alternative method is to associate each stand with a string of state codes indicating the sequence of states. Dissimilarity measures can be defined between transition matrices and strings, so we finally end with standard single dissimilarity matrix. Our problem has been reduced from many to one. However the choice of an appropriate dissimilarity measure for such structured data may have its own problems.

Another possibility has been discussed by Lehmacher & Wall (1978). They have compared response curves, asking if there is evidence for agreement between the curves. The same techniques would be applicable for successional studies, or in any case where there was a definitive ordering to the several dissimilarity matrices. Like Hollander & Sethuraman (1978) they are testing for deviation AWAY from perfect agreement i.e. deviation from +1.

Deciding what is best for you

One good heuristic for generalising methodologies is to examine the standard procedure in some detail, identify where limits apply, and seek to suggest how they might be overcome through exaggeration and analogy. Ferligoj & Batagelj (1992) have attempted this and I shall follow their arguments here.

The standard agglomerative classification procedure proceeds iteratively in the following way:

1. Search the matrix for the smallest dissimilarity value.

2. Indicate that the stands (or groups of stands) associated with this value are to be fused.

3. Recalculate the dissimilarity of all other stands (or groups) to this newly formed group, according to whatever sorting strategy has been selected.

4. Update the dissimilarity matrix with these values, deleting one of the just-fused pair

5. If there are more that 2 stands remaining in the matrix, GOTO 1
ELSE STOP

It should be noted that the *only* place at which we *need* a single measure is at the point at which we select the smallest dissimilarity as the basis for a fusion. All the other actions and activities can be applied independently to several matrices without modification. But the selection of the smallest value is simply the application of a procedure for identifying a pair to fuse! It should be remembered that some methods of clustering, such as Wishart's (1969) mode analysis or Jardine & Sibson's (1970) B_k method, do not necessarily select the smallest dissimilarity but some other, perhaps the k^{th} smallest. Thus we can replace the selection of the smallest dissimilarity by any other procedure which permits us to select one or more pairs of stands for fusion! Obviously we must choose any such procedure carefully so that we retain something of the notion of closeness, but there is nothing remarkable about selecting the smallest value. What procedure, then, should we choose?

Ferligoj & Batagelj (1992) have suggested the use of the Pareto optimal solution; they comment that there are other, laxer, solutions available if such prove desirable. If we have several matrices, each pair of stands can be associated with a vector of dissimilarity measures, one from each matrix. Let us examine one pair of stands, and its associated vector. The Pareto rule states that we shall accept this pair as suitable for fusion *iff* there is *no other* pair of stands where the corresponding elements in the vector of dissimilarities are always equal to, and *in at least one case* smaller than, the pair we are examining.

i.e. for k matrices, and items a,b,i,j then a,b are Pareto optimal

iff there is no i,j such that $\forall\ d_{ijp} \mid d_{ijp} \le d_{abp}$, $p=1....k \cap \exists\ k, d_{ijk} < d_{abk}$

Of course there may be several such acceptable pairs. Even in the single matrix case, if there are tied minimal values, we should strictly examine all the alternatives since it is possible that several different classifications could result from different tie-breaking selections. That we should have several possibilities in the Pareto case is, therefore, not alarming *sui generis*.

What is alarming is the very rapid increase in the possible number of alternative classifications, if at each stage there are several alternative fusions. Some of these may be regarded as trivial variations. Thus if we examine 3 stands, a,b,c, then we might find a possible fusion of a & b with c joining next, and equally one of b & c with a joining next. In both cases we end with a group a-b-c but with 2 routes subsumed within it. Any further grouping involving these stands will be identical in both cases. In general we are interested in the larger groupings formed near the root of the classification tree, and such minor variations in the initial fusions are often of little interest. Once the paths rejoin we can terminate (prune) one of the classifications, knowing that it will simply duplicate the other in subsequent fusions.

But even if we identify this situation and discontinue further search of one or other choice once it has been subsumed, the number of alternatives can still grow very rapidly. This is especially true if the structure in the various dissimilarity matrices is markedly different. How fast can be easily demonstrated. For 10 stands with 12 matrices each, a full Pareto analysis with pruning took about 17 minutes on a DEC workstation with a 20 MIPS performance. The time may seem unexceptionable. Unfortunately it also produced 110 Mbytes of output identifying the fusion paths for all the trees it found! This can hardly be regarded as light reading; this paper is only of the order of 60 Kbytes!

In such a predicament we must needs look for "interesting" classification trees. Somewhat surprisingly there is little information about the kinds of classification trees people find interesting! Trees have certainly been a popular mechanism for structuring, while the greater complexity of B_k classifications have not, although efficient programs have been produced for both. Diday & Bertrand's (1986) pyramids provide perhaps a more attractive method of generalising classification trees, but they too have not been widely exploited. If we examine any single matrix then the pair associated with the smallest element must be Pareto optimal; so if we have 10 matrices in general we shall have 10 optimal solutions immediately. However we are not interested so much in these, since we could more easily have analysed the individual matrices.

Consider the case of 2 matrices and construct a plot from the pairwise dissimilarity values which are selected as optimal. Each matrix will contribute a pair associated with smallest elements, and these will lie close to the ends of the axes. Any others will lie above or roughly on a hyperbolic curve joining these 2 points. One suggestion for interesting solutions is simply to select those solutions closest to the origin. These clearly form a set of compromise solutions, yet are relatively unlike the solutions obtained from the individual matrices.

Obviously if the smallest elements of several matrices coincide in position then they form a unique solution anyway. So a second possible definition of interesting would be any pairs which are minimal in more than one matrix; the more matrices the better. Others can be derived by using the actual dissimilarity values. For example we might calculate means, medians, midranges, max, min, range or other statistics and pick solutions for which these are either minimal or are within some threshold distance of the minimal solution; the standard single matrix methodology corresponds to using the mean in a static manner.

But adopting a converse viewpoint, it is possible that instead of using closeness to the origin we should select a solution in some sense midway between the individual matrix solutions. That is we should select compromise solutions which do *not* include matrix minima at all; assuming such exist. At present there is insufficient evidence to suggest which if any of these suggestions are profitable, and empirical study is necessary.

It seems that, unless the matrices are rather similar one to the other, we shall be overwhelmed with alternative solutions if we adopt the Pareto approach. The proposals for identifying interesting solutions given above will be investigated here. However a more complex procedure might be necessary. For example, Hubert (1979) has proposed a generalised measure of concordance which would enable us to assess the degree of similarity of the matrices in various situations. It is in fact closely related to the Schucany-Frawley measure discussed earlier and permits us to test the concordance within and between subsets of a *partition* of the matrices. Interesting partitions would then be those which are most concordant. However, the difficulties associated with assessing such concordance indicate that simpler methods should be examined first.

Data and Analyses

Given the large variety of sources from which multiple matrices can arise, we have chosen to restrict ourselves to a temporal source. We have examined here one set of data only, the Mt. Glorious data of Williams *et al.* (1969). This consists of 10 stands observed at 12 periods, recording at each the density of all plants present. For each period 2 dissimilarity matrices were generated, one using the Bray-Curtis coefficient and presence/absence data, the other the 'change in diversity' information coefficient termed CENTPERC 2 by Dale & Anderson (1972).

The program, written in C, involves checking for possible Pareto optimal pairs which is an $o(n^4)$ operation, for n stands. Pruning is implemented through a very efficient hash-function which permits rapid identification of any partition. The user may enter any number of matrices, each associated with a different sorting strategy. In addition a variety of 'interest' functions are available to increase the amount of pruning. They include accepting all Pareto solutions (the unrestricted solution), or minimising the mean, median, midrange, product, maximum or range of dissimilarity values. We are presently implementing other *interest* functions; suggestions are welcome! We also permit the user to specify a *slop* factor, a percentage of the selected minimal value, which provides a threshold for acceptable alternatives. However in our examples the *slop* factor was always set to 10% of the minimum identified value.

As noted earlier, the complete Pareto analysis produced a very large output which is impossible to present here. We therefore proceed as follows. Because the dissimilarity matrices are each associated with a specific time, we can order them in sequence. Using the midrange interest function to reduce possibilities, we examined the analyses of matrix 1 alone, matrices 1 & 2, 1, 2 & 3, 1,2, 3 & 4 ... ,1 & 2 &...12 until all 12 matrices were included. In one case we used only the information measures, in another we used only the Bray-Curtis measure and finally we introduced both information and Bray-Curtis measures simultaneously.

In addition we examined the use of various other interest functions. Here we shall show only the median interest function applied to all 12 Bray-Curtis matrices. The other functions either gave results similar, and in some cases identical, results or produced even larger volumes of output. It should be noted that if the various trees developed are drawn there is no associated number which can be attached to each level; only the order of fusions can be defined. But we shall not provide any trees here for reasons which will soon become apparent.

Results

Output Complexity

The reason for not providing trees is simply the quantity of output produced. One very simple measure of complexity is simply the size of the output. We shall measure this in terms of the number of lines of output, since each fusion produces the same number of lines (3 in our case), each pruning produces less (1 in our case). More lines therefore indicate more alternative fusions; pruning reduces the total number of lines. By plotting the size of output against the number of matrices included we get a picture of the complexity of the structure being examined as we proceed along the time axis.

Starting with the information or contrast measure, Fig. 1 shows there is, in general, an increase in complexity with a greater number of matrices. However there is a very marked increase when matrix 8 is added in the information analysis, followed by a fall at matrix 10. This means that, while for the first 7 periods the dissimilarities were moderately consistent, matrix 8 and 10 introduced very different competing patterns of relationship, resulting in the increase in alternative solutions. After matrix 10 the falling complexity seems to indicate a period of convergence of dissimilarity values; that is there is an introduction of some stability in the assessment of between stand dissimilarity, although the impact of the variability in matrices 7-9 cannot be lost completely of course.

This pattern seems to fit well with an interpretation that the 'pioneer' species were replaced by others, and that during the transitional period there was considerable confusion in the assessment of dissimilarity as the species composition changed independently in each sample area. With the loss of the 'pioneers' and the closure to addition of further species, a more stable regime is approached. The original paper suggested that some sort of convergence to 2 types had indeed occurred, one rain-forest proper the other dominated by *Lantana camera* although it was also suggested that the latter would in time be suppressed by the surrounding forest.

The Bray-Curtis measure, which assesses commonality of the vegetation, shows a markedly different pattern (Fig. 2). There is a minimum at the 2nd time period, then a steady increase up to period 6, followed by a fall to an essentially constant complexity. This would suggest that after an early convergence the vegetation diverges but the later appearance of forest species equalises the dissimilarities and an essentially constant diversity is approached; presumably this level of complexity reflects the rain-forest *Lantana* dichotomy noted earlier. Of course we suspect that such would not be maintained for a long period of time. It simply means that for this series of samples, which in fact is not a very large area, the secondary succession does converge at least in the short term to some stable vegetation forms.

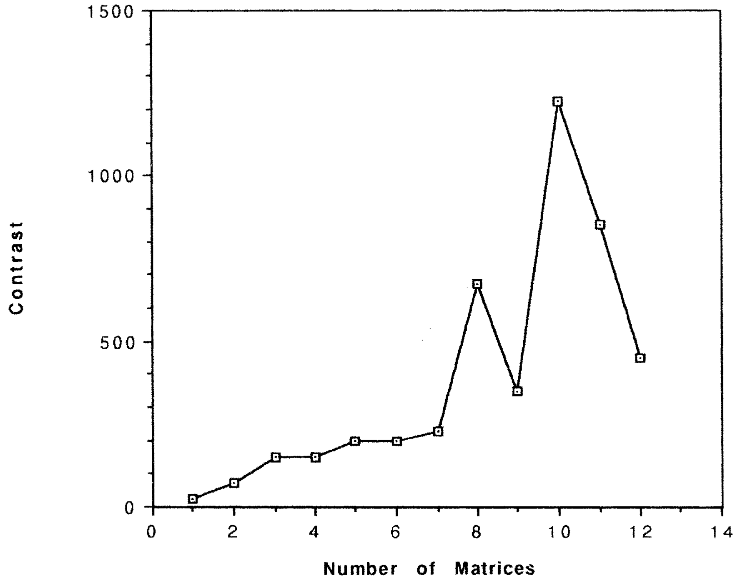

Figure 1. Increase in complexity with increasing number of matrices: contrast (information) measure.

Combined Time and Coefficient Matrices

Combining the 2 coefficients in the manner of Noest & van der Maarel (1989) shows us some of the effects suppressed by the usual summation method of combining matrices (Fig. 3). Even using only the first pair of matrices the results start to differ from those of the single coefficient analyses. Far from period 2 showing stability, as in the Bray-Curtis analysis, there is a marked increase in complexity. Here are further peaks in complexity at periods 7 and 10; only one peak therefore coincides with a peak in a single-coefficient analysis. There is little sign of any approach to stability anywhere in the series; dynamic changes are rife. While this single test should not be regarded as conclusive in any measure, proposals for using combined coefficients must clearly be regarded with considerable caution. It must call into serious doubt any proposals to combine measures in the manner suggested by Noest & van der Maarel, whatever superficial attractions such a procedure might have.

Superposing the 3 results (Fig. 4) shows the marked increase in complexity the additional matrices introduce. The combined result is markedly more complex than the single coefficients until the introduction of period 10 samples. With 10 stands, a tree will have around 30 lines of output, so the commonality analysis, which stabilises at around 250 lines is equivalent to at least 8 distinct trees although because of the pruning there may be considerably more than this. The contrast measure initally has around the same complexity, and indeed the major disturbances seem to arise at about the point where the contrast and commonality curves cross. The combined analysis rarely has so low a complexity, and in some

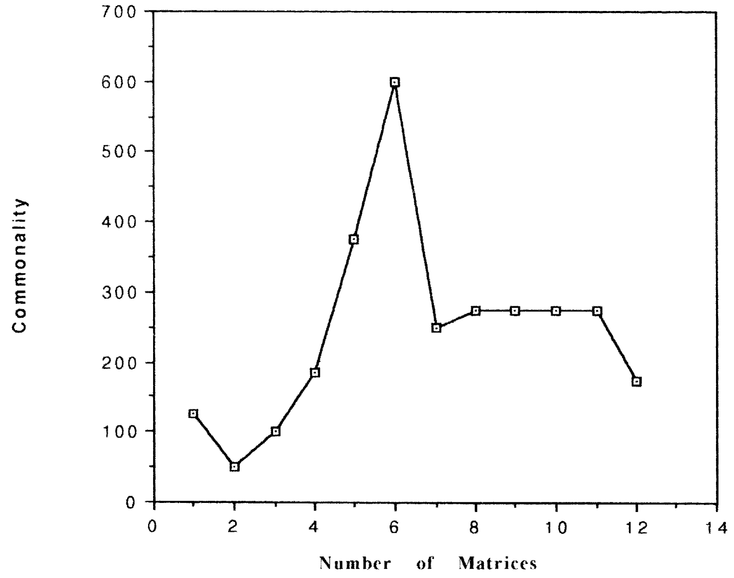

Figure 2. Increase in complexity with increasing number of matrices: commonality (Bray-Curtis) measure.

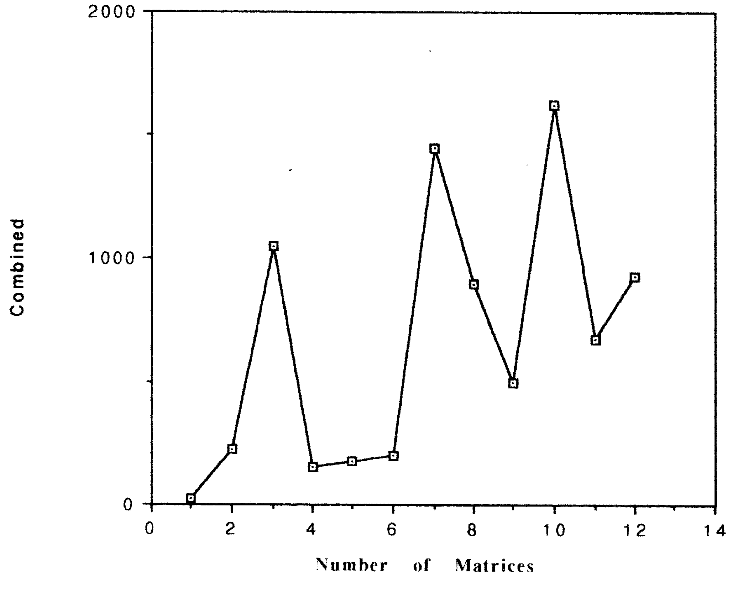

Figure 3. Increase in complexity with increasing number of matrices: Pareto-combined measures.

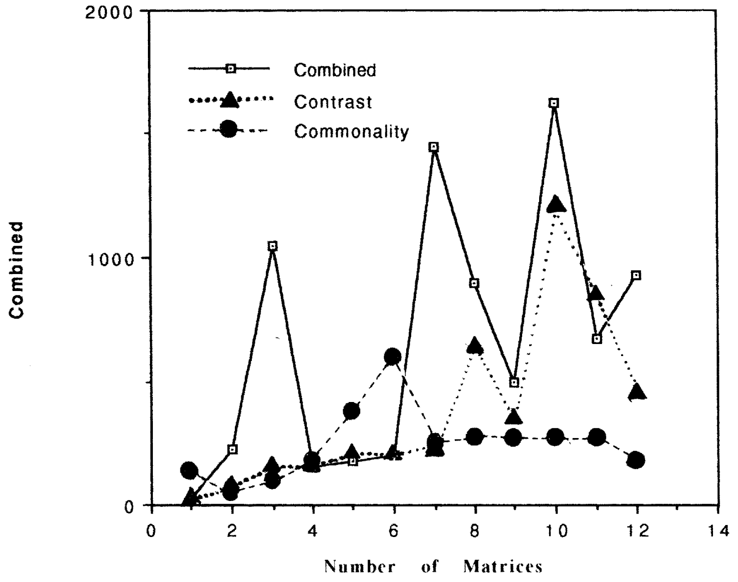

Figure 4. Superposition of results to indicate discrepancies between the various measures.

places reaches the equivalent of 50 trees or more. Such complexity is a reminder of the diversity loss which summation would impose! There is not one relevant tree structure here but many. Yet with 118 species, there are clearly more possibilities than we are actually recording; there is some redundancy in the data, but much less than we might hope.

Interest Functions

Interest functions gave some interesting results although we shall not provide details here. Using the minimum total dissimilarity, which is equivalent to the mean, gave an unwieldy output of approximately 20 pages. Using Maximum gave about half this output, while the others were considerably smaller. The midrange proved the most restrictive of those using some central tendency measure, with product and median also providing considerable pruning, but still being uncomfortably large. The range, which of course is measuring the consistency of the dissimilarity estimation and not its absolute value, also seems to provide an acceptable set of solutions in many cases.

For the 2 matrix, different dissimilarity, analysis, the unconstrained solution resulted in 13.5 Mb output, whereas the product produced only some 9 -10 pages. Range and midrange were acceptably small, and in this case gave identical results! Although further study is clearly required, the present results suggest that the method of summation of matrices to provide a single dissimilarity measure is not a particularly attractive or effective method of proceeding. Some more robust method, such as the median, is probably to be preferred.

Discussion

In the more specific case of the methodology illustrated here, the conclusion to be drawn is largely negative. While it is indeed possible to analyse multiple matrices using the Pareto approach, the resulting output is likely to swamp the user. Larger matrices, and more of them, will generally increase this volume, as well as rapidly becoming computationally intractable, unless the structure in all the matrices is very similar.

To cope with this exponential explosion, it is necessary to identify interesting classifications with a high rejection rate of uninteresting ones. Yet the identification of a classification as *interesting* would seem to depend largely on the relevance of the result to the problem of elucidating processes generating the patterns. It is not obvious that this can be done using *only* the information available to the classification procedure. Instead we are necessarily forced to examine the contextual evidence, which in general will mean the environmental semantics of the species. Practically speaking, it would appear likely that weighted combination of matrices or consensus studies provide more profitable methods for analysing multiple matrices. This is in some contrast to ordination, where multiple matrices used in an individual differences analysis can provide a great deal of interesting information and incidentally, but not necessarily less importantly, provide a unique solution!

More generally, the investigation has been more rewarding, since it has focussed attention on some fundamental problems of phytosociology. The problem of how to combine various pieces of evidence is a fundamental problem which has to be faced in quantitative phytosociological studies. Furthermore, the necessity of identifying interesting solutions has directed attention at the question of how results from exploratory analyses relate to the explanation of data. Stockwell & Sachse-Åkerlind (in prep.) have examined the forms of explanation used by ecologists to provide justifications for arguments. They are several in number, not always logically consistent, but often useful as a means of directing our realisation of possibilities.

It is equally clear that multiple matrices are very common, and very commonly ignored! It is not, in our view, sufficient to accept whatever devices are included in the locally available computer package without thought for the consequences. Indeed they strike at the heart of phytosociology, since they demand an answer to the question of when, or if, 2 or more species can be regarded as commensurate attributes of a stand. Do we really believe that some small hepatic nestled among buttress roots is really equivalent to the *Argyrodendron* species which is providing those selfsame buttresses? And if we answer yes, then it is legitimate to ask in what sense(s)? Are the environmental factors impacting on the one the same, or even comparable, to those acting on the other? Webb *et al.* (1967) identified the "big tree species" as sufficient to represent the macro-environmental patterns of North Queensland rain-forest, but what is the nature of the correlation of the other groups of species? For if the hepatic *is* important then should we not consider the microflora? Where ought we stop?

Even if we are prepared to accept commensurability, we should still examine carefully the manner in which the components of evidence, which the different matrices represent, are united. Are they independent, are they consistent and if not what should be done about it? Clearly our answers to these questions will have to be based in some theory of vegetation structure and processes. Not that these are necessarily immediately correct. Approach to correctness will almost certainly be by successive approximation and refinement, but also by change of representation and revolutionary change. We can no longer expect that our forefathers assumptions should remain unchallenged and unchallangeable. A science which cannot deny its history is no science.

References

Austin, M.P. & Belbin, L. 1982. A new approach to the species classification problem. Austral. J. Ecol. 7:75-89.

Bartels, P.H., Bahr, G.F., Calhoun, D.W. & Wied, G.L. 1970. Cell recognition by neighbourhood grouping techniques in TICAS. Acta Cytologica 14:313-324.

Batagelj, V. 1984. Agglomerative clustering methods in clustering with constraints. Preprint Ser., Dept. Math., Univ. Ljubljana 22:5-19.

Belbin, L. 1991. Semi-strong hybrid scaling, a new ordination algorithm. J. Veg. Sci. 2:491-496.

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

Breslow, N. 1970 A generalized Kruskal-Wallis test for comparing k samples subject to an arbitrary pattern of censoring. Biometrics 57:579-594

Critchlow, D. 1985. "Metric Methods for Analyzing Partially Ranked Data". Springer-Verlag, New York.

Critchlow, D.E., Fligner, M. A. & Verducci, J. S. 1991. Probability models on rankings. J. Math.Psychol. 35: 294-318.

Dale, M.B. & Anderson, D.J. 1972. Qualitative and quantitative information analysis. J. Ecol. 60:639-653.

Dale, M.B., Coutts, R. & Dale, P.E.R. 1988. Landscape classification by sequences: a study of Toohey Forest. Vegetatio 29: 113-129

Dale, M.B. & Walker, D. 1970. Information analysis of pollen diagrams. Pollen et Spores 12:21-37.

Diday, E. & Bertrand, P. 1986. An extension of hierarchical clustering: the pyramidal representation. In: E.S. Gelsema & L.N. Kanal (eds.) "Pattern Recognition in Practice II." Elsevier Science Publ., North Holand, pp. 411-424.

Ehrenberg, A.S.C. 1952. On sampling from a population of rankers. Biometrika 39:82-87.

Esty, W.E. 1992. Votes or competitions which determine a winner by estimating expected plurality. Amer. Statist. Assoc. J. 87: 373-377.

Faith, D.P., Minchin, P.R. & Belbin,L. 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69:57-87.

Farebrother, R.W. 1987. The theory of committee decisions and the double median method. Comp. Statist. Data Anal. 5: 437-442.

Feigin, P.D. & Cohen, A. 1978. On a model of concordance between judges. J. Roy. Statist. Soc. B 40:203-213.

Ferligoj, A. & Batagelj, V. 1992. Direct multicriterion clustering algorithms. J. Classif. 9:43-61.

Fligner, M.A. & Rust, S.W. 1983. Recent literature on testing for intergroup concordance. Appl. Statist. 32:134-140.

Goodall, D.W. 1966. A new similarity index based on probability. Biometrics 23:883-907.

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

Gower, J.C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27:857-874.

Greig-Smith, P. 1983. "Quantitative Plant Ecology". 3rd. ed. Blackwell, Oxford.

Havránek, T. 1981. Formal systems for mechanized statistical inference. Int. J. Man-Machine Stud. 15:333-350

Hays, W.L. 1960. A note on average Tau as a measure of concordance. Amer. Statist. Soc. J. 55:331-341.

Hollander M. & Sethuraman, J. 1978. Testing for agreement between two groups of judges. Biometrika 65:403-411.

Hubert, L.J. 1979. Generalized concordance. Psychometrika 44: 135-142.

Jardine, N. & Sibson, R. 1970. The construction of hierarchic and non-hierarchic classifications. Comput.J. 13:116

Kacprzyk, J. 1985. Group decision making with a fuzzy majority via linguistic quantifiers. I. a consensory-like pooling. Cyber. Syst. 16:119-129.

Kendall, M.G. & Babington-Smith, B. 1939. The problem of m rankings. Ann. Math. Statist. 10:275-287.

Kraemer, H.C. 1981. Intergroup concordance: definition and calculation. Biometrika 68:641-646.

Krishna-Iyer, P. 1949. The first and second moments of some probability distributions arising from points on a lattice and their application Biometrika 36:135-141.

Krumhansl, C. L. 1978. Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. Psychol. Rev. 85:445-463.

Lance, G.N. & Williams, W.T. 1967. Mixed data classificatory programs I. .Agglomerative systems. Aust. Comput. J. 1:15-20.

Lebart, L. 1978. Programme d'aggregation avec constraintes. Cah. Anal. Donn. 3:275-287.

Lefkovitch, L. P. 1978. Consensus coordinates from qualitative and quantitative attributes. Biom. J. 20:679-691.

Lehmacher, W. & Wall, K. D. 1978. A new nonparametric approach to the comparison of K independent samples of response curves. Biom. J. 20:261-273.

Lewis, G. H. 1971. Kendall's coefficient of concordance for sociometric rankings with self-exclusion. Sociometry 34:494-503

Li, L. & Schucany, W.R. 1975. Some properties of a test for concordance of two groups of rankings. Biometrika 62:417-423.

Linhart, H. 1960. Approximate test for m rankings Biometrika 47:476-480.

Lipmaa, T. 1939. The unistratal concept of plant communities (the Unions). Amer. Midl. Natur. 2:111-145.

Lyerly, S. B. 1952. The average Spearman rank correlation coefficient. Psychometrika 17:421-428.

Mäkirinta, U. 1978. Die Pflanzensoziologische Gliederung der Wasservegetation im See Kukkia, Süd-Finnland. Acta. Univ. Ouluensis Ser. A Scient. Rerum Natur. No. 75, Biologica No.5. Univ. Oulu, Oulu, 157 pp.

Margules, C. R., Faith, D. P. & Belbin, L. 1985. An adjacency constraint in agglomerative classifications of geographic data. Environ. Planning A 17: 397-412.

Meddes, R. 1980. Unified analysis of variance by ranks. Brit. J. Math. Statist. Psychol. 33:84-98.

Naplatanoff, N & Lakov, D. 1977. Decision making in vague conditions. Kybernetes 6:91-93.

Noest, V. & van der Maarel, E. 1989. A new dissimilarity measure and a new optimality criterion in phytosociological classification. Vegetatio 83:157-165.

Noy-Meir, I. & Anderson, D. J. 1970. Mutliple pattern analysis, or multiscale ordination: towards a vegetation hologram. In: Patil, G.P., Pielou, E.C. & Waters, W. E. (eds.) "Statistical Plant Ecology III" Pennsylvania State University Press, University Park.

Ozawa, K. 1983. CLASSIC: a hierarchical clustering algorithm based on asymmetric similarities. Pattern Recognition 16: 201-212.

Page, E.B. 1963. Ordered hypotheses for multiple treatments: a significance test for linear ranks. Amer. Statist. Soc. J. 58: 216-239

Palachek, A.D. & Schucany, W.R. 1984. On approximate confidence intervals for measures of concordance. Psychometrika 49:133-141.

Podani, J. 1984. Spatial processes in the analysis of vegetation: Theory and review. Acta Bot. Hung. 30:75-118.

Podani, J. 1989. A method for generating consensus partitions and its application to community classification. Coenoses 4:1-10.

Prentice, R. L. 1978. Linear rank tests with right censored data. Biometrika 65:167-179.

Quade, D. 1967. Rank analysis of covariance. Amer. Statist. Assoc. J. 62:1187-1200.

Raunkiaer, C. 1934. "The Life Forms of Plants and Statistical Plant Geography". Oxford Univ. Press, Oxford.

Ross, D., Dale, M. B., Shields, K. & Hulett, C. 1984. "TAXON Users Manual" ed. P4, CSIRO.

Schemper, M. 1984. Exact test procedures for generalized Kendall correlation coefficients. Biom. J. 26:399-406.

Schucany, W. R. & Frawley, W. H. 1973. A rank test for two group concordance. Psychometrika 38:249-258.

Sen, P. K. 1968. Estimation of the regression coefficient based on Kendall's Tau. J. Amer. Statist. Assoc. 63:1379-1389.

Snell, M. C., 1983. Recent literature on testing for intergroup concordance. Appl. Statist. 32:134-140.

Stinebrickner, R. 1984a. An extension of intersection methods from trees to dendrograms. Syst. Zool. 33:381-386.

Stinebrickner, R. 1984b. s-Consensus trees and indices. Bull. Mathem. Biol. 46:923-935.

Taylor, C. C. 1987. Akaike's information criterion and the histogram. Biometrika 74:636-639.

Tversky, A. 1977. Features of similarity. Psychol. Rev. 84: 327-352.

Webb, L. J., Tracey, J. G., Williams, W. T. & Lance, G. N. 1967. Studies on the numerical analysis of complex rain-forest communities. II. The problem of species sampling. J. Ecol. 55:525-538

Weir, A. 1970 Program NEWCLAS: a classificatory program for mixed data. CSIRO Land Use Research, Tech Memo.70/5. 10pp.

Whitfield, J.W. 1953. The distribution of total rank value for one particular object in m rankings of n objects. Brit. J. Statist. Psychol. 6:35-40.

Whitfield, J.W. 1954. The distribution of the difference in total rank value for two particular objects in m rankings of n objects. Brit. J. Statist. Psychol. 7:45-49

Williams, W. T., Lance, G. N., Webb, L. J., Tracey, J. T. & Dale, M. B. 1969. Studies in the numerical analysis of complex rain-forest communities. III. The analysis of successional data J. Ecol. 57:515-535.

Wishart, D. 1969. Mode analysis: a generalisation of nearest neighbour which reduces chaining effects. In: A.J.Cole.(ed.) "Numercial Taxonomy". Academic Press, London. pp. 282-311.

Ying, Z-L. 1990. Linear rank statistics for truncated data. Biometrika 77:909-914

Manuscript received: November 15, 1993