

# *From classical to canonical ordination*

PIERRE LEGENDRE

Département de sciences biologiques, Université de Montréal, C.P. 6128,  
succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

Short course, Université catholique de Louvain, Louvain-La-Neuve, April 18, 2012

1. Ordination objective: rotation of points while preserving distances
2. Algebra of principal component analysis (PCA) and correspondence analysis (CA)
3. Symmetrical and asymmetrical distances
4. Transformations for community composition data
5. Principal coordinate analysis (PCoA)
6. Wrap up on simple ordination for community composition data
7. Types of canonical ordination:
  - asymmetric: RDA, CCA, LDA
  - symmetric: CCorA, CoIA, MFA
8. An asymmetric canonical ordination method: redundancy analysis (RDA)
9. Eight applications of partial RDA
10. Wrap up on canonical ordination for community composition data

# Algebra of Principal Component Analysis

Data:  $\mathbf{Y} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 5 & 0 \\ 7 & 6 \\ 9 & 2 \end{bmatrix}$       Centre each column on its mean:  $\mathbf{Y}_c = [y - \bar{y}] = \begin{bmatrix} -3.2 & -1.6 \\ -2.2 & 1.4 \\ -0.2 & -2.6 \\ 1.8 & 3.4 \\ 3.8 & -0.6 \end{bmatrix}$

Covariance matrix (2 variables):  $\mathbf{S} = \frac{1}{n-1} \mathbf{Y}_c' \mathbf{Y}_c = \begin{bmatrix} 8.2 & 1.6 \\ 1.6 & 5.8 \end{bmatrix}$

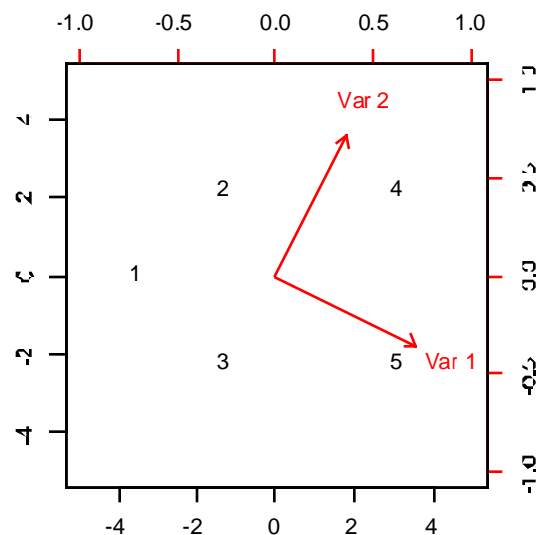
Equation for eigenvalues and eigenvectors of  $\mathbf{S}$ :  $(\mathbf{S} - \lambda_k \mathbf{I}) \mathbf{u}_k = \mathbf{0}$

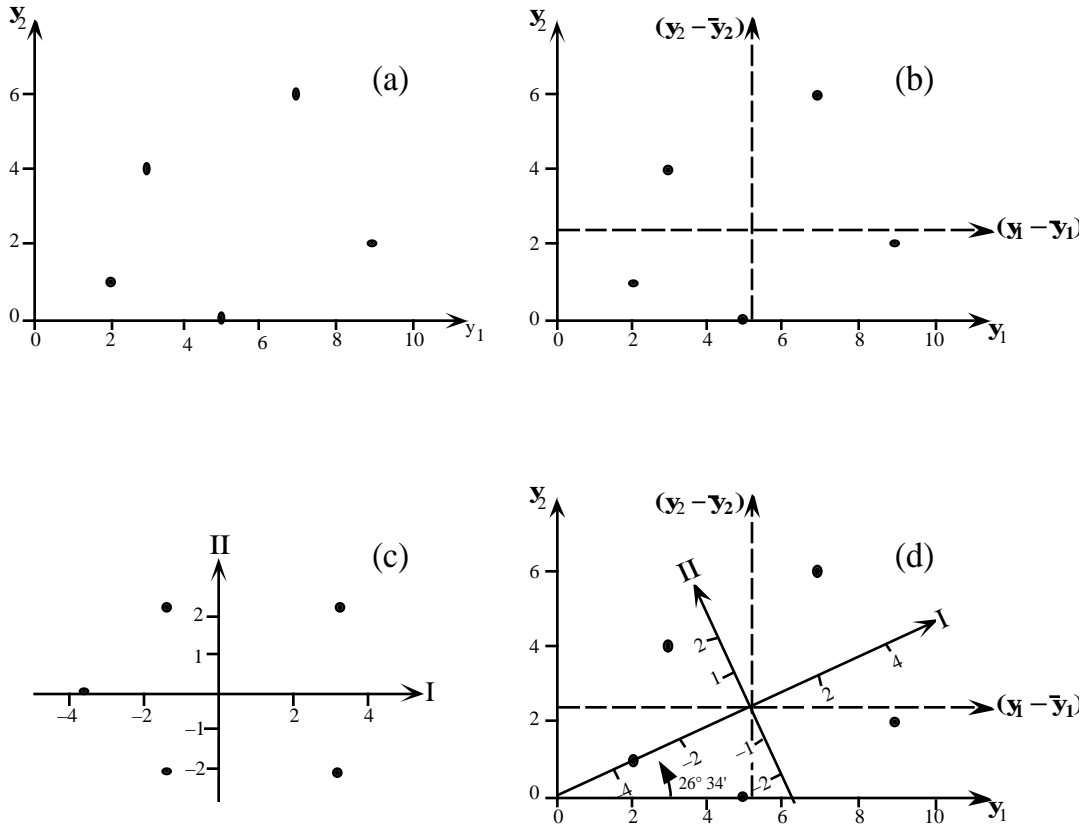
Eigenvalues:  $\lambda_1 = 9, \lambda_2 = 5$       Matrix of eigenvalues:  $\Lambda = \begin{bmatrix} 9 & 0 \\ 0 & 5 \end{bmatrix}$

Matrix of eigenvectors:  $\mathbf{U} = \begin{bmatrix} 0.8944 & -0.4472 \\ 0.4472 & 0.8944 \end{bmatrix}$

Positions of the 5 objects in ordination space:  $\mathbf{F} = [y - \bar{y}] \mathbf{U}$

$$\mathbf{F} = \begin{bmatrix} -3.2 & -1.6 \\ -2.2 & 1.4 \\ -0.2 & -2.6 \\ 1.8 & 3.4 \\ 3.8 & -0.6 \end{bmatrix} \begin{bmatrix} 0.8944 & -0.4472 \\ 0.4472 & 0.8944 \end{bmatrix} = \begin{bmatrix} -3.578 & 0 \\ -1.342 & 2.236 \\ -1.342 & -2.236 \\ 3.130 & 2.236 \\ 3.130 & -2.236 \end{bmatrix}$$





**Figure 9.2** Numerical example of principal component analysis. (a) Five objects are plotted with respect to descriptors  $y_1$  and  $y_2$ . (b) After centring the data, the objects are now plotted with respect to  $(y_1 - \bar{y}_1)$  and  $(y_2 - \bar{y}_2)$ , represented by dashed axes. (c) The objects are plotted with reference to principal axes I and II, which are centred with respect to the scatter of points. (d) The two systems of axes (b and c) can be superimposed after a rotation of  $26^\circ 34'$ .

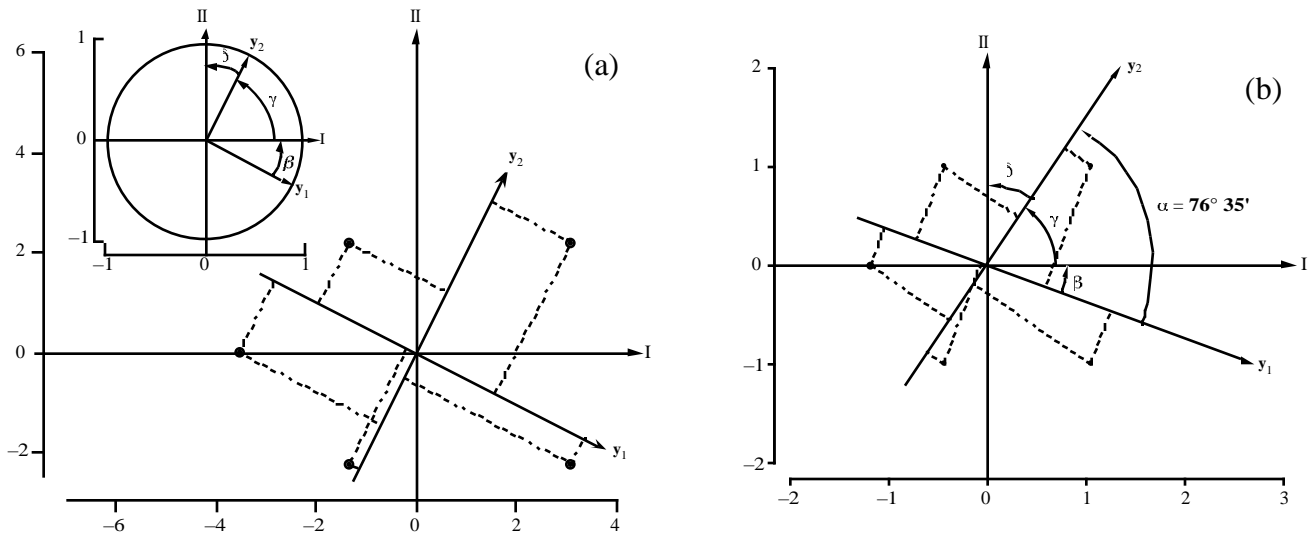


Fig. 9.3 Numerical example from Fig. 9.2. Distance and correlation biplots are discussed in Subsection 9.1.4. **(a) Distance biplot.** The eigenvectors are scaled to lengths 1. Inset: descriptors (matrix  $\mathbf{U}$ ). Main graph: descriptors (matrix  $\mathbf{U}$ ; arrows) and objects (matrix  $\mathbf{F}$ ; dots). The interpretation of the object-descriptor relationships is not based on their proximity, but on orthogonal projections (dashed lines) of the objects on the descriptor-axes or their extensions. **(b) Correlation biplot.** Descriptors (matrix  $\mathbf{U}\Lambda^{1/2}$ ; arrows) with a covariance angle of  $76^\circ 35'$ . Objects (matrix  $\mathbf{G}$ ; dots). Projecting the objects orthogonally on a descriptor (dashed lines) reconstructs the values of the objects along that descriptors, to within a multiplicative constant.

### Use the following matrices to draw biplots

*Distance biplot* (scaling 1): objects =  $\mathbf{F}$ , variables =  $\mathbf{U}$

*Correlation biplot* (scaling 2): objects =  $\mathbf{G} = \mathbf{F}\Lambda^{-1/2}$ , variables =  $\mathbf{U}_{sc2} = \mathbf{U}\Lambda^{1/2}$

These two projections respect the biplot rule, that the product of the two projected matrices reconstruct the data  $\mathbf{Y}$ :

$$\text{Distance biplot: } \mathbf{F}\mathbf{U}' = \mathbf{Y}$$

$$\text{Correlation biplot: } \mathbf{G}(\mathbf{U}\Lambda^{1/2})' = \mathbf{Y}$$

# Algebra of Correspondence Analysis

$$\text{Frequency data table } \mathbf{Y} = \begin{bmatrix} \mathbf{f}_{ij} \end{bmatrix} = \begin{bmatrix} 10 & 10 & 20 \\ 10 & 15 & 10 \\ 15 & 5 & 5 \end{bmatrix} \begin{bmatrix} \mathbf{f}_{i+} \\ 40 \\ 35 \\ 25 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{f}_{+j} \end{bmatrix} = \begin{bmatrix} 35 & 30 & 35 \end{bmatrix} \quad 100 = \mathbf{f}_{++}$$

$$\begin{aligned} p_{ij} &= \mathbf{f}_{ij} / \mathbf{f}_{++} \\ p_{i+} &= \mathbf{f}_{i+} / \mathbf{f}_{++} \\ p_{+j} &= \mathbf{f}_{+j} / \mathbf{f}_{++} \end{aligned}$$

$$\text{Matrix } \mathbf{Q} = [\bar{q}_{ij}] = \left[ \frac{p_{ij} - p_{i+}p_{+j}}{\sqrt{p_{i+}p_{+j}}} \right] = \frac{(O_{ij} - E_{ij}) / \sqrt{E_{ij}}}{\sqrt{\mathbf{f}_{++}}}$$

$$\text{Matrix } \mathbf{Q} = \begin{bmatrix} -0.10690 & -0.05774 & 0.16036 \\ -0.06429 & 0.13887 & -0.06429 \\ 0.21129 & -0.09129 & -0.12667 \end{bmatrix}$$

$$\text{Cross-product matrix: } \mathbf{Q}'\mathbf{Q} = \begin{bmatrix} 0.06020 & -0.02204 & -0.03980 \\ -0.02204 & 0.03095 & -0.00661 \\ -0.03980 & -0.00661 & 0.04592 \end{bmatrix}$$

Compute eigenvalues and eigenvectors of  $\mathbf{Q}'\mathbf{Q}$  :  $(\mathbf{Q}'\mathbf{Q} - \lambda_k \mathbf{I}) \mathbf{u}_k = \mathbf{0}$

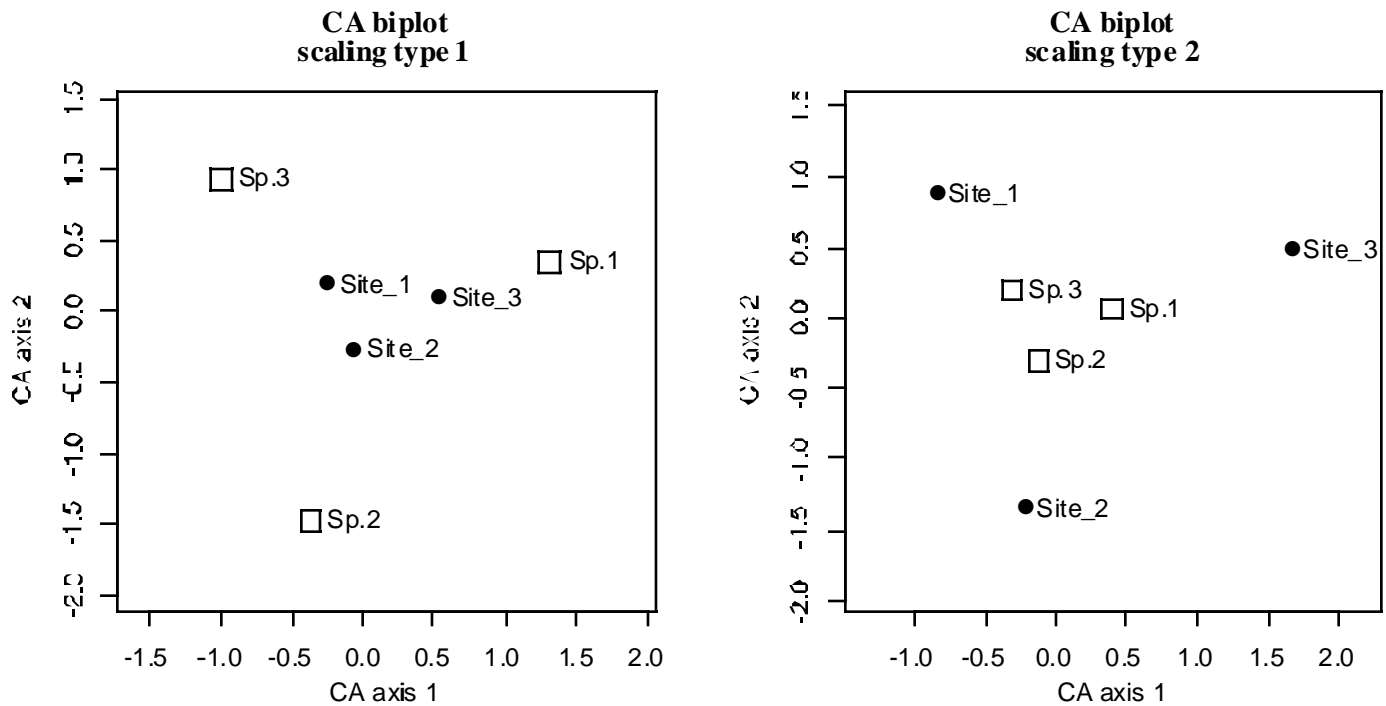
$$\text{Eigenvalues: } \lambda_1 = 0.096, \lambda_2 = 0.041 \quad \text{Matrix of eigenvalues: } \mathbf{\Lambda} = \begin{bmatrix} 0.096 & 0 \\ 0 & 0.041 \end{bmatrix}$$

*There are never more than  $k = \min(r - 1, c - 1)$  eigenvalues  $> 0$  in CA*

$$\text{Matrix of eigenvectors of } \mathbf{Q}'\mathbf{Q}_{(c \times c)} : \mathbf{U}_{(c \times k)} = \begin{bmatrix} 0.78016 & 0.20336 \\ -0.20383 & -0.81145 \\ -0.59144 & 0.54790 \end{bmatrix}$$

$$\text{Matrix of eigenvectors of } \mathbf{Q}\mathbf{Q}'_{(r \times r)} : \hat{\mathbf{U}}_{(r \times k)} = \mathbf{Q}\mathbf{U}\mathbf{\Lambda}^{-1/2} = \begin{bmatrix} -0.53693 & 0.55831 \\ -0.13043 & -0.79561 \\ 0.83349 & 0.23516 \end{bmatrix}$$

Compute matrices  $\mathbf{F}$  and  $\mathbf{V}$  for scaling 1 biplot, and  $\hat{\mathbf{V}}$  and  $\hat{\mathbf{F}}$  for scaling 2 biplot:



### Calculation details

Compute matrices  $\mathbf{V}$ ,  $\hat{\mathbf{V}}$ ,  $\mathbf{F}$ , and  $\hat{\mathbf{F}}$  used in the ordination biplots:

$$\mathbf{V}_{(c \times k)} = \mathbf{D}(p_{+j})^{-1/2} \mathbf{U} \quad \text{where } p_{+j} = f_{+j}/f_{++}$$

$$\hat{\mathbf{V}}_{(r \times k)} = \mathbf{D}(p_{i+})^{-1/2} \hat{\mathbf{U}} \quad \text{where } p_{i+} = f_{i+}/f_{++}$$

$$\mathbf{F}_{(r \times k)} = \hat{\mathbf{V}} \mathbf{\Lambda}^{1/2}$$

$$\hat{\mathbf{F}}_{(c \times k)} = \mathbf{V} \mathbf{\Lambda}^{1/2}$$

Biplot, scaling type 1: plot  $\mathbf{F}$  for sites,  $\mathbf{V}$  for species:

- This projection preserves the chi-square distance among the sites.
- The sites are at the centroids (barycentres) of the species.

Biplot, scaling type 2: plot  $\hat{\mathbf{V}}$  for sites,  $\hat{\mathbf{F}}$  for species:

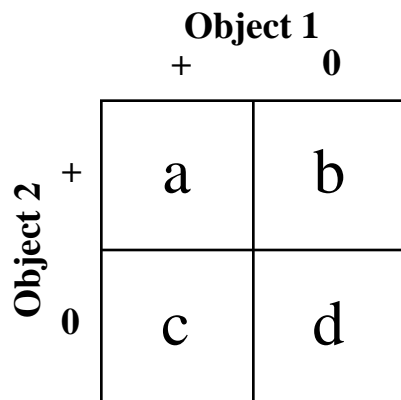
- This projection preserves the chi-square distance among the species.
- The species are at the centroids (barycentres) of the sites.

# Some measures of resemblance

References: Sneath & Sokal (1973); Legendre & Legendre (1983, 1998)

## 1) Examples of binary similarity measures

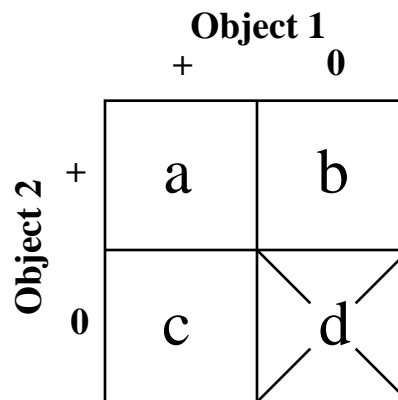
Coefficients including double-zeros  
(symmetrical coefficients)



Simple matching coefficient:

$$S(\mathbf{x}_1, \mathbf{x}_2) = \frac{a + d}{a + b + c + d}$$

Coefficients excluding double-zeros  
(asymmetrical coefficients)



Jaccard's coefficient of community:

$$S(\mathbf{x}_1, \mathbf{x}_2) = \frac{a}{a + b + c}$$

Example:

**Binary variables** ( $n = 7$ )

Object $\mathbf{x}_1$	1	0	0	1	1	0	1
Object $\mathbf{x}_2$	0	0	1	1	1	0	0
Agreement		1		1	1	1	Sum = 4
Positive agreement				1	1		Sum = 2
Presence	1		1	1	1		Sum = 5

$$S(\mathbf{x}_1, \mathbf{x}_2) = 4/7 = 0.57$$

$$S(\mathbf{x}_1, \mathbf{x}_2) = 2/5 = 0.40$$

## 2) Examples of quantitative measures

Distance not designed for shared species

No upper limit

Euclidean distance (preserved in principal component analysis):

$$D(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^p (y_{1j} - y_{2j})^2}$$

Distance designed for shared species

Bounded in the interval [0, 1]

Bray-Curtis distance:

$$D(\mathbf{x}_1, \mathbf{x}_2) = 1 - \frac{2W}{A+B}$$

Example:

**Species abundances**

A

B

W

Object  $\mathbf{x}_1$

7 3 4 5 1

20

Object  $\mathbf{x}_2$

2 4 7 6 3

22

Minima

2 3 4 5 1

15

Differences

5 1 3 1 2

$$D(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{5^2 + 1^2 + 3^2 + 1^2 + 2^2} = 6.325 \quad D(\mathbf{x}_1, \mathbf{x}_2) = 1 - \frac{2 \times 15}{20 + 22} = 0.286$$

Distance designed for shared species

No upper limit

Chi-square ( $\chi^2$ ) distance (preserved in correspondence analysis):

$$D(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^p \frac{1}{y_{+j}/y_{++}} \left( \frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right)^2}$$

Example:

$\begin{bmatrix} y_{i+} \end{bmatrix}$

$$\mathbf{Y} = \begin{bmatrix} 7 & 3 & 4 & 5 & 1 \\ 2 & 4 & 7 & 6 & 3 \\ 12 & 8 & 5 & 14 & 6 \end{bmatrix} \begin{bmatrix} 20 \\ 22 \\ 45 \end{bmatrix}$$

$$\begin{bmatrix} \frac{y_{ij}}{y_{i+}} \end{bmatrix} = \begin{bmatrix} 0.350 & 0.150 & 0.200 & 0.250 & 0.050 \\ 0.091 & 0.182 & 0.318 & 0.273 & 0.136 \\ 0.267 & 0.178 & 0.111 & 0.311 & 0.133 \end{bmatrix}$$

$$\begin{bmatrix} y_{+j} \end{bmatrix} = \begin{bmatrix} 21 & 15 & 16 & 25 & 10 \end{bmatrix} \quad 87$$

$$D(\mathbf{x}_1, \mathbf{x}_2) = 0.653$$



## Data transformations

### Transform physical variables (*Ecology*) or characters (*Taxonomy*)

- Univariate distributions are not symmetrical  
⇒ Apply skewness-reduction transformation
- Variables are not in the same physical units

⇒ Apply standardization  $z_i = \frac{y_i - \bar{y}}{s_y}$  or ranging  $y'_i = \frac{y_i - y_{min}}{y_{max} - y_{min}}$

### Transform community composition data (*Ecology*)

(species presence-absence or abundance)

- Reduce asymmetry of distributions  
⇒ Apply  $\log(y + c)$  transformation
- Make community composition data suitable for Euclidean-based ordination methods (PCA, RDA)  
⇒ Use the chord, chi-square, profile, or Hellinger transformations (Legendre & Gallagher 2001)

Species abundance paradox data  
(3 sites, 3 species)

	Species 1	Species 2	Species 3
Site 1	0	1	1
Site 2	1	0	0
Site 3	0	4	8

$$D_{\text{Euclidean}}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^p (y_{1j} - y_{2j})^2}$$

$$D_{\text{chord}}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^p \left( \frac{y_{1j}}{\sqrt{\sum_{j=1}^p y_{1j}^2}} - \frac{y_{2j}}{\sqrt{\sum_{j=1}^p y_{2j}^2}} \right)^2}$$

$$D_{2_{\text{metric}}}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^p \frac{1}{y_{+j}} \left( \frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right)^2}$$

$$D_{2_{\text{distance}}}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^p \frac{1}{y_{+j} \sqrt{y_{++}}} \left( \frac{y_{1j}}{y_{1+}} - \frac{y_{2j}}{y_{2+}} \right)^2}$$

$$D_{\text{species profiles}}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^p \frac{y_{1j} - y_{2j}}{y_{1+} y_{2+}}^2}$$

$$D_{\text{Hellinger}}(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{j=1}^p \left[ \sqrt{\frac{y_{1j}}{y_{1+}}} - \sqrt{\frac{y_{2j}}{y_{2+}}} \right]^2}$$

Transformations

$$y'_{ij} = \frac{y_{ij}}{\sqrt{\sum_{j=1}^p y_{ij}^2}}$$

$$y'_{ij} = \frac{y_{ij}}{y_{i+} \sqrt{y_{+j}}}$$

$$y'_{ij} = \sqrt{y_{++}} \frac{y_{ij}}{y_{i+} \sqrt{y_{+j}}}$$

$$y'_{ij} = \frac{y_{ij}}{y_{i+}}$$

$$y'_{ij} = \sqrt{\frac{y_{ij}}{y_{i+}}}$$

$$\mathbf{D} = \begin{bmatrix} 0.0000 & 1.7321 & 7.6158 \\ 1.7321 & 0.0000 & 9.0000 \\ 7.6158 & 9.0000 & 0.0000 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 0.0000 & 1.4142 & 0.3204 \\ 1.4142 & 0.0000 & 1.4142 \\ 0.3204 & 1.4142 & 0.0000 \end{bmatrix}$$

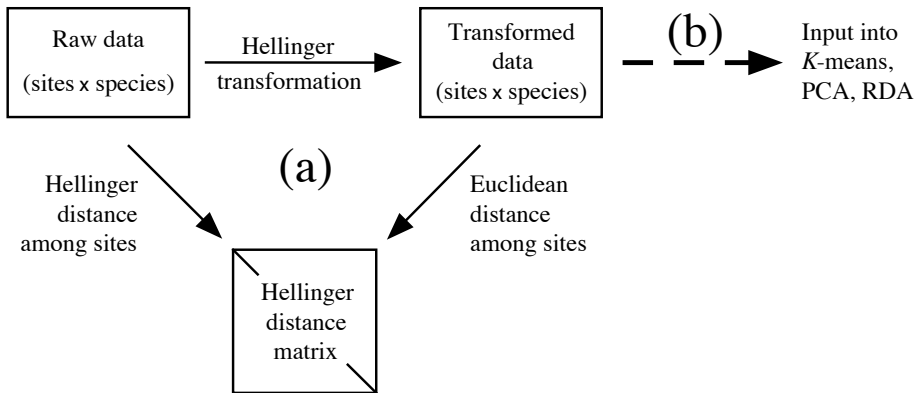
$$\mathbf{D} = \begin{bmatrix} 0.0000 & 1.0382 & 0.0930 \\ 1.0382 & 0.0000 & 1.0352 \\ 0.0930 & 1.0352 & 0.0000 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 0.0000 & 4.0208 & 0.3600 \\ 4.0208 & 0.0000 & 4.0092 \\ 0.3600 & 4.0092 & 0.0000 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 0.0000 & 1.2247 & 0.2357 \\ 1.2247 & 0.0000 & 1.2472 \\ 0.2357 & 1.2472 & 0.0000 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 0.0000 & 1.4142 & 0.1697 \\ 1.4142 & 0.0000 & 1.4142 \\ 0.1697 & 1.4142 & 0.0000 \end{bmatrix}$$

Fig. 2 (alternative). Species abundance paradox data, modified from Orłóci (1978). The paradox is that the Euclidean distance between sites 1 and 2, which have no species in common, is smaller than that between sites 1 and 3 which share species 2 and 3; this example shows that the Euclidean distance is not appropriate for species abundance data. With the other coefficients in the Table, the distance between sites 1 and 2 is larger than that between sites 1 and 3; furthermore, the distance between sites 1 and 2 is the same as between sites 2 and 3, or very nearly so.



**Figure 7.7** (a) Calculation of a distance matrix either directly from the raw data (left diagonal arrow) or through a two-step approach in which the raw data are transformed (horizontal arrow) before computation of the distance matrix (right diagonal arrow). The example shown here uses the Hellinger transformation to obtain the Hellinger distance matrix ( $D_{17}$ ). The same approach can be used to obtain the chord ( $D_3$ ), species profile ( $D_{18}$ ), chi-square metric ( $D_{15}$ ) and chi-square distance ( $D_{16}$ ) matrices, as summarized in Fig. 7.8. (b) The transformed species data can also be used as input (dashed arrow) into linear methods of analysis, in particular PCA, RDA, and  $K$ -means partitioning. Modified from Legendre & Gallagher (2001).

## Principal coordinate analysis (PCoA)

Example: a Euclidean distance matrix computed from the data of the PCA example.

$$\mathbf{D} = \begin{bmatrix} 0.00000 & 3.16228 & 3.16228 & 7.07107 & 7.07107 \\ 3.16228 & 0.00000 & 4.47214 & 4.47214 & 6.32456 \\ 3.16228 & 4.47214 & 0.00000 & 6.32456 & 4.47214 \\ 7.07107 & 4.47214 & 6.32456 & 0.00000 & 4.47214 \\ 7.07107 & 6.32456 & 4.47214 & 4.47214 & 0.00000 \end{bmatrix}$$

Transform  $\mathbf{D}$  to a new matrix  $\mathbf{A} = [a_{hi}]$ :  $a_{hi} = -0.5D_{hi}^2$

Centre matrix  $\mathbf{A}$  to produce matrix  $\mathbf{\Delta}$  with sums of the rows and columns equal to 0:

$$\mathbf{\Delta} = [\delta_{hi}] = [a_{hi} - \bar{a}_h - \bar{a}_i + \bar{a}]$$

$$\mathbf{\Delta} = \begin{bmatrix} 12.8 & 4.8 & 4.8 & -11.2 & -11.2 \\ 4.8 & 6.8 & -3.2 & 0.8 & -9.2 \\ 4.8 & -3.2 & 6.8 & -9.2 & 0.8 \\ -11.2 & 0.8 & -9.2 & 14.8 & 4.8 \\ -11.2 & -9.2 & 0.8 & 4.8 & 14.8 \end{bmatrix}$$

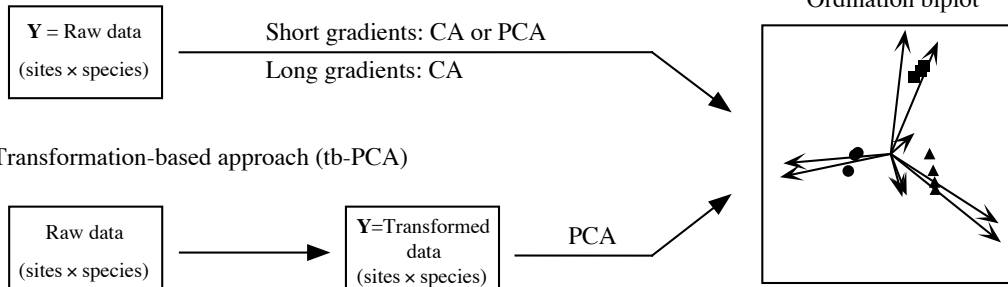
Compute the eigenvalues and eigenvectors of matrix  $\mathbf{\Delta}$ . Scale the eigenvectors to lengths equal to the square roots of their respective eigenvalues,  $\sqrt{\mathbf{u}'_k \mathbf{u}_k} = \sqrt{\lambda_k}$ .

Eigenvalues:	$\lambda_1 = 36$	$\lambda_2 = 20$
Objects	Eigenvectors	
$\mathbf{x}_1$	-3.578	0.000
$\mathbf{x}_1$	-1.342	-2.236
$\mathbf{x}_1$	-1.342	2.236
$\mathbf{x}_1$	3.130	-2.236
$\mathbf{x}_1$	3.130	2.236
Eigenvector length	$\sqrt{36} = 6.000$	$\sqrt{20} = 4.472$
PCoA eigenvalues / $(n-1)$	9	5

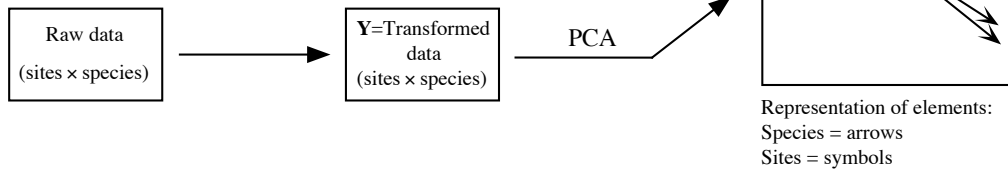
Compare the PCoA eigenvalues and eigenvectors to the PCA eigenvalues and matrix  $\mathbf{F}$ .

PCoA is used for ordination of  $\mathbf{D}$  matrices produced by functions other than the Euclidean.

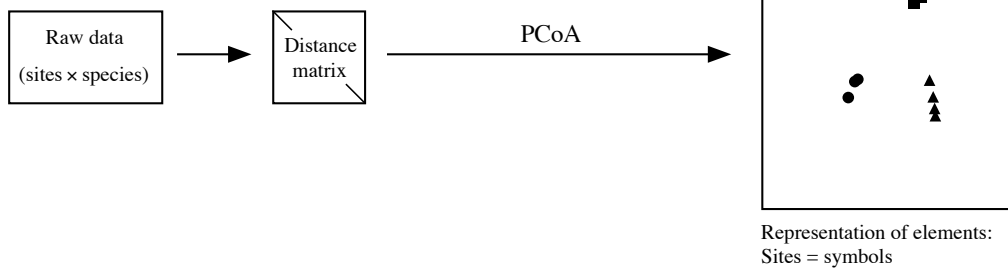
## (a) Classical approach



## (b) Transformation-based approach (tb-PCA)



## (c) Distance-based approach (PCoA)



**Figure 9.8** Different approaches are available for metric ordination of community composition data: (a) classical PCA and CA, (b) the transformation-based approach, and (c) the distance-based approach (PCoA). Metric ordination methods produce ordinations that fully preserve the distances among sites, as specified in Table 9.1. Modified from Legendre & Gallagher (2001).

## Canonical ordination methods

produce linear ordinations based upon two [or more] data matrices

---

### *Asymmetric methods*

$Y$  = response data  
 $X$  = explanatory data

### *Symmetric methods*

$X_1$  and  $X_2$  play the same role in the analysis

---

#### Redundancy analysis (RDA)

$Y$  = any multivariate data table, as in PCA. Preserves Euclidean dist.

#### Canonical Correspondence Analysis (CCA)

$Y$  = frequency data, as in CA. Preserves chi-square distances

#### Linear Discriminant Analysis (LDA)

$Y$  = a factor (classification of the objects). Preserves Mahalanobis distances

#### Canonical Correlation Analysis (CCorA)

Maximizes square correlations among data sets. Preserves Mahalanobis distances

#### Co-inertia Analysis (CoIA) Procrustes (Proc)

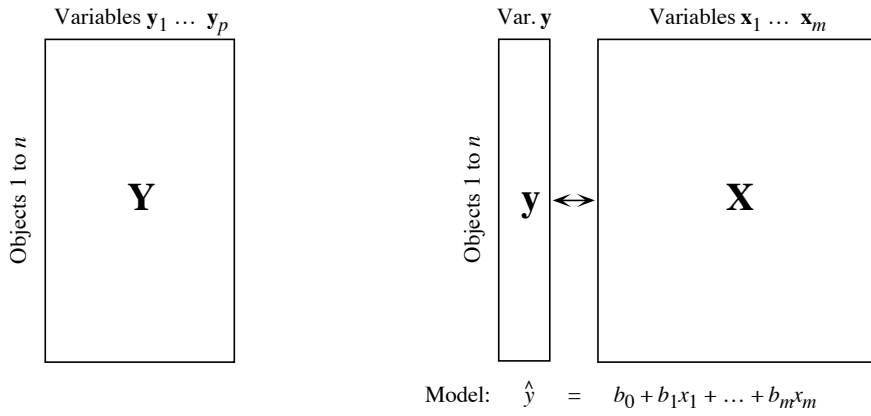
Maximizes square covariances among data sets. Preserves Euclidean distances

#### Multiple Factor Analysis (MFA)

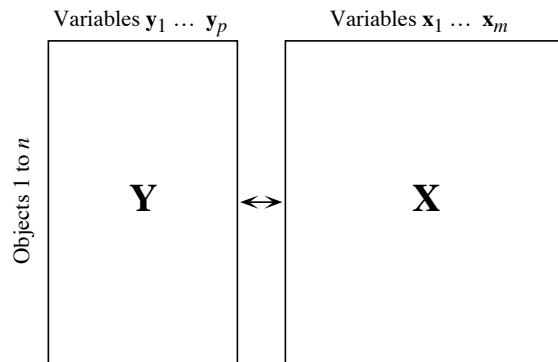
Generalization of PCA to more than two data matrices

---

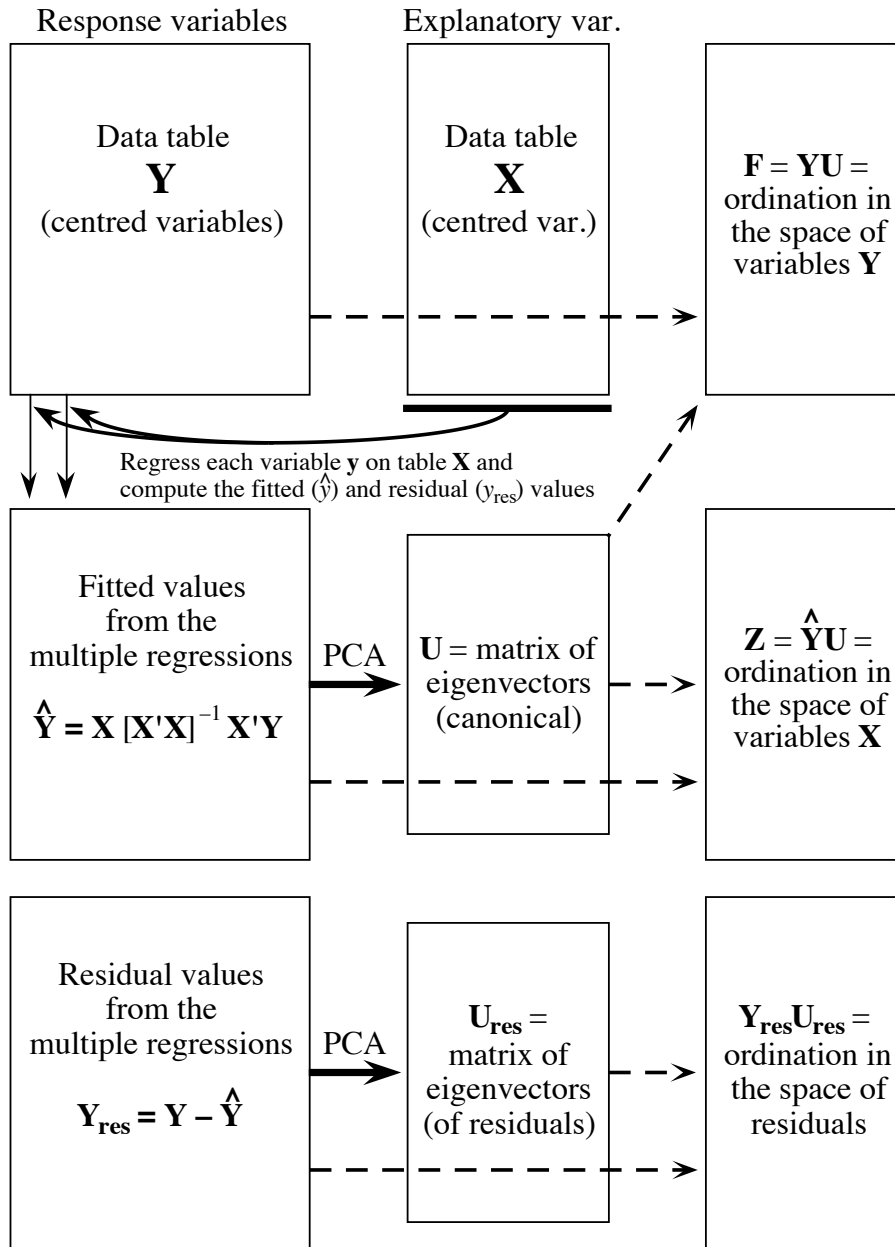
- (a) Simple ordination of matrix  $\mathbf{Y}$ :  
principal comp. analysis (PCA)  
correspondence analysis (CA)
- (b) Ordination of  $\mathbf{y}$  (single axis) under  
constraint of  $\mathbf{X}$ : multiple regression



- (c) Ordination of  $\mathbf{Y}$  under constraint of  $\mathbf{X}$ :  
redundancy analysis (RDA)  
canonical correspondence analysis (CCA)

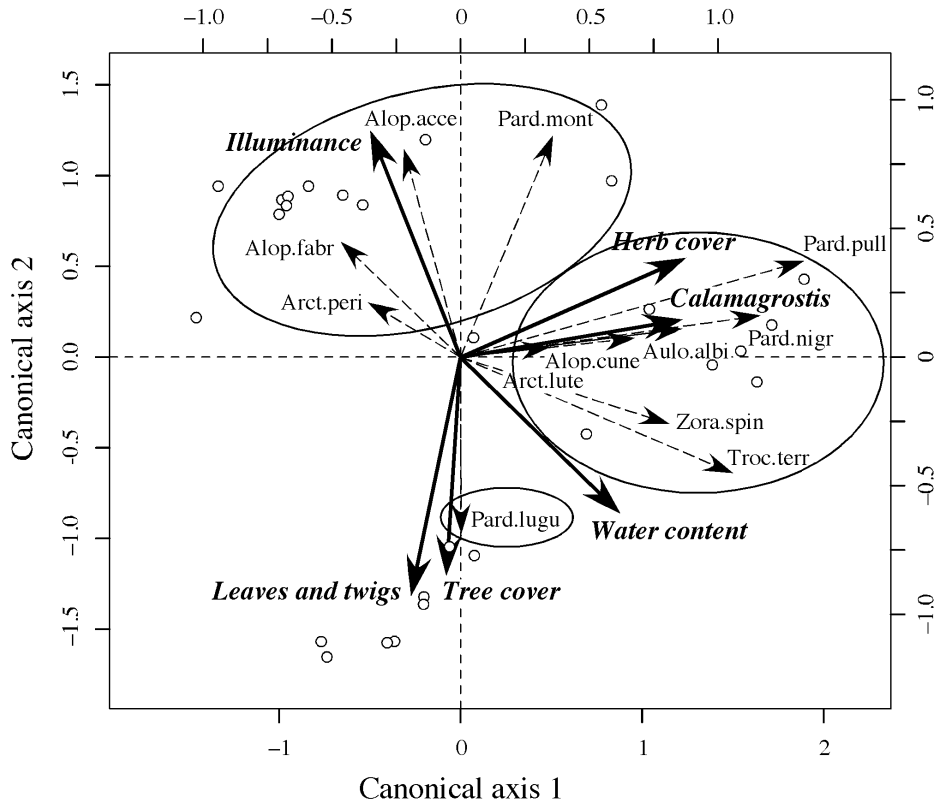


**Figure 11.1** Relationships between (a) ordination, (b) regression, and (c) two asymmetric forms of canonical analysis (RDA and CCA). In (c), each canonical axis of  $\mathbf{Y}$  is constrained to be a linear combination of the explanatory variables  $\mathbf{X}$ .



**Figure 11.2** Redundancy analysis may be understood as a two-step process: (1) regress each variable in  $\mathbf{Y}$  on all variables in  $\mathbf{X}$  and compute the fitted values; (2) carry out a PCA of the matrix of fitted values to obtain the eigenvalues and eigenvectors. Two ordinations are obtained, one ( $\mathbf{F} = \mathbf{Y}\mathbf{U}$ ) in the space of the response variables  $\mathbf{Y}$ , the other ( $\mathbf{Z} = \hat{\mathbf{Y}}\mathbf{U}$ ) in the space of the explanatory variables  $\mathbf{X}$ . Another PCA ordination can be computed for the matrix of residuals.



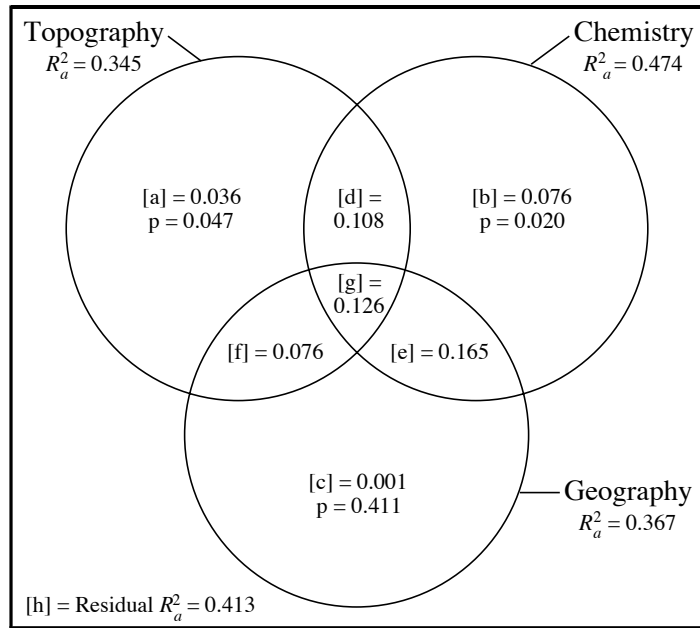


**Figure 11.7** RDA triplot relating the spider species (dashed arrows) to the selected environmental variables (full-line arrows). Scaling type 2 was used to emphasize the covariances among the species. Small open circles represent the 28 sites; site names were not printed to keep the diagram simple. The species associations are indicated by ellipses. Association 1: *Alopecosa accentuata* (abbreviation: Alop.acce), *Alopecosa fabrilis* (Alop.fabr), *Arctosa perita* (Arct.peri) and *Pardosa monticola* (Pard.mont, weakly associated with this group). Association 2: *Alopecosa cuneata* (Alop.cune), *Arctosa lutetiana* (Arct.lute), *Aulonia albimana* (Aulo.albi), *Pardosa nigriceps* (Pard.nigr), *Pardosa pullata* (Pard.pull), *Trochosa terricola* (Troc.terr) and *Zora spinimana* (Zora.spin). Single-species group: *Pardosa lugubris* (Pard.lugu).

## Some applications of partial redundancy analysis

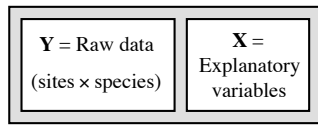
partial RDA:  $\text{rda}(\mathbf{Y}, \mathbf{X}, \mathbf{W})$

1. *Control for well-known linear effects.* —  $\mathbf{W}$  contains covariables whose effects on  $\mathbf{Y}$  are well understood.
2. *Isolate the effect of a single explanatory variable or factor.* —  $\mathbf{X}$  contains a single explanatory variable,  $\mathbf{W}$  contains all the other explanatory variables.
3. *Analysis of related samples.* —  $\mathbf{W}$  contains a matrix representing the pairing of sampling units.
4. *MANOVA by RDA.* — Partial canonical analysis may be used, instead of MANOVA, to analyse a multivariate response data matrix  $\mathbf{Y}$  in cross-factor experimental designs, including tests of significance for the main effects and the interaction terms.
5. *Principal response curves (PRC).* — *Principal response curves* is a special case of RDA for the analysis of the results of experiments conducted over time, that involved multivariate response data (e.g. community composition data).
6. *Partial PCA.* — Partial PCA is the PCA of a response data table  $\mathbf{Y}$  residualized on a set of explanatory variables.
7. *Selection of explanatory variables.* — Different selection methods are available in canonical analysis, just as in multiple regression: backward, forward, and stepwise.
8. *Multivariate variation partitioning.* — Partition the variation of multivariate response data (e.g. community composition data) with respect to two or more sets of explanatory variables.



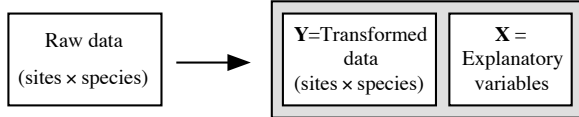
**Figure 11.6** Venn diagram illustrating the results of variation partitioning of the Doubs River fish assemblage data (29 sites) among three sets of explanatory variables: Topography, Chemistry and Geography. The fractions of variation are identified by letters [a] to [h]. The value next to each identifier is the adjusted  $R^2$  ( $R_a^2$ ). The circles, drawn by the plotting function *plot.varpart()*, are of equal sizes despite differences in the corresponding  $R_a^2$ . Circle sizes and shapes can be modified using a graphics editor prior to the publication of the partitioning results.

(a) Classical approach: RDA preserves the Euclidean distance, CCA preserves the chi-square distance

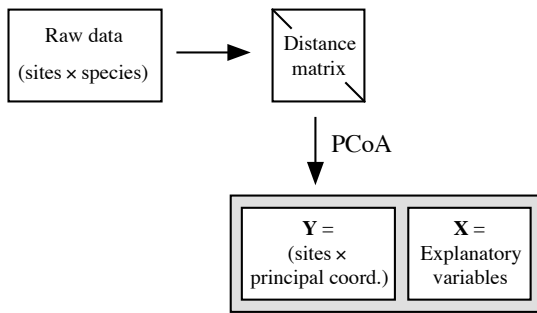


Short gradients: CCA or RDA  
Long gradients: CCA

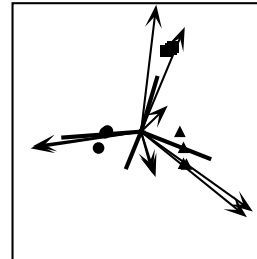
(b) Transformation-based RDA (tb-RDA) approach:  
preserves a distance obtained by data transformation



(c) Distance-based RDA (db-RDA) approach:  
preserves a pre-computed distance



Canonical  
ordination triplot



Representation of elements:  
Species = arrows  
Sites = symbols  
Explanatory variables = lines

**Figure 11.4** Comparison of (a) classical RDA and CCA, and (b and c) alternative approaches forcing RDA to preserve other distances adapted to community composition data. Modified from Legendre & Gallagher (2001).

# Numerical Ecology with R

*A practical guide to Numerical Ecology using the R language*

Links to the authors' web pages

[Daniel Borcard](#)

[François Gillet](#)

[Pierre Legendre](#)

This page provides the data sets, R scripts, R functions and several useful links related to the book entitled "Numerical Ecology with R". The R code is provided for PC and Macintosh.

## For first-time users of the book and when upgrading R

[Script to install the necessary packages](#)

### Data sets

[Doubs fish data and oribatid mite data](#)

### R scripts for all chapters

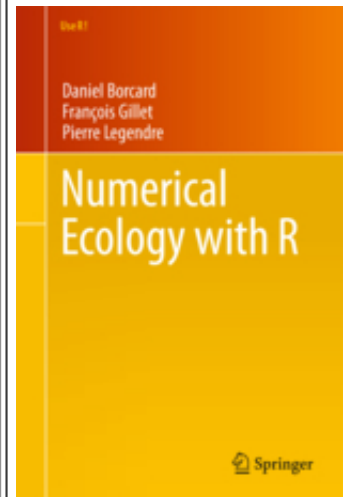
[R scripts](#) for PC and Macintosh computers

### R functions

[Functions](#) to be sourced when necessary

### R package

MVPARTwrap, a wrapper for package mvpart is now available on CRAN for OS X, Windows and Linux. To install it, proceed as for any CRAN package.



[Springer web page of the book](#)