**PBDB Intensive Summer Course 2007**
**Paleoecology Section – Thomas Olszewski**

# Non-Parametric Ordination

## Introduction

*Purpose* – graphically display data in such a way that relationships among objects are revealed/displayed; summarize complex, multidimensional data

All ordination techniques start with a similarity matrix (or its complement, the dissimilarity matrix) and proceed to project the data into a Euclidean space that requires fewer dimensions than the full data set; reducing the dimensionality loses some information, but if successful, this is minor, stochastic noise

## Non-Metric Multidimensional Scaling (NMDS)

An approach to ordination in which the scatter of points representing samples or taxa in an ordination space is iteratively allowed to evolve until it resembles the observed dissimilarity matrix as closely as possible

Method is *non-metric* because stress, the measure of closeness of fit, is based on the *ranking* of dissimilarity values rather than actual values of dissimilarity

## Steps in NMDS

1) Starting with appropriately vetted, transformed, and standardized data, generate a dissimilarity matrix, *S* (any measure of similarity/dissimilarity can be used).
2) Reorganize this square, symmetric matrix to a vector of $S_{ij}$ values in increasing order; the length of this vector will be $(p^2 - p)/2$ ($p$ = number of objects – i.e., taxa or samples) – this vector is usually labeled δ.
3) Plot *p* points in a space. Each point corresponds to an object. The dimensionality of the space must be specified by the user (generally 2 or 3 because we can conceive of and display such a space).
4) Calculate the Euclidean distances between all pairs of points in the ordination space (call this matrix *E*). *E* will have the same size as *S* and for every $S_{ij}$ value there will be a corresponding $E_{ij}$ value.
5) Reorder the elements of *E* in the same order as they appear in the vector δ – this vector is called *d*. A plot of $\delta$ vs *d* is called a Shepard Diagram and it shows the discrepancy between the points plotted in ordination space and their actual measured dissimilarities based on the data. ***The goal of NMDS is to rearrange the points in the ordination space so that their rank order matches that of the δ vector***, which cannot be changed because it is based on data.
6) The lousiness of the fit between the real dissimilarities and the ordination distances is measured using *stress*:

$$S = \sqrt{\frac{\sum_{i=1}^{n-1}\sum_{j=2}^{n}(d_{ij} - \hat{d}_{ij})}{\sum\sum d_{ij}^2}} \qquad (4)$$

$\hat{d}_{ij}$ = the discrepancy between a point's rank in the $\delta$ vector and the *d* vector

Stress ranges from 0 to 1, although it is often reported as a percentage.

7) Move the points in the ordination space so that their *d* values are closer in rank order. The problem is that improving the position of any point with respect to another point can worsen its position relative to a third point, so letting them evolve randomly is very inefficient. Without proof, it is possible to choose a *direction*, although not a distance, to move each point that will reduce the overall stress:

$$g_{kl} = S\sum_{i=1}^{p}\sum_{j=1}^{p}(\Delta^{ki} - \Delta^{kj})\left[\frac{d_{ij} - \hat{d}_{ij}}{S^*} - \frac{d_{ij}}{\sum_{i,j} d_{ij}^2}\right]\frac{(x_{il} - x_{jl})}{d_{ij}} \qquad (5)$$

$g_{kl}$ = a vector indicating direction in which point *k* should be moved

$$S^* = \sum_{i,j}(d_{ij} - \hat{d}_{ij})^2$$

*i* = index of one of the p points
*j* = index of another of the p points
*l* = index of dimension (x, y, z or however many were chosen)
*k* = index of 3rd point (the point being moved in the *l* direction)
$\Delta^{ki}$ and $\Delta^{kj}$ = Kronecker symbols ($\Delta^{ki} = 1$ if $k = i$; $\Delta^{ki} = 0$ if $k \neq i$)

8) The distance points are moved is a function of stress: If it is too large, then the ordination space never converges on a reasonable representation of the dissimilarity matrix. If it is too small, then the convergence is inefficient and slow.
9) Repeat until stress stabilizes. Clarke (1993) suggests the following guidelines for acceptable stress values: <0.05 = excellent, <0.10 = good, <0.20 = usable, >0.20 = not acceptable.

**Limitations**

1) Ordination axes are arbitrary (i.e., axis 1 does not necessarily maximize the amount of information it displays). This limitation can be easily overcome by rigidly rotating the axes to conform to some sort of criterion (e.g., use a Principal Components Analysis to find high-variance axes or an iterative rotation routine like VARIMAX).
2) Rerunning from different starting points can result in convergence on a different solution – typically just rotated or mirrored, but sometimes with critical samples in very different locations in the ordination space.
3) NMDS can get stuck at local stress minima, producing an uninterpretable solution. This requires starting from a new initial state.
4) NMDS can be very slow to converge from a random scatter of points. Its efficiency can be greatly increased if the initial solution is obtained from an initial parametric ordination like PCA or CA (which can be approximately obtained for a few axes very easily).
5) Dimensionality of ordination space can be assessed using a "scree plot" (a plot of stress versus number of dimensions): when additional dimensions do not

greatly decrease stress, then they are not needed for a reasonable description of the data.

**Final Comments**

NMDS is regarded as one of the most powerful of ordination techniques available. However, due to its unconstrained axes, it can be very difficult to interpret if more than 2 or 3 dimensions are required. If too few dimensions are chosen, even if stress is acceptable, then results can be uninterpretable. A recent development called "flexible shortest path adjustment" of dissimilarity matrices shows great promise for overcoming some ongoing problems with most current ordination techniques that are usually expressed in plots as a "horseshoe" or "arch" effect.

---

## References for Nonmetric Multi-Dimensional Scaling (NMDS)

There are numerous books on the topic of MDS (Multi-Dimensional Scaling, Metric and Nonmetric). Most of them are completely incomprehensible. Each of the articles below, either directly or indirectly, refers to some important aspect of NMDS, but I know of no satisfactory overall summary.

Clarke, K.R., 1993, Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology, v. 18, p. 117-143. (Clarke includes a description of the strength of non-parametric or rank order methods in multivariate ecological analyses.)

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

Kruskal, J.B., 1964b, Nonmetric multidimensional scaling: A numerical method. Psychometrika, v. 29, p. 115-129. (Kruskal's articles are the original and still best description of NMDS – later workers have tweaked the technique in various ways to speed them up, but they all still conform to this original and remarkably lucid description of the algorithm.)

McCune, B., and Mefford, M.J., 1999, PC-ORD. Multivariate Analysis of Ecological Data, Version 4. MjM Software Design, p.111-113. (This is a very concise synopsis of the classical NMDS method developed by Kruskal.)

Concerning Global NMDS or the "step-across" method of dealing with dissimilarity limits (these methods are not generally implemented in most commercial software packages, but routines are available in the R environment):

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

De'ath, G., 1999, Extended dissimilarity: a method of robust estimation of ecological distances from high beta diversity data. Plant Ecology, v. 144, p. 191-199.

Minchin, P.R., 1987, An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio, v. 69, p. 89-107. (This reference includes a good description of local NMDS.)

Williamson, M.H., 1978, The ordination of incidence data. Journal of Ecology, v. 66, p. 911-920.

## Appendix: Polar or Bray-Curtis Ordination

This method will not be discussed in lecture during the PBDB Intensive Short Course. It is included here simply for completeness' sake and so that you are aware of this method if you should ever run across and example. It is one of the oldest and most intuitive methods for ordination, but regarded as outdated (see Beals, 1984 for an apologia)

Advantages:

1) Easy to understand
2) Very flexible with regard to choice of similarity coefficient
3) Bulletproof – hard to crash

**Steps in Polar Ordination**

1) SEARCH the dissimilarity matrix for the highest value of $D_{AB}$ – the two samples associated with this value, A and B, are the most dissimilar and will serve as reference points for the first ordination axis. (In polar ordination, *D* is classically the Bray-Curtis measure of dissimilarity, which is the complement of the percent similarity coefficient and overall regarded as a very good coefficient for ecological data, but polar ordination is very flexible and can be implemented using any metric.)
2) PROJECT all other points onto the line defined by the reference points. This is calculated using the following equation:

$$x_i = \frac{D_{AB}^2 + D_{Ai}^2 - D_{Bi}^2}{2D_{AB}} \qquad (1)$$

$D_{AB}$ = dissimilarity of the reference points
$D_{Ai}$ = dissimilarity between *A* and sample *i*
$D_{Bi}$ = dissimilarity between *B* and sample *i*
$x_i$ = the position of *i* on the ordination axis

*Note*. This equation treats the dissimilarities as Euclidean distances in a multivariate space; this approach assumes that *D* does not violate the triangular inequality.

3) MEASURE the "lack of fit" – many points in the ordination space will not be close to the calculated ordination axis. Their lack of fit is measured using the following equation:

$$e_i = \sqrt{D_{Ai}^2 - x_i^2} \qquad (2)$$

4) SEARCH for the point with the greatest $e_i$ value and find the point that is most dissimilar from it. Use these points as reference points for the next ordination axis and repeat procedure until as many axes as desired are obtained.
5) PLOT objects as points in ordination space.

*Alternative Procedure for Higher Axes:*

Rather than using $e_i$ as the basis for choosing successive reference points, most modern implementation of polar ordination use a *residual distance matrix* to construct each successive ordination axis. The values in this matrix can be calculated using the following equation:

$$RD_{ij} = \sqrt{D_{ij}^2 - OD_{ij}^2} \qquad (3)$$

$OD_{ij}$ = Euclidean distance between $i$ and $j$ in ordination space using as many axes as have already been calculated
$RD_{ij}$ = residual distance between $i$ and $j$ – i.e., difference between $i$ and $j$ that is not explained by the currently calculated ordination axes

*Note.* If $D_{ij}$ does not follow the triangular inequality, this can cause severe distortions of ordination space and imaginary values of $RD_{ij}$. Even if $D_{ij}$ does follow the triangular inequality, if it is not linearly related to the Euclidean distance, it can warp the ordination space.

## How Good is the Ordination?

Correlate $OD_{ij}$ with $D_{ij}$ to obtain a matrix correlation (i.e., like the cophenetic correlation seen in cluster analysis).
Calculate $OD_{ij}$ using only the coordinates of a single ordination axis and correlate these values to $D_{ij}$ to assess the quality of that axis in the ordination.
Calculate $OD_{ij}$ between one point and all the other points and correlate these values to $D_{ij}$ to assess the quality of that point in the ordination.

## Alternative Strategies for Choosing Reference Points

1) *Maximum Range* – choose reference points based on maximum dissimilarity; original Bray-Curtis approach described above
2) *Correlation* – choose reference points with lowest *r* value (most negative value) when correlating based on their distances with all other points; susceptible to outliers
3) *Variance-Regression* – choose reference points using the minimum regression slope rather than lowest *r* value; appears to be robust
4) *Known Extremes* – user chooses reference points based on geographic or environmental knowledge
5) *Minimum Residual Distance* – choose reference points that define an axis that minimizes $\Sigma e_i^2$; computationally intensive because all possible pairs of points must be tried

6) *Perpendicular Axes* – after the first axis, choose the point with greatest $e_i$ and use its perpendicular projection onto the existing axes as the next axis; sensitive to outliers
7) *Centroid and Reference Point* – one reference point equals the centroid of the samples (i.e., the average composition) and the other reference point is chosen as the real sample that maximizes the variance (see method #3); computationally intensive
8) *Functional Mean* – like method 7 but instead of centroid use the real collection closest to the centroid
9) *Synthetic Reference Points* – like method 7 but user chooses multiple reference points and uses the average of this selected sample as one reference point; alternatively, use the centroids of clusters defined by cluster analysis

**Issues and Solutions**

1) What if there is a tie between two alternative sets of reference points?
   Solution: of the tied points, choose the ones with the maximum variance of distances with all other points
2) Outliers can significantly distort the orientation and direction of axes
   Solution a: discard the outliers and rerun
   Solution b: use a method of reference points that is insensitive to outliers
3) Some methods of finding reference points result in axes that are not perpendicular – this means they are redundant in terms of the information they display. The degree of redundancy can be determined by correlating axis ordination values (remember that $r = cos\theta$, where $\theta$ is the angle between axes in the ordination space)

**Final Caveat**

Polar Ordination is generally regarded as an antiquated approach that has been superceded by much more powerful, modern techniques. It is still useful for teaching the fundamental geometric concepts of ordination (and possibly as a last resort if all other methods fail or are unavailable), but it is not generally appropriate for publishable analyses.

---

## References for Polar Ordination

Beals, E.W., 1984, Bray-Curtis ordination: An effective strategy of analysis of multivariate ecological data. Advances in Ecological Research, v. 14, p. 1-55.

Bray, J.R. and Curtis, J.T., 1957, An ordination of the upland forest communities of southern Wisconsin. Ecological Monographs, v. 27, p. 325-349.

Cottam, G., Goff, F.G., and Whittaker, R.H., 1973, Wisconsin comparative ordination. *In* Whittaker, R.H., ed., Ordination and Classification of Communities, Part V of Handbook of Vegetation Science (Tüxen, R., ed.), p. 195-221. (Do not miss the erratum at the beginning of the book concerning the equation on page 202).