TOPIC 7

INTRODUCTION TO ORDINATION

Ordinations

The following methods are covered in the lecture slides. You are expected to be familiar with how to select the appropriate method, and how to execute each. Supplement your studying by accessing these sources: **Numerical Ecology with R**, **GUSTA ME,** and **Analysis of Community Ecology Data in R**
(here are the specific links):

- Principal Component Analysis (PCA) -- https://sites.google.com/site/mb3gustame/indirect-gradient-analysis/pca; https://www.davidzeleny.net/anadat-r/doku.php/en:pca
- Correspondence Analysis (CA) -- https://sites.google.com/site/mb3gustame/indirect-gradient-analysis/ca; https://www.davidzeleny.net/anadat-r/doku.php/en:ca_dca
- Principal Coordinate Analysis (PCoA) -- https://sites.google.com/site/mb3gustame/dissimilarity-based-methods/principal-coordinates-analysis; https://www.davidzeleny.net/anadat-r/doku.php/en:pcoa_nmds
- non-Metric Multidimensional Scaling (nMDS) -- https://sites.google.com/site/mb3gustame/dissimilarity-based-methods/nmds; https://www.davidzeleny.net/anadat-r/doku.php/en:pcoa_nmds
- Redundancy Analysis (RDA) -- https://sites.google.com/site/mb3gustame/constrained-analyses/rda; https://www.davidzeleny.net/anadat-r/doku.php/en:rda_cca
- Canonical Correspondence Analysis (CCA) -- https://sites.google.com/site/mb3gustame/constrained-analyses/cca and https://www.davidzeleny.net/anadat-r/doku.php/en:rda_cca

Ordinations

- clustering find discontinuities, but **ordinations highlight gradients**
- gradients are particularly present in ecological communities
- suited to multivariate data, which represent…
    - …a **space** (e.g. the landscape) comprised of many **sites** (i.e. the samples forming the rows), each one occupied by many **variables** (i.e. species or environmental variables in the columns)
    - …a **time series** (e.g. the landscape) comprised of many **repeated samples** (again in the rows), each one occupied by many **variables** (i.e. species or environmental variables in the columns)
    - there are as many dimensions (columns) as variables (hence multidimensional or multivariate)
    - it would be a silly and not very revealing to analyse it all using a series of univariate or bivariate analyses
    - $(27 \times 26) / 2 = 351$ in the Doubs River data set to be precise

Ordinations

- ordinations represent the data along a reduced number of **orthogonal axes** (linearly independent, uncorrelated), constructed in such a way that they represent, in decreasing order, the main trends of the data — examples will clarify (below)
    - each one of these orthogonal axes will capture some of the variation due to the original variables (i.e. the columns).
- interpretation is aided by visualisations, regressions, and clustering
- **unconstrained** ordinations are not statistical (no inference testing); they are purely descriptive
    - also called indirect gradient analysis
    - based only on the environment × sites matrix or the species × sites matrix, each seen in isolation
- **constrained** ordination adds a level of statistical testing (next topic)
    - also called direct gradient analysis or canonical ordination
    - typically uses explanatory variables (in the env. matrix) to explain the patterns seen in the spp. matrix

Ordinations

- basically, ordinations geometrically arrange sites or species so that distances between them in the cartesian 2D or 3D graph represent their ecological distances
- the further the distances between sites or species are on the graph, the larger the ecological differences between them

- **they are therefore dimension reduction methods**
    - **take high-dimensional data (i.e. the many columns)**
    - **apply scaling and rotation**
    - **reduces the complexity to low-dimensional space (i.e. the orthogonal axes)**

Why ordination?

according to Gauch (1982): "Ordination primarily endeavours to represent sample and species relationships as faithfully as possible in a low-dimensional space". But why is this objective desirable? There are a number of answers, but most are derived from the 'properties of community' data as described above:

- it is impossible to visualise multiple dimensions simultaneously…
    - …while physicists grumble if dimensions exceeds four dimensions (space + time), ecologists typically grapple with dozens or hundreds of dimensions (species and/or samples)
- a single multivariate analysis saves time
    - imagine doing a separate univariate analysis for each one of the 800+ species of SA seaweeds!
- ideally and typically, dimensions of this 'low dimensional space' will represent important and interpretable environmental gradients

http://ordination.okstate.edu/overview.htm

Why ordination?

- if statistical tests are desired, problems of multiple comparisons are diminished when species composition is studied in its entirety
  - read about Type I and Type II errors
- statistical power is enhanced when species are considered in aggregate, because of redundancy
- by focusing on 'important dimensions', we avoid interpreting (and misinterpreting) noise; thus, ordination is a 'noise reduction technique' (Gauch 1982)
- we can **determine the relative importance of different gradients**; this is virtually impossible with univariate techniques
- community patterns may differ from population patterns
- some techniques provide a measure of beta diversity
- the graphical results from most techniques often lead to ready and intuitive interpretations of species-environment relationships

Ordinations

- **Principal Component Analysis (PCA):** the main eigenvector-based method. Works on raw, quantitative data. Preserves the Euclidean (linear) distances among sites and therefore mainly used for environmental data (but can be applied to species dissimilarities).
- **Correspondence Analysis (CA):** works on data that must be frequencies or frequency-like, dimensionally homogeneous, and non-negative. Preserves the χ^2 distances among rows or columns. Mainly used in ecology to analyse species data tables.
- **Principal Coordinate Analysis (PCoA):** devoted to the ordination of dissimilarity or distance matrices, most often in the Q mode, instead of site-by-variables tables. Hence, great flexibility in the choice of association measures.
- **Non-metric Multidimensional Scaling (nMDS):** unlike the three others, this is not an eigenvector-based method. nMDS tries to represent the set of objects along a predetermined number of axes while preserving the ordering relationships among them.
- PCoA and nMDS can produce ordinations from any square dissimilarity or distance matrix.

Taken verbatim from 5.2.2 in Numerical Ecology with R

Interpretation of ordination plots

- the **direction of the axes** (e.g. left vs. right; up vs. down) **is arbitrary** and should not affect the interpretation (i.e. no independent and dependent variables)
- the **numeric scale on the axis is not very useful for the interpretation** (an exception for this is Detrended Correspondence Analysis (DCA), in which the scales are in units of beta diversity)
- in most techniques (but not NMDS), **the order of the axes is important**. Thus, axis 1 is more important than axis 2, etc. The meaning of 'importance' depends on the technique employed, but ideally **related to the relative influence of environmental gradients**
- **third and higher axes can be constructed**. The choice of 'when to stop' interpreting new axes is largely a matter of taste, the quantity and quality of the data, and the ability to interpret the results
  - most of the techniques presented later provide supplemental statistics that can assist in the task
- it is desirable that axes not be correlated, because you would like them to represent different gradients, but **most techniques automatically result in uncorrelated** (or orthogonal) axes
- **a biologist's insight, experience, understanding of ecology, and knowledge of the literature are the most important tools for interpreting indirect gradient analysis**

Taken verbatim from http://ordination.okstate.edu/overview.htm

Unconstrained ordinations

- recap:
    - unconstrained ordinations are not statistical (no inference testing); they are purely descriptive
        - i.e. no inferences of causation
    - also called indirect gradient analysis
    - based only on the **environment × sites** matrix or the **species × sites** matrix, each seen in isolation

- ...apply to one multivariate (>>2) matrix (spp. or env.)
    - ✔ Principal Component Analysis (**PCA**); eigenvector | Euclidian distance (i.e. linear responses)
    - ✔ Correspondence Analysis (**CA**); eigenvector | χ^2-distance (i.e. non-linear responses)
    - ✘ Detrended Correspondence Analysis (**DCA**)
    - ✔ Principal Coordinates Analysis (**PCoA**); eigenvector | dissimilarities
    - ✔ Non-Metric Multidimensional Scaling (**NMDS**); non-metric | dissimilarities

- (Cluster analysis also applies to one matrix (spp. or env.) with >>2 variables)

Multivariate Techniques

Obs	Group	X-set				Y-set			
1	A	a_{11}	a_{12}	a_{13}	... a_{1p}	b_{11}	b_{12}	b_{13}	... b_{1m}
2	A	a_{21}	a_{22}	a_{23}	... a_{2p}	b_{21}	b_{22}	b_{23}	... b_{2m}
3	A	a_{31}	a_{32}	a_{33}	... a_{3p}	b_{31}	b_{32}	b_{33}	... b_{3m}
.	.	.	.	.	... .	.	.	.	... .
.	.	.	.	.	... .	.	.	.	... .
n	A	a_{n1}	a_{n2}	a_{n3}	... a_{np}	b_{n1}	b_{n2}	b_{n3}	... b_{nm}
n+1	C	c_{11}	c_{12}	c_{13}	... c_{1p}				
n+2	C	c_{21}	c_{22}	c_{23}	... c_{2p}				
n+3	C	c_{31}	c_{32}	c_{33}	... c_{3p}				
.	.	.	.	.	... .				
.	.	.	.	.	... .				
N	C	c_{n1}	c_{n2}	c_{n3}	... c_{np}				

Unconstrained ordination
Principal Component Analysis (**PCA**)
Correspondence Analysis (**CA**)
Detrended Correspondence Analysis (**DCA**)
Principal Coordinate Analysis (**PCOA**)
Non-Metric Multidimensional Scaling (**nMDS**)

also…

Cluster Analysis

http://www.umass.edu/landeco/teaching/multivariate/schedule/summary.handouts.pd