Chapter 11
CANONICAL Analysis

Wha whaa

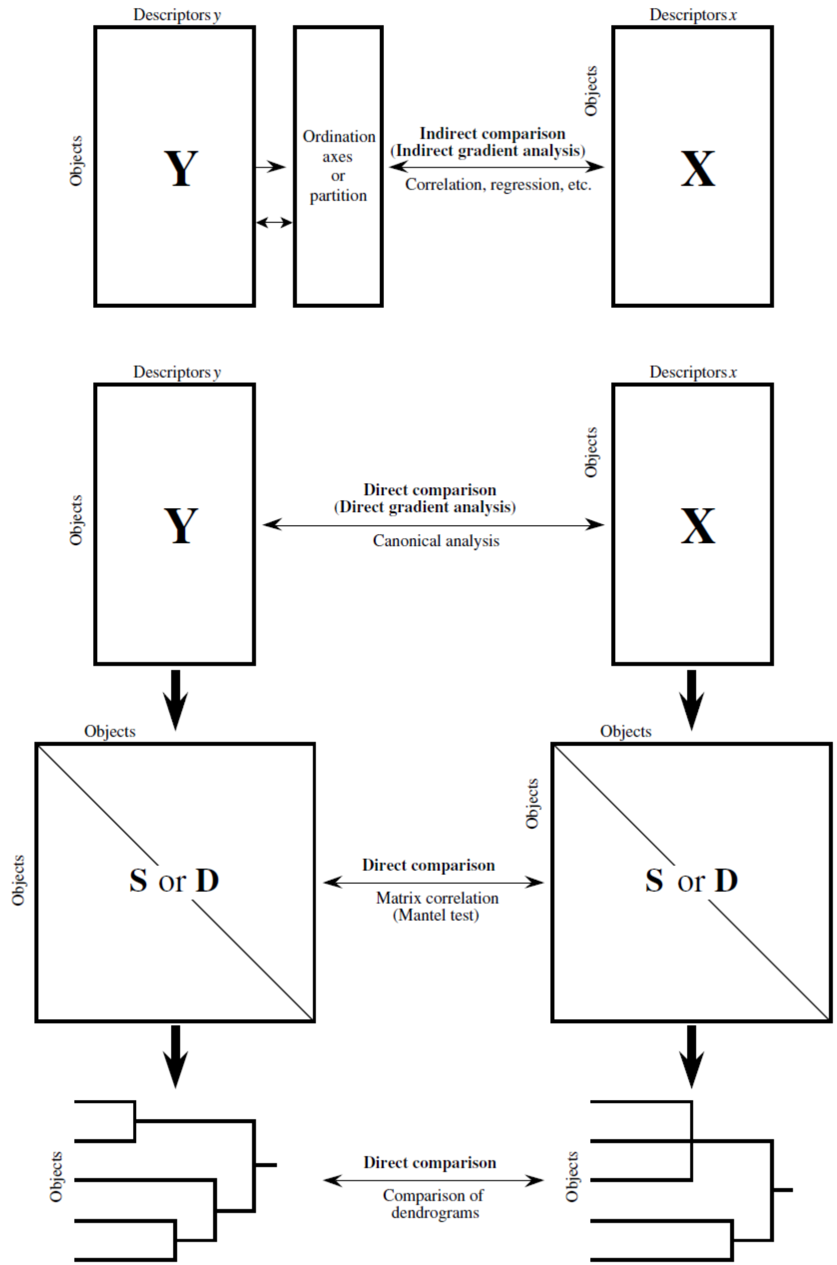


Figure 10.4 Indirect and direct comparison approaches for analysing and interpreting the structure of ecological data. Single thin arrow: inference of structure. Double arrow: interpretation strategy.

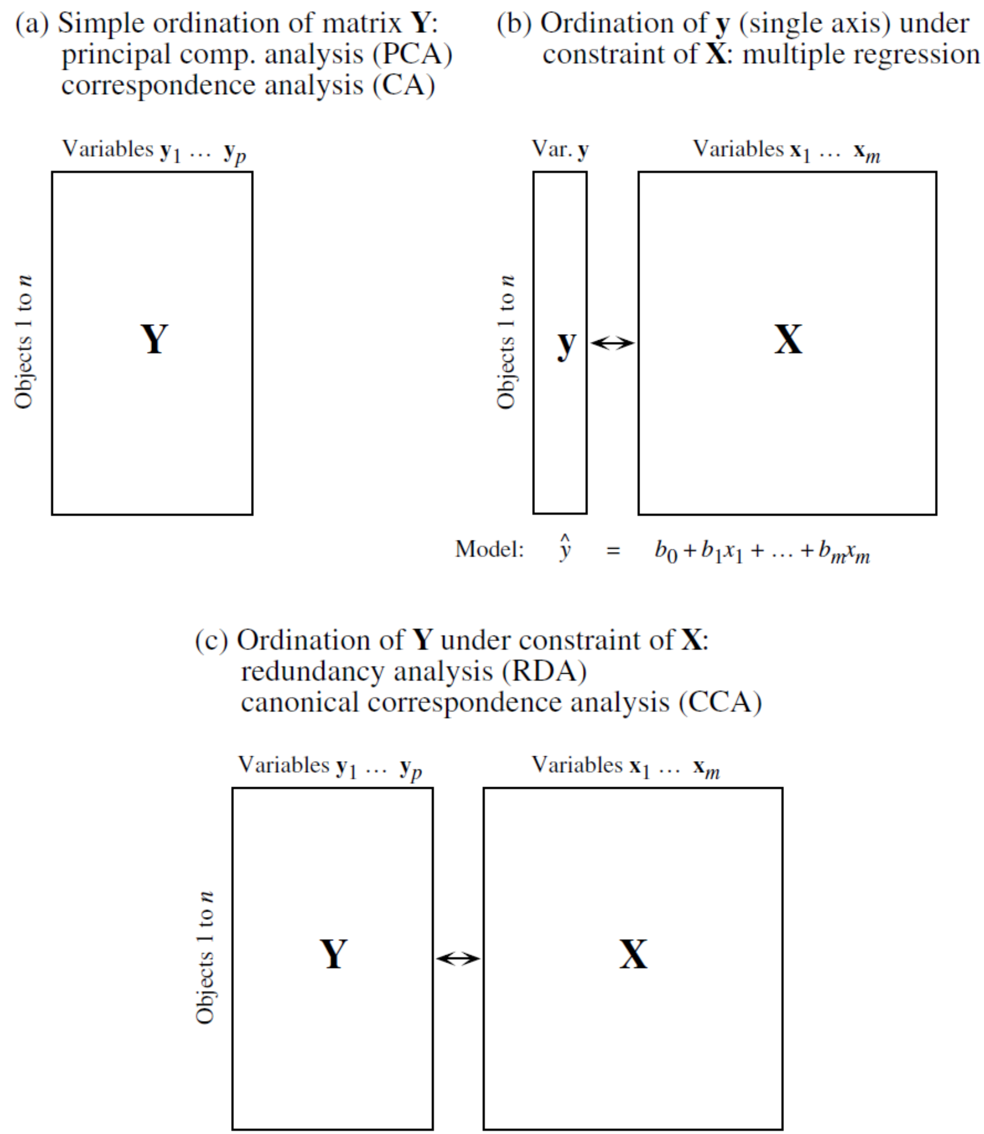


Figure 11.1 Relationships between (a) ordination, (b) regression, and (c) the asymmetric forms of canonical analysis (RDA and CCA). In (c), each canonical axis of **Y** is constrained to be a linear combination of the explanatory variables **X**.

Advantages of constrained ordinations

- Hypothesis testing!
- Proportion of variation explained by your explanatory variables
- Significance testing

- Canonical Correlation Analysis (CCorA)
  - Examines relationship between two sets of variables
- Redundancy Analysis (RDA)
  - Examines how set of dependent variables relates to set of independent variables
- Canonical Correspondence Analysis (CCA)
  - Counterpart of Canonical Correlation and Redundancy Analyses when relationship between sets of variables is Gaussian not linear

$y_1 <=> y_2$ Correlation Analysis

$x => y$ Simple Regression Analysis

$\mathbf{X} => y$ Multiple Regression Analysis

($\mathbf{X}=\{x_1,x_2,\ldots\}$)

$\mathbf{Y}_1 <=> \mathbf{Y}_2$ Canonical Correlation Analysis

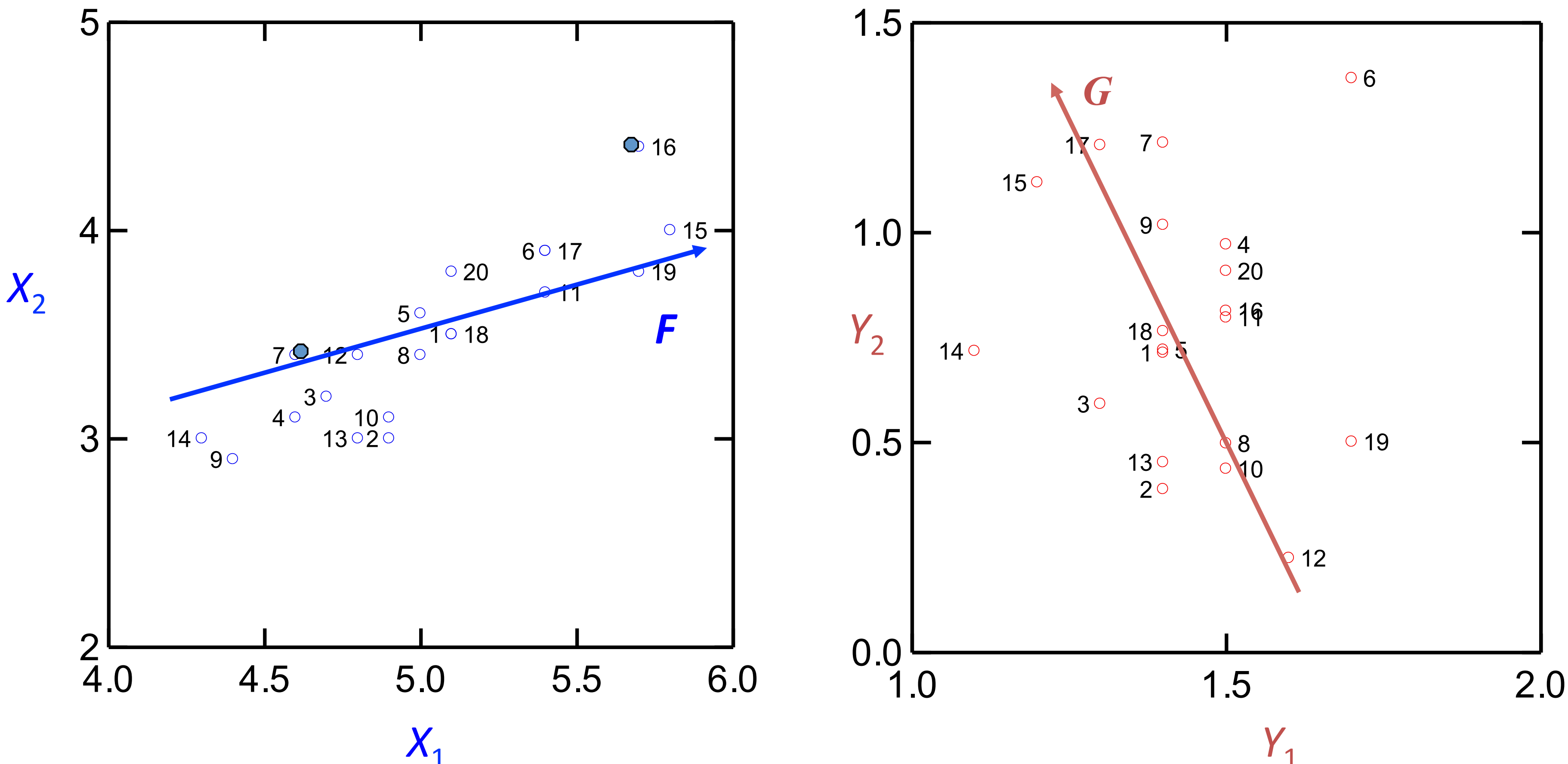
$\mathbf{X} => \mathbf{Y}$ Redundancy Analysis

$\mathbf{X} => \mathbf{Y}$ Canonical Correspondence Analysis

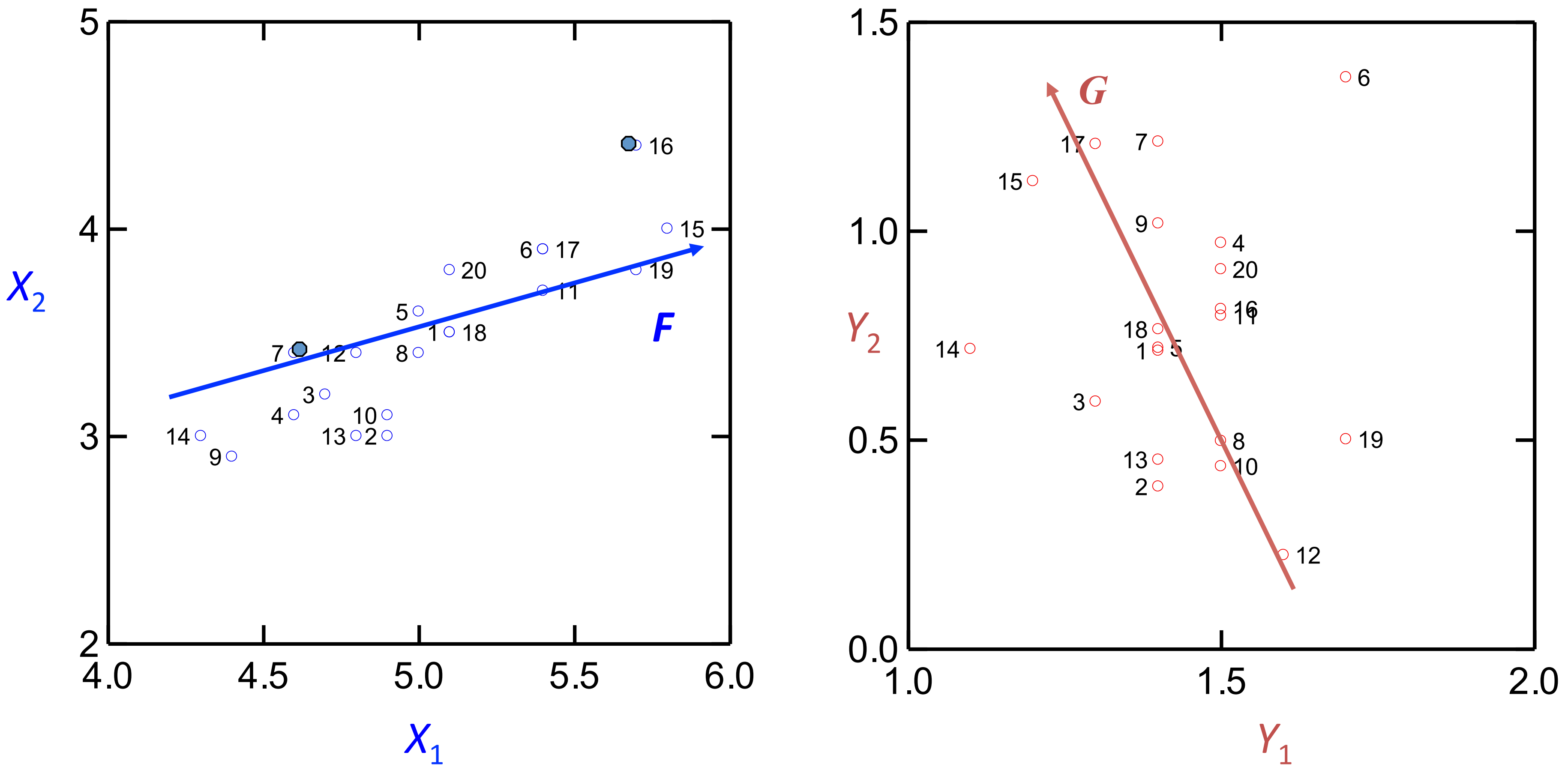
(Gaussian distribution)

Canonical Correlation Analysis (CCorA)

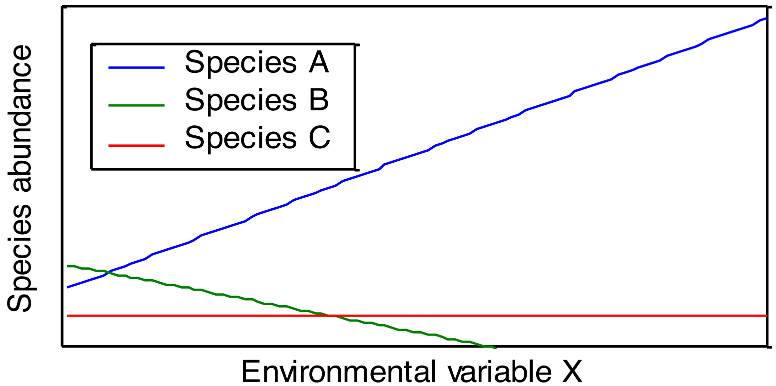
- Multivariate extension of correlation analysis
- Looks at relationship between two *sets* of variables

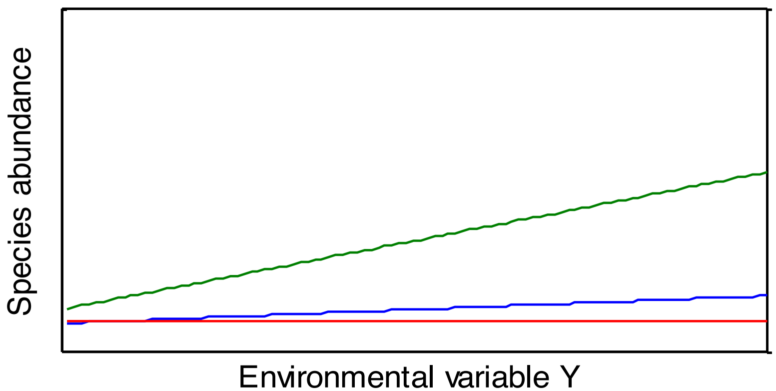


The **first canonical correlation** is:
Maximum correlation coefficient between F ($F = f_1 X_1 + f_2 X_2 + \ldots + f_p X_p$)
and G ($G = g_1 Y_1 + g_2 Y_2 + \ldots + g_q Y_q$),
for all F and G

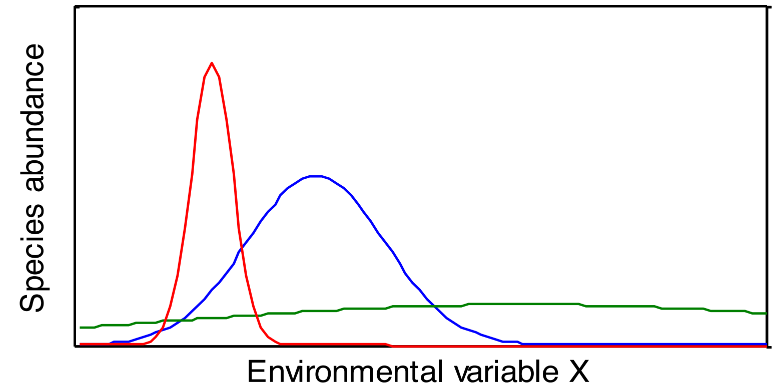


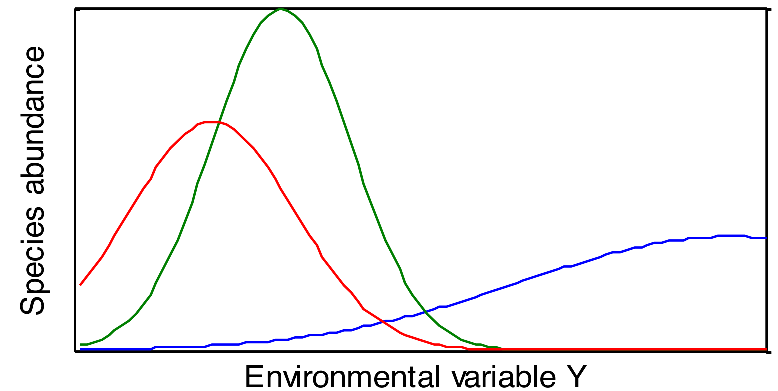
Canonical Correlation Analysis And RDA



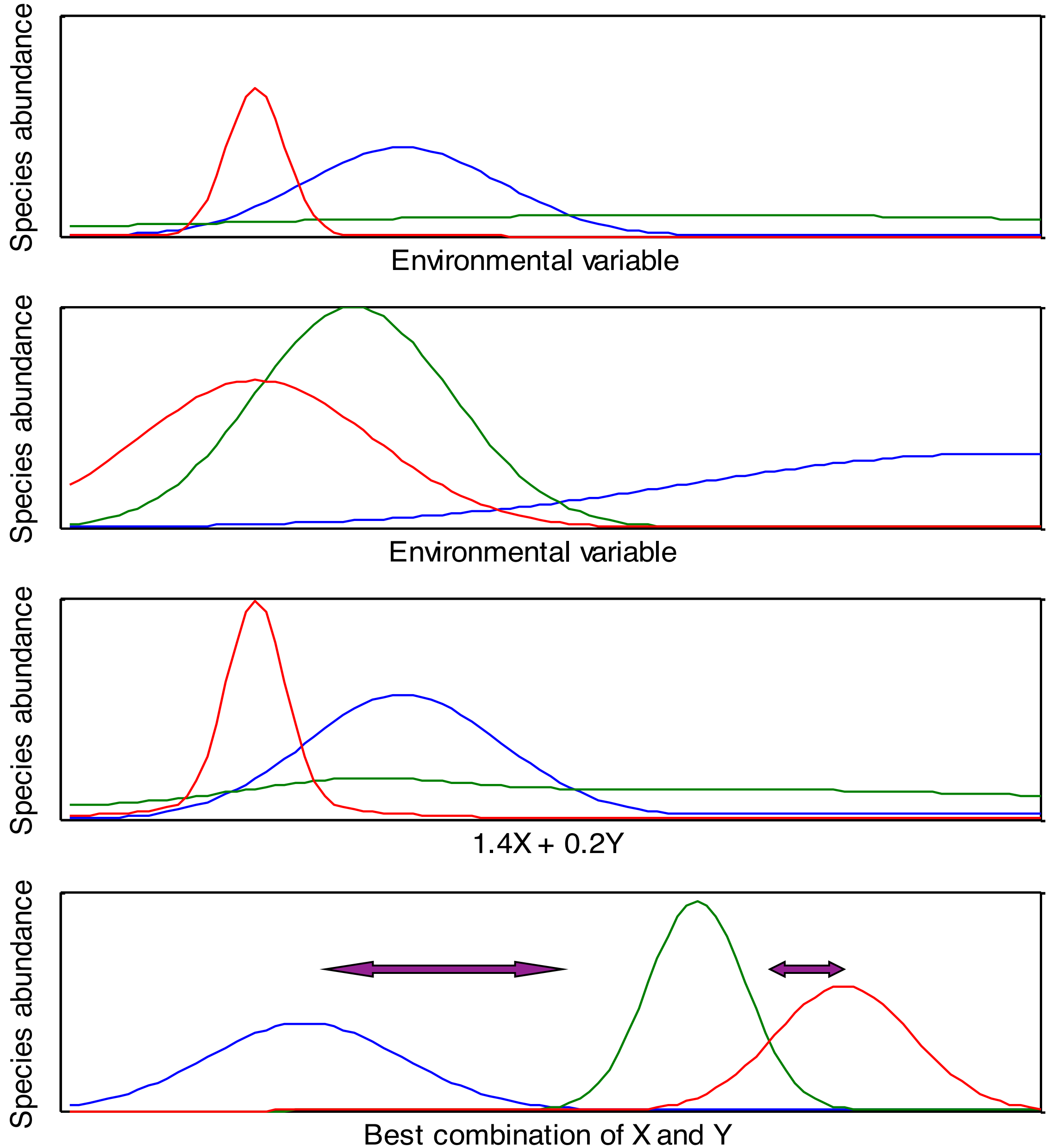


Canonical Correspondence Analysis

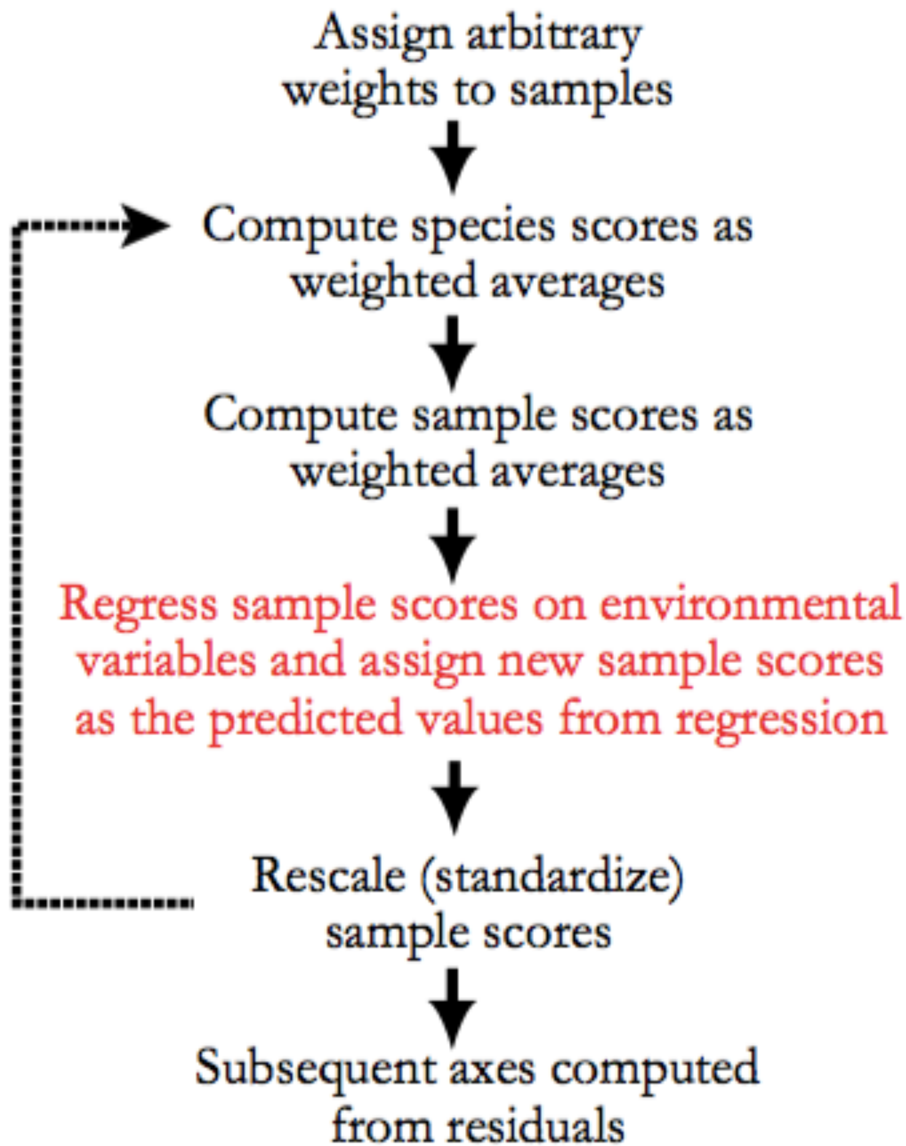




CCA


Species abundance
Environmental variable
Species abundance
Environmental variable
Species abundance
1.4X + 0.2Y
Species abundance
Best combination of X and Y

CCA Algorithm



- Effectiveness of the Constrained Ordination:

Total and constrained inertia, eigenvalues, percent variance explained, significance tests, species-environment correlation, cumulative species/sample fit.

Major Differences Between RDA and CCA

CCA

- *Chi-square distance* among samples preserved (approximates unimodal species responses).
- *Rare species* are given disproportionate weight in the analysis.
- *Unequal weighting of sites* in the regression (equal to site totals).

RDA

- *Euclidean distance* among samples preserved (appropriate with linear species respones and can approximate unimodal responses with appropriate data transformations).
- *Rare species* are not given high weight in the analysis.
- *Equal weighting of sites* in the regression.

Major Differences Between RDA and CCA

CCA

- Species-environment correlation equals the correlation between the site scores that are weighted *averages* of the species scores (WA's) and the site scores that are a linear combination of the environmental variables (LC's).
- Ordination diagram can be interpreted using the *centroid principle.*

RDA

- Species-environment correlation equals the correlation between the site scores that are weighted *sums* of the species scores (WA's) and the site scores that are a linear combination of the environmental variables (LC's).
- Ordination diagram can be interpreted using the *biplot rule.*

RDA

1) Center and standardize variables in matrices X and Y
2) Regress each variable in Y on all variables in X and compute fitted values
3) Carry out a PCA on matrix of fitted values to reduce dimensionality
4) Calculate scores

Partial canonical analysis

Use when:

- Control for effect of a covariate
- Isolate the effect of one variable
- Measure the effect of experimental factors when you have 2 or more

Discriminant Analysis

Dependent variable is categorical, metric independent variables (types of data for DV and IVs opposite of ANOVA)

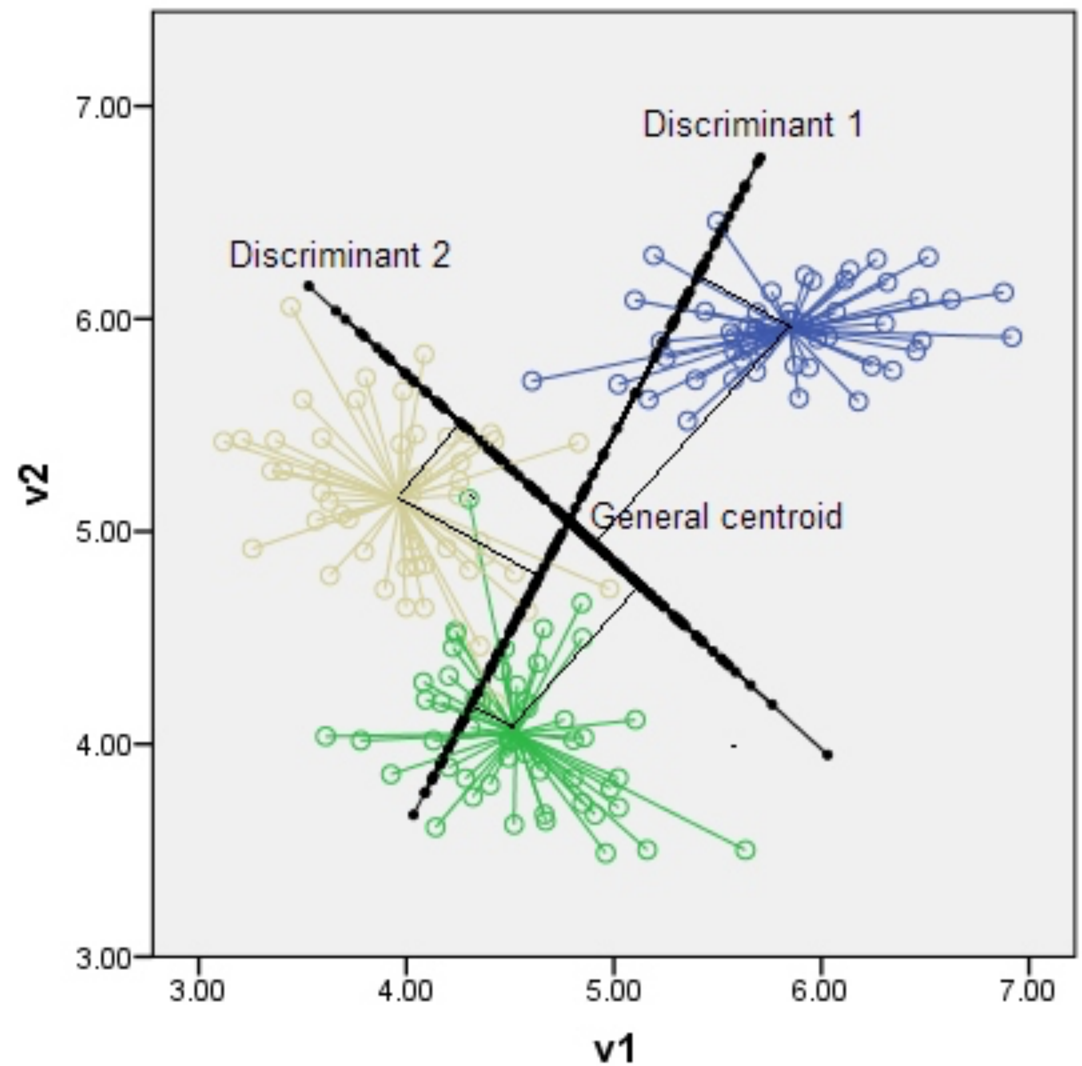
We’re essentially predicting the probability of falling into one level of the category or another => predicting group membership

It is a method of linear modelling

Discriminant Analysis

- Two steps:
  - (1) test for differences in explanatory variables (**X**) among predefined groups (like MANOVA)
  - (2) If have significant differences, find linear combinations (discriminant functions) of **X** that best discriminate between groups

Discriminant Analysis



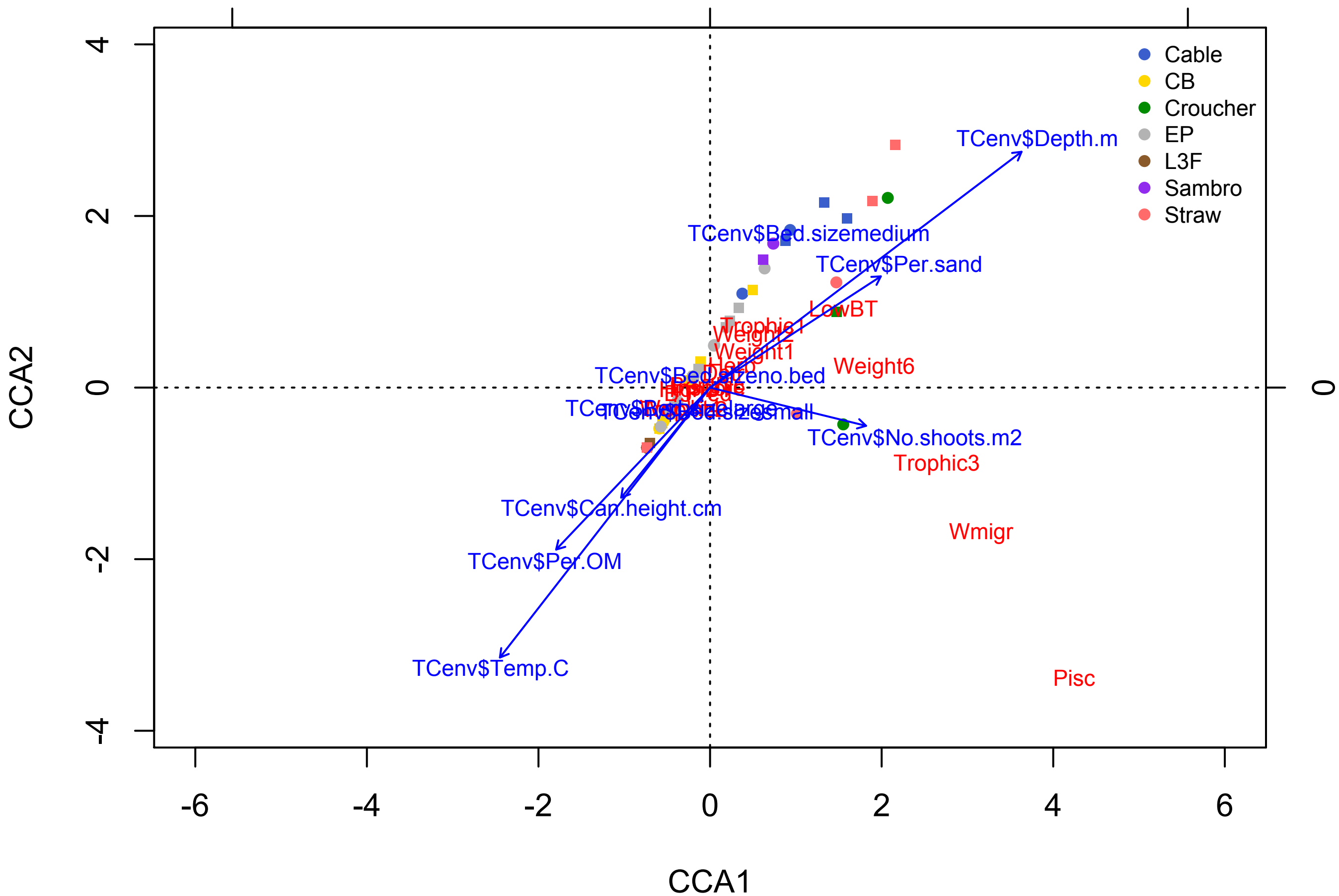
Length of an arrow => Importance of environmental variable

Direction of arrow => how well environment is correlated with the various axes

Angle between arrows => correlations between variables

Location of site scores relative to arrows => environmental characteristics of the sites

Location of species scores relative to the arrows => environmental preferences of each species



Significance testing (RDA, CCorA, CCA)

- Not possible for ordinary ordination techniques (Ch. 9)

- Permutation methods

- Significance tests performed at two levels: overall model and individual canonical eigenvalues