A COMPARISON OF NONMETRIC MULTIDIMENSIONAL SCALING, PRINCIPAL COMPONENTS AND RECIPROCAL AVERAGING FOR THE ORDITION OF SIMULATED COENOCLINES, AND COENOPLANES

M. J. R. FASHAM
Institute of Oceanographic Sciences, Wormley, Godalming, Surrey, England

Abstract. The use of nonmetric multidimensional scaling as an ordination method has been studied by the use of simulated coenoclines and coenoplanes. It was found that the method always produced better ordinations than principal components analysis and in most cases better than reciprocal averaging, providing the multidimensional scaling was calculated using a space of the same dimensions as the simulated data. For real data of which the dimensionality is unknown, the minimum spanning tree can provide a useful means of estimating this dimensionality. Nonmetric multidimensional scaling was also less susceptible to distortion of the ordination by single gradients of high beta diversity and two-gradient situations when each gradient is of a different beta diversity. A number of different similarity measures were evaluated for use in conjunction with nonmetric scaling and the cos theta coefficient was found to generally give good results. The use of a log transformation of the species values had a marginal effect while the use of a ‘Manhattan’ distance metric in ordination space produced inferior results.

INTRODUCTION

It is now well established that when the species response to a single environmental gradient is bell shaped (Gaussian) in form then most ordination methods will produce a curvilinear distortion of this gradient in a two-dimensional ordination space (Swan 1970, Noy-Meir and Austin 1970, Gauch and Whittaker 1972a). This effect has also been noticed in studies of the seriation problem in archaeology and has been termed the ‘horseshoe’ effect by Kendall (1971, 1975). This curvature of the ordination into a space of dimensions higher than the true dimensionality of the data has also been observed for two-gradient data (Austin and Noy-Meir 1971, Gauch et al. 1976).

It has been further observed (Gauch and Whittaker 1972b, Whittaker and Gauch 1973) that if the species turnover along a gradient is increased then the distortion becomes more pronounced. There are a number of factors which are thought to contribute to this effect, although their relative importance is still not clear. These are,

1) The response of a species to a physical gradient is usually nonlinear, while most ordination techniques (e.g., principal components) assume a linear model.
2) Most measures of sample similarity are not linear functions of the separation of the samples along the gradient.
3) For high beta diversity, samples widely separated will have few, if any, species in common and most commonly used similarity measures will be insensitive measures of sample separation.

Of these problems (2) has been investigated by Gauch (1973a) who proposed an inverse error function as a measure of similarity, while (3) has been studied by Kendall (1971).

The problem of devising a nonlinear method of ordination has been attempted by Gauch et al. (1974) and Ihm and Groenewoud (1975) who both used a very specific model of vegetation distribution, and by Noy-Meir (1974) who applied Carroll’s continuity analysis (Shepard and Carroll 1966). In this paper, the more widely used technique of nonmetric multidimensional scaling (MDS) will be compared with principal components analysis (PCA) and reciprocal averaging (RA) (Hill 1974), using simulated one- and two-dimensional species distributions (Gauch and Whittaker 1972a, personal communication).

THE EVALUATION OF ORDITION METHODS

The rational evaluation of different ordination methods has been made possible by the development of computer programs for simulating the spatial distribution of species in response to environmental gradients. Thus Swan (1970), Noy-Meir and Austin (1970) and Gauch and Whittaker (1972a) have developed programs for one-dimensional gradients, while Austin and Noy-Meir (1971) have produced two-dimensional programs. The programs of Gauch and Whittaker have been made freely available (Gauch 1973b) and so have been used in the present work for evaluating a number of ordination techniques.

The basic assumptions of the Gauch and Whittaker programs are:

1) The distribution of a species along a gradient takes the form of a Gaussian curve. For the two-dimensional simulation, bivariate Gaussian surfaces are used.
2) The modes of the minor species are assumed to be randomly distributed along an axis, but those of major species are evenly distributed to simulate competition for resources between major species.
3) The maxima of the curves follow a log-random distribution.

4) The standard deviation of the Gaussian curves are themselves assumed to be normally distributed with a mean value inversely proportional to the level of beta diversity (species changeover) along a given gradient. Thus, as the specified level of beta diversity is increased, the species distributions narrow, which in turn increases the species turnover along the gradient.

These assumptions were derived mainly from the study of forest vegetation data but they are likely to apply in many ecological situations.

In a number of studies (Gauch and Whittaker 1972b, Kessell and Whittaker 1976, Gauch et al. 1976) the simulation programs have been used to evaluate a number of different ordination methods. It was found that all the methods produced a curvilinear distortion of the one- or two-dimensional gradients, but that the degree of distortion produced, especially at high values of beta diversity, varied greatly between methods. The most successful was Reciprocal Averaging (RA) which gave good results for beta diversities up to 5 to 10 half-changes (HC) and was rather robust against effects of sampling errors and sample clusters. (A half-change is a measure of beta diversity defined as follows [Whittaker 1972]. A measure of inter-sample similarity is first defined and then the sample separation along the transect is determined, at which this similarity measure falls to half of that expected for two samples at zero distance apart. The reciprocal of this separation as a fraction of the total sample length is defined as the beta diversity in half-change units. In this paper, percentage similarity was used to define the half-change.) A disadvantage of the method is that coordinates along the second ordination axis bear a quadratic relationship to those on the first axis (Hill 1974).

The Bray-Curtis method (Bray and Curtis 1957) was the next most successful but suffered from the disadvantage that the investigator had to define end-point samples which were assumed to be at the extreme ends of a gradient. Gauch and Whittaker (1972b) stated that a good choice of end points can usually be made from the ecological background knowledge of the investigator. However, it is obviously preferable that an ordination technique should not require any subjective input from the ecologist. The worst ordination method was Principal Components Analysis (PCA), which was especially susceptible to increases in beta diversity, often producing ordinations that were meaningless and uninterpretable in terms of the underlying gradient. The method was also much more affected by sample clustering.

Analysis of Nonlinear Data Structures

The problem of the analysis of nonlinear data structures has been well discussed by Shepard and Carroll (1966). They observe that, if there is a strong nonlinear relationship among the variables (species in this case), then the samples will not form an ellipsoidal distribution in multivariate space, but will tend to lie on a manifold of lower intrinsic dimensionality that may twist and turn through the space to give the appearance of filling an ellipsoidal volume. Such structures are often found in the physical sciences as well as ecology and psychology; an example being the familiar temperature-salinity diagram of physical oceanography. As has been discovered by ecologists, the application of PCA to such a structure may produce misleading results.

Shepard and Carroll (1966) used two methods to tackle nonlinear data sets, namely nonmetric multidimensional scaling (MDS) and a method called 'parametric mapping.' The latter method attempts to place samples in an ordination space such that the response curves of the species to the new ordination axes are as smooth or continuous as possible. Noy-Meir (1974) has suggested the term 'continuity analysis' for this method and has applied it with some success to both simulated and real data, where it gave superior results to PCA. However, the method presents some computational problems when applied to large data sets. Nonmetric multidimensional scaling was developed by Shepard (1962a, b), and Kruskal (1964a, b), and Shepard (1974) recently reviewed the state of the art. It has been used in zoogeography by Holloway and Jardine (1968) and Thorrington-Smith (1971) but, apart from Anderson (1971), it does not appear to have been used for ecological ordination although its use has been suggested by Orloci (1973).

The principle behind the method is basically simple although the algorithm required to make the calculations is fairly complex. It is assumed that a matrix of dissimilarities between samples has been calculated and it is required to represent this matrix by means of an ordination or mapping of the samples in a space of m dimensions (where of course m is less than the original number of samples). The basic assumption of MDS is that for a good ordination, there should be a rank-order relationship between intersample dissimilarity and intersample distance in the ordination space. This means that the more similar two samples are, the closer they should be in the ordination space and this can be expressed in mathematical terms by stating that there should be a monotonic increasing relationship between intersample dissimilarity and intersample ordination distance. Note that the exact functional form of this relationship need not be specified and this distinguishes the method from metric multidimensional scaling (Torgerson 1958) and principal coordinates analysis (Gower 1966). Thus the MDS algorithm only uses the rank order of the intersample dissimilarities and not their magnitude. Let us assume that some initial configuration of the samples in the ordination space of a specified dimension has been defined and let the distance in this space between samples i and j be $d_{ij}$. 

ANALYSIS OF NONLINEAR DATA STRUCTURES

The problem of the analysis of nonlinear data structures has been well discussed by Shepard and Carroll (1966). They observe that, if there is a strong nonlinear relationship among the variables (species in this case), then the samples will not form an ellipsoidal distribution in multivariate space, but will tend to lie on a manifold of lower intrinsic dimensionality that may twist and turn through the space to give the appearance of filling an ellipsoidal volume. Such structures are often found in the physical sciences as well as ecology and psychology; an example being the familiar temperature-salinity diagram of physical oceanography. As has been discovered by ecologists, the application of PCA to such a structure may produce misleading results.

Shepard and Carroll (1966) used two methods to tackle nonlinear data sets, namely nonmetric multidimensional scaling (MDS) and a method called 'parametric mapping.' The latter method attempts to place samples in an ordination space such that the response curves of the species to the new ordination axes are as smooth or continuous as possible. Noy-Meir (1974) has suggested the term 'continuity analysis' for this method and has applied it with some success to both simulated and real data, where it gave superior results to PCA. However, the method presents some computational problems when applied to large data sets. Nonmetric multidimensional scaling was developed by Shepard (1962a, b), and Kruskal (1964a, b), and Shepard (1974) recently reviewed the state of the art. It has been used in zoogeography by Holloway and Jardine (1968) and Thorrington-Smith (1971) but, apart from Anderson (1971), it does not appear to have been used for ecological ordination although its use has been suggested by Orloci (1973).

The principle behind the method is basically simple although the algorithm required to make the calculations is fairly complex. It is assumed that a matrix of dissimilarities between samples has been calculated and it is required to represent this matrix by means of an ordination or mapping of the samples in a space of m dimensions (where of course m is less than the original number of samples). The basic assumption of MDS is that for a good ordination, there should be a rank-order relationship between intersample dissimilarity and intersample distance in the ordination space. This means that the more similar two samples are, the closer they should be in the ordination space and this can be expressed in mathematical terms by stating that there should be a monotonic increasing relationship between intersample dissimilarity and intersample ordination distance. Note that the exact functional form of this relationship need not be specified and this distinguishes the method from metric multidimensional scaling (Torgerson 1958) and principal coordinates analysis (Gower 1966). Thus the MDS algorithm only uses the rank order of the intersample dissimilarities and not their magnitude. Let us assume that some initial configuration of the samples in the ordination space of a specified dimension has been defined and let the distance in this space between samples i and j be $d_{ij}$. 

ANALYSIS OF NONLINEAR DATA STRUCTURES

The problem of the analysis of nonlinear data structures has been well discussed by Shepard and Carroll (1966). They observe that, if there is a strong nonlinear relationship among the variables (species in this case), then the samples will not form an ellipsoidal distribution in multivariate space, but will tend to lie on a manifold of lower intrinsic dimensionality that may twist and turn through the space to give the appearance of filling an ellipsoidal volume. Such structures are often found in the physical sciences as well as ecology and psychology; an example being the familiar temperature-salinity diagram of physical oceanography. As has been discovered by ecologists, the application of PCA to such a structure may produce misleading results.

Shepard and Carroll (1966) used two methods to tackle nonlinear data sets, namely nonmetric multidimensional scaling (MDS) and a method called 'parametric mapping.' The latter method attempts to place samples in an ordination space such that the response curves of the species to the new ordination axes are as smooth or continuous as possible. Noy-Meir (1974) has suggested the term 'continuity analysis' for this method and has applied it with some success to both simulated and real data, where it gave superior results to PCA. However, the method presents some computational problems when applied to large data sets. Nonmetric multidimensional scaling was developed by Shepard (1962a, b), and Kruskal (1964a, b), and Shepard (1974) recently reviewed the state of the art. It has been used in zoogeography by Holloway and Jardine (1968) and Thorrington-Smith (1971) but, apart from Anderson (1971), it does not appear to have been used for ecological ordination although its use has been suggested by Orloci (1973).
while the dissimilarity between these samples is defined as $\delta_{ij}$. The samples are then ordered such that their dissimilarities are in ascending order. If the initial configuration happened to represent the samples perfectly, we would find that, when the intersample ordination distances were arranged in the same sample order, they would also be monotonically increasing. Thus if the dissimilarities and distances were plotted on a scatter diagram for this ideal solution, a monotonically increasing curve could be drawn through all the points. In any real case, this ideal solution is unlikely to occur and so the next step in the algorithm is to calculate a monotone increasing curve that gives a best fit to the actual scatter diagram of distances and dissimilarities. This is done by the technique of nonparametric regression (Kruskal 1964b). Thus for each sample pair, $i$ and $j$, the regression produces a value, $\delta_{ij}$, on the regression curve for which the similarity takes the value $\delta_{ij}$. Kruskal then defines a measure of the departure of the configuration from monotonicity called "stress" (3) which is given by the equation

$$S = \left[ \frac{\sum_{i<j} (d_{ij} - \delta_{ij})^2}{\sum_{i<j} d_{ij}^2} \right]^{1/2}. \quad (1)$$

The smaller the value of stress, the closer the relation between dissimilarities and distance approaches monotonicity. The MDS algorithm then attempts iteratively to move the configuration of samples in ordination space in order to minimise the stress. This is done using the minimisation process called the method of steepest descent, the mathematical details of which are described by Kruskal (1964a). If the original matrix had contained similarities rather than dissimilarities, the algorithm is easily modified by specifying that the monotone regression curve be descending rather than ascending.

The choice of an initial configuration gives rise to the main problem of the method. If it is inappropriately chosen, then it is possible for the steepest descent method to become trapped in a local minimum, thereby missing the global minimum. Two solutions have been suggested to this problem. The first is to start with a number of replicate random starting configurations and choose the result that produces the lowest value of stress. However, simulation runs (Shepard 1974) have shown that although the globally minimum solution often appears after 5 starts in some cases as many as 20 to 30 may be required; this obviously requires extensive computer time. The second method is to start with a configuration that will hopefully be close to the global minimum by using coordinates derived from a metric scaling or PCA or RA ordination.

The MDS differs from the PCA in that the solutions in spaces of different dimensions have to be calculated separately. This means that a solution in a given dimensional space is not necessarily a projection of solutions in higher dimensions. It is similar to factor analysis in this respect.

Another interesting possibility of nonmetric scaling is that the metric used to measure distance in the solution space need not be Euclidean. More generally we could use a Minkowski r-metric defined as

$$d_{ij} = \left( \sum_{k=1}^{m} |X_{ik} - X_{jk}|^r \right)^{1/r}, \quad (2)$$

where $X_{ik}, X_{jk}$ are the $k$th coordinates of the $i$th and $j$th sample and $m$ is the number of dimensions of the ordination space. Normal Euclidean distance is obtained using $r = 2$, while $r = 1$ gives the Manhattan or 'city-block' metric. At this point it is perhaps worth mentioning a certain confusion that can arise from the term nonmetric multidimensional scaling. The nonmetric in this term refers to the use of distance and dissimilarity rank orders in determining the best configuration and does not refer in any way to the distance measures used in the configuration space, which can be either metric or nonmetric. A better title for the method would probably be nonparametric multidimensional scaling but the original title has now become firmly established.

We have not discussed here the problem of determining the number of dimensions to use in the ordination space; as with simulated data, this is known as a priori. For real data, this problem is discussed by Kruskal (1964a), Shepard (1974) and in the conclusions of this paper.

A number of programs for calculating MDS configurations have been published but probably the most used is that of Kruskal and Carmone (1971). When using this program it should be noted that it allows the use of two different formulae for the definition of stress. These are the one described above (called formula 1) and a second in which the denominator is replaced by $\sum (d_{ij} - \delta)^2$ where $\delta = \Sigma d_{ij}/m$. It has been found that this second stress formula gives inferior results for ordination purposes and so the first stress formula should always be specified.

**Methods**

*Simulation of coenoclines and coenoplanes*

The CEP1 program (Gauch and Whittaker 1972a) was used to simulate four coenoclines having different levels of beta diversity. In this program, the beta diversity is defined by the parameter $Z$ which sets the average standard deviation ($\sigma$) of the species distribution curves by the relation $\sigma = 100/Z$ (the total length of the coenoclinc is 100 units). The four simulations had $Z$ values of 3, 6, 12 and 30 which are equivalent to beta diversities of 2.25, 4.5, 9 and 22 half-changes of the percentage similarity coefficient. All these data sets were simulated without the addition of sample noise or competition effects and the species profiles for $Z = 6$ and 30 are plotted in Fig. 1. All four coenoclincs had 50 samples evenly spaced and 65 initial species of which between 10% and 20% were assumed.
to be major species. The same random number sequence was used for all four simulations.

The coenoplanes were produced using an unpublished CEP21 program of H. G. Gauch and R. H. Whittaker. The main results were produced using three coenoplanes having beta diversities of $1.5 \times 1.5$ HC, $1.5 \times 4.5$ HC, and $4.5 \times 4.5$ HC. These are the identical coenoplanes used in the study of Gauch et al. (1976) and consist of 30 species and 40 samples placed on a regular $8 \times 5$ grid.

**Transformations**

The problem of standardization and transformation of the raw data is a complex one (Noy-Meir 1973, Noy-Meir et al. 1975) and we will thus restrict ourselves to one small aspect. A number of workers, especially in marine biology (Williamson 1961, Angel and Fasham 1975), have applied the log $(1 + x)$ transformation on the raw data, before doing a PCA. In the case of an R analysis, this is intended to normalize the data, which are often found to be distributed as a lognormal or negative binomial distribution (Cassie 1962). For a Q analysis, the transformation would have the effect of reducing the weight of the abundant species relative to the rare ones. In order to test the usefulness of the log transform, the analysis of the $4.5$ HC coenocline was run both with and without it.

**Similarity measures**

Nonmetric multidimensional scaling, like principal coordinates analysis, can be applied to any similarity matrix. It should be remembered however that the calculation of stress depends only on the rank order of the similarities and not on their magnitude. Thus two different similarity coefficients could produce the same MDS ordination if they had the same rank order over all the sample pairs. This would suggest that the choice of a similarity measure might be less critical for MDS than principal coordinates analysis. To test this, the $Z = 6$ coenocline has been analysed using a number of different measures of intersample similarity. These follow:

1) Pearson’s product-moment correlation coefficient. The use of this coefficient in a Q analysis has been criticised by Eades (1965), Minkoff (1965) and Beals (1973). Part of Minkoff’s criticism, relating to the instability of the coefficient to the change of direction of a variable axis, does not apply in the present case in which this direction is naturally defined.

2) Cos theta (angular separation). This is defined (Gower 1967) as

$$S_{ik} = \sum_{p=1}^{s} X_{ip} X_{kp} \left( \sum_{p=1}^{s} X_{ip}^2 \frac{\sum_{p=1}^{s} X_{kp}^2}{2} \right)^{-1/2}$$

where $X_{ip}$ and $X_{kp}$ are the counts (or some transformation of them) for the $p$th species from samples $i$ and $k$, and $s$ is the total number of species found in all the samples being analysed. The measure has also been called the coefficient of proportional similarity (Jöreskog et al. 1976).

3) Percentage similarity. This is defined as

$$P_{ik} = \left[ \sum_{p=1}^{s} \min(X_{ip}, X_{kp}) \right] / \left[ 2 \sum_{p=1}^{s} (X_{ip} + X_{kp}) \right].$$

Gauch and Whittaker (1972b) found that this coefficient gave better results than Euclidean distance when used with Bray-Curtis ordination.

4) Kendall’s coefficient of rank correlation. This uses the rank order of the species within the sample rather than their magnitude. This would therefore be an appropriate similarity measure if sampling problems gave rise to inaccurate estimates of quantitative species abundance. For example, Angel (1969) has shown that the abundance of certain ostracod species, obtained from repeated trawls at a constant depth and position in the Atlantic, changed, sometimes by an order of magnitude. However, the rank order of the species remained relatively constant.

5) Jaccard’s coefficient of community. This is a presence-absence index and is defined (Jaccard 1912) as

$$J_{ik} = c / (a_i + a_k - c)$$

where $a_i$ and $a_k$ are the number of species in samples $i$ and $k$, and $c$ is the number of species in common. Kessell and Whittaker (1976) have suggested that a presence-absence similarity measure may be better than a quantitative measure in samples of high alpha and beta diversity in the presence of high levels of sampling noise.
Ordination performance statistics

In order to determine the usefulness of an ordination method, it is necessary to derive a statistic that measures how well the resulting ordination compares with the known positions of the samples in the simulated coenocline or coenoplane. In their study of simulated coenoclones, Gauch and Whittaker (1972b) proposed a number of statistics, one of which was the product-moment correlation between the position of the samples along the first ordination axis and their simulated position along the coenocline. This statistic was found to be a rather insensitive measure of ordination efficiency and also it is not easily generalised to the two-dimensional case. Kessell and Whittaker (1976) proposed using the mean and standard deviation of the absolute displacement of the ordinated positions of each sample from their true positions along the coenocline. We will use a related measure that has the advantage of being easily generalised to two dimensions. Let \( X_i \) be the positions of sample \( i \) on the simulated coenocline and let \( Y_i \) be its position on the first ordination axis. It is assumed that the \( X_i \) and \( Y_i \) have been range standardised to lie between 0 and 100. We can then define the RMS displacement error (\( D_1 \)) as

\[
D_1 = \left[ \frac{1}{n} \sum_{i=1}^{n} (X_i - Y_i)^2 \right]^{1/2},
\]

where \( n \) is the total number of samples. The observed value of \( D_1 \) for an ordination can be compared with the separation of the samples, which for the case of 50 samples is two units.

In the particular case of one-dimensional coenoclines, it is also desirable that an ordination should preserve the rank order of the samples along the first ordination axis. This has been measured using the Kendall rank correlation coefficient (\( T \)) between the true order of the samples and their ordination order.

The problem of devising a performance statistic for the ordination of coenoplanes is complicated by the fact that the two ordination axes are often rotated with respect to the original simulated axes (Gauch et al. 1976). Let us assume that the coordinates of the ordination axes have been range standardised to lie between 0 and 100. In order to obtain a performance statistic, it is first necessary to overlay the two-dimensional ordination on the simulated position of the samples and rotate it so that it gives the best fit to the true sample positions. The distance between the ordinated and original positions could then be measured for each sample and a RMS average of these distances calculated in an analogous way to the one-dimensional case.

This operation could be done manually but the statistical technique of Procrustes analysis (Schonemann and Carroll 1970) can be used to achieve the effect mathematically. This method uses least squares to fit one matrix to another by means of a central dilation and a rigid motion. Let the simulated coordinates of the samples be contained in an \( n \times m \) matrix \( B \) (in our particular case \( m \) takes the value 2) and let the ordinated coordinates be contained in a matrix \( A \). Procrustes analysis fits \( A \) to \( B \) using the model

\[
B = cAT + J\delta' + E,
\]

where \( T \) is a \( m \times m \) orthogonal rotation matrix, \( J \) is a unit vector, \( \delta \) is an \( n \times 1 \) translation vector, \( c \) a dilation scalar and \( E \) the \( n \times m \) error matrix. \( T \), \( \delta \) and \( c \) are all estimated in order to minimise the squared elements of the error matrix \( E \). It can be seen from Eq. 7 that Procrustes analysis provides a statistical method for carrying out the manual operation described above. The range scaling is achieved by the scalar \( c \) and the rotation for best fit to the simulated sample positions is achieved by the matrix \( T \). Furthermore, the RMS average of the displacement errors can be calculated from the matrix \( E \) as

\[
D_2 = \left[ \text{tr}(E'E)/n \right]^{1/2}.
\]

This quantity is the \( m \)-dimensional generalisation of the quantity \( D_1 \). Programs to carry out Procrustes analysis are becoming more generally available but, in fact, the method is reasonably easy to program given the existence of matrix handling subroutines.

Ordination Results Using Simulated Coenoclones

Evaluation of various ordination methods applied to the 4.5-HC coenocline

In this section the ordinations of the 4.5-HC coenocline produced by PCA, RA and MDS will be compared and the effect on the MDS ordination of using different similarity measures, log transformations and nonmetric distance measures in the ordination space will be studied. In the following section the effect of varying beta diversity will be investigated.

Two varieties of PCA were applied to the 4.5-HC coenocline, centered and uncentered. The centered version is the standard method of PCA, while the uncentered version uses the cos theta coefficient in the similarity matrix rather than the correlation coefficient. The latter method has been used often in the geological sciences for Q analysis (Jöreskog et al. 1976). In both cases, a log transform of the data was made before analysis.

Both varieties of PCA explained between 85% and 90% of the variance and produced similar values of \( D_1 \) (Table 1). The uncentered version gave a marginally better value for \( T \), although neither recovered the true order of the samples. The RA ordination was better than either PCA, perfectly preserving the rank order of the samples. The RA and uncentered PCA ordinations have been plotted in Figs. 2a and 2b where the horseshoe effect can be plainly seen.

Seven MDS ordinations of the 4.5-HC coenocline
TABLE 1. Performance statistics of various ordinations of the 4.5-HC coenocline. The statistic $D_1$ measures the RMS deviation of the ordinated position of the samples from their true position and $T$ is the rank correlation between their true and ordinated positions. The distinction between one- and two-dimensional ordinations is not relevant for PCA and RA and their performance statistics have been placed arbitrarily in the one-dimensional column.

<table>
<thead>
<tr>
<th>Ordination method</th>
<th>Performance statistics, stress values for MDS</th>
<th>Euclidean metric used to calculate $d_{ij}$ for MDS ordination</th>
<th>‘Manhattan’ metric used to calculate $d_{ij}$ for MDS ordination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$D_1$</td>
<td>Two-dimensional</td>
<td>One-dimensional</td>
</tr>
<tr>
<td>Centered PCA</td>
<td>$13.1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.89$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncentered PCA</td>
<td>$13.0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.90$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td>$8.7$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$1.00$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDS using cos theta</td>
<td>$7.8$</td>
<td>$4.5$</td>
<td>$10.5$</td>
</tr>
<tr>
<td></td>
<td>$1.00$</td>
<td>$1.00$</td>
<td>$0.93$</td>
</tr>
<tr>
<td>MDS using cos theta with log transformation</td>
<td>$7.9$</td>
<td>$1.9$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$1.00$</td>
<td>$1.00$</td>
<td></td>
</tr>
<tr>
<td>MDS using percentage similarity</td>
<td>$7.7$</td>
<td>$2.4$</td>
<td>$12.1$</td>
</tr>
<tr>
<td></td>
<td>$1.00$</td>
<td>$1.00$</td>
<td>$0.90$</td>
</tr>
<tr>
<td>MDS using percentage similarity with log transformation</td>
<td>$7.2$</td>
<td>$3.5$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$1.00$</td>
<td>$1.00$</td>
<td></td>
</tr>
<tr>
<td>MDS using correlation coefficient with log transformation</td>
<td>$8.8$</td>
<td>$2.4$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.99$</td>
<td>$1.00$</td>
<td></td>
</tr>
<tr>
<td>MDS using rank correlation coefficient</td>
<td>$9.6$</td>
<td>$6.0$</td>
<td>$10.0$</td>
</tr>
<tr>
<td></td>
<td>$0.99$</td>
<td>$1.00$</td>
<td>$0.93$</td>
</tr>
<tr>
<td>MDS using Jaccard index</td>
<td>$4.9$</td>
<td>$2.4$</td>
<td>$6.5$</td>
</tr>
<tr>
<td></td>
<td>$0.96$</td>
<td>$0.98$</td>
<td>$0.94$</td>
</tr>
<tr>
<td></td>
<td>$0.8%$</td>
<td>$2.7%$</td>
<td>$2.3%$</td>
</tr>
</tbody>
</table>

were made using the five similarity measures, with cos theta and percentage similarity being calculated both with and without log transformations of the raw data. It has already been mentioned that MDS differs from most ordination methods in that a solution in $m$ dimensions is not the same as a solution in $m + 1$ dimensions projected down on to $m$-dimensional space. Accordingly, each MDS ordination was calculated in both one and two dimensions to investigate the possible distortion in the ordination produced by incorrectly estimating the underlying dimensionality of the data. In all cases, the first and second principal components of the uncentered PCA were used for the initial configuration. The performance statistics and stress values are given in Table 1 and the two-dimensional ordination for cos theta with log transformation is shown in Fig. 2c.

The following conclusions can be derived from the results of Table 1:

1) In all seven cases, the one-dimensional (1D) solutions gave lower values of $D_1$ than the two-dimensional (2D) solutions, although the stress values were higher. Thus stress cannot reliably be used as a measure of ordination efficiency (the stress value will in fact always decrease with increasing dimensionality of the ordination or solution space irrespective of the true dimensionality of the data).

2) The best overall performance was obtained using the quantitative similarity measures, which recovered the true rank order of the samples in the 1D and 2D ordinations. The Jaccard index gave the lowest values of $D_1$ for the 2D ordinations but failed to recover the true sample order ($T = 0.961$). The best quantitative measure in the 1D case was cos theta (with log transformation) and it should be noted that this is not the ordination with lowest stress.

3) The effect of a log transformation seemed marginal in the 2D ordination and ambiguous in the 1D case. However the best value of $D_1$ (= 1.9), which is less than the true separation on the samples, was obtained using the log transformation with the cos theta similarity index.

4) All the 1D MDS ordinations yielded performance statistics that were better than that produced by either RA or PCA, while the 2D MDS ordinations gave performance statistics better than that of PCA and of the same order as that of RA. Thus, using MDS with a quantitative measure is likely to give as good a result...
as RA and will give a better result if the true dimensionality of the data is known.

In the foregoing MDS ordination, the standard Euclidean formula was used to calculate the distances $d_{ij}$. However MDS allows the use of any distance metric for this calculation and so four 2D runs were made using the Manhattan metric to test the effect of using non-Euclidean distances. The results (Table 1) show that for each similarity measure, the resulting ordinations were worse than those obtained using the Euclidean formula, failing even to recover the true rank order of the samples. This result is in disagreement with Beals (1973) who suggested that the underlying metric of ecological ordination space might be the Manhattan metric.

**The effect of increasing beta diversity**

As beta diversity is increased the species change-over along the gradient increases, resulting in more low and zero similarity values in the similarity matrix. To test the effect of this on ordination efficiency, RA, PCA and MDS ordinations were made on coenoclines having beta diversities of 2.25, 9 and 22 HC which would be compared with the 4.5-HC coenocline already analysed. Beta diversity values as high as 22 HC are only likely to be encountered in zoogeographic studies but this coenocline should give a “worst possible” case with which to test the methods.

In order to keep the computing to a minimum, only four ordinations were made on each coenocline. These were RA, centered PCA and MDS using a cos theta similarity coefficient (with log transformation); the MDS ordination was run using both the RA and PCA ordinations as the initial configuration, and the performance statistics are given in Table 2.

It is obvious that PCA is most affected by increasing beta diversity. For the 22-HC coenocline, the first two principal components accounted for only 40% of the total variance, while it needed eight components to explain more than 85%. Thus, even assuming the applicability of a linear model, the first two components would be unlikely to recover the true relation-
The RA is less affected by increasing beta diversity and the RA ordinations always succeed in recovering the true order of the samples. Hill (1974) has shown that if the distribution of species (or any other data) along a gradient takes the form of unimodal curves, then RA will always reproduce the correct sample order along the first ordination axis. One surprising point of the RA ordination is that the value of $D_1$ for the 2.25-HC coenocline is higher than that for 4.5 HC. The RA ordination for 22 HC, plotted in Fig. 3b, is a great improvement on PCA, although there is a considerable bunching of the samples at a number of points.

If we consider next the 1D-MDS ordinations it can be seen that these are the least affected by beta diversity. They always recovered the true order of the samples and gave very similar values of $D_1$ for the 2.25-, 4.5- and 22-HC coenoclines. The value for 9 HC was much higher (although still lower than the RA and PCA ordinations) and the reason for this is not understood. The 2D-MDS ordinations are generally worse than the 1D solutions and for the 9-HC and 22-HC coenoclines they failed to reproduce the correct sample sequence. They are however, always better than the PCA ordinations. The 2D-MDS ordination for the 22-HC coenocline (RA ordination starting configuration) is plotted in Fig. 3c and shows a double horseshoe effect which is presumably a function of the very high beta diversity of this coenocline.

It can be seen from Table 2 that the choice of initial configuration had no effect on the 1D-MDS ordinations and only a small effect on the 2D solutions. The reason for using two starting configurations was that it is known that 1D MDS is susceptible to trapping in local minima. Thus Shepard (1974) has stated that "a point that is initially situated on the wrong side of some other points can gradually work its way round in a space of two dimensions but, when confined to a single line, it is unable to move through these points owing to forces of mutual repulsion." This would suggest the likelihood that PCA starting configurations (having some samples out of their correct sequence) would be trapped in local minima whereas RA starting configurations (which are in the correct sample order) would not. It can be seen, however, that 1D-MDS ordinations recover the true sample order even when the order was initially wrong. It would be dangerous to generalise from these few results and it is probably safer, when the investigator suspects a one-dimensional situation, to use an RA ordination as the starting configuration.

**Evaluation of ordination methods applied to coenoplanes**

Three coenoplanes were used for this study, having beta diversities for the two axes of $1.5 \times 1.5$ HC, $1.5 \times 4.5$ HC, and $4.5 \times 4.5$ HC, and were the identi-
For the MDS ordinations, the log transformation and the cos theta similarity coefficient were used. The MDS algorithm is less likely to become trapped in local minima when working in two-dimensional space and so the effect of using a random pattern for the initial configuration (as well as the RA ordination) was investigated. The ordinations were compared with the original sample positions using Procrustes analysis as previously described and the values of $D_2$ and stress values for the MDS ordinations are given in Table 3 and plots of the ordinations in Fig. 4. When considering the values of $D_2$, it should be remembered that the original sample positions were on an $8 \times 5$ grid and so the sample separation for the $1.5 \times 1.5$ HC- and $4.5 \times 4.5$ HC-coenoplanes was 25 units along one axis and 14.3 along the other. For the $4.5 \times 1.5$ HC-coenoplane the separation was 8.2 and 14.3 units on each axis.

The first observation is that MDS gives better ordinations than RA for the $4.5 \times 1.5$ HC- and $4.5 \times 4.5$ HC-coenoplanes but a worse ordination for the $1.5 \times 1.5$ HC-coenoplanes. This ordination was repeated using percentage similarity as the similarity measure and the same value of $D_2$ was obtained. From the plots (Fig. 4), it can be seen that the MDS ordination has been stretched along a diagonal compared to the RA ordination and on studying the similarity matrix, it was found that this asymmetry along diagonals was observed in the similarity measures. In this case therefore, the MDS ordination is reflecting a real quantitative feature of the simulation, which is not picked up by the RA ordination, but which results in a slightly worse fit to the true sample positions. The second point to be observed is that when using cos theta, the

![Fig. 4. Two dimensional ordinations of $1.5 \times 1.5$ HC, $1.5 \times 4.5$ HC and $4.5 \times 4.5$ HC coenoplanes using reciprocal averaging and nonmetric multidimensional scaling. (HC = half-change.)](image-url)
same results are obtained using random and RA initial configurations. If this were generally true for two or more dimensions it would eliminate the necessity of calculating RA initial configurations.

As a small further test of the effects of various similarity measures on the MDS ordination, the $4.5 \times 1.5$ HC-coenoplane was ordinated using percentage similarity, correlation coefficient, Jaccard's presence-absence index and Kendall's rank correlation coefficient. The values of $D_2$ obtained were respectively, 7.1, 5.0, 12.6, and 5.0. It will be noticed that the relative efficacity of the measures in this study is not the same as observed in the 1D-MDS ordination of the 4.5 HC-coenocline. However, the cos theta produced the better ordinations in both cases.

**Conclusions**

The results of the above one-dimensional and two-dimensional simulations show that in nearly all cases, providing the true dimensionality of the data is known and quantitative similarity measures were used, nonmetric multidimensional analysis produced better ordinations than either reciprocal averaging or principal components analysis. In practice, of course, the investigator does not initially know the true dimensionality of the data and may have to make a judgement based on his ecological background knowledge. Even if the investigator believes the true situation to be onedimensional, Kendall (1971) has suggested that a two-dimensional scaling should first be made as a test of this hypothesis. If a horseshoe or similar structure is obtained then this lends weight to the hypothesis. A better method is perhaps to use the dissimilarity matrix to calculate the linkages of the minimum spanning tree (Gower and Ross 1969). This tree connects each sample in such a way that its total length (estimated from the dissimilarity matrix) is a minimum. If the sample points in a two-dimensional ordination are connected in the same order as that of the minimum spanning tree, then one-dimensional data structures will be displayed as a line curving around the two-dimensional space, whereas a two-dimensional data structure will consist of one or more main lines with any number of side branches. If a large number of these side branches cross over each other it indicates the dimensionality is greater than two.

MDS is less susceptible to high beta diversities than either RA or PCA, and for two-dimensional coenoplanes it can cope better when each axis has different beta diversities. It does not, however, manage to completely straighten out the horseshoes when ordinating coenoclines. In the Introduction, the cause of the horseshoe effect was suggested as due to a combination of using a linear model in a nonlinear situation and using similarity measures that were not linear with sample separation and were insensitive at large sample separations. The results of this paper showed that using a nonlinear method of analysis in two dimensions still produced horseshoe ordinations, although they were a great improvement on PCA. This suggests that it is the similarity measures that are the prime cause of the horseshoe effect. Kendall (1971) has shown that if the data is truly one-dimensional, it is possible to define a similarity measure that overcomes the insensitivity of the similarity measures at large sample separations. He was able to straighten out the horseshoe by using this measure in conjunction with nonmetric scaling. This would seem a promising line of attack for the ecological problem. An alternative approach investigated by Shepard and Carroll (1966) consisted of relaxing the requirement of global monotonicity during the steepest descent iterative process of the scaling algorithm. This has the result of ignoring the insensitive similarity measures (i.e., those at large sample separation) and so indirectly brings about the same result achieved by Kendall. However, in a limited number of tests on the 22 HC-coenocline, this did not produce any improvement in the performance statistics.

In the tests reported here, the choice of initial configuration for the MDS did not appear to be critical. However, as it is known that locally minimum solutions can give rise to problems, it is probably safest to run each MDS ordination with at least two different starting configurations, e.g., a random pattern and an RA ordination.

The results of the evaluation of the similarity measures were ambiguous, although the most consistent performer in terms of ordination efficacy was the cos theta index. This problem requires more simulations using replicated coenoclines and coenoplanes having different initial random numbers.

The effects of sample outliers, sample clusters and sample noise on MDS ordinations has not yet been studied in any detail. Some preliminary observations have suggested that MDS is not affected by sample clusters (as might be expected from a consideration of the algorithm) or irregularly spaced samples and is not seriously affected by noise levels up to 30% (defined so that percentage similarity amongst replicates was 70%).

**Acknowledgments**

I am very grateful to Dr. H. G. Gauch, Jr. for providing programs for reciprocal averaging and coenocline and coenoplane simulation. I also thank the two referees for stimulating comments on the first draft of this paper which led, I hope, to a better presentation.

**Literature Cited**


