History of canonical correspondence analysis
C. J. F. ter Braak

Please cite this publication as follows:

ter Braak, C. J. F. 2014. History of canonical correspondence analysis. Pages 61-75 in J. Blasius and M. Greenacre, editors. Visualization and verbalization of Data. Chapman and Hall/CRC, London.

You can see the published version at:
http://www.crcpress.com/product/isbn/9781466589803
and
http://books.google.nl/books?id=qLu_kgEACAAJ

This version of the paper uses notation that follows ter Braak (1986), whereas the book version uses uniform notation for consistency of the book. The mnemonic for the book version is that **t** refers to taxon/taxa and **s** refers to site/sites.

The main correspondence list is

meaning	here	Published version
Response data table	**Y**	**Y** and**,** in relative form, **P**
Row totals as diagonal matrix	**R**	$\mathbf{D_r}$
Column totals as diagonal matrix	**K**	$\mathbf{D_c}$
Explanatory variables table	**Z**	**X**
Species score, taxon score	**u**	**t**
Constrained site score	**x**	**s**
Site score	$\mathbf{x}^*$	$\mathbf{s}^*$
Canonical weight	**b**	**b**

The history of canonical correspondence analysis

Cajo J. F. ter Braak

Biometris,

Wageningen University and Research Centre,

the Netherlands

E-mail: `cajo.terbraak@wur.nl`

1. Introduction

Canonical correspondence analysis (CCA) was introduced in ecology by ter Braak (1986) as a new multivariate method to relate species communities to known variation in the environment. The CARME 2011 conference celebrated not only 50 years of correspondence analysis (CA) but also the 25th anniversary of CCA. In this period, four CCA papers (ter Braak 1986, 1987, 1988c, ter Braak and Verdonschot 1995) have been cited more than 3000 times in Web of Science, ~10% of which in 2010, and 6% (=180) outside ecology. Jean-Dominique Lebreton and Daniel Chessel (JDL-DC) independently invented CCA (Chessel et al. 1987, Lebreton et al. 1988a, Lebreton et al. 1988b).

CCA extends correspondence analysis (CA) with predictor variables. If CA is applied to the $n \times m$ matrix $\mathbf{Y}$ ($y_{ik} \geq 0$), CCA treats this matrix as a matrix of multivariate responses and requires a second $n \times p$ matrix $\mathbf{Z}$ with predictor variables (columns of $\mathbf{Z}$). CCA integrates CA and regression analysis. As in regression analysis, response and predictors must be measured at the same set of n sites, and predictors can be both quantitative, binary and nominal, whereby nominal ones are converted into sets of indicator (1/0) variables. In ecology, $\mathbf{Y}$ typically contains the species data with y_{ik} the presence-absence (1/0) or abundance (or another related nonnegative index such as biomass) of species k in site i and $\mathbf{Z}$ contains environmental variables with z_{ij} the measurement of environmental variable j in site i. Whereas the common procedure for 'indirect gradient analysis'(Prodon and Lebreton 1981, ter Braak and Prentice 1988)
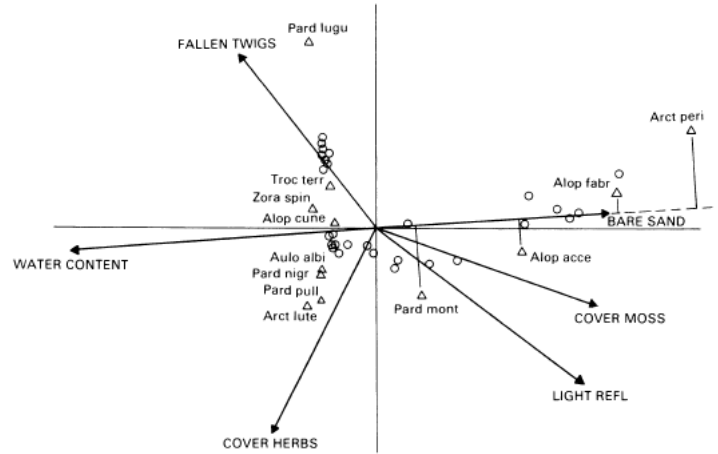
Figure 1: Ordination diagram of CCA (factorial plane) for 12 species of hunting spiders (triangles: species scores **u**) in pitfalls (circles: site scores $\mathbf{x}^*$) with respect to five environmental variables (arrows: environmental biplot scores **c**). The pair **u** and $\mathbf{x}^*$ form a joint plot of **Y** as in CA (ruled by the centre of gravity principle) and a biplot of the contingency ratios $\mathbf{R}^{-1}\mathbf{Y}\mathbf{K}^{-1}$. The pair **u** and **c** form a weighted least-squares biplot of $\mathbf{M} = \mathbf{K}^{-1}\mathbf{Y}^t\mathbf{Z}$, the 12×5 matrix of weighted averages of species with respect to the environmental variables. When the **Z**-row points **x** are plotted (instead of $\mathbf{x}^*$), with **x** in standard row coordinates (and **u** in principal coordinates), and **Z** column-normalized, the pair **x** and **c** and form a biplot of **Z**, as **c** is then the intra-set correlation and the projection of **Z** on the factorial plane.

consists of 1) applying CA or detrended CA (Hill and Gauch 1980) to the species data and 2) interpreting the factorial axes in terms of (external) variables, particularly environmental variables, CCA integrated these two steps into one, whereby treating the external variables as predictors. Prodon and Lebreton (1994) show that CCA is more efficient than the two-step approach. A major product of CCA is an ordination diagram (factorial plane) that displays the pattern of community variation that can be explained best by the known environment (Figure 1).

The shortest summary is perhaps that CCA is (multiple) CA with external linear restrictions on the row points (Gifi 1990). As such, CCA is the counterpart of redundancy analysis, which is principal components analysis with external linear restrictions on the row points (Sabatier et al. 1989, Takane and Hunter 2001). The eigen equation of CCA is

$$(\mathbf{Z}^t\mathbf{Y}\mathbf{K}^{-1}\mathbf{Y}^t\mathbf{Z} - \lambda\mathbf{Z}^t\mathbf{R}\mathbf{Z})\mathbf{b} = \mathbf{0} \tag{1}$$

with **K** and **R** diagonal matrices, containing the column and row totals of **Y**, λ and **b** the eigen value and eigen vector (of length p), giving the CCA site scores **x** = **Zb** (a linear combination of the predictor variables) and the CCA species scores **u**, which are at the centre of gravity of the site scores where they occur. Equation (1) looks like the eigen equation of canonical correlation analysis except for **K** and **R** not being **Y′Y** and **I**$_n$, respectively. What is the virtue of this equation and how was it discovered?

This chapter presents the history of CA and CCA in ecology, the definition of CCA in terms of statistical triplets and associated biplots, the relation of CCA to canonical variate analysis (multi-group linear discriminant analysis) and concludes with extensions and ramifications.

2. History of correspondence analysis (CA) in ecology

The first application of CA to ecological data was presumably by Roux and Roux (1967) and the second by Hatheway (1971), but CA became popular by its introduction by Hill (1973) under the name of reciprocal averaging. Hill (1973) acknowledges John Gower for pointing out that the method was indeed a 'Analyse Factorielle des Correspondances' and refers to Benzécri (1969) and Escofier-Cordier (1969), with more references to older work in Hill (1974). Tenenhaus and Young (1985) point to Richardson and Kuder (1933) as the inventors of the reciprocal averaging approach (in psychology). Hill (1973) is arguably more general as the denominators in his averages are unequal in both rows and columns, instead of being equal in one mode.

The popularity of CA in ecology is, in my view, due to the unimodal response that CA can discover in data. The possibility of analysing unimodal relationships with CA was first noticed by Mosteller (1948 in Torgerson 1958 : p. 338). Unimodal response is common in ecology and derives from two well-known laws. The first law, Liebig's law, states that each species requires a minimum amount of resource (*e.g.* nitrogen). Fertilizer usage in agriculture builds on this law. The second law, Shelford's (1919) law of tolerance (Allaby 1998) states that, in addition, each species tolerates no more than a certain maximum. These two laws yield the niche of a species, that is the region in resource space where the species can actually grow and reproduce. Niches vary among species because the required minima and tolerated maxima differ among species. Species also tend to prosper best around the centre of the ecological niche,
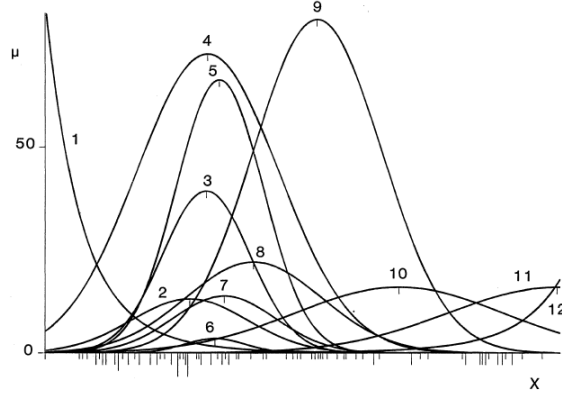
Figure 2: The niche model: unimodal response curves of species of hunting spiders with respect to a latent environmental variable (x).

yielding unimodal (single peaked) response (Figure 2). Unimodal response is also an important feature in psychology with the ideal point model (Coombs and Avrunin 1977) and unfolding method (Coombs and Kao 1960, Heiser 1981, 1987). An example in personal preference is 'I like coffee, but only at the right temperature, not too cold and not too hot' and persons differ in their ideal temperature.

Ecologists (Gause 1930, Ellenberg 1948, Whittaker 1956, Zelinka and Marvan 1961, Ellenberg et al. 1991) developed the method of weighted averaging in this niche context. The preference or indicator value of a species (u_k) with respect to a physical gradient, *e.g.* acidity (pH) can be estimated by the average of the pH values of the sites where the species occurs. If abundance is taken into account, a weighted average is taken with weight proportional to abundance. In formula,

$$u_k = \sum_{i=1}^{n} y_{ik} \, x_i \big/ \sum_{i=1}^{n} y_{ik} \tag{2}$$

with x_i the measured pH at the site i. Conversely, when indicator values are known, but the acidity of a site is not, an estimate of the acidity at the site can be estimated on the basis of the species that are recorded there. The estimate is simply the average of the indicator values of the species recorded there. If abundance is taken into account, a weighted average is taken with weight proportional to abundance. In formula,

$$\lambda x_i = \sum_{k=1}^{m} y_{ik} \, u_k \big/ \sum_{k=1}^{m} y_{ik} \tag{3}$$

with u_k the indicator value of species k and, in weighted averaging, $\lambda = 1$. For presence-absence data the weighted average formulas hold true with $y_{ik} = 0$ or 1. Note

that absences ($y_{ik} = 0$) do not count in weighted averaging. This is a distinctive and useful feature of the method, as a species may not occur at a site just because of another unfavourable factor.

Hill (1973) extended the method of weighted averaging to CA by proposing to iterate between the weighted averaging equations (2) and (3) until convergence (for a fixed range of *x*, for example). By this process, the meaning of the manifest variable, in our example pH, is lost and the variable obtained after convergence is a latent one, the best hypothetical environmental variable. It is well known to be the first non-trivial axis of CA. The well-known transition formulae of CA are essentially equations (2) and (3) with the first nontrivial eigenvalue λ ($0<\lambda\leq1$). Weighted averaging is, of course, nothing more than the centre of gravity principle (*le principe barycentrique* in French) and CA takes it both ways, although not exactly if $\lambda<1$.

In ecological context (and many more, see chapter de Leeuw) the second axis has often an almost quadratic relation to the first. This is known as the arch effect (Gauch 1982) or Guttman effect (it is *not* a horseshoe as the ends do not bend in). An easy, nonmathematical explanation of this effect is presented in Jongman et al. (1987, 1995). Hill and Gauch (1980) proposed to remove this relationship by detrending within the iteration algorithm and their efficient computer program DECORANA (Hill 1979a) implementing this has allowed major usage of both CA and detrended CA in ecology. The arch affect, which is mathematical artefact of CA, is in my view *the* reason that Guttman never wanted to go beyond the first CA axis (see chapters Lebart/de Leeuw). Detrending is somewhat controversial (Wartenberg et al. 1987), but is not the only way to avoid the arch effect. External linear constraints as in CCA are another (Palmer 1993).

But, what does CA optimize to make it suited for ecological applications and the unimodal response model in particular? CA has many nice optimality properties when applied to contingency tables (Greenacre 1984), but is the species data matrix **Y** a contingency table? Often it is a incidence matrix with just zeros and ones, or if **Y** contains counts, these counts arise from sampling a number of individuals from each of a fixed number of sites. **Y** is thus not a contingency table, unless 'sites' are treated as (random) nominal variable. Which might be the ecological interpretation of the fact that CA finds the row and column quantification that maximizes the correlation

between rows and columns (Hill 1974, Tenenhaus and Young 1985)? What else could motivate CA for use in ecology, rather than being just principal component analysis with some fancy pre- and post-transformations? First, CA is useful for seriation of incidence matrices (Torgerson 1958): "if the rows and columns of an incidence matrix of an incidence matrix $\mathbf{Y}$ can be permuted so that all the 1's in every row and column come together (a Petrie matrix), then there is (apart from degeneracies) a unique correct ordering of rows and columns which is generated by the first non-trivial axis of CA of $\mathbf{Y}$" (free after Hill, 1974). Heiser (1981) extended this result to so-called complete matrices with consecutive 1's in the rows (or columns) only. Second, CA is able to find disjoint blocks in two-way tables (Hill 1974) and this feature is used the cluster analysis program TWINSPAN (Hill 1979b, Jongman et al. 1995). The features of CA that make it less suited for ecological applications (Jongman et al. 1995) are its sensitivity to rare species and outlier sites (a feature related to block structure detection) and the arch effect (see above). The ecological meaning of the analysis of variance approach to CA with its correlation ratio η^2 (Guttman 1941, Nishisato 1980, Tenenhaus and Young 1985) escaped me until ter Braak (1987). For the relation of CA to the Gaussian response model see the next section.

3. History of canonical correspondence analysis (CCA)

In this section I present my route to CCA followed by notes on what I know of Jean-Dominique Lebreton and Daniel Chessel's route. My personal prehistory is as follows. In 1978 I worked for the Research Institute of Nature Management (RIN, nowadays part of Alterra, Wageningen University and Research Centre), while being employed by TNO (www.tno.nl) in a statistics group in Wageningen, the Netherlands, where regression analysis and ANOVA were the workhorses of statistics. My RIN colleague Hans van Biezen had the two Benzécri volumes (Benzécri 1973) on his shelves, a gift by the French Embassy in the Netherlands, and I studied them in some detail. I dived into statistical ecology at the 1978 Satellite Program in Statistical Ecology (Patil 1995). Kooijman (1977) and Kooijman and Hengeveld (Kooijman and Hengeveld 1979) introduced me to the Gaussian model for ordination and pointed out the numerical problems with the maximum likelihood approach when applied in several dimensions (see also Goodall and Johnson (1982)). In 1979-80 I was on leave and did an MSc in statistics (by research) with Peter Diggle in Newcastle upon Tyne (UK) where I learned about maximum likelihood equations in the exponential family

from Professor R.L. Placket. I also met Colin Prentice with whom I discussed ordination methods current at the time (Prentice 1977, 1980a, b) and Mark Hill who kindly pointed out that he particularly disliked my draft Principal Component analysis biplot and diversity paper (ter Braak 1983) as 1) ecology is not linear and 2) differences in niches locations create diversity as in DECORANA (Hill 1979a). In 1981 I attended the second Gifi course (Gifi 1981, 1990). Back at the RIN, I was asked statistical advice on the usage of species indicator values in nature management and acidification research (van Dam et al. 1981, ter Braak and Gremmen 1987, ter Braak and van Dam 1989). This led me to the study of the properties of the method of weighted averaging (outlined above). I put weighted averaging in the context of response curve modelling and asked myself the question for which shape of species response curves and under which conditions, weighted averaging would be close to maximum likelihood (ter Braak and Barendregt 1986, ter Braak and Looman 1986). These papers show the link of weighted averaging with the ecological niche model and the Gaussian response model in particularly. The paper on CA of incidence and abundance models in terms of a unimodal response model (Gaussian ordination) followed naturally (ter Braak 1985), despite being published earlier.

Independently Ihm and van Groenewoud (1984) compared correspondence analysis and a variant of Gaussian ordination (model B) that is attractive when sites vary in 'size' so that only relative abundance is meaningful. They show that their Model B is identical to Goodman's RC model (Goodman 1979) and, while citing Escoufier (1982), that a first order Taylor expansion yields the reconstitution formulae of CA. This result, earlier discovered by Goodman (1981), applies to small λ (close to 0; data close to row-column independence) whereas the result of ter Braak (1985) holds true for large λ (close to 1).

With this background, how was CCA derived? In October 1984 Colin Prentice and I made an outline in Uppsala (Sweden) for "A theory of gradient analysis", which was eventually published as ter Braak and Prentice (1988). We searched for something like canonical correlation analysis for niche models. The night before leaving Uppsala I got the idea to linearly constrain the scores in Gaussian ordination and approximate the maximum likelihood equations, just as I did for CA (ter Braak 1985). This derivation given in the appendix of ter Braak (1986) yielded the precise row and

		Formula in words (with ecological term)	formula
1	**Y**-column points	'Species scores': weighted averages of **Z**-row points	$\mathbf{u} = \mathbf{K}^{-1}\mathbf{Y}^t\mathbf{x}$
2	**Y**-row points	'Site scores': weighted averages of the **Y**-column points	$\mathbf{x}^* = \mathbf{R}^{-1}\mathbf{Y}\mathbf{u}$
3	**Z**-weights	Coefficients of regression of $\mathbf{x}^*$ on **Z** with row weights **R** (canonical weights)	$\mathbf{b} = (\mathbf{Z}^t\mathbf{R}\mathbf{Z})^{-1}\mathbf{Z}^t\mathbf{R}\mathbf{x}^*$
4	**Z**-row points	'Constrained site scores': row points obtained as a linear combination of **Z**	$\lambda\mathbf{x} = \mathbf{Z}\mathbf{b}$

Table 1: Transition formulae of canonical correspondence analysis (CCA) of a response matrix **Y** with respect to a predictor matrix **Z**. The matrices **R** and **K** are diagonal and contain the row and column totals of **Y**, respectively.

columns weights that are particular for CCA (and CA), the transition formulae (Table 1) and from these, the eigen equation (1). This explains the adjective 'canonical' in the name of the method. A better, later motivation of the term is that CCA transforms a quadratic into its canonical form (ter Braak 1988c). The close relationship to canonical variate analysis provides additional motivation.

In an attempt to find a derivation without explicit use of the Gaussian model, I derived CCA as a linear combination of predictors **Z** that best separate species niches (ter Braak 1987). The maximized criterion is the dispersion of the species scores (weighted averaged site scores) with respect to standardized site scores, which is precisely the correlation ratio η^2 in the analysis of variance approach to CA (Nishisato 1980, Tenenhaus and Young 1985). The difference is, that is the criterion is now maximized subject to linearly constrained sites scores ($\mathbf{x} = \mathbf{Z}\mathbf{b}$), which is achieved by optimizing **b**.

I presented partial CCA (a CCA with the effects of nuisance variables partialled out) in 1987 at the first IFCS (International Federation of Classification Societies)

conference in Aachen (Germany), where Yves Escoufier strongly express how much he disliked the name 'partial CCA' and the regression structure in the context of correlated predictors, but also kindly invited me to Montpellier (France) to get to know statistical triplets, duality diagrams and his French colleagues who independently invented CCA. Early work related to partial CCA (ter Braak 1988c) is Yoccoz and Chessel (1988), Cazes et al. (1988), Sabatier et al. (1989) and Lebreton et al. (1991).

Jean-Domique Lebreton (JDL) worked at the Université Lyon I (Laboratoire de Biométrie), where he met Daniel Chessel and Nigel Yoccoz, and then moved to C.E.P.E./C.N.R.S at Montpellier where he developed further work with Yves Escoufier and Robert Sabatier. JDL was not only interested in multivariate data analysis but also in survival analysis of animals by capture-mark-recapture sampling and had been exposed to generalized linear models from the onset of these models in the 70s. Capture and survival probabilities are very difficult to estimate by this sampling method with, as a result, extremely wide confidence intervals. In that context he invented the trick to borrow strength across years and sampling occasions by constraining the probabilities of the model by logistic models with environmental covariates (Clobert and Lebreton 1985, Lebreton et al. 1992). This gave JDL the idea of applying the same trick in CA (December 1984). Daniel Chessel contributed the general perspective brought by the duality diagram approach to multivariate analysis which was commonly used in France at that time (Caillez and Pagès 1976). He then developed CCA under the point of view of linearly constrained CA with Daniel Chessel, Nigel Yoccoz and Robert Sabatier (Chessel et al. 1987, Lebreton et al. 1988a, Lebreton et al. 1988b, Sabatier et al. 1989). In the process they discovered a 1985 TNO-report, a preprint to ter Braak (1986). The connection between the unimodal model and CA was well known to JDL and Daniel Chessel (Chessel et al. 1982). Chessel was well aware of orthogonal and oblique projectors (Afriat, 1975) and Rao's (1964) "principal components of instrumental variables", later reinvented as redundancy analysis (van den Wollenberg, 1977) . Neither paper was known to me when I invented CCA. For biplots in redundancy analysis see ter Braak and Looman (1994).

CCA was first implemented on a computer as an extension of DECORANA (Hill 1979a) that was already quite popular in ecology because of the large data sets it could handle. DECORANA used sparse matrix algebra and an efficient eigenvector routine (ter Braak and de Jong 1998). CANOCO version 1.0 (1985) did CA, detrended CA, CCA and detrended CCA. From version 2.1 it included (partial) principal component analysis and redundancy analysis as well (ter Braak 1988a, b) and did permutation testing of the null hypothesis that the responses (species) were independent of the predictors (environmental variables), given particular covariables (nuisance predictors, which are not under test) if present. Onno van Tongeren contributed with the user interface, a dedicated data editor (Cedit) and a program that produced line printer plots. From version 3.1 (ter Braak 1990), Canoco included a plotting program (Canodraw) created by Petr Šmilauer, for plotting ordination diagrams (biplot, CA plots and many more). Canoco for Windows (from version 4) started in 1998 (ter Braak and Šmilauer 1998, 2012). Colin Prentice, John Birks and Paul van den Brink were the ambassadors of Canoco by presenting its methods for ecological audiences. CCA has also been included in the R packages vegan (Oksanen et al. 2011), ade4 (Dray and Dufour 2007) and anacor (de Leeuw and Mair 2009). CCA in vegan allows a formula interface. The statement cca(Y~ A*B + Condition(C)) specifies a partial CCA with covariables C and with predictors A*B, that is the main effects of factors (or variables) A and B and their interaction A.B.

4. Triplets and biplots in CCA[1]

This section points to the connections between statistical triplets (Cailliez and Pagès 1976, Tenenhaus and Young 1985, Escoufier 1987), reduced rank least-squares approximation (Greenacre 1984) and the biplot (Gabriel 1971) and uses these to provide triplets and biplots for CCA. Triplets belong to the French multivariate data analysis tradition, and reduced rank approximations and biplots to the English-speaking one.

A statistical triplet ($\mathbf{D}$, $\mathbf{K}$, $\mathbf{R}$) of an $n \times m$ matrix $\mathbf{D}$ (for 'data') and positive semi-definite metrics (weight matrices) $\mathbf{K}$ ($m \times m$) and $\mathbf{R}$ ($n \times n$) corresponds to a rank r

---

[1] This section draws from a seminar in 1988 in Montpellier where I presented CCA to statisticians trained in the French data analysis tradition. With thanks to Yves Escoufier, Jean-Dominique Lebreton and Maurice Roux for discussions and making this a very pleasant visit.

approximation of $\mathbf{D} \approx \mathbf{GH}^t$ with $\mathbf{G}$ $n \times r$ and $\mathbf{H}$ $m \times r$ ($r<\min(m,n)$) by minimizing the Frobenious norm (Greenacre 1984)

$$\left\|\mathbf{R}^{1/2}(\mathbf{D} - \mathbf{GH}^t)\mathbf{K}^{1/2}\right\|^2 = tr[(\mathbf{D} - \mathbf{GH}^t)^t\mathbf{R}(\mathbf{D} - \mathbf{GH}^t)\mathbf{K}] \tag{4}$$

with $\|\mathbf{A}\|^2 = \sum_{i,j} a_{ij}^2$. The solution matrices can be obtained from the generalized singular value problem (Greenacre 1984). This result is related to the famous Eckhart-Young theorem (see chapter Gower). The biplot graphic display is nothing more than a vector representation of the rows of both $\mathbf{G}$ and $\mathbf{H}$ with the mutual inner products providing the rank r least squares approximation to $\mathbf{D}$ (Gabriel 1971). The key statistical triplets of CCA are given in Table 2. A duality diagram (Escoufier 1987) of CCA with the corresponding transition formulae is given in Figure 3. See Chessel et al. (1987) for related duality diagrams.

'data table'	triplet	biplot	
Fitted contingency ratios	$(\mathbf{\Pi_Z}\mathbf{R}^{-1}\mathbf{YK}^{-1}, \mathbf{K}, \mathbf{R})$	$\mathbf{x}_i$	$\mathbf{u}_k$
Table of weighted averages of species ($\mathbf{Y}$-columns) with respect to $\mathbf{Z}$	$(\mathbf{K}^{-1}\mathbf{Y}^t\mathbf{Z}, (\mathbf{Z}^t\mathbf{RZ})^{-1}, \mathbf{K})$	$\mathbf{c}_j$	$\mathbf{u}_k$
Regression coefficients of contingency ratios with respect to $\mathbf{Z}$	$\left((\mathbf{Z}^t\mathbf{RZ})^{-1}\mathbf{Z}^t\mathbf{R}(\mathbf{R}^{-1}\mathbf{YK}^{-1}), \mathbf{K}, \mathbf{Z}^t\mathbf{RZ}\right)$	$\mathbf{b}_j$	$\mathbf{u}_k$

Table 2: Key statistical triplets and corresponding biplots of CCA. The matrices $\mathbf{R}$ and $\mathbf{K}$ are defined in Table 1 and $\mathbf{\Pi_Z} = \mathbf{Z}(\mathbf{Z}^t\mathbf{RZ})^{-1}\mathbf{Z}^t\mathbf{R}$, the $\mathbf{R}$-weighted projection operator on $\mathbf{Z}$.

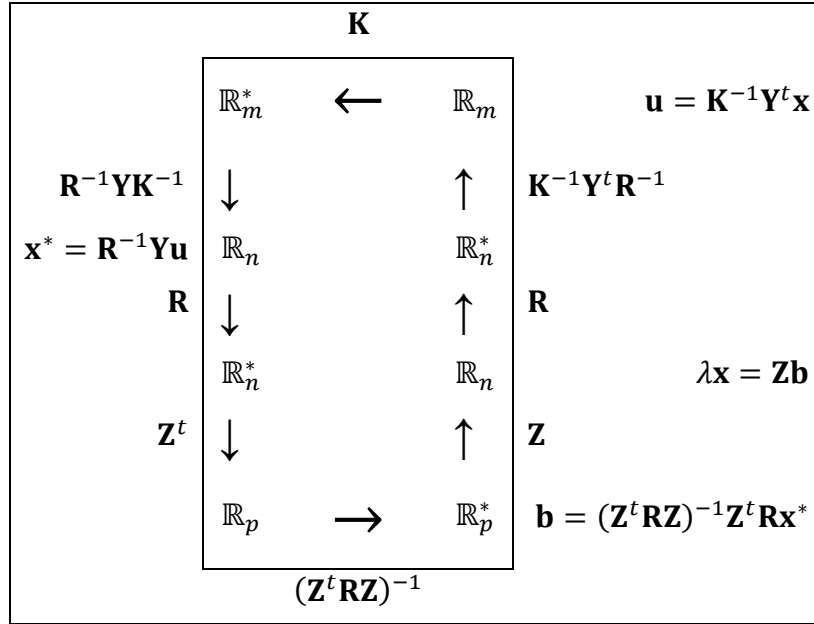*Figure 3:* A duality diagram of CCA of response matrix **Y** and predictor matrix **Z** with the corresponding transition formulae of Table 1. The matrices **R** and **K** are defined in Table 1. Note that **x** is the projection of $\mathbf{x}^*$ on **Z**, in short, $\mathbf{x} = \mathbf{\Pi_Z x}^*$.

5. Relation of CCA with canonical variate analysis (CVA)

Chessel et al. (1987) and Lebreton et al. (1988a) were the first to recognize the formal equivalence between CCA and linear discriminant analysis (CVA) with *m* groups (species) on reformatted (inflated) data. With the knowledge that CCA is a form of CVA, CCA has an early precursor in the ecological literature in the form of Green's (1971, 1974) multi-group linear discriminant analysis, as noted by ter Braak and Verdonschot (1995). Interest in Green's method was lost, ironically in the same period as the popularity of CA surged. The main difference between two methods is that the unit of statistical analysis in CVA is the individual (a row in the inflated data matrices), whereas it is the site (row of **Y**) in CCA. Statistical tests designed for CVA as used by Green (1971, 1974) are thus invalid in the context of CCA (ter Braak and Verdonschot 1995). Valid tests can be based on Monte Carlo permutation of sites (instead of individuals), as implemented in CANOCO. This part of the history is similar to that of CA, as CA is also a particular form of CVA and canonical correlation analysis on data inflated to indicator matrices.

The analysis of variance approach to CA is essentially one of discriminant analysis/CVA. CVA extracts orthogonal components of the form $\mathbf{x} = \mathbf{Zb}$ such that the between (B) to within (W) groups sum of squares is maximized. As the total sum of squares is the sum of the two (T = B + W), maximizing B/W is identical to maximizing B/T, which is the correlation ratio (Nishisato 1980), and identical to minimizing W/T. The W/T criterion puts CCA (and CA) in the context of unfolding models (Heiser 1981, 1987). With a rescaling so that T = 1, W can be rewritten as

$$W = \sum_{i,k} y_{ik}(x_i - u_k)^2 \tag{5}$$

because for any fixed choice of $\mathbf{x}$, equation (5) is minimized for $\mathbf{u} = \mathbf{K}^{-1}\mathbf{Y}^t\mathbf{x}$, so that is the right hand side of equation (5) becomes the usual within group sum of squares. In CA, W is minimized subject to the constraints $\mathbf{1}^t\mathbf{Rx} = 0$ and $\mathbf{x}^t\mathbf{Rx} = 1$. CCA imposes the additional constraint $\mathbf{x} = \mathbf{Zb}$. Note that the data enter as weights. A second component can be extracted by requiring it to be orthogonal to the first ($\mathbf{x}^t\mathbf{Rx}_1 = 0$). Heiser (1987) noted that the orthogonality constraint is "not free from arbitrariness *under the present rationale* of the method" (CA). Hence, the arch effect artefact and the occasional need for detrending in CA. In CVA, and thus CCA, the orthogonality is not so arbitrary, for example, if we take as starting point the least-squares approximation of the group means $\mathbf{M}$ (second triplet in Table 2, detailed below). Zhu et al. (2005) showed the equivalence between CCA and CVA via a weighted sample model.

The usual CCA ordination diagram (Figure 1) is similar to that of CVA. The scores for the variables in $\mathbf{Z}$ are usually not based on the eigenvector $\mathbf{b}$ (a vector with regression coefficients) as these are very unstable if predictors show high mutual correlation. Instead, scores for predictor variables are based on the weighted least-squares approximation of the matrix of weighted averages (conditional means) of the m species with respect to the p predictor variables, represented by the $m \times p$ matrix $\mathbf{M} = \mathbf{K}^{-1}\mathbf{Y}^t\mathbf{Z}$ with elements ($\mathbf{K}$ defined as a diagonal matrix containing column sums)

$$m_{kj} = \sum_{i=1}^n y_{ik} z_{ij} / \sum_{i=1}^n y_{ik}. \tag{6}$$

The matrix M corresponds to the group means when CCA is viewed as a CVA on inflated data. The corresponding triplet for CCA in Table 2 is $(\mathbf{M}, (\mathbf{Z}^t\mathbf{RZ})^{-1}, \mathbf{K})$, whereas that of CVA would use $(\mathbf{M}, \mathbf{S}^{-1}, \mathbf{K})$ with $\mathbf{S}$ the pooled within-group

covariance matri This difference corresponds to the fact that CCA maximizes B/T, whereas CVA maximizes B/W. If the eigenvalue of CCA is λ, the corresponding eigenvalue of CVA is therefore λ/(1- λ), a relation surprisingly absent in Chessel et al. (1987). This difference does not influence the biplot of **M** in either method, as the biplot does not change when we rescale the pair $(\mathbf{u}_k, \mathrm{c}_\mathrm{j})$ in Table 2 to $(s\mathbf{u}_k, \mathrm{c}_\mathrm{j}/s)$ for any scalar *s*. However, it does influence the interpretation of, for example, the distances between the species (Y-column points)(see chapter Gower). In a default CCA diagram in the column-preserving metric, the inter-species distances approximate chi-square distances (from below (Meulman 1986)), whereas in CVA, the inter-group distances then approximate Mahalanobis distances. The CCA diagram can be transformed into the CVA form by using the so-called Hill's scaling which equalizes the within-species variance W across different components. In this scaling the species scores are scaled to $\mathbf{u}^t\mathbf{Ku} = \lambda/(1 - \lambda)$ with λ the eigenvalue of the corresponding component (Hill 1979a, ter Braak and Verdonschot 1995). The Mahalanobis distance is then a natural measure of distance between niches of species (Green 1971, 1974).

6. Later landmarks

With many predictors CCA does no longer constrain the CA and is then identical to CA. To avoid this, Dolédec and Chessel (1994) developed co-inertia analysis, which ignores the correlations among predictors. As an alternative, ter Braak and Verdonschot (1995) developed a PLS variant of CCA. The regression setting of CCA allows easy variance (inertia) partitioning (Borcard et al. 1992, Okland and Eilertsen 1994). Doubly linearly constrained CA or double CCA for short was developed by Böckenholt and Böckenholt (1990) whereas the co-inertia version is by Dolédec et al. (1996). These variants are endpoints (usage or no usage of the within set covariance matrix) in the framework of Tenenhaus and Tenenhous (2011). Lavorel et al. (1999) noted that double CCA is essentially a weighted canonical correlation analysis. Later assessments of CCA are by Palmer (1993), Johnson and Altman (1999), Graffelman (2000, 2001), Graffelman and Tuft (2004) and Zhang and Thas (2012). A modern view on CCA is Zhu, Hastie and Walther (2005). A related method is co-correspondence analysis, a method designed to relate two species communities (ter Braak and Schaffers 2004, Schaffers et al. 2008). This method can also be used to related two sets of (many) nominal variables from small samples. It avoids the

multicollinearity problems in such data and is as such an alternative to CANALS (van der Burg and de Leeuw 1983, 1990). It appears that all these CA-related methods can be obtained from standard (un-weighted) linear methods by inflating the data matrices to (super)indicator matrices as in the dual scaling approach (Nishisato 1980, Gifi 1990). After inflation, the unit of statistical analysis is the individual instead of being the site. Alternatively, they can be obtained from un-weighted linear methods by pre- and post-transformation (ter Braak and Verdonschot 1995) as in the principal components approach to CA (Tenenhaus and Young 1985, Chessel et al. 1987).

In this chapter I emphasized the unimodal properties of CA and CCA, but the linear ones are explicit in the first triplet of Table 2 that forms the basis of the reconstitution formulae of CA (Greenacre 1984, Van der Heijden et al. 1994). So, CA and CCA are chameleons: sometimes they show up as linear methods as in the reconstitution formulae and associated biplot of the contingency ratios, and sometimes as unimodal methods as in their relationship to CVA and unfolding (equation (5)). How is that possible; how can this be explained? In the RC model this duality is fully mathematically understood (Ihm and van Groenewoud 1984, ter Braak 1988c, de Rooij 2007), but its understanding in (C)CA is still limited. When there is a strong arch effect (not very likely in CCA with few predictors), the rank 1 reconstitution appears bad. Perhaps power transformation (Greenacre 2009) can shed light (and remove the arch)? The relation between CA, CCA and the (constrained) RC-model, ideal point discriminant analysis and the Anderson's stereotype model (Anderson 1984) has been further developed in de Rooij and Heiser (2005) and van der Heijden et al. (1994).

References

Allaby, M. 1998. Shelford's law of tolerance. A dictionary of ecology. <http://www.encyclopedia.com>.

Anderson, J. A. 1984. Regression and ordered categorical variables. J. R. Statist. Soc. B **46**:1-30.

Afriat, S. N. 1957. Orthogonal and oblique projectors and the characteristics of pairs of vector spaces. Proceedings of the Cambridge Philosophical Society **53**:800-816.

Benzécri, J.-P. 1973. L'analyse des donnees. II L'analyse des correspondances. Dunod, Paris.

Benzécri, J. P. 1969. Statistical analysis as a tool to make patterns emerge from data.*in* S. Watanabe, editor. Methodologies of pattern recognition. Academic Press, New York.

Böckenholt, U. and I. Böckenholt. 1990. Canonical-analysis of contingency-tables with linear constraints. Psychometrika **55**:633-639.

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. Ecology **73**:1045-1055.

Cailliez, F. and J.-P. Pagès. 1976. Introduction á l'Analyse des Données. Societé de Mathématiques Appliquées et de Sciences Humaines, Paris.

Cazes, P., D. Chessel, and S. Dolédec. 1988. L'Analyse des correspondances internes d'un tableau partitionné: son usage en hydrobiologie. Revue de Statistique Appliquée **36**:39-54.

Chessel, D., J.-D. Lebreton, and R. Prodon. 1982. Mesures symétriques d'amplitude d'habitat et de diversité, intra-échantillon dans un tableau espèces-relevés: cas d'un gradient simple. C.R. Acad. Sc. Paris, Series III **295**:83-88.

Chessel, D., J. B. Lebreton, and N. Yoccoz. 1987. Propriétés de l'analyse canonique des correspondances; une illustration en hydrobiologie. Revue de Statistique Appliquée **35**:55-72.

Clobert, J. and J. D. Lebreton. 1985. Dépendance de facteurs de milieu dans les estimations de taux de survie par capture-recapture. Biometrics **41**:1031-1037.

Coombs, C. H. and G. S. Avrunin. 1977. Single-peaked functions and the theory of preference. Psychological Review **84**:216-230.

Coombs, C. H. and R. C. Kao, editors. 1960. On a connection between factor analysis and multidimensional unfolding. John Wiley, New York.

de Leeuw, J. and P. Mair. 2009. Simple and Canonical Correspondence Analysis Using the R Package anacor. Journal of Statistical Software **31**:1-18.

de Rooij, M. 2007. The distance perspective of generalized biadditive models: Scalings and transformations. Journal of Computational and Graphical Statistics **16**:210-227.

De Rooij, M. and W. J. Heiser. 2005. Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data. Psychometrika **70**:99-122.

Dolédec, S. and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biology **31**:277-294.

Dolédec, S., D. Chessel, C. J. F. ter Braak, and S. Champely. 1996. Matching species traits to environmental variables: a new three-table ordination method. Environmental and Ecological Statistics **3**:143-166.

Dray, S. and A. B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software **22**:1-20.

Ellenberg, H. 1948. Unkrautgesellschaften als Mass fuer den Saeuregrad, die Verdichtung und andere Eigenschaften des Ackerbodens. Berichten der Landtechnik **4**:130-146.

Ellenberg, H., H. E. Weber, R. D ll, V. Wirth, W. Werner, and D. Paulissen. 1991. Indicator values of plants in Central Europe. Scripta Geobotanica **18**:0-0.

Escofier-Cordier, B. 1969. L'analyse factorielle des correspondances. Cah. Bur. Univ. Rech. opér Univ. Paris **13**.

Escoufier, Y. 1982. L'analyse des tableaux de contingence simple at multiples Metron **40**:53-77.

Escoufier, Y. 1987. The duality diagram: a means for better practical applications. Pages 139-156 *in* P. Legendre and L. Legendre, editors. Developments in numerical ecology. Springer-Verlag, Berlin.

Gabriel, K. R. 1971. The biplot graphic display of matrices with application to principal component analysis. Biometrika **3**:453-467.

Gauch, H. G. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge.

Gause, G. F. 1930. Studies on the ecology of the Orthoptera. Ecology **11**:307-325.

Gifi, A. 1981. Nonlinear multivariate analysis. DWSO-press, Leiden.

Gifi, A. 1990. Nonlinear multivariate analysis. Wiley, New York.

Goodall, D. W. and R. W. Johnson. 1982. Non-linear ordination in several dimensions. A maximum likelihood approach. Vegetatio **48**:197-208.

Goodman, L. A. 1979. Simple models for the analysis of association in cross-classifications having ordered categories. Journal of the American Statistical Association **74**:537-552.

Goodman, L. A. 1981. Association models and canonical correlation in the analysis of cross-classifications having ordered categories. JASA **76**:320-334.

Graffelman, J. 2000. Use of the Moore-Penrose inverse in canonical correspondence analysis - Solution. Econometric Theory **16**:792-793.

Graffelman, J. 2001. Quality statistics in canonical correspondence analysis. Environmetrics **12**:485-497.

Graffelman, J. and R. Tuft. 2004. Site scores and conditional biplots in canonical correspondence analysis. Environmetrics **15**:67-80.

Green, R. H. 1971. A multivariate statistical approach to the Hutchinsonian niche: bivalve mollucs of central Canada. Ecology **52**:543-556.

Green, R. H. 1974. Multivariate niche analyis with temporally varying environmental factors. Ecology **55**:73-83.

Greenacre, M. 2009. Power transformations in correspondence analysis. Computational Statistics and Data Analysis **53**:3107-3116.

Greenacre, M. J. 1984. Theory and applications of correspondence analysis. Academic Press, London.

Guttman, L. 1941. The quantification of a class of attributes: A theory and method of scale construction. Pages 319-348 *in* P. Horst, editor. The prediction of personal adjustment. Social Science Research, New York.

Hatheway. 1971. Contingency-table analysis of rain forest vergetation.*in* G. P. Patil, E. C. Pielou, and W. E. Waters, editors. Statistical Ecology. Pennsylvania State University Press, University Part, Pennsylvania.

Heiser, W. J. 1981. Unfolding analysis of proximity data. Thesis. University of Leiden, Leiden.

Heiser, W. J., editor. 1987. Joint ordination of species and sites: the unfolding technique. Springer-Verlag, Berlin.

Hill, M. O. 1973. Reciprocal averaging: an eigenvector method of ordination. Journal of Ecology **61**:237-249.

Hill, M. O. 1974. Correspondence analysis: a neglected multivariate method. Applied Statistics **23**:340-354.

Hill, M. O. 1979a. DECORANA - A FORTRAN program for detrended correspondence analysis and reciprocal averaging. Ecology and Systematics. Cornell University, Ithaca, New York.

Hill, M. O. 1979b. TWINSPAN - a FORTRAN program for detrended correspondence analysis and reciprocal averaging. Cornell University, Ithaca.

Hill, M. O. and H. G. Gauch. 1980. Detrended correspondence analysis, an improved ordination technique. Vegetatio **42**:47-58.

Ihm, P. and H. van Groenewoud. 1984. Correspondence analysis and Gaussian ordination. Compstat Lectures **3**:5-60.

Johnson, K. W. and N. S. Altman. 1999. Canonical correspondence analysis as an approximation to Gaussian ordination. Environmetrics **10**:39-52.

Jongman, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren. 1987. Data analysis in community and landscape ecology. Pudoc, Wageningen.

Jongman, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren. 1995. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge.

Kooijman, S. A. L. M. 1977. Species abundance with optimum relations to environmental factors. Annals of System Research **6**:123-138.

Kooijman, S. A. L. M. and R. Hengeveld. 1979. The description of a non-linear relationship between some carabid beetles and environmental factors. Pages 635-647 *in* G. P. Patil and M. L. Rosenzweig, editors. Contemporary Quantitative Ecology and Related Econometrics. International Co-operative Publishing House, Fairland MD.

Lavorel, S., C. Rochette, and J. D. Lebreton. 1999. Functional groups for response to disturbance in Mediterranean old fields. Oikos **84**:480-498.

Lebreton, J.-D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Monographs **62**:67-118.

Lebreton, J. D., D. Chessel, R. Prodon, and N. Yoccoz. 1988a. L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. I. Variables de milieu quantitatives. Acta Oecologia Generalis **9**:53-67.

Lebreton, J. D., D. Chessel, M. Richardot-Coulet, and N. Yoccoz. 1988b. L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. II. Variables de milieu qualitatives. Acta Oecologia Generalis **9**:137-151.

Lebreton, J. D., R. Sabatier, G. Banco, and A. M. Bacou. 1991. Principal component and correspondence analysis with respect to instrumental variables: an overview of their role in studies of structure-activity and species-environment relationships. Pages 85-114 *in* J. Devillers and W. Karcher, editors. Applied Multivariate Analysis in SAR and Environmental Studies. Kluwer, Dordrecht.

Meulman, J. 1986. A distance approach to nonlinear multivariate analysis. DSWO Press, Leiden.

Nishisato, S. 1980. Analysis of categorical data: dual scaling and its applications. Toronto University Press, Toronto.

Okland, R. H. and O. Eilertsen. 1994. Canonical correspondence analysis with variation partitioning: some comments and an application. J. Veg. Sci. **5**:117-126.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2011. vegan: Community Ecology Package. R package version 1.17-12. http://CRAN.R-project.org/package=vegan.

Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. Ecology **74**:2215-2230.

Patil, G. P. 1995. Editorial: Statistical ecology and related ecological statistics — 25 years. Environmental and Ecological Statistics **2**:81-89.

Perriere, G., J. R. Lobry, and J. Thioulouse. 1996. Correspondence discriminant analysis: A multivariate method for comparing classes of protein and nucleic acid sequences. Computer Applications in the Biosciences **12**:519-524.

Prentice, I. C. 1977. Non-metric ordination methods in ecology. J. Ecol. **65**:85-94.

Prentice, I. C. 1980a. Multidimensional scaling as a research tool in quaternary palynology: a review of theory and methods. Review of Palaeobotanay and Palynology **31**:71-104.

Prentice, I. C. 1980b. Vegetation analysis and order invariant gradient models. Vegetatio **42**:27-34.

Prodon, R. and J.-D. Lebreton. 1981. Breeding avifauna of a Mediterranean succession: the holm oak and cork oak series in the eastern Pyrenees, 1. analysis and modelling of the structure gradient. Oikos **37**:21-38.

Prodon, R. and J.-D. Lebreton. 1994. Analyses multivariees des relations especes-milieu: structure et interpretation ecologique. Vie milieu **44**:69-91.

Richardson, M. and G. F. Kuder. 1933. Making a rating scale that measures. Personnel Journal **12**:36-40.

Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. Sankhya **26**:329-358.

Roux, G. and M. Rou 1967. A propos de quelques méthodes de classification en phytosociologie. Revue Statistique Appliqué **15**:59-72.

Sabatier, R., J.-D. Lebreton, and D. Chessel. 1989. Multivariate analysis of composition data accompanied by qualitative variables describing a structure. Pages 341-352 *in* R. Coppi and S. Bolasco, editors. Multiway data tables. North-Holland, Amsterdam.

Schaffers, A. P., I. P. Raemakers, K. V. Sỳkora, and C. J. F. ter Braak. 2008. Arthropod assemblages are best predicted by plant species composition. Ecology **89**:782-794.

Takane Y, Bozdogan H, Shibayama T. 1987. Ideal point discriminant analysis. Psychometrika 52:371-392.

Takane, Y. and M. A. Hunter. 2001. Constrained principal component analysis: A comprehensive theory. Applicable Algebra in Engineering, Communications and Computing **12**:391-419.

Tenenhaus, A. and M. Tenenhaus. 2011. Regularized Generalized Canonical Correlation Analysis. Psychometrika:1-28.

Tenenhaus, M. and F. W. Young. 1985. An analysis and synthesis of multiple correspondence analyis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. Psychometrika **50**:91-119.

ter Braak, C. J. F. 1983. Principal components biplots and alpha and beta diversity. Ecology **64**:454-462.

ter Braak, C. J. F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. Biometrics **41**:859-873.

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology **67**:1167-1179.

ter Braak, C. J. F. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio **69**:69-77.

ter Braak, C. J. F. 1988a. CANOCO - a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis (version 2.1). Report LWA-88-02. Agricultural Mathematics Group, Wageningen.

ter Braak, C. J. F. 1988b. CANOCO - an extension of DECORANA to analyze species-environment relationships. Vegetatio **75**:159-160.

ter Braak, C. J. F. 1988c. Partial canonical correspondence analysis. Pages 551-558 *in* H. H. Bock, editor. Classification and related methods of data analysis. North-Holland, Amsterdam.

ter Braak, C. J. F. 1990. Update notes: CANOCO version 3.1. Agricultural Mathematics Group, Wageningen.

ter Braak, C. J. F. and L. G. Barendregt. 1986. Weighted averaging of species indicator values: its efficiency in environmental calibration. Mathematical Biosciences **78**:57-72.

ter Braak, C. J. F. and S. de Jong. 1998. The objective function of partial least squares regression. Journal of Chemometrics **12**:41-54.

ter Braak, C. J. F. and N. J. M. Gremmen. 1987. Ecological amplitudes of plant species and the internal consistency of Ellenberg's indicator values for moisture. Vegetatio **69**:79-87.

ter Braak, C. J. F. and C. W. N. Looman. 1986. Weighted averaging, logistic regression and the Gaussian response model. Vegetatio **65**:3-11.

ter Braak, C. J. F. and I. C. Prentice. 1988. A theory of gradient analysis. Advances in Ecological Research **18**:271-317.

ter Braak, C. J. F. and Looman, C. W. N. (1994) Biplots in reduced-rank regression. *Biometrical Journal*, **36,** 983-1003.

ter Braak, C. J. F. and A. P. Schaffers. 2004. Co-correspondence analysis: a new ordination method to relate two community compositions. Ecology **85**:834-846.

ter Braak, C. J. F. and P. Šmilauer. 1998. CANOCO Reference Manual and User's Guide to Canoco for Windows: Software for Canonical Community Ordination (version 4). Microcomputer Power, Ithaca, NY, USA.

ter Braak C.J.F. and Šmilauer P. (2012). Canoco reference manual and user's guide: software for ordination, version 5.0. Microcomputer Power, Ithaca, USA, 496 pp.

ter Braak, C. J. F. and H. van Dam. 1989. Inferring pH from diatoms: a comparison of old and new calibration methods. Hydrobiologia **178**:209-223.

ter Braak, C. J. F. and P. F. M. Verdonschot. 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. Aquatic Sciences **57**:255-289.

Torgerson, W. S. 1958. Theory and methods of scaling. Wiley, New York.

van Dam, H., G. Suurmond, and C. J. F. ter Braak. 1981. Impact of acidification on diatoms and chemistry of Dutch moorland pools. Hydrobiologia **83**:425-459.

van den Wollenberg A.L. (1977) Redundancy analysis. An alternative for canonical correlation analysis. Psychometrika 42:207-219.

van der Burg, E. and J. de Leeuw. 1983. Non-linear canonical correlation. British Journal of Mathematical and Statistical Psychology **36**:54-80.

van der Burg, E. and J. de Leeuw. 1990. Non-linear redundancy analysis. British Journal of Mathematical and Statistical Psychology **43**:217-230.

Van der Heijden, P. G. M., A. Mooijaart, and Y. Takane. 1994. Correspondence analysis and contingency models. Pages 79-111 *in* M. J. Greenacre and J. Blasius, editors. Correspondence analysis in the social sciences. Academic Press, New York.

Wartenberg, D., S. Ferson, and F. J. Rohlf. 1987. Putting things in order: a critique of detrended correspondence analysis. American Naturalist **129**:434-448.

Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. Ecol. Monogr. **26**:1-80.

Yoccoz, N. G. and D. Chessel. 1988. Ordination sous contraintes de relevés d'avifaune:elimination d'effets dans un plan d'observation à deux facteurs. Compte-Rendus de l'Académie des Sciences,Paris,Série III **307**:189-194.

Zelinka, M. and P. Marvan. 1961. Zuer Praezisierung der biologischen Klassifikation der Reinheit fliessender Gewaesser. Arch. Hydrobiol. **57**:389-407.

Zhang, Y. and O. Thas. 2012. Constrained ordination analysis in the presence of zero inflation. Statistical Modelling, 12, 463-485.

Zhu, M., T. J. Hastie, and G. Walther. 2005. Constrained ordination analysis with flexible response functions. Ecological modelling 187:524-536.

Appendix A
Canonical correspondence analysis via singular value decomposition
The two-sided eigen-equation (1) of canonical correspondence analysis can equally
well be obtained via a singular value decomposition (SVD). On using the notation of
the main text, define

$$\mathbf{Y}^* \equiv (\mathbf{Z}^\mathsf{T}\mathbf{R}\mathbf{Z})^{-1/2} \mathbf{Z}^\mathsf{T}\mathbf{Y}\mathbf{K}^{-1/2} \tag{A.1}$$

and obtain the SVD of $\mathbf{Y}^*$,

$$\mathbf{Y}^* = \mathbf{P} \, \boldsymbol{\Sigma} \, \mathbf{Q}^\mathsf{T} \tag{A.2}$$

with $\mathbf{P}$ and $\mathbf{Q}$ orthonormal matrices and $\boldsymbol{\Sigma}$ a diagonal matrix with singular values that
are the square-root of the CCA eigen values ($\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^{1/2}$). The canonical weights ($\mathbf{b}$) and
the taxon scores ($\mathbf{u}$) for the various dimensions are then in the columns of the matrices

$$\mathbf{B} = (\mathbf{Z}^\mathsf{T}\mathbf{R}\mathbf{Z})^{-1/2} \, \mathbf{P} \, \boldsymbol{\Sigma}^\alpha \tag{A.3}$$

and

$$\mathbf{U} = \mathbf{K}^{-1/2} \, \mathbf{Q}\boldsymbol{\Sigma}^{1-\alpha}, \tag{A.4}$$

respectively, where α is a biplot scaling constant ($0\le \alpha \le 1$); see Table 1. The third
triplet in Table 2 and associated biplot follows directly from these equations. If $\mathbf{Z}$
contains a column with ones for the intercept, as usual in multiple regression, the first
singular value is 1, and the first column of $\mathbf{B}$ is $\mathbf{b}_1 = (1,0,0,\ldots, 0)^\mathsf{T}$ and that of $\mathbf{U}$ is
$(1,1,1,\ldots,1)^\mathsf{T}$, similar to the trivial first solution in CA. It can be avoided by first $\mathbf{R}$-
centring the columns of $\mathbf{Z}$ such that $\mathbf{Z}^\mathsf{T}\mathbf{R}\mathbf{1}_n=0$, as can be seen as follows.

As $\mathbf{P}$ and $\mathbf{Q}$ are orthonormal, $\mathbf{B}^\mathsf{T}(\mathbf{Z}^\mathsf{T}\mathbf{R}\mathbf{Z})\mathbf{B} = (\mathbf{Z}\mathbf{B})^\mathsf{T}\mathbf{R}(\mathbf{Z}\mathbf{B})= \boldsymbol{\Sigma}^{2\alpha}$ and $\mathbf{U}^\mathsf{T}\mathbf{K}\mathbf{U} = \boldsymbol{\Sigma}^{2(1-\alpha)}$
showing that the site scores $\mathbf{X} = \mathbf{Z}\mathbf{B}$ and species scores are $\mathbf{R}$- and $\mathbf{K}$-orthogonal,
respectively. With the first site score vector $\mathbf{x}_1 = \mathbf{Z}\mathbf{b}_1 = \mathbf{1}_n$, we thus have $\mathbf{x}^\mathsf{T}\mathbf{R}\mathbf{1}_n =$
$\mathbf{b}^\mathsf{T}\mathbf{Z}^\mathsf{T}\mathbf{R}\mathbf{1}_n = 0$, for any higher numbered canonical weight vector $\mathbf{b}$ and its associated
site score vector $\mathbf{x}$. Similarly, for any higher numbered taxon score vector $\mathbf{u}$, $\mathbf{u}^\mathsf{T}\mathbf{K}\mathbf{1}_m=$
0. These null equations hold trivially true, when we apply $\mathbf{R}$-centring to $\mathbf{Z}$ and thus
entail no additional constraint to the eigenvalue problem. If $\mathbf{Z} = \mathbf{I}_n$, the SVD (A.1) of
CCA reduces to that of CA and the centring the identity matrix amounts to the usual
centring of the matrix $\mathbf{Y}$ in CA. The trivial solution and the way to avoid it comes as
no surprise when you understand CA.

Equations (A.1)-(A.4) reduce to those of principal components analysis with respect
to instrumental variables (Rao, 1964), alias redundancy analysis, by redefining $\mathbf{R}= \mathbf{I}_n$
and $\mathbf{K}= \mathbf{I}_m$ with corresponding unweighted column centring of By contrast, in CCA
$\mathbf{R}$ and $\mathbf{K}$ contain the row and column totals of $\mathbf{Y}$ without which there is no weighted
averaging (equations (2) and (3)) and no link to the unimodal model (Figure 2). This
shows that CCA is more than just a minor generalization of redundancy analysis.
Section 5 serves to show that CCA is much closer related to multi-group linear
discriminant analysis.