COMMUNITY ECOLOGY 1(1): 43-56, 2000
1585-8553 © Akadémiai Kiadó, Budapest

On plexus representation of dissimilarities

M. B. Dale

Faculty of Environmental Sciences, Griffith University, Nathan, Qld. 4111 Australia.
Fax: +61 7 3870 5681, E-mail: m.dale@mailbox.gu.edu.au

Keywords: Correlation, Euclidean representation, Gradients, Graph theory, Plexus, Riemannian representation, Spanning tree, SplitsTree.

Abstract: Correspondence analysis has found widespread application in analysing vegetation gradients. However, it is not clear how it is robust to situations where structures other than a simple gradient exist. The introduction of instrumental variables in canonical correspondence analysis does not avoid these difficulties. In this paper I propose to examine some simple methods based on the notion of the plexus (*sensu* McIntosh) where graphs or networks are used to display some of the structure of the data so that an informed choice of models is possible. I show that two different classes of plexus model are available. These classes are distinguished by the use in one case of a global Euclidean model to obtain well-separated pair decomposition (WSPD) of a set of points which implicitly involves all dissimilarities, while in the other a Riemannian view is taken and emphasis is placed locally, i.e., on small dissimilarities. I show an example of each of these classes applied to vegetation data.

Abbreviations: MST - Minimal Spanning Tree; CA - Correspondence Analysis; w-b-c - Williams, Bunt and Clay (1991).

Introduction

Both correspondence analysis (CA) and canonical correspondence analysis (cf. ter Braak 1986) have proved extremely popular as means of analysing vegetation data and with Yanai (1988) and Escofier et al. (1990) developing extensions to correct for spatial correlation, their usage is likely to grow further. Birks et al.'s (1994) bibliography records 379 entries and, being geographically biased, underestimates the total.

Recently, incorrect specification of convergence tolerances in a specific widely-used computer program was identified by Oksanen and Minchin (1997), as a cause of data-order dependence in the results[1]. Such dependence is not unique to correspondence analysis, and the algorithmic fault is easily remedied, but the discovery provides food for thought. A similar, earlier recognition (Beals 1973) that all was not well with ordination procedures in vegetation studies led to the development of new methods and ultimately to the spread of CA and various related methods of Gaussian ordination (cf. Hill 1973, Ihm and van Groenewoud 1975, Goodall and Johnson 1982, 1987). It was also one (distal) cause for my distinguishing two major objectives for ordination, dimensionality reduction and gradient seeking (Dale 1975). How many other problems are users likely to find in these ordination procedures?

Before examining the potential for problems, I should make it clear that my objective is not to denigrate CA and similar methods. It is rather to suggest that other, possibly simpler, methods might be useful in indicating if any problems are likely to occur. Some modifications and alternatives have already been studied. Famili and Turney (1991) and Fayyad et al. (1996) suggested preprocessing techniques for improving CA and there have been suggestions that other methods provide better recovery of gradient structure with fewer assumptions. Such methods include non-metric multidimensional scaling (Minchin 1987), principal curves (De'ath, 1999) and appropriately standardised Principal Components (Karadžić and Popović 1994)).

In this paper I propose to examine some procedures for providing visual assessment using what are variously

1 'Data dependence' here means that the results obtained changed if the order in which the data entered the computer program was changed (see Tausch et al. 1995). Mathematically such dependence is impossible but computationally it is another matter. In this case the mis-specification led to an early, and incorrect, termination of the iterative procedure for calculating the eigensolution.

termed graphs, hypergraphs, networks or plexuses. However, I shall not be concerned with the problems of automatically drawing such networks; for material on this see Tamassia and Tollis (1995). For special cases, such as planar graphs and trees, effective algorithms are available but the general case is very difficult.

Potential problems with correspondence analysis

We can distinguish at least four classes of potential problems, none of which are restricted to correspondence analysis:

- interpretational problems;
- algorithmic problems;
- user choices; and
- incorrect assumptions.

Interpretational problems

These arise when the nature of the result is misinterpreted by the user. A simple example is shown in Fig. 1 where the principal axes of a rectangular shape are indicated. Since they are chosen to maximise variation, the axes do NOT bear a simple relation to the rectangular shell, running diagonally rather than parallel to the 'sides'. Another common case is where the samples form two or more discrete clusters (see e.g., Duckworth et al. 2000), when the result is likely to compromise between intra- and inter- cluster structure. For this class of problems, the solution requires only education of users.

Algorithmic problems

Users of particular computer programs should be wary of possible flaws in the algorithms implemented. Projection pursuit methodology, which provides another interesting class of ordinations, has only recently become viable with the development of appropriate and effective algorithms (Posse 1995). It seems that almost all the early studies were decidedly suboptimal and sometimes totally erroneous! This is certainly not the case for CA but some problems might still appear.

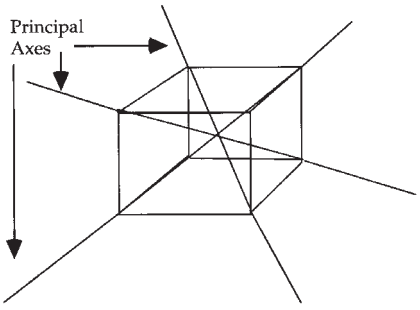

Figure 1. Principal axes of a cuboid.

Sequential extraction of axes. The eigenanalysis which underlies CA is commonly made using the reciprocal averaging algorithm, which identifies axes sequentially, starting with the axis with largest eigenvalue, which is obviously faster. But there is a slight risk, (I have had it happen!) that the initialisation approximation is closer to some vector other than those which interest us and then we may not obtain the axes we are seeking, instead converging to the nearer axis.

Coarse data. CA was designed for use with binary and frequency data whereas ecologists commonly use coarse, ordered category variables such as cover-abundance scales. The possible effects of this are simply neglected. Such neglect may be benign but I do not believe it should be ignored. Naouri (1970) has considered quantitative CA so a comparison of numeric and ordered category data should be possible.

Robustness. Robustness is primarily concerned with the effects of outliers or other information with a high leverage. We might expect to use algorithmic procedures which protect us against such effects such as that of Gabriel and Odoroff (1984), rather than relying on notoriously non-robust least squares methodology. A note of caution is necessary, though, for Naga and Antille (1990) showed that robust axes were not necessarily more effective in recovering known axes in Principal Component Analysis.

Resistance. Resistance is concerned with the effects of missing values which are normally not of great concern with vegetation data but are possibly more common with associated environmental data. There is, however, a well known logical dependency between presence and abundance; you do not know how much something is absent. One technique for treating such dependency is to partition the data into binary 'presence/ absence' and numeric 'abundance when present' categories with absences regarded as 'missing values' in the numeric category. Solutions are offered by Gabriel et al. (1988). Data can also be partially available, censored, ambiguous or erroneous as well as actually missing or inapplicable (Babad and Hoffer 1984).

User choices

Most analytic methods involve the user in making choices and CA is no exception. The problems arise when default values are adopted without due consideration.

Rarity. If a species occurs only in a single stand with no other species present, then we would expect an eigenvalue of unity. In fact, this is rarely seen in analyses because down-weighting of rare species is common practice (see Eilertson et al. 1989). Whether such down-weighting is entirely justified is another matter. I would suggest that it would be desirable to know of the existence of such axes before totally discarding them. The first CA I ever attempted, studying relationships between Lepidoptera and their host plants, commenced with 11 unit eigenvalues!

The number of axes. Van Groenewoud's (1992) empirical study reported that CA and its detrended offspring were both very bad at recovering any axis except the first. Indeed, examining the literature, it is rare for a sensible second axis to be reported, and not uncommon for a diagonal axis to be apparent in point plots. This finding is the more interesting because of the recent demonstration by Hubert and Arabie (1992) that ANY pair of axes in a CA can be replaced by a single, different, axis. They argue from this that the size of the eigenvalues is NOT a good measure of whether more than one axis is necessary. Even if we argue that a 2-dimensional solution is necessary because we have bivariate Gaussian responses for each species, to expect that the recovered axes will conveniently reflect the bivariate nature is to hugely constrain the solution. I doubt that realised niches are so conveniently spaced in an environmental universe that their projections will conform. Indeed if they did, it would be of enormous significance. For linear factor analysis, methods have been proposed for determining the appropriate number of factors (Wallace 1995), but these have not so far been extended to other ordination methods.

Incorrect assumptions

The gradient model may not in fact be the best choice; for example I have one set of data which fits a trivariate normal distribution very well. Simply because we have sampled along what appears visually to be a gradient is no guarantee that there is but one simple gradient present or indeed any gradient at all! The gradient we are investigating is a *vegetation* gradient and this need not correspond to any environmental gradient since it may be self-generated.

With disjoint data the order of the stands and species cannot be uniquely recovered. It is also possible for 2 or more gradients to be appended into a single axis. This is illustrated in Fig. 2 where, because wet serpentine is similar to dry granite, it is possible to arrange the stands on a single gradient. We then lose any relationship with moisture unless we happen to include appropriate interaction terms. Of course we can have much more substantial departures from a single gradient than this simple case of abutment. The introduction of instrumental variables in canonical correspondence analysis does not, of itself, alleviate the problem and introduces the difficulties of choosing an appropriate regression function.

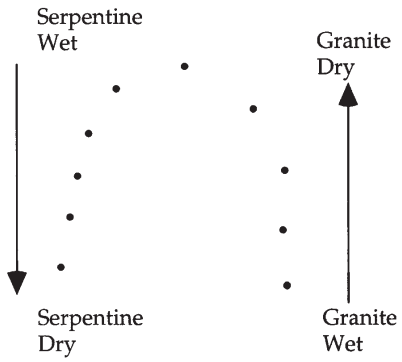

Figure 2. A 'horseshoe' confounding interpretation. Redrawn from a Principal Components Analysis of some Zimbabwean data.

Response function. CA presumes a specific form of response function with equally spaced identical bell-shaped curves. Empirically this does not seem realistic (see Austin 1990) and alternatives have been proposed (see e.g., Austin 1976, Huisman et al. 1993).

Global and local solutions. One feature of using eigenvalue techniques for identifying gradients often overlooked is the global nature of the solution. What this means is that, no matter how long the gradient, the stands and species at one end still influence the positioning of those at the other. It is not clear that this is a desirable feature, especially for a long gradient. It is certainly true that dissimilarities will be best estimated between near neighbours, and that disjunct samples will have very badly estimated dissimilarities. The obvious solution is to emphasise local information and this is the approach adopted in principal curves. De'Ath (1999) has shown the effectiveness of such analyses, but even here it is common to use relatively long sections of the sequence, say 1/4 of its total length, in the smoothing procedure. Still greater emphasis on small dissimilarities may be desirable.

Environmental relationships. If we knew of the existence of nonlinearities, serial dependencies and interactions in the relationship between vegetation and environmental factors, then we could construct an appropriate regression model. But how do we know this especially if we are conducting an exploratory analysis? Methods such as linked vector plots (Taguri et al. 1976) analysis might suggest that a relationship exists for only part of the data, say only

the upper or lower ranks of the stand series while quite different relationships exist elsewhere in the data.

It can be seen from the above that in many cases appropriate solutions to these problems are known even if rarely used. However, I would argue that it is desirable to investigate simple methods for confirming the existence of a gradient as opposed to other structures, before applying correspondence analysis. At present formal comparison of models is difficult.

Plexus procedures

Aristotle is credited with the remark that "in order to think, we must speculate in images". While I might disagree with this as a universal rule, it is true that very often we need to present graphs derived by mathematical analysis in a visually effective manner. Very few ecologists would, I feel, be attracted to the use of formal languages as suggested by Culik and Maurer (1978) although these might have considerable advantages in describing developmental sequences.

Given a complete dissimilarity matrix, we seek some method of effectively portraying its structure to identify departures from a simple gradient situation. If possible the technique applied should be robust and rapid as befits an exploratory method. Keil and Gutwin (1992) have discussed some approximating graphs and related work is presented by Klauer (1989) and Orth (1988) but here I shall concentrate on "plexus" methods.

McIntosh (1973) surveyed a number of techniques which he termed "plexus methods" but which are mathematically graph representations. In fact the idea had been used much earlier, especially in Russia. For example, Vasilevich (1967) used a graph representation as a mechanism for clustering stands, while de Vries (1953) clustered species. Essentially these methods use plexus representations (graphs or networks) to display dissimilarity (or correlational) relationships between stands or between property descriptors, both species and environmental variables. I shall for convenience refer to stands throughout, in which case the procedure is as follows.

1 Calculate for all pairs of stands some measure of dissimilarity. This gives a complete graph with every stand connected to every other.

2. Select a subset of the linkages between stands

3. Use this subset as a basis for drawing a network representation of the whole matrix of values.

4. We may further wish to examine inter-relationships between several plexus representations; for example, Barkman (1965) gives a separate plexus for each of vascular plants, mosses, lichens and algae, higher fungi and Carabid beetles and relationships between these would clearly be of some interest. One possibility is to combine the graphs and develop a hypergraph representation (Godehardt and Herrmann 1988) but there are also formal methods though they will not be discussed further here.

There are obviously three sets of problems here.

- We must construct a dissimilarity measure appropriate to the problem and to the data description.
- We must identify the interesting subset of dissimilarities which should be represented; i.e., we establish those pairs of stands which are regarded as "linked" in the plexus diagram.
- We must actually draw the network, preferably in a way which represents both the presence of the linkage in the subset and its magnitude.

This is not the place for an extended discussion on the nature of dissimilarity and its measurement. My primary concern is with the selection of the set of dissimilarities to be represented and a wide range of possibilities exists. In some circumstances, the selection can be imposed using external criteria, which provides an analogue of analysis of variance when coupled with various data-based significance testing procedures (see e.g., de Vries et al. 1954, and Gimingham 1961). More often we must rely on internal criteria to establish the set of linkages to be displayed. Which set depends on the objectives of the analysis.

Expectations and objectives

If our stands are indeed ordered along a gradient, then we might expect low dissimilarity between adjacent stands. Indeed, if the assumptions of CA were fully met the minimal spanning tree (hereafter MST) would simply be a linear, unbranched sequence! We might even use the largest eigenvalue of the incidence matrix (which is the binary matrix showing the existence of edges between stands) as a measure of the degree to which the data do NOT meet the assumptions (cf. Murtagh 1983).

In any case, if we select the set of small dissimilarities and draw the associated graph we should see most stands linked to their neighbours in the gradient sequence. This suggests that approaches concentrating on small dissimilarities should be most appropriate and indicates that such methods as determining maximal spanning trees (Agarwal et al. 1992) or most or least uniform trees (Camerini et al. 1986) are unlikely to be helpful, whereas min-max

trees (Camerini 1978) which minimise the single maximum value might be valuable as an alternative to the MST. I shall concentrate here on very simple procedures using the dissimilarity values, but there are other possibilities, such as using angles formed by successive triple of stands in the presumed gradient (see Eppstein 1992 and Keil and Gutwin 1992).

However, we might also ask that selected subsets adequately represent the entire dissimilarity matrix instead of concentrating only on local conditions. For environmental variables it may be true that even large dissimilarity values can be regarded as well-estimated and that the information they supply should be used to develop the representation. Thus, just as noted earlier for ordination (Dale 1975), there are two different classes of plexus model available. These are distinguished by the use in one case of a global Euclidean model to obtain well-separated pair decomposition (WSPD) of a set of points which implicitly involves all dissimilarities, while in the other case a Riemannian view is taken and emphasis is placed locally on small dissimilarities. In the present paper I shall look at both possibilities, since they are not necessarily mutually exclusive.

Concentration on small dissimilarities is also a characteristic of correlation. To say that two variables are (positively) correlated implies that samples which have values which are "close" with respect to one variable are more likely than chance to have values which are "close" with respect to the second variable. (Friedman and Rafsky 1979, 1983, Critchlow 1985). However, correlation does NOT require that samples which are "distant" with respect to one variable are more likely than chance to have values which are "distant" with respect to the second variable. All the standard coefficients of correlation conform to this definition.

Euclidean or global approach

The Euclidean approach involves a global approximation. Let d_{ij} be the original dissimilarity and let a_{ij} be the approximation obtained from our selected subset. This approximation is made by calculating the shortest path between i and j using only links which are included in the selected subset. We seek to make the selection so as to minimise some function relating d_{ij} and a_{ij}. Various criteria can be used to assess the relationship between d_{ij} and a_{ij}. Three such are[2]:

- $\Sigma\,\Sigma\, f(|d_{ij} - a_{ij}|)$ leading to least squares or similar criteria, such as is used in metric scaling. Least Squares is also used by de Soete (1988) when fitting additive trees.
- Maximise $\tau(d_{ij},a_{ij})$ where τ is some measure of rank correlation such as is used in nonmetric scaling.
- $t*d_{ij} = a_{ij}$ where t is some selected multiplier and we are asking that no individual dissimilarity is extremely distorted.

SplitsTree. In the present paper I shall examine two methods only. One is a generalisation of additive or Steiner trees due to Dress et al. (1996; see also Bandelt and Dress 1992) called "SplitsTree". A tree is simply a graph such that all stands form leaves and are connected by some path but without loops. An additive tree (also known as a Steiner tree) adjusts the lengths of the edges in a tree so that the shortest path-length between any 2 leaves provides give a good approximation to the original dissimilarity between that pair of items. "Splits-Tree" further generalises the additive tree to form a graph where this seems desirable; i.e., extra linkages may be added to a tree structure which necessarily form loops if this allows the approximation of dissimilarities to be improved. The rationale is as follows:

In any tree if any link is broken, then two subsets of the nodes are formed. If we break a link A we obtain sets A^+ and A^-. Similarly if we break the tree using link B we get sets B^+ and B^-. Now if we examine the intersection sets formed by $(A^+ \leftrightarrow B^+)$, $(A^+ \leftrightarrow B^-)$, $(A^- \leftrightarrow B^+)$ and $(A^- \leftrightarrow B^-)$ we find that one of these sets is empty. This is, in fact, a sufficient characterisation of a tree.

What "SplitsTree" does is to relax this condition and replace it with some other, weaker, conditions. For example, we might demand only that the intersections of one of the triples, such as $(A^+ \leftrightarrow B^+ \leftrightarrow C^+)$ need be empty. This permits some parts of the representation to diverge from a tree form.

Spanners. The second method, due to Althöfer et al. (1993), uses the third criterion given above, based on a threshold or stretch value t. This determines tolerable errors in the following way. Any observed dissimilarity, d_{ij} whose approximation in the graph has a length exceeding (d x stretch) will cause a new edge to be inserted in the graph between nodes i and j. Thus, for the final graph we can say that no observed dissimilarity d is represented by a path which is longer than (stretch x d). Obviously small

2 Other methods of generalising trees exist, including Diday and Bertrand's (1986) "pyramids", Al Ayouti's (1992) additive forests and Mulder and Schrijver's (1979) Median graphs and Helly hypergraphs but I shall not consider these here.

values of stretch result in a closer approximation to the dissimilarities, at the expense of an increase in the number of edges in the graph.

Cai (1994) has shown the NP-completeness of finding the minimum such spanner, but the approximations seem quite good. One algorithm for computing an approximation to this is a simple modification of a MST algorithm. The dissimilarities are first sorted into ascending order. Starting from the smallest, a link between i and j is entered in the approximating graph if the dissimilarity calculated by tracing a path through the extant links in that graph exceeds the original dissimilarity by the threshold i.e., $t^*d_{ij} = a_{ij}$, or equivalently $t = a_{ij}/d_{ij}$.

Riemannian or local approach

The Riemannian approach (cf. Dale 1994) differs from the Euclidean in that it is not concerned with overall global approximation at all. Instead it concentrates on small dissimilarity values, and seeks to identify a set of neighbours for each stand. Links between near neighbours form the basis of the graph. Note that this emphasis on neighbours will avoid many of the problems associated with bad estimation of dissimilarities due to absence of species.

d-neighbours and k-neighbours. The problem is to select a suitable set of neighbours, and there are many ways in which this might be done. One obvious method is simply to choose a threshold and accept all edges associated with dissimilarities which are less than this value; this defines *d*-neighbours. In some cases we might use a significance test to establish the critical value. For cases where this is not possible, Deichsel (1980) suggested a procedure for identifying a suitable threshold value.

A second approach makes use of nearest neighbour relationships. Williams (1980) suggested using 2-neighbours and found this useful, although later extending this to many neighbours; this defines *k*-neighbours. Simply using the k nearest neighbours can often lead to a disconnected graph; i.e., it is impossible to get from some stand to some other through graph links. Using only the nearest neighbours, for example, such disconnection will occur whenever two stands are mutually nearest neighbours.

Trees and tesselations. To overcome this disconnection, it is common to employ the MST, which is simply the shortest tree which connects all stands - and of course a nearest neighbour classification as well. An early exponent was Falinski (1960). The MST contains only *n*-1 edges which is perhaps somewhat sparse for our present purposes and several generalisations have been proposed, the Relative Neighbourhood graph (Toussaint 1980) and the Gabriel graph (Gabriel and Sokal 1969). In fact all of these are subgraphs of the Delaunay or Dirichlet tessellation (see Aurenhammer 1991), and these (and more) are themselves subgraphs of the γ-graph (Veltkamp 1992). The Delaunay tessellation solves a problem raised by O'Callaghan (1974) who required that neighbours did not obscure each other. Ash and Bolker (1986) have further generalised the Delaunay tessellation, incorporating various weightings dependent on the nature of the interactions and the simultaneous or successive incorporation of stands into the system, while Levcopoulos and Lingas (1989, see also Vaidya 1991) have found that there are other planar graphs which can well approximate observed values, yet are not much more complex than the MST.

The stands linked in such a tessellation are termed "natural neighbours" and possess a variety of optimality properties. Somewhat surprisingly, Dobkin et al. (1990) have shown that this tessellation, although based on local neighbourhoods, is also a good approximation to the complete graph and hence is, in a sense, a global solution as well. The major problem with using this tessellation is that the computational cost increases rapidly with dimensionality (but see Vaidya 1991 for some possibilities), and this is true of the many other tessellations in the literature. For the purposes of this paper I have therefore examined only *k*-neighbours.

Multiple MST's. Generalisation of the MST is the most obvious way in which to reduce the sparseness of its sampling of edges. There is, however, another method deriving from the MST which was suggested by Friedman and Rafsky (1979) precisely for the purpose of sampling the small dissimilarities. If we extract a MST using a given dissimilarity matrix, all those links which appear in the tree can have their values in the original matrix replaced by infinities. We can now extract a second MST orthogonal to the first in the sense that no linkage can appear in both. We can now set links for the second tree to infinite values and extract a third tree, and so on for (*n*-1) trees. The first few such trees will effectively sample all the small dissimilarity values, although we might wish to increase the number of orthogonal trees as the number of stands grows larger.

Before concluding this description of procedures one other must be briefly noted they was used in examining the exemplary data. Williams et al. (1991) described a simple procedure for approximately ordering a sequence of species when the sequence of stands was already known. This is based on a ranking model, but it does permit the identification of two anomalous cases, where the species are bimodally distributed (i.e., occur at both ends

of a "gradient" but not in the centre) or alternatively are very common and hence, of necessity, placed centrally.

Data and analyses

To illustrate the procedures I shall make use of data collected by Dr. L. Mucina. It reports floristic descriptions of 22 stands of chalk grassland in Slovakia, with 46 species present in total. It was believed to represent a gradient and was ordered appropriately; such belief may well be based on extra-floristic information unavailable to me. However, I have accepted that the stands are arranged in a suitable order, but suspected that the species were not well arranged.

The data were examined using both Euclidean and Riemannian plexus methods, specifically a spanning graph approximation of the dissimilarity matrix, the "SplitsTree" procedure, multiple MSTs and multiple neighbour graphs.

The dissimilarity measure used was Williams (1973) partitioned information coefficient for frequency data, which unites qualitative and quantitative differences between stands. Assume that each sampling unit has been subdivided into N subunits and that we are considering a group of g such units, in s of which a species k occurs. Let the total number of subunits in which the species occurs be a of the available gN. The likelihood for species k is then given by

$$L_{\max} = (s/g)^{Ns}\,((g-s)/g)^{N(g-s)}\,(a/Ns)^{a}\,(Ns-a)/Ns)^{Ns-a}$$

which can be converted to an information measure

$$I_c = N[g \lg (g) - s \lg (s) - (g - s) \lg (g - s)] + [Ns \lg (Ns) - a \lg (a) - (Ns - a) \lg (Ns - a)]$$

where $\lg(x)$ is some convenient logarithmic function usually to base 2 or base e. We can sum over species to obtain a measure of homogeneity with the group of g units and then use the change in information consequent on fusing two such groups to represent the dissimilarity between them.

From this, it can be seen that the first part of the right-hand side is concerned solely with the presence of the species in the units, while the second part is concerned with its abundance within subunits. With ordered category data, such as cover-abundance codes we simply take the highest code value to be the number of subunits per unit.

In all cases it seems likely that some desirable properties might be emphasised by appropriate choice of matrix-iterative calculations in the manner of Kendall (1971). It is, for example, possible to ensure that the dissimilarity matrix is Euclidean or to emphasise serial structure. These techniques were not applied in this first examination

Results

Table 1 shows the presence/absence form of the data after the species have been re-ordered according to the Williams et al. (1991) procedure. The rank order relationship between Mucina's original ordering and the re-ordered species list is highly significant (τ_b = 0.58, $p<<0.001$), although perhaps less high than we might have hoped for.

If we examine the relationship between original and constructed ordering, shown in Fig. 3, there are two areas of difference. At the start of the sequence some species are moved to unexpected positions, notably species 9 to 15 of the Mucina ordering which are moved to the start of the w-b-c sequence. At the end of the sequence there seems to be a confusion, or even an absence, of the ordering. These differences suggest that there may have been some slight misordering of the stands at the start and that the end of the sequence is not part of any sequence. Rare species are the commonest aberrant form, although ubiquitous and bimodal distributions are both present. One species, *Tithymalus cyparissias*, might be regarded as bimodal or ubiquitous depending on the precise decision criteria - w-b-c used a minimal separation of five steps along the gradient as sufficient to infer bimodality, but this number was arbitrarily chosen.

Spanning graphs for the stands were constructed for a variety of stretch values, ranging from 2 to 5 and a few are shown in Fig. 4. A simple graph is obtained with value of 5 (Fig 4a) and 3 (Fig 4b). Further trials suggested that a

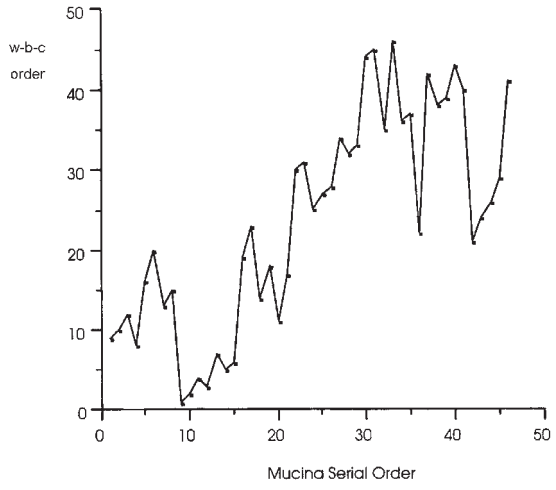

Figure 3. Comparison of Mucina and Williams-Bunt-Clay ordering of species along the presumed stand gradient.

Table 1. Mucina Data re-ordered according to the Williams-Bunt-Clay criterion. **Bold** indicates probable ubiquitous species. *Italic* indicates probable bimodal species. ***Bold italic*** indicates rare species which have an uncertain position. **1** = present, **.** = absent.

Stand label																						
									1	1	1	1	1	1	1	1	1	2	2	2	
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	1	
1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Jovibarba glabrescens
1	1	1	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Sedum album
1	.	1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Draba lasiocarpa
1	1	1	1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Allium montanum
1	.	1	.	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Pilosella bauhinii
.	.	1	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Poa badensis
1	.	1	1	1	1	1	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	Dianthus lumnitzeri
1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.	.	.	.	.	.	Fumana procumbens
1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.	.	.	.	.	.	Globularia aphyllanthes
1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.	.	.	.	.	.	Linum tenuifolium
.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	***Carex caryophyllea***
.	.	.	.	.	.	1	1	1	1	.	.	.	.	.	.	.	.	.	.	.	.	Scabiosa canescens
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.	.	.	Leontodon incanus
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.	.	.	.	Potentilla arenaria
1	.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	.	.	Teucrium montanum
1	1	1	1	1	1	1	1	1	1	1	1	1	.	.	.	1	1	1	.	.	.	Thymus praecox
.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	***Euphrasia stricta***
.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	***Juniperus communis***
1	.	1	.	1	.	1	.	1	1	.	.	1	1	.	.	1	.	1	1	.	.	**Seseli elatum**
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	.	**Festuca pallens**
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**Poterium sanguisorba**
.	.	.	.	.	.	.	.	.	1	1	.	1	1	1	1	1	1	1	1	1	1	Plantago media
.	.	.	.	1	1	1	.	1	1	1	1	.	1	.	.	1	.	.	1	.	.	Trinia glauca
1	.	.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**Helianthemum nummularium**
.	.	.	.	.	.	.	.	.	1	.	1	1	1	1	.	.	.	.	.	.	.	Pilosella officinalis
.	.	.	.	.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Carex humilis
.	.	.	.	.	.	.	.	.	.	1	.	1	1	.	.	.	.	.	.	.	.	Polygala brachyptera
.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	***Rosa sp.***
.	.	.	.	.	.	1	.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Genista pilosa
.	.	.	.	.	.	.	.	.	1	1	1	1	1	1	1	1	1	1	1	.	1	Linum catharticum
.	.	.	.	.	.	.	.	.	.	.	.	1	1	1	.	1	1	1	.	.	.	Anthyllis polyphylla
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	***Taraxacum laevigatum***
.	.	.	.	.	.	.	.	1	.	.	.	.	1	1	.	1	.	.	1	.	1	*Taraxacum vulgare*
.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	1	.	1	.	.	Plantago lanceolata
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	***Brachypodium pinnatum***
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	***Acosta rhenana***
.	.	.	.	1	.	.	.	.	.	.	1	1	1	1	1	1	1	1	1	1	1	Biscutella austriaca
.	.	.	.	.	.	.	.	.	.	.	.	1	1	1	1	1	1	1	1	.	.	Teucrium chamaedrys
1	1	.	.	1	1	1	1	.	.	.	1	1	1	1	1	1	1	1	1	1	1	**Tithymalus cyparissias**
.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	1	.	1	.	.	.	1	Galium austriacum
.	.	.	.	.	.	.	.	.	1	1	.	.	.	1	.	1	1	1	1	1	1	Sesleria albicans
.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	1	1	1	1	1	1	1	Hypochaeris radicata
.	.	.	.	1	.	.	.	.	.	.	.	.	1	1	1	1	1	1	.	1		*Hippocrepis comosa*
.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	1	1	1	1	1	1	Seseli annuum
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1		Koeleria macrantha
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1		Campanula moravica

value of 2.75 was at or about the position where a simple graph first appeared and such a graph is shown in Fig. 4c. It can be seen to be planar (no lines cross) and appears to be composed of 3 major clusters. Examining this and the other simple graphs does suggest that there is an initial sequence but that the later stands are much less clearly aligned in order. Even with a stretch value of 5, which is a very liberal tolerance, stands 13-20 do not form anything akin to a simple sequence. In contrast, with a value of 2 a rather complex graph is formed; Fig. 4d, based on the Hamiltonian cycle, gives a reasonably clear representation of a complex structure.

The "SplitsTree" results appear to confirm this interpretation. In Fig. 5 in order to reduce overwriting of labels, the graph is drawn with all links of equal length; i.e., without regard for the actual dissimilarity values associated with the links. It is clear that the early stages of the sequence (stands 1-9) are not in a simple sequence, but this pales into insignificance compared with intermediate stands 10-13 and still more the later stands 14-22! Both these groups come closer to clusters than sequences. This explanation would be more convincing if the graph were a better fit to the data; the additive tree accounts for some 16% of the variation, and the SplitsTree graph for around

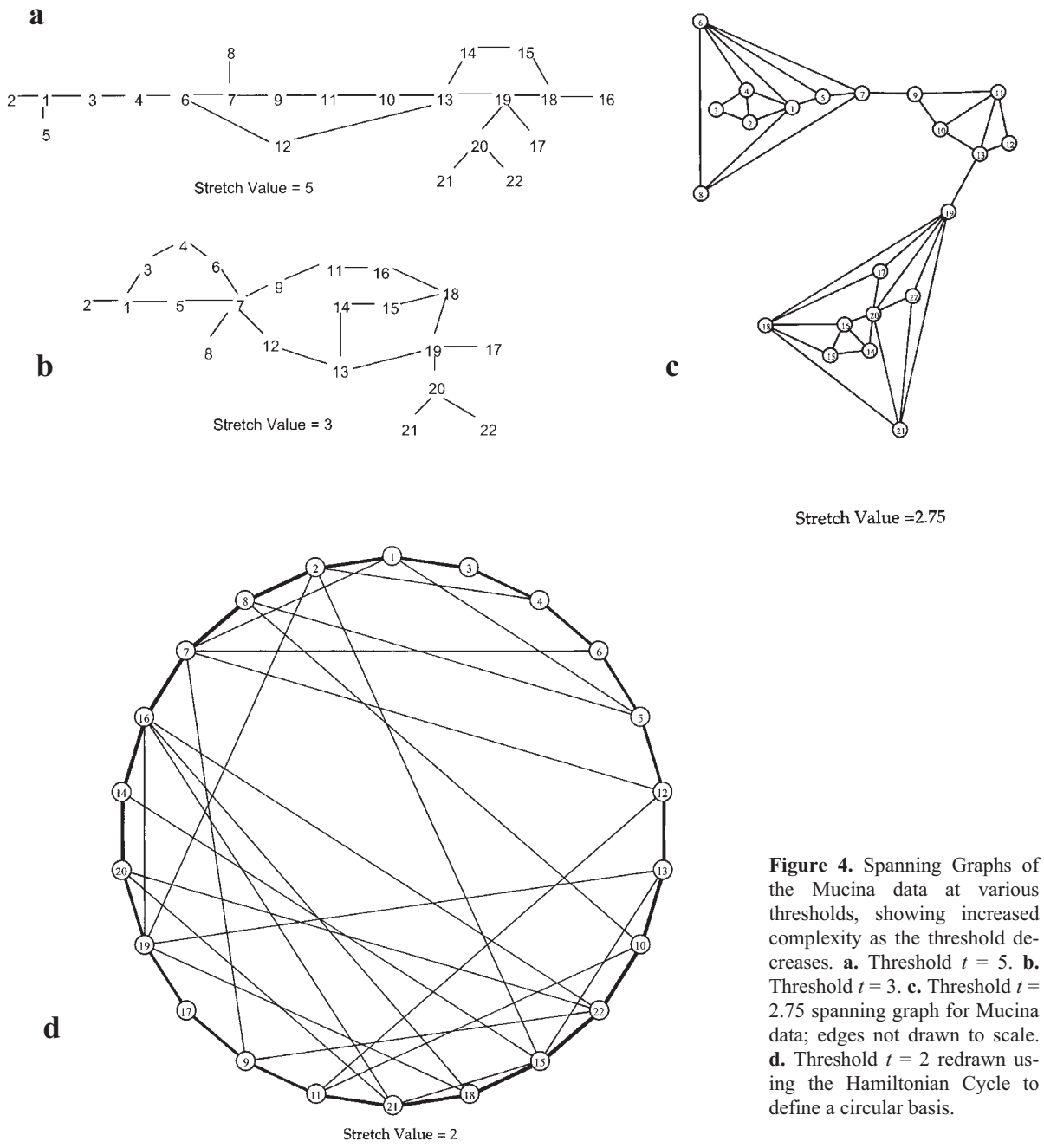

Figure 4. Spanning Graphs of the Mucina data at various thresholds, showing increased complexity as the threshold decreases. **a.** Threshold $t = 5$. **b.** Threshold $t = 3$. **c.** Threshold $t = 2.75$ spanning graph for Mucina data; edges not drawn to scale. **d.** Threshold $t = 2$ redrawn using the Hamiltonian Cycle to define a circular basis.

28%, though note the improvement in fit as we move away from a tree.

When we draw the links to scale (Fig. 6) the result is a clear separation into 2 clusters, neither of which is particularly associated with any simple sequence. The global methods are of course dominated by the larger dissimilarity values which makes the apparent lack of fit of somewhat less important since these are the most likely values to be ill-estimated. In summary, the global methods seem to suggest that the early part of the sequence is acceptable but that the later stands do not fall into a simple linear sequence.

Turning to local methods, three orthogonal MSTs were extracted, and these are presented in Fig. 7. Overall, in Fig. 7a there does seem to be a reasonable approximation to a gradient interpretation, although the first tree shows a less than perfect sequence for the later stands. Obviously the MST will not provide a good fit globally (it is known that nearest neighbour clustering is usually more distorting than furthest neighbour for example; see Hubert and Schultz 1975) and might be expected to em-

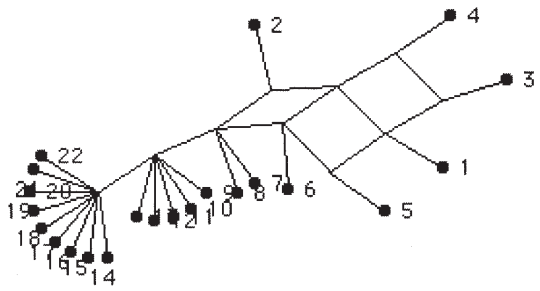

Figure 5. "SplitsTree" graph for Mucina data; edges not to scale.

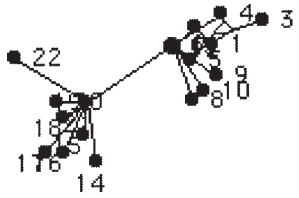

Figure 6. "SplitsTree" graph for Mucina data; edges to scale.

phasise any serial structure, for a common problem with this method concerns "chaining" (e.g., Wishart 1969 specifically designed a method to reduce such effects). However looking at Fig. 7b immediately suggests that 2 clusters are present with a common connection through stand 13. The graph does not seem to be planar although I have not submitted it to a formal test of planarity.

Turning to multiple neighbours the graph resulting from examining the 3 nearest neighbours is shown in Fig. 8. The major disjunction into two clusters is obvious and there appears to be another possible disjunction between stands 7 and (9, 12). As usual the earlier stands can be regarded as a somewhat malleable gradient, but this interpretation does not seem to be acceptable for the later stands.

Discussion

Reliance on dissimilarities

CA and its derivatives directly employ the observed performance values for species in stands, whereas the plexus methods used here are based on dissimilarities between stands. How much do we lose by changing from

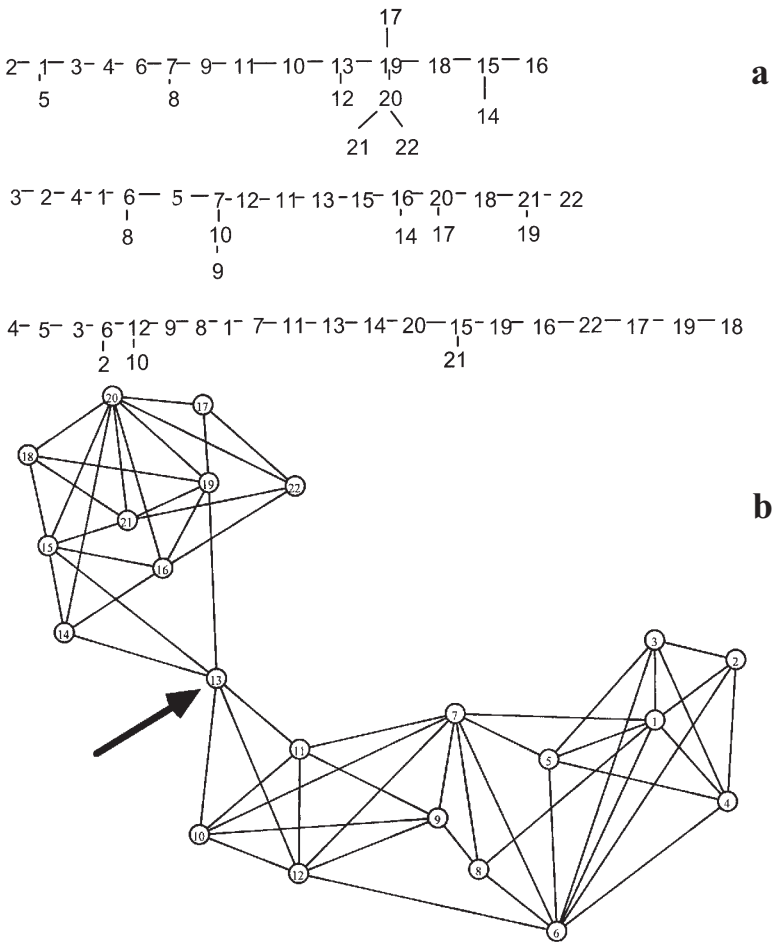

Figure 7. Multiple Orthogonal Minimal Spanning Trees for the Mucina data. **a.** the separate trees, **b.** the composite graph.

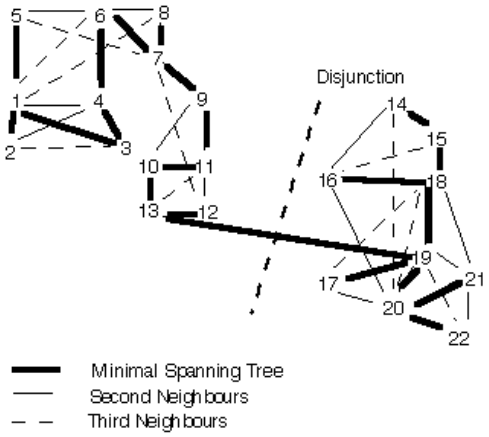

Figure 8. Multiple 3- neighbour Graph; Mucina data.

2-mode data on stands and species to 1-mode involving only dissimilarities? Maa et al. (1996) have shown that dissimilarities are quite powerful in analogues of analysis of variance, which suggests that not much information is actually being lost. Additionally the ability to choose a dissimilarity measure allows the user to emphasise or de-emphasise various aspects of the relationships between stands. The question is similar to that of selecting a particular performance measure, such as projected cover, density or biomass, in that each particular measure will have advantages and disadvantages. Indeed the chi-square distance which underlies CA may not be a particularly useful measure of the relationships which we wish to examine.

Gradients: for and against

While the data set shows some signs that a gradient might exist in some sections, this is not a clear indication, even when the original author has accepted a sequential structure as appropriate. Any gradient present seems to be one involving clusters of stands rather than individual stands and I suggest the data more strongly support separation into at least 2 clusters, one of which might internally be sequentially arranged.

One supposed advantage of CA is that each species is assigned a location and spread along the presumed gradient. The strict model is even stronger in that only the location of the curves differs, spread and spacing being equal. The results obtained here suggest that these values are not a great deal of use, since the sequential bell-shaped curve response model is unlikely to be apposite. However, nonmetric scaling is less likely to be affected and can cope with the implied higher dimensionality.

This leaves the question of when a gradient model might be chosen in preference to any other. It is clear that the choice will be scale dependent - the smoothing effect of using larger sampling areas will smooth local discontinuities. A referee has (correctly) suggested that "the choice should depend on the ability of the method to capture and summarise reality concisely" and not on any preconception that vegetation is continuous or formed from disjoint classes. Apart from some philosophical problems with 'reality' - I would prefer to substitute 'observed data' - this is true but it avoids the question of how this might be done! There is a distinct lack of methods for making such evaluation and so far as I am aware no phytosociological applications. Good fit is not sufficient by itself, for a more complex model can always fit data better. It is necessary to balance the complexity of the model against its adequacy of fit, as is done in the minimal message length proposals of Allison and Wallace (1994), Edgoose and Allison (1999) and Wallace and Dowe (2000).

Euclidean or Riemannian

Both Euclidean (global) and Riemannian (local) approaches seem to be effective. However, unless users are prepared to define the structure which they wish to fit and adopt Hubert and Arabie's (1994) methodology, the single best representation would appear to be the multiple neighbour graph. This could well be a salutary experience for those who unthinkingly have adopted a gradient methodology. Theoretically, the Riemannian approach is to be preferred as it does not rely on (possibly badly estimated) large dissimilarity values. Furthermore, it seems that a stretch factor circa 3 is sufficient to permit effective analysis, and avoid the threshold problems associated with re-estimation procedures such as "step-across" (Bradfield and Kenkel 1987)

Methodological

In choosing between the Spanner and "SplitsTree" graphs, I would argue that as a visual representation the "SplitsTree" graphs are probably more interesting and useful. However there are a great number of other methods for obtaining global representations and it would be foolish to accept either of the two demonstrated here as optimal.

Choosing between MSTs and Multiple Neighbours is equally difficult, and again there are a large number of alternative techniques which need further examination. But these two methods are computationally tractable,

whereas many of the alternatives, such as Dirichlet tessellations and Gabriel graphs, are less attractive in high dimensionalities. Overall I believe that the MST is likely to overemphasise sequential properties at the expense of more complex structures and therefore I would opt for the multiple neighbour method.

It should be added that clustering is a useful means of reducing complexity - any short enough interval can be regarded as flat. Using clusters does not necessarily mean accepting discontinuity in the representational space, although methods for assessing dissimilarity in such case are now available (Chatterjee and Narayanan 1992). By assembling the clusters into sequences and networks we can gain a greater insight into the structure of the vegetation without being overly influenced by sampling problems.

Conclusion

The main conclusion here is that prior to applying CA with its implicit assumption of a simple gradient underlying the vegetation patterns, it is desirable to first determine if such a model is pertinent. To do otherwise is to be Procrustean, forcing the data to fit our preconceptions. Plexus methods permit a visual appreciation of the dissimilarities while being easily appreciated and used. My suggestions on the basis of limited experience would suggest examining a graph of multiple nearest neighbours to see if a single sequential ordering is really acceptable.

As with all exploratory methods, they do not answer all the questions. In particular, if two gradients are appended into a single "vegetation performance" sequence examination of the vegetation dissimilarities alone is unlikely to reveal the possible confounding of environmental gradients.

Acknowledgments: To Laco Mucina for permission to use his grassland data. To my anonymous referees for their suggestions and patience with an ill-prepared manuscript.

The SplitsTree program is available as a shareware program for Macintosh. Contact huson @ mathematik.uni-bielefeld.de or download from
ftp://ftp.uni-bielefeld.de/pub/math/splits/splitstree2.

All other programs were written by myself in Language Systems FORTRAN for a Macintosh computer.

References

Agarwal, P. K., J. Matousek and S. Suri. 1992. Farthest neighbors, maximum spanning trees, and related problems in higher dimensions. *Comput. Geom.: Theory and Appl.* 4:189-201.

Al Ayouti, B. 1992. New forms of graphical representation in data analysis: additive forests. *Proc Conf. Distancia*, Rennes.

Allison, L. & C. S. Wallace. 1994. The posterior probability distribution of alignments and its application to parameter estimation of evolutionary trees and to optimisation of multiple alignments. *J. Molecular Evolution* 39:418-430.

Althöfer, I., G. Das, D. Dobkin, D. Joseph and J. Soares. 1993 On sparse spanners of weighted graphs. *Discrete Comput. Geom.* 9:81-100.

Ash, P. F. and E. D. Bolker. 1986. Generalized Dirichlet tessellation. *Geometriae Dedicata* 20: 209-243.

Aurenhammer, F. 1991. Voronoi diagrams - a survey of a fundamental geometric data structure. *A. C. M. Computing Surveys* 23: 345-405.

Austin, M. P. 1976. On nonlinear species response models in ordination. *Vegetatio* 33: 33-41.

Austin, M. P. 1990. Community theory and competition in vegetation. In: J. B. Grace and D. Tilman (eds.), *Perspectives on Plant Competition.* Academic Press, San Diego. pp. 215-238.

Babad, Y. M. & J. A. Hoffer. 1984. Even no data has value. *Commun. Assoc. Comput. Mach.* 27: 748-756.

Bandelt, H-J. and A. W. M. Dress. 1992. A canonical decomposition theory for metrics on a finite set. *Adv. Math.* 92: 47-105.

Barkman, J. J. 1965. Die Kryptogamenflora einiger Vegetationstypen in Drente und ihr Zusammenhang mit Boden und Mikroklima. In: R. Tuxen (ed.), *Biosoziologie, Ber. Symp. Int. Ver. Vegetskunde*. Stolzenau/Weser 1960, pp. 157-171.

Beals, E. W. 1973. Ordination: mathematical elegance and ecological naivete. *J. Ecol.* 61: 23-35.

Birks, H. J. B., S. M. Peglar and H. A. Austin. 1996. An annotated bibliography of canonical correspondence analysis and related constrained ordination methods 1986-1993. *Abstracta Botanica* 20: 17-36.

Bradfield, G. E. and N. C. Kenkel. 1987. Nonlinear ordination using flexible shortest path adjustment of ecological distances. *Ecology* 68: 750-753.

Cai, L. 1994. NP-completeness of minimum spanner problems. *Discrete Appl. Math.* 48: 187-194.

Camerini, P. M. 1978. The min-max spanning tree problem and some extensions. *Inform. Proc. Lett.* 7: 10-14.

Camerini, P. M., F. Maffioli, S. Martello and P. Toth. 1986. Most and least uniform spanning trees. *Discrete Applied Math.* 15: 81-187.

Chatterjee, S. and A. Narayanan. 1992. A new approach to discrimination and classification using a Hausdorff type metric. *Austral. J. Statist.* 34: 391-406.

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

Culik II, K. and H. A. Maurer. 1978. String representations of graphs. *Internt. J. Computer Math. Sect.* A 6: 272-301.

Dale, M. B. 1975. On the objectives of ordination. *Vegetatio* 30: 15-32.

Dale, M. B. 1994. Straightening the horseshoe: a Riemannian resolution? *Coenoses* 9: 43-53

De'ath, G. 1999. Principal curves: a new technique for indirect and direct gradient analysis. *Ecology* 80: 2237-2253.

de Soete, G. 1988. Tree representations of proximity data by least squares methods. In: H. H. Bock (ed.), *Classification and Related Methods of Data Analysis.* North Holland, Amsterdam, pp. 147-156.

de Vries 1952. Objective combination of species. *Acta Bot. Neerl.* 1: 497-499.

de Vries, D. M., J. P. Baretta and G. Haming. 1954. Constellation of frequent herbage plants based on their correlation in occurrence. *Vegetatio* 5/6: 105-111.

Deichsel, G. 1980. Random walk clustering in large data sets. *COMPSTAT 1980*. Physica-Verlag, Vienna pp. 454-459.

Diday, E. and P. Bertrand. 1986. An extension to hierarchical clustering: the pyramidal presentation. In: E. S. Gelsema and L. N. Kanak (eds.), *Pattern Recognition in Practice*. Elsevier Science, Amsterdam. pp. 411-424.

Dobkin, D., S. J. Friedman. and K. J. Supowit. 1990. Delaunay graphs are almost as good as complete graphs. *Discrete Comput. Geom.* 5: 399-407.

Dress, A. W. M., D. H. Huson and V. Moulton. 1996. Analyzing and visualizing sequence and distance data using "SplitsTree". *Discrete Applied Math.* 71: 95-109.

Duckworth, J. C., R. G. H. Bunce and A. J. C. Malloch. 2000. Vegetation-environment relationships in Atlantic European calcareous grasslands. *J. Veg. Sci.* 11: 15-22.

Edgoose, T. and L. Allison. 1999. MML Markov classification of sequential data. *Statistics and Computing* 9: 269-278.

Eilertson, O., R. H. Økland, T. Økland and O. Pederson. 1989. The effects of scale range, species removal and downweighting of rare species on eigenvalue and gradient length in DCA ordination. *J. Veg. Sci.* 1: 261-270.

Eppstein, D. 1992. The farthest point Delaunay triangulation minimizes angles. *Computational Geometry Theory and Applications* 1: 143—148.

Escofier, B., H. Benali and K. Bachar. 1990. Comment introduire la contiguité en analyse des correspondances? Application en segmentation d'image. Rapport de recherche de l'INRIA - Rennes, RR-1191, 22 pages - Mars 1990.

Falinski, J. 1960. Zastosowanie taksonomii wroclawskiej do fitosocjologii. *Acta Soc. bot. Pol.* 29: 333-361.

Famili, A. and P. Turney. 1991. *Intelligently Helping Human Planner in Industrial Process Planning.* AIEDAM 5: 109-124.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth. 1996. From Data Mining to Knowledge Discovery. In: U. Fayyad et al. (eds.), *Advances in Knowledge Discovery and Data Mining.* AAAI/MIT Press, Menlo Park, CA, pp. 1-34.

Friedman, J. H. and L. C. Rafsky. 1979. Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* 7: 697-717.

Friedman, J. H. and L. C. Rafsky. 1983 Graph-theoretic measures of multivariate association and prediction. *Ann. Statist.* 11: 377-391.

Gabriel, K. R. and C. L. Odoroff. 1984. Resistant lower rank approximation of matrices. In: E. Diday, M. Jambu, L. Lebart, J. Pagés, and R. Tomassone (eds.), *Data Analysis and Informatics III.* North Holland, Amsterdam. pp. 23-30.

Gabriel, K. R., C. L. Odoroff and S. Choi. 1988. Fitting lower dimensional ordinations to incomplete similarity data. In: H. H. Bock (ed.), *Classification and Related Methods of Data Analysis*. Elsevier- North Holland , pp. 445- 454.

Gabriel, K. R. and R. R. Sokal. 1969. A new statistical approach to geographical analysis. *Syst. Zool.* 18: 54-64.

Gimingham, C. H. 1961. North European heath communities: a network of variation. *J. Ecol.* 49: 655-694.

Godehardt, E. and H. Herrmann. 1988. Multigraphs as a tool for numerical classification. In: H. Bock (ed.), *Classification and related methods of Data Analysis.* Elsevier, North Holland. pp. 219-229.

Goodall, D. W. and R. W. Johnson. 1982. Non-linear ordination in several dimensions: a maximum likelihood approach. *Vegetatio* 48: 197-208.

Goodall, D. W. and R. W. Johnson. 1987. Maximum likelihood ordination: some improvements. *Vegetatio* 73: 3-13.

Hill, M. O. 1973. Reciprocal averaging: an eigenvector method of ordination. *J. Ecol.* 61: 237-249.

Hubert, L. and P. Arabie. 1992. Correspondence analysis and optimal structural representations. *Psychometrika* 56: 119-140.

Hubert, L. and P. Arabie. 1994. The analysis of proximity matrices through sums of matrices having (anti-)Robinson forms. *Brit. J. Math. Statist. Psychol.* 47: 1-40.

Hubert, L. and J. Schultz. 1975. Hierarchical clustering and the concept of space distortion. *Brit. J. Math. Statist. Psychol.* 28:121-133.

Hubert, L. and J. Schultz. 1976. Quadratic assignment as a general data analysis strategy. *Brit. J. Math. Statist. Psychol.* 29:190-241.

Huisman, J., H. Olff and L. F. M. Fresco. 1993. A hierarchical set of models for species response analysis. *J. Veg. Sci.* 4: 37-46.

Ihm, P and H. van Groenewoud. 1975A multivariate ordering of vegetation data based on Gaussian type gradient response curves *J. Ecol.* 63: 767-777.

Karadžić, B. and R. Popović. 1994. A generalized standardization procedure in ecological ordination: test with Principal Components Analysis. *J. Veg. Sci.* 5: 259-262.

Keil, J. M. and C. A. Gutwin. 1992. Classes of graphs which approximate the complete Euclidean graph. *Computational Geometry* 7: 13-28.

Kendall, D. G. 1971. Seriation from abundance matrices. In: F. R. Hodson, D. G. Kendall and P. Tautu (eds.), *Mathematics in the Archaeological and Historical Sciences*. Edinburgh Univ. Press. pp. 215-252.

Klauer, K. C. 1989. Ordinal network representation: representing proximities by graphs. *Psychometrika* 54: 737-750.

Levcopoulos, C. and A. Lingas. 1989. There are planar graphs almost as good as complete graphs and about as cheap as the minimal spanning tree. Proc. Internatl. Symp. Optimal Algorithms. *Lecture Notes in Computer Science* 401, Springer, Berlin. pp. 9-13.

Maa, J-F, D. K. Pearl and R. Bartoszyński. 1996. Reducing multidimensional two-sample data to one-dimensional interpoint comparisons. *Annals Statistics* 24:1069-1074.

McIntosh, R. P. 1973. Matrix and plexus techniques. In: R. H. Whittaker (ed.), *Ordination and Classification of Communities.*Dr. W. Junk, den Haag. pp. 157-191.

Minchin, P. R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69: 89-107.

Mulder, H. M. and A. Schrijver. 1979. Median graphs and Helly hypergraphs. *Discrete Mathematics* 25: 41-50.

Murtagh, F. 1983. A probability theory of hierarchic clustering using random dendrograms. *J. Statist. Comput. Simul.* 18: 145-157.

Naga, R. A. and G. Antille. 1990. Stability of robust and non-robust principal components analysis. *Comput. Statist. Data Anal.* 10: 169-174.

Naouri, J-C. 1970. Analyse factorielle des correspondances continues. *Publ. l'Inst. Statist. Univ. Paris* 19:1-100.

O'Callaghan, J. 1974. An alternative definition for the neighbourhood of a point. *I. E. E. E. Trans. Comput.* C-24: 1121-1125.

Oksanen, J. and R. R. Minchin. 1997. Instability of ordination results under changes in input data order: explanations and remedies. *J. Veg. Sci.* 8: 447-454.

Orth, B. 1988. Representing similarities by distance graphs: monotone network analysis (MONA). In: H. H. Bock (ed.), *Classification and Related methods of Data Analysis*. North Holland, Amsterdam. pp. 489-496.

Posse, C. 1995. Tools for two-dimensional exploratory projection pursuit. *J. Computer Graphics Statist* . 4: 83-100.

Taguri, M., M. Hiramatsu, T. Kittaka and K. Wakimoto. 1976. Graphical representation of correlation analysis of ordered data by linked vector pattern. *J. Jap. Statist. Soc.* 6: 17-25.

Tamassia, R. and I. G. Tollis. 1995. Graph Drawing. DIMACS Internatl. Workshop, Princeton 1994. *Lecture Notes in Computer Science* 894, Springer, Berlin.

Tausch, R. J., D. A. Charlet, D. A. Weixelman and D. C. Zamudio. 1995. Patterns of ordination and classification instability resulting from changes in input data order. *J. Veg. Sci.* 6: 897-902.

ter Braak, C. J. F. 1986. Canonical correspondence analysis a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67: 1167-1179.

Toussaint, G. T. 1980. The relative neighbourhood graph of a finite planar set. *Patt. Recog.* 12: 261-268.

Vaidya, P. M. 1991. A sparse graph almost as good as the complete graph on points in K dimensions. *Discrete Comput. Geom.* 6: 369-381.

Van Groenewoud, H. 1992. The robustness of Correspondence, Detrended Correspondence and TWINSPAN analysis. *J. Veg. Sci.* 3: 239-246.

Vasilevich, V. I. 1967. A continuum in the coniferous and parvifoliate forest of the Karelian isthmus. *Bot. Zhur. SSSR* 52: 45-53 (in Russian).

Veltkamp, R. C. 1992. The γ-neighbourhood graph. *Computational Geometry: Theory and Applications* 1: 227-246.

Wallace, C. S. 1995. Multiple factor analysis by MML estimation Tech. Rep. 95/218, Dept Computer Science, Monash University, Clayton Victoria 3168, Australia 21 pp.

Wallace, C. S. & D. L. Dowe. 2000. MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. *Statistics and Computing* 10: 73-83.

Williams, W. T. 1973. Partition of information: the CENTPERC problem. *Austral. J. Bot.* 21: 277-281.

Williams, W. T. 1980. TWONET: A new program for the computation of a two-neighbour network. *Austral. Comput. J.* 12: 70.

Williams, W. T., J. S. Bunt and H. J. Clay. 1991. Yet another method of species-sequencing. *Marine Ecol. Prog. Ser.* 72: 283-287

Wishart, D. 1969. Mode analysis: A generalisation of nearest neighbour which reduces chaining effects. In: A. J. Cole (ed.), *Numerical Taxonomy*. Academic Press, New York. pp. 282-308.

Yanai, H. 1988. Partial correspondence analysis and its properties. In: E. Diday, C. Hayashi, M. Jambu and N. Ohsumi (eds.), *Recent Developments in Clustering and Data Analysis*. Academic Press, New York and London. pp. 259-266.