Chapter 11 *Canonical analysis*

11.0 Principles of canonical analysis

Canonical analysis is the simultaneous analysis of two, or possibly several data tables. Canonical analyses allow ecologists to perform *direct comparisons* of two data matrices (also called "direct gradient analysis"; Fig. 10.4, Table 10.1). Typically, one may be interested in the relationship between a first table describing species composition and a second table containing environmental descriptors, observed *at the same locations*; or two tables of environmental descriptors, e.g. a table about the chemistry of lakes and another about drainage basin geomorphology.

Indirect comparison

Direct comparison

In *indirect comparison* (also called "indirect gradient analysis"; Fig. 10.4), the matrix of explanatory variables **X** does not intervene in the calculation producing the ordination of **Y**. Correlation or regression of the ordination vectors on **X** are computed *a posteriori*. In *direct comparison analysis* (canonical analysis) on the contrary, matrix **X** intervenes in the calculation, forcing the ordination vectors to be maximally related to combinations of the variables in **X**. This description applies to all forms of canonical analysis and in particular to the asymmetric forms described in Sections 11.1 to 11.3.

There is a parallel in cluster analysis, when clustering results are constrained to be consistent with explanatory variables in multivariate regression trees (MRT, Section 8.11) or with structural relationships among observations, either temporal (Subsection 12.6.4) or spatial (Subsection 13.3.2), which are inherent to the sampling design. In constrained clustering or canonical ordination, the results differ in most instances from those of unconstrained analysis and are, hopefully, more readily interpretable. Furthermore, direct comparison analysis allows one to directly test *a priori* ecological hypotheses by (1) bringing out *all* the variance of **Y** that is related to **X** and (2) allowing formal tests of these hypotheses to be performed, as detailed below. Further examination of the unexplained variability may help generate new hypotheses, to be tested using new field observations (Section 13.5).

Canonical form

In mathematics, a *canonical form* (from the Greek κανων, pronounced "kanôn", rule) is the simplest and most comprehensive form to which certain functions, relations, or expressions can be reduced without loss of generality. For example, the

canonical form of a covariance matrix is its matrix of eigenvalues. In general, methods of canonical analysis use eigenanalysis (i.e. calculation of eigenvalues and eigenvectors), although some extensions of canonical analysis have been described that use multidimensional scaling (nMDS) algorithms (Section 9.4).

There are two main families of canonical ordination methods: asymmetric and symmetric. In the asymmetric forms of analysis, there is a response data set and an explanatory data set, which are represented by $\mathbf{Y}$ and $\mathbf{X}$, respectively, in this chapter. The asymmetric methods are redundancy analysis (RDA), canonical correspondence analysis (CCA), and linear discriminant analysis (LDA). In contrast, symmetric methods are used in cases where the two data sets, called $\mathbf{Y}_1$ by $\mathbf{Y}_2$ to mark the symmetry, play the same role in the study; this means that an analysis of $\mathbf{Y}_1$ by $\mathbf{Y}_2$ produces the same result as an analysis of $\mathbf{Y}_2$ by $\mathbf{Y}_1$. These methods include canonical correlation analysis (CCorA), co-inertia analysis (CoIA), Procrustes analysis (Proc), and some others.

Interrelationships among the variables involved in canonical analysis may be represented by the following partitioned covariance matrix, resulting from the concatenation of the $\mathbf{Y}$ (or $\mathbf{Y}_1$, order $n \times p$) and $\mathbf{X}$ (or $\mathbf{Y}_2$, $n \times m$) data sets. The joint dispersion matrix $\mathbf{S_{Y+X}}$ contains blocks that are identified as follows for convenience:

$$\mathbf{S_{Y+X}} = \left[\begin{array}{ccc|ccc} s_{y_1,y_1} & \cdots & s_{y_1,y_p} & s_{y_1,x_1} & \cdots & s_{y_1,x_m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ s_{y_p,y_1} & \cdots & s_{y_p,y_p} & s_{y_p,x_1} & \cdots & s_{y_p,x_m} \\ \hline s_{x_1,y_1} & \cdots & s_{x_1,y_p} & s_{x_1,x_1} & \cdots & s_{x_1,x_m} \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots \\ s_{x_m,y_1} & \cdots & s_{x_m,y_p} & s_{x_m,x_1} & \cdots & s_{x_m,x_m} \end{array}\right] = \left[\begin{array}{cc} \mathbf{S_{YY}} & \mathbf{S_{YX}} \\ \mathbf{S_{XY}} & \mathbf{S_{XX}} \end{array}\right] = \left[\begin{array}{cc} \mathbf{S_{YY}} & \mathbf{S_{YX}} \\ \mathbf{S'_{YX}} & \mathbf{S_{XX}} \end{array}\right] \qquad \textbf{(11.1)}$$

Submatrices $\mathbf{S_{YY}}$ (order $p \times p$) and $\mathbf{S_{XX}}$ ($m \times m$) concern each of the two sets of descriptors, respectively, whereas $\mathbf{S_{YX}}$ ($p \times m$) and its transpose $\mathbf{S'_{YX}} = \mathbf{S_{XY}}$ ($m \times p$) account for the covariances among the descriptors of the two groups, as in eq. 4.27.

Asymmetric, canonical analysis

Asymmetric canonical analysis combines the concepts of ordination and regression. It involves a response matrix $\mathbf{Y}$ and an explanatory matrix $\mathbf{X}$. As it was the case with the simple ordination methods (Chapter 9 and Fig. 11.1a), the asymmetric methods of canonical analysis produce a single ordination of the objects, which may be plotted in a scatter diagram. With the symmetric methods on the contrary, two different ordinations of the objects are produced, one for each data set; see below.

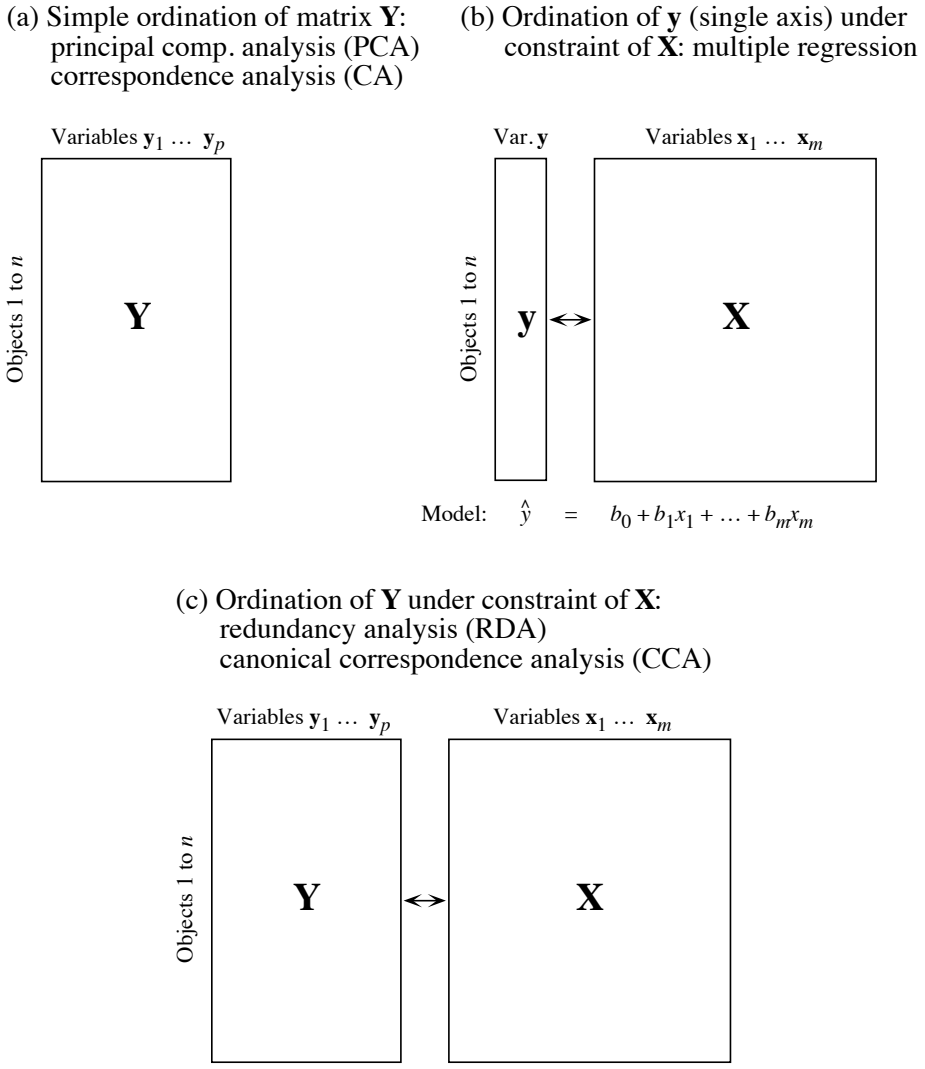


Figure 11.1 Relationships between (a) ordination, (b) regression, and (c) two asymmetric forms of canonical analysis (RDA and CCA). In (c), each canonical axis of **Y** is constrained to be a linear combination of the explanatory variables **X**.

Redundancy analysis (RDA, Section 11.1) and canonical correspondence analysis (CCA, Section 11.2) are related to multiple linear regression. In Subsection 10.3.3, multiple regression was described as a method for modelling a response variable **y** using a set of explanatory variables assembled into a data table **X**. Another aspect of regression analysis must be stressed: while the original response variable **y** provides,

by itself, an ordination of the objects in one dimension, the vector of fitted values (eq. 10.15)

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_p x_{ip}$$

creates a new one-dimensional ordination of the same objects (Fig. 11.1b). The ordinations corresponding to $\mathbf{y}$ and $\hat{\mathbf{y}}$ differ; the square of their correlation is the coefficient of determination (eq. 10.20) of the multiple regression model:

$$R^2_{\mathbf{y}|\mathbf{X}} = [r(\mathbf{y}, \hat{\mathbf{y}})]^2 \qquad \textbf{(11.2)}$$

So, multiple regression creates a correspondence between ordinations $\mathbf{y}$ and $\hat{\mathbf{y}}$, because ordination $\hat{\mathbf{y}}$ is constrained to be optimally (in the least-squares sense) and linearly related to the variables in $\mathbf{X}$. The constraint implemented in multiple regression maximizes R^2. The asymmetric methods of canonical analysis share this property.

Asymmetric canonical analysis combines the properties of two families of methods, i.e. ordination and regression (Fig. 11.1c). It produces ordinations of $\mathbf{Y}$ that are constrained to be linearly related to a second set of variables $\mathbf{X}$, and the results are plotted in reduced space. The way in which the relationship between $\mathbf{Y}$ and $\mathbf{X}$ is established differs among methods of asymmetric canonical analysis.

- In redundancy analysis (RDA, Section 11.1), each canonical ordination axis corresponds to a direction, in the multivariate scatter of objects, that is maximally related to a linear combination of the explanatory variables $\mathbf{X}$. A canonical axis is thus similar to a principal component (Box 9.1). Two ordinations of the objects may be plotted along the canonical axes: (1) linear combinations of the $\mathbf{Y}$ variables (matrix $\mathbf{F}$, eq. 11.17), as in PCA, and (2) linear combinations of the fitted $\hat{\mathbf{Y}}$ variables (matrix $\mathbf{Z}$, eq. 11.18), which are thus also linear combinations of the $\mathbf{X}$ variables. RDA preserves the Euclidean distances among objects in matrix $\hat{\mathbf{Y}}$, which contains values of $\mathbf{Y}$ fitted by regression to the explanatory variables $\mathbf{X}$ (Fig. 11.2); variables in $\hat{\mathbf{Y}}$ are therefore linear combinations of the $\mathbf{X}$ variables.

- Canonical correspondence analysis (CCA, Section 11.2) is similar to RDA. The difference is that CCA preserves the χ^2 distance (as in correspondence analysis), instead of the Euclidean distance among objects in matrix $\hat{\mathbf{Y}}$. Calculations are a bit more complex since matrix $\hat{\mathbf{Y}}$ contains fitted values obtained by weighted linear regression of matrix $\overline{\mathbf{Q}}$ of correspondence analysis (eq. 9.24) on the explanatory variables $\mathbf{X}$. As in RDA, two ordinations of the objects may be plotted.

- In linear discriminant analysis (Section 11.3), the objects are divided into k groups, described by a qualitative descriptor (factor) forming the response matrix $\mathbf{Y}$. The method seeks linear combinations of explanatory variables (matrix $\mathbf{X}$) that explain the classification in $\mathbf{Y}$ by maximizing the dispersion of the centroids of the k groups. This is obtained by maximizing the ratio of the among-object-group dispersion over the pooled within-object-group dispersion (eq. 11.33).

Symmetric, canonical analysis

The *symmetric forms of canonical analysis* described in this book are the following:

- In canonical correlation analysis (CCorA, Section 11.4), the canonical axes maximize the correlation between linear combinations of the two sets of variables $\mathbf{Y}_1$ and $\mathbf{Y}_2$. This is obtained by maximizing the squared among-variable-set correlations (Table 11.10). Two different ordinations of the objects are obtained, one for data set $\mathbf{Y}_1$ and the other for $\mathbf{Y}_2$.

- Co-inertia analysis (CoIA) and Procrustes analysis (Proc) (Section 11.5) search for common structures between two data sets $\mathbf{Y}_1$ and $\mathbf{Y}_2$ describing the same objects. Each object has two representations in the joint plot, one from $\mathbf{Y}_1$ and the other from $\mathbf{Y}_2$.

The application of the various methods of canonical analysis to ecological data was briefly discussed in Section 10.2. In summary, when one of the data sets (**Y**) is to be explained by another (**X**), the asymmetric forms of canonical analysis should be used; the methods are redundancy analysis (RDA) and canonical correspondence analysis (CCA) when **Y** is a full table of response variables, and linear discriminant analysis (LDA) when **Y** contains a classification of the objects. RDA is used when the **X** variables display linear relationships with the **Y** variables, whereas CCA can be used in the cases where correspondence analysis (CA, Section 9.2) would be appropriate for an ordination of **Y** alone. Linear discriminant analysis is applicable when the response data set contains a classification of the objects or an ANOVA factor; in ecology, LDA is used mostly to discriminate among groups of sites using descriptors of the physical environment (Section 11.3). In contrast, canonical correlation analysis (CCorA), co-inertia analysis (CoIA) and Procrustes analysis (Proc) are used to relate two data sets describing the same objects in a correlative framework (Sections 11.4 and 11.5).

Canonical analysis has become an instrument of choice for ecological analysis. A bibliography on the applications of canonical analysis to ecology, covering the period 1986 to 1996, contains a total of 804 entries (Birks *et al.*, 1998). CCorA and discriminant analysis are available in most commercial statistical packages. For RDA, CCA, CoIA and Proc, one must rely on specialized ordination packages and R functions. CANOCO (ter Braak, 1988b) was the first ordination package that made RDA and CCA available to users. These methods are also available in PC-ORD and SYN-TAX 2000. See Section 11.7.

11.1 Redundancy analysis (RDA)

Redundancy analysis (RDA) is the direct extension of multiple regression to the modelling of multivariate response data. The analysis is asymmetric: **Y** ($n \times p$) is a table of response variables and **X** ($n \times m$) is a table of explanatory variables. In RDA, the ordination of **Y** is constrained in such a way that the resulting ordination axes (matrix **Z** below) are linear combinations of the variables in **X**. The difference between

RDA and canonical correlation analysis (CCorA, Section 11.4) is the same as that between simple linear regression (asymmetric analysis) and linear correlation analysis (symmetric); see Box 10.1.

In RDA, the ordination axes are obtained by principal component analysis (PCA, Section 9.1) of a matrix $\hat{\mathbf{Y}}$, computed by fitting the **Y** variables to **X** by multivariate linear regression (details in Subsection 11.1.1). So, in scaling type 1 plots (Subsection 11.1.3), RDA preserves the Euclidean distance among objects (D_1, Chapter 7): the ordination of the points in matrix **Z** is a PCA rotation of the points in $\hat{\mathbf{Y}}$. The ordination axes in **Z** differ, of course, from the principal components that could be computed directly from the **Y** data table because they are constrained to be linear combinations of the variables in **X**. Prior to RDA, the data in **Y** must be at least centred, or transformed following the same principles as in PCA.

1 — Simple RDA

Canonical redundancy analysis was first described by Rao (1964). In his 1973 book (p. 594-595), he proposed the topic to readers as an exercise at the end of his Chapter 8 on multivariate analysis. Rao called the method *Principal components of instrumental variables*. RDA was later rediscovered by Wollenberg (1977) who called the method *Redundancy analysis* by reference to the *redundancy index* of Stewart & Love (1968), which is the proportion of the variance of the response data matrix **Y** that is accounted for by the explanatory matrix **X**. Redundancy is synonymous with explained variance (Gittins, 1985). In his paper, Wollenberg did not refer to Rao's paper (1964) and book (1973). Wollenberg's equation, which only applied to correlation matrices, was less general than that of Rao which involved covariance matrices in general.

Redundancy

Redundancy analysis (RDA) of a response matrix **Y** (with n objects and p variables) by an explanatory matrix **X** (with n objects and m variables) is called simple RDA in Subsections 11.1.1 to 11.1.5, by opposition to partial RDA, described in Subsections 11.1.6 to 11.1.10, which involves a matrix of covariables **W**. Simple RDA involves two computational steps (Fig. 11.2). In the algebraic development that follows, the columns of matrices **Y** and **X** are centred to have means of 0. In computer software, the columns of **X** may be standardized for programming convenience, but this has no effect on the results of the analysis since the matrix of fitted values $\hat{\mathbf{Y}}$ is identical when computed from centred or standardized **X** variables. As in PCA, the variables in **Y** should be standardized if they are not dimensionally homogeneous (e.g. if they are a mixture of temperatures, concentrations, and pH values). Transformations applicable to community composition data (presence-absence or abundance) are described in Section 7.7. As in multiple regression analysis, matrix **X** can contain explanatory variables of different mathematical types: quantitative, multi-state qualitative (e.g. ANOVA factors), or binary variables; see the last five paragraphs of Subsection 10.3.3. If present, collinearity among the **X** variables should be reduced prior to RDA using the methods described for multiple regression in Subsection 10.3.3. Chapters 13 and 14 will show how different expressions of spatial relationships can be used as the explanatory matrix **X** in RDA.

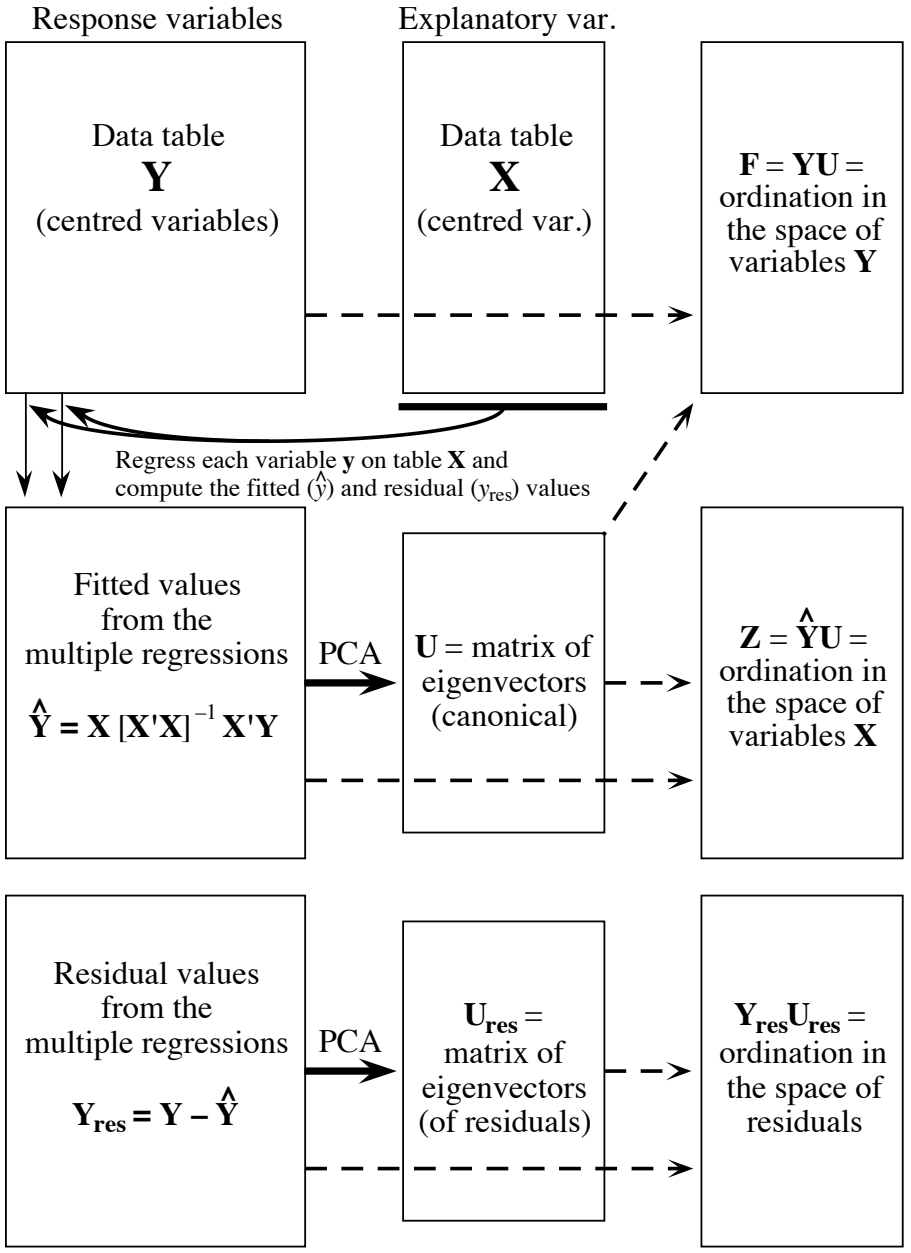


Figure 11.2 Redundancy analysis may be understood as a two-step process: (1) regress each variable in **Y** on all variables in **X** and compute the fitted values; (2) carry out a PCA of the matrix of fitted values to obtain the eigenvalues and eigenvectors. Two ordinations are obtained, one (**F**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

The variable distributions should be examined for normality at this stage, as well as bivariate plots within and between the sets **Y** and **X**. Because RDA is a linear model based on multiple linear regression, data transformations (Section 1.5) should be applied as needed to linearize the relationships and make the frequency distributions as symmetric as possible, thus reducing the effect of outliers.

- Step 1 is a *multivariate linear regression* of **Y** on **X** (eq. 10.16), which produces a matrix of fitted values $\hat{\mathbf{Y}}$ through the linear equation:

$$\hat{\mathbf{Y}} = \mathbf{X}\,[\mathbf{X}'\mathbf{X}]^{-1}\,\mathbf{X}'\mathbf{Y} \qquad \textbf{(11.3)}$$

This is equivalent to a series of multiple linear regressions of the individual variables of **Y** on **X** to calculate vectors of fitted values followed by stacking these column vectors side by side into matrix $\hat{\mathbf{Y}}$. In principle, model II regression should be used when the explanatory variables **X** are *random*, by opposition to *controlled* (Subsection 10.3.2). Ordinary least squares (OLS) are used in eq. 11.3 because, among the model II regression methods, OLS produces fitted values with the smallest error for given values of the predictors (Table 10.4). For efficiency reasons in computer software, matrix $\hat{\mathbf{Y}}$ may be computed through QR decomposition instead of eq. 11.3.

- Step 2 is a principal component analysis of $\hat{\mathbf{Y}}$. This PCA produces the canonical eigenvalues and eigenvectors, as well as matrix **Z** containing the canonical axes (object ordination scores, like matrix **F** in PCA). That step is performed to obtain reduced-space ordination diagrams displaying the objects, response variables, and explanatory variables for the most important axes of the canonical relationship. The PCA step is pertinent only if a significant canonical relationship has been found between **Y** and **X** through an appropriate test of significance (Subsection 11.1.2).

Like the fitted values of a multiple linear regression, which are linear combinations of the explanatory variables, the canonical axes (object ordination scores) are also linear combinations of the explanatory variables in **X**. That RDA axes are linear combinations of the explanatory variables is the fundamental property of RDA (ter Braak, 1987c; ter Braak and Prentice, 1988). Individual canonical axes can be tested for significance to determine which ones are important enough to warrant consideration, plotting, and detailed analysis.

2 — Statistics in simple RDA

After step 1 of RDA, one can compute the following informative statistics.

Redundancy statistic (R^2)

1. From matrices **Y** and $\hat{\mathbf{Y}}$, one can calculate the canonical R^2, which Miller and Farr (1971) called the *bimultivariate redundancy statistic*. This statistic measures the strength of the linear relationship between **Y** and **X**:

$$R^2_{\mathbf{Y}|\mathbf{X}} = \frac{\mathrm{SS}\,(\hat{\mathbf{Y}})}{\mathrm{SS}\,(\mathbf{Y})} \qquad \textbf{(11.4)}$$

where SS ($\hat{\mathbf{Y}}$) is the total sum of squares (or sum of squared deviations from the means) of $\hat{\mathbf{Y}}$ and SS($\mathbf{Y}$) is the total sum of squares of $\mathbf{Y}$. The canonical R^2 is constructed in the same way and has the same meaning as the R^2 statistic in multiple regression (eq. 10.20): it is the proportion of the variation of $\mathbf{Y}$ explained by a linear model of the variables in $\mathbf{X}$.

Note: in the absence of relationship between $\mathbf{Y}$ and $\mathbf{X}$, the expected value of R^2 in multiple regression and in RDA is not 0 but $m/(n-1)$. This is because a matrix $\mathbf{X}$ containing $m = (n-1)$ columns of random numbers produces an R^2 of 1; this surprising fact can easily verify numerically by computing a multiple regression or a RDA with a matrix $\mathbf{X}$ containing $(n-1)$ columns of random numbers. Hence, the expected value (E) of the R^2 produced by a single explanatory variable made of random numbers is $E(R^2) = 1/(n-1)$, and $E(R^2) = m/(n-1)$ for m explanatory variables. This is illustrated in the numerical simulation results presented by Peres-Neto *et al*. (2006).

Adjusted R^2

2. The adjusted R^2 (R^2_a) is computed as in eq. 10.21 (Ezekiel, 1930):

$$R^2_a = 1 - (1 - R^2_{\mathbf{Y}|\mathbf{X}}) \frac{(n-1)}{(n-m-1)} \qquad \textbf{(11.5)}$$

where m is the number of explanatory variables in $\mathbf{X}$ or, more precisely, the rank of the variance-covariance matrix of $\mathbf{X}$.

F-statistic
Overall test

3. The F-statistic for the overall test of significance is constructed as follows (Miller, 1975):

$$F = \frac{R^2_{\mathbf{Y}_{stand}|\mathbf{X}} / mp}{(1 - R^2_{\mathbf{Y}_{stand}|\mathbf{X}}) / (n-m-1)\, p} \qquad \textbf{(11.6)}$$

This statistic is used to perform the overall test of significance of the canonical relationship. The null hypothesis of the test is H_0: the strength of the linear relationship, measured by the canonical R^2, is not larger than the value that would be obtained for unrelated $\mathbf{Y}$ and $\mathbf{X}$ matrices of the same sizes.

When the variables of $\mathbf{Y}$ are standardized ($\mathbf{Y}_{stand}$) and the error distribution is normal, the F-statistic (eq. 11.6) can be tested for significance using the Fisher-Snedecor F-distribution with degrees of freedom $\nu_1 = mp$ and $\nu_2 = p(n-m-1)$. p is the number of response variables in $\mathbf{Y}$. Because m parameters were estimated for each of the p multiple regressions used to compute the vectors of fitted values forming the p columns of $\hat{\mathbf{Y}}$, a total of mp parameters were estimated. This is why there are $\nu_1 = mp$ degrees of freedom attached to the numerator of F. Each multiple regression equation has residual degrees of freedom equal to $(n-m-1)$, so the total number of degrees of freedom of the denominator, ν_2, is p times $(n-m-1)$. Miller (1975) conducted numerical simulations in the multivariate normal case, with combinations of m and p

from 2 to 15 and sample sizes of $n = 30$ to 160. He showed that eq. 11.6 produced distributions of F values that were very close to theoretical F-distributions with the same numbers of degrees of freedom. Additional simulations conducted by Legendre *et al*. (2011, Appendix A) confirmed that the parametric test of significance had correct levels of type I error when **Y** was standardized. This was not the case, however, for non-standardized matrices of response variables **Y** generated with equal or unequal population variances, especially when the error was not normal. Permutation tests always had correct levels of type I error in these simulations. The effect of correlations among the standardized response variables in **Y** on the validity of the parametric test remains to be investigated.

In many instances, the response variables should not be standardized prior to RDA. With community composition data (species abundances), for example, the variances of the species should be preserved in most analyses since abundant and rare species do not play the same roles in ecosystems. A permutation test should always be used in that case. For permutation tests, one can simplify eq. 11.6 of the F-statistic by eliminating the constant p from the numerator and denominator:

Permutation test

$$F = \frac{R^2_{\mathbf{Y}|\mathbf{X}} / m}{(1 - R^2_{\mathbf{Y}|\mathbf{X}}) / (n - m - 1)} \qquad \textbf{(11.7)}$$

While the numerator and denominator of eq. 11.6 indicate the numbers of degrees of freedom for a correct parametric test of F, eliminating p from both does not change the computed value of F. Equation 11.7 is the one used for permutation tests in programs of canonical analysis such as CANOCO and VEGAN's ***rda()***. Actually, the degrees of freedom can be entirely eliminated from statistic equations used in permutation tests since they are invariant across all permutations of the data. However, most computer programs and functions that carry out permutation tests display them to allow comparison with the F-statistic used in parametric tests.

Test of individual axes

4. Individual canonical axes can be tested for significance. Since one deals with complex, multivariate data influenced by many factors, several independent structures may coexist in the response data. If these structures are linearly independent, they should appear on different canonical axes. The results of the tests of individual axes allow researchers to determine which of the canonical axes represent variation that is more structured than random. Canonical axes that do not explain more variation than random should be identified since they do not need to be further considered in the interpretation of the results.

Two methods, called the *forward* and *marginal* testing procedures, can be used for testing individual axes. The forward method was developed by Cajo J. F. ter Braak and implemented in the CANOCO package since version 3.10 (ter Braak, 1990). The marginal method was developed by Jari Oksanen for the ***permutest.cca()*** function of the VEGAN R package; that function carries out tests of significance of the canonical axes when users call the ***anova.cca()*** function with parameter *by="axis"*, after

canonical analysis by functions ***rda()*** or ***cca()***. In a simulation study, Legendre *et al*. (2011) showed that these two methods had correct levels of type I error and comparable powers. This latter study also investigated a third method, the simultaneous test of all canonical axes, which was shown to be invalid.

The null hypothesis for the test of significance of the j^{th} canonical axis is H_0: the linear dependence of the response variables **Y** on the explanatory variables **X** is less than *j*-dimensional. More informally, the null hypothesis is that the j^{th} axis under test explains no more variation than a random axis of the same order (*j*), given the variation explained by the previously tested axes. The test of individual canonical axes can also be carried out in partial RDA (Subsection 11.1.6), a form of RDA that incorporates a matrix of covariables **W**.

3 — The algebra of simple RDA

The eigenanalysis equation for redundancy analysis, which is an asymmetric form of analysis, can be obtained from eq. 11.48 of canonical correlation analysis (CCorA, Section 11.4), which is a symmetric form of analysis, by changing the $\mathbf{S}_{\mathbf{YY}}^{-1}$ matrix (called $\mathbf{S}_{11}^{-1}$ in eq. 11.48) into an identity matrix **I**. The latter does not have to be written after matrix $\mathbf{S}_{\mathbf{YX}}'$ and thus disappears from the equation (Rao, 1973; ter Braak, 1987c):

$$(\mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}_{\mathbf{YX}}' - \lambda_k\mathbf{I})\,\mathbf{u}_k = \mathbf{0} \qquad \textbf{(11.8)}$$

The covariance relationships among the explanatory variables, $\mathbf{S}_{\mathbf{XX}}^{-1}$, remains included in the equation. Equation 11.8 differs from the original formulations by Rao (1964, 1973) and Wollenberg (1977), but it produces the same canonical eigenvalues.

Equation 11.8 is the end result of carrying out the two steps described in the previous subsection, which characterize RDA: (1) a multivariate regression of **Y** on **X** to obtain a matrix of fitted values $\hat{\mathbf{Y}}$, followed by (2) a PCA of that matrix of fitted values. The asymmetric nature of RDA comes from the fact that multivariate regression (eqs. 10.16 and 11.3) is an asymmetric analysis, just as its univariate counterpart, multiple linear regression, where **y** is the response vector and **X** is the explanatory matrix. The developments that follow show that these two computational steps produce eq. 11.8.

1) For *each* response variable in matrix **Y**, compute a multiple linear regression on all variables in matrix **X**. For each regression, the coefficients are computed as follows (eq. 2.19):

$$\mathbf{b} = [\mathbf{X}'\mathbf{X}]^{-1}\,\mathbf{X}'\mathbf{y}$$

The matrix containing all regression coefficients can be obtained by a single matrix operation (equation without number above eq. 10.16 in Subsection 10.3.3):

$$\mathbf{B} = [\mathbf{X}'\mathbf{X}]^{-1}\,\mathbf{X}'\mathbf{Y} \tag{11.9}$$

where $\mathbf{B}$ ($m \times p$) is the matrix of regression coefficients of all p response variables $\mathbf{Y}$ on the m explanatory variables $\mathbf{X}$.

As in multiple regression, the fitted values $[\hat{y}]$ can be computed by a single matrix operation:

$$\hat{\mathbf{Y}} = \mathbf{X}\,\mathbf{B} \tag{11.10}$$

This is the multivariate extension of eq. 10.1. Replacing $\mathbf{B}$ by the expression from eq. 11.9, eq. 11.10 becomes:

$$\hat{\mathbf{Y}} = \mathbf{X}\,[\mathbf{X}'\,\mathbf{X}]^{-1}\,\mathbf{X}'\,\mathbf{Y} \tag{11.11}$$

which is the multivariate linear regression equation (eq. 10.16). Because the variables in $\mathbf{X}$ and $\mathbf{Y}$ were centred on their respective means, there are no intercept parameters in the column vectors of regression coefficients forming $\mathbf{B}$, and the column vectors in $\hat{\mathbf{Y}}$ are also centred.

2) The covariance matrix corresponding to the table of fitted values $\hat{\mathbf{Y}}$ is computed using eq. 4.6:

$$\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}} = [1/(n-1)]\;\hat{\mathbf{Y}}'\hat{\mathbf{Y}} \tag{11.12}$$

Replacing $\hat{\mathbf{Y}}$ by the expression from eq. 11.11, eq. 11.12 becomes:

$$\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}} = [1/(n-1)]\,\mathbf{Y}'\,\mathbf{X}\,[\mathbf{X}'\,\mathbf{X}]^{-1}\,\mathbf{X}'\,\mathbf{X}\,[\mathbf{X}'\,\mathbf{X}]^{-1}\,\mathbf{X}'\,\mathbf{Y} \tag{11.13}$$

This equation reduces to:

$$\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}} = \mathbf{S}_{\mathbf{YX}}\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}'_{\mathbf{YX}} \tag{11.14}$$

where $\mathbf{S}_{\mathbf{YY}}$ is the ($p \times p$) covariance matrix among the response variables, $\mathbf{S}_{\mathbf{XX}}$ the ($m \times m$) covariance matrix among the explanatory variables (it is actually a matrix $\mathbf{R}_{\mathbf{XX}}$ when the $\mathbf{X}$ variables have been standardized), and $\mathbf{S}_{\mathbf{YX}}$ is the ($p \times m$) covariance matrix among the variables of the two sets; the order of its transpose $\mathbf{S}'_{\mathbf{YX}} = \mathbf{S}_{\mathbf{XY}}$ is ($m \times p$). If the $\mathbf{Y}$ variables had also been standardized, this equation would read $\mathbf{R}_{\mathbf{YX}}\mathbf{R}_{\mathbf{XX}}^{-1}\mathbf{R}'_{\mathbf{YX}}$, which is the multivariate form of the equation for the coefficient of multiple determination (eq. 4.31).

3) The matrix of fitted values $\hat{\mathbf{Y}}$ is subjected to principal component analysis to reduce the dimensionality of the solution. This corresponds to solving the eigenvalue problem:

$$(\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}} - \lambda_k \mathbf{I})\, \mathbf{u}_k = \mathbf{0} \qquad \textbf{(11.15)}$$

which, using eq. 11.14, translates into:

$$(\mathbf{S}_{\mathbf{YX}} \mathbf{S}_{\mathbf{XX}}^{-1} \mathbf{S}_{\mathbf{YX}}' - \lambda_k \mathbf{I})\, \mathbf{u}_k = \mathbf{0} \qquad \textbf{(11.16)}$$

This is the equation for redundancy analysis (eq. 11.8). Different programs may express the eigenvalues in different ways: raw eigenvalues, fractions of the total variance in **Y**, or percentages; see Tables 11.2 and 11.4 for examples.

The matrix containing the normalized canonical eigenvectors $\mathbf{u}_k$ is called **U**. The eigenvectors give the contributions of the descriptors in matrix $\hat{\mathbf{Y}}$ to the various canonical axes. Matrix **U**, of size $(p \times p)$, contains only $\min[p, m, n-1]$ eigenvectors with non-zero eigenvalues, since the number of canonical eigenvectors cannot exceed the minimum of p, m and $(n-1)$:

- It cannot exceed p, which is the dimension of the reference space of matrix **Y**. This is obvious in multiple regression where matrix **Y** contains a single variable; the ordination given by the fitted values $\hat{y}$ is one-dimensional.

- It cannot exceed m, which is the number of variables in **X**. Consider an extreme example: if **X** contains a single explanatory variable ($m = 1$), regressing all p variables in **Y** on this single explanatory variable produces p fitted vectors $\hat{y}$ which all point in the same direction of the p-dimensional space; a principal component analysis of matrix $\hat{\mathbf{Y}}$ of these fitted vectors can only produce one common (canonical) axis.

- It cannot exceed $(n-1)$, which is the maximum number of dimensions required to represent n points in Euclidean space.

The canonical coefficients in the normalized matrix **U** give the contributions of the variables of $\hat{\mathbf{Y}}$ to the canonical axes. They should be interpreted as in PCA. Matrix **U** is used to produce scaling 1 biplot or triplot diagrams, described below. For scaling 2 plots, **U** is rescaled in such a way that the length of each eigenvector is $\sqrt{\lambda_k}$.

If **X** and **Y** are made to contain the same data (i.e. $\mathbf{X} = \mathbf{Y}$), eq. 11.16 becomes $(\mathbf{S}_{\mathbf{YY}} - \lambda_k \mathbf{I})\, \mathbf{u}_k = \mathbf{0}$, which is the equation for principal component analysis (eq. 9.1). The result of RDA is then a principal component analysis of data table **Y**, a fact that was pointed out by Rao (1964, 1973) and by Wollenberg (1977). Another way to look at this point is to say that a RDA of **Y** by **Y** is a PCA of **Y** because $\hat{\mathbf{Y}} = \mathbf{Y}$ in that case.

Additional computations must be done to produce the RDA triplot diagram (below), which contains three types of elements: response variables (e.g. species), objects (e.g. sites), and explanatory variables.

4) The ordination of objects in the space of the response variables **Y** is obtained directly from the centred matrix $\mathbf{Y}_c$, using the standard equation for principal components (matrix **F**, eq. 9.4) and matrix **U** of the eigenvectors $\mathbf{u}_k$ found in eq. 11.16:

$$\mathbf{F} = \mathbf{Y}_c\mathbf{U} \qquad \textbf{(11.17)}$$

Site scores

The ordination vectors (columns of **F**) defined in eq. 11.17 are called the vectors of "site scores". They have variances that are close, but not equal to the corresponding eigenvalues. How to represent matrix **F** in biplots is discussed in point 8 (below).

5) Likewise, the ordination of objects in space **X** is obtained as follows:

$$\mathbf{Z} = \hat{\mathbf{Y}}\,\mathbf{U} = \mathbf{X}\mathbf{B}\mathbf{U} \qquad \textbf{(11.18)}$$

Fitted site scores

As stated above, the vectors in matrix $\hat{\mathbf{Y}}$ are centred on their respective means. The right-hand part of eq. 11.18, obtained by replacing $\hat{\mathbf{Y}}$ by it value in eq. 11.10, shows that this ordination is a linear combinations of the **X** variables. For that reason, these ordination vectors (columns of matrix **Z**) are also called "fitted site scores", or "sample scores that are linear combinations of environmental variables" in program CANOCO. The ordination vectors, defined in eq. 11.18, have variances equal to the corresponding eigenvalues. The representation of matrix **Z** in biplots is discussed in point 8 (below).

The "site scores" of eq. 11.17 are obtained by projecting the original data (matrix **Y**) onto axis *k*; they approximate the observed data, which contain residuals ($\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{Y}_{res}$, Fig. 11.2). In contrast, the "fitted site scores" of eq. 11.18 are obtained by projecting the fitted values of the multiple regressions (matrix $\hat{\mathbf{Y}}$) onto axis *k*; they approximate the fitted data. Either set may be used in biplots; different programs offer one or the other as the default option. These plots may look very different, so users must decide which one they want to obtain and report in published papers. The practical difference between "site scores" and "fitted site scores" is further discussed in the second example below.

6) The correlation r_k between the ordination vectors in spaces **Y** (from eq. 11.17) and **X** (from eq. 11.18) for dimension *k* is called the "species-environment correlation". It measures the strength of the relationship between the two data sets as expressed by each canonical axis *k*. It should be interpreted with caution because a canonical axis with high species-environment correlation may explain but a small fraction of the variation in **Y**, which is given by the amount (or proportion) of variance of matrix **Y** explained by each canonical axis; see example in Table 11.2.

7) The last important information needed for interpretation is the contribution of the explanatory variables **X** to the canonical ordination axes. Either the regression or the correlation coefficients may be considered:

- Matrix **C** of the canonical coefficients,

$$\mathbf{C} = \mathbf{B}\,\mathbf{U} \qquad \textbf{(11.19)}$$

gives directly the weights of the explanatory variables **X** in the formation of the matrix of fitted site scores. The ordination of objects in the space of the explanatory variables can be found directly by computing **XC**; these vectors of site scores are the same as in eq. 11.18. The coefficients in the columns of matrix **C** are identical to the regression coefficients of the ordination scores from eq. 11.18 on the matrix of standardized explanatory variables **X**; they may thus be interpreted in the same way.

• Correlations may also be computed between the variables in **X**, on the one hand, and the ordination vectors, in either space **Y** (from eq. 11.17) or space **X** (from eq. 11.18), on the other hand. The correlations between **X** and the ordination vectors in space **X**, $\mathbf{R_{XZ}} = \text{cor}(\mathbf{X}, \mathbf{Z})$, are used to represent the explanatory variables in biplots.

Biplot
Triplot

8) In RDA, one can draw *biplot diagrams*, called *biplots*, which contain two sets of points as in PCA (Subsection 9.1.4), or *triplot diagrams* (*triplots*) which contain three sets: the site scores (matrices **F** or **Z**, from eqs. 11.17 and 11.18), the response variables from **Y,** and the explanatory variables from **X**. Each pair of sets of points may be drawn in a biplot. Biplots help interpret the ordination of objects in terms of **Y** and **X**. When there are too many objects, or too many variables in **Y** or **X**, separate ordination diagrams for the response and explanatory variables may be drawn and presented side by side. The construction of RDA biplot diagrams is explained in detail in ter Braak (1994); his conclusions are summarized here. As in PCA, two main types of scalings may be used (Table 9.2):

Scalings in RDA

RDA scaling type 1. — The eigenvectors in matrix **U**, representing the scores of the response variables along the canonical axes, are scaled to lengths 1. The site scores in space **X** are obtained from equation $\mathbf{Z} = \hat{\mathbf{Y}}\mathbf{U}$ (eq. 11.18); these vectors have variances equal to λ_k. The site scores in space **Y** are obtained from equation $\mathbf{F} = \mathbf{YU}$; the variances of these vectors are usually slightly larger than λ_k because **Y** contains both the fitted and residual components and has thus more total variance than $\hat{\mathbf{Y}}$. Matrices **Z** and **U**, or **F** and **U**, can be used together in biplots because the products of the eigenvectors with the site score matrices reconstruct the original matrices perfectly: $\mathbf{ZU}' = \hat{\mathbf{Y}}$ and $\mathbf{FU}' = \mathbf{Y}$, as in PCA (Subsection 9.1.4).

Matrix of biplot scores

In scaling type 1, a quantitative explanatory variable **x** is represented in the biplot or triplot using the vector of correlations of **x** with the fitted site scores, $\mathbf{r_{xZ}} = \text{cor}(\mathbf{x}, \mathbf{Z})$, modified by multiplying each correlation by $\sqrt{\lambda_k / \text{Total variance in } \mathbf{Y}}$ where λ_k is the eigenvalue of the corresponding axis *k*. The whole *matrix of biplot scores* in scaling type 1 ($\mathbf{BS}_1$) for the explanatory variables is computed as follows:

$$\mathbf{BS}_1 = (\text{Total variance in } \mathbf{Y})^{-1/2} \mathbf{R_{XZ}} \boldsymbol{\Lambda}^{1/2} \qquad \textbf{(11.20)}$$

This correction accounts for the fact that, in this scaling, the variances of the site scores differ among axes. The correlation matrix $\mathbf{R_{XZ}}$ was obtained in calculation step 7.

The consequences of this scaling, for PCA, are summarized in the central column of Table 9.2. The graphs resulting from this scaling, called *distance biplots* or *triplots*,

focus the interpretation on the ordination of objects because the distances among objects approximate their Euclidean distances in the spaces corresponding to matrices $\mathbf{Y}$ or $\hat{\mathbf{Y}}$.

Distance triplot

The main features of a distance biplot or triplot are the following: (1) Distances among objects in a biplot are approximations of their fitted Euclidean distances. (2) Projecting an object at right angle on a response variable **y** approximates the fitted value (e.g. abundance) of the object along that variable, as in Fig. 9.3a. (3) The angles among variables **y** are meaningless. (4) The angle between two variables **x** and **y** in the biplot reflect their correlation. (5) Binary explanatory variables **x** may be represented as the centroids of the objects possessing state "present" or "1" for that variable. Examples are given in Subsection 11.1.4. Since a centroid represents a "mean object", its relationship to a variable **y** is found by projecting it at right angle on the variable, as for an object. Distances among centroids, and between centroids and individual objects, approximate Euclidean distances.

RDA scaling type 2. — Alternatively, one obtains response variable scores by rescaling the eigenvectors in matrix $\mathbf{U}$ to lengths $\sqrt{\lambda_k}$, using the transformation $\mathbf{U\Lambda}^{1/2}$ as in PCA (eq. 9.10). The site scores in space $\mathbf{X}$ obtained for scaling 1 (eq. 11.18) are rescaled to unit variances using the transformation $\mathbf{Z\Lambda}^{-1/2}$; this is the same transformation as used in PCA (eq. 9.14) to obtain matrix $\mathbf{G}$ of site scores in scaling 2. Likewise, the site scores in space $\mathbf{Y}$ obtained for scaling 1 are rescaled using the transformation $\mathbf{F\Lambda}^{-1/2}$; the variances of these vectors are usually slightly larger than 1 for the reason explained in the case of scaling 1. Matrices $\mathbf{Z\Lambda}^{-1/2}$ and $\mathbf{U\Lambda}^{1/2}$, or $\mathbf{F\Lambda}^{-1/2}$ and $\mathbf{U\Lambda}^{1/2}$, can be used together in biplots because the products of the eigenvectors with the site score matrices reconstruct the original matrices perfectly: $\mathbf{Z\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2}\mathbf{U}' = \hat{\mathbf{Y}}$ and $\mathbf{F\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2}\mathbf{U}' = \mathbf{Y}$, as in PCA (Subsection 9.1.4).

In scaling type 2, a quantitative explanatory variable **x** is represented in the biplot using the vector of correlations of **x** with the fitted site scores, $\mathbf{r_{xZ}} = \text{cor}(\mathbf{x}, \mathbf{Z})$, obtained in calculation step 7, without further transformation. The matrix of biplot scores ($\mathbf{BS}_2$) for the explanatory variables is then:

$$\mathbf{BS}_2 = \mathbf{R_{XZ}} = \text{cor}(\mathbf{X}, \mathbf{Z}) \qquad \textbf{(11.21)}$$

Note that $\text{cor}(\mathbf{X}, \mathbf{Z\Lambda}^{-1/2})$ produces the same correlations as $\text{cor}(\mathbf{X}, \mathbf{Z})$.

The consequences of this scaling, for PCA, are summarized in the right-hand column of Table 9.2. The graphs resulting from this scaling, called *correlation biplots* or *triplots*, focus on the relationships among the response variables (matrix $\mathbf{Y}$ or $\hat{\mathbf{Y}}$).

Correlation triplot

The main features of a correlation biplot or triplot are the following: (1) Distances among objects in the biplot *are not* approximations of their fitted Euclidean distances. (2) Projecting an object at right angle on a response variable **y** approximates the fitted value (e.g. abundance) of the object along that variable. (3) The angle between two variables **x** and **y** in the biplot reflects their correlation. (4) Projecting an object at right angle on a variable **x** approximates the value of that object along the variable.

Table 11.1 Maximum number of non-zero eigenvalues and corresponding eigenvectors that may be obtained from canonical analysis of a matrix of response variables $\mathbf{Y}(n \times p)$ and a matrix of explanatory variables $\mathbf{X}(n \times m)$ using redundancy analysis (RDA) or canonical correspondence analysis (CCA).

	Canonical eigenvalues and eigenvectors	Non-canonical eigenvalues and eigenvectors
RDA	$\min[p, m, n-1]$	$\min[p, n-1]$
CCA	$\min[(p-1), m, n-1]$	$\min[(p-1), n-1]$

(5) Binary explanatory variables may be represented as described above. Their interpretation is done in the same way as in scaling type 1, except for the fact that the distances in the biplot among centroids, and between centroids and individual objects, do not approximate Euclidean distances.

The type of scaling depends on the purpose of the plot: displaying the distances among objects or the correlations among variables. When most explanatory variables are binary, scaling type 1 is probably the most interesting; when most of the variables in set **X** are quantitative, one may prefer scaling type 2. When the first two eigenvalues are nearly equal, the two scalings lead to very similar plots.

9) Redundancy analysis usually does not completely explain the variation in the response variables (matrix **Y**). During the regression step (Fig. 11.2), regression residuals may be computed for each variable **y**; the residuals are the differences between the observed values y_{ij} in matrix **Y** and the corresponding fitted values $\hat{y}_{ij}$ in matrix $\hat{\mathbf{Y}}$. The matrix of residuals ($\mathbf{Y}_{\text{res}}$ in Fig. 11.2) is also a matrix of size $(n \times p)$. Residuals may be analysed by principal component analysis, leading to $\min[p, n-1]$ non-canonical eigenvalues and eigenvectors (Fig. 11.2, bottom). So, the full analysis of matrix **Y** (i.e. the analysis of fitted values and residuals) may lead to more eigenvectors than a principal component analysis of matrix **Y**: there is a maximum of $\min[p, m, n-1]$ non-zero canonical eigenvalues and corresponding eigenvectors, plus a maximum of $\min[p, n-1]$ non-canonical eigenvalues and eigenvectors, the latter being computed from the matrix of residuals (Table 11.1). When the variables in **X** are good predictors of the variables in **Y**, the canonical eigenvalues may be larger than the first non-canonical eigenvalues, but this is not always the case. If the variables in **X** are not good predictors of **Y**, the first non-canonical eigenvalues, computed on the residuals, may be larger than their canonical counterparts.

In the case where **Y** contains a single response variable, redundancy analysis is simply a multiple linear regression analysis. This is why variation partitioning

(Subsection 11.1.11) can be obtained for a single response variable using an R function, ***varpart()***, which was designed for the analysis of multivariate response data.

Different algorithms can be used in computer programs to compute RDA. One may go through the multiple regression and principal component analysis steps described in Fig. 11.2, or calculate the matrix corresponding to $\mathbf{S_{YX}S_{XX}^{-1}S_{YX}'}$ in eq. 11.8 and decompose it into eigenvalues and eigenvectors using standard eigen-analysis (Section 2.9). Computation of the matrix of fitted values $\hat{\mathbf{Y}}$ can be done by QR decomposition, as explained in Subsection 11.1.1, and eigen-decomposition can be replaced by singular value decomposition (SVD, Section 2.11) as shown for PCA (Subsection 9.1.9). Instead of eigen-decomposition or SVD, an iterative algorithm is used in the program CANOCO to calculate the first four canonical eigenvalues and eigenvectors (ter Braak, 1987c).

4 — Numerical examples, simple RDA

As a first example, consider again the data presented in Table 10.6. For RDA, the first five variables were assembled into matrix **Y** whereas the three spatial variables made up matrix **X**. The **Y** variables were standardized at the beginning of the calculations because they were dimensionally heterogeneous. The results of RDA are presented in Table 11.2. There are min[5, 3, 19] = 3 canonical eigenvectors in this example, and 5 non-canonical PCA axes computed from the residuals. This is a case where the canonical analysis is not very successful: the three canonical eigenvalues account together for only 28% (R^2 = 0.2807) of the variation present in the standardized response data **Y**. The first non-canonical eigenvalues are larger than any of the canonical eigenvalues. The correlations shown in Table 11.2 between the two sets of ordination axes (matrices **F** and **Z**) are rather weak. The ordination of objects along the canonical axes (calculation steps 4 and 5 of the previous subsection) as well as the contributions of the explanatory variables to the canonical ordination axes (calculation step 6) are not reported in the table.

A second example was constructed to illustrate the calculation and interpretation of redundancy analysis. In this artificial example, fish have been observed at 10 sites along a transect perpendicular to the beach of a tropical island, with water depths going from 1 to 10 m (Table 11.3). The first three sites are on sand while the other sites alternate between coral and "other substrate". The first six species avoid the sandy area, possibly because there is little food for them there, whereas the last three are ubiquitous. The sums of abundances for the 9 species are in the last row of the table. Species 1 to 6 come in three successive pairs, with distributions forming opposite gradients of abundance between sites 4 and 10. Species 1 and 2 are not associated with a single type of substrate. Species 3 and 4 are found in the coral areas only while species 5 and 6 are found on other substrates only (coral debris, turf, calcareous algae, etc.). The distributions of abundances of the ubiquitous species (7 to 9) have been produced using a random number generator, fitting the frequencies to a predetermined sum; these species will only be used to illustrate CCA in Section 11.2.

Table 11.2 Results of redundancy analysis (selected output). Matrix **Y** contained the first five variables of Table 10.6 and matrix **X**, the last three.

	Canonical axes			Non-canonical axes				
	I	II	III	IV	V	VI	VII	VIII
Eigenvalues (with respect to total variance of the standardized variables in **Y** = 5)								
	0.8044	0.5864	0.0124	1.4517	1.1165	0.5469	0.3715	0.1101
Fraction of total variance in **Y**								
	0.1609	0.1173	0.0025	0.2903	0.2233	0.1094	0.0743	0.0220
Correlations between the ordination vectors in spaces **Y** and **X**								
	0.7996	0.5936	0.1301					
Normalized eigenvectors (the rows correspond to the five standardized variables in matrix **Y**)								
1	0.2977	0.6173	–0.3441	–0.3345	0.5904	–0.1631	–0.5570	–0.4502
2	–0.6286	0.3455	0.0471	0.1753	0.5936	–0.4738	0.1769	0.6010
3	0.1664	0.4049	0.8922	–0.7254	–0.0735	0.3017	–0.2000	0.5808
4	0.6414	–0.2740	0.0928	0.4459	–0.2856	–0.1018	–0.7857	0.3031
5	0.2778	0.5105	–0.2735	0.3638	0.4605	0.8047	–0.0331	0.0832

RDA was computed using the first six species as matrix **Y**. Had the data been real, they would have been subjected to a Hellinger, chord, or chi-square transformation (Section 7.7) prior to RDA, because of the large proportion of zeros in the data. This is not done here in order to simplify the task of readers who would like to replicate the results. These same data, augmented with species 7 to 9, will be analysed using CCA in Section 11.2. Comparison of the RDA results about species 1 to 6 (Tables 11.4 and Fig. 11.3), on the one hand, to the CCA results about species 1 to 9 (Table 11.7 and Fig. 11.9), on the other hand, allows some comparison of the two methods.

The **Y** variables were not standardized: species abundances do not require standardization since they are all in the same physical dimensions. In most ecological studies, it is important to preserve the variances of the individual species in the analyses because abundant and rare species play different roles in ecosystems. Among the **X** variables, the three binary variables coding for substrate types form a collinear group. Including all three in the cross-product matrix [**X**'**X**] would prevent its inversion because the matrix would be singular (Section 2.8); this would jeopardize

Table 11.3 Artificial data set representing observations (fish abundances) at 10 sites along a tropical reef transect. The variables are further described in the text.

Site No.	Sp. 1	Sp. 2	Sp. 3	Sp. 4	Sp. 5	Sp. 6	Sp. 7	Sp. 8	Sp. 9	Depth (m)	Substrate type Coral	Sand	Other
1	1	0	0	0	0	0	2	4	4	1	0	1	0
2	0	0	0	0	0	0	5	6	1	2	0	1	0
3	0	1	0	0	0	0	0	2	3	3	0	1	0
4	11	4	0	0	8	1	6	2	0	4	0	0	1
5	11	5	17	7	0	0	6	6	2	5	1	0	0
6	9	6	0	0	6	2	10	1	4	6	0	0	1
7	9	7	13	10	0	0	4	5	4	7	1	0	0
8	7	8	0	0	4	3	6	6	4	8	0	0	1
9	7	9	10	13	0	0	6	2	0	9	1	0	0
10	5	10	0	0	2	4	0	1	3	10	0	0	1
Sum	60	50	40	30	20	10	45	35	25				

the calculation of the regression coefficients (eq. 11.9) and of the matrix of fitted values $\hat{\mathbf{Y}}$ (eq. 11.11). It is not necessary, however, to eliminate one of the dummy variables: in well-designed programs for canonical analysis, the last dummy variable is automatically eliminated from the calculations leading to $\hat{\mathbf{Y}}$, but its position in the ordination diagram is estimated in the final calculations. A group of dummy variables coding for a qualitative variable, like the substrate types here, can be replaced by a single factor-type variable in R functions such as VEGAN's ***rda()***.

Results of the analysis are presented in Table 11.4. Scaling type 1 was selected for the biplot in order to illustrate the extra calculation step required to transform the correlations into biplot scores for scaling type 1. The data could have produced 3 canonical axes and up to 6 non-canonical eigenvectors. In this example, only 4 of the 6 non-canonical axes had variances larger than 0. An overall test of significance (Subsection 11.1.2) showed that the canonical relationship between matrices **X** and **Y** was very highly significant ($p = 0.001$ after 999 permutations). The canonical axes explained 66%, 22% and 8% of the variance of the response data, respectively, for a total R^2 of 0.9597 and $R^2_a = 0.9396$. The three canonical axes were all significant ($p < 0.05$) and displayed strong species-environment correlations ($r = 0.999$, 0.997, and 0.980, respectively).

In Table 11.4, the eigenvalues are first shown with respect to the total variance of matrix **Y**, as is customary in principal component analysis. They are also presented as proportions of the total variance of **Y**; these are the eigenvalues provided by CANOCO for PCA and RDA. The species and sites are scaled for a distance triplot (RDA scaling type 1). The eigenvectors, normalized to length 1, provide the "species scores". The

Table 11.4 Results of redundancy analysis of the data in Table 11.3 (selected output). Matrix **Y**: species 1 to 6. Matrix **X**: depth and substrate classes.

	Canonical axes			Non-canonical axes			
	I	II	III	IV	V	VI	VII
Eigenvalues (with respect to total variance of **Y** = 112.88889)							
	74.52267	24.94196	8.87611	4.18878	0.31386	0.03704	0.00846
Fraction of total variance of **Y**							
	0.66014	0.22094	0.07863	0.03711	0.00278	0.00033	0.00007
Cumulative fraction of total variance of **Y** accounted for by axes 1 to *k*							
	0.66014	0.88108	0.95971	0.99682	0.99960	0.99993	1.00000
Normalized eigenvectors ("species scores"): mat. **U** for canonical, $\mathbf{U}_{\mathbf{res}}$ for non-canonical portions (Fig. 11.2)							
Species 1	0.30127	–0.64624	0.39939	–0.00656	–0.40482	0.70711	–0.16691
Species 2	0.20038	–0.47265	–0.74458	0.00656	0.40482	0.70711	0.16691
Species 3	0.74098	0.16813	0.25690	–0.68903	–0.26668	0.00000	0.67389
Species 4	0.55013	0.16841	–0.26114	0.58798	0.21510	0.00000	0.68631
Species 5	–0.11588	–0.50594	0.29319	0.37888	–0.66624	0.00000	0.12373
Species 6	–0.06292	–0.21535	–0.25679	–0.18944	0.33312	0.00000	–0.06187
Matrix **Z** for the canonical part ("fitted site scores", eq. 11.18) and **F** for the non-canonical part (eq. 9.4)							
Site 1	–6.79498	5.49498	2.24897	0.24712	1.14353	0.23570	0.01271
Site 2	–6.96197	5.91719	0.63774	0.00000	0.00000	–0.47140	0.00000
Site 3	–7.12895	6.33941	–0.97349	–0.24712	–1.14353	0.23570	–0.01271
Site 4	–3.55205	–6.52301	4.39356	2.14250	–0.28230	0.00000	0.00141
Site 5	12.69996	0.24686	3.17159	–3.80923	–0.14571	0.00000	0.10360
Site 6	–3.88603	–5.67858	1.17109	0.71417	–0.09410	0.00000	0.00047
Site 7	12.36599	1.09129	–0.05088	0.22968	0.08889	0.00000	–0.22463
Site 8	–4.22000	–4.83415	–2.05138	–0.71417	0.09410	0.00000	–0.00047
Site 9	12.03201	1.93572	–3.27335	3.57956	0.05682	0.00000	0.12103
Site 10	–4.55398	–3.98972	–5.27384	–2.14250	0.28230	0.00000	–0.00141
Correlations of environmental variables with the **Z** site scores							
Depth	0.42265	–0.55914	–0.71325				
Coral	0.98850	0.15079	–0.01178				
Sand	–0.55652	0.81760	0.14771				
Other subs.	–0.40408	–0.90584	–0.12715				
Biplot scores of environmental variables							
Depth	0.34340	–0.26282	–0.20000				
Coral	0.80314	0.07088	–0.00330				
Sand	–0.45216	0.38431	0.04142				
Other subs.	–0.32831	–0.42579	–0.03565				
Centroids, in the triplot, of the sites with code "1" for the BINARY environmental variables							
Coral	12.36599	1.09129	–0.05088				
Sand	–6.96197	5.91719	0.63774				
Other subs.	–4.05301	–5.25636	–0.44014				

"fitted site scores" (matrix **Z**) are obtained from eq. 11.18. They provide the ordination of the objects, computed from $\hat{\mathbf{Y}}$, in the space of the explanatory variables **X**. These axes are orthogonal to one another because they directly result from the PCA of $\hat{\mathbf{Y}}$. The "site scores" (matrix **F**, not shown) in the space of **Y** would be obtained by eq. 11.17. The columns of matrix **F** are, however, not orthogonal to one another because **Y** contains the "residual" components of the multiple regressions (Fig. 11.2). Both the "site scores" (matrix **F**) and "fitted site scores" (matrix **Z**) may be used in RDA triplots.

Correlations of the environmental variables with the ordination vectors can be obtained in two forms: with respect to either the "site scores" (eq. 11.17) or the "fitted site scores" (eq. 11.18). The latter set of correlations is used to draw triplots containing the sites as well as the variables from **Y** and **X**, as done in Fig. 11.3. There were three binary variables in Table 11.3. Each such variable may be represented by the centroid of the sites possessing state "1" for that variable (or else, the centroid of the sites possessing state "0"). These three variables are represented by both arrows (correlations) and symbols (centroids) in Fig. 11.3 to show the difference between these representations. In real-case triplots, only one of the two representations is used.

The fitted site scores in Table 11.4 have much larger ranges of values than the species scores and the biplot scores of environmental variables. Drawing triplots from these tables of values would produce graphs in which the arrows representing the species and environmental variables would be minute and clustered in the centre of the graph. Two strategies are used in computer software: either the tables of output results are modified to make the three sets of values to be drawn (species, sites, environmental variables) commensurable in the graph (this is the case in CANOCO and in VEGAN's function ***rda()***), or the output tables are those produced by the equations of Subsection 11.1.3 but the species and environmental variable arrows are drawn using a different scale than for the site scores (as done in Fig. 11.3).

5 — RDA and CCA of community composition data

Different approaches are available for the canonical analysis of community composition data (Fig. 11.4): the classical approaches (RDA and CCA), transformation-based RDA (tb-RDA), and distance-based RDA (db-RDA). The three approaches are discussed here in turn.

In the classical approach (Fig. 11.4a), the species-environment relationship is analysed by RDA (this section) or by CCA (Section 11.2). In the early applications of canonical analysis to community ecology, the latter was considered preferable for species data tables sampled in highly diversified regions ("long gradients"), which contain many zeros. This is the case, for example, when sampling communities along extensive spatial or temporal gradients, where the species composition may differ greatly between the two ends of the gradient. For groups of sites that were fairly homogeneous in species composition ("short gradients"), RDA was considered appropriate. A wider array of options is now available.

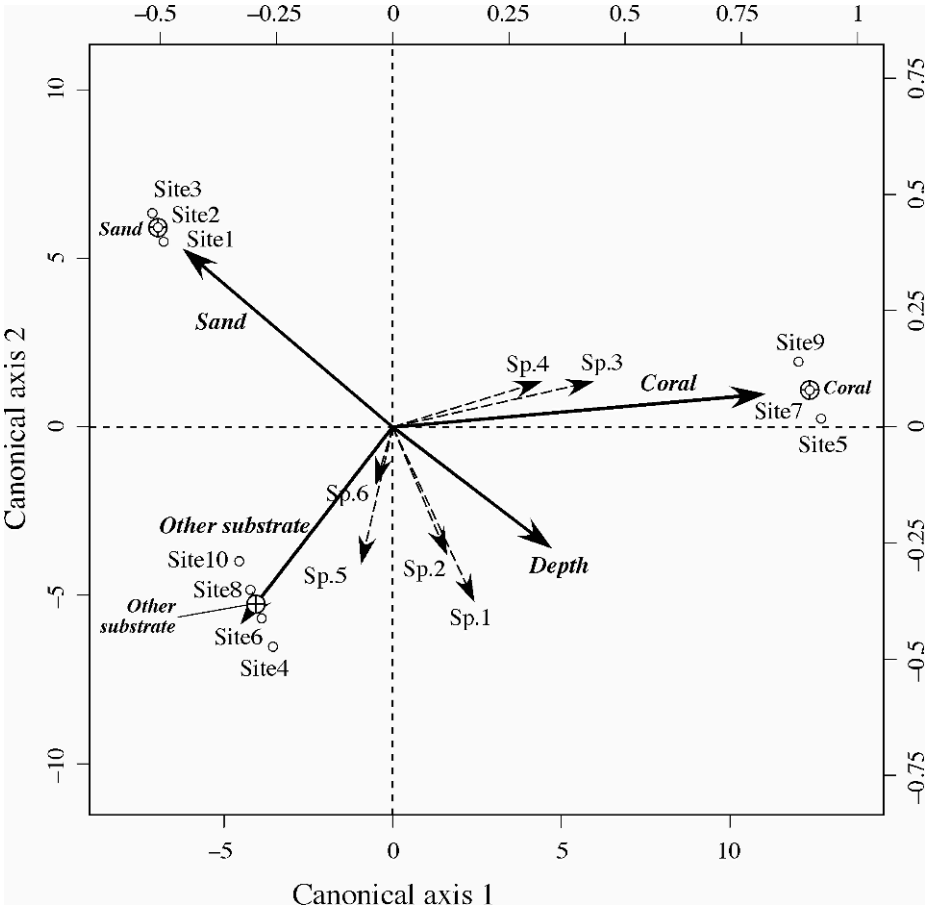


Figure 11.3 RDA triplot of the data in Table 11.3, scaling 1; the numerical results are in Table 11.4. Open circles represent the sites; the site numbers correspond to the site water depths (in m). Dashed arrows are the species. Full-line arrows represent the environmental variables. The sites are positioned in the diagram using the lower and left-hand scales, whereas the species and environmental variables are positioned using the top and right-hand scales. The "centroids of the sites with code 1 for the [three] binary environmental variables" are represented by crossed circles. Binary environmental variables are usually represented by *either* arrows *or* symbols, not both as in this triplot.

Like PCA (Fig. 9.8), RDA can be made to preserve some distance that is appropriate to study composition data along gradients, instead of the Euclidean distance. Figure 11.4b shows that composition data can be transformed using the transformations described in Section 7.7. This is the transformation-based RDA

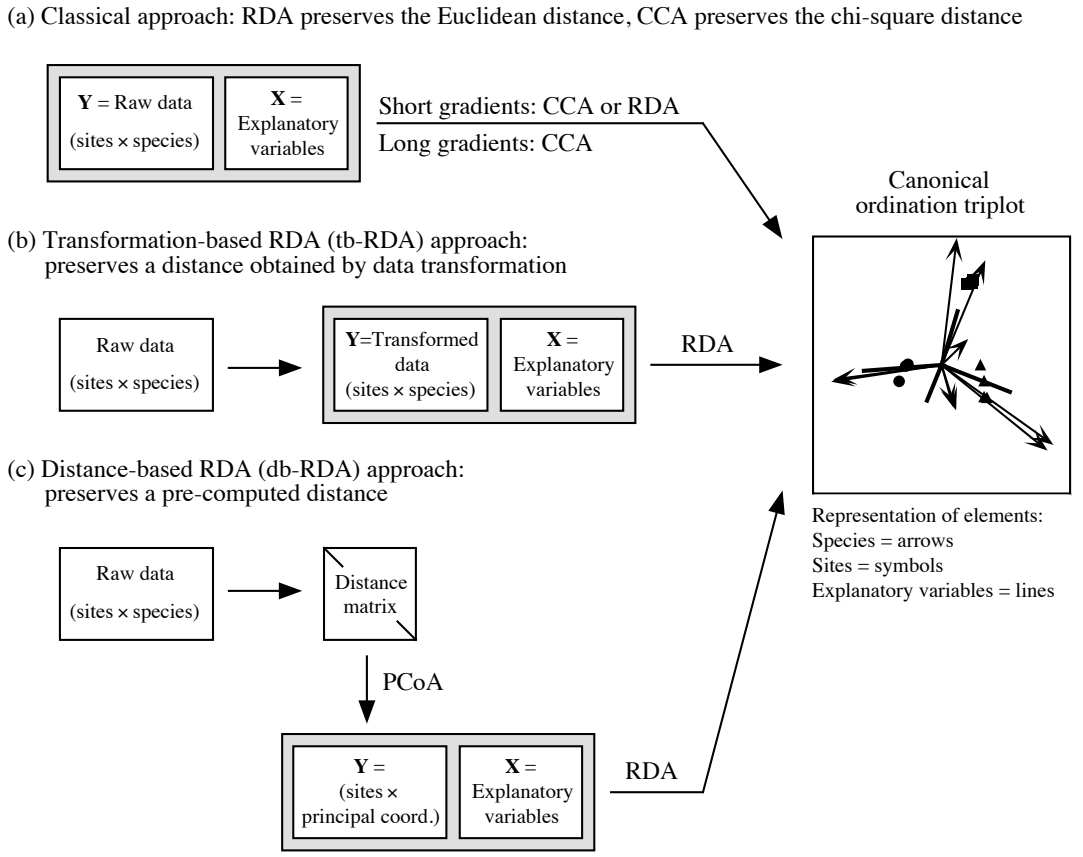


Figure 11.4 Comparison of (a) classical RDA and CCA, and (b and c) alternative approaches forcing RDA to preserve other distances adapted to community composition data. Modified from Legendre & Gallagher (2001).

tb-RDA

(Legendre & Gallagher, 2001), or tb-RDA, approach. RDA computed on data transformed by these equations will actually preserve the chord, profile, Hellinger, chi-square distance or chi-square metric among sites, depending on the transformation used.

One can also (Fig. 11.4c) compute one of the distance functions appropriate for community composition data (Table 7.4), carry out a principal coordinate analysis (PCoA) of the distance matrix, and use *all* the PCoA eigenvectors as input into a RDA.

db-RDA

This is the distance-based RDA, or db-RDA, approach advocated by Legendre & Anderson (1999).

The db-RDA approach must be used in analyses involving distance functions that cannot be obtained by a data transformation followed by RDA (tb-RDA). Among these are most of the coefficients designed for binary data, e.g. Jaccard ($\sqrt{1 - S_7}$) and Sørensen (D_{13} or $\sqrt{1 - S_8}$), as well as quantitative distance measures like the asymmetric Gower coefficient ($\sqrt{1 - S_{19}}$), the geodesic metric (D_4), Whittaker (D_9), Canberra (D_{10}), Clark (D_{11}), percentage difference (D_{14}), and mean character difference modified for species data D_{19}. Distance coefficients intended for other types of data, e.g. symmetric Gower ($\sqrt{1 - S_{15}}$), Estabrook-Rogers ($\sqrt{1 - S_{16}}$), and the generalized Mahalanobis distance for groups of observations, can also be used in canonical ordination through db-RDA. Published studies involving db-RDA include Anderson (1999), Geffen *et al*. (2004) and Lear *et al*. (2008).

6 — Partial RDA

Partial RDA is the analysis of response variables **Y** by explanatory variables **X** in the presence of additional explanatory variables, **W**, called covariables. In partial RDA, the linear effects of the explanatory variables **X** on the response variables **Y** are adjusted for the effects of the covariables **W**, as was done in partial linear regression (Subsection 10.3.5). Partial RDA was first proposed by Davies & Tso (1982, their Section 10.3).

In multiple regression, the partial regression of **y** on **X** in the presence of covariables **W** can be computed in two different ways that were described in Subsection 10.3.5. After computing the residuals of **y** on **W** (noted $\mathbf{y}_{res|\mathbf{W}}$) and the residuals of **X** on **W** (noted $\mathbf{X}_{res|\mathbf{W}}$), one could either (1) regress $\mathbf{y}_{res|\mathbf{W}}$ on $\mathbf{X}_{res|\mathbf{W}}$ or (2) regress **y** on $\mathbf{X}_{res|\mathbf{W}}$. The same partial regression coefficients were obtained in both cases. Between calculation methods, the vectors of fitted values only differed by the value of the intercept of the regression of **y** on $\mathbf{X}_{res|\mathbf{W}}$, which was also the mean of **y**. The R^2 of the first analysis was the partial R^2, whereas that of the second analysis was the semipartial R^2; their square roots were the partial and semipartial correlation coefficients described in Box 4.1.

The same two approaches can be used for partial RDA, which is the extension of partial linear regression to a multivariate response matrix **Y**. First, one computes the residuals of **Y** on **W** (noted $\mathbf{Y}_{res|\mathbf{W}}$) and the residuals of **X** on **W** ($\mathbf{X}_{res|\mathbf{W}}$). Then, one can compute either (1) the RDA of $\mathbf{Y}_{res|\mathbf{W}}$ by $\mathbf{X}_{res|\mathbf{W}}$ or (2) the RDA of **Y** by $\mathbf{X}_{res|\mathbf{W}}$. The two approaches produce the same canonical eigenvalues, eigenvectors and axes. In both approaches, the significance of the canonical axes can be tested using the forward and marginal methods described in Subsection 11.1.2 (paragraph 4). In partial RDA, the canonical axes (matrix **Z**) are linear combinations of the residuals of the explanatory variables **X**, $\mathbf{X}_{res|\mathbf{W}}$, and are orthogonal to the covariables in **W**. The R^2 obtained in the first approach is the partial canonical R^2, whereas that of the second analysis is the semipartial canonical R^2; these two statistics are described in the next subsection. In computer programs, it is customary to use as matrix **F** (eq. 11.17) the matrix obtained from the RDA of $\mathbf{Y}_{res|\mathbf{W}}$ by $\mathbf{X}_{res|\mathbf{W}}$, not the matrix computed in the RDA of **Y** by $\mathbf{X}_{res|\mathbf{W}}$.

Table 11.5 Algorithm for partial RDA in the R language. This is the skeleton of the algorithm used in the ***rda()*** function of the VEGAN package. (Jari Oksanen, personal communication.)

```
pRDA <- function(Y, X = NULL, W = NULL, scale.Y = FALSE)
{
    Y <- scale(as.matrix(Y), center = TRUE, scale = scale.Y)

    if (!is.null(W)) {                  # If covariables W are present
        W <- scale(as.matrix(W), center = TRUE, scale = FALSE)
        Y <- qr.resid(qr(W), Y)
    }
    if (!is.null(X)) {                  # If there are explanatory variables X
        X <- scale(as.matrix(X), center = TRUE, scale = FALSE)
        X <- cbind(X, W)
        Q <- qr(X)
        RDA <- svd(qr.fitted(Q, Y))
        RDA$w <- Y %*% RDA$v %*% diag(1/RDA$d)
        Y <- qr.resid(Q, Y)
    } else {                            # No explanatory variables X nor covariables W
        RDA <- NULL
    }
    RES <- svd(Y)                       # PCA of the residuals
    list(RDA = RDA, RES = RES)
}
```

Table 11.5 presents a very short algorithm for partial RDA, designed by Prof. Jari Oksanen (University of Oulu, Finland). This algorithm handles different cases. (1) If there are covariables (**W**) in the analysis, **Y** is regressed on **W** and residuals $\mathbf{Y}_{\text{res}|\mathbf{W}}$ are computed using QR decomposition (function ***qr()*** in R), which is faster than multivariate regression by matrix inversion (eqs. 10.16 and 11.11). (2) If there are explanatory variables (**X**), RDA is the eigen-decomposition (by SVD through function ***svd()*** in R, Section 2.11) of the fitted values of the multivariate regression of **Y** on **X**. If **X** and **W** are both present, regressing $\mathbf{Y}_{\text{res}|\mathbf{W}}$ on the column concatenation of **X** and **W** produces the same result as a partial regression of $\mathbf{Y}_{\text{res}|\mathbf{W}}$ on $\mathbf{X}_{\text{res}|\mathbf{W}}$ because $\mathbf{Y}_{\text{res}|\mathbf{W}}$ is orthogonal to **W**. (3) A PCA of the residuals is computed. (4) If there are neither explanatory variables **X** nor covariables **W** in the analysis, the result only contains a PCA of **Y** and no RDA is computed.

7 — Statistics in partial RDA

Partial *F*-statistic

For analysis in the presence of **W** containing q covariables (partial RDA), the partial F-statistic is constructed as follows (ter Braak & Smilauer, 2002):

$$F = \frac{\mathrm{SS}(\mathbf{Y}_{\mathrm{fit}}) / m}{\mathrm{SS}(\mathbf{Y}_{\mathrm{res}}) / (n - m - q - 1)} \qquad \textbf{(11.22)}$$

There are several ways of computing the sum of squares of the fitted values $\mathrm{SS}(\mathbf{Y}_{\mathrm{fit}})$ and residuals $\mathrm{SS}(\mathbf{Y}_{\mathrm{res}})$ in the partial RDA case. The most convenient are the following:

$$\mathrm{SS}(\mathbf{Y}_{\mathrm{fit}}) = \mathrm{SS}(\mathbf{Y}_{\mathrm{fit}|(\mathbf{X}+\mathbf{W})}) - \mathrm{SS}(\mathbf{Y}_{\mathrm{fit}|\mathbf{W}})$$

and

$$\mathrm{SS}(\mathbf{Y}_{\mathrm{res}}) = \mathrm{SS}(\mathbf{Y}) - \mathrm{SS}(\mathbf{Y}_{\mathrm{fit}|(\mathbf{X}+\mathbf{W})})$$

where (**X**+**W**) designates the concatenation of **X** and **W** in a single matrix; this is obtained by the operation ***cbind(*****X**,**W***)* in the R language. $\mathbf{Y}_{\mathrm{fit}}$ was noted $\hat{\mathbf{Y}}$ in eq. 11.3 which did not involve covariables **W**.

Semipartial R^2

The semipartial R^2, $R^2_{\mathbf{Y}|\mathbf{X}_{\mathrm{res}|\mathbf{W}}}$, is the proportion of explained variation with respect to the total variation in **Y**. This is the most widely used R^2 statistic in partial RDA because the denominator, which is the total variation in **Y**, forms a common basis for comparisons among analyses using different explanatory matrices **X** and different matrices of covariables **W**. It is the R^2 of the simple RDA of **Y** by $\mathbf{X}_{\mathrm{res}|\mathbf{W}}$:

$$R^2_{\mathbf{Y}|\mathbf{X}_{\mathrm{res}|\mathbf{W}}} = \frac{\mathrm{SS}(\mathbf{Y}_{\mathrm{fit}})}{\mathrm{SS}(\mathbf{Y})} \qquad \textbf{(11.23)}$$

Partial R^2

The partial R^2, $R^2_{\mathbf{Y}_{\mathrm{res}|\mathbf{W}}|\mathbf{X}_{\mathrm{res}|\mathbf{W}}}$, is the proportion of explained variation with respect to the total variation in $\mathbf{Y}$ residualized on the matrix of covariables **W**. Although more rarely used than the semipartial R^2, it is computed as the R^2 of the simple RDA of $\mathbf{Y}_{\mathrm{res}|\mathbf{W}}$ by $\mathbf{X}_{\mathrm{res}|\mathbf{W}}$:

$$R^2_{\mathbf{Y}_{\mathrm{res}|\mathbf{W}}|\mathbf{X}_{\mathrm{res}|\mathbf{W}}} = \frac{\mathrm{SS}(\mathbf{Y}_{\mathrm{fit}})}{\mathrm{SS}(\mathbf{Y}_{\mathrm{res}|\mathbf{W}})} \qquad \textbf{(11.24)}$$

8 — Tests of significance in partial RDA

Permutation test

Tests of significance in partial RDA, using the F-statistic described in eq. 11.22, involve either permutation of the raw data, unrestricted permutation of the residuals of the reduced model (a method proposed by Freedman & Lane, 1983), or unrestricted permutation of the residuals of the full model (a method proposed by ter Braak, 1990, 1992). These methods are described in Anderson & Legendre (1999) for multiple linear regression, which is RDA with a single response variable.

Permute raw data

• In permutation of the raw data (method = "direct" in VEGAN's ***permutest.cca()***), the rows of **Y** are permuted at random to produce the matrix of permuted response data **Y***. This permutation method is used in simple RDA. It can also be used in partial RDA when the covariables do not contain outlying values, e.g. when they represent experimental factors (Subsection 11.1.10, point 4).

Permute residuals of reduced model

• In permutation of the residuals of the reduced model (method = "reduced" in VEGAN's ***permutest.cca()***), one computes the matrix of fitted values $\mathbf{Fit_{Y|W}}$ and the matrix of residuals $\mathbf{Res_{Y|W}}$ of the multivariate regression of **Y** on the matrix of covariables **W**. The rows of $\mathbf{Res_{Y|W}}$ are permuted, producing matrix $\mathbf{Res^*_{Y|W}}$. The matrix of permuted response data, **Y***, is obtained by adding $\mathbf{Fit_{Y|W}}$ (unpermuted) to $\mathbf{Res^*_{Y|W}}$.

Permute residuals of full model

• In permutation of the residuals of the full model (method = "full" in VEGAN's ***permutest.cca()***), one computes the matrix of fitted values $\mathbf{Fit_{Y|XW}}$ and the matrix of residuals $\mathbf{Res_{Y|XW}}$ of the multivariate regression of **Y** on the matrix obtained by concatenation of **X** and **W** by columns into a single matrix. The rows of $\mathbf{Res_{Y|XW}}$ are permuted, producing matrix $\mathbf{Res^*_{Y|XW}}$. The matrix of permuted response data, **Y***, is obtained by adding $\mathbf{Fit_{Y|XW}}$ (unpermuted) to $\mathbf{Res^*_{Y|XW}}$.

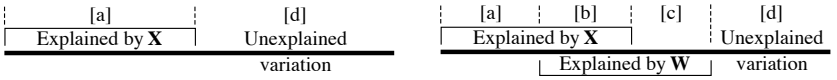
Permutation of the residuals of the reduced and full models were found by Anderson and Legendre (1999) to produce equivalent results. Permutation of the raw data should not be used in partial RDA when the covariables contain outliers. It can, however, be used when partial RDA is used as a form of 2-way MANOVA (Subsection 11.1.10, point 4): in tests of individual factors or the interaction, matrix **W** contains variables coding for the factors or the interaction, and these variables do not have outlier values.

Restricted permutation

Besides these methods, one can also permute the rows of **Y** in a way imposed by the logic of the problem at hand. The most important methods of restricted permutation are: permutation within the levels of a factor or block which is used as a covariable in the study, loop permutation along a time series, and toroidal permutation of the points on a geographic surface (Lotwick & Silverman, 1982).

Methods of permutation of raw data or residuals are compared in Table 11.6 in terms of the permuted portions of variation, in the presence or absence of covariables **W**. *Without covariables*, permutation of raw data involves fraction [a + d] of variation partitioning (Subsection 10.3.5) whereas permutation of residuals of the full model involves [d]. No residual can be computed under a reduced model in the absence of covariables; the method becomes a permutation of raw data. *With covariables*, permutation of residuals may only involve the residuals of the reduced model of the covariables (fraction [a + d]), or the residuals of the full model of the explanatory variables and covariables (fraction [d]). Permutation of the raw data may result in unstable (often inflated) type I error when the covariable contains outliers. This does not occur, however, when using restricted permutations of raw data within groups of a qualitative covariable, which produces an exact test.

Table 11.6 Tests of statistical significance in canonical analysis. Comparison of the methods of permutation of raw data or residuals in terms of the permuted fractions of variation, in the presence or absence of a matrix of covariables **W**. Fractions of variation are noted as in Fig. 10.10: [a] is the variation of matrix **Y** explained by **X** alone, [c] the variation explained by **W** alone, [b] the variation explained jointly by **X** and **W**, and [d] the residual variation.



	Without covariables	With matrix **W** of covariables
Permute raw data	Permute [a + d]	Permute [a + b + c + d]
Permute residuals:		
• reduced model	Equivalent to permuting raw data	Permute [a + d]
• full model	Permute [d]	Permute [d]

9 — Numerical example, partial RDA

Partial RDA provides an answer to the question: what is the partial contribution of one set of explanatory variables when controlling for the effect of another set?

Example 1. — Consider the data in Table 11.3. In that data table, the species can be analysed with respect to *substrate types* while controlling for the effect of *depth*, which is correlated with substrate types. The semipartial R^2 of the analysis is 0.73271; the partial effect of substrate types is highly significant (p = 0.001 after 999 random permutations of the residuals of the reduced model). The two canonical axes produce the triplot shown in Fig. 11.5a.

Example 2. — The converse analysis of the partial effect of *depth* on the distributions of species across the sites while controlling for *substrate types* is also interesting. The semipartial R^2 of this analysis is 0.08274. This is a much weaker effect than that of substrate types, but the partial effect of depth remains significant (p = 0.002 after 999 random permutations of the residuals of the reduced model). A single canonical axis (abscissa of the triplot, Fig. 11.5b) is produced, with the explanatory variable *depth* pointing to the right. Since there is no second canonical axis available, the first axis of the PCA of the residual variation is used as the ordinate of the diagram. This axis separates the coral sites 5, 7 and 9 from the other sites.

These two effects will be considered jointly within the framework of variation partitioning in Subsection 11.1.11 below.

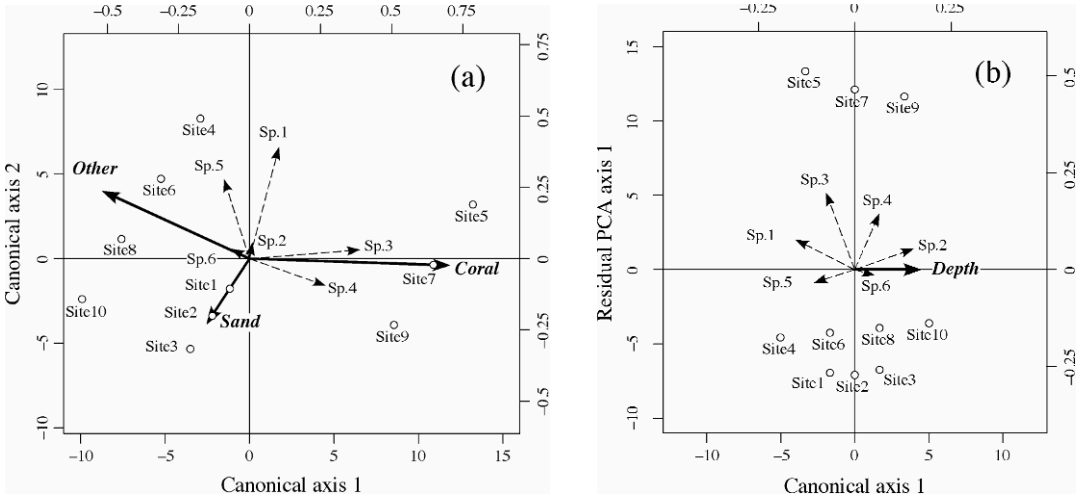


Figure 11.5 Partial RDA triplots of the data in Table 11.3. (a) The explanatory matrix **X** is substrate types, the covariable **W** is depth; (b) the explanatory matrix **X** is depth, the covariable **W** is substrate types. The sites are represented by open circles, the species by dashed arrows, and the explanatory variables in **X** by bold arrows.

10 — Some applications of partial RDA

Partial canonical analysis can be used to investigate a variety of problems. Here are some examples. In most of these applications, CCA (Section 11.2) can be used instead of RDA when **Y** contains frequency data and one wants the analysis to preserve the chi-square distance instead of the Euclidean distance.

Control for effect of **W**

1. *Control for well-known linear effects*. — Consider the case where **W** contains variables whose effects on **Y** are well understood. One wants to control for these well-known effects when analysing the effect of a set of variables **X** on **Y**. For example, one may want to control for the well-known co-variation between phytoplankton assemblages and salinity in a river estuary when analysing the linear effect of nutrient concentrations on phytoplankton. Partial RDA should be used in that case.

Partial effect of a variable

2. *Isolate the effect of a single explanatory variable or factor.* — After conducting a standard RDA as in Subsections 11.1.4, one may want to isolate the partial effect of a single explanatory variable, as in the two examples presented in Subsection 11.1.9 (example 1: a factor with 3 levels; example 2: a quantitative variable). Using all the other explanatory variables as covariates produces a single canonical axis that represents the partial effect of a single quantitative explanatory variable on **Y**. The

corresponding canonical eigenvalue divided by the total variance of **Y** quantifies the partial fraction of the variation of **Y** that is accounted for by that variable (semipartial R^2). The effect of a factor with more than two levels can be isolated by the same method, but then more than one canonical axis are produced because a factor with k levels produces $(k - 1)$ canonical axes.

Related samples

3. *Analysis of related samples*. — Ecological sampling often results in related samples (Box 1.1), where each observation at a site shares some properties with observations at other sites. This is the case, for instance, when sampling different lakes at several depths, the same in all lakes, to study the variation in zooplankton composition. A large portion of the variation in community composition may be associated with the different depths, possibly more than among lakes. Partial RDA offers a way to take this source of variation into account in the analysis of the species-environment relationships. Depth can be coded as a factor or a series of dummy variables, or else as Helmert or polynomial contrasts (Subsection 1.5.7). (Ecologists usually do not hypothesize that zooplankton composition is linearly related to depth, so the covariable depth structuring the sample should be treated as a multi-level factor instead of a quantitative variable.) Including the coding variables in the analysis as a matrix of covariables **W** will effectively control for the effect of the structuring variable. The semipartial R^2 will correctly estimate the partial effect of the environmental variables included in the analysis while controlling for the effect of the structuring variables. Carrying out the analysis for one factor (lakes in this example) while controlling for the effect of the other (depths), and then the opposite, is a form of two-way analysis-of-variance without replication.

Related samples are also obtained when sampling a single lake at different dates and at several depths, or a set of lakes at different dates, the same for all lakes. One may wish to control for the effect of the sampling dates in an analysis of the effect of depths, or lakes, on species composition, bacterial production, or other response variables of interest. As in the previous paragraph, this can be done by using the variable(s) describing the sampling dates as covariable(s) in the analysis. Dates may be represented by dummy variables, or by a quantitative variable whose effect on **Y** is assumed to be linear, or by a sine transformation of the "day of year" (also called "ordinal date", and often "Julian date"[*]), etc. The analysis will effectively control for the effect of dates (days, weeks, years, …) if they only affect the means of the response variables and nothing else. If there is an interaction between sampling dates and the other environmental or spatial variables included in the analysis, the effect of dates cannot be controlled through this simple approach. In the presence of an interaction, the interaction terms must remain in the analysis for the model to be valid (see

[*] The "day of year", also called "ordinal date", is a calendar date starting on 1st January and ranging between 001 and 366. The "Julian day" is used in the "Julian date" system of time measurement, mostly by the astronomy community, where the interval of time is stated in days and fractions of a day since 1st January 4713 BC Greenwich noon. The use of "Julian date" to refer to the day of year, although technically incorrect, is widespread in ecology and other natural sciences. Readers may check the entries "Julian day" and "ordinal date" on Wikipedia.

paragraph 4 below). How to test the space-time interaction in the absence of replication is described in Subsection 14.5.1.

In the same way, one can control for the effect of the sampling locations. Sampling locations may be represented by dummy variables, or by a trend-surface polynomial (Chapter 13) or a set of spatial eigenfunctions (Chapter 14) derived from the geographic coordinates of the sites. The caveat of the previous paragraph concerning interactions applies here as well.

MANOVA by RDA

4. *MANOVA by RDA*. — Partial canonical analysis may be used, instead of MANOVA, to analyse a multivariate response data matrix **Y** in cross-factor experimental designs, including tests of significance for the main effects and the interaction term. For a single experimental factor, the analysis can be conducted using simple RDA or CCA. For two or more factors and their interactions, partial RDA or CCA must be used.

In MANOVA by RDA involving two or more crossed factors, the factors and their interactions are coded by Helmert contrasts (Subsection 1.5.7). The interaction between factors A and B, for example, is represented by a series of variables obtained by the Hadamard product of each Helmert variables coding for factor A by each Helmert variable coding for factor B. Three partial RDAs are necessary to conduct an analysis involving two crossed factors:

- Test the interaction through a RDA of **Y** with the interaction variables in the explanatory matrix **X** and the Helmert variables coding for factors A and B together in the matrix of covariables **W**. If the interaction is significant, analyse separately the effect of factor A in each class of factor B, and conversely the effect of factor B in each class of factor A, because a significant interaction indicates that the effects of factor A on **Y** depend on the levels of factor B, and conversely. If the interaction is not significant, proceed with the next two steps.

- Test the effect of factor A on **Y** through a RDA of **Y** with the variables coding for A in **X** in the presence of a matrix of covariables **W** containing all variables coding for B and the interaction.

- Test the effect of factor B on **Y** through a RDA of **Y** with the variables coding for B in **X** in the presence of a matrix of covariables **W** containing all variables coding for A and the interaction.

The results of the three tests of significance can be assembled in a MANOVA table.

The condition of homogeneity of the variance-covariance matrices applies to this form of MANOVA, as it does to regular MANOVA. It can be tested by the function ***betadisper()*** of the VEGAN package, which implements the testing method described by Anderson (2006). A fully worked out example of MANOVA by RDA, including a test of homogeneity of the multivariate dispersion, is given in Section 6.3.2.8 of Borcard *et al*. (2011).

As stated in Subsection 11.1.5, when **Y** is a matrix of species presence-absence or abundance data, one can either transform **Y** prior to MANOVA by RDA using the transformations described in Section 7.7 (transformation-based RDA, tb-RDA) to force the partial RDA to preserve the distance that is implicit in the transformation, or use partial CCA to preserve the chi-square distance among sites. Else, one can use the distance-based RDA method (db-RDA, Subsection 11.1.5) to preserve some other distance function appropriate for community data.

Principal response curves

5. *Principal response curves (PRC).* — *Principal response curves* is another form of MANOVA; it was developed by van den Brink and co-authors (1998, 1999, 2003, 2009) to analyse the results of experiments conducted over time, that involved multivariate response data (e.g. community composition data). PRC is a special case of RDA with a single factor for treatments and a single factor representing the time series of repeated observations. The method studies the changes in the multivariate (e.g. species) response variables associated with the treatments over time. In this type of analysis, one is interested in displaying the values of the coefficients (contrasts against the control level) computed for the first RDA axis representing the effects of treatment along time. Significance of the canonical relationship and of the first axis can be tested when there is replication in the experimental design. This is an omnibus test: H_0 corresponds to 'no treatment effect'. H_1 includes all functional forms that the treatment effects can take, i.e. main effect and/or interaction. No effect at all produces coinciding treatment lines in the plot. One can also test separately the effect of the main factor (treatment) and, when there is replication, the treatment-time interaction. A significant interaction indicates that the treatment levels had different effects on the response data at different times; it is displayed as non-parallel or crossing treatment lines in the plot.

Partial PCA

6. *Partial PCA.* — Partial principal component analysis is the PCA of a response data table **Y** residualized on a set of explanatory variables. This method allows researchers to examine the multivariate structure of the data after removing the effect of the **X** variables on **Y**, which may already be well understood, by computing residuals. Note that the results of a partial PCA differ from those of a partial RDA.

The three tables represented at the bottom of Fig. 11.2 illustrate how partial PCA is carried out: the residuals of **Y** by **X** are computed, followed by a PCA of the matrix of residuals. Alternatively, since RDA of **Y** by **Y** is a PCA of **Y**, as shown in Subsection 11.1.3, partial PCA can be obtained by computing a partial RDA of **Y** by **Y** with **X** as covariables. In R, the regression function ***lm()*** can be used to easily obtain a matrix of residuals: res = ***residuals(lm(*****Y** ~ **X)).** With the data of Table 11.3 for instance, one could examine the residual structure, after controlling for depth and substrate, by plotting a PCA biplot of the non-canonical axes shown in Table 11.4. In spatial analysis, one could detrend the data by computing the regression residuals of **Y** on the geographic coordinates of the sites before computing a PCA.

Selection of explanatory variables

7. *Selection of explanatory variables.* — Different selection methods are available in canonical analysis, as well as in multiple regression (Subsection 10.3.3): backward,

forward, and stepwise. Function ***ordistep()*** in VEGAN offers all three methods of selection. In *forward selection*, the significance of the partial *F*-statistics associated with all candidate variables is tested using permutations, and the explanatory variable that has the most significant partial effect is selected if its p-value satisfies the "p-to-enter" significance level; in case of equality, the variable that has the lowest value of the Akaike Information Criterion (*AIC*, eq. 10.22)* is selected for inclusion in the model. The *backward* option sequentially drops variables from the model using the same criteria of significance (the highest p-value is compared to a "p-to-exclude" significance level) and *AIC* in case of equality (the variable whose removal produces the model with the lowest *AIC* value is excluded). The *stepwise* option tries to eliminate variables from the model (*backward*) after each *forward* step. In this function, "best" refers to the most significant variable.

Functions ***ordiR2step()*** of VEGAN and ***forward.sel()*** of PACKFOR offer the forward method. In these functions, the basic algorithm, developed by ter Braak (1990), is the same as in CANOCO: considering the variables already selected, the explanatory variable with the highest partial R^2 is selected if the additional contribution of that variable is significant (permutation test) at a pre-selected significance level. In these functions, "best" refers to the variable that explains the largest portion of the remaining unexplained variance of **Y**. These two functions offer the option of applying a second stopping criterion proposed by Blanchet *et al*. (2008b): the selection stops either when the tested variable has a p-value higher than the pre-selected significance level or when the adjusted R^2 of the full model, before any selection, is exceeded.

Before applying these variable selection methods, one should look at the collinearity among the variables in **X** by computing variance inflation factors (VIF, eq. 10.17), and remove variables as needed to reduce collinearity. Borcard *et al*. (2011) present examples of forward selection prior to RDA.

11 — Variation partitioning by RDA

Variation partitioning, described in Subsection 10.3.5 for univariate response data, was originally developed for the analysis of multivariate response tables (Borcard *et al*., 1992; Borcard & Legendre, 1994). It is especially useful for partitioning the variation of community composition data with respect to two or more sets of explanatory variables.

Ecological application 11.1a

The method is illustrated here using fish assemblage data (27 species) from 29 sampling sites along the Doubs River in eastern France. The calculations reported in this application were done

* The *AIC* criterion is not meant to identify the "true" model (which is only known in simulation studies) among several alternative models, but to find the best predictive model for new observations.

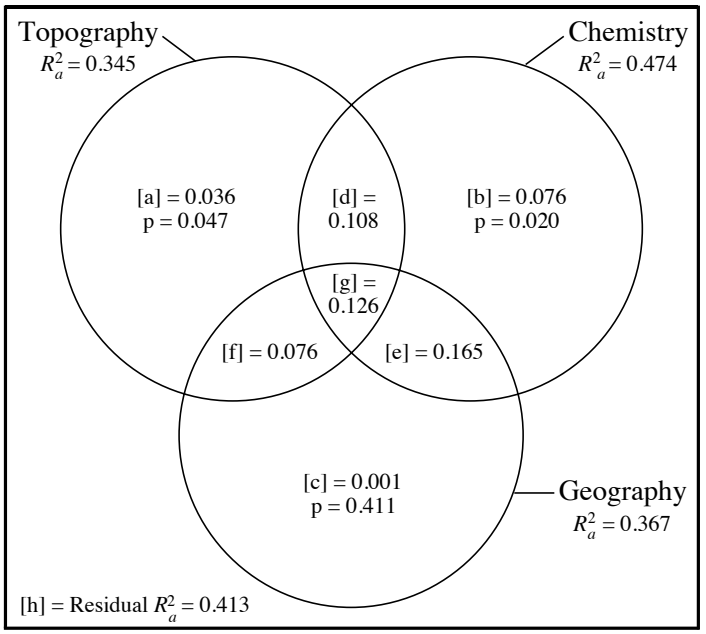


Figure 11.6 Venn diagram illustrating the results of variation partitioning of the Doubs River fish assemblage data (29 sites) among three sets of explanatory variables: Topography, Chemistry and Geography. The fractions of variation are identified by letters [a] to [h]. The value next to each identifier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

by RDA, whereas they involved multiple regression in Subsection 10.3.5. The partitioning example reported here uses the species and environmental data collected by Verneaux (1973), which are available in the R package ADE4. The data were reanalysed for variation partitioning by Borcard *et al*. (2011, Section 6.3.2.7); here as in that book, site 8, where no fish were caught, was removed from the original 30-site species and environment data tables.

Partitioning involved three data sets: Topography (variables: altitude, slope, water flow), Chemistry (variables: pH, hardness, concentrations of phosphate, nitrate, ammonia, dissolved O_2, and biological oxygen demand), and Geography (variable: linear distance from the source along the course of the river). The partitioning results, obtained from function ***varpart()*** of the VEGAN package, are illustrated by a Venn diagram (Fig. 11.6); the decomposition into fractions [a] to [h] was done from the adjusted R^2 values (R^2_a) calculated by RDAs involving 1, 2, and all 3 explanatory data tables, as in Subsection 10.3.5. The first finding was that each of the three data sets explained approximately the same fraction of the spatial variation in the fish assemblage along the river (R^2_a = 0.345, 0.474, and 0.367, respectively). A great deal of the variation was shared among two or all three sets of explanatory variables. There was a small but significant portion of the fish variation explained by Topography that was not shared with the

two other data sets (fraction [a]: R^2_a = 0.036, p = 0.047), and likewise for Chemistry (fraction [b]: R^2_a = 0.076, p = 0.020). However, all the fish variation among sites explained by Geography was also explained by one of the other two explanatory data tables, or by both, leaving no significant portion of variation explained solely by Geography (R^2_a = 0.001, p = 0.411). Whereas the three explanatory data sets explained jointly 58.7% of the species variation, it was the Topography and Chemistry data that were the most informative and complementary, adding to the model portions of variation that were not explained by Geography alone. Results of a partitioning involving soil mite assemblages by four explanatory data sets are presented in Borcard *et al*. (2011, Section 7.4.2.5).

In Chapter 14, which describes multiscale spatial analysis, variation partitioning is used to partition the variation of data **Y** between two components, environmental (**X**) and spatial (**W**). Two ecological applications (14.1a and 14.1b) involving variation partitioning by partial canonical analysis are presented.

Ecological application 11.1b

In a classical study of spider community ecology, Aart & Smeenk-Enserink (1975) used canonical correlation analysis (CCorA, Section 11.4) to analyse a portion of the hunting spider data collected in pitfall traps at 100 sites in the Bierlap dune valley of the Netherlands. The paper related the species to environmental descriptors obtained at 28 of the 100 sites. The authors used canonical correlation analysis, a symmetric method of canonical analysis that was available in computer packages in the 1970s, to describe the influence of environmental conditions on the spider assemblages; their objective was to test the hypothesis of an asymmetric relationship between species and environmental conditions. The present example will show original results that we obtained by RDA, which is a more appropriate method to study and test asymmetric relationships. An additional advantage is that RDA can be carried out on unstandardized species data, thus preserving the original variances of the individual species in the analysis (Subsection 11.1.5), whereas the species data are always standardized in CCorA (Subsection 11.4.1). The Aart & Smeenk-Enserink spider data have been reanalysed, after recoding, by ter Braak (1986)* using CCA. The same data (recoded by ter Braak, 1986) were also analysed by De'ath (2002) using multivariate regression tree analysis (MRT, Ecological application 8.11).

At the 28 sites included in the canonical correlation analysis of Aart & Smeenk-Enserink (1975), the community composition data were the abundances of 12 hunting spider species normalized by logarithmic transformation, log(y + 1). Among the 27 environmental descriptors characterizing the light, vegetation, and soil that had been observed, only the 15 that were linearly correlated with the species variables were used by these authors for their canonical correlation analysis in order to ensure linearity of the relationships between the two sets of descriptors.†

* Warning to users: the 28 sites for which environmental data are provided in the Aart & Smeenk-Enserink (1975) paper are presented in a different order in Table 3 of ter Braak (1986).

† The spider (28 sites, 12 species) and environmental data (28 sites, 15 variables) used in this Application are available on the Web page http://numericalecology.com/data.

For the present application, the 15 environmental variables selected by Aart & Smeenk-Enserink (1975) were used as matrix **X** to insure comparability of the present results with theirs. The adjusted R^2 (R^2_a) of the RDA provided a criterion to select the best transformation for the species data: after computing RDA of the spider data, transformed in different ways, with the 15 environmental variables, R^2_a was higher for the log-transformed species data than for any of the other transformations of Section 7.7; so the log-transformed data were used in the RDA. Forward selection (with the stopping criterion $p \leq 0.05$) was carried out among the 15 environmental variables (Subsection 11.1.10, point 7). A parsimonious model containing six environmental variables was selected, which provided the same explanation as the full set of explanatory variables: $R^2_a = 0.761$ for the full set of 15 environmental variables, $R^2_a = 0.768$ for the subset of six variables, i.e. water content of the soil, illuminance under cloudless sky, ground cover by leaves and twigs, cover by the herb layer, cover by *Calamagrostis epigejos* (a grass, family Poaceae), and cover by the tree layer.

A search for species associations was carried out using concordance analysis, described in Subsection 8.9.2. The first statistically significant association comprised three species: *Alopecosa accentuata*, *Alopecosa fabrilis* and *Arctosa perita*; a fourth species, *Pardosa monticola*, was weakly associated with this group. The second significant association contained seven species: *Alopecosa cuneata*, *Arctosa lutetiana*, *Aulonia albimana*, *Pardosa nigriceps*, *Pardosa pullata*, *Trochosa terricola* and *Zora spinimana*. The species *Pardosa lugubris* formed a single-species group.

The RDA triplot (Fig. 11.7) shows the relationships between the species and the environmental variables. The species belonging to association 1 (upper ellipse) were found in greater abundances at very dry and more intensely lit sites. Those belonging to association 2 (right ellipse) were found at sites with higher soil humidity and higher cover by herbs and by *Calamagrostis epigejos*. The single-species association *Pardosa lugubris* exhibited preference for shaded sites with higher soil humidity and higher cover by trees and by leaves and twigs.

The total species variation, which is a measure of beta diversity (Section 6.5), was partitioned between the physical (soil, light) and vegetation influences using variation partitioning (Fig. 11.8). Fractions [a] and [c] were both statistically significant (tested by partial RDA), but fraction [c], which depicted the fraction of beta diversity explained exclusively by vegetation ($R^2_a = 0.43$), was much larger than fraction [a], which corresponded to the variation explained only by the physical environment ($R^2_a = 0.07$). Most of the explanation ([b] = 0.27) provided by the physical variables was shared with the vegetation variables.

11.2 Canonical correspondence analysis (CCA)

Canonical correspondence analysis is a canonical asymmetric ordination method developed by ter Braak (1986, 1987a, 1987c). First implemented in the program CANOCO (ter Braak, 1988b, 1988c, 1990; ter Braak & Smilauer, 1998), it is now available in several computer packages and R functions for community ecology. It is the canonical form of correspondence analysis. Any data table that could be subjected to correspondence analysis (CA, Section 9.2) is a suitable *response matrix* **Y** for CCA; this is the case, in particular, for species presence-absence or abundance data (Subsection 9.2.4).

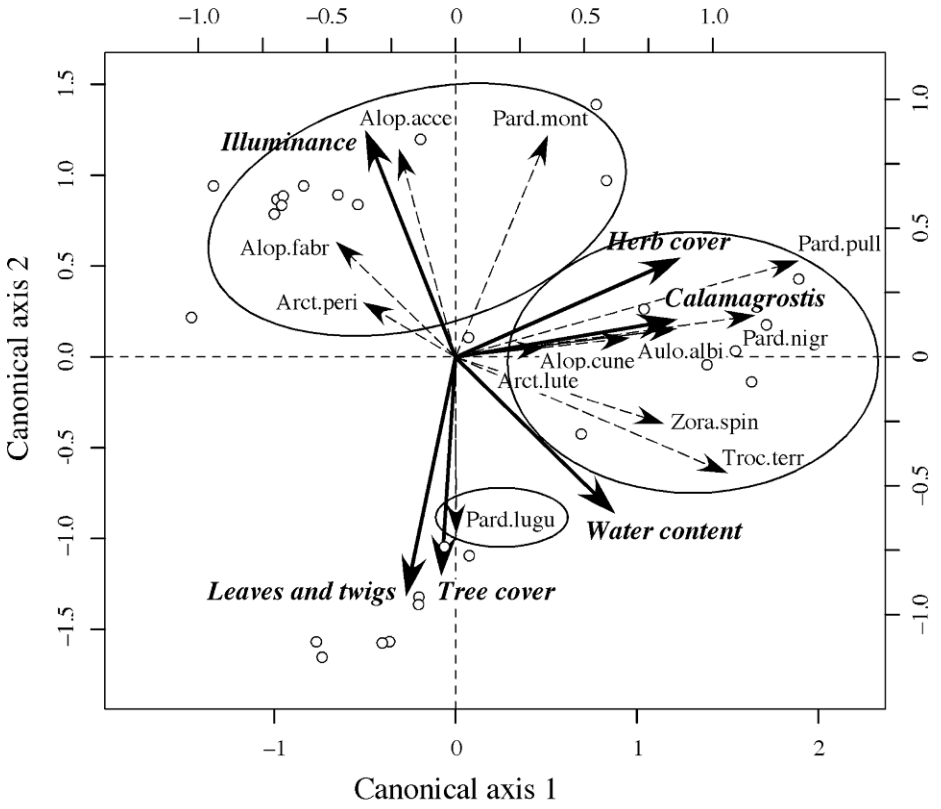


Figure 11.7 RDA triplot relating the spider species (dashed arrows) to the selected environmental variables (full-line arrows). Scaling type 2 was used to emphasize the covariances among the species. Small open circles represent the 28 sites; site names were not printed to keep the diagram simple. The 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).

1 — The algebra of canonical correspondence analysis

The mathematics of CCA is derived from that of RDA. The first difference is that matrix $\overline{\mathbf{Q}}$ is used instead of **Y** as the response matrix in the calculations, as it was the case in correspondence analysis (Section 9.2). The second difference is that a diagonal matrix of row weights, $\mathbf{D}(p_{i+})$, is used in the regression portion of the calculation. For

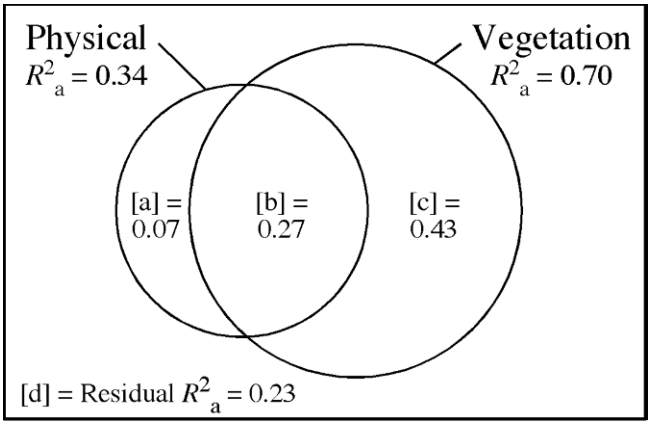


Figure 11.8 Venn diagram partitioning the total spider species variation (rectangle) between physical (water content of the soil, illuminance under cloudless sky) and vegetation influences (ground cover by leaves and twigs, cover by the herb layer, cover by *Calamagrostis epigejos*, and cover by the tree layer). The fraction identifiers [a], [b] and [c], are as in Fig. 10.10. The fractions are expressed as R^2_a, as in Fig. 11.6. Circle sizes are approximate.

each row of **Y**, f_{i+} is the sum of the values in row i, and p_{i+} is f_{i+} divided by the grand total, f_{++}, of the frequencies in **Y**.

Inflated data matrix

To obtain a CCA, the regression portion of the calculation is modified, in eq. 11.25 (below), in such a way as to emulate a RDA carried out on *inflated data matrices* $\mathbf{Y}_{infl}$ and $\mathbf{X}_{infl}$ constructed as follows. **Y** ($n \times p$) contains frequency data, such as species presences or abundances of p species observed at n sites, and **X** ($n \times m$) contains explanatory, e.g. environmental, variables. The presence of *a single individual* in **Y** produces a new row in $\mathbf{Y}_{infl}$, so that there are as many rows in $\mathbf{Y}_{infl}$ as there are individual organisms, or presences, in **Y**. The number of rows of $\mathbf{Y}_{infl}$ is thus f_{++}. $\mathbf{Y}_{infl}$ still has p columns for the p species, but a single individual is present in each row. In $\mathbf{X}_{infl}$, the row vectors of explanatory data are duplicated as many times as needed to make every individual organism (i.e. every species presence) in $\mathbf{Y}_{infl}$ face, in $\mathbf{X}_{infl}$, a copy of the appropriate vector of explanatory data. Compute $\overline{\mathbf{Q}}_{infl}$ from $\mathbf{Y}_{infl}$ using eq. 9.24. CCA is the RDA of $\overline{\mathbf{Q}}_{infl}$ by $\mathbf{X}_{infl}$: the eigenvalues* and matrix of eigenvectors are the same.

* The eigenvalues of RDA of $\overline{\mathbf{Q}}_{infl}$ by $\mathbf{X}_{infl}$ computed on the covariance matrix of $\hat{\mathbf{Y}}$, instead of the cross-product matrix, are smaller than those of CCA by a multiplicative factor ($f_{++} - 1$).

In computer programs, it is possible to use matrices $\overline{\mathbf{Q}}$ and $\mathbf{X}$ for the calculations instead of $\overline{\mathbf{Q}}_{infl}$ and $\mathbf{X}_{infl}$, which would be cumbersome to compute when $\mathbf{Y}$ has a large sum f_{++}. The modified algorithm is the following:

• The dependent data matrix is not $\mathbf{Y}$ centred by columns as in RDA. In this algorithm, CCA uses matrix $\overline{\mathbf{Q}}$ of the contributions to chi-square, also used in correspondence analysis, as the response matrix. $\overline{\mathbf{Q}}$ is derived from matrix $\mathbf{Y}$ through eq. 9.24.

• Matrix $\mathbf{X}$ is standardized to $\mathbf{X}_{stand}$ using weights $\mathbf{D}(f_{i+})$. To achieve this, the inflated matrix $\mathbf{X}_{infl}$ is constructed as described above; it contains f_{++} rows. Then the mean and standard deviation of each column of $\mathbf{X}_{infl}$ are computed and used to standardize the explanatory variables in $\mathbf{X}$. For the standard deviations (eq. 4.5), the maximum likelihood estimator of the variance is used instead of the usual unbiased estimator (eq. 4.3); in other words, the sum of squared deviations from the mean of the variables in $\mathbf{X}_{infl}$ is divided by the number of rows of that matrix (which is equal to f_{++}), instead of the number of rows minus 1.

• To obtain the regression coefficients, weighted multiple regression is used instead of conventional multiple regression. The row weights, written in diagonal matrix $\mathbf{D}(p_{i+})^{1/2}$ (Subsection 9.2.1), are applied to matrix $\mathbf{X}$ everywhere it occurs in the multivariate regression equation, which becomes:

$$\mathbf{B} = [\mathbf{X}_{stand}' \, \mathbf{D}(p_{i+}) \, \mathbf{X}_{stand}]^{-1} \, \mathbf{X}_{stand}' \, \mathbf{D}(p_{i+})^{1/2} \, \overline{\mathbf{Q}}$$

and

$$\hat{\mathbf{Y}} = \mathbf{D}(p_{i+})^{1/2} \, \mathbf{X}_{stand} \, \mathbf{B}$$

The equation for computing $\hat{\mathbf{Y}}$ is then:

$$\hat{\mathbf{Y}} = \mathbf{D}(p_{i+})^{1/2} \, \mathbf{X}_{stand} \, [\mathbf{X}_{stand}' \, \mathbf{D}(p_{i+}) \, \mathbf{X}_{stand}]^{-1} \, \mathbf{X}_{stand}' \, \mathbf{D}(p_{i+})^{1/2} \, \overline{\mathbf{Q}} \qquad \textbf{(11.25)}$$

The matrix of residuals is computed as $\overline{\mathbf{Q}}_{\mathbf{res}} = \overline{\mathbf{Q}} - \hat{\mathbf{Y}}$. This is the equivalent, for CCA, of equation $\mathbf{Y}_{\mathbf{res}} = \mathbf{Y} - \hat{\mathbf{Y}}$ found in Fig. 11.2 for RDA.

• Eigenvalue decomposition (eqs. 11.15 and 11.16) is carried out on matrix $\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}}$ which, in this case, is simply the matrix of sums of squares and cross products, without division by the number of degrees of freedom — as in correspondence analysis:

$$\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} \qquad \textbf{(11.26)}$$

One can show that $\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}}$ (eq. 11.26) is equal to $\mathbf{S}_{\overline{\mathbf{Q}}\mathbf{X}}\mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}'_{\overline{\mathbf{Q}}\mathbf{X}}$ if the covariance matrices $\mathbf{S}_{\overline{\mathbf{Q}}\mathbf{X}}$ and $\mathbf{S}_{\mathbf{XX}}$ are computed as follows, with weights on $\mathbf{X}$ given by matrix $\mathbf{D}(p_{i+})^{1/2}$:

$$\mathbf{S}_{\overline{\mathbf{Q}}\mathbf{X}} = \overline{\mathbf{Q}}' \, \mathbf{D}(p_{i+})^{1/2} \, \mathbf{X} \quad \text{and} \quad \mathbf{S}_{\mathbf{XX}} = \mathbf{X}' \, \mathbf{D}(p_{i+}) \, \mathbf{X}$$

without division by degrees of freedom.

In the modified algorithm, CCA is the eigen-decomposition of $\mathbf{S}_{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}}$ (eq. 11.26). It produces matrices $\mathbf{\Lambda}$ of eigenvalues and $\mathbf{U}$ of eigenvectors. Canonical correspondence analysis is thus a weighted form of redundancy analysis, applied to response matrix $\overline{\mathbf{Q}}$. The solution approximates the chi-square distances among the rows (objects) of the dependent data matrix, subject to the constraint that the canonical ordination vectors be maximally related to weighted linear combinations of the explanatory variables. The method is well suited to analyse the relationships between species presence/absence or abundance data matrices and tables of environmental variables. The number of canonical and non-canonical axes expected from the analysis are shown in Table 11.1. Tests of significance for the total canonical variation and for individual canonical axes are carried out in the same way in CCA as described for RDA in Subsections 11.1.2 and 11.1.8.

- The normalized matrix $\hat{\mathbf{U}}$ is obtained using eq. 9.30:

$$\hat{\mathbf{U}} = \overline{\mathbf{Q}}\,\mathbf{U}\mathbf{\Lambda}^{-1/2}$$

In CCA, matrix $\hat{\mathbf{U}}$ defined here does not contain the loadings of the rows of $\hat{\mathbf{Y}}$ on the canonical axes. It contains instead the loadings of the rows of $\overline{\mathbf{Q}}$ on the ordination axes, as in CA. It will be used to find the site scores (matrices $\mathbf{F}$ and $\hat{\mathbf{V}}$) in the space of the original variables $\mathbf{Y}$. The site scores in the space of the fitted values $\hat{\mathbf{Y}}$ will be found using $\mathbf{U}$ instead of $\hat{\mathbf{U}}$.

Scalings in CCA

- Matrix $\mathbf{V}$ of species scores (for scaling type 1) and matrix $\hat{\mathbf{V}}$ of site scores (for scaling type 2) are obtained from $\mathbf{U}$ and $\hat{\mathbf{U}}$ using the transformations described for correspondence analysis (Subsection 9.2.1):

eq. 9.33 (species scores, scaling 1): $\mathbf{V} = \mathbf{D}(p_{+j})^{-1/2}\mathbf{U}$

and eq. 9.34 (site scores, scaling 2): $\hat{\mathbf{V}} = \mathbf{D}(p_{i+})^{-1/2}\hat{\mathbf{U}}$

or combining eqs. 9.30 and 9.34: $\hat{\mathbf{V}} = \mathbf{D}(p_{i+})^{-1/2}\overline{\mathbf{Q}}\,\mathbf{U}\mathbf{\Lambda}^{-1/2}$

Scalings 1 and 2 are the same as in correspondence analysis (Subsection 9.2.1). Matrices $\mathbf{F}$ (site scores for scaling type 1) and $\hat{\mathbf{F}}$ (species scores for scaling type 2) are found using eqs. 9.35a and 9.36a:

$$\mathbf{F} = \hat{\mathbf{V}}\mathbf{\Lambda}^{1/2} \quad \text{and} \quad \hat{\mathbf{F}} = \mathbf{V}\mathbf{\Lambda}^{1/2}$$

Equations 9.35b and 9.36b cannot be used here to find $\mathbf{F}$ and $\hat{\mathbf{F}}$ because the eigenanalysis has been conducted on a covariance matrix (eq. 11.26) computed from the matrix of fitted values $\hat{\mathbf{Y}}$ (eq. 11.25) and not from $\mathbf{Q}$ defined in Subsection 9.2.1.

As mentioned in Subsection 9.2.1 about correspondence analysis, scaling type 3, which is called "symmetric scaling" in program CANOCO, is a compromise between

scalings 1 and 2. This scaling does not preserve the chi-square distances among the species or among the site scores. It is obtained by drawing together matrices $\hat{\mathbf{V}}\,\mathbf{\Lambda}^{1/4}$ (or $\mathbf{F}\mathbf{\Lambda}^{-1/4}$) for sites and $\mathbf{V}\mathbf{\Lambda}^{1/4}$ (or $\hat{\mathbf{F}}\,\mathbf{\Lambda}^{-1/4}$) for species.

The site scores that are linear combinations of the environmental variables, corresponding to eq. 11.18 of RDA, are found from $\hat{\mathbf{Y}}$ using the following equations:

For scaling type 1: $$\mathbf{Z}_1 = \mathbf{D}(p_{i+})^{-1/2}\,\hat{\mathbf{Y}}\,\mathbf{U} \qquad \textbf{(11.27)}$$

For scaling type 2: $$\mathbf{Z}_2 = \mathbf{D}(p_{i+})^{-1/2}\,\hat{\mathbf{Y}}\,\mathbf{U}\mathbf{\Lambda}^{-1/2} \qquad \textbf{(11.28)}$$

For scaling type 3: $$\mathbf{Z}_3 = \mathbf{D}(p_{i+})^{-1/2}\,\hat{\mathbf{Y}}\,\mathbf{U}\mathbf{\Lambda}^{-1/4} \qquad \textbf{(11.29)}$$

Before computing the biplot scores, matrix $\mathbf{Z}_1$ (or $\mathbf{Z}_2$ or $\mathbf{Z}_3$: identical results) must be standardized to $\mathbf{Z}_{stand}$ using the procedure described for the standardization of $\mathbf{X}$: generate the inflated matrix $\mathbf{Z}_{1\,infl}$, compute the vectors of column means and standard deviations (in the computation of the variances, divide the sums of squares by f_{++} instead of $(f_{++} - 1)$) for $\mathbf{Z}_{1\,infl}$, and use these vectors to standardize $\mathbf{Z}_1$. Applying this concept, computational shortcuts can be used to obtain matrix $\mathbf{Z}_{stand}$ without actually generating matrix $\mathbf{Z}_{1\,infl}$. The matrices of biplot scores (**BS**) for the explanatory variables can now be computed using $\mathbf{X}_{stand}$, $\mathbf{Z}_{stand}$, and the diagonal matrix of row weights $\mathbf{D}(p_{i+})$:

For scaling type 1: $$\mathbf{BS}_1 = \mathbf{X}_{stand}'\,\mathbf{D}(p_{i+})\,\mathbf{Z}_{stand}\,\mathbf{\Lambda}^{1/2} \qquad \textbf{(11.30)}$$

For scaling type 2: $$\mathbf{BS}_2 = \mathbf{X}_{stand}'\,\mathbf{D}(p_{i+})\,\mathbf{Z}_{stand} \qquad \textbf{(11.31)}$$

For scaling type 3: $$\mathbf{BS}_3 = \mathbf{X}_{stand}'\,\mathbf{D}(p_{i+})\,\mathbf{Z}_{stand}\,\mathbf{\Lambda}^{1/4} \qquad \textbf{(11.32)}$$

For scaling type 1, triplots are drawn using matrix $\mathbf{V}$ for the species, either $\mathbf{Z}_1$ or $\mathbf{F}$ for the sites, and $\mathbf{BS}_1$ for the explanatory variables. For scaling type 2, matrix $\hat{\mathbf{F}}$ is used for the species, either $\mathbf{Z}_2$ or $\hat{\mathbf{V}}$ for the sites, and $\mathbf{BS}_2$ for the explanatory variables. For scaling type 3, matrix $\mathbf{V}\mathbf{\Lambda}^{1/4}$ is used for the species, either $\mathbf{Z}_3$ or $\hat{\mathbf{V}}\,\mathbf{\Lambda}^{1/4}$ for the sites, and $\mathbf{BS}_3$ for the explanatory variables. The construction and interpretation of CCA triplots is discussed in more detail in ter Braak & Verdonschot (1995).

• Residuals can be analysed by applying eigenvalue decomposition (eq. 11.15) to matrix $\bar{\mathbf{Q}}_{\mathbf{res}}$, producing a matrix of eigenvalues $\mathbf{\Lambda}$ and a matrix of eigenvectors $\mathbf{U}$. Matrix $\hat{\mathbf{U}}$ is obtained using eq. 9.30: $\hat{\mathbf{U}} = \bar{\mathbf{Q}}\,\mathbf{U}\mathbf{\Lambda}^{-1/2}$. Species and site scores are obtained for scaling types 1 and 2 (eqs. 9.33, 9.34, 9.35a and 9.36a) using the matrices of row and column sums $\mathbf{D}(p_{i+})^{-1/2}$ and $\mathbf{D}(p_{+j})^{-1/2}$ of the original matrix $\mathbf{Y}$.

CCA can be computed following the algorithm described in the present subsection*. One may also use the iterative algorithm proposed by ter Braak (1986, 1987a) and implemented in the program CANOCO. The latter algorithm, which has historical significance, is described in Table 11.6 of Legendre & Legendre (1998).

Partial CCA Developed by ter Braak (1988a), partial CCA is computed essentially like partial RDA (Subsection 11.1.6), after residualizing $\mathbf{Q}$ and $\mathbf{X}$ on the covariables $\mathbf{W}$. The weights $\mathbf{D}(p_{i+})^{1/2}$ are used in the computation of these residuals.

CCA can be used for variation partitioning (Subsections 10.3.5 and 11.1.11). The difficulty with CCA resides in the calculation of the adjusted R^2, which is necessary to obtain unbiased estimates of the fractions of explained variation. A method to compute the adjusted R^2 in CCA, involving a permutation procedure, was described by Peres-Neto *et al*. (2006). In Supplements to their paper, Peres-Neto *et al*. (2006) provided a MATLAB package and an executable program to conduct variation partitioning in CCA. At the time this paragraph is written, however, that method has not been incorporated into any major package for community ecology, with the consequence that variation partitioning is not yet generally available for CCA.

2 — Numerical example

Table 11.3 will now be used to illustrate the computation and interpretation of CCA. The 9 species were used in matrix **Y**. Matrix **X** comprised the four columns shown in the right-hand portion of Table 11.3. CCA results are presented in Table 11.7 and Fig. 11.9; the CANOCO program and the CCA functions in R* provide more output tables than presented here. There was a possibility of 3 canonical and 8 non-canonical axes. It turned out that the last 2 non-canonical axes had zero variance; they are consequently not displayed. An overall test of significance showed that the canonical relationship between matrices **X** and **Y** was very highly significant ($p = 0.001$ after 999 permutations of residuals under a full model; Subsection 11.1.8). The canonical axes explained 47%, 24% and 10% of the response table's inertia, respectively. They were all significant ($p < 0.05$) and displayed strong row-weighted species-environment correlations ($r = 0.998, 0.940$, and 0.883, respectively).

Scaling type 2 (Subsection 11.2.1) was used, in this example, to emphasize the relationships among species. As a result, the species (matrix $\hat{\mathbf{F}}$) are at the centroids of the sites (matrix $\hat{\mathbf{V}}$) in Fig. 11.9a, and distances among species approximate their chi-square distances. Species 3 and 4 characterize the sites with coral substrate, whereas species 5 and 6 indicate the sites with "other substrate". Species 1 and 2, which occupy an intermediate position between the sites with coral and other substrate, are not well represented in the biplot of canonical axes I and II; axis III is needed to adequately represent the variance of these species. Among the ubiquitous species 7 to 9, two are well represented in the subspace of canonical axes I and II; they fall near the middle of the area encompassing the three types of substrate. The sites are not perfectly ordered along the depth vector; the site ordering along this variable mainly reflects differences in species composition between the shallow sandy sites (1, 2 and 3) and the other sites.

* Function CCA.R was written to demonstrate the CCA algorithm described in this subsection. It produces results identical to those of CANOCO 4.x. The function is available on the Web page http://numericalecology.com/rcode.

Table 11.7 Results of canonical correspondence analysis of the data in Table 11.3 (selected output). Matrix **Y**: species 1 to 9; **X**: depth and 3 substrate classes. Non-canonical axes VIII and IX not shown.

	Canonical axes			Non-canonical axes			
	I	II	III	IV	V	VI	VII
Eigenvalues (their sum is equal to the total inertia in matrix $\overline{\mathbf{Q}}$ of species data = 0.78417)							
	0.36614	0.18689	0.07885	0.08229	0.03513	0.02333	0.00990
Fraction of the total variance in $\overline{\mathbf{Q}}$							
	0.46691	0.23833	0.10055	0.10494	0.04481	0.02975	0.01263
Cumulative fraction of total inertia in $\overline{\mathbf{Q}}$ accounted for by axes 1 to k							
	0.46691	0.70524	0.80579	0.91072	0.95553	0.98527	0.99791
Eigenvectors ("species scores", scaling 2): matrices $\hat{\mathbf{F}}$ for the canonical and non-canonical portions (eq. 9.36a)							
Species 1	–0.11035	–0.28240	–0.20303	0.00192	0.08223	0.08573	–0.01220
Species 2	–0.14136	–0.30350	0.39544	0.14127	0.02689	0.14325	0.04303
Species 3	1.01552	–0.09583	–0.19826	0.10480	–0.13003	0.02441	0.04647
Species 4	1.03621	–0.10962	0.22098	–0.22364	0.24375	–0.02591	–0.05341
Species 5	–1.05372	–0.53718	–0.43808	–0.22348	0.32395	0.12464	–0.11928
Species 6	–0.99856	–0.57396	0.67992	0.38996	–0.29908	0.32845	0.21216
Species 7	–0.25525	0.17817	–0.20413	–0.43340	–0.07071	–0.18817	0.12691
Species 8	–0.14656	0.85736	–0.01525	–0.05276	–0.35448	–0.04168	–0.19901
Species 9	–0.41371	0.70795	0.21570	0.69031	0.14843	–0.33425	–0.00629
Site scores ("sample scores", scaling 2): matrices $\hat{\mathbf{V}}$ for the canonical and the non-canonical portions (eq. 9.34)							
Site 1	–0.71059	3.08167	0.21965	1.24529	1.07293	–0.50625	0.24413
Site 2	–0.58477	3.00669	–0.94745	–2.69965	–2.13682	0.81353	0.47153
Site 3	–0.76274	3.15258	2.13925	3.11628	2.30660	–0.69894	–1.39063
Site 4	–1.11231	–1.07151	–1.87528	–0.66637	1.10154	1.43517	–1.10620
Site 5	0.97912	0.06032	–0.69628	0.61265	–0.98301	0.31567	0.57411
Site 6	–1.04323	–0.45943	–0.63980	–0.28716	0.57393	–1.44981	1.70167
Site 7	0.95449	0.08470	0.13251	0.42143	0.11155	–0.39424	–0.67396
Site 8	–0.94727	0.10837	0.52611	0.00565	–1.26273	–1.06565	–1.46326
Site 9	1.14808	–0.49045	0.47835	–1.17016	1.00599	0.07350	0.08605
Site 10	–1.03291	–1.03505	2.74692	1.28084	–0.36299	1.98648	1.05356
Correlations of environmental variables with site scores							
Depth	0.18608	–0.60189	0.65814				
Coral	0.99233	–0.09189	–0.04614				
Sand	–0.21281	0.91759	0.03765				
Other subs.	–0.87958	–0.44413	0.02466				
Correlations of environmental variables with fitted site scores (for biplot, scaling 2)							
Depth	0.18636	–0.64026	0.74521				
Coral	0.99384	–0.09775	–0.05225				
Sand	–0.21313	0.97609	0.04263				
Other subs.	–0.88092	–0.47245	0.02792				
Centroids of sites with code "1" for the BINARY environmental variables, scaling 2							
Coral	1.02265	–0.10059	–0.05376				
Sand	–0.66932	3.06532	0.13387				
Other subs.	–1.03049	–0.55267	0.03266				

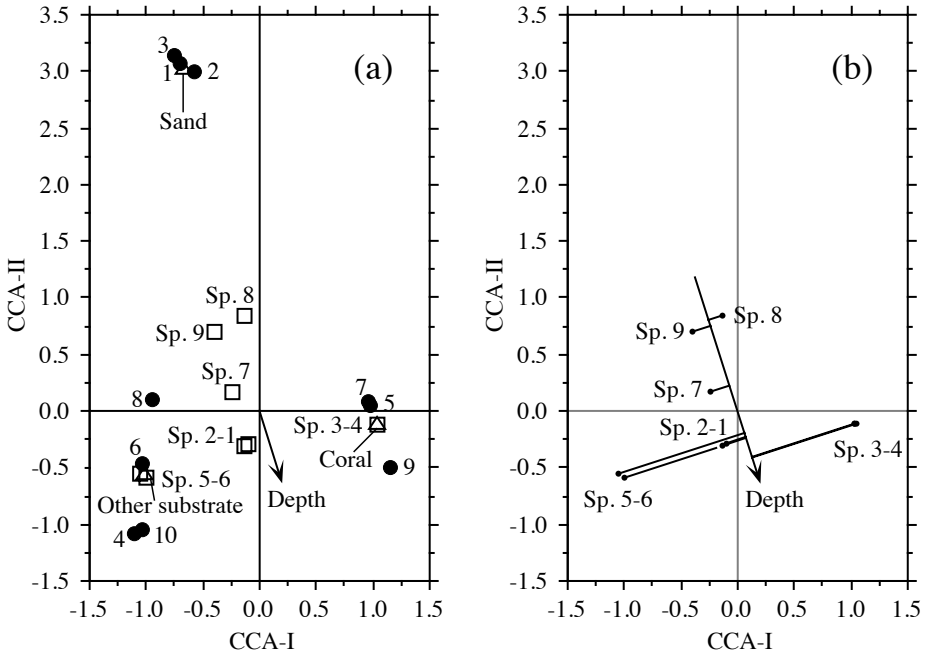


Figure 11.9 CCA ordination triplot (scaling type 2) of the artificial data in Table 11.3; the numerical results of the analysis are in Table 11.7. (a) Triplot representing the species (squares), sites (dots, with site identifiers that also correspond to water depths in m), and environmental variables (full arrow for depth, triangles for the three binary substrate variables). (b) Ranking of the species along the quantitative environmental variable (depth) is inferred by projecting the species at right angle onto the arrow representing that variable.

Figure 11.9b shows how to infer the ranking of species along a quantitative environmental variable. Depth is used in this example. The graphical method simply consists in projecting (at right angle) the species onto the arrow representing that variable. This gives an approximation of the weighted averages of the species with respect to environmental variables. Ecologists like to interpret this ranking as representing the niche optima for the species under consideration. It is important to realize that three rather strong assumptions are made when attempting such an interpretation:

- that the various species have unimodal distributions along the environmental variable of interest (Subsection 9.2.4);

- that the species distributions are under environmental control (Whittaker, 1956; Bray & Curtis, 1957), so that the mode of each species is at its optimum along each environmental variable; and

- that the environmental gradient under study is long enough to allow each species to go from some less-than-optimum low frequency to its high-frequency optimum, and back to some past-optimum low frequency.

In the data of the present example (Table 11.3), only species 1, 3 and 5 were constructed to approximately correspond to these criteria. Species 7, which may also look like it has a unimodal distribution, has actually been constructed using a pseudo-random number generator; so its optimum along depth is fortuitous.

To investigate the similarities among sites or the relationships among species after controlling for the linear effects of depth and type of substrate, one could draw ordination biplots of the *non-canonical axes* in Table 11.7. These axes correspond to a correspondence analysis of the table of regression residuals, as shown in Fig. 11.2.

Ecological application 11.2a

Ecological application 9.2b described the spatial distribution of chaetodontid fish assemblages (butterflyfishes) around a tropical island, using correspondence analysis. This application is continued here. Cadoret *et al*. (1995) next described the relationships between the fish species (quantitative relevés) and some environmental variables, using canonical correspondence analysis. The environmental variables were: the type of environment (qualitative descriptor: bay, lagoon, or outer slope of the reef on the ocean side), geomorphology (qualitative: reef flat, crest, and reef wall of the fringing reefs of bays; fringing reef, shallow, barrier reef, and outer slope for transect sites), depth (quantitative: from 0.5 to 35 m), and exposure to swell (qualitative: low, high, or sites located in bays).

The ordination of sampling sites by CCA is virtually identical to that in Fig. 9.14; this indicates that the first two CA axes are closely related to the environmental variables. The canonical axes account together for 35% of the variation in the species data ($p = 0.001$ after 999 permutations). The description of the ordination of sites presented in Ecological application 9.2b may be compared to Fig. 11.10. This figure shows which types of environment are similar in their chaetodontid species composition and which species are associated with the various types of environment. It indicates that the reef flats of the fringing reefs in bays are similar in species composition to the fringing reefs in the lagoon; likewise, the crests of the fringing reefs in bays are similar to the barrier reefs in the lagoon. The species composition along the reef walls in bays and that on the outer slopes differ, however, from all the other types of environment. The authors discuss the ecology of the most important chaetodontid species in their paper.

Ecological application 11.2b

Canonical correspondence analysis is widely used in palaeoecology, together with regression and calibration, to infer past ecological conditions (climatic, limnological, etc.) from palaeo-assemblages of species. The first 10 years of that literature (1986-1996) was summarized in a bibliography assembled by Birks *et al*. (1998), under the headings *limnology*, *palaeoecology*, *palaeolimnology*, etc. Several applications of CCA are described in a chapter by Legendre & Birks (2012) in a book about numerical methods in palaeoecology edited by Birks *et al*. (2012).

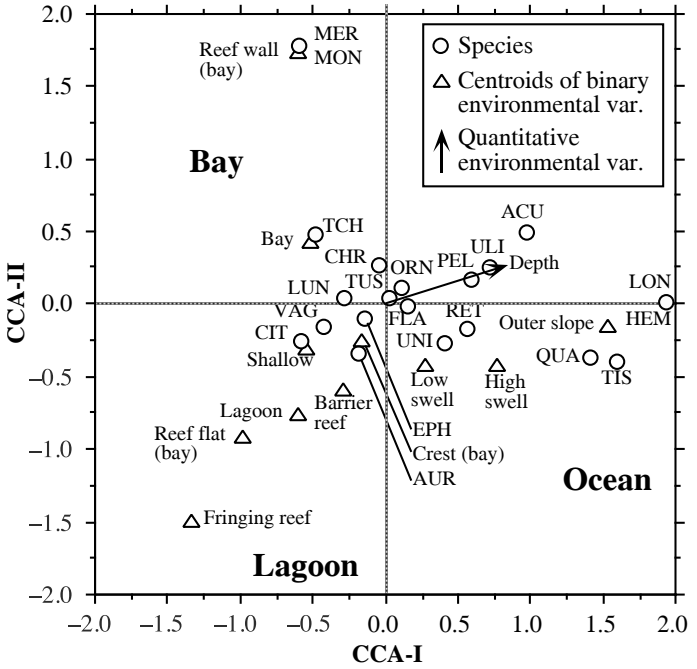


Figure 11.10 CCA ordination diagram: presence/absence of 21 Chaetodontid fish species at 42 sampling sites around Moorea Island, French Polynesia, related to environmental variables. The species (names abbreviated to 3 letters) are represented by circles for readability of the diagram. Axis I: 14.6% of the variation ($p = 0.001$ after 999 permutations); axis II: 7.4% ($p = 0.010$). Redrawn from the original data of Cadoret *et al*. (1995).

One of the classical papers on the subject was written by Birks *et al*. (1990a). Palaeolimnological reconstruction involves two main steps: modelling from a *training data set*, followed by the construction of forecasting models that are then applied to the palaeo-data. In this paper, diatoms were used to reconstruct past water chemistry. The training data set consisted of diatom assemblages comprising 287 species, from present-day surface samples from 138 lakes in England, Norway, Scotland, Sweden, and Wales. Data were also available on pH, conductivity, Ca, Mg, K, SO_4, Cl, alkalinity, total Al, and DOC. Data from more lakes were available for subsets of these variables. CCA was used to relate species composition to water chemistry. The first two canonical eigenvalues were significant and displayed strong species-environment correlations ($r = 0.95$ and 0.84, respectively). The first axis expressed a significant diatom gradient which was strongly and positively correlated with alkalinity and its close correlates, Ca and pH, and negatively but less strongly correlated with total Al; the second axis corresponded to a significant gradient strongly correlated with DOC. This result indicated that pH (or alkalinity), Al, and DOC were potentially reconstructible from the fossil diatom assemblages.

The fossil data set contained 101 slices of a sediment core from a small lake, the Round Loch of Glenhead, in Galloway, southwestern Scotland. The data series covered the past 10000 years. The fossil data (292 diatom taxa) were included in the CCA as passive objects (called *supplementary objects* in Subsection 9.1.9) and positioned in the ordination provided by canonical axes I and II. All fossil objects were well-fitted in that space (they had low squared residual distances), indicating that the pattern of variation in diatom composition can be linked to the modern chemical variables.

Reconstruction of past surface-water chemistry involved two steps. First, the training set was used to model, by regression, the responses of modern diatoms to the chemical variables of interest (one variable at a time). Secondly, the modelled responses were used to infer past chemistry from the composition of fossil diatom assemblages; this phase is called *calibration* (ter Braak, 1987b; ter Braak & Prentice, 1988). Extensive simulations led Birks *et al*. (1990b) to prefer weighted averaging (WA) over maximum likelihood (ML) regression and calibration. Consider pH in lakes, for example. *WA regression* simply consists in applying eq. 9.39 to estimate the pH optimum of each taxon of the training set as the weighted average of all the pH values for lakes in which this taxon occurs, weighted by the taxon's relative abundance. *WA calibration* consists in applying eq. 9.38 to estimate the pH of each lake as the weighted average of the pH optima of all the taxa present. Taxa with a narrow pH tolerance or amplitude may, if required, be given greater weight in WA regression and calibration than taxa with a wide pH tolerance (Birks *et al*., 1990b).

Application of eqs. 9.39 and 9.38 to the data resulted in shrinkage of the range of pH scores. Shrinkage occurred for the same reason as in the TWWA algorithm for correspondence analysis; in step 6.4 of that algorithm (Table 9.8), the eigenvalue was actually estimated from the amount of shrinkage incurred by the site scores after each iteration through eqs. 9.39 and 9.38 (steps 3 and 4). Deshrinking may be done in at least two ways; the relative merits of the two methods are discussed by Birks *et al*. (1990b).

• Deshrinking by classical regression proceeds in two steps. (1) The pH values inferred by WA regression and calibration ($\hat{x}_i$) are regressed on the observed values x_i for the training set, using the linear regression model $\hat{x}_i = b_0 + b_1 x_i + \varepsilon_i$. (2) The parameters of that model are then used to deshrink the $\hat{x}_i$ values, using the equation: final $\hat{x}_i = (\hat{x}_i - b_0)/b_1$. This method was used to deshrink the inferred pH values.

• Another way of deshrinking, advocated by ter Braak & van Dam (1989) for palaeolimnological data, is to use "inverse regression" of x_i on $\hat{x}_i$ (ter Braak, 1987b). Inverse regression was used to deshrink the inferred Al and DOC values.

Training sets containing different numbers of lakes were used to infer pH, total Al, and DOC. Past values of these variables were then reconstructed from the palaeo-assemblages of diatoms, using the pH optima estimated above (eq. 9.39) for the various diatom species, followed by deshrinking. Reconstructed values were plotted against depth and time, together with error estimates obtained by bootstrapping. The past history of the Round Loch of Glenhead over the past 10000 years is discussed in the paper.

This approach involving CCA, WA regression, and WA calibration, is now widely used in palaeolimnology to reconstruct, for example, surface-water temperatures from fossil chironomid assemblages, as well as lake salinity, lake water phosphorus concentrations, or surface water chlorophyll *a* concentrations from fossil diatom assemblages. The WA regression and WA calibration method was further improved by ter Braak & Juggins (1993). ter Braak (1995) made a theoretical comparison of reconstruction methods. For a recent presentation, see the chapter by Birks (2010) in a book edited by Smol & Stoermer (2010). How to carry out the calculations was

described by ter Braak & Juggins (1993) and Line *et al*. (1994). R functions for palaeoenvironmental reconstruction are available in package RIOJA (Juggins, 2009).

A little-known application of CCA is worth mentioning here. Consider a qualitative environmental variable and a table of species presence-absence or abundance data. How can one "quantify" the qualitative states, i.e. give them values along a quantitative scale that would be related in some optimal way to the species data? CCA provides an easy answer to this problem. The species data form matrix **Y**; the qualitative variable, which may be coded as a factor or recoded as a set of dummy variables, is placed in matrix **X**. Compute CCA and take the fitted site scores (or "site scores that are linear combinations of environmental variables"): they provide a quantitative rescaling of the qualitative variable, maximizing the weighted linear correlation between the dummy variables and matrix $\mathbf{\bar{Q}}$. In the same way, RDA may be used to rescale a qualitative variable (factor) with respect to a table of quantitative variables of the objects if linear relationships can be assumed.

McCune (1997) warns users of CCA against inclusion of noisy or irrelevant explanatory variables in the analysis: they may lead to misleading interpretations.

11.3 Linear discriminant analysis (LDA)

A situation that often occurs is to start with an already known grouping of the objects, considered to form a qualitative response variable **y** in this type of analysis, and try to determine to what extent a set of quantitative descriptors, which are the explanatory variables **X**, can actually explain this grouping. In this type of analysis, the grouping is known at the start of the analysis. It may be the result of a cluster analysis computed from a *different* data set, or reflect an ecological hypothesis to be tested. The problem no longer consists in delineating groups, as in cluster analysis, but in interpreting them.

Linear discriminant analysis is a method of linear modelling, like the analysis of variance, multiple linear regression, redundancy analysis, and canonical correlation analysis. It proceeds in two steps. (1) First, one tests for differences in the predictor variables (**X**) among the predefined groups using Wilks' lambda (eq. 11.42). This part of the analysis is identical to the overall test performed in MANOVA. (2) If the test supports the alternative hypothesis of significant differences among groups in the **X** variables, the analysis proceeds to find the linear combinations (called *discriminant functions* or *identification functions*) of the **X** variables that best discriminate among the groups.

Like one-way analysis of variance, discriminant analysis considers a single classification criterion (i.e. division of the objects into groups) and allows one to test whether the explanatory variables can discriminate among the groups. Testing for differences among group means, in discriminant analysis, is identical to ANOVA for a single explanatory variable and to MANOVA for multiple variables (**X**).

When it comes to modelling, i.e. finding the linear combinations of the predictors (**X**) that best discriminate among the groups, discriminant analysis is a form of "inverse analysis" (ter Braak, 1987b), where the classification criterion is considered to be the response variable (**y**) whereas the quantitative variables (matrix **X**) are predictors of the classification. In ANOVA, on the contrary, the objective is to estimate if the variation in a quantitative response descriptor **y** is significantly explained by one or several classification criteria (explanatory variables **X**). As in multiple regression, the discriminatory power of **X** is the same in LDA for **X** standardized or not.

As in multiple regression, discriminant analysis estimates the parameters of a linear model of the explanatory variables that may be used to forecast the response variable (states of the classification criterion). While inverse multiple regression would be limited to two groups (expressed by a single binary variable **y**), discriminant analysis can handle several groups. Discriminant analysis is a canonical method of analysis; its link to canonical correlation analysis (CCorA) is explained at the end of Subsection 11.3.1, after some necessary concepts have been introduced.

After the overall test of significance, the search for discriminant functions may be conducted with two different purposes in mind. One may be interested in obtaining a linear equation to allocate new objects to one of the states of the classification criterion (identification), or simply in determining the relative contributions of various explanatory descriptors to the distinction among these states (discrimination).

Discriminant analysis is also called *canonical variate analysis* (CVA). The method was originally proposed by Fisher (1936) for the two-group case ($g = 2$). Fisher's results were extended to $g > 2$ by Rao (1948, 1952). Fisher (1936) illustrated the method using a famous data set describing the morphology (lengths and widths of sepals and petals) of 150 specimens of irises (Iridaceae) belonging to three species. The data had originally been collected in the Gaspé Peninsula, eastern Québec (Canada), by the botanist Edgar Anderson of the Missouri Botanical Garden who allowed Fisher to publish and use the raw data in his 1936 paper. Fisher showed how to use these morphological measurements to discriminate among the species. The data set is sometimes — erroneously — referred to as "Fisher's irises".

The analysis is based upon an explanatory data matrix **X** of size ($n \times m$), where n objects are described by m descriptors. **X** is meant to discriminate among the groups defined by a separate classification criterion vector (**y**). As in regression analysis, the explanatory descriptors must in principle be quantitative, although qualitative descriptors coded as dummy variables may also be used (Subsection 1.5.7). Other methods are available for discrimination using non-quantitative descriptors (Table 10.1). The objects, whose membership in the various groups of **y** is known before the analysis is undertaken, may be sites, specimens, quadrats, etc.

One possible approach would be to examine the descriptors one by one, either by hand or using analyses of variance, and to note those which have states that characterize one or several groups. This information could be transformed into an

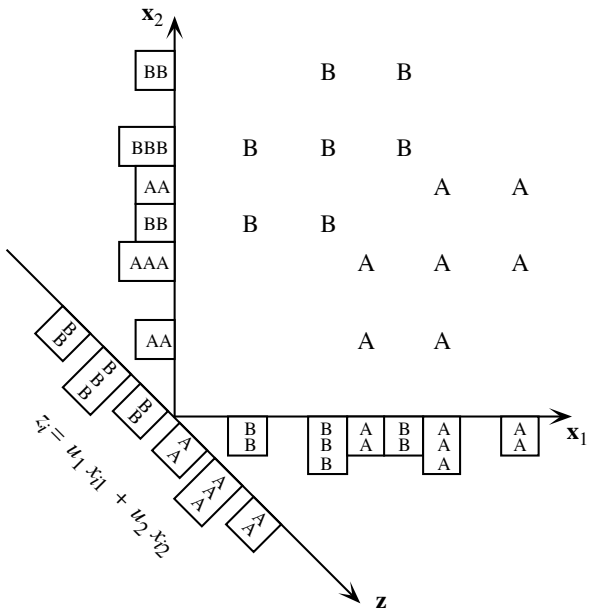


Figure 11.11 Two groups, A and B, with 6 objects each, are overlapping on both axes, $\mathbf{x}_1$ and $\mathbf{x}_2$, as shown by the histograms on the axes. They are perfectly separated, however, along a discriminant axis **z**. The position of each object i is calculated along **z** using the equation $z_i = (\cos 45°)\, x_{i1} - (\cos 45°)\, x_{i2}$. Adapted from Jolicoeur (1959).

identification key, for example. It often occurs, however, that no single descriptor succeeds in separating the groups completely. The next best approach is then to search for a linear combination of descriptors that provides the most efficient discrimination among groups. Figure 11.11 shows an idealized example of two groups (A and B) described by two descriptors only. The groups cannot be separated on either of the two axes taken alone. The solution is a new discriminant descriptor **z**, drawn on the figure, which is a linear combination of the two original descriptors. Along **z**, the two groups of objects are perfectly separated. Note that discriminant axis **z** is parallel to the direction of greatest variability *between groups*. This suggests that the weights u_j used in the discriminant function could be the elements of the eigenvectors of a between-group dispersion matrix. The method can be generalized to more than two groups and several descriptors.

Discriminant function

- *Discriminant functions* (also called standardized discriminant functions) are computed from *standardized descriptors*. The coefficients of these functions are used to assess the relative contributions of the descriptors to the final discrimination.

Identification function

• *Identification functions* (also called unstandardized discriminant functions) are computed from the *original descriptors* (not standardized). They may be used to compute the group to which a new object is most likely to belong. Discriminant analysis is seldom used for this purpose in ecology, whereas it is widely used in that way in taxonomy.

When there are only two groups of objects, the method is called *Fisher's*, or *simple discriminant analysis* (a single function is needed to discriminate between two groups), whereas the case with several groups is called *multiple discriminant analysis* or *canonical variate analysis*. Because the simple discriminant analysis model (two groups) is a particular case of multiple discriminant analysis, it will not be developed here. The solution can be entirely derived from the output of a multiple regression using a dummy variable defining the two groups (used as the dependent variable **y**) against the table of predictor variables **X**.

Analysis of variance is often used for screening variables prior to discriminant analysis: each variable in matrix **X** is tested for its capacity to discriminate among the groups of the classification **y**. Figure 11.11 shows however that there is a danger in this approach; any single variable may not discriminate groups well, although it may have high discriminating power in combination with other variables. One should be careful when using univariate analysis to eliminate variables. If the analysis requires that poorly discriminating variables be eliminated, one should use stepwise discriminant analysis instead, which allows users to identify a subset of good discriminators. Bear in mind, though, that stepwise selection of explanatory variables does not guarantee that the "best" set of explanatory variables will necessarily be found. This is equally true in discriminant analysis and regression analysis (Subsection 10.3.3).

1 — The algebra of discriminant analysis

The problem consists in finding linear combinations of the predictors in matrix **X** ($n \times m$) that maximize the differences among groups while minimizing the variation within the groups. As in regression analysis, the descriptors must be quantitative or binary since they are combined into linear functions. Each descriptor may have already been transformed to meet the condition of multinormality or at least to reduce the asymmetry of its distribution. The discriminant analysis model is robust to departures from this condition, but the parametric statistical tests assume *within-group normality* of the descriptors.

Computations are carried out using either dispersion matrices (**V**, **A**) or matrices of sums of squares and cross-products of centred descriptors (**W**, **B**) (Table 11.8). These matrices are constructed in the same way as in analysis of variance, except that here the predictors form a multivariate data matrix. Matrix **T** is the matrix of scalar products of the centred descriptors, $[x - \bar{x}]$, for all objects irrespective of the groups: $\mathbf{T} = [x - \bar{x}]'[x - \bar{x}]$ (total sums of squares and cross-products). When divided by the total number of degrees of freedom $n - 1$, it becomes the total dispersion matrix **S** used in principal component analysis.

Table 11.8 Discriminant analysis is computed on either dispersion matrices (right-hand column) or matrices of sums of squares and cross-products (centre). Matrices in the right-hand column are simply those in the central column divided by their respective numbers of degrees of freedom. The dimension of all matrices is ($m \times m$).

	Matrices of sums of squares and cross-products	Dispersion matrices
Total dispersion	$\mathbf{T}$	$\mathbf{S} = \mathbf{T}/n - 1$
Pooled within-group dispersion	$\mathbf{W} = \mathbf{W}_1 + \ldots + \mathbf{W}_g$	$\mathbf{V} = \mathbf{W}/n - g$
Among-group dispersion	$\mathbf{B} = \mathbf{T} - \mathbf{W}$	$\mathbf{A} = \mathbf{B}/g - 1$

Matrix **W**, which pools the sums of squares within all groups, is computed by adding up matrices $\mathbf{W}_1$ to $\mathbf{W}_g$ of the sums of squares and cross-products for each of the g groups. Each matrix $\mathbf{W}_j$ is computed from descriptors that have been *centred for the objects of that group only*, just as in ANOVA. In other words, matrix $\mathbf{W}_j$ is the product $[x - \bar{x}]'[x - \bar{x}]$ for the objects that belong to group j only. Dividing the pooled within-group matrix **W** by the within-group number of degrees of freedom, $n - g$, produces the pooled within-group dispersion matrix **V**. In ANOVA involving a single explanatory variable, **W** contains a single value, the residual sum of squares, which is the sum of the within-group sums of squares ($SS_{\text{within groups}}$).

Matrix **B** of the sums of squares among groups is computed by subtracting the pooled within-group matrix **W** from the total matrix of sums of squares **T**. Since $\mathbf{B} = \mathbf{T} - \mathbf{W}$, the number of degrees of freedom by which **B** must be divided to obtain the among-group dispersion matrix **A** is: $(n - 1) - (n - g) = g - 1$. In ANOVA involving a single explanatory variable, **B** contains a single value, the among-group sum of squares ($SS_{\text{among groups}}$).

In analysis of variance involving a single explanatory variable, sums of squares (SS) among and within groups are used to construct the F-statistic to test the hypothesis of no difference between the means of the groups:

$$F = \frac{\text{MS}_{\text{Among groups}}}{\text{MS}_{\text{Within groups}}} = \frac{SS_{\text{Among groups}} / (g - 1)}{SS_{\text{Within groups}} / (n - g)}$$

where MS stands for "mean square".

The matrix of predictor variables, **X**, is multivariate in discriminant analysis. The numerator of F is matrix **A** and its denominator is matrix **V**. One cannot compute **A/V**

but one can compute $\mathbf{V}^{-1}\mathbf{A}$. The eigenvalues of $\mathbf{V}^{-1}\mathbf{A}$ are the canonical *F*-statistics, which can be tested for significance. The eigenvalues are found by eigen-decomposition (eq. 2.22) using the following equation:

$$(\mathbf{V}^{-1}\mathbf{A} - \lambda_k\mathbf{I})\,\mathbf{u}_k = \mathbf{0} \qquad \textbf{(11.33)}$$

This equation maximizes the variation among groups while minimizing the variation within groups. It also produces the matrix of normalized eigenvectors $\mathbf{U} = [\mathbf{u}_k]$. Matrix $\mathbf{V}^{-1}\mathbf{A}$ is asymmetric, as in eqs. 11.48 and 11.50 of CCorA, so its eigenvectors are not orthogonal. The maximum number of discriminant axes needed for the ordination of g groups is $(g - 1)$, so the number of canonical eigenvalues is at most *min*(number of predictors in $\mathbf{X}$, $(g–1)$).

Eigen-decomposition could have been carried out using the matrices of sums of squares and cross-products $\mathbf{W}$ and $\mathbf{B}$ instead of the dispersion matrices $\mathbf{V}$ and $\mathbf{A}$. The eigen-decomposition of

$$(\mathbf{W}^{-1}\mathbf{B} - l_k\mathbf{I})\,\mathbf{u}_k = \mathbf{0} \qquad \textbf{(11.34)}$$

produces the same matrix of eigenvectors $\mathbf{U}$ as eq. 11.33. The eigenvalues l_k are modified by a factor corresponding to the degrees of freedom shown in the equation of the *F*-statistic and in Table 11.8:

$$l_k = \frac{g-1}{n-g}\,\lambda_k \qquad \textbf{(11.35)}$$

which leaves unchanged the percentage of the variance of $\mathbf{V}^{-1}\,\mathbf{A}$ or $\mathbf{W}^{-1}\,\mathbf{B}$ explained by each canonical eigenvalue.

When the non-orthogonal eigenvectors are plotted at right angles, they straighten the reference space and, with it, the ellipsoids of the within-group scatters of objects. If the eigenvectors are now rescaled in an appropriate manner, the within-group scatters of objects can be made circular (Fig. 11.12), insofar as the within-group cross-product matrices $\mathbf{W}_j$ are homogeneous (same dispersion in all groups). This is done by rescaling the eigenvectors (matrix $\mathbf{U}$) using the following formula. The result is matrix $\mathbf{C}$ containing the discriminant function coefficients:

$$\mathbf{C} = \mathbf{U}\left(\mathbf{U}'\frac{\mathbf{W}}{n-g}\mathbf{U}\right)^{-1/2} = \mathbf{U}\,(\mathbf{U}'\mathbf{V}\mathbf{U})^{-1/2} \qquad \textbf{(11.36)}$$

where $(\mathbf{U}'\mathbf{V}\mathbf{U})^{-1/2}$ is a diagonal matrix.

Matrix $\mathbf{C}$ contains the rescaled eigenvectors defining the *canonical space* of the discriminant analysis. After this transformation, the variance among group centroids is maximized even if the group dispersion matrices are not homogeneous. This leads to the conclusion that the principal axes describe the dispersion *among groups*. The first

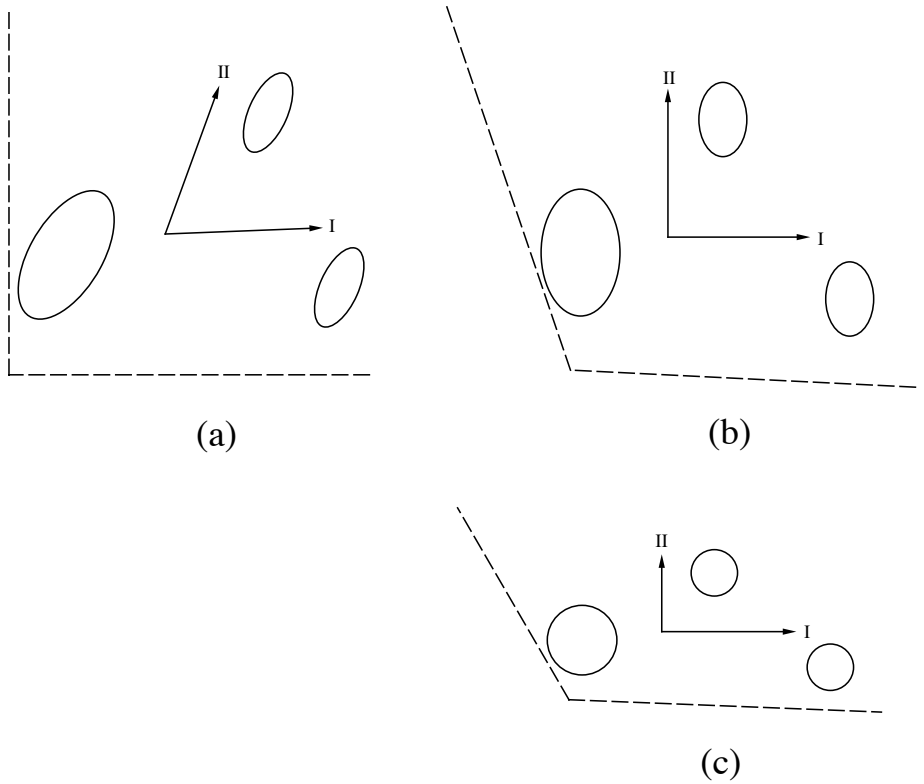


Figure 11.12 Basic principles of multiple discriminant analysis. Dashed: two original descriptors. Full lines: canonical axes. The within-group dispersion matrices are homogeneous in this example. (a) The canonical axes are not orthogonal in the reference space of the original descriptors. (b) When they are used as the orthogonal reference space, the ellipsoids of the within-group scatters of objects are straightened up. (c) Rescaling the eigenvectors to produce **C** (eq. 11.36) makes the within-group dispersions circular if they are homogeneous.

principal axis indicates the direction of largest variation among group centroids, and so on for the successive canonical axes, after the reference space has been straightened up to make each group spherical. The SAS and STATISTICA packages, among others, as well as function ***lda()*** of the MASS package in R, offer the normalization of eq. 11.36.

Other methods for normalizing the eigenvectors are found in the literature, i.e. to lengths 1 or $\sqrt{\lambda}$. Some statistical packages unfortunately compute the positions of the objects along the canonical axes (matrix **F**, eq. 11.37) directly from matrix **U** of the eigenvectors normalized to length 1. In that case, the group dispersions remain nonspherical; it is then difficult to compare the eigenvectors because they describe a combination of within-group and among-group dispersion. It is not always easy to

understand, from the documentation, what a specific statistical program does. A simple way to find out what kind of normalization is used by a program or R function is to run the small example presented in Subsection 11.3.3.

The last step of the computation is to find the positions of the objects in the space of the canonical axes. The matrix of discriminant scores **F** is obtained by multiplying the matrix of centred data with the matrix of normalized eigenvectors **C**:

$$\mathbf{F} = [x - \bar{x}]\,\mathbf{C} \qquad \textbf{(11.37)}$$

Since the matrix of centred data $[x - \bar{x}]$ is used in eq. 11.37, the origin of the discriminant axes is located at the centroid of all objects, as in Fig. 11.12. It is common practice to also compute the positions of the centroids of the g groups of objects in canonical space, by multiplying the matrix of the original group centroids (computed from data centred over all objects in the analysis) with matrix **C**. The centroid of a group is a point whose coordinates are the mean values of the objects of that group for all descriptors. The matrix of group centroids therefore has g rows and m columns.

As in principal component analysis, equation $\mathbf{F} = [x - \bar{x}]\,\mathbf{C}$ contains the scores of the objects, i, on each canonical axis k:

$$f_{ik} = (x_{i1} - \overline{x_1})\, c_{1k} + \ldots + (x_{ip} - \overline{x_p})\, c_{pk} \qquad \textbf{(11.38)}$$

The columns of matrix **F** are called *canonical variates* in discriminant analysis. The distances among objects in discriminant space are Mahalanobis distances (eq. 7.38). The positions of the group centroids in discriminant space can be found by computing these same functions for the mean values of the groups along the **X** variables.

If the analysis is carried out on the *non-standardized* descriptors, the columns of matrix **C** are called *identification functions*. Identification functions are used to place new objects in the canonical space. To do so, values of the various descriptors of a new object are centred using the same descriptor means as in eq. 11.38, and the centred values are multiplied by the weights c_{jk}. This provides the position of this object on the canonical axes. By plotting the point representing this object in the canonical ordination space together with the original set of objects, it is possible to identify the group to which the new object most likely belongs.

There are other ways of assigning objects to groups. *Classification functions*[*] are linear equations that can be used for that purpose. A separate classification function is computed as follows for each group j:

$$\text{Classification function for group } j = -0.5\ \bar{\mathbf{x}}'_j \mathbf{V}^{-1} \bar{\mathbf{x}}_j + \mathbf{V}^{-1} \bar{\mathbf{x}}_j \qquad \textbf{(11.39)}$$

[*] This terminology is unfortunate. In biology, classification consists in forming groups, using clustering methods for instance (Chapter 8), whereas identification is to assign objects to preestablished groups.

where $\mathbf{V}$ is the pooled within-group dispersion matrix (Table 11.8) and $\bar{\mathbf{x}}_j$ is the vector describing the centroid of group j for all m variables of matrix $\mathbf{X}$. Each classification function looks like a multiple regression equation; eq. 11.39 provides the weights $(\mathbf{V}^{-1}\bar{\mathbf{x}}_j)$ to apply to the various descriptors of matrix $\mathbf{X}$ combined in the linear equation, as well as a constant $(-0.5\ \bar{\mathbf{x}}'_j\mathbf{V}^{-1}\bar{\mathbf{x}}_j)$. The classification score of each object is calculated for each of the g classification functions; an object is assigned to the group for which it receives the highest classification score. Another way is to compute Mahalanobis distances (eq. 7.38) of the objects from each of the group centroids. An object is then assigned to the group to which it is the closest.

Confusion or classification table

A *confusion* or *classification table* (also called *confusion or classification matrix*) can be constructed; this is a contingency table comparing the original assignment of objects to groups (usually in rows) to the group assignments made by the classification functions (in columns). From that table, users can determine the number and percentage of the objects correctly classified by the discriminant functions.

An alternative way to obtain the eigenvalues and eigenvectors in discriminant analysis is through the canonical correlation equation (eq. 11.50). The method is described by Tatsuoka & Lohnes (1988, Section 7.8). The vector of classification levels is first transformed into a matrix $\mathbf{Y}$ containing $(g-1)$ binary or Helmert-coded variables (Subsection 1.5.7). One then computes the eigenvalues and eigenvectors of matrix $\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$. The notation is the same as in eq. 11.46: $\mathbf{S}_{11}$ is the covariance matrix of $\mathbf{Y}$, $\mathbf{S}_{22}$ is the covariance matrix of $\mathbf{X}$, and $\mathbf{S}_{12}$ is the covariance matrix crossing the two groups of variables:

$$(\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}\mathbf{S}_{12} - h_k\mathbf{I})\,\mathbf{u}_k = 0 \qquad \textbf{(11.40)}$$

$\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ is the transpose of the matrix in eq. 11.50 (Tatsuoka & Lohnes, 1988, eq. 7.26). Because this matrix is of order $(m \times m)$, the computed matrix of eigenvectors $\mathbf{U}$ has m rows and contains the weights associated with the predictors $\mathbf{X}$.

The discriminant eigenvalues l_k are found by transforming the eigenvalues h_k as follows:

$$l_k = h_k / (1 - h_k)$$

It is the eigenvalues l_k of eq. 11.34 that are found here because eq. 11.40 does not involve the degrees of freedom of Table 11.8. The discriminant eigenvalues λ_k are obtained by applying eq. 11.35 in reverse. The matrix of eigenvectors $\mathbf{U}$ is the same as obtained from eqs. 11.33 and 11.34. Matrix $\mathbf{C}$ of discriminant function coefficients can then be computed using eq. 11.36.

This approach is further supported by the demonstration made by Gittins (1985) that $\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ is matrix $\mathbf{B}$ of Table 11.8, while $\mathbf{S}_{22}$ is matrix $\mathbf{T}$. Hence eq. 11.40, which is the form of the CCorA equation applied to discriminant analysis, can be written $(\mathbf{T}^{-1}\mathbf{B} - \lambda_k\mathbf{I})\,\mathbf{u}_k = 0$.

2 — Statistics in linear discriminant analysis

Spherical within-group dispersions are obtained only if the condition of homogeneity of the within-group dispersion matrices is fulfilled. Even if discriminant analysis is moderately robust to departures from this condition, it remains advisable to examine whether this condition is met prior to LDA. Several statistics have been developed to test the hypothesis of homogeneity of the within-group dispersion matrices. One of them is Kullback's statistic (1959) which is approximately distributed as χ^2:

$$\chi^2 = \sum_{j=1}^{g} \frac{(n_j - 1)}{2} \log_e \frac{|\mathbf{V}|}{|\mathbf{V}_j|} \qquad \textbf{(11.41)}$$

with $(g-1)m(m+1)/2$ degrees of freedom, where n_j is the number of objects in group j, $|\mathbf{V}|$ is the determinant of the pooled within-group dispersion matrix $\mathbf{V}$, and $|\mathbf{V}_j|$ is the determinant of the within-group dispersion matrix of group j. When the test value is larger than or equal to the critical χ^2 value, the hypothesis of homogeneity is rejected. Another method, which is robust to departures from normality, is the test of homogeneity of multivariate dispersions developed by Anderson (2006); this is the multivariate analogue of Levene's (1960) univariate test for homogeneity of variances. Anderson's test can be computed for any dissimilarity measure of choice. It is available in VEGAN's function ***betadisper()***.

Several important tests in discriminant analysis are based on Wilks' Λ (lambda) statistic (1932). This statistic can be used in an overall test to assess if the groups significantly differ in the positions of their centroids, given the within-group dispersions. Λ is computed as the ratio of the determinants of the matrices of sums of squares and cross-products $\mathbf{W}$ and $\mathbf{T}$:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} \qquad \textbf{(11.42)}$$

This ratio produces values in the range from near 0 (maximum dispersion of the centroids) to 1 (no dispersion among groups). It can be transformed to a X^2 (chi-square) statistic with $m(g-1)$ degrees of freedom (Bartlett, 1938):

$$X^2 = -\left[(n-1) - \frac{1}{2}(m+g) \right] \log_e \Lambda \qquad \textbf{(11.43)}$$

Alternatively, Wilks' Λ can be transformed into an F-statistic following Rao (1951). It is a generalization of Student's t-test to several groups and several explanatory variables. Another multidimensional generalization of t, for two groups, is Hotelling's T^2 (eq. 7.41), which has been discussed with reference to the Mahalanobis generalized distance (eq. 7.39).

As explained above, discrimination among g groups requires a maximum of $(g - 1)$ discriminant functions. To test the significance of the $(g - k - 1)$ eigenvalues that remain after examining the first k, Wilks' ratio is computed as the product

$$L = \prod_{j=k+1}^{g-1} \frac{1}{1 + l_j} \tag{11.44}$$

where the l_j are the eigenvalues of eq. 11.34. The value L computed for all eigenvalues produces the value Λ of eq. 11.42. Transformation of this statistic to X^2, as above (eq. 11.43), allows one to estimate the significance of the discriminant power of the axes remaining after accepting the first k eigenvalues as significant (Bartlett, 1948):

$$X^2 = \left[(n-1) - \frac{1}{2}(m+g) \right] \log_e \left[\prod_{j=k+1}^{g-1} (1 + l_j) \right] \tag{11.45}$$

with $(m - k)(g - k - 1)$ degrees of freedom. (The logarithm of L from eq. 11.44 is equal to minus the logarithm of the product of the terms $(1 + l_j)$ in the denominator.) When the last $(g - k - 1)$ canonical eigenvalues, taken together, do not reach the chosen critical χ^2 value, the null hypothesis that the centroids of the groups do not differ on the remaining $(g - k - 1)$ discriminant functions cannot be rejected. This indicates that the detectable discriminant power is limited to the first k discriminant functions.

3 — Numerical example

Discriminant analysis is illustrated by means of a numerical example in which seven objects, allocated to three groups, are described by two descriptors. The calculation of *identification functions* is shown first (raw data), followed by *discriminant functions* (standardized data). Normally, these data should not be submitted to discriminant analysis since the variances of the group matrices are not homogeneous; they are used here to illustrate the steps involved in the computation. The data set is the following:

$$\begin{array}{lccccccccc} \text{Groups} = & & \underline{\quad 1 \quad} & & \underline{\quad 2 \quad} & & \underline{\quad 3 \quad} & & \text{Means} \\ \mathbf{X}' = & \left[\begin{array}{ccccccc} 1 & 2 & 2 & 8 & 8 & 8 & 9 \\ 2 & 2 & 1 & 7 & 6 & 3 & 3 \end{array}\right] & & & & & & & \left[\begin{array}{c} 5.42857 \\ 3.42857 \end{array}\right] \end{array}$$

The centred data for the objects and the group centroids are the following:

$$\text{Groups} = \quad \underline{\quad 1 \quad} \quad \underline{\quad 2 \quad} \quad \underline{\quad 3 \quad}$$

$$[\mathbf{X}\text{ centred}]' = [x - \bar{x}]' = \begin{bmatrix} -4.429 & -3.429 & -3.429 & 2.571 & 2.571 & 2.571 & 3.571 \\ -1.429 & -1.429 & -2.429 & 3.571 & 2.571 & -0.429 & -0.429 \end{bmatrix}$$

$$[\text{Centroids}]' = \begin{bmatrix} -3.762 & 2.571 & 3.071 \\ -1.762 & 3.071 & -0.429 \end{bmatrix}$$

The matrix of sums of squares and cross-products is:

$$\mathbf{T} = [x - \bar{x}]'[x - \bar{x}] = \begin{bmatrix} 75.71429 & 32.71429 \\ 32.71429 & 29.71429 \end{bmatrix}$$

The pooled within-groups matrix **W** is computed by adding up the three group matrices of sums of squares and cross-products $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$:

$$\mathbf{W} = \begin{bmatrix} 0.66667 & -0.33333 \\ -0.33333 & 0.66667 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0.5 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1.16667 & -0.33333 \\ -0.33333 & 1.16667 \end{bmatrix}$$

The determinants of matrices **W** and **T** are 1.25000 and 1179.57, respectively. The ratio of these two values is Wilks' Λ (eq. 11.42: $\Lambda = 0.00106$; eq. 11.43: $X^2 = 23.97$, $p < 0.001$). This indicates that there are significant differences among the groups in the **X** variables. Hence, the analysis can proceed with the calculation of identification and discriminant functions.

To obtain the *identification functions*, the matrix of sums of squares among groups is computed as:

$$\mathbf{B} = \mathbf{T} - \mathbf{W} = \begin{bmatrix} 74.54762 & 33.04762 \\ 33.04762 & 28.54762 \end{bmatrix}$$

The characteristic equation $|\mathbf{B} - l\mathbf{W}| = 0$ is used to calculate the two eigenvalues:

$$l_1 = 106.03086 \quad \Rightarrow \quad \lambda_1 = \frac{(7-3)}{(3-1)} \times 106.03086 = 212.06171\,(93.13\%)$$

$$l_2 = 7.81668 \quad \Rightarrow \quad \lambda_2 = 2 \times 7.81668 = 15.63336\,(6.87\%)$$

In this example, canonical axes 1 and 2 explain 93.13 and 6.87% of the among-group variation, respectively. The two eigenvalues are used to compute the eigenvectors, by

means of matrix equation $(\mathbf{B} - l_k\mathbf{W})\,\mathbf{u}_k = \mathbf{0}$. These eigenvectors, normalized to length 1, are the columns of matrix **U**:

$$\mathbf{U} = \begin{bmatrix} 0.81202 & -0.47849 \\ 0.58363 & 0.87809 \end{bmatrix}$$

The vectors are not orthogonal since $\mathbf{u}_1' \mathbf{u}_2 = 0.12394$. In order to bring the eigenvectors to their final lengths, the following scaling matrix is computed:

$$\left(\mathbf{U}'\frac{\mathbf{W}}{n-g}\mathbf{U}\right)^{1/2} = \begin{bmatrix} 0.46117 & 0 \\ 0 & 0.60141 \end{bmatrix}$$

The component terms of each eigenvector $\mathbf{u}_j$ are *divided* by the corresponding diagonal term from this matrix, to obtain the final vectors (identification functions):

$$\mathbf{C} = \begin{bmatrix} 1.76077 & -0.79562 \\ 1.26553 & 1.46006 \end{bmatrix}$$

Multiplication of the centred matrices of the raw data and centroids by **C** gives the positions of the objects (matrix **F**) and centroids in canonical space (Fig. 11.13):

$$\text{Groups} = \quad \underline{\quad 1 \quad} \quad \underline{\quad 2 \quad} \quad \underline{\quad 3 \quad}$$

$$\mathbf{F} = [\mathbf{X}\ \text{centred}]\,\mathbf{C} = [x - \bar{x}]\,\mathbf{C} = \begin{bmatrix} -9.606 & -7.845 & -9.111 & 9.047 & 7.783 & 3.985 & 5.747 \\ 1.438 & 0.642 & -0.818 & 3.169 & 1.708 & -2.672 & -3.466 \end{bmatrix}'$$

$$[\text{Centroids}]\,\mathbf{C} = \begin{bmatrix} -8.854 & 8.415 & 4.866 \\ 0.420 & 2.438 & -3.069 \end{bmatrix}'$$

One can verify that, in canonical space, the among-group dispersion matrix **A** is equal to the matrix of eigenvalues and that the pooled within-groups dispersion matrix **V** is the identity matrix **I**. Beware: some computer programs calculate the discriminant scores as **XU** instead of [**X** centred] **U** or [**X** centred] **C**.

The *classification functions*, computed from eq. 11.39, are the following for descriptors x_1 and x_2 of the example:

Group 1: $\text{Score}_i = -13.33333 + 8.00000\, x_{i1} + 8.00000\, x_{i2}$

Group 2: $\text{Score}_i = -253.80000 + 36.80000\, x_{i1} + 32.80000\, x_{i2}$

Group 3: $\text{Score}_i = -178.86667 + 34.93333\, x_{i1} + 20.26667\, x_{i2}$

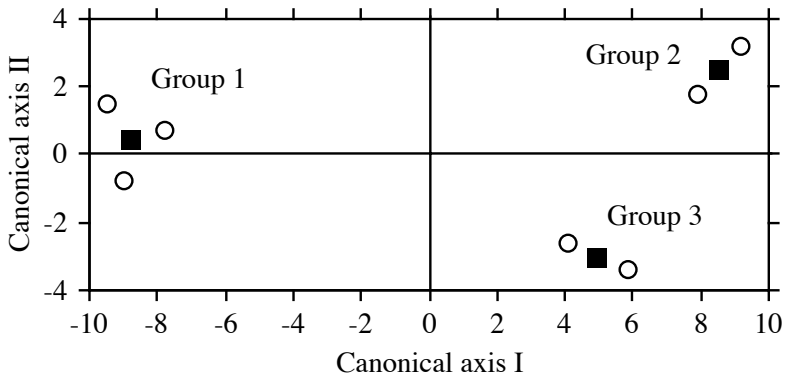


Figure 11.13 Ordination diagram of the seven objects (circles) and group centroids (squares) of the example in the canonical discriminant space.

The scores of the 7 objects i, computed from these functions, are the following:

Object number	Observed group	Function 1	Function 2	Function 3	Assigned to group
1	1	10.66667	–151.40000	–103.40000	1
2	1	18.66667	–114.60000	–68.46667	1
3	1	10.66667	–147.40000	–88.73334	1
4	2	106.66667	270.20000	242.46666	2
5	2	98.66667	237.40000	222.19999	2
6	3	74.66667	139.00000	161.39998	3
7	3	82.66667	175.80000	196.33331	3

Each object is assigned (right-hand column) to the group corresponding to the function giving it the highest score. The *classification table* can now be constructed; this is a contingency table comparing the original group assignment of the objects (from the second column in table above) to the assignment made from the classification functions (last column in table above):

Observed group	Assigned to group			Total and % correct
	1	2	3	
1	3	0	0	3 (100%)
2	0	2	0	2 (100%)
3	0	0	2	2 (100%)
Total	3	2	2	7 (100%)

In order to compute *discriminant functions*, the descriptors are standardized at the start of the analysis:

$$\text{Groups} = \quad \underline{\quad 1 \quad} \quad \underline{\quad 2 \quad} \quad \underline{\quad 3 \quad}$$

$$[\mathbf{X}\ \text{standardized}]' = \left[\frac{x-\bar{x}}{s_x}\right]' = \begin{bmatrix} -1.247 & -0.965 & -0.965 & 0.724 & 0.724 & 0.724 & 1.005 \\ -0.642 & -0.642 & -1.091 & 1.605 & 1.155 & -0.193 & -0.193 \end{bmatrix}$$

$$[\text{Centroids}]' = \begin{bmatrix} -1.059 & 0.724 & 0.865 \\ -0.792 & 1.380 & -0.193 \end{bmatrix}$$

The remaining calculations are the same as for the identification functions (above):

$$\mathbf{T} = \left[\frac{x-\bar{x}}{s_x}\right]' \left[\frac{x-\bar{x}}{s_x}\right] = \begin{bmatrix} 6.00000 & 4.13825 \\ 4.13825 & 6.00000 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 0.05283 & -0.04217 \\ -0.04217 & 0.13462 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0.10096 \end{bmatrix} + \begin{bmatrix} 0.03962 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.09246 & -0.04217 \\ -0.04217 & 0.23558 \end{bmatrix}$$

$$\mathbf{B} = \mathbf{T} - \mathbf{W} = \begin{bmatrix} 5.90755 & 4.18042 \\ 4.18042 & 5.76441 \end{bmatrix}$$

$$l_1 = 106.03086 \quad \Rightarrow \quad \lambda_1 = \frac{(7-3)}{(3-1)} \times 106.03086 = 212.06171\ (93.13\%)$$

$$l_2 = 7.81668 \quad \Rightarrow \quad \lambda_2 = 2 \times 7.81668 = 15.63336\ (6.87\%)$$

The amounts of among-group variation explained by the canonical axes (93.13 and 6.87%) are the same as those obtained above with the unstandardized data.

$$\mathbf{U} = \begin{bmatrix} 0.91183 & -0.65630 \\ 0.41057 & 0.75450 \end{bmatrix}$$

$$\left(\mathbf{U}'\frac{\mathbf{W}}{n-g}\mathbf{U}\right)^{1/2} = \begin{bmatrix} 0.14578 & 0 \\ 0 & 0.23221 \end{bmatrix} \quad \Rightarrow \quad \mathbf{C} = \begin{bmatrix} 6.25473 & -2.82627 \\ 2.81631 & 3.24918 \end{bmatrix}$$

$$\text{Groups} = \quad \underline{\quad 1 \quad} \quad \underline{\quad 2 \quad} \quad \underline{\quad 3 \quad}$$

$$\mathbf{F} = [\mathbf{X} \text{ standardized}]\,\mathbf{C} = \begin{bmatrix} -9.606 & -7.845 & -9.111 & 9.047 & 7.783 & 3.985 & 5.747 \\ 1.438 & 0.642 & -0.818 & 3.169 & 1.708 & -2.672 & -3.466 \end{bmatrix}'$$

$$[\text{Centroids}]\,\mathbf{C} = \begin{bmatrix} -8.854 & 8.415 & 4.866 \\ 0.420 & 2.438 & -3.069 \end{bmatrix}'$$

The raw and standardized data produce exactly the same ordination of the objects and group centroids.

The classification functions computed using the standardized descriptors differ from those reported above for raw data, but the classification table is the same in both cases.

Computer packages usually have an option for variable selection using forward entry, backward elimination, or stepwise selection, as in multiple regression (Subsection 10.3.3). These procedures are useful for selecting only the descriptors that significantly contribute to discrimination, leaving the others out of the analysis. This option must be used with caution. As it is the case with any stepwise computation method, the step-by-step selection of s successively most discriminant descriptors does not guarantee that they form the most discriminant set of s descriptors.

The following ecological application is an example of multiple discriminant analysis among groups of observations, using physical and chemical descriptors as discriminant variables. Steiner *et al*. (1969) applied discriminant analysis to the agronomic interpretation of aerial photographs, based upon a densimetric analysis of different colours.

Ecological application 11.3

Sea ice is an environment with a rich and diversified biota. This is because ice contains a network of brine cells and channels in which unicellular algae, heterotrophic bacteria, protozoa, and small metazoa can develop and often reach very high concentrations. Legendre *et al*. (1991) investigated the environmental factors controlling the growth of microscopic algae in the sea ice of southeastern Hudson Bay, Canadian Arctic.

Ice cores were taken at eight sites along a transect that extended from the mouth of the Great Whale River to saline waters 25 km offshore. Ice thickness ranged from 98 to 125 cm. The cores were used to determine the crystallographic structure of the ice, at 2 cm intervals from the top to the bottom of each core, together with several chemical and biological variables (nutrients, algal pigments, and taxonomic composition of algal assemblages) along the cores. The chemical and biological variables were determined on melted 10-cm thick sections of the cores; using crystallographic information, the chemical and biological data were transformed into values per unit of brine volume. The rate of ice growth for each 10-cm interval of each core was calculated

Table 11.9 Standardized canonical coefficients for the first two canonical variates.

Discriminant variable	Canonical variate 1	Canonical variate 2
Nitrate	–0.63	0.69
Phosphate	0.55	–0.08
Silicate	0.29	0.44
Rate of ice growth	0.89	0.54

by combining the mean daily air temperatures since the start of ice formation with the ice thickness at the date of sampling. Data on taxonomic composition of the algal assemblages in the brine cells were analysed as follows: (1) Similarities (χ^2 similarity; S_{21}, eq. 7.28) were computed among all pairs of core sections, on the basis of their taxonomic composition. (2) The similarity matrix was subjected to flexible clustering (Subsection 8.5.10) to identify groups of core sections that were taxonomically similar. (3) Discriminant analysis was used to determine which environmental variables best accounted for differences among the groups of core sections. Chlorophyll *a* is not a descriptor of the environment but of the ice algae, so that it was not used as discriminant variable; it is, however, the response variable in the path analysis mentioned below. Another approach to this question would have been to look directly at the relationships between the physical and chemical data and the species, using RDA or CCA.

Cluster analysis evidenced five groups among the 10-cm ice sections. The groups were distributed at various depths in the cores, sometimes forming clusters of up to 5 adjacent ice sections from within the same core. Discriminant analysis was conducted on standardized descriptors. The first canonical variate accounted for 62% of the variation among groups, and the second one 29%.

The standardized canonical coefficients for the first two canonical variates (Table 11.9) indicate that the environmental descriptors that best accounted for the among-group variation were the rate of ice growth (first variate) and nitrate (second variate). Figure 11.14 shows the position of the centroids of the 5 groups of core sections, plotted in the space of the first two canonical axes, with an indication of the role played by the environmental variables in discriminating among the groups of core sections. According to the figure, the groups of core sections are distributed along two gradients, one dominated by ice growth rate (with groups 1, 3 and 5 in faster-growing ice) and the other by nitrate (with group 1 in low-nitrate and group 5 in high-nitrate environments). These results are consistent with those of a path analysis (Section 10.4) conducted on the same data, showing that algal biomass (chl *a*) was inversely related to the rate of ice growth. The paper concluded that slower ice growth favoured the colonization of brine cells by microalgae (path analysis) and that the rate of ice growth had a selective effect on taxa, with nutrient limitation playing a secondary role in some brine cells (discriminant analysis).

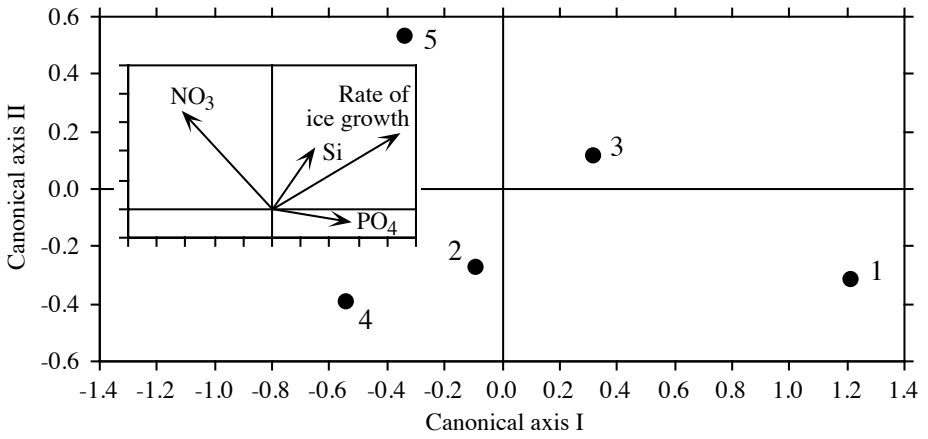


Figure 11.14 Centroids of the five groups of taxonomically similar core sections plotted along the first two canonical axes. Insert: contributions (from Table 11.9) of the four environmental variables (arrows) to the formation of the canonical axes. The groups of core sections are distributed along two gradients, one dominated by ice growth (groups 4, 2 and 3), the other by nitrate (groups 1, 3 and 5). Modified from Legendre *et al*. (1991).

11.4 Canonical correlation analysis (CCorA)

Canonical correlation analysis (CCorA; Hotelling, 1936), differs from redundancy analysis (RDA) in the same way as linear correlation differs from simple linear regression (Box 10.1). In CCorA, the two matrices under consideration are treated in a symmetric way whereas, in RDA, the **Y** matrix is considered to be dependent on an explanatory matrix **X**. The algebraic consequence is that, in CCorA, the matrix whose eigenvalues and eigenvectors are sought (eq. 11.48) is constructed from all four parts of eq. 11.1 whereas, in the asymmetric RDA method, eq. 11.8 does not contain the $\mathbf{S_{YY}}$ matrix.

In CCorA, the objects (sites) under study are described by two sets of quantitative descriptors between which a general form of correlation is sought; for example, a first set $\mathbf{Y}_1$ of p_1 chemical and a second set $\mathbf{Y}_2$ of p_2 geomorphological descriptors of the sampling sites. The dispersion matrix **S** of these $p_1 + p_2$ descriptors contains four sub-matrices, as in eq. 11.1:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S'}_{12} & \mathbf{S}_{22} \end{bmatrix} \qquad \textbf{(11.46)}$$

The data matrices are designated $\mathbf{Y}_1$ and $\mathbf{Y}_2$, instead of $\mathbf{Y}$ and $\mathbf{X}$, to emphasize the fact that the two matrices play equivalent roles in CCorA. Matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are designates by numbers (1, 2) to simplify the writing in the present section. Submatrices $\mathbf{S}_{11}$ (order $p_1 \times p_1$) and $\mathbf{S}_{22}$ (order $p_2 \times p_2$), represent, respectively, the covariance matrices of $\mathbf{Y}_1$ and $\mathbf{Y}_2$, whereas $\mathbf{S}_{12}$ (order $p_1 \times p_2$) and its transpose $\mathbf{S}'_{12} = \mathbf{S}_{21}$ (order $p_2 \times p_1$) represent the covariance matrix between the two sets of descriptors.

Gittins (1985) presents a comprehensive review of the theory and applications of CCorA in ecology. CCorA has limited applications nowadays because many two-matrix problems encountered in ecology are asymmetric and should be analysed by RDA (Section 11.1) or CCA (Section 11.2), whereas symmetric analyses are often conducted using the more flexible method of co-inertia analysis (Section 11.5).

1 — The algebra of canonical correlation analysis

Consider two response data sets $\mathbf{Y}_1$ ($n \times p_1$) and $\mathbf{Y}_2$ ($n \times p_2$), containing different variables about the same objects. They are to be related and compared in an analysis. CCorA does not invoke the directional hypothesis that $\mathbf{Y}_1$ may influence $\mathbf{Y}_2$ or the opposite.

The correlation coefficient between two variables is computed as $r_{jk} = s_{jk}/(s_j s_k)$ (eq. 4.7). Matrix $\mathbf{K}$ is constructed like a correlation coefficient:

$$\mathbf{K} = \mathbf{S}'^{-0.5}_{11}\mathbf{S}_{12}\mathbf{S}^{-0.5}_{22} \qquad \textbf{(11.47)}$$

$\mathbf{K}$ summarizes the correlation structure between data matrices $\mathbf{Y}_1$ and $\mathbf{Y}_2$. In this equation, $\mathbf{S}^{-0.5}_{11}$ is the inverse of the Cholesky root* of $\mathbf{S}_{11}$, and similarly for $\mathbf{S}^{-0.5}_{22}$. $\mathbf{K}$ would be identical if computed from correlation matrices $\mathbf{R}_{11}$, $\mathbf{R}_{12}$ and $\mathbf{R}_{22}$; this is why the same eigenvalues and eigenvectors are found in CCorA based on either covariance (raw data) or correlation matrices (standardized data). In this symmetric analysis, matrix $\mathbf{K}$ would be transposed if computed after inverting the roles of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ in the equation. The algebra in the present section applies equally well to $\mathbf{S}$ matrices defined as matrices of sums of squares and cross products ($\mathbf{S_{YY}} = \mathbf{Y}'\mathbf{Y}$) or dispersion (variance-covariance) matrices ($\mathbf{S_{YY}} = (1/(n-1))\,\mathbf{Y}'\mathbf{Y}$).

The canonical correlation approach consists in maximizing the between-set dispersion with respect to the two within-set dispersions. The expression to be optimized is $\mathbf{S}_{12}\mathbf{S}^{-1}_{22}\mathbf{S}'_{12}\,\mathbf{S}^{-1}_{11}$ since $\mathbf{S}_{12}\mathbf{S}'_{12} / \mathbf{S}_{11}\mathbf{S}_{22}$ does not exist in matrix algebra.

* The Cholesky root of a matrix $\mathbf{A}$ is an upper triangular matrix $\mathbf{L}$ such that $\mathbf{L}'\mathbf{L} = \mathbf{A}$. Cholesky factorization is easier than computing the true square root of $\mathbf{A}$ using eq. 2.29.

Finding solutions to this optimization problem calls for eigenvalues and eigenvectors. Canonical eigenvalues are obtained by solving the characteristic equation:

$$\left|\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1} - \lambda_k\mathbf{I}\right| = 0 \qquad \textbf{(11.48)}$$

In this equation, $\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}$ is an asymmetric matrix, as was that of numerical example 2 in Section 2.9. Its eigenvalues can be used in turn in the following equation, which results from the multiplication of the left and right members of eq. 11.48 by $\mathbf{S}_{11}$:

$$(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}'_{12} - \lambda_k\mathbf{S}_{11})\,\mathbf{v}_k = 0 \qquad \textbf{(11.49)}$$

This equation is used to estimate the matrix of eigenvectors $\mathbf{V} = [\mathbf{v}_k]$ of the *first data set*. Because the analysis is symmetric, the same non-zero eigenvalues are found in the solution of the following equation, which is the dual of eq. 11.48:

$$\left|\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1} - \lambda_k\mathbf{I}\right| = 0 \qquad \textbf{(11.50)}$$

The eigenvalues are now used in the following equation, which results from the multiplication of both sides of eq. 11.50 by $\mathbf{S}_{22}$:

$$(\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}\mathbf{S}_{12} - \lambda_k\mathbf{S}_{22})\,\mathbf{u}_k = 0 \qquad \textbf{(11.51)}$$

This equation is used to estimate the matrix of eigenvectors $\mathbf{U} = [\mathbf{u}_k]$ of the *second data set*. Matrix $\mathbf{U}$ is also the matrix of eigenvectors of $\mathbf{K}'\mathbf{K}$ whereas $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{K}\mathbf{K}'$.

Equations 11.49 and 11.51 cannot be solved using regular eigenvalue decomposition as described in Section 2.9. So in practice, the solution is found by singular value decomposition of $\mathbf{K}$ (SVD, Section 2.11):

$$\mathbf{K}(p_1\times p_2) = \mathbf{V}(p_1\times c)\,\mathbf{W}(\text{diagonal}, c\times c)\,\mathbf{U}'(c\times p_2)^{*} \qquad \textbf{(11.52)}$$

The canonical correlations (r_k) are the singular values found on the diagonal of $\mathbf{W}$. The eigenvalues are the squared singular values, so the diagonal matrix of eigenvalues is:

$$\mathbf{\Lambda} = \mathbf{W}^2$$

The eigenvalues found here are the same as those of eqs. 11.48 to 11.51. The rank of the solution (i.e. the number of canonical axes) is equal to the number, c, of eigenvalues larger than 0, where $c \leq min(p_1, p_2)$.

---

* As explained in Section 2.11, the symbolism for SVD used in this book differs from that of the R language. Matrix $\mathbf{V}$ is component \$u of the output object of R function ***svd()*** while matrix $\mathbf{U}$ is component \$v.

The *canonical coefficients* give the contributions of the two sets of variables to the canonical axes. They are computed as follows:

$$Coeff_{\mathbf{Y}1} = \mathbf{S}_{11}^{-0.5}\mathbf{V} \quad \textbf{(11.53)}$$

$$Coeff_{\mathbf{Y}2} = \mathbf{S}_{22}^{-0.5}\mathbf{U} \quad \textbf{(11.54)}$$

The scores of the objects on the canonical axes form matrices $\mathbf{C}_{\mathbf{Y}1}$ and $\mathbf{C}_{\mathbf{Y}2}$, called the *canonical variates*. These matrices, with order $(n \times c)$, are computed as follows:

$$\mathbf{C}_{\mathbf{Y}1} = \mathbf{Y}_1 Coeff_{\mathbf{Y}1} = \mathbf{Y}_1 \mathbf{S}_{11}^{-0.5}\mathbf{V} \quad \textbf{(11.55)}$$

$$\mathbf{C}_{\mathbf{Y}2} = \mathbf{Y}_2 Coeff_{\mathbf{Y}2} = \mathbf{Y}_2 \mathbf{S}_{22}^{-0.5}\mathbf{U} \quad \textbf{(11.56)}$$

The canonical correlations found above are the Pearson correlations between the object scores in corresponding columns of $\mathbf{C}_{\mathbf{Y}1}$ and $\mathbf{C}_{\mathbf{Y}2}$. The interpoint distances in the canonical space are Mahalanobis distances. Indeed, CCorA of a data set by itself produces identical matrices $\mathbf{C}_{\mathbf{Y}1}$ and $\mathbf{C}_{\mathbf{Y}2}$; in these matrices, the Euclidean distances among objects are the Mahalanobis distances (D_5, eq. 7.38) among the objects in the original data matrices.

The variables of both sets are drawn in the canonical space using correlation matrices computed as follows:

plot variables $\mathbf{Y}_1$ in space $\mathbf{Y}_1$ using $\text{cor}(\mathbf{Y}_1,\mathbf{C}_{\mathbf{Y}1})$ **(11.57)**

plot variables $\mathbf{Y}_2$ in space $\mathbf{Y}_2$ using $\text{cor}(\mathbf{Y}_2,\mathbf{C}_{\mathbf{Y}2})$ **(11.58)**

One could also plot the variables of one set in the space of the other set, although this is rarely done:

plot variables $\mathbf{Y}_1$ in space $\mathbf{Y}_2$ using $\text{cor}(\mathbf{Y}_1,\mathbf{C}_{\mathbf{Y}2})$

plot variables $\mathbf{Y}_2$ in space $\mathbf{Y}_1$ using $\text{cor}(\mathbf{Y}_2,\mathbf{C}_{\mathbf{Y}1})$

These equations explain why analyses based upon unstandardized or standardized descriptors produce CCorA biplots with the same projections of the objects and variables: the projections of the objects are the same because the covariance matrices, which differ between unstandardized and standardized descriptors, are included in eqs. 11.55 and 11.56, and the projections of the variables are computed using correlations (eqs. 11.57 and 11.58), which are the same for unstandardized and standardized descriptors.

2 — Statistics in canonical correlation analysis

Several statistics derived from multivariate analysis of variance (MANOVA) can be used to test the significance of the canonical correlation between two matrices. The most commonly used is Wilks' Lambda likelihood ratio test, which is also used in discriminant analysis (eq. 11.42). These statistics may provide diverging diagnostics especially when they are tested parametrically. Pillai & Hsu (1979) showed that

Pillai's trace

Pillai's trace (V) is quite robust to non-normality; it also performs well in MANOVA. It is computed as follows:

$$V = \text{trace}(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}) = \text{sum of the canonical eigenvalues} \qquad \textbf{(11.59)}$$

Permutation test

For normal data, V can be tested parametrically with reference to the F-distribution. For non-normal data, a permutation test of V is available in VEGAN's function ***CCorA()*** in addition to the parametric test. For that test, the rows of either $\mathbf{Y}_1$ or $\mathbf{Y}_2$ are permuted, $\mathbf{S}_{12}$ is recomputed using the permuted data, and $V = \text{trace}(\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1})$ is recomputed. This operation is repeated a large number of times to obtain the sampling distribution of V under H_0.

When one of the matrices only contains one descriptor ($p_1 = 1$ and $p_2 > 1$ for example), there is only one positive eigenvalue. The canonical correlation problem reduces to the problem of finding the linear combination of variables in $\mathbf{Y}_2$ that is maximally correlated with the single variable $\mathbf{y}_1$; this is simply a problem of multiple correlation (Subsection 4.5.1). The general equation from which eigenvalues are computed in eq. 11.48 simplifies to:

$$\lambda = r^2 = \mathbf{s}_{12}\mathbf{S}_{22}^{-1}\mathbf{s}'_{12} / s_1^2 \qquad \textbf{(11.60)}$$

where $\mathbf{s}_{12}$ is a vector of covariances. This equation corresponds to that of multiple correlation (eq. 4.31), expressed in Chapter 4 in terms of $\mathbf{r}$ instead of $\mathbf{s}$. Finally, when the two sets contain only one descriptor each ($p_1 = p_2 = 1$), eq. 11.60 becomes:

$$\lambda = r^2 = s_{12}s_{22}^{-1}s_{12}s_{11}^{-1} = \frac{(s_{12})^2}{s_{11}s_{22}} = \frac{(s_{12})^2}{s_1^2 s_2^2} \qquad \textbf{(11.61)}$$

which is the formula for the square of the Pearson linear correlation (eq. 4.7). The parametric F-test of Pillai's trace gives the exact same p-value as the test of the Pearson correlation coefficient in that case.

3 — Applications of CCorA

CCorA cannot handle data sets with p_1 or p_2 greater than $(n - 1)$ because the covariance matrices $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ must be inverted (eqs. 11.48 to 11.51). This precludes the analysis of community composition data that contain more species than sites.

The interpretation of canonical correlation analyses is more difficult than that of other multidimensional analyses. The main use of this technique is to test the significance of the correlation between two multidimensional data sets, then explore the structure of the data by computing the correlations (which are the square roots of the CCorA eigenvalues) that can be found between linear functions of two groups of descriptors (Kendall & Stuart, 1966). The detailed (graphical) study of pairs of eigenvectors is usually restricted to the first few canonical correlations, although Blackith & Reyment (1971) give an example taken from Blackith & Albrecht (1959) where the lowest canonical correlations were of interest; the corresponding canonical eigenvectors made it possible to isolate a "phase" vector in locusts which was independent of the "size" vector.

When using CCorA, one should remember that strong canonical correlations do not necessarily mean that the corresponding vectors of ordination scores $\mathbf{C}_{\mathbf{Y}1}$ and $\mathbf{C}_{\mathbf{Y}2}$ explain a large fraction of the variation in $\mathbf{Y}_1$ or $\mathbf{Y}_2$; indeed, strong canonical correlations may be produced between members of a pair of canonical variates that may not explain large portions of the variance of the two data sets. *Redundancy coefficients* are used in CCorA to measure the proportion of the variance of $\mathbf{Y}_1$ (or $\mathbf{Y}_2$) that is explained by a linear combination of the variables in $\mathbf{Y}_2$ (or $\mathbf{Y}_1$); they should always be computed together with canonical correlations to help interpret them (Stewart & Love, 1968).

Ecological application 11.4

The Doubs river data of Verneaux (1973) were described and analysed by variation partitioning in Ecological application 11.1a. In the present example, CCorA is used to compare 3 geographic and topographic (linear distance from the source along the course of the river, slope, mean minimum discharge) to 7 water chemistry variables (pH, hardness, concentrations of phosphate (PO_4), nitrate (NO_3), ammonia (NH_4) and dissolved oxygen (O_2), and biological oxygen demand (BOD)) observed at 30 sites along the main course of the river. A similar analysis is presented by Borcard *et al*. (2011), where one topographic variable differs from the present analysis and some of the variables are pre-transformed to make their distributions more symmetrical.

The data sets individually explain fairly well the variation in fish assemblages along the river: a RDA of the Hellinger-transformed fish abundances by geography and topography (site 8 removed, where no fish were captured) produced $R^2_a = 0.45$, whereas a RDA of fish by chemistry produced $R^2_a = 0.47$. The present analysis will try to determine to what extent the water chemistry variables are correlated with the geographic and topographic variables.

Pillai's trace ($V = 1.54387$) is very highly significant, which shows that there is a significant correlation between the two groups of variables. RDA of geography and topography on chemistry produced a very high R^2_a of 0.49, whereas the opposite analysis, RDA of chemistry by geography and topography, produced a fairly high R^2_a of 0.29. The canonical correlations are high on the first two canonical axes: 0.93 for axis 1 and 0.72 for axis 2.

Two biplots (Fig. 11.15) show the canonical relationships. In both biplots, the sites are clearly divided between sites 1-15 on the left (highest portion of the river), associated with high

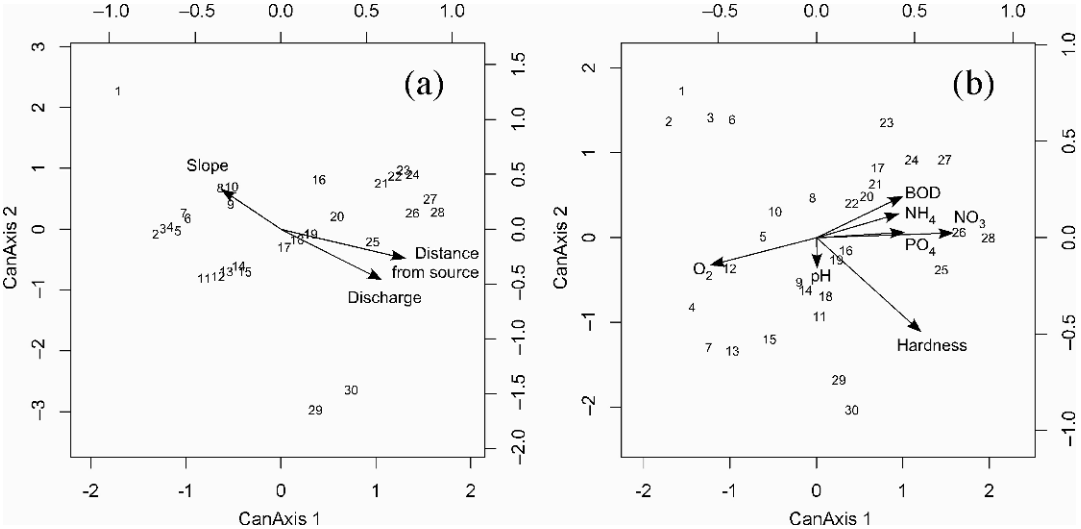


Figure 11.15 (a) CCorA biplot of the sites and the geographic and topographic variables, canonical axes 1 and 2. (b) Biplot of the sites and the water chemistry variables, same canonical axes. Results computed using function ***CCorA()*** of the VEGAN package. The sites are drawn in the graph using the lower and left-hand scales whereas the variables are positioned using the top and right-hand scales.

values of slope in Fig. 11.15a and O_2 in Fig. 11.15b, and sites 16-30 on the right (lowest portion of the river) which are associated with high values of distance from the source and discharge in Fig. 11.15a and BOD, NH_4, PO_4, NO_3 and hardness in Fig. 11.15b.

11.5 Co-inertia (CoIA) and Procrustes (Proc) analyses

Co-inertia analysis (CoIA) is an alternative to canonical correlation analysis,; it was proposed by Dolédec & Chessel (1994) to search for common structures between two data sets describing the same objects.The method is closely related to Procrustes analysis (Proc), described below. CoIA and Proc are symmetric forms of analysis, meaning that they are appropriate when either data set can be equally used as $\mathbf{Y}_1$ or $\mathbf{Y}_2$ in the analysis. This characteristic distinguishes the symmetric canonical ordination methods (CCorA, CoIA, Proc) from the asymmetric methods (RDA, CCA, LDA).

In CoIA, the variables of both data sets are projected onto the axes obtained by eigen-analysis of the cross-set covariance matrix. Various transformations can be used to correctly model the structure in each data set prior to CoIA (Dray *et al*., 2003).

1 — The algebra of co-inertia analysis (CoIA)

Consider two response data sets $\mathbf{Y}_1$ ($n \times p_1$) and $\mathbf{Y}_2$ ($n \times p_2$), containing different variables about the same n objects. The analysis will relate and compare the two sets. Like CCorA, CoIA does not invoke the directional hypothesis that $\mathbf{Y}_1$ may influence $\mathbf{Y}_2$, or the opposite. CoIA is compared to canonical correlation analysis (CCorA) in Subsection 11.5.3.

The analysis for two data sets is conducted as follows:

Total co-inertia

- Compute the covariance matrix crossing the variables of the two data sets. The sum of the squared covariances is the total co-inertia. Compute the eigenvalues and eigenvectors of that matrix. The eigenvalues represent a partitioning of the total co-inertia.

- Project the objects and variables of the two original data sets on the co-inertia axes. Different graphs are produced to compare the projections of the two data sets in the common co-inertia space.

Here is the algebraic development. Co-inertia analysis is implemented in R by the ADE4 function ***coinertia()***, and some of its computational details are mentioned below to clarify what the function does. Compute the matrix of covariances crossing the two data sets $\mathbf{Y}_1$ ($n \times p_1$) and $\mathbf{Y}_2$ ($n \times p_2$):

$$\mathbf{Cov}_{12} = \frac{1}{n-1}\mathbf{Y}'_{1.\text{cent}}\mathbf{Y}_{2.\text{cent}} \qquad \textbf{(11.62)}$$

The notation $\mathbf{Y}_{1.\text{cent}}$ and $\mathbf{Y}_{2.\text{cent}}$ indicates that the two matrices are centred to have column means of 0 before computing $\mathbf{Cov}_{12}$. $\mathbf{Cov}_{12}$ is matrix $\mathbf{S}_{12}$ of Subsection 11.5.1. Compute the singular value decomposition (Section 2.11) of $\mathbf{Cov}_{12}$, with the following result:

$$\mathbf{Cov}_{12}\ (p_1 \times p_2) = \mathbf{V}\ (p_1 \times c)\ \mathbf{W}(\text{diagonal}, c \times c)\ \mathbf{U}'\ (c \times p_2) \qquad \textbf{(11.63)}$$

The value c is defined a few lines down. As mentioned above, the total co-inertia is the sum of the squared covariances in $\mathbf{Cov}_{12}$. It is partitioned among the CoIA eigenvalues, which are the squares of the singular values found on the diagonal of $\mathbf{W}$. So, the diagonal matrix of eigenvalues is:

$$\mathbf{\Lambda} = \mathbf{W}^2 \qquad \textbf{(11.64)}$$

One could have carried out an eigen-decomposition of $\mathbf{Cov}_{12}'\ \mathbf{Cov}_{12} = \mathbf{S}'_{12}\,\mathbf{S}_{12}$ or $\mathbf{Cov}_{12}\mathbf{Cov}_{12}' = \mathbf{S}_{12}\mathbf{S}'_{12}$, instead of a SVD of $\mathbf{Cov}_{12}$: the eigenvalues of these decompositions are the same as those found by squaring the singular values in $\mathbf{W}$; the matrix of eigenvectors of $\mathbf{Cov}_{12}'\ \mathbf{Cov}_{12}$ is matrix $\mathbf{U}$ whereas that of $\mathbf{Cov}_{12}\mathbf{Cov}_{12}'$ is matrix $\mathbf{V}$. The rank of the solution (i.e. the number of co-inertia axes) is equal to the number (c) of eigenvalues larger than 0, where $c \leq \min(p_1, p_2)$.

The objective of co-inertia analysis is to project the objects and variables of the two data sets onto this common multivariate space and compare their positions. This is done as follows.

• To obtain the positions of the objects of $\mathbf{Y}_1$ in the common space, compute $\mathbf{F}_1 = \mathbf{Y}_{1.\text{cent}}\mathbf{V}$, then normalize each column of $\mathbf{F}_1$ to length 1 (eq. 2.7). Multiply the resulting matrix by $\sqrt{n-1}$ and by $\mathbf{\Lambda}^{1/2}$. As a result, column k of $\mathbf{F}_1$ has variance λ_k; this preserves the Euclidean distance among the objects as in PCA scaling type 1. Proceed in the same way for the second data set: compute $\mathbf{F}_2 = \mathbf{Y}_{2.\text{cent}}\mathbf{U}$, then normalize $\mathbf{F}_2$ and multiply by $\sqrt{n-1}$ and by $\mathbf{\Lambda}^{1/2}$. Use the normalized $\mathbf{F}_1$ and $\mathbf{F}_2$ to construct *a single plot* showing the two sets of objects; add arrows going from the representation of each object in $\mathbf{F}_1$ to the representation of the corresponding object in $\mathbf{F}_2$; invert the order of the input data sets to obtain arrows going in the other direction. The objects that have very close representations (i.e. short arrows) in the joint plot contribute more to the co-inertia (overall similarity) between the data sets than objects that are linked by long arrows; see the ecological application below. One may choose to further norm the columns of $\mathbf{F}_1$ and $\mathbf{F}_2$ to variances of 1 and use the resulting matrices for plotting, thus preserving the Mahalanobis distances among objects in the joint plot; this is done in ADE4 function ***coinertia()***, as discussed in the notes below.

• Project the variables of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ onto the canonical axes: draw arrows anchored at the zero-origin of the plot using the coordinates provided by matrices $\mathbf{V}$ for the variables of $\mathbf{Y}_1$ and $\mathbf{U}$ for the variables of $\mathbf{Y}_2$.

Function ***coinertia()*** of ADE4 carries out the computation slightly differently from the description above. The differences are mentioned here to allow comparison between the results obtained with the above equations and those of the ADE4 function.

• In ***coinertia()***, the covariance matrix is computed as $\mathbf{Cov}_{21} = \frac{1}{n}\mathbf{Y}'_{2.\text{cent}}\mathbf{Y}_{1.\text{cent}}$.

• Function ***coinertia()*** was designed to handle data sets where the rows within each set may have different weights, representing for instance different sizes of sampling units or different row sums of abundances in CA. CoIA requires, however, that the weights be the same for $\mathbf{F}_1$ and $\mathbf{F}_2$; the row weights are scaled to sum to 1. In ***coinertia()***, the lengths of the column vectors are computed with these row weights. When the weights are all equal, this amounts to multiplying all values in $\mathbf{F}_1$ and $\mathbf{F}_2$ by $\sqrt{n}$; the column vectors normalized in that way have lengths of $\sqrt{n}$. For equal row weights, this multiplication has no incidence on the joint plot other than changing the numerical scales along the axes of the graph.

• The final normalization of matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ to variances equal to the respective eigenvalues is not done in function ***coinertia()***. With equal row weights, the vectors in $\mathbf{F}_1$ and $\mathbf{F}_2$ are normalized to constant lengths of $\sqrt{n}$ (or variance = 1); this preserves the Mahalanobis distance among objects as in PCA scaling type 2. Compared to scaling type 1, this operation shrinks the objects along axis 1 and stretches them along axis 2. The plot no longer preserves the Euclidean distances among objects, but the two sets of objects are more easily represented into a square plot and the arrows joining the corresponding objects are more easily seen. Co-inertia plots resulting from different pre-treatments of the data sets, e.g. a PCA and a PCoA, are also easier to read.

• Function ***coinertia()*** requires the prior computation of separate ordinations, one for $\mathbf{Y}_1$ and the other for $\mathbf{Y}_2$, using a ***dudi.xxx()*** function. The raw data are included in the ***dudi.xxx()*** output lists. The ***coinertia()*** function retrieves them from these output objects to compute the covariance matrix. For ordinary data sets or community composition data pre-transformed using for instance the Hellinger transformation, use ***dudi.pca()***. To apply a distance function other than the Euclidean distance, use ***dudi.pco()*** which carries out principal coordinate analysis (PCoA).

• Function ***coinertia()*** also produces graphs of the principal axes of the two data sets $\mathbf{Y}_1$ and $\mathbf{Y}_2$. These additional graphs, which are not an essential part of co-inertia analysis, indicate how the principal axes of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are related to the axes of the common co-inertia solution.

Data may have to be transformed prior to co-inertia analysis: standardize the data (eq. 1.12) if a set contains variables expressed in different physical units, or carry out a Hellinger, chord, or chi-square transformation (Section 7.7) for presence-absence or abundance community composition data with many zeros. No transformation is required if the Euclidean distance among objects is to be preserved.

Else, compute distance matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ using distance coefficients appropriate to each type of data (Chapter 7). Carry out a principal coordinate analysis (PCoA, Section 9.3) of each $\mathbf{D}$ matrix and obtain tables of principal coordinates $\mathbf{PC}_1$ and $\mathbf{PC}_2$. If negative eigenvalues are present, retain only the axes corresponding to the positive eigenvalues, or apply a correction method to make all eigenvalues positive (Subsection 9.3.4). Use matrices $\mathbf{PC}_1$ and $\mathbf{PC}_2$ as input into co-inertia analysis.

Note that if the preliminary analysis incorporates vectors of row weights (row weight can be imposed in functions ***dudi.pca()*** and ***dudi.pco()*** of ADE4), CoIA requires that the weights must be the same for both data sets, a condition that precludes using CoIA with the results of two correspondence analyses (CA). Indeed, CA is a weighted regression method, and the row weights differ between data sets since they depend on the data in each matrix. Applying one or the other vector of weights to both data sets would lead to different CoIA solutions, and there seems to be no logical way of deciding between these two solutions. Co-correspondence analysis (ter Braak & Schaffers, 2004) offers a way to handle that problem; it is available in function ***cocorresp()*** of package COCORRESP. Another way of producing an analysis of two community composition data sets that preserves chi-square distances (eq. 7.55) is to apply a chi-square transformation (eq. 7.70) to each table and use them as input into CoIA; with ADE4, pre-process these tables using ***dudi.pca()***.

RV coefficient

Two overall statistics of co-inertia are available. The first one is the *RV* coefficient (Escoufier, 1973; Robert & Escoufier, 1976), which is a multivariate generalization of the Pearson correlation coefficient; for two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, $RV(\mathbf{x}_1,\mathbf{x}_2) = \text{cor}(\mathbf{x}_1,\mathbf{x}_2)^2$. The second one is the Procrustes statistic described in the next subsection. Coefficient *RV* is computed as follows for two rectangular data matrices with corresponding objects as rows, centred to column means of 0:

$$RV(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{trace\,(\mathbf{Y}_1\mathbf{Y}'_1\mathbf{Y}_2\mathbf{Y}'_2)}{\sqrt{trace\,(\mathbf{Y}_1\mathbf{Y}'_1\mathbf{Y}_1\mathbf{Y}'_1)\;trace\,(\mathbf{Y}_2\mathbf{Y}'_2\mathbf{Y}_2\mathbf{Y}'_2)}} \qquad \textbf{(11.65)}$$

The *RV* coefficient can also be computed from the matrices of sums of squares and cross-products $\mathbf{SS}_{12} = \mathbf{Y}_1'\mathbf{Y}_2$, $\mathbf{SS}_{11} = \mathbf{Y}_1'\mathbf{Y}_1$, and $\mathbf{SS}_{22} = \mathbf{Y}_2'\mathbf{Y}_2$:

$$RV(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{sum\,(\mathbf{SS}_{12}^{(2)})}{\sqrt{sum\,(\mathbf{SS}_{11}^{(2)})\; sum\,(\mathbf{SS}_{22}^{(2)})}} \qquad \textbf{(11.66)}$$

where the notation $\mathbf{SS}_{12}^{(2)}$ means that each element of $\mathbf{SS}_{12}$ is squared before summation. Significance of the *RV* coefficient can be tested by permutation (Heo & Gabriel, 1998) or parametrically (Josse *et al*., 2008). The null hypothesis of the test is: the two data sets are no more related than random data sets would be; this is the same kind of null hypothesis as in correlation analysis.

Permutation test

Ecological application 11.5, part 1

Oribatid mites (Acari: Oribatida) are a very diversified group of small (0.2 to 1.2 mm) soil-dwelling arthropods. In June 1989, Daniel Borcard collected 70 soil cores from the peat blanket of a bog lake on the territory of *Station de biologie des Laurentides* of Université de Montréal, Québec, Canada. During the following weeks, he extracted, identified, and counted 9800 individuals found therein, which he separated into 35 morphospecies. The mite and environmental data are fully described in a paper by Borcard & Legendre (1994); the data set is available in the VEGAN and ADE4 R packages. For the present ecological application, Dr. Borcard divided the species into a group of 23 panphytophagous species, which eat vegetation debris (most of their regime) as well as algae, fungi, spores, pollen grains, and bacteria, and 12 microphagous species which feed mostly on algae, fungi, spores, and pollen grains. The presence of vegetation debris in the diet of the panphytophagous species differentiates the two groups, which were almost equally represented in the data sets: there were 5667 panphytophagous and 4133 microphagous individuals.

The two data sets, which were considered to represent different communities, were Hellinger-transformed (eq. 7.69) separately, then subjected to co-inertia analysis using the ADE4 function ***coinertia()***. All row weights were equal. The *RV* coefficient (*RV* = 0.36038) was highly significant (p = 0.0001 after 9999 permutations), indicating a strong relationship between the two data sets. The first two canonical axes represented respectively 83% and 10% of the co-inertia. Figure 11.16a shows the 70 sites from the two data sets projected in the co-inertia space and linked by arrows (tail of each arrow = panphytophagous, head = microphagous). The low-numbered sites were physically located near the margin of the forest surrounding the bog lake whereas the high-numbered sites were near the free water portion of the bog. In Fig. 11.16a, most sites are equally distant from the origin of the plot and their arrows have about the same lengths, showing that they contribute fairly equally to the total co-inertia; only a few sites, with high site numbers, have long arrows. Panels b and c show that most of the 12 microphagous species contribute to the dispersion of the sites in the co-inertia plane (long arrows), whereas only 4 or 5 of the 23 panphytophagous species contribute strongly to that dispersion.

Two mite species associations were identified by Legendre (2005). It is interesting to note that the species with fairly long arrows (important contributions) in Fig. 11.16 (panels b and c), found in the same quadrant of the two species projections, also belong to the same species association: species 9, 31, 34 and 35 in quadrant 1, species 1, 10 and 15 in quadrant 2, species 13, 14 and 27 that point left along axis 1, species 7 and 11 in quadrant 3, and species 16 and 23

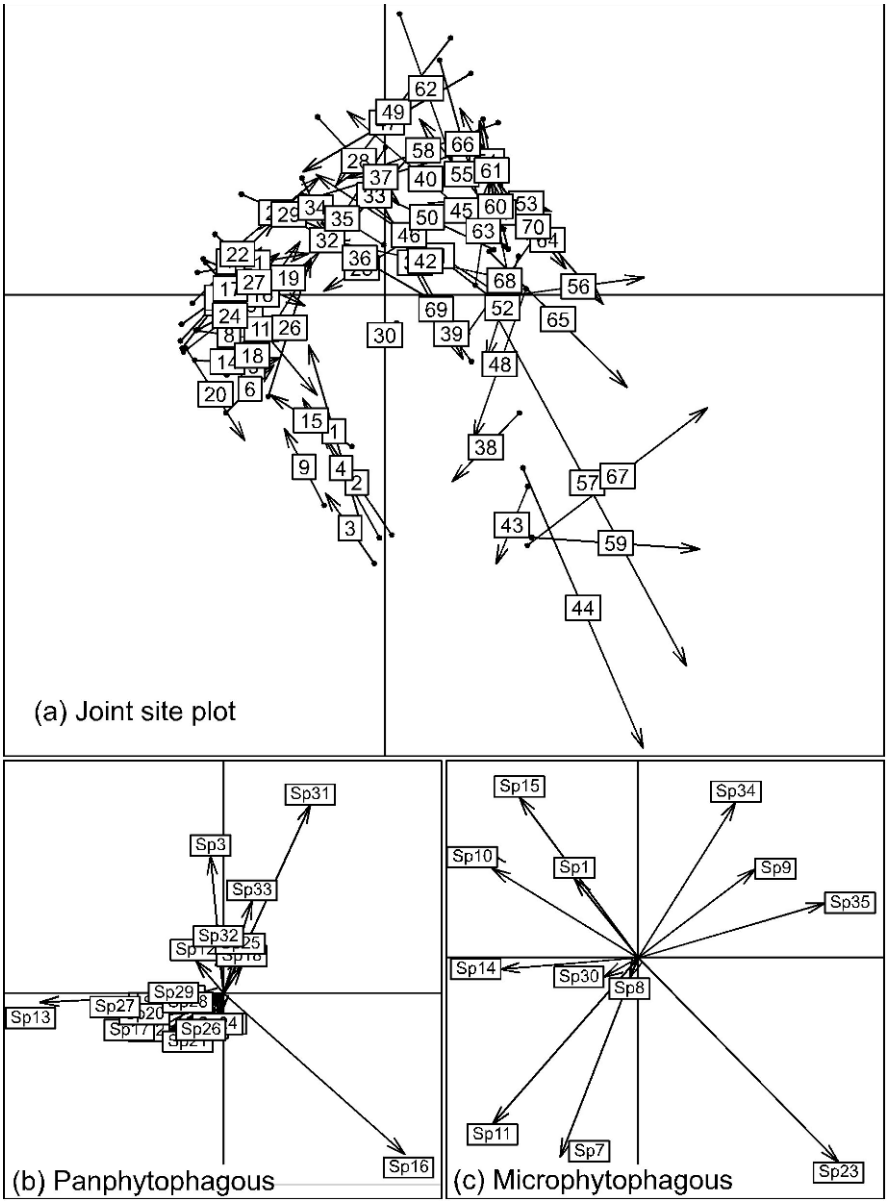


Figure 11.16 Results of co-inertia analysis of the mite data. (a) Joint site plot for the two sets of objects represented in the plane of canonical axes 1 and 2, which accounts for 93% of the total co-inertia. Arrows link objects from set 1 (panphytophagous species at the arrow tails) to the corresponding objects from set 2 (microphagous species at the arrow heads). (b) Projection of the panphytophagous and (c) of the microphagous species onto the co-inertia plane. Figure produced by the ***plot.coinertia()*** function of ADE4, then modified using a graphics editor.

in quadrant 4. In addition, a majority of the panphytophagous species point toward the left-hand site of their plot (panel b), where, in panel a, are found the sites that are closer to the forest and contain coarser vegetation debris.

Interestingly, site-species associations are also identifiable in Fig. 11.16. For example, species 16 (in panel b) and 23 (in panel c) dominate the communities in sites 38, 43, 44, 59 and 67. These site-species associations can be verified in the matrices of Hellinger-transformed mite data.

It is also interesting to look at arrow directions in panel (a): for most of the low-numbered sites (with the exceptions of sites 1 to 4, 9 and 15), the arrows point from left to right, indicating that the panphytophagous species put these sites at larger distances from the other sites than the microphytophagous species do. So for these sites, the panphytophagous species contribute more to beta diversity than the microphytophagous species. Indeed, for sites 5-8, 10-14 and 16-29, the sum of the variances of the Hellinger-transformed data, which is a measure of beta diversity (Legendre *et al*., 2005), is 0.25638 for the panphytophagous and 0.13374 for the microphytophagous species. This method of calculation produces a value of 1 when the sites have completely different species compositions; this corresponds to the situation where beta diversity is maximum (Subsection 6.5.3).

Co-inertia analysis is appropriate to compare pairs of data sets that play equivalent roles in the analysis. The method finds a common space onto which the objects and variables of these data sets can be projected and compared. The analysis may, for example, concern two segments of the species forming an ecological community (as in Ecological application 11.5, part 1). One could also compare data sets representing the physical or biological characteristics of organisms (individuals, species) to their behavioural characteristics. Moretti & Legg (2009) used co-inertia analysis to compare plant and invertebrate animal functional traits across forest sites with different fire and cutting histories. Several other examples are described in Dray *et al*. (2003). Compared to CCorA, co-inertia analysis imposes no constraint regarding the number of variables in the two sets, so that it can be used to compare ecological communities even when they are species-rich; see the comparison of methods in Subsection 11.5.3.

Co-inertia analysis is not well-suited, however, to analyse pairs of data sets that contain the same variables, because the analysis does not establish one-to-one correspondences between variables in the two data sets; the method does not 'know' that the first variable is the same in the first and the second data sets, and likewise for the other variables. Data of that type are found in before-after (BA) or control-impact (CI) studies. When the two data sets contain most or all of the same species, they can be analysed by placing the data 'before' on top of those 'after' in a joint data file, and computing a PCA of the combined data; the before-after pairs can then be linked by arrows in the common PCA ordination graph. Else, the difference between the two sections of the data table can be tested by RDA for the effect of a 'before-after' factor, in the presence of covariables representing the pairing of the sampling sites; see *Analysis of related samples* in Subsection 11.1.10, point 3. For data of that type, the RDA test has greater power to detect a difference than a co-inertia test because it uses the information more efficiently. The null hypothesis of the RDA test is H_0: there is no difference between 'before' and 'after' for data described by the same variables,

whereas the null hypothesis in CoIA is H_0: the two data sets have no more co-inertia structure than random data sets would have, without any reference to the variables being the same in the two data sets.

Multiple factor analysis

Multiple factor analysis (MFA) can be used to compare several data sets describing the same objects (Escofier & Pagès, 1994). MFA consists in projecting objects and variables of two or more data sets on a global PCA, computed from all data sets, in which the sets receive equal weights. For the comparison of two data sets, the algebra of MFA differs from that of CoIA. This method is implemented in functions ***mfa()*** of ADE4 and ***MFA()*** of FACTOMINER; the latter offers more options. A summary of the theory as well as an ecological application are presented in Section 6.10 of Borcard *et al*. (2011).

2 — Symmetric Procrustes analysis (Proc)

Orthogonal Procrustes analysis

Co-inertia analysis is closely related to the orthogonal Procrustes analysis of two data sets. The orthogonal Procrustes problem was first formulated by Hurley & Cattell (1962) who called their computer program PROCRUSTES after the villain of Greek mythology*; later authors referred to the method by that name. The problem consists in finding the best superposition of two sets of corresponding objects (i.e. the *n* objects of the two sets are the same) by rotation and mirror reflection, if necessary, of one of the data sets with respect to the other, in such a way as to minimize the sum of squared distances between the corresponding objects. A general least-squares solution was described by Schönemann (1966) and Schönemann & Carroll (1970) and perfected by Gower (1971b). The Procrustes rotation solution can be asymmetric, meaning that one matrix is rotated to maximum fit while the other is kept fixed, or symmetric in the sense described below.

Asymmetric Procrustes rotation

In symmetric Procrustes rotation, described here, each of the two data sets, $\mathbf{Y}_1$ ($n \times p_1$) and $\mathbf{Y}_2$ ($n \times p_2$), is standardized to have its total variance equal to 1 prior to rotation. This is obtained by Gower's standardization (Gower, 1971b), which consists in dividing each value in a column-centred data matrix by the square root of the total variance of the matrix, which is also the square root of the sum of its eigenvalues. The covariance of the two Gower-standardized matrices, $\mathbf{Y}_{1.\text{Gower}}$ and $\mathbf{Y}_{2.\text{Gower}}$, is then computed using the same covariance formula as in co-inertia analysis (Subsection 11.5.1):

$$\mathbf{Cov}_{12} = \frac{1}{n-1}\mathbf{Y}'_{1.\text{Gower}}\mathbf{Y}_{2.\text{Gower}} \qquad \textbf{(11.67)}$$

Note that co-inertia analysis (Subsection 11.5.1) of two Gower-standardized matrices produces the same relative eigenvalues, *RV* coefficient, and plots, as the CoIA of the

* For a brief description of the story of Procrustes in Greek mythology, see the first paragraph of Subsection 10.5.4.

original data sets. Procrustes analysis differs from co-inertia analysis in that it uses different output matrices for the joint plot of the two sets of objects.

The singular values of $\mathbf{Cov}_{12}$ are computed by singular value decomposition (Section 2.11):

$$\mathbf{Cov}_{12}\ (p_1 \times p_2) = \mathbf{V}\ (p_1 \times c)\ \mathbf{W}(\text{diagonal}, c \times c)\ \mathbf{U}'\ (c \times p_2) \qquad \textbf{(11.68)}$$

and the trace of **W** is computed:

$$\text{Trace}\mathbf{W} = \sum(\text{singular values}) \qquad \textbf{(11.69)}$$

The singular values are the diagonal values of **W**. Because singular values are positive or null, Trace**W** is non-negative. The Procrustes residual sum-of-squares statistic (Gower, 1971b, 1975*; Davis, 1978) is

$$m_{12}{}^2 = 1 - \text{Trace}\mathbf{W}^2 \qquad \textbf{(11.70)}$$

Following that, the rotation matrix that provides the best adjustment of the objects of $\mathbf{Y}_2$ to the objects of $\mathbf{Y}_1$ is computed as:

$$\mathbf{H} = \mathbf{U}\mathbf{V}' \qquad \textbf{(11.71)}$$

The rotated matrix $\mathbf{Y}_{2.\text{rot}}$ is computed as follows:

$$\mathbf{Y}_{2.\text{rot}} = \text{trace}\mathbf{W}\ \mathbf{Y}_{2.\text{Gower}}\ \mathbf{H} \qquad \textbf{(11.72)}$$

Symmetric Procrustes rotation

where trace**W** acts as a scaling factor. The Procrustes analysis is called symmetric when the two data sets are subjected to Gower standardization; it remains asymmetric in the fact that $\mathbf{Y}_{2.\text{Gower}}$ is projected after optimal rotation ($\mathbf{Y}_{2.\text{rot}}$) onto $\mathbf{Y}_{1.\text{Gower}}$. Objects are plotted on a graph using matrices $\mathbf{Y}_{1.\text{Gower}}$ and $\mathbf{Y}_{2.\text{rot}}$. It is thus more interesting (but not compulsory) in most instances to start Procrustes analysis with a matrix $\mathbf{Y}_{1.\text{ord}}$ that represents an ordination of $\mathbf{Y}_1$, e.g. by PCA or PCoA. Differences in positions between corresponding objects of the two data sets can be interpreted as in co-inertia analysis (Subsection 11.5.1).

Permutation test

For permutation testing, one can use either the $m_{12}{}^2$ statistic, or its complement the Procrustes $R^2 = (1 - m_{12}{}^2) = \text{Trace}\mathbf{W}^2$, or else Trace**W**. The latter is a Procrustean form of the correlation coefficient; its value is always positive or null. This permutation test, called PROTEST (Jackson, 1995; Peres-Neto & Jackson, 2001), is available in VEGAN's ***protest()*** function described in Subsection 10.5.4. It tests the same hypothesis as the test of the *RV* coefficient in CoIA. Two-matrix (or *Classical*) Procrustes rotation has been extended to *m* matrices in *Generalized Procrustes analysis* (Gower, 1975).

Generalized Procrustes analysis

* Gower called the residual sum of squares statistic R^2 in 1971b and $m_{12}{}^2$ in his 1975 paper.

Symmetric Procrustes analysis is appropriate for the same types of questions as co-inertia analysis, the difference between the two methods residing in the matrices used to plot the objects. Likewise, the situations where symmetric Procrustes analysis is inappropriate are the same as for CoIA analysis. Procrustes analysis is also appropriate to compare the results of ordinations derived from two sets of distances, for example the distances computed among sets of morphological landmarks (which form the data rows) measured on two organisms, or ordinations obtained by different methods, e.g. PCA and CA of the same data; no test of significance is possible in that case, however, because the original data are the same in the two ordinations.

Ecological application 11.5, part 2

The mite data used to illustrate co-inertia analysis (Subsection 11.5.1) were subjected to a symmetric Procrustes rotation using VEGAN's function ***procrustes()***. Principal components of the Hellinger-transformed panphytophagous species were compared to the rotated (Hellinger-transformed) microphagous species. The value of Trace**W**, which is used as the test statistic in PROTEST (Subsection 10.5.4), was 0.53994; the test was highly significant (p = 0.0001 after 9999 permutations). Figure 11.17 illustrates the symmetric Procrustes rotation results. Note that the axes of the graph are not the canonical axes of the co-inertia analysis, Fig. 11.16. The rotated vectors (species in this example) of the microphagous data set are shown as crossed hairs with numbers in the centre of the plot.

3 — Canonical correlation, Procrustes, or co-inertia analysis?

Which method should be used for symmetric analysis of two data sets? Table 11.10 compares the properties and requirements of canonical correlation analysis, on the one hand, to those of co-inertia and Procrustes analyses, on the other hand. A particularly interesting feature of CoIA and Proc, compared to CCorA, is that they allow the joint analysis of two community composition data sets with more species than there are sites. In addition, as in PCA, collinearity among the variables in one or the other data sets produces no problem in CoIA and Proc. This is not the case in CCorA where collinearity may prevent the computation of the inverses of the covariance matrices of the separate data sets, $\mathbf{S}_{11}^{-1}$ and $\mathbf{S}_{22}^{-1}$.

Note that CoIA carried out between two sets of principal components or principal coordinates is not equivalent to CCorA of these same matrices: the eigenvalues and eigenvectors of the two-table analyses are not the same. The reason is that $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ are diagonal matrices of eigenvalues in that case, not identity matrices $\mathbf{I}$, so that the matrix $[\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}]$, which is decomposed in CCorA (eq. 11.48), is not equal to matrix $[\mathbf{S}_{12}\mathbf{S}'_{12}]$ which is subjected to eigen-decomposition in CoIA.

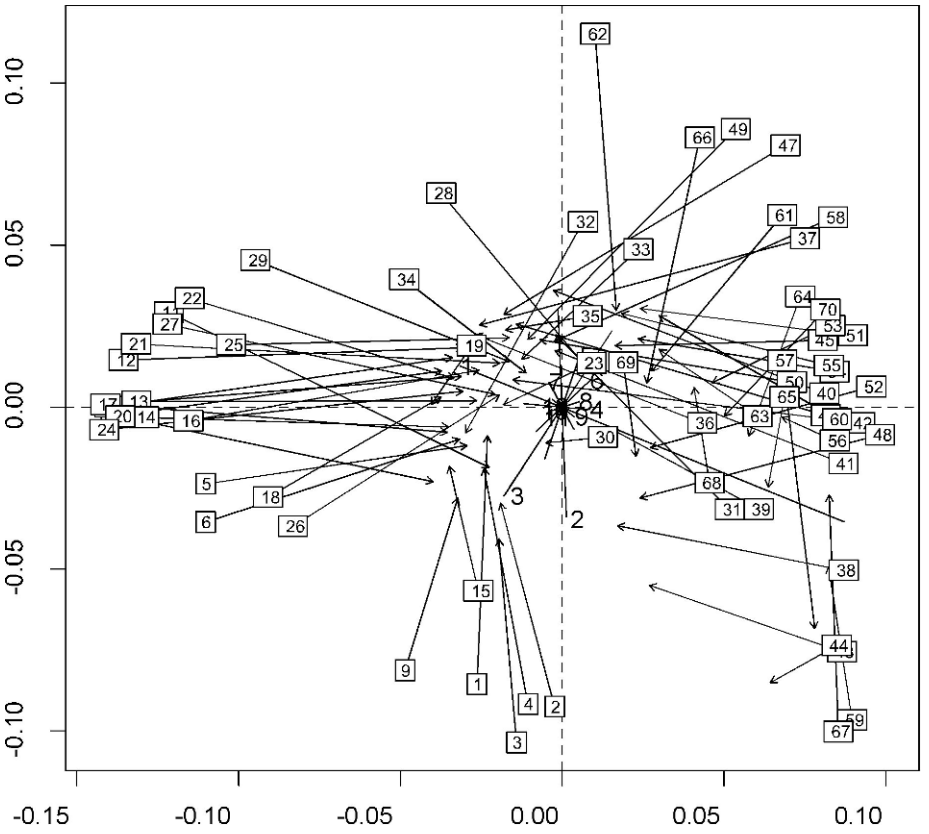


Figure 11.17 Results of symmetric Procrustes rotation of the mite data showing the two sets of objects along axes 1 and 2. Arrows link objects (numbers in boxes) from set 1 (principal components of the panphytophagous species at the arrow tails) to the corresponding objects from set 2 (microphagous species at the arrow heads). The plot was produced by VEGAN's function ***plot.procrustes()***.

11.6 Canonical analysis of community composition data

In early numerical ecology papers, canonical correlation analysis and discriminant analysis were used to analyse tables of species presence/absence or abundance data. In many applications, however, the assumptions of linearity and the algebraic constraints imposed by the models make these methods unsuitable for such data. RDA and CCA provide alternatives that are often more appropriate. Let us consider different types of situations that may involve species data.

Table 11.10 Comparison of canonical correlation analysis (CCorA), on the one hand, to co-inertia analysis (CoIA) and symmetric Procrustes analysis (Proc), on the other.

	CCorA	CoIA, Proc
Matrix sizes	p_1 and $p_2 < n$	No constraint
Physical dimensions	The variables in each set are standardized in CCorA	All variables in each set must have the same physical dimensions*
Eigen-analysis of …	$\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}'_{12}\,\mathbf{S}_{11}^{-1}$	$\mathbf{S}_{12}\mathbf{S}'_{12}$ (or SVD of $\mathbf{SS}_{12}$)
The canonical axes …	Maximize squared correlations among sets	Maximize squared covariances (co-inertia) among sets
Distances among objects preserved	Mahalanobis distances**	Euclidean distances
Test of significance of the relationship	Statistic: Pillai's trace, etc.	Statistics: RV, Trace$\mathbf{W}$

* Standardize (eq. 1.12) the variables of data sets that have heterogeneous physical dimensions. This can be done before the analysis, or by the computer programs performing the analysis.

** Mahalanobis distances among points (eq. 7.38) are preserved in CCorA. They are equal to Euclidean distances computed from matrix $\mathbf{G}$ used to position the objects in PCA scaling 2 (correlation biplot). In this scaling, the ordination axes are stretched to account for the correlations among variables (Subsection 9.1.4).

1. *Species in* $\boldsymbol{Y}$. — The first case involves a matrix $\mathbf{Y}$ of species presence-absence or abundance data and a matrix $\mathbf{X}$ of habitat characteristics. One may wish to find support for the ecological hypothesis of environmental control of the species distributions (Whittaker, 1956; Bray & Curtis, 1957), and/or describe in what way the species are related to the environmental variables. The analysis is not symmetric; the species clearly form the response variables, to be explained by the environmental variables. Hence a symmetric form of analysis such as CCorA or CoIA is not appropriate; one should rely instead on the asymmetric forms of analysis, RDA and CCA.

In some applications, it is interesting to compare two groups of species found together at a series of sampling sites. There are often more species than sites when these analyses involve species-rich communities. CCorA is unable to analyse such data because it cannot handle more variables in any one of the data sets than there are sites minus 1. Rare species would have to be dropped from the analysis to satisfy the requirements of the method. Co-inertia and Procrustes analyses do not have this limitation and can be used for this type of analysis.

When **X** contains dummy variables describing e.g. types of habitat (qualitative variable) recoded as in Subsection 1.5.7, RDA or CCA may be used to test the hypothesis that groups of sites, identified *a priori*, do not differ in species composition. The question is of the same type as in multivariate analysis of variance. Likewise, when **X** codes for factors of an experiment, RDA or CCA may be used for analysis.

2. *Species presence-absence*. — Ecologists may wish to use the environmental variables in **X** to forecast a classification criterion (**y**) representing the presence or absence of a single species, a group of species, or a functional trait at various locations.

Linear discriminant analysis (LDA) is a suitable choice to separate the sites where the species is present from those where it is absent. The classification functions (eq. 11.39) will attribute the observations to the rightful group if the descriptors allow it. Another appropriate statistical model is logistic regression (Subsection 10.3.7) because the forecasted response is binary, 0 (absence) or 1 (presence).

Between linear discriminant analysis (Section 11.3) and logistic regression (Subsection 10.3.7), which method is the most appropriate? Efron (1975) has shown that when the groups are drawn from populations with multinormal distributions in the space of the explanatory variables, discriminant analysis is more effective than logistic regression. On the other hand, logistic regression is more robust than discriminant analysis to departures from multivariate normality. This finding is important for the analysis of species with unimodal distributions along environmental gradients (Subsection 9.2.4): a species may be absent *under both low and high values* of an environmental variable. One can plot scatter diagrams of the presence/absence of the species of interest against each of the environmental variables in matrix **X**. When a unimodal response is detected, a quadratic [orthogonal] polynomial function of that explanatory variable should be used in the logistic model (see Gaussian logistic response, Subsection 10.3.7). The multivariate dispersions of groups of observations representing the presence and absence of a species in the space of an explanatory variable to the powers 1 and 2 (or 1, 2 and 3) cannot be multivariate normal. So in this situation, Gaussian logistic regression should be preferred to discriminant analysis.

3. *Indicator species*. — Species may represent the explanatory variables (matrix **X**). What are the species that characterize different types of habitat? In such cases, the types of habitat form the classification criterion **y**. The method of choice is indicator species analysis (Subsection 8.9.3).

Discriminant analysis is ill-adapted to this type of problem because it requires that there be more objects (n) than descriptors (m) in **X**. Actually, Williams & Titus (1988) recommend that the total number of observations *per group* be at least three times the number of variables in **X**; ter Braak (1987c) recommends that n be much larger than the number of species (m) plus the number of groups (g).

4. *Inverse analysis*. — RDA may be used as a form of inverse analysis to relate species assemblages to types of habitat. The classification criterion (e.g. types of habitat) is represented by a set of dummy variables, written into response matrix **Y**.

The species data are the explanatory variables **X**; they should most likely be transformed using one of the transformations of Section 7.7. The condition of more objects (n) than species (m) must be satisfied in this analysis. For species-rich communities, a solution may be to replace the abundances of individual species by the abundances of species associations identified using an appropriate statistical method; see Section 8.9.

Matrix **X**, in which each species is represented by a vector, may be transformed prior to RDA or discriminant analysis, by replacing the m species vectors by m ordination axes produced by PCA of the transformed species data, or by $(m - 1)$ ordination axes obtained by CA of the raw species data. An alternative is to compute a similarity or distance matrix among sites using the species data and obtain new axes by principal coordinate analysis (PCoA). PCA, CA or PCoA axes might relate to the environmental descriptors better than the original species data. The solution using PCoA is implemented in the CAP method of Anderson & Willis (2003).

One may wish to use the species data in **X** to predict or reconstruct one or more environmental variables in **Y**. This case, which is related to Ecological application 11.2b, is like CCA but with **X** and **Y** interchanged. A solution, which circumvents the too-many-species-problem, is Weighted Averaging Partial Least Squares (WA-PLS), which extends PLS regression in the correspondence analysis framework (ter Braak, 1995).

11.7 Software

Among the methods of canonical analysis, commercial statistical packages usually offer canonical correlation analysis and linear discriminant analysis. RDA and CCA are available in CANOCO* as well as in other packages, in particular PC-ORD and SYN-TAX 2000†.

The R language offers functions for all methods described in this chapter:

1. Redundancy analysis (RDA). — Simple and partial RDA is available in function ***rda()*** in VEGAN, and in package RDATEST found on the Web page http://numericalecology.com/rcode. Selection of explanatory variables in RDA: ***ordistep()*** and ***ordiR2step()*** in VEGAN, ***forward.sel()*** in PACKFOR. Principal response curves are computed by function ***prc()*** in VEGAN. Function ***varpart()*** is available in VEGAN for variation partitioning by RDA. db-RDA: function ***capscale()*** in VEGAN

* CANOCO is available from Plant Research International, Wageningen, The Netherlands. http://www.canoco.com/.

† PC-ORD (http://www.pcord.com) is available from MjM Software, P.O. Box 129, Gleneden Beach, Oregon 97388, USA. SYN-TAX 2000 (http://ramet.elte.hu/~podani) is available from Exter Software, 47 Route 25A, Suite 2, Setauket, New York 11733-2870, USA.

offers db-RDA based on any of the distance functions in ***vegdist()***. Multivariate analysis of variance: function ***manovRDa()*** for two-way crossed-factor MANOVA, for fixed or random factors, is available on the Web page http://www.elaliberte.info/. A similar function ***anova.2way.unbalanced()*** for two fixed factors, balanced or unbalanced designs, is available on the Web page http://numericalecology.com/rcode; type III sums-of-squares are used in the analysis of unbalanced designs. Function ***nested.anova.dbrda()*** for nested MANOVA with two levels (the main factor and one nested factor) is available in the BIODIVERSITYR package.

2. Canonical correspondence analysis (CCA). — Simple and partial CCA: function ***cca()*** in VEGAN, ***cca()*** in ADE4. Function ***CCA()*** is available on the Web page http://numericalecology.com/rcode; this function was written to demonstrate the CCA algorithm described in Subsection 11.2.1; it is fully functional for calculation of CCA and plotting triplots, but tests of significance are not available in that function.

3. Linear discriminant analysis (LDA). — ***lda()*** in MASS, ***discrimin()*** in ADE4. Test of homogeneity of multivariate dispersions: ***betadisper()*** in vegan. Selection of explanatory variables in LDA can be carried out with function ***stepclass()*** of package KLAR (direction = "forward", "backward" or "both").

4. Canonical correlation analysis (CCorA). — Functions ***cancor()*** of STATS, ***CCorA()*** of VEGAN, and ***cc()*** of CCA.

5. Co-inertia (CoIA) and Procrustes (Proc) analyses. — Co-inertia analysis in ***coinertia()*** of ADE4. Test of the *RV* coefficient: permutational test in ***RV.rtest()*** of ADE4, parametric test in ***coeffRV()*** of FACTOMINER. One can also apply function ***randtest.coinertia()*** to the output of function ***coinertia()***. Asymmetric and symmetric Procrustes analysis in ***procrustes()*** of VEGAN and ***procustes()*** of ADE4. Permutation test of the Procrustes statistic: function ***protest()*** of VEGAN. Co-correspondence analysis in ***cocorresp()*** of package COCORRESP. Multiple factor analysis (MFA) in ***mfa()*** of ADE4 and ***MFA()*** of FACTOMINER; the latter offers more options.

6. Miscellaneous methods. — Functions for palaeoenvironmental reconstruction, in particular ***MAT()***, ***MLRC()***, ***WA()*** and ***WAPLS()***, are available in package RIOJA. QR decomposition is computed by ***qr()*** of BASE; this is an efficient computation method for regression coefficients in linear models, e.g. in RDA.

PERMANOVA (permutational ANOVA/MANOVA) is an add-on package for PRIMER 6* that carries out permutational multivariate analysis of variance. This program tests the simultaneous response of one or more variables to one or more factors in an ANOVA experimental design on the basis of any distance measure, using permutation methods. The latest version of the program can handle any balanced ANOVA design up to nine factors.

* Available from PRIMER-E Ltd., 3 Meadow View, Lutton, Ivybridge, PL21 9RH, England.