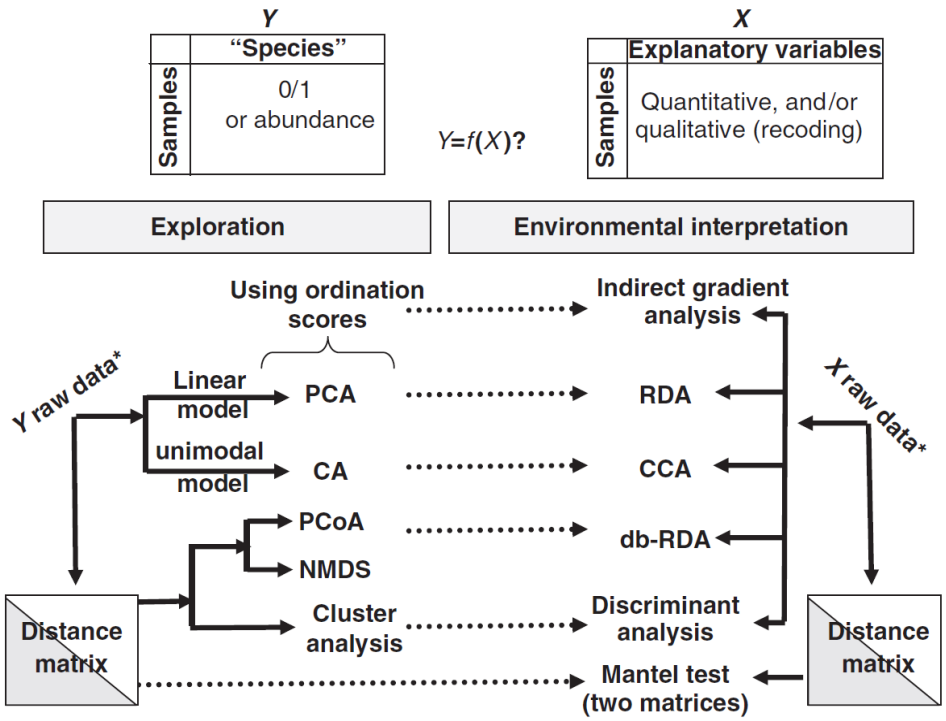
Auxiliary multivariate analyses

- Linear Discriminant Analysis (LDA)
- Mantel correlation (Mantel correlogram)
- Procrustes rotation and superposition
- ANOSIM (Analysis of Similarity)
- SIMPROF (Similarity Profiles)
- Multivariate permutational ANOVA (PERMANOVA)
- Simper (similarity percentage)
- variance partitioning

…

Linear Discriminant Analysis (LDA)



Note: assumption of homogeneity of within-group covariance (see PERMANOVA)

Ordination technique maximizing group separation.

LDA uses **a single variable** classifying sites into groups (different to RDA & CCA).

This grouping may represent a hypothesis or be obtained using cluster analysis of **another** dataset.

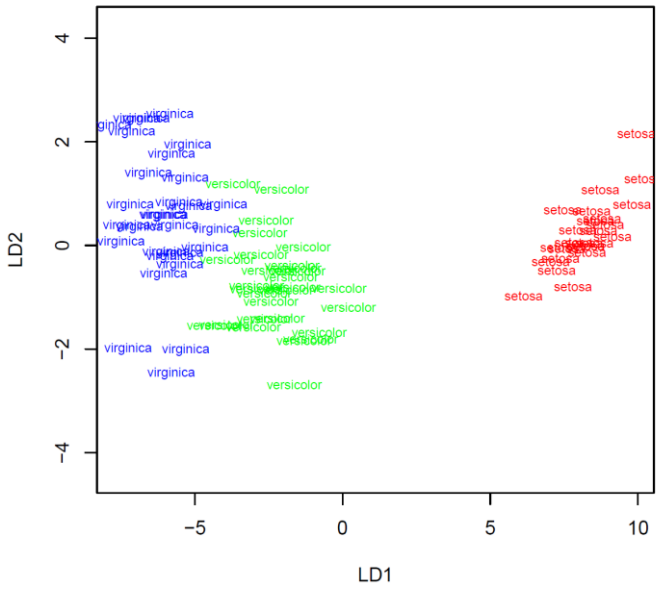
(important: clustering must have been obtained independently from the variables used in the LDA; otherwise the procedure would be circular)

EPFL

LDA example: iris dataset

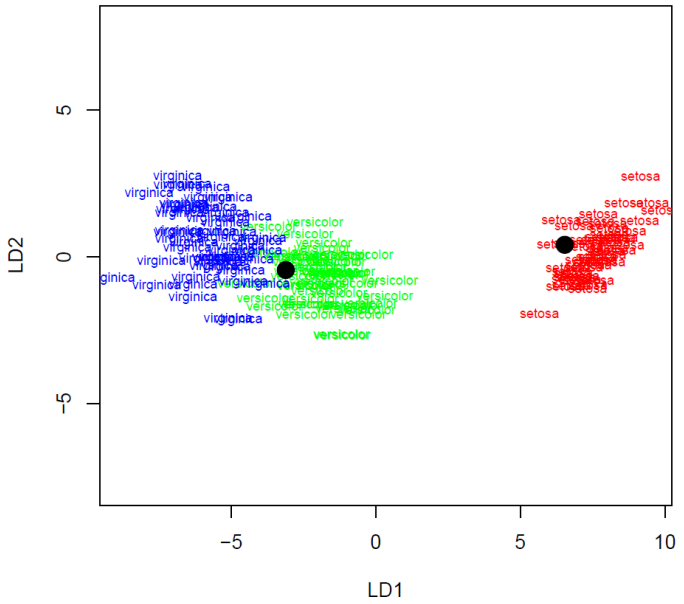
- Iris dataset: measurements of sepal and petal length and width…
- LDA on iris dataset maximizes the separation of the three species

```
>library(mass)
>iris.lda <- lda(Species~., iris)
>plot(iris.lda, col=c("red","green","blue")[iris$Species])
>summary(iris.lda)
```



Can also be used to classify new objects:

```
>predict.unknown.iris<-predict(iris.lda,
newdata=unknown.iris)
>points(predict.unknown.iris$x, pch=16, cex=2)
```



EPFL

Mantel test

- test of correlation between two distance or similarity matrices
- Uses permutations to establish significance
- Different types of correlation statistics possible (e.g. Pearson, Spearman, Kendall)
- example:
  - Genetic distance between taxa
  - Environmental dissimilarity or spatial distance
  - Q: is genetic distance between taxa correlated with dissimilarity of the environment, or does genetic distance between taxa increase with increasing spatial distance?

>mantel(xdis, ydis, method="pearson", permutations=999)

Mantel correlogram

example: mite dataset (35 mites at 70 locations, xy coordinates available)

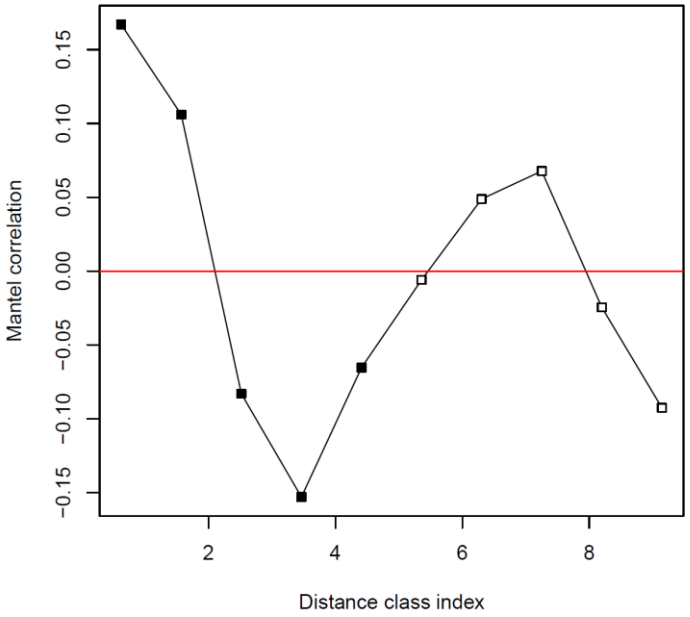
- A correlogram is (typically) a graph in which correlation statistic (i.e. Mantel correlation) is plotted as a function of geographic distance classes (or time lags).

=> test of spatial (or temporal) autocorrelation

- In a Mantel correlogram the correlation is computed between a multivariate distance matrix (e.g. species dissimilarity) and a design matrix (typically spatial/temporal distance classes).

>mite.correlog <-mantel.correlog(mite.hel.D, XY=mite.xy, n.class=10, r.type="pearson", nperm=999, cutoff=FALSE)

	class.index	n.dist	Mantel.cor	Pr(Mantel)	Pr(corrected)
D.cl.1	0.6234366	532.0000000	0.1670561	0.001	0.001 ***
D.cl.2	1.5703098	966.0000000	0.1059971	0.001	0.002 **
D.cl.3	2.5171829	914.0000000	-0.0830506	0.001	0.003 **
D.cl.4	3.4640561	706.0000000	-0.1529721	0.001	0.004 **
D.cl.5	4.4109293	522.0000000	-0.0653801	0.001	0.005 **
D.cl.6	5.3578024	470.0000000	-0.0059279	0.374	0.374
D.cl.7	6.3046756	318.0000000	0.0488553	0.026	0.052 .
D.cl.8	7.2515488	236.0000000	0.0677753	0.008	0.024 *
D.cl.9	8.1984220	120.0000000	-0.0244946	0.217	0.434
D.cl.10	9.1452951	46.0000000	-0.0925184	0.003	0.015 *

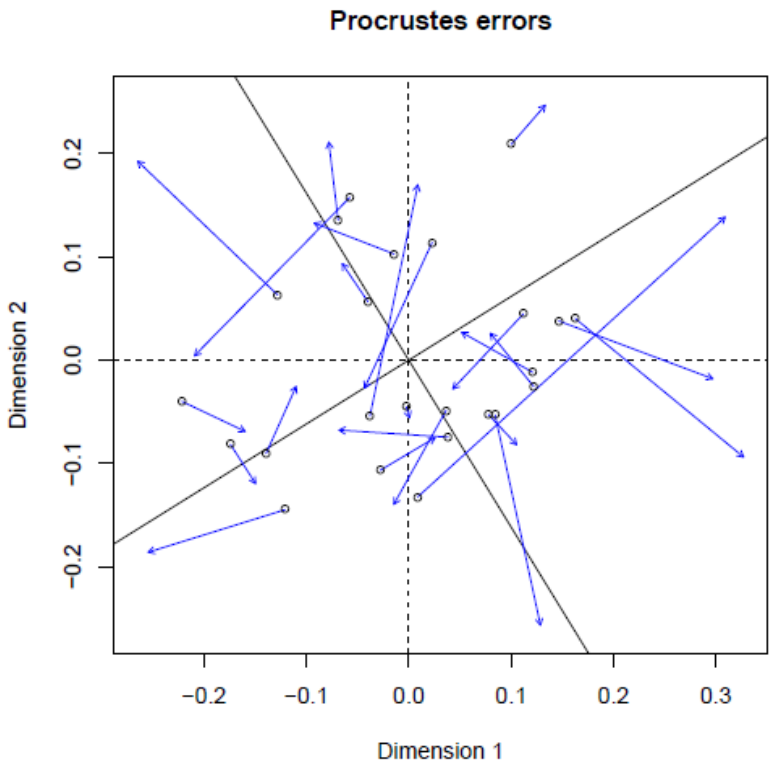


Procrustes analysis

- *Procrustes rotation* rotates (and rescales) an ordination to maximize similarity with a target ordination.
- Procrustes rotation is typically used for comparison of 2 independent ordinations (superimposition, see example).
- Allows (visual) estimation of the congruence between two ordinations (e.g. two ordinations using different taxa, or two different ordination techniques applied to the same dataset)

```
>procrustes(X, Y, scale = TRUE)
>protest(X,Y)
```

See also: co-inertia ade4::coinertia;
cocoresp::coca



ANOSIM (Analysis of Similarity)

- A non-parametric (rank) test of difference (equivalent to ANOVA) between groups based on any distance/dissimilarity measure
- significance testing using permutations
- No assumption regarding normality (non-parametric); useful for skewed species abundance data
- example Q: are there differences in similarity between groups of samples?

```
>anosim(species_site_matrix, grouping_sites, permutations = 999, distance = "bray“)
```

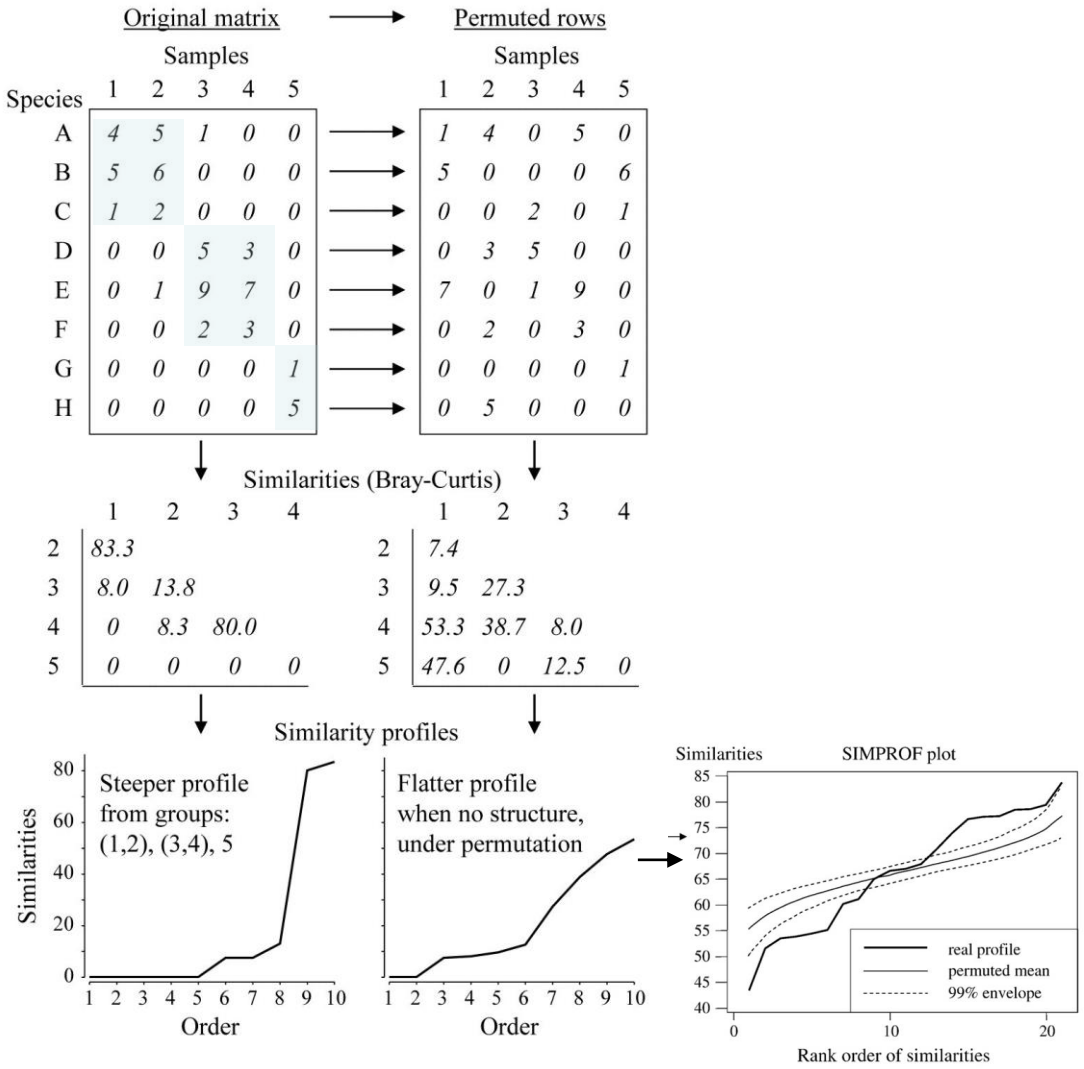
SIMPROF - similarity profile routine

Contents lists available at ScienceDirect
Journal of Experimental Marine Biology and Ecology
journal homepage: www.elsevier.com/locate/jembe

Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage

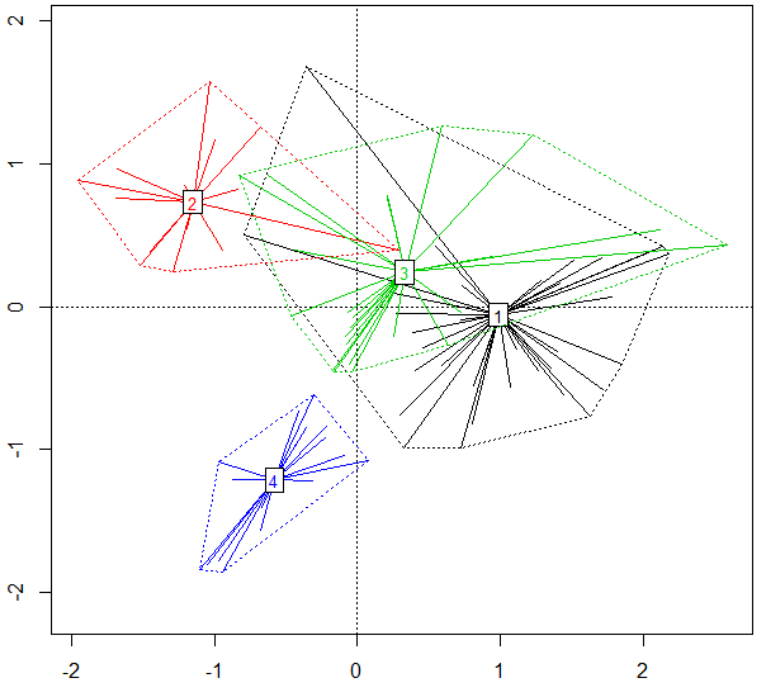
K. Robert Clarke [a,b,*], Paul J. Somerfield [a], Raymond N. Gorley [b]

- general test for multivariate structure
  (does not require a priori grouping)
- Q: is there non-random structure in the dataset?
- well-suited in addition to cluster analysis/exploratory analysis



PERMANOVA/NPMANOVA

- Non-parametric permutation-based test of differences between two or more groups based on any distance measure
- Useful to describing how variation is attributed to different experimental treatments.



- Result depends on **location** and **dispersion!** Location and dispersion effects can be confounded: differences may be caused by different within-group variation (dispersion) instead of different centroids.
- Dispersion (distances of samples from group centroid) is often used for analysis of beta-diversity (*betadisper* in vegan).

Note: betadisper can be used to check assumption in LDA!

PERMANOVA (test of differences in centroids):
> adonis2(dune ~ Management, data=dune.env, permutations=99)

TEST OF DISPERSION
> betadisper(d, group, type = c("centroid"), bias.adjust = FALSE)

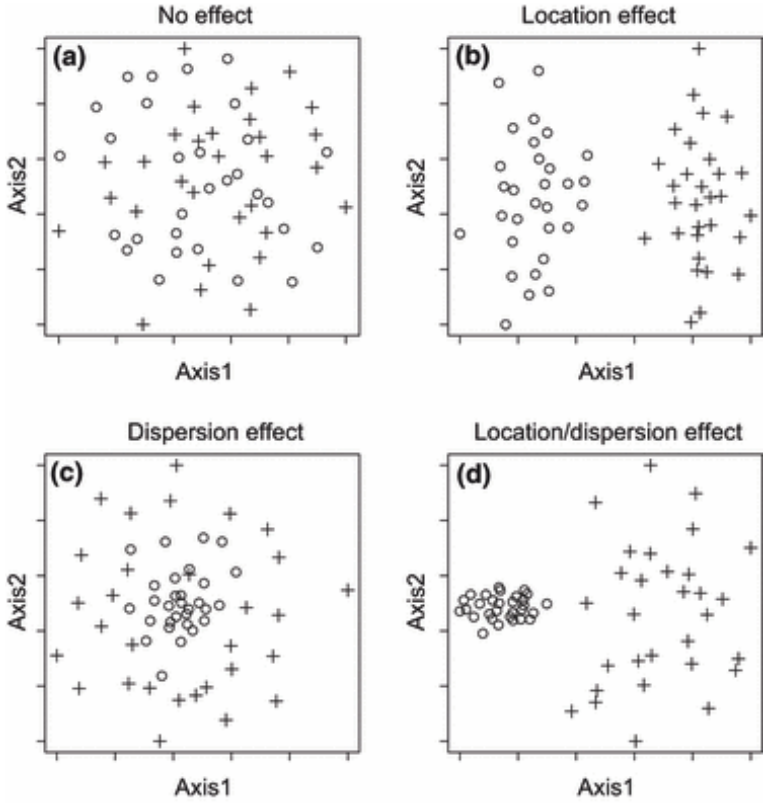
EPFL

Ecological Monographs, 83(4), 2013, pp. 557–574

PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing?

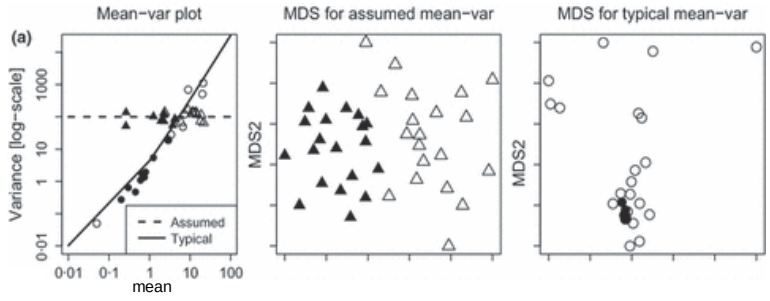
MARTI J. ANDERSON[1,3] AND DANIEL C. I. WALSH[2]

[1]*New Zealand Institute for Advanced Study (NZIAS), Massey University, Albany Campus, Private Bag 102 904, Auckland 0745 New Zealand*
[2]*Institute of Natural and Mathematical Sciences (INMS), Massey University, Albany Campus, Private Bag 102 904, Auckland 0745 New Zealand*



- PERMANOVA typically the most powerful and robust technique to test for differences in community structure.

- generalised linear models (GLM) can deal with mean-var relationships
  - R::*mvabund*
  - manyglm(abund~treatment, family="negative.binomial")



Simper (similarity percentage)

- method for assessing which taxa are primarily responsible for an observed difference between groups of samples.

- typically used with Bray-Curtis similarity.

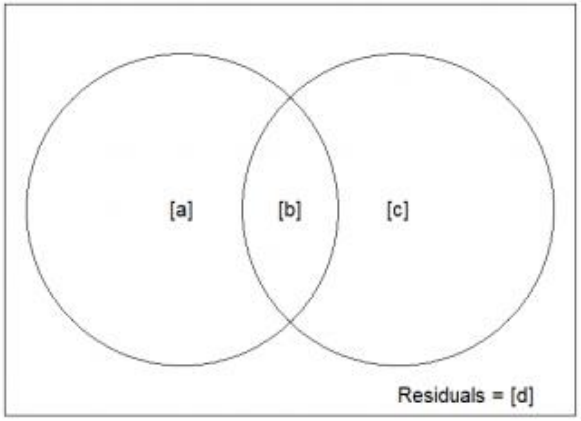
- example: there are two groups of samples – which taxa contribute to dissimilarity between samples?

```
>simper(comm, group, permutations=999)
```

Variance partitioning

- Can be used to explain the relative importance (% variance explained) between two or more two sets of explanatory variables (see constrained ordinations).
- Common example: Q: how much variation in a species dataset is explained by the environment compared to spatial distance?
- Results can be visualized using Venn diagram

Venn diagram



[a] - variation explained by dataset X1
[c] - variation explained by dataset X2
[b] - shared variation explained by both
[d] - unexplained variation.

```
>browseVignettes("vegan") -> partitioning
>var_part_results<-varpart(Y,X1,X2, data)
>showvarparts(var_part_results)
```

Note: for spatial data, consider ordination of the spatial relationships among sampling locations using "spatial eigenvectors (e.g. PCNM, MEM)