BIO 682
Quantitative Biology
Spring 2010

Steve Shuster

http://www4.nau.edu/shustercourses/BIO682/index.htm

Lecture 17

Indirect Gradient Analysis

Refers to several distance-based approaches:

• Polar ordination, PO (Bray-Curtis ordination)

• Principal Coordinates Analysis, PCoA (Metric multidimensional scaling)

• Nonmetric Multidimensional Scaling, NMDS

Distance Based Approaches

• Rely on a *distance matrix* as input so are highly sensitive to choice of distance measure

• They "hide information" - Information about species is *collapsed* (we don't know what individual species are doing, we just know how they relate to all other species)

Polar Ordination - PO
(Bray and Curtis 1957)

Arranges samples between endpoints or "poles" according to the distance matrix.

In the earliest versions of PO, these endpoints were the two samples with the highest ecological distance between them, *or* two samples which are suspected of being at opposite ends of an important gradient (subjective?)
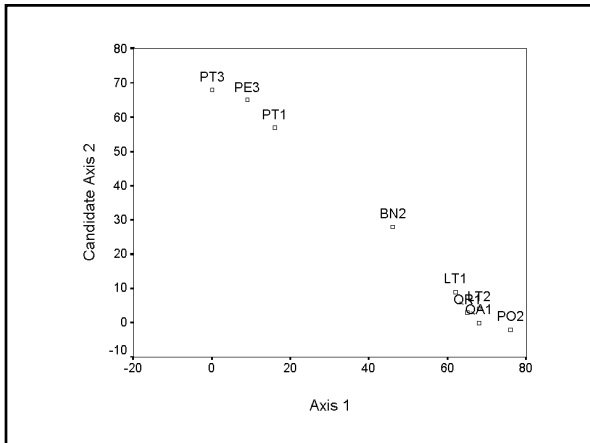

Polar Ordination

- These samples became the endpoints for the primary axis
- The 2nd axis was determined by either the two samples with the next highest ecological distance
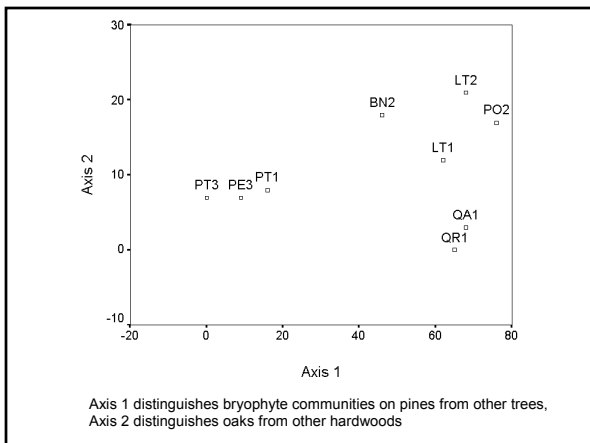- Or two samples creating an axis with the lowest correlation with axis 1

Polar Ordination Axis scores for the bryophyte data.

	Axis 1	First candidate for Axis 2	Second candidate for Axis 2
Endpoint 1	PT3	QA1	QR1
Endpoint 2	PO2	PE3	LT2
BN2	46	28	18
LT1	62	9	12
LT2	68	4	21
PE3	9	65	7
PO2	76	-2	17
PT1	16	57	8
PT3	0	68	7
QA1	68	0	3
QR1	65	3	0
Correlation with Axis 1	1.000	-0.9961	0.292841

Polar Ordination

- Axis 1 score = $(D^2 + D1^2 - D2^2)/2D$

- Where D is the distance between the endpoints,
- D1 is the distance between a sample and the first endpoint
- D2 is the distance between a sample and the second endpoint

Axis 1 distinguishes bryophyte communities on pines from other trees, Axis 2 distinguishes oaks from other hardwoods
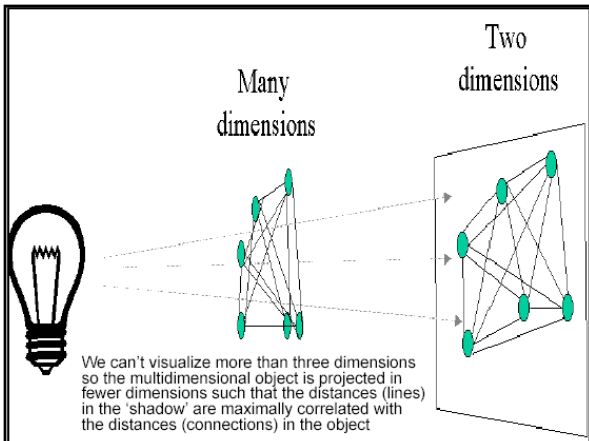
Polar Ordination

- In PO the order of the axes are important (axis one is more important than axis two), but the direction and the numeric scale are arbitrary

- Important that axes are *not correlated*
  - other techniques (MMDS, NMDS) automatically give uncorrelated or orthogonal axes

Principal Coordinates Analysis – PCoA or MMDS

- Maximizes the *linear correlation* between the distances in the distance matrix, and

- Displays the distances in a space of low dimension (typically, 2 or 3 axes)

- Distances between N objects are represented in N-1 dimensions

- When distance is Euclidean it is equivalent to PCA

Many dimensions

Two dimensions

We can't visualize more than three dimensions so the multidimensional object is projected in fewer dimensions such that the distances (lines) in the 'shadow' are maximally correlated with the distances (connections) in the object

Disadvantages

PCoA maximizes linear correlation so with moderate to high beta diversity an "arch effect" occurs

---

Arch Effect

A and B are 60 units apart

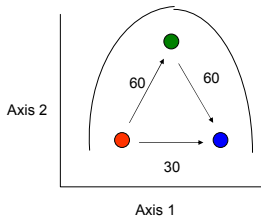B and C are 60 units apart

---
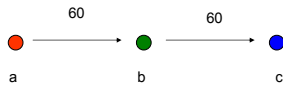
Arch Effect

A and B are 60 units apart

B and C are 60 units apart

A and C are 30 units apart

This becomes a problem when axes are important (i.e. order is important or the axes indicate gradients)

Non-metric Multidimensional Scaling - NMDS

- Whereas PCoA maximizes *linear correlation*, NMDS maximizes a *rank order correlation*

- This difference effectively takes away the arch effect

NMDS

1. User selects the # of dimensions (N) and the distance measure

2. The distance matrix is calculated

3. An initial configuration of samples in N dimensions is selected. This configuration can be random

4. A measure of 'stress' (mismatch between the rank order of distances in the data, and the rank order of distances in the ordination) is calculated

5. The samples are moved slightly in a direction that decreases the stress

6. 4 and 5 are repeated until 'stress' appears to reach a minimum. The final configuration of points may be rotated if desired.
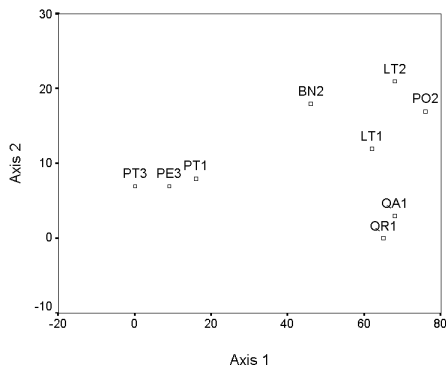
NMDS

- Configuration depends on the # of axes (axes 1 and 2 in a three dimensional solution will not be the same as a 2 dimensional solution)

- Axis # does not reflect importance

NMDS of bryophyte data
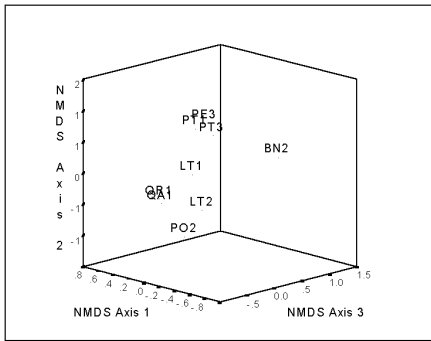
Polar Ordination of bryophyte data

Indirect Gradient Analysis, cont.

- Eigenanalysis-based approaches
  - **Linear model**
    - **Principal Components Analysis, PCA**

  - **Unimodal model**
    - **Correspondence Analysis, CA (Reciprocal Averaging)**

    - **Detrended Correspondence Analysis, DCA**

Eigenanalysis-based indirect gradient analysis (PCA, CA, and DCA)

1. An eigenanalysis is performed on a square, symmetric matrix derived from the data matrix (matrix algebra)

2. There is a unique solution to the eigenanalysis, no matter the order of data.

3. Axes are ranked by their eigenvalues. Thus, the first axis has the highest eigenvalue, the second axis has the second highest eigenvalue, etc.

Eigenanalysis-based indirect gradient analysis (PCA, CA, and DCA)

4. In principal components analysis, eigenvalues are 'variance extracted'. In methods related to correspondence analysis, eigenvalues are 'inertia extracted', or equivalently, correlation coefficients (species and sites).

5. There are a potentially large number of axes (usually, the number of samples minus one, or the number of species minus one, whichever is less) so there is no need to specify the dimensionality in advance. However, the number of dimensions worth interpreting is usually very low.

6. Species and samples are ordinated simultaneously, and can hence both be represented on the same ordination diagram (if this is done, it is termed a *biplot*).

Principal Components Analysis

Geometrically, PCA is a rigid rotation of the original data matrix

Can be defined as a projection of samples onto a new set of axes, such that the maximum variance is projected or "extracted" along the first axis,

The maximum variation uncorrelated with axis 1 is projected on the second axis,

The maximum variation uncorrelated with the first and second axis is projected on the third axis, etc.

Principal Components Analysis

- This is why typically few axes are important because the first two or three axes often explain close to 90% of the variance

- Especially if samples are collected with gradients in mind

Principal Components Analysis

The abundance of the species is continuously increasing in the direction of the arrow,

Decreasing in the opposite direction. Thus PCA is a 'linear method' and is subject to the arch or horseshoe effect

Principal Components Analysis

- Eigenvalues represent the variance described by each axis and are usually given as percentages

- The axis cut-off occurs when adding new axes doesn't appreciably add to the variance explained

Disadvantages

PCA assumes linear relationships between samples and environmental gradients and this is hardly ever the case
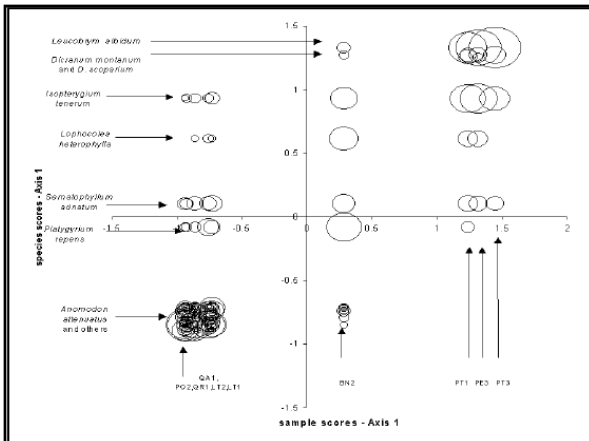
Horseshoe effect

Correspondence Analysis (CA)

- Also known as reciprocal averaging, because one algorithm for finding the solution involves the repeated averaging of sample scores and species scores

- Instead of maximizing 'variance explained', CA maximizes the *correlation* between species scores and sample scores

Correspondence Analysis (CA)

- Correspondence analysis is better for count data and more suitable if species have unimodal responses to the underlying parameters
  - They favor a certain range of the parameter and become more rare for lower and higher values

- The eigenvalue of the CA axis is equivalent to the correlation coefficient between species scores and sample scores

Disadvantages

- CA is prone to the arch effect and gradient compression which can seriously affect ecological interpretation

- These problems have led to Detrended Correspondence Analysis

Detrended Correspondence Analysis (DCA) Hill and Gauch 1980

- DCA eliminates the arch effect by *detrending* (Hill and Gauch 1982)

- Detrending can be done by polynomials or by segments

Detrended Correspondence Analysis

Detrending by polynomials is the more elegant of the two:

A regression is performed in which the second axis is a polynomial function of the first axis,

The second axis is replaced by the residuals from this regression.

Similar procedures are followed for the third and higher axes.
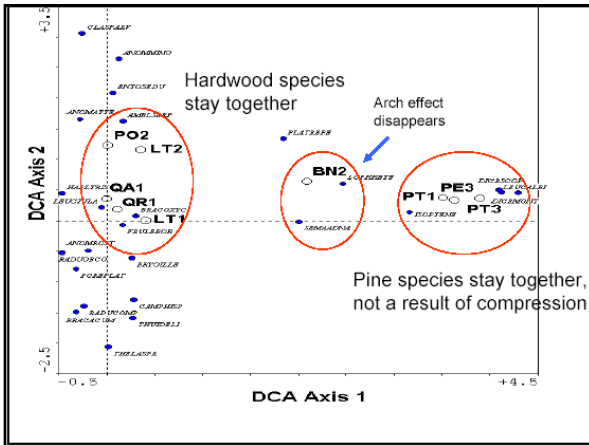
Detrended Correspondence Analysis

- To detrend the second axis by segments, the first axis is divided up into segments (usually 26), and the samples within each segment are centered to have a zero mean for the second axis
- The procedure is repeated for different 'starting points' of the segments

Detrended Correspondence Analysis

The compression of the ends of the gradients is corrected by *nonlinear rescaling*.

Rescaling shifts sample scores along each axis such that the average width (or 'tolerance') is equal to 1
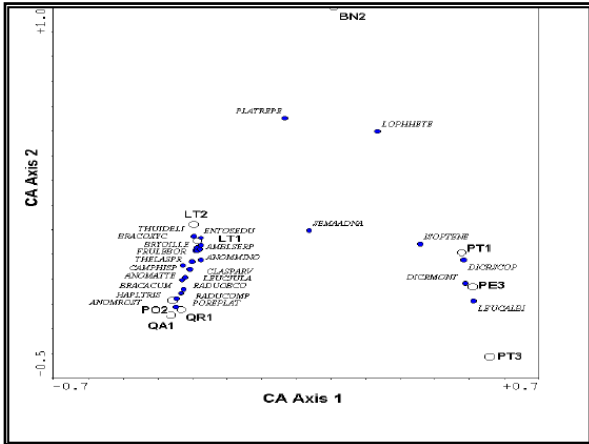
Detrended Correspondence Analysis

Rescaling has a beneficial consequence: the axes are scaled in units of beta diversity (SD units, or units of species standard deviations).

Thus if the underlying gradient is important or well known, it is possible to plot the DCA scores as a function of the gradient, and

Thereby determine whether the species 'perceive' the gradient differently than we measure it.

Disadvantages

- Because of the detrending process DCA will destroy any true arch in a data set

- Performs poorly with skewed species distributions (i.e. not unimodal)

- Does not handle complex sampling designs well

Detrended Correspondence Analysis

- DCA is based on an underlying model of species distributions (the unimodal model), while NMDS is not.

- Use DCA when you want to see how species relate to each other along a gradient

Indirect Gradient Analysis

Distance-based approaches

- **Polar ordination, PO (Bray-Curtis ordination)**

- **Principal Coordinates Analysis, PCoA (Metric multidimensional scaling)**

- **Nonmetric Multidimensional Scaling, NMDS**

Indirect Gradient Analysis.

- Eigenanalysis-based approaches
  - **Linear model**
    - **Principal Components Analysis, PCA**

  - **Unimodal model**
    - **Correspondence Analysis, CA (Reciprocal Averaging)**

    - **Detrended Correspondence Analysis, DCA**

_____

_____

_____

_____

_____

_____

_____

_____