MODELS IN INVASION ECOLOGY - CHALLENGES AND APPLICATIONS

March, 18th-20th of 2019

# Multivariate statistics in ecology

A. Sofia Vaz

Joana R. Vicente

# Outline

- Introduction
- Gradient analysis

Dissimilarities and permutation tests

- Cluster analysis
- Unconstrained ordination
- Constrained ordination
- Ordination modelling
- Hands on…

# Introduction to multivariate analysis

Terminology

**Explanatory data** (predictors/independent) – environmental data

**Covariables** (covariates) – environmental data with an acknowledged/ hypothesized influence on species data

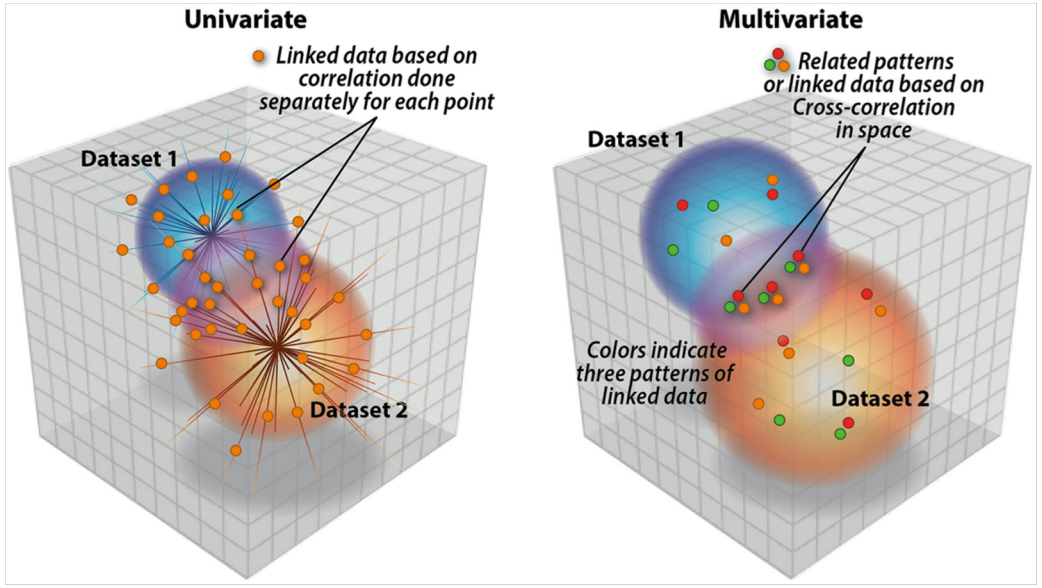
**Primary data** (response/dependent variable) – **species diversity data**

# Introduction to multivariate analysis

**Univariate**: one variable of interest (descriptive)

**Bivariate**: two variables of interest (regression, correlation)

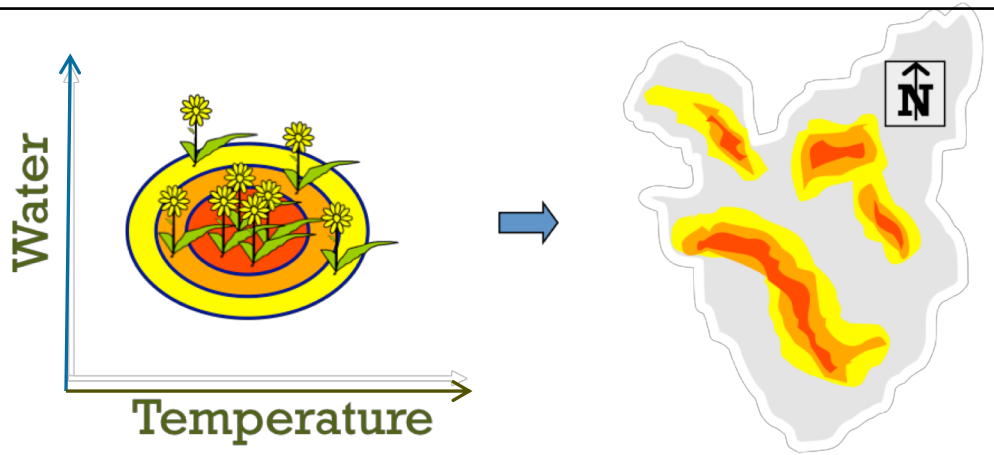
**Multivariate**: at least two variables of interest (multiple regression, ordination, clustering)



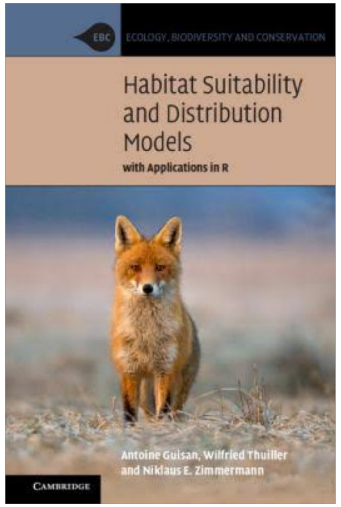
# Introduction to multivariate analysis

| Response variable | Predictors | |
|---|---|---|
| | Absent | Present |
| One | Distribution summary | - |
| Two | Correlation | Simple regression |
| Multiple | Unconstrained ordination, Cluster analysis | Multiple regression, Constrained ordination, Discriminant analysis |


Univariate: descriptive




Bivariate & Multivariate: exploratory and predictive



# Introduction to multivariate analysis

**Species diversity data**

1 - Diversity on a sample plot, or ‘point’ diversity, or site (α)

2 - Diversity along ecological gradients (β)

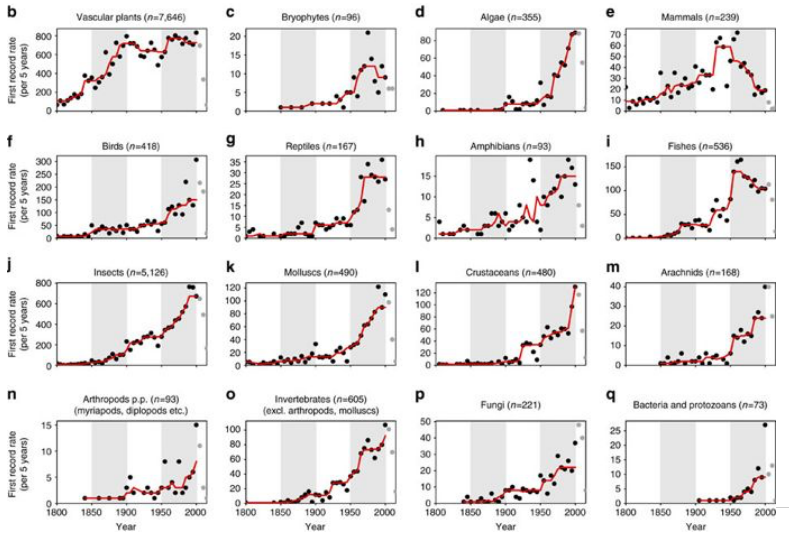
3 - Diversity among parallel gradients or classes of environmental variables (γ)

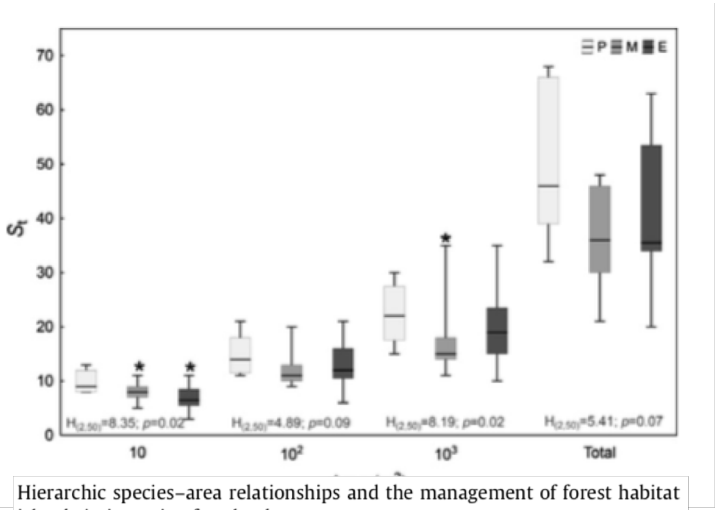
The total diversity of a landscape: sum of all previous.

# Local diversity

## Indices and models

- Shannon diversity
- Simpson diversity
- Hill numbers
- Evenness
- …




Hierarchic species–area relationships and the management of forest habitat islands in intensive farmland

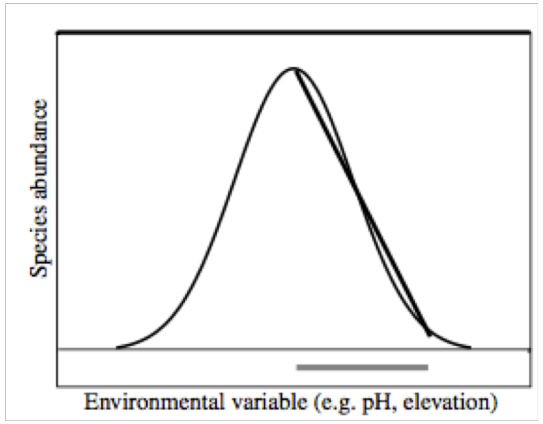
Angela Lomba [a,b,c,*], Ana Sofia Vaz [a,d], Francisco Moreira [e], João Pradinho Honrado [a,d]

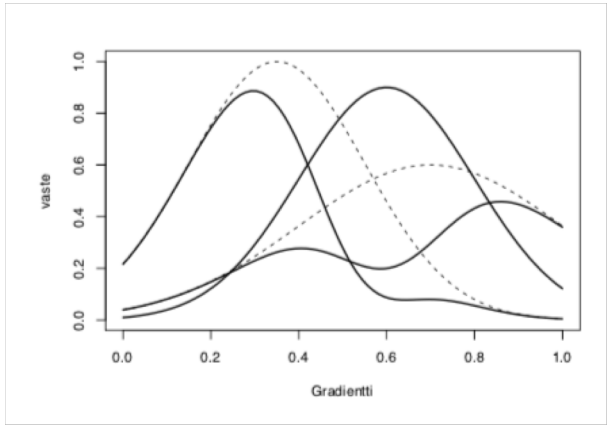

Rarefaction curves

Species-area relationships

…

# Gradient analysis




Species respond to complex and interrelated environmental gradients

Species have non-linear responses along gradients...but Gaussian…with many shapes!

# Gradient analysis

**Multivariate**: multiple regression

Simple regression: $Y = b0 + b1\ x$.

Multiple regression: $Y = b0 + b1\ x1 + b0 + b1\ x2 \ldots b0 \ldots b1\ xn$.

*E.g. GLM, GAM*

Focusing on one gradient or more gradients (fitted individually) and their interactions

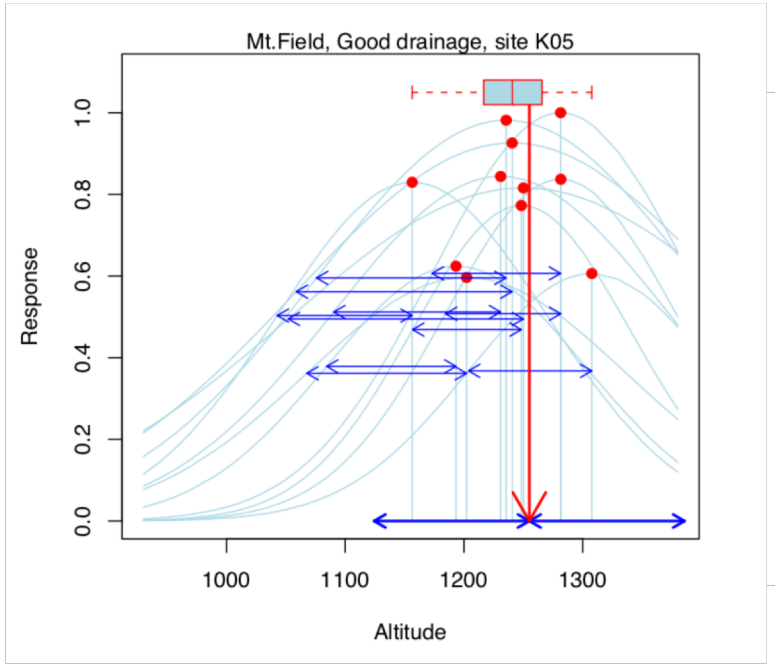
# Gradient analysis

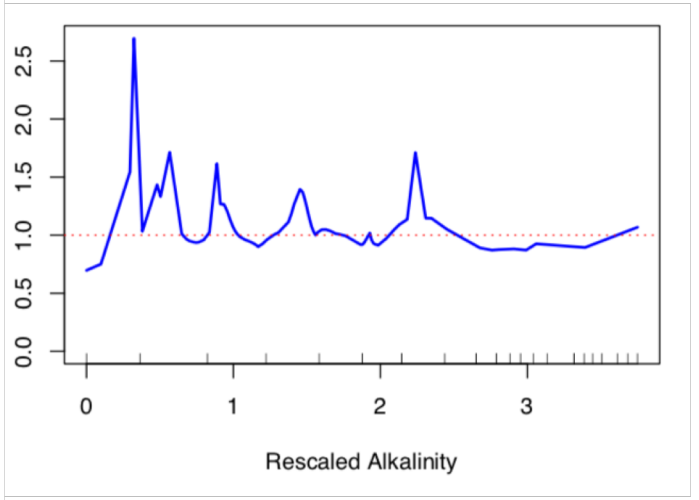
## Smoothing

Hill rescaling

Weighted average

Likelihood approach




Goodness of fit

# Gradient analysis

**Goodness of fit tests**

Permutation (null model)

Cross-validation

Bootstrap

Jacknife


**Permutation**
Randomization test

**Bootstrap**

**Jacknife**

# Gradient analysis

**Permutation-based tests**

Test for significance on groups' differences

Assess relationships between species and environmental data

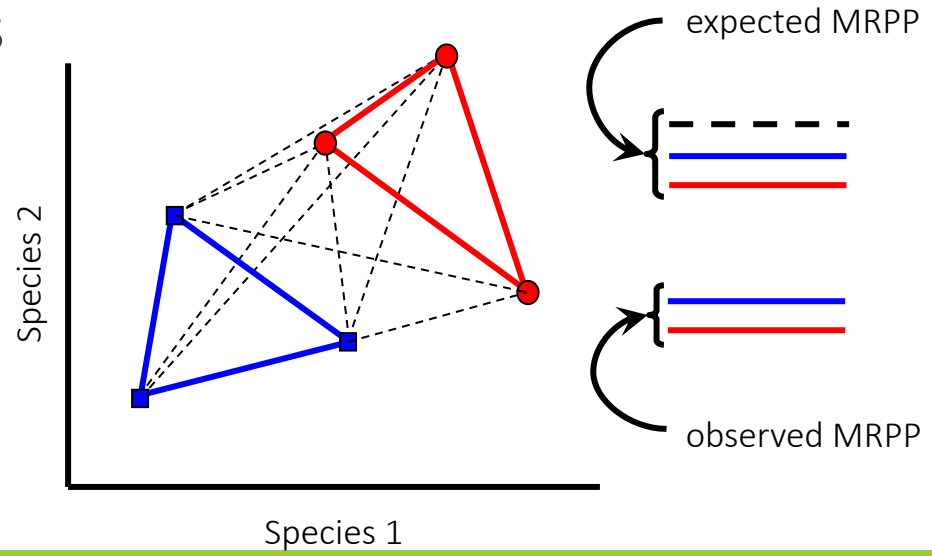
- NPMANOVA (Non-Parametric MANOVA)
- MRPP (MultiResponse Permutation Procedure)
- ANOSIM (Analysis of Similarity)
- Mantel Test

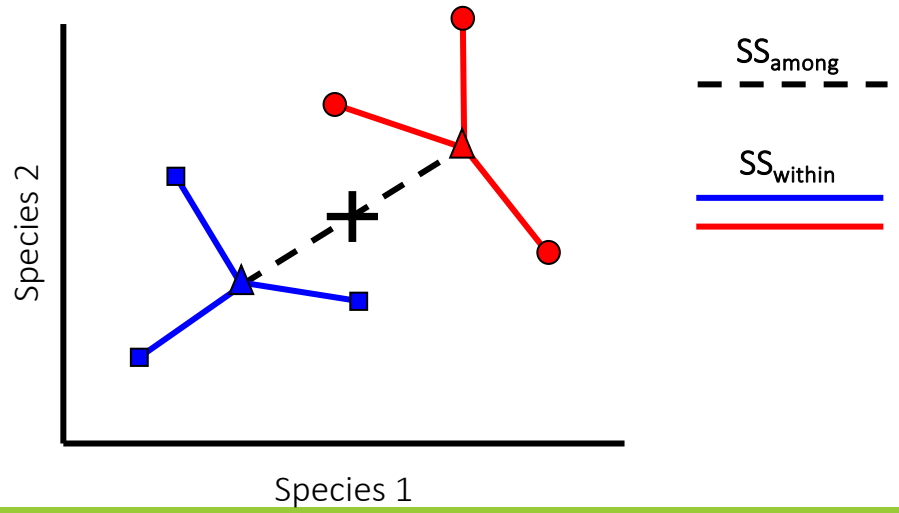
# Permutation-based tests

**Mantel test:** testing correlation between two/three distance matrices

**ANOSIM**: testing for difference between multivariate groups, based on ranked distance measures (minimum of 2 factors, 2 levels each)

**MRPP** and **NPMANOVA :** testing differences and interactions in means among vectors



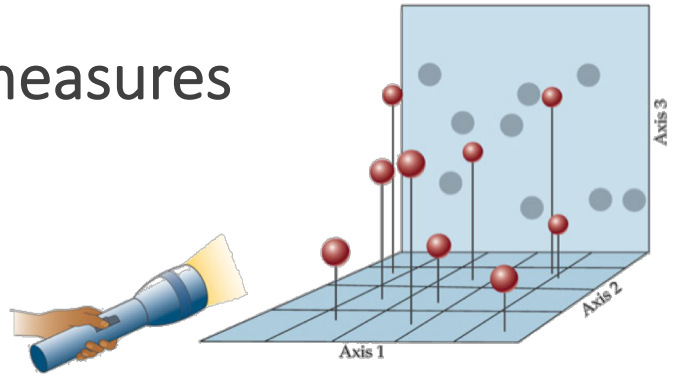



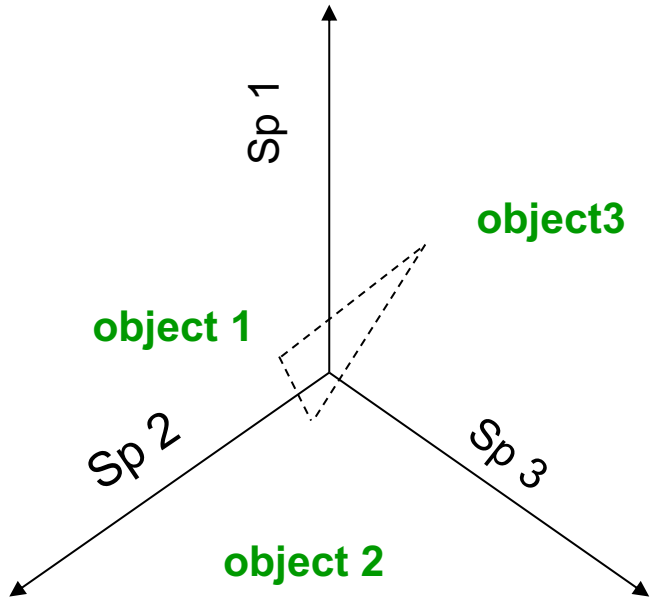
# Ordination analysis

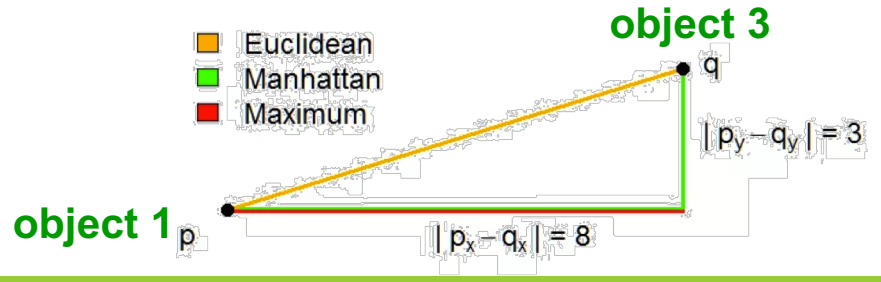
## Proximity implies similarity!

### Distance/dissimilarity measures

- Euclidean
- Jaccard index
- Manhattan
- Bray-Curtis
- Morisita
- Geographic coordinates
- 1-Correlation…



# Ordination analysis ordination

**Unconstrained ordination** – indirect gradient analysis

- Only considers dissimilarity data on computations

- Environmental data may be overlaid but do not influence the calculation

**NMDS** – Non-metric multidimensional scaling

**PCA** - Principal components analysis

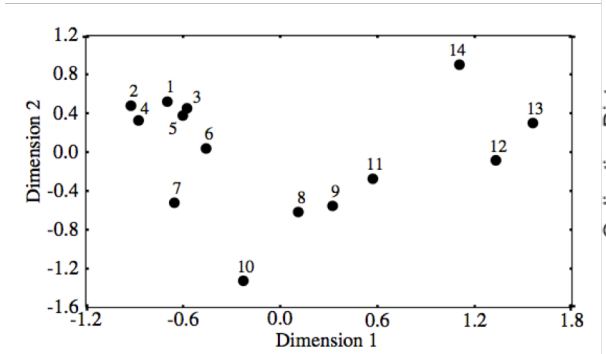
**CA** - Correspondence analysis

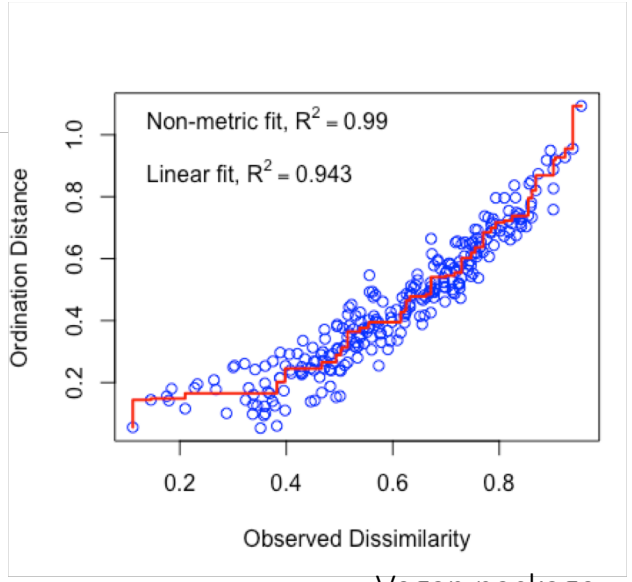
**DCA** - Detrended correspondence analysis

# Unconstrained ordination

**NPMDS** – a.k.a. NMDS or NMS – few number of axes are chosen prior to the analysis and the data are fitted to those dimensions (there are no hidden axes)

- Visualization method
- Different solutions to minimize stress
- Goodness-of-fit called “stress”
- Can take any configuration
- Small assumptions on data
- Any distance measure



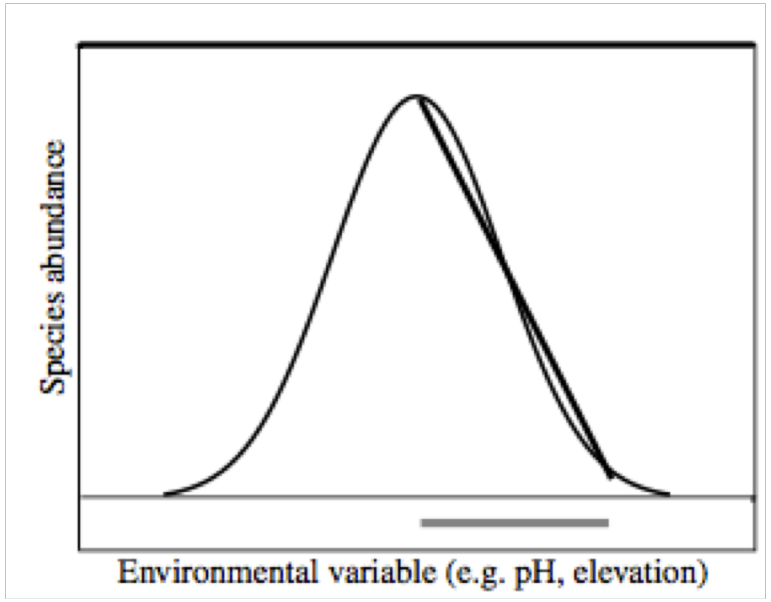

Vegan package

# Unconstrained ordination

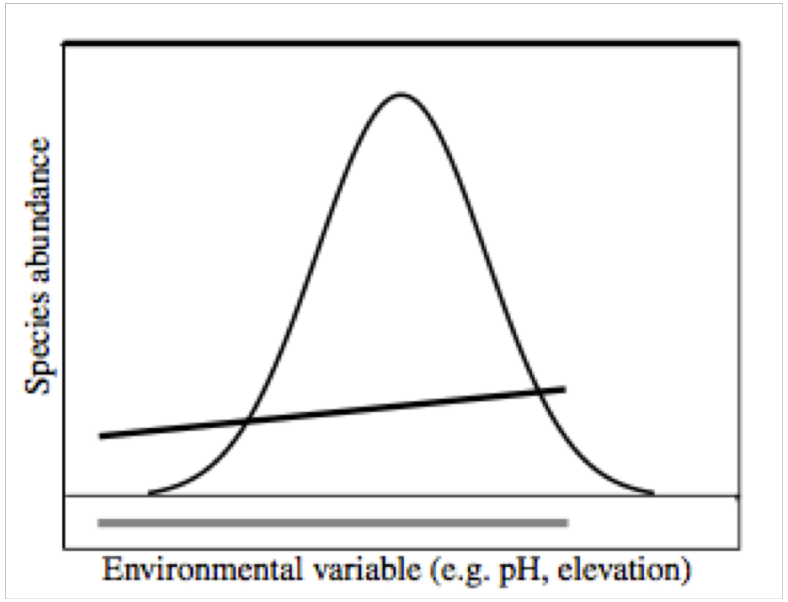
**MDS** – a.k.a. PCoA – principal coordinate analysis or multidimensional scaling

- Visualization method
- It aims at represent real distances/similarities in the first axis
- Any distance measure
- Can be done based on real distances

# Unconstrained ordination



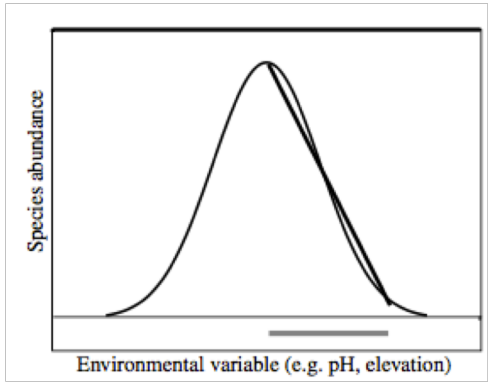

linear approximation of a unimodal response curve over a short part of the gradient - **PCA**



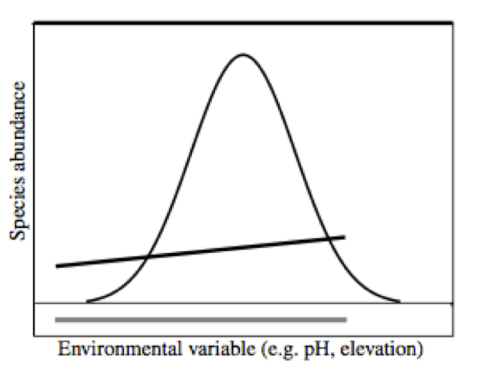

linear approximation of a unimodal response curve over a large part of the gradient - **CA**

Lepš and Šmilauer, 2003

# Unconstrained ordination




Lower turnover rates




High turnover rates

```
decorana(veg = trees)

Detrended correspondence analysis with 26 segments.
Rescaling of axes with 4 iterations.

                 DCA1      DCA2      DCA3      DCA4
Eigenvalues     0.03901 0.0020688 0.0009308 0.0007487
Decorana values 0.03904 0.0005059 0.0001835 0.0001051
Axis lengths    0.52369 0.1261756 0.0948353 0.0843756
```

Estimation of species optimum from the response curve

< 3 : low turnover rates

> 4 : high turnover rates

# Unconstrained ordination

PCA – produces linear combinations of the original variables to generate the axes

- Visualization and reduction method
- Usually based on the Euclidean distance
- New axes are formed to capture data variance

# Unconstrained ordination

CA – produces weighted combinations of the original variables to generate the axes

- Visualization and reduction method

- Usually based on the Chi-square distance

- New axes are formed to capture data inertia

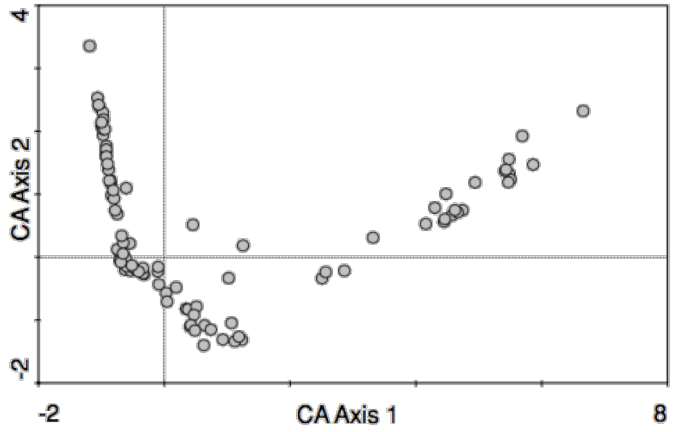
# Unconstrained ordination

DCA – detrended weighted combinations of the original variables to generate the axes

- Visualization and reduction method

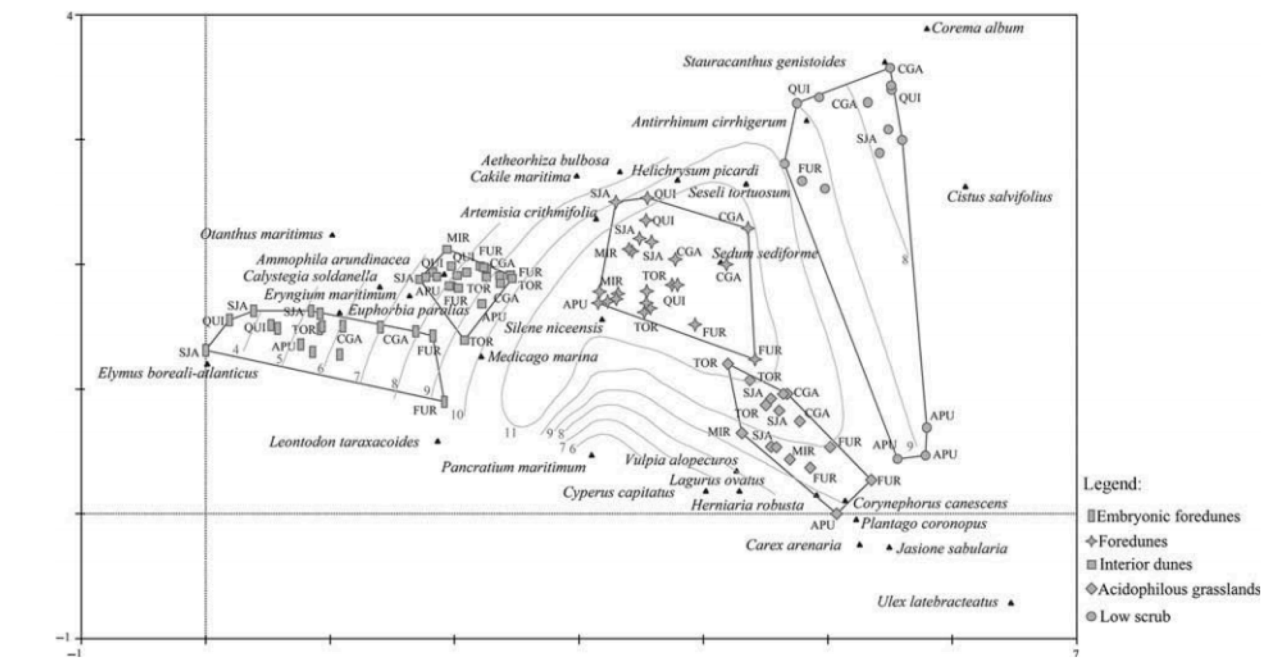
- Based on the CA, but follows detrended to rescale axes

- Used to avoid arch effects in data (artifact by CA)



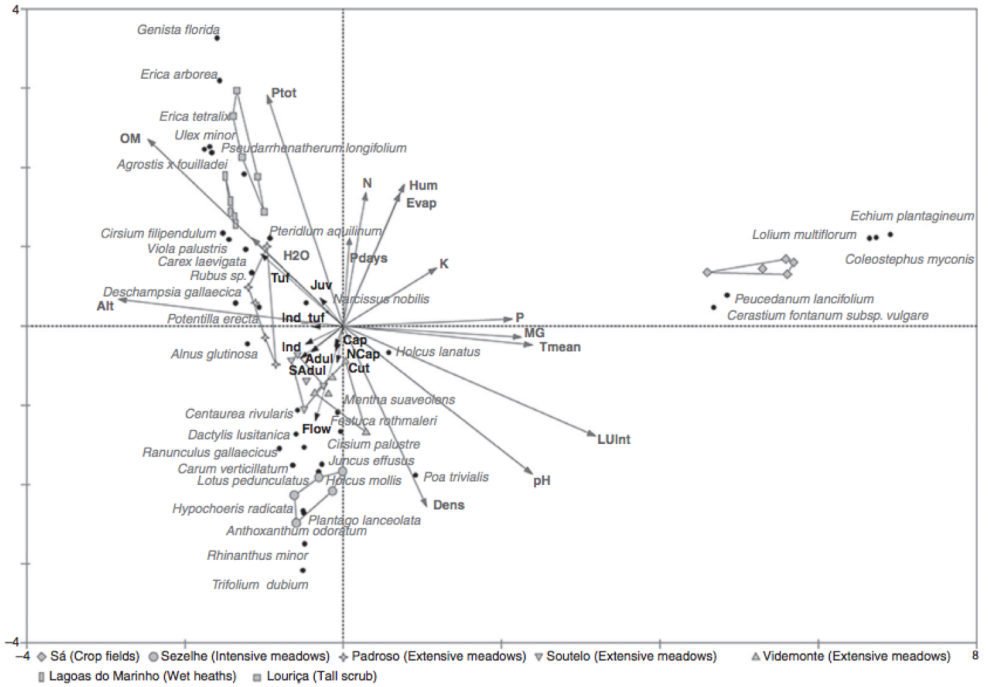

Lepš and Šmilauer, 2003

# Unconstrained ordination




**Plant species segregation in dune ecosystems emphasises competition and species sorting over facilitation**

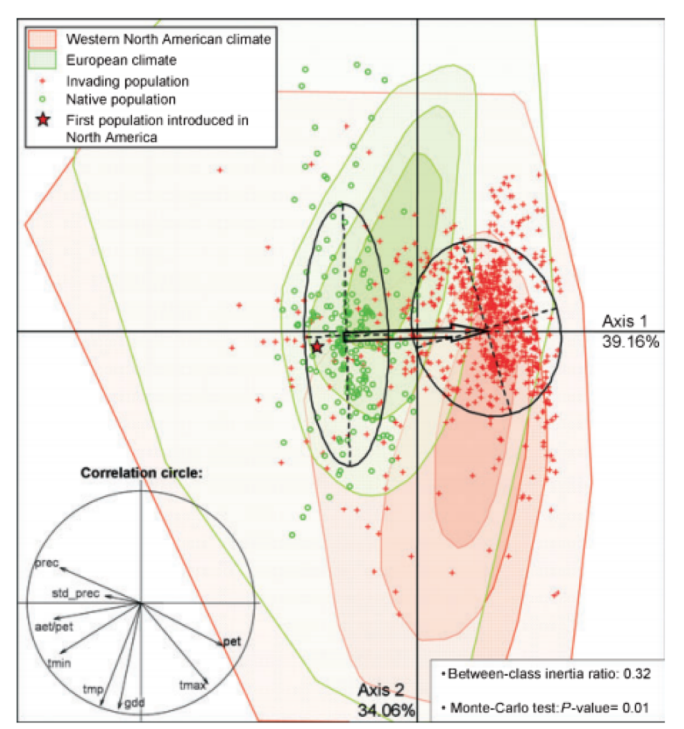
Ana S. Vaz[a*], José A. Macedo[a,b], Paulo Alves[a], João P. Honrado[a,b] and Angela Lomba[a,c,d]




Evaluating population and community structure against climate and land-use determinants to improve the conservation of the rare *Narcissus pseudonarcissus* subsp. *nobilis*

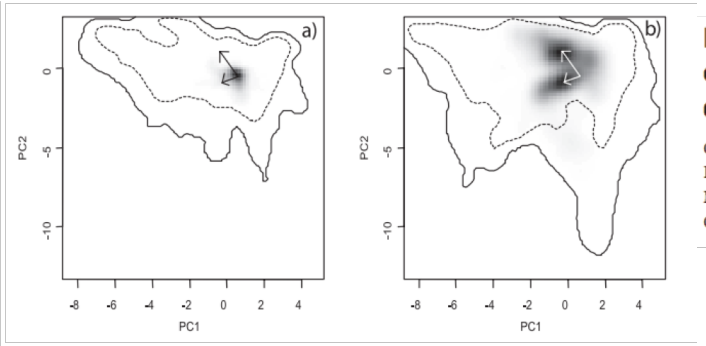
Ana Sofia Vaz*[1,2], Duarte Silva[3], Paulo Alves[1], Joana Raquel Vicente[1], Francisco Barreto Caldas[1,2], João Pradinho Honrado[1,2] & Angela Lomba[1,4,5]

# Unconstrained ordination



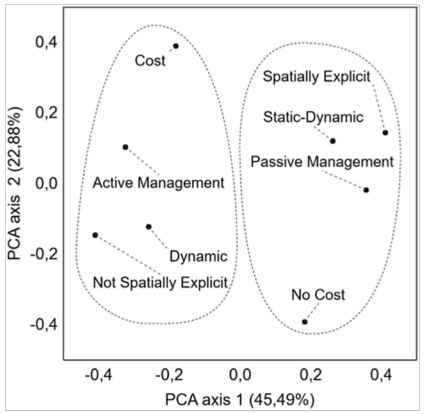

*Ecology Letters*, (2007) **10**: 701–709 doi: 10.1111/j.1461-0248.2007.01060.x

**Evidence of climatic niche shift during biological invasion**



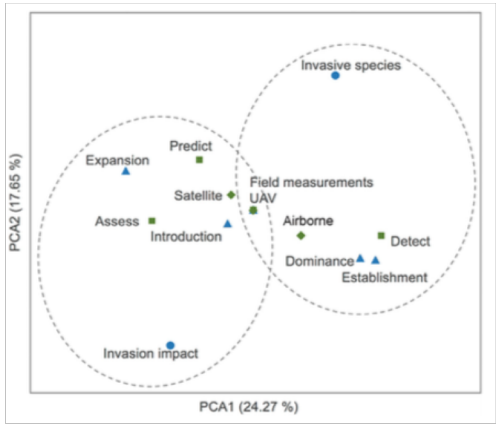

**Measuring ecological niche overlap from occurrence and spatial environmental data**

Olivier Broennimann[1]*,†, Matthew C. Fitzpatrick[2]†, Peter B. Pearman[3]†, Blaise Petitpierre[1], Loïc Pellissier[1], Nigel G. Yoccoz[4], Wilfried Thuiller[5], Marie-Josée Fortin[6], Christophe Randin[7], Niklaus E. Zimmermann[3], Catherine H. Graham[8] and Antoine Guisan[1]




Dynamic models in research and management of biological invasions

Ana Buchadas [a,*], Ana Sofia Vaz [a], João P. Honrado [a], Diogo Alagador [b], Rita Bastos [c], João A. Cabral [c], Mário Santos [c], Joana R. Vicente [a,c]




Managing plant invasions through the lens of remote sensing: A review of progress and the way forward

Ana Sofia Vaz [a,b,*], Domingo Alcaraz-Segura [c,d,e], João C. Campos [a], Joana R. Vicente [a,f], João P. Honrado [a,b]

# Constrained ordination

**Constrained ordination** – direct gradient analysis

- Considers species and environmental data on computations

**RDA -** redundancy analysis

**CCA -** canonical correspondence analysis

**DCCA –** detrended canonical correspondence analysis

…

# Constrained ordination

RDA and CCA – the new gradients must be linear combinations of the provided explanatory (environmental) variables

- Reduction, explanatory (and predictive) method

- Ordination axes capture the effect of predictors on species data

- Method resembles multivariate regressions

# Ordination modelling

**Handling missing data** – absence is different from a missing value

- Remove samples with missing values
- Remove variables with missing values
- Apply an imputation method:
  - Average value of the variable
  - Multiple regression model
  - …

# Ordination modelling

**Data transformation** – species and environmental data

- Rare species versus over-represented species

- Interpret species - predictors relations
  - Log-transformation
  - Square-root transformation

- Deal with cover/abundance classes (e.g. Braun-Blanquet)
  - Number ordered classes
  - Use the value of the centre of each class

# Ordination modelling

**Data standardization** – species and environmental data

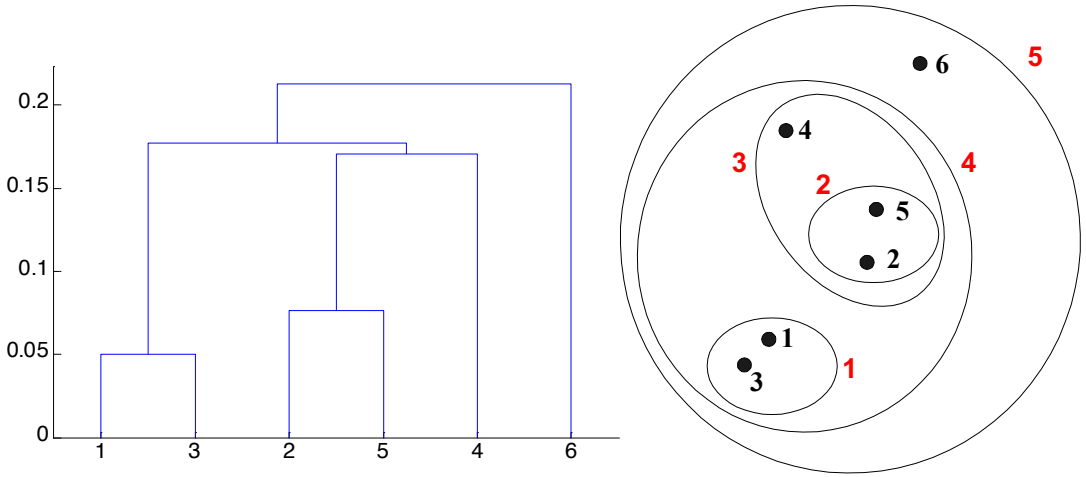
- Species and predictors with different orders of magnitude
  - Standardization to unit variance
  - Vector norm
  - Divide by margin total
  - Hellinger standardization
  - Chi-square
  - …

# Ordination modelling

**Applied to constrained analysis**

- RDA, CCA, DCCA
- Predictor selection is done through permutations
- Model evaluated through AIC (comparing different predictor combinations )
- The contribution of covariates is only accounted as explained variance/inertia

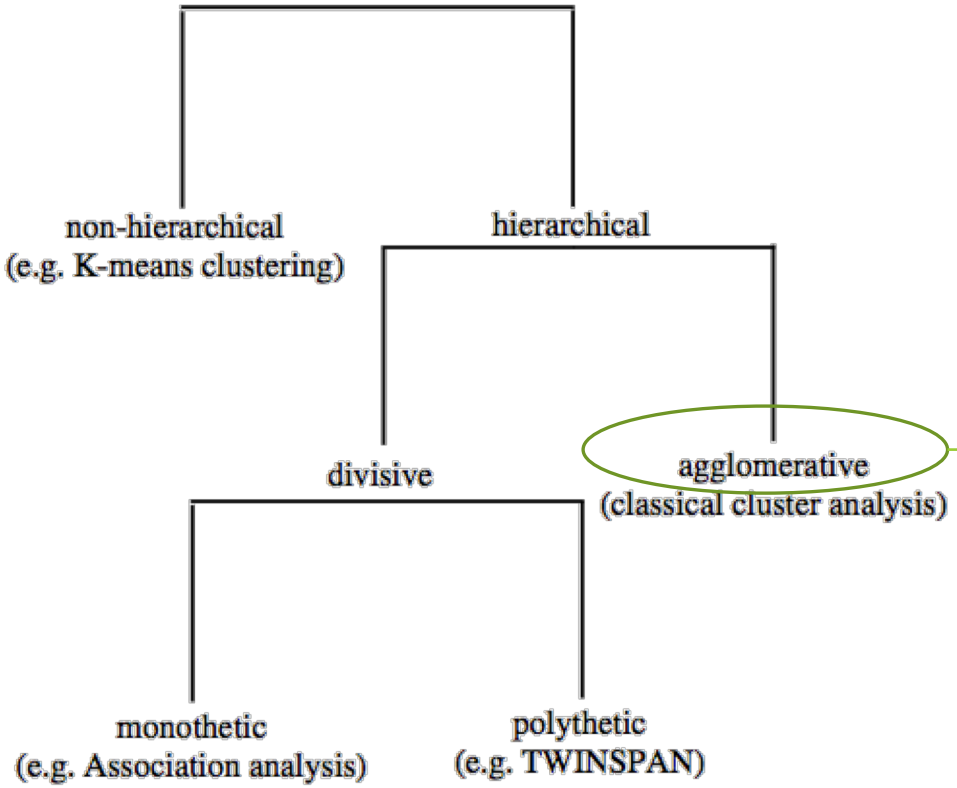
# Cluster analysis

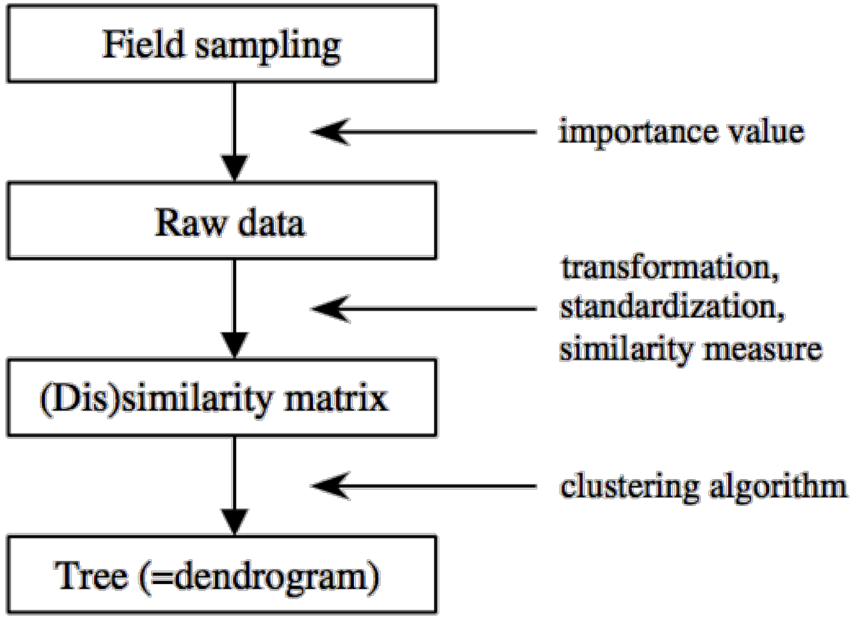



Classification of the similarity/distance among species/sites

# Cluster analysis

## Hierarchical classification: the basics…

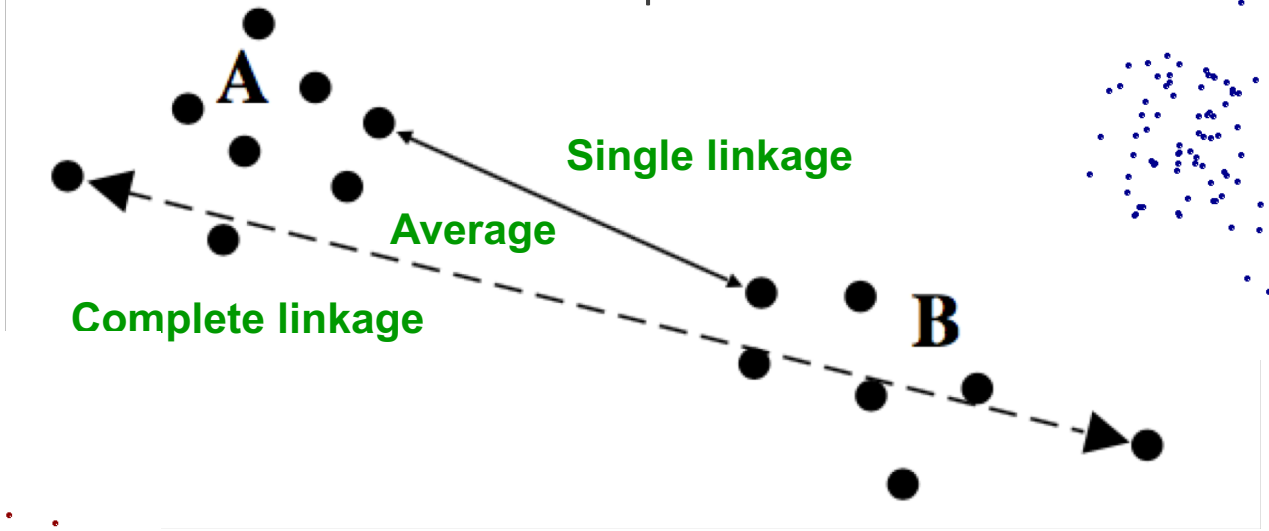


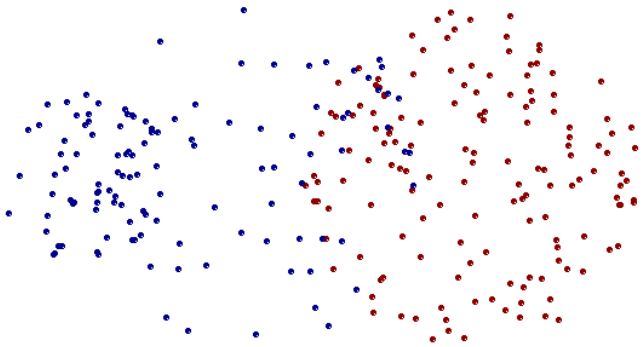



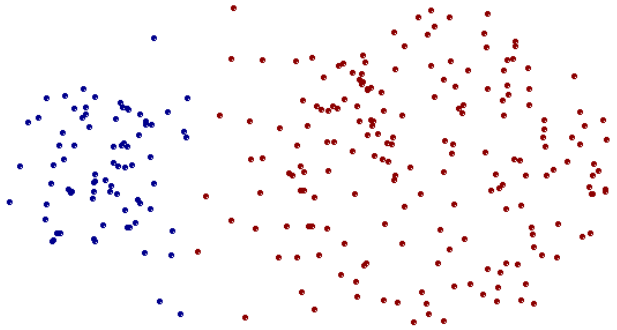

Lepš and Šmilauer, 2003

# Cluster analysis

Agglomerative classification: bottom-up



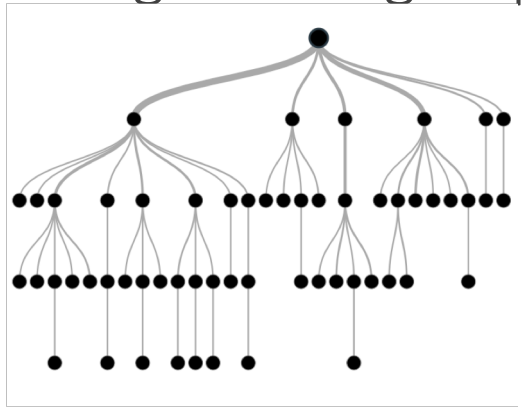


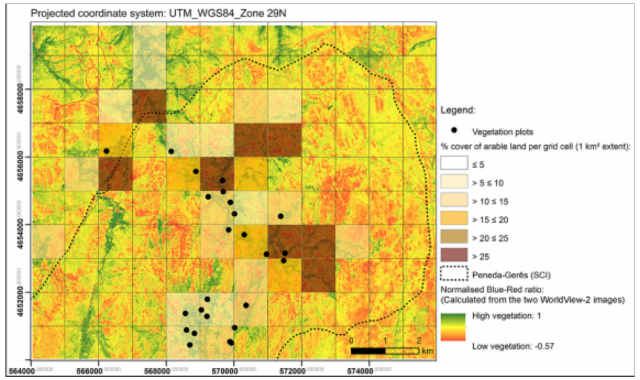
UPGMA
Centroid distance
Ward's distance
K-means....

# Cluster analysis

**Cluster analysis**: clustering/classification of objects into homogeneous groups

- Machine learning *(CTA, BRT in BIOMOD2)*
- Image classification *(K-means, in remote sensing)*
- Pattern recognition *(in psychology and medicine)*
- Bioinformatics *(in phylogeny)*
- …

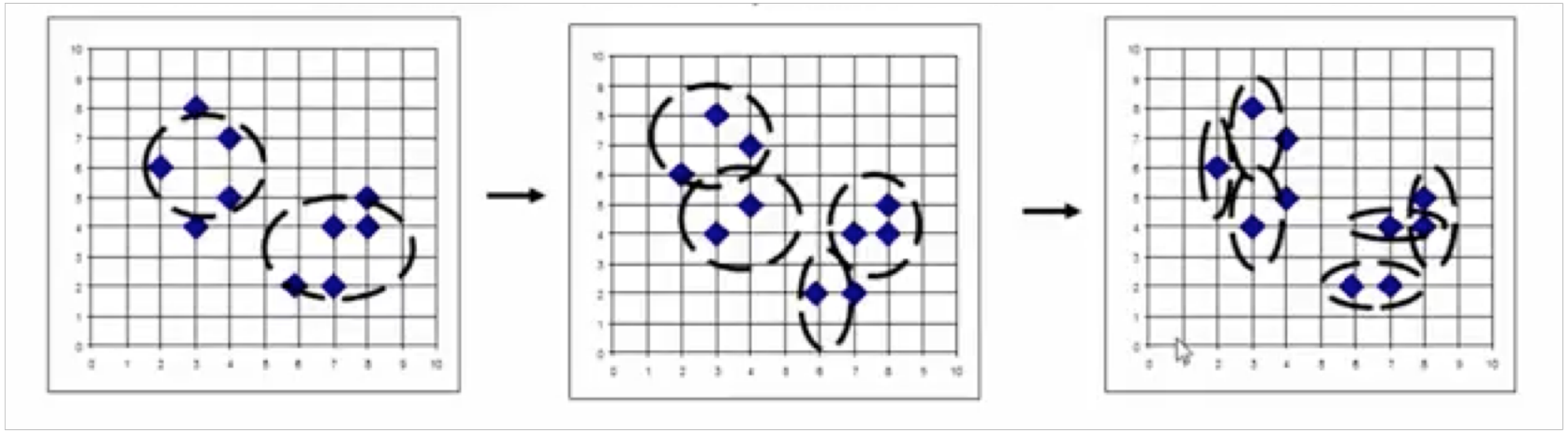





Vaz et al. 2015. *JAE*

# Cluster analysis

**Divisive classification**: top-down



Monothetic versus polythetic

Ward's criterion
GINI index
...

# Hands on – Part I

- **Vegan Package (needs MASS)**
- Species data – 44 tree species (non-native and native) per 22 sites
- Environmental data – soil data for 22 sites + management and invasion levels
- Unconstrained ordination – NMDS, PCA, CA, DCA
- Constrained ordination – CCA, predictor selection, and covariable modelling
- Factor analysis – MANOVA, Mantel, Procrustes
- Basic clustering + dendograms