COENOSES 9(1): 43-53, 1994
© C.E.T.A., Gorizia, Italy

STRAIGHTENING THE HORSESHOE: A RIEMANNIAN RESOLUTION?

M. Dale

Strie Pty. Ltd., 10/217 Hawken Dr., St. Lucia, Qld., 4067, Australia.

Keywords: Dissimilarity, Horseshoe effect, Nonmetric multidimensional scaling, Ordination, Riemannian space.

Abstract. In this paper I examine some of the causes of the "horseshoe" effect in ordination, and some of the proposals made to mitigate or remove it. I then introduce a Riemannian unfolding methodology whereby the horseshoe is embedded in the curvature of the space itself. By using monotonic regression methods the curvature of this space need not remain constant. This means first, that the nonlinear responses of species are incorporated into the curvature of the space itself, and second, that differential sampling densities along a gradient can be accepted without excessive distortion of the results.

Although much more evaluation is necessary, it seems that the use of a Riemannian space might avoid many of the problems presently found with ordination methods. It also raises several questions of fundamental importance in vegetation science.

Introduction

Ordination is one of the most commonly applied multivariate techniques in numerical vegetation science. Yet, to adopt a cynical view, it seems that users of such methods all too often sample along a marked environmental gradient, apply their preferred technique, recover a somewhat twisted version of the gradient and proclaim the analysis successful! Rarely, an attempt may be made to estimate rates of change in composition or to utilise the modal positions of the species in some predictive manner.

Ordination methodology has, of course, been subject to considerable review concerning the choice of the best method (Anderson 1971, Fasham 1977, Clymo 1980, Minchin 1987), as well as more specific reviews and comments on currently popular methods (Sherman 1972, Olsson 1979, Arabie & Soli 1982, Marquand 1982, Wartenberg, Ferson & Rohlf 1987, Hubert & Arabie 1988, Naga & Antille 1990, van Groenewoud 1992). In fact, for a rather popular multivariate technique, there seems to be considerable dissatisfaction with present performance. Kenkel & Orlóci (1986) concluded that none of the currently available ordination strategies is appropriate under all circumstances and specifically indicate that detrending could lead to distortion. This conclusion is supported by Oksanen (1988). But when are they appropriate? To answer that question we need some evaluation functions.

Wilson (1981) pointed out that there are at least 2 different properties involved in evaluating ordinations.

1. The accuracy with which the ordination recovers the "true" species or stand order. Wilson estimates the true ordering using a maximum likelihood method based on a non-null rank model, which involves no Euclidean assumptions.

2. The consistency with which it performs, which is essentially a measure of the risk of obtaining aberrant results. Wilson estimates this by crossvalidation, using the variance of replicate ordinations based on subsets, although a bootstrap method might be preferable.

In Wilson's example, detrended correspondence analysis was more accurate but less consistent than principal component analysis; perhaps a multivariate analogue of the nursery rhyme *"and when she was good she was very, very good, but when she was bad she was horrid"*. However, Wilson's proposals are notorious for their absence from subsequent ordination studies!

Where we have a single gradient, then some modest success is likely, in that we can hope to estimate the locations of stands and/or species along the gradient with some semblances of accuracy and precision. With more than a single gradient the results become more entangled. This is hardly surprising! The notion of ordering, or seriation is strictly a one-dimensional one. This can be easily seen when attempting to estimate the median, perhaps the quintessential order statistic. It is well defined in one-dimension, but there are at least 4 competing definitions in multidimensional space, each with its own effective definition of order (cf. Small 1990)[1]. It can easily be seen that the minimal spanning tree in one dimension is simply the sorted data so there is a case to be made for the minimal spanning tree as an extension of the concept of ordering in more than one dimension. Friedman & Rafsky (1983) make use of this in defining their non-

---

[1] Gilula & Ritov (1990) comment that even stochastic ordering is best understood in the univariate case, and that multivariate extensions, such as that of Sampson & Whitaker (1989) did not appear useful as a means of extending correspondence analysis.

parametric multivariate t-test analogue and, as we shall see later, the minimal spanning tree provides a basis for one of the more successful methods of unbending the horseshoe.

Very often an expected gradient appears in a phytosociological ordination as a curve, commonly known as a horseshoe. Various reasons exist for this occurrence, but a primary problem is that, if the β diversity is large, many of the pairwise dissimilarities may be poorly estimated. Estimates of dissimilarity for pairs of stands which share few or no species can be extremely inaccurate, for our estimates of species performance are truncated at 0! A zero value can mean either that a species is absent but could survive in the stand, or that it is impossible for the species ever to survive. As W. T. Williams once remarked, some noughts are noughtier (naughtier) than others.

The precise effect of such lack of commonality in species complement depends on the coefficient in use. In some cases pairs of stands could appear to be extremely similar when they share few or no species. In other cases the dissimilarities between such pairs of stands may be large but non-discriminating; the value indicates only that the stands are indeed markedly different. Thus the value of the dissimilarity we obtain as our estimate is not a secure guide to the reliability of that estimate.

However, all the methods have another feature in common which we have so far ignored. They all seek a projection of the data into a Euclidean space. Even Johnson & Goodall (1979; see also Goodall & Johnson 1982,1987), who made a valiant attempt at incorporating the non-linear relationships, albeit as Gaussian curves approximated by quadratic functions, expect their results to be Euclidean in nature. But what is it that is so definitive about a Euclidean solution? In this paper I wish to examine another possibility which has some attractions, a negatively curved Riemannian space. But first I want to look briefly at methods by which the horseshoe effect can be mitigated, modified, removed or accepted.

Mitigation Methodologies

I shall discuss the various methodologies under the following headings:

- Grin and Bare It
- How Many Miles to Babylon?
- Hammer and Tongs
- Jus Primus Noctis
- Coupling
- Cohabiting
- Miscegenation
- Winnowing

Grin and Bare It

Feoli & Feoli Chiapella (1980) have argued that the appearance of a horseshoe should not be regarded as a total catastrophe! It should, they suggest, pose no insuperable problems provided the user is aware of the existence of a nonlinear response. Furthermore, they point out that misguided attempts to remove a non-existent horseshoe may prove more damaging than accepting an existent one. For it is not always the case that the second axis is a quadratic function of the first, the third a cubic function, and so on, if only because this assumes that there is only a single interesting axis anyway. I have myself found examples where the 1st and 2nd axes were essentially independent but the 3rd and 4th were quadratically related, the 3rd with the 1st, the 4th with the 2nd. Application of any unbending methodology should clearly be selective and not routinely applied without consideration, a point made by Kenkel & Orlóci (1986) and reinforced by Jackson & Somers (1991). However, if the user determines that such a quadratic response is present, then the interpretation can be adjusted. While this can hardly be regarded as a strategy for unbending a horseshoe, it is useful to remember that to do nothing is not necessarily bad.

How Many Miles to Babylon?

The most obvious solution to the problem of badly estimated dissimilarities is to change the estimation procedure to obtain a better (combination of) dissimilarity measure(s). We can then possibly employ a parametric scaling method, such as Gower's (1966) principal coordinates analysis, giving a very thorough analysis. Although most, if not all, of the dissimilarity measures employed in vegetation studies can be encompassed within the general concept of a Hausdorff distance (Hausdorff 1927; Bednarek & Ulam 1979), they can be usefully divided into 2 or 3 conceptual types. These are the difference metrics and the intersection metrics, with correlations providing a possible third category of shape metrics.

In difference metrics we use functions such as

$$\sum_{k=1}^{p} \mathrm{abs}\,(x_{ik} - x_{jk})^{\tau}$$

which are then scaled or further modified. The dissimilarity is clearly a function of the difference of corresponding values in the description vector. A typical such measure would be Gower's (1971) metric, where the difference for each variable is scaled by its range.

With intersection metrics we employ instead functions of the form

$$\sum_{k=1}^{p} \min\,(x_{ik}, x_{jk})$$

which are scaled and then complemented to convert them from similarity to dissimilarity. Typical measures here would be the Sørensen (1948) measure or Lance & Williams (1966, 1967) Canberra metric with double-zero suppression, although more complex forms can be found (Dale 1989a). The difference between intersection and difference metrics is most easily seen with binary data where it corresponds to the inclusion, or exclusion of double-zero matches. If we examine 2 pairs of values (1,2) and (1,7) we can see that in both cases the contribution to the intersection metric would be 1, whereas the contribution to the difference metric would be a function of 1 or 6 respectively.[2]

However it is also necessary to consider the scale of measure of the data themselves. In many cases phytosociologists have used cover-abundance scales which avoid the over-precise appearance, the tedium and probably some of the associated noise, which direct measurement involves. These scales are strictly ordered category scales and subtraction is not an appropriate operation, since arithmetic operations are undefined for such measures. In such cases the Gower metric is inappropriate. However Critchlow (1985) has shown that several intersection metrics can be derived using the Levenshtein minimal mutation approach. Dale, Coutts & Dale (1988) used this to provide a landscape classification.

There are other measures which might be preferable in certain circumstances. Thus Dale, (1989a) describes a measure suitable for the ordered category data typically collected by phytosociologists. Another example is Calhoun's (Bartels *et al.* 1970, see also Dale 1989b) "betweenness" measure, which relies only on the order properties of the observed data. In essence it counts the number of other points which lie between the pair whose distance is being estimated and as such it is likely to be sensitive to sampling intensity. Like Goodall's (1966) probabilistic coefficient it is solely a function of the specific data set being examined.

Shape metrics are more commonly used in inverse classification of species, although others have been suggested (e.g., Austin & Belbin 1982). Most take the form of correlation coefficients which are generally based on products, although some rank correlations are based on differences. Wilson, Partridge & Sykes (1990) recommend use of the Cole/Hurlbert C_8 coefficient as a means of addressing certain types of rarity. However, they remark:

> *none of the methods give a classification that is consistently intuitive ... None give a classification that agrees completely with the autecological and distributional information available.*

If there is an expectation that stands and species should be referable to the same space, recommending different coefficients and different analyses for stands and species is not going to accomplish this. However many analyses result in the stands being represented by points, and the species by vectors, with logical relationships to distance and correlation measures respectively.

Selecting a dissimilarity measure will almost always involve some compromise (cf. Gordon 1990). It would seem sensible to determine the scale of measure of the data, and perhaps the components of similarity which the user regards as important and select a measure within these constraints. Using nonmetric scaling methods avoids great reliance on the magnitude of the dissimilarity values, since these methods rely only on their order. Faith, Minchin & Belbin's (1987) results do suggest that many coefficients do get the order approximately correct, although with considerable noise.

Robustness can also be maintained in scaling by regarding the zero values as missing (Spence 1982). This reduces the maximal number of dimensions which can be safely estimated from our data. As a rule of thumb we require about 2.5-3 times as many data values as parameters to estimate. For m sites and n species in k dimensions we have k(m+n) parameters to estimate. Most phytosociological data tables seem to have no more than 20% nonzero entries, so we should have $mn/5 \geq 3k(m+n)$. This gives as an estimate of the maximal dimensionality in an unfolding of $k \leq mn/15(m+n)$. We may seek less but we should certainly not try to seek more.

With ordered category data even this is unnecessary since the 0 values are simply another category whose ordinal position is clear. However, these scales remain folded. Only the largest value has a unique position since the other values could represent positions on an underlying axis either above or below. Thus, however careful our choice and our consideration of the properties of the data, it is unlikely that any direct measure of similarity will be successful in situations of high β diversity where such folding is of greatest importance. It would seem more sensible to avoid calculating such measures entirely and regard the abundances of species in stands as direct estimates of the similarity of species and stand. Such data can then be analysed using multidimensional unfolding techniques.

Hammer and Tongs

The second, and most popular, method of addressing the horseshoe is to attempt to flatten out the ordination. Two procedures have been suggested for this unbending of horseshoes, or detrending as it is more commonly known. The first was developed by Hill & Gauch (1980) for correspondence analysis and involves a smoothing operation within consequent fragments of the presumed gradient. Hill's procedure, while somewhat *ad hoc*, appears to work quite well on occasion, even if the second axis produces no curvature.

The second, due to ter Braak (1985) and again for correspondence analysis, attempted to fit polynomial functions of the earlier axes to remove quadratic and higher order effects in later axes. The problem with ter Braak's method is that the second axis need not contain any quadratic component of the first, while the third or some later axis might do so, or might reflect a polynomial relationship with an independent second axis. I should add that it is not at all clear what the mathe-

---

[2] The normalisation adopted can also have a marked effect. The Bray Curtis (1957) coefficient, given by $\Sigma(\min)/\Sigma(\max)$, would evaluate to 0.22, whereas the Canberra metric, using $\Sigma(\min/\max)$, would give 0.64.

matical nature of the detrended axes is! The first axis is an eigenvector of the data matrix, but after detrending the nature of the subsequent axes is decidedly obscure.

Provided there is only a single gradient, either method might be expected to perform well enough, since then they do nothing; we are after all in an exploratory mode of analysis and it does not behove us to be too particular![3] But if there is a quadratic effect then it will contaminate the first axis anyway by compressing the ends, and neither method does anything to alleviate this. In practice ter Braak (pers. commun.) has admitted that the results obtained by his procedure were inferior to those given by Hill's method.

Correspondence analysis does make use of the unknowable zero values, and can be extremely fragile if there are outlying values. I have seen cases where the first 11 latent roots were all unity, giving trivial axes in such cases. Hill & Gauch (1980) downweight rare species for precisely this reason; the weighting is so effective the axes disappear altogether! Interestingly Eilertson *et al.* (1990) modified the weighting scheme when attempting to estimate β-diversity from detrended correspondence analysis results.

More robust algorithms can be defined with greater resistance to outliers. Gabriel & Odoroff (1984; see also Emerson & Wong 1978) propose a method using medians in the reciprocal averaging algorithm, and this gives results very similar to the logarithmic transformation proposed by Ihm & van Groenewoud (1975). But using medians is not without problems. For rare species, that is those which occur in less than 50% of the stands, the median will always be zero and the species effectively contributed nothing to the analysis. Worse, if any stand has less than half the total number of species then it too will have a median of zero; with much vegetation data this would have disastrous consequences! Gabriel & Odoroff suggest that any other percentage point might be used, for example the 95%le, but this simply shifts the problem to the more rarefied species and the very species poor stands. In selecting a percentile the analyst is in effect deciding at what level of rarity a species becomes unimportant. The use of trimmed means to provide robustness meets similar problems. It would be possible to use weighted medians but this has not been examined in practice.

Jus Primus Noctis

There are 2 alternative proposals for incorporating some specific aspects of the purported species responses into the model which have had phytosociological application; Johnson & Goodall's (1979) maximum likelihood method, together with its later polishings (Goodall & Johnson 1982, 1987), and the Parametric mapping method of Shepard & Carroll (1966), used by Noy-Meir (1974). Some of the more flexible nonmetric scaling techniques might also be included here (see e.g., Holman 1978, Friedman, Tukey & Tukey 1980).

Johnson & Goodall argue throughout that the species response is Gaussian; by this they mean that the envelope of the plot of abundance against position on a gradient is bell-shaped. This shape is symmetric, which seems unlikely even if convenient. They approximate the shape by a quadratic, which ignores the tails of the distribution entirely, while separately estimating a presence/absence component. My own limited tests of their program have had little success, at quite large cost in computer time.

The same can be said of parametric mapping, which has problems avoiding local minima, though it can be effective in one dimensional cases. The problems here may in part be due to the use of least squares criteria since these are known to be highly sensitive to outlying observations. Since the shape of the response curve is not prescribed in parametric mapping, it might be profitable to examine more robust estimation methods, along the lines of Heiser's (1987) use of absolute deviations

The explicit representation of the response of the species does show a major problem with using any dissimilarity-based method. As noted above, any particular performance value for a species corresponds to 2 different positions along the gradient, with the exception of the unique maximum if it exists. Obviously more complex response curves could permit more than 2 positions to be present. We therefore require techniques of analysis which will cope with this ambiguity, so that rather than scaling methods we should choose unfolding methods.

Coupling

The suggestion here is to combine several measures representing intersection and difference components of dissimilarity. Probably the most obvious method of combination is simply to average the measures, and Noest & van der Maarel (1989) have suggested just this, adding a Bray-Curtis measure to what is effectively a Gower metric component. They further suggested that a weighted mean might be more appropriate, although they indicate no mechanism for establishing the values for the weights. In fact such summation had been earlier suggested by Weir (1970) who used it to equalise the contribution of blocks of attributes in numerical taxonomy. Each block was weighted by the reciprocal of the number of attributes in it or by some user-specified value. Tversky (1977) and Krumhansl (1978) have also examined composite dissimilarity measures which could be pertinent as phytosociologically apposite models.

Most dissimilarity measures sum the contributions of species to provide an overall measure, and this is often implicitly a weighted summation - the Canberra metric is a

---

[3] Gilula & Ritov (1990) have examined some problems of inferential statistics in correspondence analysis. It seems likely that bootstrap and permutation methods (see e.g. Biondini *et al.* , 1988) will be useful.

good example of this. It is interesting that in all cases summation has been used, although this is clearly not the only possible means of combining the components. For example the product would have the advantage of being scale-free.[4] Again, using summation is the same as using the mean dissimilarity, and we could argue for using the mode, as more representative, or the median, as more robust, or even the maximum as most pessimistic.

Another possibility is to use one matrix as a source of weights to apply to the other (see e.g., Borg & Lingoes 1980). In essence multidimensional scaling methods seek a set of coordinates such that the distances between stands in the coordinate space is a good approximation to the observed dissimilarities. The measure of goodness in metric analyses is a sum of squared differences, and it is obviously possible to replace this by a weighted sum of discrepancies. Indeed many of the criteria used in the computer programs already include weights of various kinds to normalise the statistics in desirable ways. In our case we could simply use an intersection similarity measure as a weight for the difference measure, thus combining the reliability and difference estimates. While such analyses are possible with available computer programs, I can find no example of their use in the ecological literature. Note that in this approach there is an asymmetry in the use of the 2 measures, since one is used simply to weight the importance of the other.

Cohabiting

Rather than summing the matrices or using one as a source of weights for the other, it is also possible to analyse them jointly. The idea here is that the distance in the recovered ordination space should approximate all the available dissimilarity measures simultaneously. Restriction to 2 matrices is clearly unnecessary. But this does not mean that the compromise ordination so recovered will necessarily be useful. I have at home a postcard showing an elephant and a giraffe surrounded by 'loving hearts' and followed by a curious offspring combining characteristics of both! Compromise can lead to bizarre conclusions!

Methods using multiple matrices include McGee (1968), Jöreskog (1971), Cunningham & Shepard (1974), Holman (1978), Friedman, Tukey & Tukey (1980), Lastovicka (1981), Kakusho & Mizoguchi (1983), and Klauer (1989). Using multiple matrices can introduce complications; e.g., Flury (1987) has provided two generalisations of principal components analysis which do not give identical results. No phytosociological evaluation has yet been made of these to my knowledge.

Somewhat more subtle is the individual differences approach of Carroll & Chang (1970; see also Kenkel & Booth 1987) where the ordination axes are regarded as forming a set common to all the dissimilarity matrices. For each matrix a set of weights indicates the importance of each of the axes for that specific case. In some cases the weight may be zero and then the dimensionality for that matrix is effectively reduced. It is not impossible that each matrix could have only one pertinent axis, and none be shared between them. I cannot find an example using this method in the ecological literature; I hope to provide one later.

A last alternative would be to analyse each matrix independently and then attempt some kind of canonical correlation or Procrustean analysis to rotate the various sets of axes to congruence. This is an analogue of the consensus methods (cf. Nelson 1979, Stinebrickner 1984a,b, Day 1986, Lefkovitch 1978), which have loomed large in cladistic taxonomic studies. The advantages here are, that in deriving the axes for each matrix, we reject some of the information in each as noise. The canonical relationship is then made between axes which have in effect already been designated as important and any noise has, hopefully, been reduced or removed from the system. The disadvantage is that the rejected information might not be entirely noise, and can include the only components shared with the other matrices! But having the results of several individual analyses is often desirable anyway.

Miscegenation

By using multiple matrix methods, we can combine metric and nonmetric analyses, to obtain the strengths of both. Faith, Minchin & Belbin (1987) use a simultaneous nonmetric analysis of the full data using Kruskal's (1964a,b) methodology. They first determine a threshold dissimilarity value, then calculate 2 matrices. The first includes all the data but relies only on the ordinal values of the dissimilarities, that is on their rank order. The second matrix retains the metric information but regards values above the threshold as missing information. They then fit a scaling model simultaneously to both matrices, using a nonmetric criterion with the ordinal data and a metric criterion with the partial metric data. Belbin (1991) further modified this procedure by splitting the data into 2 parts at some threshold dissimilarity, with values below the threshold being analysed metrically, and those above nonmetrically[5]. In both cases we have to decide which elements of the data should be treated metrically. and both procedures apply a single threshold. Belbin (pers. commun.) indicates that the threshold values used were chosen from plots of observed against "true" dissimilarities for simulated data, which is then applied to real situations where the "true" dissimilarities are unknown.

Disregarding the possibility that the simulated data which provided the 'true' values might not be entirely representative of vegetation data, I have previously observed that

---

[4] Summation is not applicable where the variables are incommensurate, whereas the product has no such limitation. This is usually only relevant in vegetation studies where floristic and environmental variables are combined in a single analysis. In general I would not recommend such a procedure since it is in identifying the commonality between these that our substantive interpretational gains are made.

[5] Belbin further introduced an alternative nonmetric procedure based on Guttman's (1968) image approach but we have as yet too little information to decide if this is preferable to Kruskal's.

the assumption that large dissimilarities are badly estimated is only partially true. Combining the two analyses may have the advantage that the metric analysis constrains the non-metric solution so that problems with local optimality and degeneracy may be minimised. Given the general sensitivity of scaling results to the larger distances, which will be represented here by the ordinal values only, Belbin's method is less likely perhaps to find an 'elephant-giraffe' compromise solution since the local and global patterns are segregated. But compromise rarely satisfies everyone.

Winnowing

If the dissimilarity values vary in their reliability then it seems sensible to identify the unreliable ones and replace them with better estimates. We might do this by transformation, as in the logarithmic transforms mentioned in a previous section. If we expect only a single gradient, then we could employ Kendall's (1971) circle-up procedure to emphasise the gradient. Neither transformation explicitly identifies the bad estimates but they will both tend to reduce the effects of large dissimilarities, and these are known to be most influential contributors to the solution (see Graef & Spence 1979).

In contrast Williamson (1978) proposed removing all ill-estimated values from the dissimilarity matrix and replacing them by minimal paths estimated from the remaining values via a minimal path metric. This would generally lead to an overestimate of the unknown values. Presumably, since the original values are generally underestimates and the final values overestimates, combinations of the original and recalculated values might be used with profit. Indeed several minimal path estimates based on different selections of ill-estimated dissimilarities could be combined.

But how to identify the contaminated values? The simplest method is to select a threshold, on the assumption that large dissimilarities are always in error. But wielding a knife on the dissimilarity matrix is a rough and ready business. Is no more sophisticated method available than this somewhat crude surgery? Bradfield & Kenkel (1987) have suggested instead that the number of shared species be used, i.e., that a fixed number of species should be shared before the dissimilarity is acceptable. It seems sensible that the number shared should be normalised by the maximum possible for that pair of stands, and this is just an intersection measure. It would appear that we are interested in combining difference and intersection measures.

Bending the Space or "Lie back and Enjoy It"

All of the previous approaches, except the first, have essentially tried to mitigate the effects of the horseshoe. All of them also seek to represent species-stand relationships in a Euclidean space and it is here we propose to differ. Euclidean space is one possible example of a Minkowski space, specifically one with a power coefficient of 2; to measure distances in Euclidean space requires summing squares of differences. Although Lissitz & Robinson (1977) reported no particular advantages to non-Euclidean solutions, Gregson (1965) found that a Minkowski space with a coefficient of 1 was more appropriate where there were several independent components contributing to the dissimilarity. I would not like to argue for such independence in ecological factors, but the possibility of looking at more general spaces than Euclidean seems attractive. For Euclidean space is a special case of a Riemannian space.

This final methodology accepts that our sampling will be uneven in density along any gradients that may exist and that our estimation of dissimilarity is less than perfect. In general estimation error will tend to place points which are distant too close together. But this is not uncommon in human estimation. Most people do underestimate large distances; after all Paris is just beyond Bilbao, and Copenhagen is only just beyond that again, as seen from Toledo! And human eyesight is such that items much beyond circa 10m are focussed at infinity; our perception of their relationship thereafter is not based on the degree of accommodation, but on binocular vision and semantic clues.

There is a mathematical construct which also has exactly these properties; this is a negatively curved Riemannian space. A Euclidean space is a Riemannian space of zero curvature. It is used primarily, I believe, because we are all comfortable with its geometry which we learn at school. But we are also comfortable with the geometry of positively curved Riemannian space, since we live on the surface of a sphere, the Earth, which is an example of such. Locally there is little noticeable difference between curved and flat space. It is only when we have to deal with large distances that discrepancies become important. So while I do not notice curvature when walking to the local pub, I do notice it when travelling from Australia to Europe and I fly a Great Circle route. Further, a positively curved space can be conveniently embedded in a Euclidean space of one higher dimension. However if we try to map the surface of the Earth onto a flat map, which is of the same dimensionality, we find that we can only preserve some of the properties.

Can we perhaps identify a more formal objective for this minimal path method? Essentially we are saying that we can only estimate small distances precisely, and that in order to obtain estimates of large distances we must sum a series of small distances. This is the basis of the re-estimation procedures outlined above which utilise the spanning tree. Our distances are composed of changes in value for the several species. These changes cannot always be regarded as equivalent ; a change of one unit in species A may not indicate the same distance as an identical change in species B. Furthermore, we know empirically that the changes of the various species are often correlated.

If we have two neighbouring points with coordinates x^i and x^i+dx^i then let us define the distance between them as given by $ds^2=g_{ij}dx^idx^j$ which forms the metric of the space. g_{ij} is called the fundamental tensor of the space. For a Euclidean space all elements of g_{ij} are zero except those for g_{ii} which equal one, i.e., g_{ij} takes the form of the identity matrix. Thus in a Euclidean space all components of the distance measure ds^2 have equal weight and they are uncorrelated. In addition the metric is positive definite, that is the values of ds^2 are either positive or zero. But earlier we noted that we did not want every species to be regarded as equivalent, and that species are indeed correlated. This suggests that a non-Euclidean geometry might be desirable as a means of embedding the relationships between stands which we seek to represent. Such geometries are called Riemannian, and Euclidean geometry is simply one special case of them.

I do not wish to present a detailed discussion of Riemannian geometry which has many properties well known to us from Euclidean geometry. For example there is an analogue to principal components obtained by solving the determinental equation, $|A-\lambda g_{ij}|=0$. We calculate distances by summing the elements ds^2 for adjacent points along a line, but we must introduce an indicator e, with values of +/-1, so that $ds^2=eg_{ij}dx^i dx^j$ is positive (where it is nonzero). The line of minimal length between 2 points is called a geodesic, and there can be several different geodesics between 2 points. The length of the geodesic can include elements ds^2 with positive or negative values of e and others with null distances! There are also differences in the definition of parallelism. However, we shall return to Riemannian geometries later in the paper.

Negatively curved spaces have some peculiar properties as we shall see, and are possibly less familiar. But are they in fact so rare? The human visual system, mentioned above, can be modelled by a negatively curved Riemannian space (Indow 1982). But a negatively curved space can be embedded in a Euclidean space of one greater dimension only if one of the axes is complex; that is has a factor $\sqrt{-1}$. This means that when we use Pythagoras' theorem to calculate distances, we have to subtract these differences[6].

What I am arguing here is that, as the environment becomes more and more remote from the species optimum, the species response will finally decrease to zero; often, but not necessarily, monotonically. No species is concerned with environments beyond its tolerance; such are all effectively placed at the same (infinite) distance even if they are completely different from each other. On the other hand, species are extremely discriminatory in environments which are close to their optimum.

Thus the solution I propose that nothing need be done about the horseshoe. We simply embed the solution in a non-Euclidean space with negative curvature. The horseshoe is then simply a visual representation of the curvature of the space forced into existence because we cannot adequately map a curved space onto a flat, Euclidean map! The complex axis necessary to embed the negatively curved Riemannian space in a Euclidean one represents the correction needed to introduce the horseshoe. The other real axes represent the structure in which we are interested. However, just as the surface of the Earth poses problems in mapping, so do the Riemannian axes here, and we have to choose some appropriate projection.

This embedding is, surprisingly, very simple to add to scaling programs. Briefly these programs are given an observed dissimilarity matrix, and seek to produce a set of coordinates for all the items (stands and/or species) a set of coordinates in a space of the required dimensionality. This set of coordinates is chosen to minimise the discrepancy between the observed dissimilarities and those calculated from the calculated coordinates; the discrepancy measure itself is obtained from a regression relating observed and calculated dissimilarities. If these coordinates represent a Euclidean solution then we simply sum the squared differences between coordinate values. If the coordinate system is regarded as a negatively curved Riemannian space, all it requires is a different way of calculating the distances betweens points based on the coordinates we are estimating. The formulæ involve hyperbolic functions and the curvature of the space. It also involves somewhat special treatment of the peculiar axis whose elements are complex[7]. The distance calculated from the real axes has to be reduced as the differences along the imaginary axis increase. This is precisely the effect we require if our observed dissimilarities are such that large values are underestimated. However, in this case the correction needed is individually provided by the appropriate difference on the complex axis. No arbitrary thresholds need to be invoked.

A metric scaling solution requires an estimate of the curvature of the space, and that this curvature remain constant throughout the space. While we could add an extra parameter to estimate curvature, and Lindman & Caelli (1978) do exactly that, constancy of curvature seems an unlikely property of ecological spaces, be they environmental or niche spaces. We expect it to be negative to cope with the horseshoe, but it is extremely unlikely that our sampling will be equally dense along the horseshoe, nor that we shall sample at equal intervals along the gradient. Both of these certainly affect our estimation of the local curvature of the space. Remember that at this stage we do not know how many gradients, if any, exist in the data.

But if we now introduce a nonmetric analysis as well, we do not have to estimate the magnitude of the curvature! We need only postulate its sign. All it requires is a change of regression procedure used to relate the observed and the derived dissimilarities to some monotone procedure. (I believe the space is now what is known as a Finsler space). It is this Riemannian space which the rest of this paper will explore in a preliminary manner. But permitting a variable curvature also encompasses the possibility of non-monotonic responses by the species! Pieszko (1975) did permit

[6] The complex axis appears in coordinate analysis (Gower 1966) as a latent root of negative sign. Interestingly such roots have been known for some time, but were originally thought to be related to missing values in the data. It seems more likely that they appear when quasi- or pseudo-metric dissimilarities are used, and it is these which Faith, Minchin & Belbin (1987) specifically recommend for phytosociological use.

[7] For a space of constant curvature K with distance in the Riemannian space d_{ij} and in the embedding space δ_{ij}, the relationship is given by $\delta_{ij} = 2/\sqrt{(|K|)}\ \sinh(\sqrt{|K|}\, d_{ij})$. Note that δ_{ij} need not be a metric, although d_{ij} is.

variation in curvature but in such a manner that he could not avoid fitting noise and was forced to admit somewhat arbitrary constraints to obtain an acceptable solution

Practical Considerations.

I do not propose in this review to examine the results of applying such a methodology. I am presently experimenting with a Riemannian unfolding program, but there remain numerical stability problems which make the search procedure used inefficient, and liable to produce degenerate results. There are however some practical problems which need discussion, and I shall use Wallnöfer's (1989) data on oak woods for illustration. This records the abundance of 386 species in 136 stands, using a cover-abundance scale.

We have a general space in which to embed our data. The curvature is such that horseshoe effects should not appear, and the effects of differential sampling along any particular gradient should also be reduced. But what data are we to use? We know that estimation of between-stand dissimilarity is difficult and in any case the values obtained are ambiguous. We avoid those problems by using the raw stand-species data and an unfolding method. The abundance of a species in a stand is taken as a measure of the similarity of stand and species. We can convert this to a dissimilarity by subtracting the values from some maximum value, defined for the entire matrix or for each species independently, by using reciprocals or simply by modifying the regression procedure. I have used the largest value for unbounded measures, or the maximum possible value for bounded measures such as ordered category scales.

Using an unfolding method derived ultimately from that of Poole (1984), and applying it to the transformed data matrix, the space recovered can show both species and stands simultaneously[8]. Some generalisation is possible. First, we need not use all the values in the data matrix, and, in particular, the absences can be regarded as missing values which do not contribute to the solution. However this means that very rare species with fewer occurrences than the dimensionality, will not be well estimated, and should be removed from the analysis. Second, we can still use multiple matrices if these seem appropriate so that combining difference and intersection measures could be done, at the cost of extra computation.

Using Trunk's (1976) procedure to estimate the topological dimensionality (Dale, unpublished) has shown a dimensionality of no more than 3 for most phytosociological data. Occasionally, for example with rain forest data, the dimensionality may be higher, but the estimation procedure becomes increasingly insensitive. In practice the available data after exclusion of zero values is rarely sufficiently profuse to allow estimates of such high dimensionalities. Each new dimension will require an additional number of non-zero values of roughly 2.5-3.0 times the sum of the number of stands and species. This sum, for the Italian data, is initially 516 so we need minimally 1250-1300 values per axis!

If we remove the rarer species then the number of species becomes 225 for 3 dimensions, and 201 for 4, so there are less parameters to estimate (361, respectively 337, per axis) There is also an increase in relative density of nonzero values but this does not have a great effect; even the 4-d reduction has only 4811 nonzero entries, which is 17.6% of the table. I would therefore request a dimensionality of 4, of which one axis, the first, would be complex.

Presentation of the results of such an analysis involves two further difficulties. There is clearly a problem concerned with the presentation of the complex axis. However the complex axis is effectively a mechanism for overcoming the inexactness of our dissimilarity estimations and probably does not have substantive interest. While experience may suggest otherwise, for the moment this axis can just be ignored. There is also a question of how to best project the real axes onto a plane; the problem is similar to the representation of the Earth's surface as a map. We have a choice of equal angle and equal area projections and since we are expecting our environmental variables to form directions in the space, my initial choice would be for an equal angle projection. The complexity of using 3-D representations, can be further reduced by transforming to spherical coordinates essentially ignoring the distance form the origin. The abundance of the species is of less interest than its associates along the same, or close, radii. It must be remembered that an equal angle projection expands the polar regions, and that the left and right sides are adjoining. There are some further complexities because we are in a negatively curved space and the projection formulae are not those used for the positively curved geometry of the Earth's surface.

Having obtained our representation, we shall need to explore the distribution of species on the surface. Testing for significant departures from randomness on a sphere depends on specifying the alternative hypotheses rather precisely (Mardia 1972). It may be possible to test if the species form many rather small clusters on the surface of the sphere, but as yet I have not completed a program for this task. Testing for uniformity is also a possibility (cf Keilson *et al.* 1983); acceptance of uniformity would mean that the environmental space was partitioned equally between all the species, which seems *a priori* somewhat unlikely. Tests on the stand distribution would probably be less helpful, since they would reflect sampling density and the uniformity of the environment. However, they could also reflect successional relationships, hopefully as linear structures.

Relationship with Environmental Variables

Having obtained our axes we would usually like to relate them to some environmental, potentially causal, variables, although often the axes represent vegetation variation such

---

[8] I hope to experiment in future with other methods such as Sixtl's (1973) probabilistic unfolding, as I am presently unhappy with the rates of convergence and stability of the present algorithm.

as increasing richness whose relationship to any environmental variable may be very complex. Grime (1979) distinguishes stresses - growth limitation factors, from disturbances - biomass-destroying factors, but each might result from changes in several variables. If the vegetation is, as is likely, in some process of change, then other phytosociological process and characteristics may be present; e.g., Westman & O'Leary (1986) distinguish 5 components of resilience.

The commonest method of relating ordination results to environmental factors is to regress the axes scores onto the available variables. Often this is coupled to variable selection procedures to try and identify a "best" subset. Disregarding any numerical problems, Frank & Friedman (1993) have shown that the choice of regression method should depend on whether we can assume that we have measured exactly the variables we require, or whether we regard them as surrogates for a complex. If the former then ordinary linear regression and variable subset selection are indeed useful. In the latter case, methods using variable subset selection are NOT appropriate, and ridge regression, partial least squares or principal component regression are to be preferred.

The other approach would be to adopt something similar to ter Braak's (1987) constraint of correspondence analysis by incorporating regression into the iterative cycle, and therefore forcing identification of the vegetation axes with the environmental variables. Obviously this emphasises the relationship to the particular environmental variables available, but it is not appropriate if we believe them to be surrogates for a complex. Furthermore it is not obvious how such constraints can be introduced into the unfolding calculation, which does not resemble reciprocal averaging, although further work on alternative algorithms may make this possible.

Discussion and Conclusions

At the outset it is necessary to emphasise that this approach needs testing on synthetic data to establish its performance under known conditions, before making any too outrageous claims. The program which searches for a solution is so far little tested for accuracy and consistency, so its problems are largely unknown. The method also needs bootstrap or permutation procedures (cf. Biondini *et al.* 1988) to establish confidence limits and provide appropriate significance tests. Yet the simplicity of Euclidean representation should not blind us to its faults and the apparent complexity of a Riemannian solution should not bar its use. I have argued earlier for the Riemannian representation as having properties which more closely approximate those we might expect of vegetation. The complex axis provides a means for incorporating the distortions due to sampling density and inappropriate estimation of dissimilarities without introducing somewhat arbitrary algorithmic activity. It is still possible to use multiple matrix methods and to seek Riemannian solutions for some or all of the matrices.

The suggestion that the species is not the best descriptor for phytosociological investigation is not novel. It is often defined largely by reproductive characteristics, ignoring vegetative properties, it is subject to taxonomic fads of splitting and lumping, the range of variation encompassed differs markedly between different taxa, and it ignores the possibility that the ecological requirements of the plant may vary during its lifetime. All in all it seems a fragile basis on which to build a science of vegetation! Yet this is precisely the assumption of the Braun-Blanquet approach (Westhoff & van der Maarel 1978). Alternatives, such as Gillison's 1988) plant functional attributes, emphasising the functional aspects of morphology and anatomy would seem to be a preferable choice, or the iconic representations of Halloy (1990).

If it is accepted then the clustering of species might well be regarded as an artefact of the projection. It is interesting that the stands do not cluster. Instead they reflect fuzziness, where any stand might contain elements of several types. Unfortunately the standard fuzzy classification methods are not suitable; they insist that all types should exist within the envelope of the stands, and this is obviously not the case here! Most of the species lie peripheral to the stands.

As a final intriguing possibility for future investigation, I should note that Antonelli, Elliot & Seymour (1987) have shown a relationship between negative curvature and stability. So perhaps it is in the realms of negatively curved Riemannian space that we shall find the key to the importance of diversity.

Acknowledgment. To Susi Wallnöfer for listening to the early versions.

References

Anderson, A. J. B. 1971. Ordination methods in ecology. J. Ecol. 59: 713-726.

Antonelli, P. L., Elliott, R. J. & Seymour, R. M. 1987. Nonlinear filtering and Riemannian scalar curvature. K. Adv. Appl. Math.. 4: 1-23.

Arabie, P. & Soli, S. D. 1982. The interface between the types of regression and methods of collecting proximity data. In: R.C. Golledge & J.N.Rayner (eds.) "Proximity and Preference: Problems in the Multidimensional scaling of large datasets", Univ. Minnesota Press, Minneapolis. pp. 90-115.

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.

Bednarek, A.R. & Ulam, S.M. 1979. An integer valued metric for patterns. In "Fundamentals of Computational Theory". Academie-Verlag, Berlin. pp. 52-57.

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

Biondini, M. E., Mielke, P. W. Jr. & Redente, E. F. 1988. Permutation techniques based on Euclidean analysis spaces: a new and powerful statistical method for ecological research. Coenoses 3:155-175.

Borg, I & Lingoes, J. C. 1980. A model and algorithm for multidimensional scaling with external constraints on the distances. Psychometrika 45: 25-36.

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

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

Carroll, J. D. & Chang, J-J. 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckhart-Young" decomposition. Psychometrika 35: 283-319.

Clymo, R. S. 1980. Preliminary survey of the peat-bog Hummell Knowe Moss using various numerical methods. Vegetatio 42: 129-148.

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

Cunningham, J. P. & Shepard, R. H. 1974. Monotone mappings of similarities into a general metric space. J. Mathem. Psychol. 11: 335-363.

Dale, M. B. 1989a. Dissimilarity for partially ranked data and its application to cover-abundance data. Vegetatio 82: 1-12 .

Dale, M. B. 1989b. Mutational and nonmutational Similarity Measures. Coenoses 3: 121-133.

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

Day, W. H. E. 1986 Consensus classification. J. Classif. 3: 183-185.

DeSarbo, W. S., Johnson, M. D., Manrai, A. K., Manrai, L. A. & Edwards, E. A. 1992. TSCALE: a new multidimensional scaling procedure based on Tversky's contrast model. Psychometrika 57: 43-69.

Eilertson, O., Økland, R. H., Økland, T. & Pederson, O. 1990. Data manipulation and gradient length estimation in DCA ordination. J. Veg. Sci. 1: 261-270.

Emerson,J.D. & Wong G.Y. 1978 Resistant nonadditive fits for two-way tables In: E. E. Leamer (ed.) "Specification searches: ad hoc inference with nonexperimental data". Wiley, New York. pp. 67-124.

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

Fasham, M. J. R. 1977. A comparison on nonmetric multidimensional scaling, principal components and reciprocal averaging for the ordination of simulated coenoclines and coenoplanes. Ecol. 58: 551-561.

Feoli, E. & Feoli Chiapella, L. 1980. Evaluation of ordination methods through simulated coenoclines: some comments. Vegetatio 42:35-41.

Flury, B. K. 1987. Two generalizations of the common principal component model. Biometrika 74: 59-69.

Frank, I. E. & Friedman, J. H. 1993 A statistical view of some chemometrics regression tools. Technometrics 35: 109-135.

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

Friedman, J.H., Tukey, J.W. & Tukey, P.A. 1980. Approaches to analysis of data that concentrate near intermediate-dimensional manifolds. In: E. Diday et al. (eds.) "Data Analysis and Informatics". North Holland. pp. 3-13.

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

Gillison, A. N. 1988. A Plant Functional Proforma for Dynamic Vegetation Studies and Natural Resource Surveys. Tech. Mem. 88/3 CSIRO Division of Water Resources, Canberra.

Gilula, Z. & Ritov, Y 1990. Inferential ordinal correspondence analysis: motivation, derivation and limitations. Interntl. Statist. Rev. 58: 99-108.

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

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

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

Gordon, A. D. 1990. Constructing dissimilarity measures J. Classif. 7: 257-269.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis Biometrika 53: 325-338.

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

Graef, J. & Spence, I. 1979. Using distance information in the design of large multidimensional scaling experiments. Psychol. Bull. 86: 60-66.

Gregson, R. A. M. 1965. Representation of taste mixture cross-modal matching in a Minkowski r-metric. Austral. J. Psychol. 17: 195-204.

Grime, J. P. 1979. "Plant Strategies and Vegetation Processes", Wiley, Chichester.

Guttman, L. L. 1968. A general nonmetric technique for finding the smallest co-ordinate space for configuration of points. Psychometrika 33: 469-506.

Halloy, S. 1990 A morphological classification of plants, with special reference to the New Zealand alpine flora. J. Veg. Sci. 1: 291-304.

Hausdorff, F. 1927. "Mengenlehre". W. de Gruyter, Berlin.

Heiser, W. J. 1987. Correspondence Analysis with least absolute residuals. Computational Statistics and Data Analysis 5: 337-356.

Hill, M. O. & Gauch, Jr. H. G. 1980. Detrended correspondence analysis: an improved ordination technique. Vegetatio 42: 47-58.

Holman, E. W. 1978. Completely nonmetric multidimensional scaling. J. Mathem. Psychol. 18: 39-51.

Hubert, H. & Arabie, P. 1988. Relying on necessary conditions for optimization: unidimensional scaling and some extensions. In: H.H.Bock (ed.) "Classification and Related Methods of Data Analysis". North Holland, Amsterdam. pp. 463-472.

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

Indow, T 1982. An approach to geometry of visual space with no a priori mapping functions: multidimensional mapping according to a Riemannian metric. J. Mathem. Psychol. 26: 204-236.

Jackson, D.A. & Somers, K. M. 1991. Putting things in order: the ups and downs of detrended correspondence analysis Amer. Natur. 137: 704-712.

Johnson, R. W. & Goodall, D. W. 1979. A maximum likelihood approach to non-linear ordination. Vegetatio 41: 133-142.

Jöreskog, K. G. 1971. Simultaneous factor analysis in several populations. Psychometrika 36: 409-426.

Kakusho, O. & Mizoguchi, R. 1983. A new algorithm for non-linear mapping with applications to dimension and cluster analysis. Patt. Recog. 16: 109-117.

Keilson, J., Petrondas, D., Sumita, U. & Wellner, J. 1983. Significance points for some tests of uniformity on the sphere. J. Statist. Comput. Simul. 17: 195-218.

Kendall. D. G. 1971. Seriation from abundance matrices. In: F. R. Hodson, D. G. Kendall & P. Tautu (eds.) "Mathematics in the

Archæological and Historical Sciences", Edinburgh Univ. Press, Edinburgh. pp. 215-252.

Kenkel, N. C. & Booth, T. 1987. A comparison of presence-absence resemblance coefficients for use in biogeographical studies. Coenoses 2: 25-30.

Kenkel, N. C. & Orlóci, L. 1988. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecol. 67: 919-928.

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

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

Kruskal, J. B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29: 1-27.

Kruskal, J. B. 1964b. Multidimensional scaling: a numerical method. Psychometrika 29: 115-129.

Kulczynski, S. 1928 Die Pflanzenassoziattionen der Penninen. Bull. Inst. Acad. Pol. Sci. Lettr. Cl. Sci. Mat. Nat. Ser. B. S. 2: 57-203.

Lance, G. N. & Williams, W. T. 1966. Computer programs for hierarchical polythetic classification. ("Similarity Analysis") Comput. J. 9: 373-380.

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

Lastovicka, J. L. 1981. The extension of component analysis to four-mode matrices. Psychometrika 46: 47-57.

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

Lindman, H. & Caelli, T. 1978. Constant curvature Riemannian scaling. J. Math. Psychol. 17: 89-109.

Lissitz, R. W. & Robinson, E. J. 1977. An examination of some factors related to using different Minkowski models in non-metric multidimensional scaling. J. Multivar. Behav. Res. 12:69-73.

Mardia, K.V. 1972. "Statistics of Directional Data". Academic, NY.

Marquand, B. 1982. Recovering the dimensions through Multidimensional scaling, remarks on 2 problems. In: R. C. Golledge & J. N. Rayner (eds.) "Proximity and Preference: Problems in the Multidimensional scaling of large datasets". Univ. Minnesota Press, Minneapolis. pp. 267-271.

McGee, V.E. 1968. Multidimensional scaling of N sets of similarity measures: a nonmetric individual differences approach. Multivar. Behav. Res. 3: 233-248.

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

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

Nelson, G. J. 1979. Cladistic analysis and synthesis: principles and definitions, with a historical note on Adanson's Familes des Plantes (1763-64). System. Zool. 28: 1-21.

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. 1974. Catenation: quantitative methods for the definition of coenoclines. Vegetatio 29: 89-99.

Oksanen, J. 1988. A note on the occasional instability of detrending in Correspondence analysis. Vegetatio 74: 29-32.

Olsson, U. 1979. On the robustness of factor analysis against crude classification of observations. Multivar. Behav. Res. 14: 485-500.

Pieszko, H. 1975. Multidimensional scaling in Riemannian space. J. Mathem. Psychol. 12: 449-477.

Poole, K. T. 1984. Least squares metric, unidimensional unfolding. Psychometrika 49: 311-323.

Sampson, R. A. & Whitaker, L. R. 1989. Estimation of multivariate distributions under stochastic ordering. J. Amer. Statist. Assoc. 84: 541-548.

Shepard, R. N. & Carroll, J. D. 1966. Parametric representation of nonlinear data structures. In: P. R. Krishnaiah (ed.) "Multivariate Analysis". Academic Press, London. pp. 561-592.

Sherman, C. R. 1972. Nonmetric multidimensional scaling: a Monte Carlo study of the basic parameters. Psychometrika 37: 323-355.

Sixtl, F. 1973 Probabilistic unfolding. Psychometrika 38: 235-248.

Small, C. G. 1990. A survey of multidimensional medians. Internatl. Statist. Rev. 58: 263-277.

Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of vegetation on Danish commons. Biol. Skr. K. danske Vidensk. Selsk. 5: 1-34.

Spence, I. 1982. Incomplete experimental designs and multidimensional scaling. In: R. C. Golledge & J. N. Rayner (eds.) "Proximity and Preference: Problems in the Multidimensional Scaling of Large Datasets". Univ. Minnesota Press, Minneapolis. pp. 29-66.

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.

ter Braak, C. J. F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. Biometrics 41: 859-873.

ter Braak, C. J. F. 1987. CANOCO, a FORTRAN program for canonical community ordintion by [partial] [detrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis Version 2.1 TNO Inst. Appl. Comput. Sci., Wageningen, The Netherlands.

Trunk, G. V. 1976 Statistical estimation of the intrinsic dimensionality of a noisy signal collection. IEEE Trans. Comput. C-25: 165-171.

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

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

Wallnöfer, S. 1989. Vergesellschaftung von *Ruscus aculeatus* in dem Flaumeichen- und Hopfenbuchenwäldern in Südtirol. Diplomarbeit, Univ. Wien.

Wartenberg, D., Ferson, S. & Rohlf, F. J. 1987. Putting things in order: a critique of detrended correspondence analysis. Amer. Nat. 129: 434-448.

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

Westhoff, V. & vand der Maarel, E. 1978. The Braun-Blanquet approach. In: R.H. Whittaker, (ed.) "Handbook of Vegetation Science V. Ordination and Classification of Vegetation". Junk, den Haag. pp. 619-726.

Westman, W. E. & O'Leary, J. F. 1986. Measures of resilience: the response of coastal sage scrub to fire. Vegetatio 65: 179-189.

Williamson, M. H. 1978. The ordination of incidence data J. Ecol. 66: 911-920.

Wilson, J. B., Partridge, T. R. & Sykes, M.T. 1990. The use of the Cole/Hurlbert C8 association coefficient in inverse ecological classification. J. Veg. Sci. 1: 3676-374.

Wilson, M. V. 1981. A statistical test of the accuracy and consistency of ordinations. Ecol. 62: 8-12.

Manuscript received: November 15, 1993