



Topics in constrained and unconstrained ordination

Cajo J.F. ter Braak and Petr Šmilauer

This is a "Post-Print" accepted manuscript, which has been published in
Plant Ecology.

This version is distributed under the [Creative Commons Attribution 3.0 Netherlands](https://creativecommons.org/licenses/by/3.0/nl/) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Please cite this publication as follows:

ter Braak, C. J. F. and P. Šmilauer. 2014. Topics in constrained and unconstrained ordination. Plant Ecology. DOI: 10.1007/s11258-014-0356-5

The final publication is available at Springer via:

<http://dx.doi.org/10.1007/s11258-014-0356-5>

Topics in constrained and unconstrained ordination

Cajo J.F. ter Braak^a and Petr Šmilauer^b

^aBiometris, Wageningen University and Research Centre, Box 100, 6700 AC, Wageningen, the Netherlands

^bFaculty of Science, University of South Bohemia, Branišovská 31, České Budějovice, Czech Republic

Corresponding author:

Cajo J.F. ter Braak

e-mail: Cajo.terBraak@wur.nl

phone: +31317480803

fax: +31317483554

Abstract

In this paper we reflect on a number of aspects of ordination methods: how should absences be treated in ordination and how do model-based methods, including Gaussian ordination and methods using generalized linear models, relate to the usual least-squares (eigenvector) methods based on (log-)transformed data. We defend detrended correspondence analysis by theoretical arguments and by reanalysing data that previously gave bad results. We show by examples that constrained ordination can yield more informative views on effects of interest compared to unconstrained ordination (where such effects can be invisible) and show how constrained axes can be interpreted. Constrained ordination uses an ANOVA/regression approach to enable the user to focus on particular aspects of species community data, in particular the effects of qualitative and quantitative environmental variables. We close with an analysis examining the interaction effects between two factors and we demonstrate how principal response curves can help in their visualisation. Example data and Canoco 5 projects are provided as Supplementary Material.

Keywords multivariate analysis . ordination . constrained ordination . interaction . experimental design . data transformation . principal response curves

Electronic supplementary material: The online version of this article (doi:10.1007/s11258-014-0356-5) contains supplementary material, which is available to authorized users.

Introduction

In 1954 (CtB's year of birth), David Goodall published "An essay in the use of factor analysis" and introduced therein the term "ordination" as we use it nowadays, that is, for "Putting things in order" (Peet et al. 1988; Palmer 2014). David Goodall used the term as antonym of "classification" and introduced the term as he felt that in factor analysis the emphasis was on the variables being analysed, whereas he was interested "in arrangement of the vegetational data in a multidimensional series" (Goodall 1954), in his situation "sample areas" or quadrats. Goodall actually used principal component analysis (PCA) remarking that "with ordinary computing methods it is bound to be very laborious, and is likely to be used in fundamental investigations into the structure of vegetation rather than in more descriptive work". How things have changed. Now PCA, an unconstrained ordination method, is an easy thing to do with many more complex methods as competitors!

Let's briefly introduce, in historical order, all ordination methods and their abbreviations, referred to in this paper. Ten years after 1954, C.R. Rao (1964) introduced the method of principal component analysis with respect to instrumental variables that was reinvented several times (including by CtB) and is known to many as redundancy analysis (RDA) (van den Wollenberg 1977) or (least-squares) reduced-rank regression (Davies and Tso 1982). RDA is like a PCA but it uses case scores that are constrained by linear combinations of external, explanatory variables. (We use the term "case" for quadrat/sampling unit.) RDA is thus the constrained counterpart of PCA. Also in 1964, Joseph B. Kruskal introduced non-metric multidimensional scaling (NMDS), an unconstrained method using only the rank order of inter-case dissimilarity values, whereas Gower (1966) proposed a PCA of dissimilarity values under the name of principal coordinate analysis. Seven years later, Mark O. Hill (1973) extended the method of weighted averaging (WA), invented by vegetation scientists, to reciprocal averaging, alias correspondence analysis (CA), so as to extract (indirectly) from community data the gradients reflecting species turnover. Species turnover and the niche concept are closely related to unimodal models (ter Braak 1987b). Some of the faults of CA were repaired in detrended correspondence analysis (DCA) by Hill and Gauch (1980). The constrained counterpart of CA is canonical correspondence analysis (CCA) by CtB (ter Braak 1986, 1987a). A history of CA, DCA and CCA can found in ter Braak (2014). The last abbreviation in this paragraph is PRC, which stands for Principal Response Curves (van den Brink and ter Braak 1998; van den Brink and ter Braak 1999), a method designed to display time-dependent treatment effects on species communities.

David Goodall started with PCA but in his later work he worked hard to popularize maximum likelihood Gaussian ordination, giving particular attention to how to treat the qualitative aspect of presence-absence in combination with cover abundance (Johnson and Goodall 1979; Goodall and Johnson 1982, 1987). This paper therefore starts with a discussion of this topic. The numerical and computational challenges in unconstrained multi- or even two-dimensional Gaussian ordination are in our opinion not yet completely solved today. Promising new attempts and approaches are stimulated by the advent of generalized linear models (GLM) and generalized linear mixed models (GLMM) (Breslow and Clayton 1993; Engel and Keen 1994; Zuur et al. 2008). The alternative is to use least-squares methods on transformed data, and use

permutation methods for statistical significance testing. We review two papers comparing transformed least-squares and GLM. Our own view and experience in this area is well represented by the remark in Warton (2005) “Surprisingly, transformed least squares appeared to fit data about as well as” a GLM model. We therefore continue with transformed least-squares as a computationally attractive approach in ordination context, leading to consideration of PCA, CA, DCA and CCA, PRC and distance-based relatives.

We then show the advantage of using a constrained approach as compared to a distance-based (NMDS) approach, using an example that Warton et al (2012) used to criticize distance-based methods, RDA and CCA. We advocate RDA and CCA here, as the transformed least-squares approach using constrained ordination works surprisingly well in our opinion.

DCA has both been applauded and criticized, as reviewed in Jackson and Somers (1991) and the same applies to CCA (Palmer 1993; Johnson and Altman 1999). Transformed PCA and RDA are perhaps easier to understand than DCA and CCA (Legendre and Legendre 2012; Borcard et al. 2011) and their theory is simpler, but this does not preclude that many variants of PCA cannot recover the true underlying gradients in data with species turnover, that is data with long gradients (Oksanen and Tonteri 1995) and many zeroes. We reanalyse a simulated data set used in Minchin (1987) for which untransformed DCA gave a particularly bad result. We extend this data set with simulated explanatory variables so as to illustrate the power of constrained ordination, that is, an unconstrained ordination combined with multiple regression. We conclude with another option provided by the regression approach, which is the analysis of interaction effects, for which we use RDA and PRC.

Data sets

Recently we published the Canoco 5 program (ter Braak and Šmilauer 2012; Šmilauer and Lepš 2014) which has a modern user interface with multiple numerical and graphical methods. All example data sets and Canoco 5 projects are provided as Supplementary Material. The Canoco 5 projects detail how analyses are carried out; these details can be inspected by any reader by asking for an one-month trial copy of Canoco 5 at trial@microcomputerpower.com. All numerical analyses can also be executed in R using *vegan* (Oksanen et al. 2013).

Tikus Island data set

The Tikus Island data set (Warwick et al. 1990) is on coral assemblages and is available in the R package *mvabund* (Wang et al. 2012). As Warton et al. (2012), we use the cases from 1981 and 1983 (ten in each year) containing the counts of 61 taxa. The same ten locations were sampled each year (Warwick et al. 1990).

Minchin87_2b data set

Minchin87_2b is a simulated data set (Minchin 1987) with 48 cases and 56 species, which we choose because its DCA ordination (figure 2b in Minchin 1987) is particularly bad. The true configuration of the 48 cases is a regular 12×4 grid forming a $1 \times 0.33 R$ coenoplane where “ $1R$ is equal in length to the mean range of species occurrence” (Minchin 1987). We pretend that the data represents the deposition of pollen from various tree and shrub species in bogs in different parts of Europe and add three simulated explanatory variables to mimic an analysis originally performed by Colin Prentice (personal comm.) in the late 1980s on genuine pollen data across Europe. The variables were on climate: mean temperatures in January and July ($^{\circ}\text{C}$) and annual precipitation (mm), abbreviated as JanT, JulyT and Precip, respectively and were related (but with noise) to the true gradients. For two cases, supposedly paleo-samples from a core and dated 3000 years B.C., no climate data was available and one of the aims was to infer the climate 3000 years B.C. from the pollen data. To this aim it is important to understand which climate variables appear ecologically important.

Sod data set

The Sod data is from an experiment reported in Baar and ter Braak (1996) investigating the decline of ectomycorrhizal fungi species in Scots pine stands in the Netherlands during the 20th century and is also part of the Canoco 5 examples (Splitplot example). The aim of the experiment was to investigate whether removal/addition of litter and humus layers would influence the number of species and sporocarps of fruiting ectomycorrhizal fungi in Scots pine stands (Baar and ter Braak 1996). To this aim, an experiment was carried out in six Scots pine stands of different age (3-66 years). In each stand 12 plots were laid out to which three treatments (Control, Sod-cutting, sod-Addition) were applied in 4 replicates, except that Sod-addition plots were not created in the two oldest stands. In total there are 64 plots and 42 taxa. The design variables are stand, treatment and age of the stand. Age was log-transformed and thereafter treated either quantitative or as a factor with four classes (1-4) of increasing age. The names of taxa are abbreviated to 8 letter codes; for full names see table 7 of Baar and ter Braak (1996).

Topics

Model-based ordination and the treatment of absences

In ecology, unconstrained ordination based on the unimodal (niche) model, in particular the Gaussian model, has been explored with three important contributions by David Goodall (Johnson and Goodall 1979; Goodall and Johnson 1982, 1987). A key aspect of Goodall's work was the combination of the qualitative and quantitative features of community data, in the form of a combined likelihood for presence-absence data and the (cover) abundance data. Goodall did not yet refer to GLMs nor to a GLM-type algorithm, but he did use a link function (arctangent) to

link probabilities to the linear predictor (a second order polynomial). His approach never became really popular, presumably because of its complexity, numerical problems in more than one dimension and limited availability of software dealing with this complexity and these problems. Nowadays some R packages facilitate the use of multiple likelihoods (Martins et al. 2013), so perhaps Goodall's approach may revive.

With a limited number of predictors, maximum likelihood (model-based) *constrained* ordination is actually much easier, because the number of unknown parameters is much smaller. This led to new attempts to popularize the approach (Johnson and Altman 1999; Zhang and Thas 2012) and Yee (2004, 2006) developed the R package VGAM that allows GLM-based dimension reduction and ordination including (constrained) Gaussian ordination. These approaches have had little follow-up according to the Web of Science, presumably due to limitations on the number of species, steepness of the learning curve and low numerical stability of the algorithm. The VGAM documentation on Gaussian quadratic ordination says, for example: "The number of species should be kept reasonably low, e.g., 12 max. Feeding in 100+ species wholesale is a recipe for failure" and Yee (2004) recommends CCA over constrained Gaussian ordination "for large and/or noisy/dirty data sets due to its lower computational expense and robustness".

The simulations in Zhang and Thas (2012) show the importance of going beyond the Poisson and negative binomial distributions and associated likelihoods in the case of zero inflation, that is for data with more zeroes than expected under these distributions. Zero inflation is a logical pattern in community data as a species may be absent due to unfavorable values of unmeasured variables, of which there are likely many. By contrast, Warton (2005) found little need for zero-inflated models (without dimension reduction) because zeroes "are more likely to arise from negative binomial distributions with small means than from zero-inflated negative binomial distributions", as judged by comparisons on real data. Constrained ordination necessarily simplifies the predictor space and may have difficulty fitting small means for every particular species where these are needed; allowing zero-inflation in constrained ordination may be a way out. Note that weighted averaging methods disregard the zeroes in the species data, and thus (D)CA and CCA are not hampered by zero inflation (Zhang and Thas 2012). Distance-based methods that use coefficients that disregard double absences might also be relatively insensitive to zero inflation.

Data transformation and generalized linear models

The variables measured in ecological research often take non-negative values only and are often right-skewed, with some large values that look like outliers on the original scale. Such outliers have a large impact on the statistical analysis. This remark also applies to cover data with one dominant species in some plots. In order to ensure that the statistical analysis summarizes the main body of the data, such variables need to be transformed by an *a priori* data transformation (often square root or log) or be analyzed by assuming an appropriate (non-normal) statistical distribution as in generalized linear models (GLM, (McCullagh and Nelder 1989). We discuss here the transformation of individual variables and not transformation and

standardization across species and cases as discussed in Legendre and Gallagher (2001) and Legendre and Legendre (2012).

The data transformation that makes most ecological sense is the logarithmic transformation as it turns ecologically plausible multiplicative models into mathematically attractive additive ones and also makes the outliers less extreme. With zero values in the data, it makes sense to use $\log(y+c)$. A natural choice for c is the minimum non-zero value in the data, which for count data leads to the $\log(y+1)$ transformation. There are two recent papers discussing this transformation in relation to GLM. O'Hara and Kotze (2010) already state in the title of their paper "Do not log-transform count data". Their conclusion is supported by simulations of negative binomial data and analysing it either by GLM or by least-squares after log-transformation. Their results showed bias and higher root-mean square error in the mean on log-scale in the transformation approach as compared to the GLM approach based on the quasi-Poisson and negative binomial distribution.

But O'Hara and Kotze (2010) do not cite the second paper, Warton (2005). As a side result of an investigation on the need for zero-inflated distributions in 20 ecological data sets, Warton (2005) expresses his surprise that $\log(y+1)$ "transformed least-squares, although a relatively crude approach, can fit abundance data reasonably well". His "first choice model for abundance [was] negative binomial log-linear models, and [transformed] least-squares only [to] be used in complex models where the negative binomial cannot be used easily, ...". Comparing weighted averaging and GLM-based Gaussian model fitting, ter Braak and Looman (1986) similarly recommended GLM, but that recommendation did not preclude the use of weighted averaging in the multivariate case where Gaussian model fitting is much more difficult (ter Braak 1985, 1986). But perhaps generalized linear mixed models may help change this case (Jamil and ter Braak 2013). Multivariate versions of GLMs are being proposed under the name of manyGLM (Warton et al. 2012; Wang et al. 2012) but these go as yet without dimension reduction; but see Yee (2004, 2006). Comparing GLM and CCA, Guissan et al. (1999) concluded that "In general GLM will provide better species-specific models, but CCA will provide a broader overview of multiple species, diversity and plant communities".

In the examples in this paper we adopt the log-transformed least-squares approach throughout. We even transform, as in ter Braak (1986), in the case of (D)CA and CCA where the theory formally starts from the (quasi) Poisson distribution (ter Braak 1985). This avoids overweighing of abundant species, as the species total is the implicit weight in these methods (see Jongman et al 1995 and below). Note that presence-absence data cannot be usefully log-transformed.

Constrained versus unconstrained ordination

Often the ordination diagrams of unconstrained and constrained analysis do not differ much. This happens in cases where the main variation in community composition can be explained by the constraining variables (for example compare Fig. 2h with Fig. 2b). However, in other cases, unmeasured variables may so strongly influence the community or yield so much additional variation that large differences occur between constrained and unconstrained low dimensional representations.

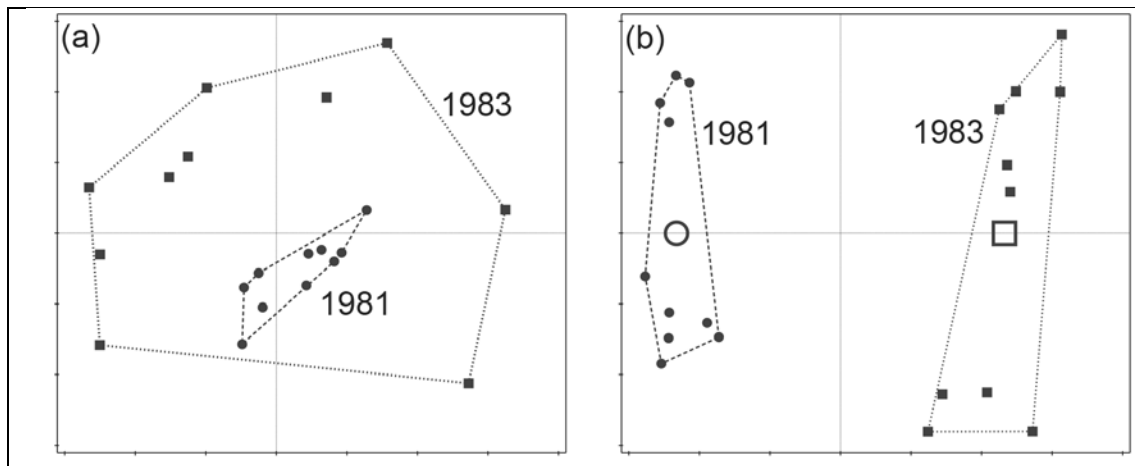


Fig. 1

Tikus island data set: (a) NMDS ordination and (b) distance-based redundancy analysis with constraining factor year using percentage dissimilarity calculated from the $\ln(y+1)$ transformed data with the centroids per year indicated by larger and open symbols of the same shape as the closed symbols representing the individual cases.

Fig. 1 is an extreme example using the Tikus Island data set. The counts are highly variable and therefore we first log-transformed the data using $\ln(y+1)$. The ordination diagram based on NMDS using percentage difference (Bray-Curtis distance, D14 in Legendre and Legendre (2012)) suggests that the main difference between the two years is that the cases in 1983 are much more spread out (thus much more variable) than those from 1981 (Fig. 1a). It is thus tempting to conclude that there are dispersion differences, but hardly any difference in location, that is no differences among mean abundance values of the species between the years.

The problem of the NMDS is that it does not allow the user to focus on particular effects of interest, in this case the differences in location between years. But this is what constrained ordination allows the user to do. Because constrained NMDS is not very well developed, we use an alternative, namely, distance-based redundancy analysis (Legendre and Anderson 1999). The ordination diagram based on this method using the same dissimilarity measure and using the factor year as a constraining variable shows a big difference in location of the cases of different years along the first (horizontal) axis (Fig. 1b). As we consider only two years, the constraint represents only one degree of freedom and therefore constrains the first axis only. The second (vertical) axis is again unconstrained and shows a slightly higher dispersion of the cases in 1983 than in 1981. The differences in location is highly significant as judged by a paired permutation test ($P \leq 0.003$). In this paired test the cases at the same locations but from different years are randomly reshuffled in each permutation. We used this test, rather than an unpaired test, because the same locations were sampled each year (Warwick et al. 1990). This test extends the usual paired t-test to multivariate data or multivariate (dis)similarity data. The analysis of related cases is described as an application of partial RDA in Legendre & Legendre (2012, section 11.1.10.3, p. 655).

Fig. 1b shows in fact the scores, which are derived from the dissimilarity values (called CaseR scores in Canoco, as they are derived from the dissimilarity data that form the Response data in this analysis). The constrained scores (called CaseE, as

they derived from the Explanatory data) coincide with the mean scores per year, which are the centroids in Fig. 1b. These terms replace the terms WA scores and LC scores (e.g. Graffelman and Tuft 2004), which are specific for CCA. The advantage of using the CaseR scores is that they display the within-year variability along the first axis and so allow us to see how well the years are separated in terms of community composition.

Warton et al. (2012) argued against distance-based methods and many other ordination methods including RDA and CCA, because they appear to confuse the differences in dispersion and differences in mean, whereas their GLMs consider the differences in dispersion as a natural result of the differences in mean. However, constrained ordination, including the distance-based version, easily displays differences in mean, as in Fig. 1b, because the constraint acts on location (and not on dispersion) and the log-transformation used before calculating the dissimilarities takes away much of the differences in dispersion in the constrained ordination, but not so in the unconstrained NMDS.

DCA, CCA and the chi-square distance

One of the reasons that Goodall's (1979-1987) method of non-linear ordination did not really take off may have been the rise of approximate methods (DCA and its constrained form CCA) and of distance-based methods such as NMDS (Minchin 1987; Faith et al. 1987).

Ter Braak and Looman (1986) and ter Braak and Barendregt (1986) showed the link of the method of weighted averaging with the ecological niche model and the Gaussian response model in particular. Ter Braak (1985) combined these results to show that the estimation equations of (D)CA (its transition formulas) are an approximation of the maximum likelihood equations of the Gaussian ordination model for Poisson distributed counts and presence-absence data. A similar rationale was used for CCA (ter Braak 1986). The approximation works best for the Gaussian model in which species have equal niche width, but that is not the same as stating that it does not work for non-Gaussian unimodal models and unequal tolerances. This was pointed out theoretically for weighted averaging in the context of calibration (ter Braak and Barendregt 1986) and, in real data sets, tolerance weighing of taxa led to little gain or loss in efficiency in some environmental reconstruction studies (ter Braak and van Dam 1989; Birks et al. 1990) but not in all (Juggins and Birks 2012). Moreover, the approximation works also for data sets with little structure (small eigenvalues, large tolerances) as pointed out by Ihm and van Groenewoud (1984); see ter Braak (2014) for details. Weighted averaging methods disregard the zeroes in the species data, and thus (D)CA and CCA are not hampered by zero inflation.

Note that there is no mention of the chi-square distance in this motivation of DCA and CCA, despite the fact that, in full dimensional space, the inter-case distances in ordination by CA are equal to the calculated chi-square distances (Greenacre 1984; Legendre and Legendre 2012). The reason is one of the serious shortcomings of CA, namely that, for data with unimodal structure, the full space of CA contains spurious polynomial axes that all contribute to the chi-square distance. In two dimensions, the CA configuration of cases (e.g. Fig. 2d) and species then shows an arch (Legendre and Legendre, 2012, p. 483). Even when there is only a single true ecological gradient, the chi-square distances are better represented in two dimensions, whereas

the ecological distance is better measured along the first (horizontal) axis only. Jongman et al. (1995, p. 105) explain why and when the arch effect can crop up, thereby supporting its removal as in DCA (Hill and Gauch 1980). With strong unimodal structure, the relation between chi-square distance and the ecological distance is worse compared to many other distance measures (Faith et al. 1987; Legendre and Gallagher 2001), but this is not necessarily so after removal of the spurious axes by detrending as in DCA (Hill and Gauch 1980). A downside of DCA is that it can be sensitive to the number of segments used in the detrending algorithm (Jackson and Somers 1991). In CCA the spurious axes are often suppressed by the external linear constraints, so then the detrending is not needed (ter Braak 1986). Due to the removal of spurious polynomial axes either by detrending (as in DCA) or by the external constraints (CCA), distances among cases and among species reproduced by the ordination axes do no longer aim to represent their chi-square distances. Instead, they aim to represent ecological distance. For the same reason, the percentage variation accounted for (where the denominator is the total inertia, that is, chi-square in full space) is a lower bound for the true percentage explained, and the same applies to the adjusted percentage (Peres-Neto et al. 2006).

(D)CA is said to have a second major shortcoming, namely, that rare species may have an unduly large influence on the analysis” (Legendre and Gallagher 2001) as appears visible from the chi-square distance in which “A difference between abundance values for a common species contributes less to the distance than the same difference for a rare species” (Legendre and Gallagher 2001). This influence happens in particular when there are cases with rare species only. This shortcoming is the consequence of the nice property that CA is able to discover sets of cases that have no species in common, a property of CA that is exploited in the two-way cluster analysis program TWINSpan (Hill 1979; Jongman et al. 1995). Greenacre (2013) argues that such disconnected sets are an important aspect of the data that CA rightly discovers, but he also demonstrates on real data sets that the contribution of rare species to the chi-square distance is small. This is in line with the idea that the species total is the implicit weight in CA (see above) as is clear from the reciprocal averaging algorithm for CA and definition of the dispersion that it maximizes (ter Braak 1987a; Jongman et al. 1995). This weighing, which is not at all obvious from the definition of the chi-square distance, as indicated above, implies that rare species do not have an unduly large influence on the analysis. We will illustrate this below using the Minchin87_2b data set.

The “fault” that remains is that rare species occupy outlying positions in the (D)CA ordination. Greenacre (2013) therefore proposes an alternative way of plotting, in which the species positions are (unfortunately) no longer related to niche centers. Rare species can be suppressed in ordination diagrams based on their total (weight). The influence of rare species in CCA is even smaller because the explanatory variables effectively “combine” the individual cases. Take, for example, a single factor as explanatory variable. Then the data table effectively reduces (in its representation on constrained CCA axes) to a smaller table. The rows of the table are the classes of the factor and its entries are the total abundance in the classes.

There exists another derivation for CCA that does not use the chi-square distance either. Ter Braak (1987a) derived CCA as a method that finds linear combinations of explanatory variables that best separate the species niches as measured by the weighted variance of the niche centers expressed as a fraction of the total variance of the case scores. This derivation can actually start from the Heiser (1987) unfolding

model. In this derivation CCA draws strength from canonical discriminant analysis (Zhu et al. 2005; ter Braak 2014). Zhu et al (2005) and Zhang and Thas (2012) used this connection to derive model-based versions of CCA.

Illustration using the Minchin87_2b data set

In this subsection we illustrate some of the theory of the previous sections using the Minchin87_2b data set.

Fig. 2a shows the sample ordination diagram of a DCA on the untransformed Minchin87_2b data set. Recall that the ideal configuration is here a 12×4 regular grid. In Fig. 2a, the variation at the left end of the primary gradient is clearly compressed as compared to its variation at right end. In other words, DCA did not recover the effect of the secondary gradient at the left end of the primary gradient. However, the count-type data are highly variable with many extreme values and with species having very unequal total counts. It appears therefore prudent to use the $\log(y+1)$ transformation before applying DCA. The DCA of the log-transformed data does not have the compression (Fig. 2b); the variation at the left end of the primary gradient that is due to the secondary gradient is even somewhat exaggerated.

The contribution of species to the inter-case chi-square distances (as defined in Greenacre 2013) steeply increases with total abundance for the untransformed data, but far less so in the transformed data (Fig. 3). In this data set there are 13 species that occur in less than five cases (and 1 species occurs once) and these species also have low relative abundance. Fig. 3 shows that these species do not have a high influence on the analysis. The $\log(y+1)$ transformation made the contribution of the species much more equitable. In the subsequent analyses, the data are always first $\log(y+1)$ transformed.

For comparison, the PCA of transformed data (Fig. 2c) shows what is known as the horseshoe effect (Jongman et al. 1995); the order of cases along the horizontal axis is at either side greatly in conflict with the order along the true primary gradient. The CA shows the famous arch or Guttman effect with its scores showing a parabolic relationship with the scores of the first axis (Fig. 2d). The order of cases along the horizontal axis is in much better agreement with the true order along the primary gradient, but the CA did not recover the secondary gradient. Note that the order of the cases along the horizontal axis in DCA is identical to that of CA and that the DCA does recover much of the true secondary gradient.

It has been suggested that PCA of Hellinger-transformed data, which preserves the Hellinger distances among cases, is suited for ecological community data (Legendre and Gallagher 2001). The Hellinger distance transformation, prominently present in (Borcard et al. 2011) and in Canoco 5 as a case standardization option, expresses abundances as fractions of the case total, takes a square-root and then calculates a Euclidean distance from the so-transformed data. With (Fig. 2e) or without prior log-transformation, PCA on Hellinger-transformed data does not recover the secondary gradient nor does it recover the correct order of the cases along the first gradient. Fig. 2e shows even a nicer example of a horseshoe than the initial PCA (Fig. 2c). Our final unconstrained ordination for this data is NMDS using percentage difference (Fig. 2f) which does slightly better than both the DCA and the local NMDS applied in Minchin (1987). Some users may be happy with an ordination like Fig. 2e, because it correctly

reflects the calculated dissimilarities, but we are happier with an ordination like Fig. 2f, because it represents the true underlying ecological gradients.

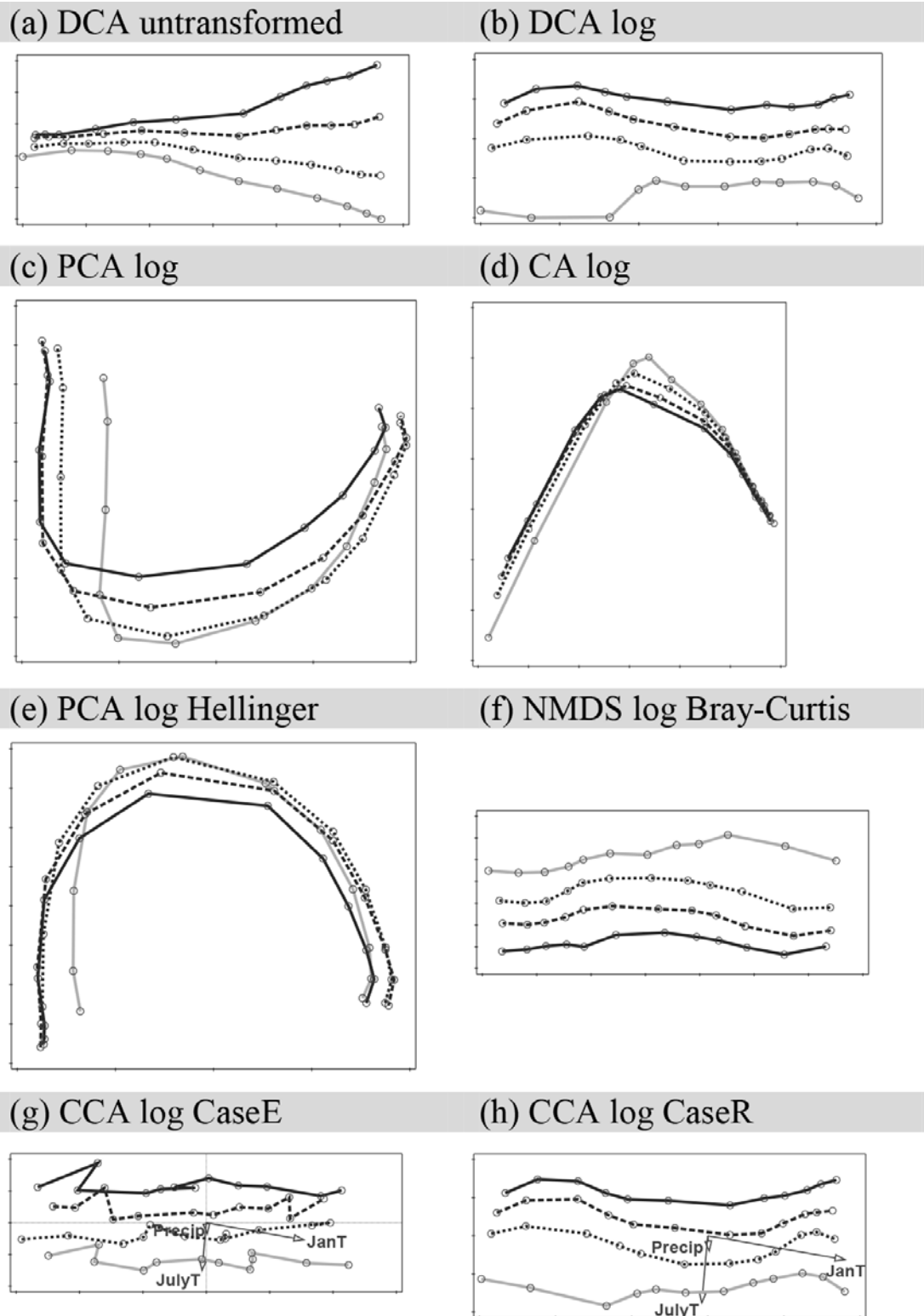


Fig. 2 Minchin87_2b: Two-dimensional sample ordinations of the 12×4 coenoplane data by different methods and transformations. The scaling of axes had focus on inter-case distances. Hill's scaling is used in d, g and h. The arrows in g and h represent the individual effects of the climate variables on the case configuration.

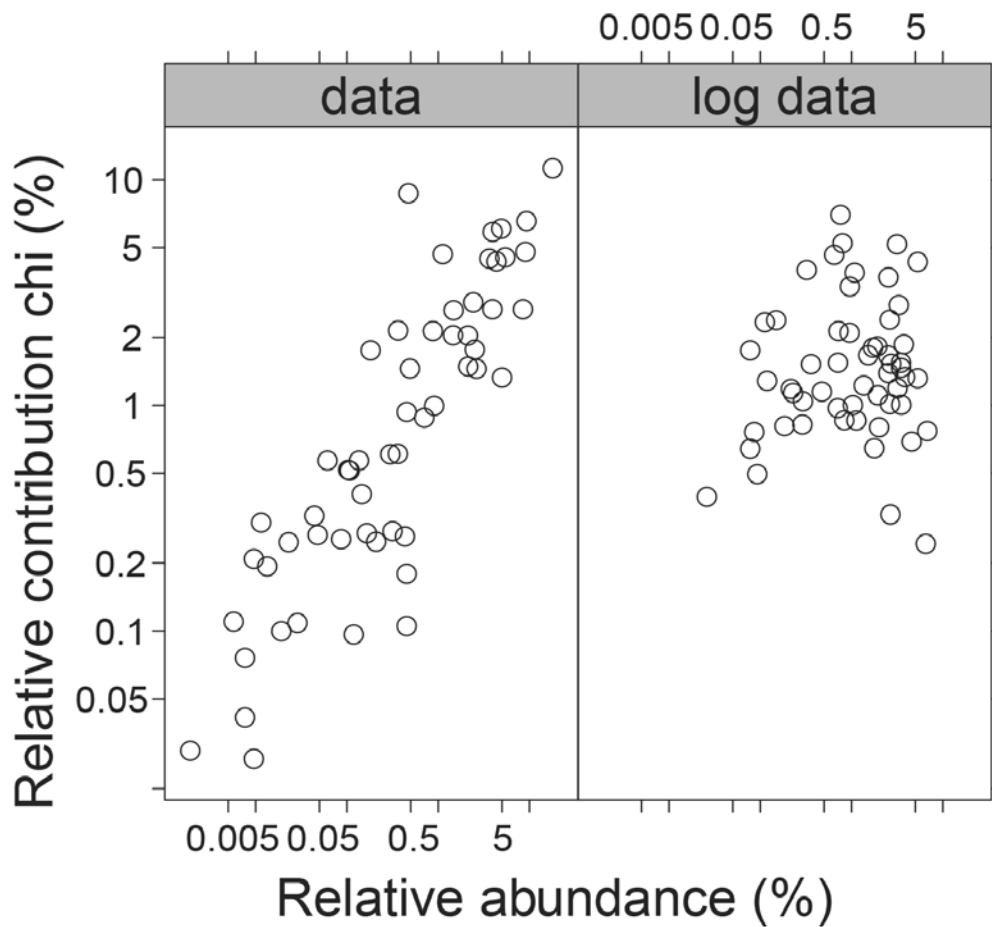


Fig. 3 Relative contributions to chi-square distance against the relative abundances for the 56 species in the Minchin87_2b data set. Both axes have logarithmic scales.

The summary statistics of the DCA include eigenvalues and gradient lengths. The gradient length is, like Minchin's R, a measure of beta-diversity. Gradient length is in standard deviation units (SD, also called turnover units); 1 SD is equal to the mean within-species variance of the case scores along the axis. By linking variance to range of occurrence, we expect $1SD \approx 4R$. The estimated gradient lengths of the first two axes are 4.77 and 1.67 and translate to a 1.2×0.4 R coenoplane. The ratio of the gradient lengths ($1.67/4.77$) is 0.35, which corresponds even better with the ratio of the R measures ($0.33/1$) of the simulated coenoplane. With gradient lengths greater than $4SD$ we expect the data to show clear unimodal (niche) structure. See Oksanen and Tonteri (1995) for more on gradient length calculations. The gradient length of the third axis (1.52) is comparable to that of the second and thus fails to indicate that there is no true third gradient in this data. The first two eigenvalues of the DCA (0.70 and 0.16) are much larger than those predicted by the broken stick model (Frontier 1976)(Legendre and Legendre 2012) whereas the third eigenvalue (0.08) is close to the value predicted, suggesting that the true number of gradients is two.

The first two eigenvalues of the CCA with respect to the three climate variables (0.64 and 0.17) are of the same order of magnitude as the eigenvalues of the DCA, but the third eigenvalue (0.003) is close to zero, showing that the three climate variables effectively define two gradients only. This is confirmed by significance tests for the

constrained axes (Legendre et al. 2011), yielding $P < 0.001$ for the first two axes and $P \approx 1$ for the third axis. The fourth axis is unconstrained and has a large eigenvalue (0.38). A plot of the case scores of axis 4 against the first axis (not shown) displays a clear arch, very much like Fig. 2d, confirming that axis 4 does not represent a true gradient. A detrended CCA yields similar eigenvalues for the first three axes, but the fourth eigenvalue is reduced to 0.12.

Finally, Fig. 2g shows the sample ordination diagram of a CCA in which the case scores are constrained to be linear combination of the noisy climate variables whereas Fig. 2h shows the WA scores (CaseR scores in Canoco 5), which can be shown to better approximate the community data in conjunction with the species scores (Graffelman 2001) and do here better in recovering the underlying coenoplane (Fig. 2h); see also McCune (1997). The biplot arrows for the climate variables in the CCA graphs (Fig. 2g,h) suggest that the first gradient is related to mean January temperature and the second gradient to mean July temperature, with precipitation having a negligible influence. However, this is not the full story as we will explain in the following section.

We also attempted constrained quadratic ordination (cgo) in VGAM on the data, so as to go beyond the equi-tolerance assumption of CCA, but VGAM did not converge for the two-dimensional model and we did not obtain useful output.

Interpreting constrained axes

There are two major ways for interpreting constrained axes in terms of the explanatory variables. The first and easiest way is based on correlations between explanatory variables and the constrained axes, and quantities derived from these, such as the usual environmental biplot arrows and class centroids (ter Braak 1994; ter Braak and Šmilauer 2012). The second way is based on the standardized regression coefficients (and associated t-values) of the explanatory variables that together define the constrained axes. Regression coefficients and correlations are scaled versions of one another, when the explanatory variables are uncorrelated, but not so if the explanatory variables are (highly) correlated, as happens often in exploratory studies. Then, the interpretations may differ and some caution is prudent, particularly if the t-values are small in absolute value (< 2), as an explanatory variable with a small associated t-value has no significant contribution to the axis.

Table 1 Intra-set correlations (CorE) and regression coefficients (RegrE) with associated t-values (TValE) of the climate variables describing axes 1 and 2 of the CCA of taxa with respect to the three climate variables in the Minchin87_2b data set.

	<i>CorE.1</i>	<i>CorE.2</i>	<i>RegrE.1</i>	<i>RegrE.2</i>	<i>TValE.1</i>	<i>TValE.2</i>
JanT	0.91	-0.20	1.38	-0.05	25.23	-1.88
JulyT	-0.04	-0.56	-0.50	-0.91	-4.62	-18.22
Precip	0.01	-0.13	0.04	0.76	0.39	15.05

As expected the correlations of the climate variables with the constrained axes (Table 1) tell the same story as the biplot arrows in Fig. 2g: the first axis is highly correlated to the January temperature and the second axis is strongly negatively correlated to July temperature and slightly negative to precipitation. The regression coefficients and t-values of first axis again show the importance of January temperature, but the July temperature is also of some importance as its t-value is larger than 2 in absolute value. The regression coefficients and t-values of second axis show negligible values for January temperature, large negative values for July temperature and, surprisingly, large positive values for precipitation. The (weighted) correlation between July temperature and precipitation is 0.88 (whereas the other pairwise correlations are about 0.30). So some caution is prudent, but the t-values are large in absolute value (>15) so that the change in sign of the precipitation effect is not simply an artefact due to the bouncing beta problem (Montgomery and Peck 1982; Jongman et al. 1995).

An ecological interpretation may thus be warranted (if this were real data). The first axis reflects the importance of winter temperature, which is a well-known determinant for the range of several trees, particularly beech. The second axis is a weighted difference between precipitation and July temperature. As evaporation increases with summer temperature, the second axis could thus be interpreted as net precipitation (precipitation minus evaporation). Phrased in terms of a limiting factor, the second axis is negatively related to the precipitation deficit (evaporation minus precipitation). The consequence for paleo-climate reconstruction is that July temperature or precipitation can be less well reconstructed than the ecological index of precipitation deficit.

The lesson learned for the interpretation of constrained axes is that in analyses using multiple, mutually correlated explanatory variables, it is worthwhile to also check a regression biplot (displaying regression coefficients) in addition to a standard ordination diagram with biplot scores of explanatory variables (representing their correlations with ordination axes).

Main and interaction effects and principle response curves (PRC)

In this section we illustrate how to investigate main and interaction effects in constrained ordination using RDA. With interactions, the resulting biplots easily get rather cluttered and it appears prudent to focus on one main effect with one interaction term at a time. This can be done using PRC which represents the results of a partial RDA using a specialized diagram that shows how the effect of one factor depends on the levels of another factor. We use the Sod data set to illustrate these approaches.

Baar and ter Braak (1996) used an RDA with the main effects only to show the effects of age and treatment on the fungi. The interaction between age and treatment was significant ($P < 0.05$), but was not included in the constraining regression model because it contributed a low fraction of the variance, about 4% of the total variation. But how did this interaction look like?

Fig. 4 shows the ordination diagram based on RDA with model age + treatment + age.treatment. The main effect of age (quantitative) is represented by an arrow. We could have represented the raw interaction with treatment by three treatment-dependent arrows for age, but it is more interesting to look at the reverse, the age-dependent treatment effects. These are displayed in Fig. 4 by discretizing age into four classes and plotting the 4×3 centroids for all combinations of age class and

treatment (of which one is missing as sod addition was not applied in the highest age class). The centroids (larger symbols) are averages of the corresponding plot locations (CaseR scores using smaller symbols). The three treatments in the same age class are connected by lines in the order sod-addition (A), control (C) and sod-cutting (S). In Fig. 4 the line connecting C and S is shorter for age class 1 than the lines for the higher age classes and has a slightly different orientation, both being an indication for an age-dependent effect of sod-cutting. The effect of sod-cutting appears larger in older stands than in young ones. The effect of sod-addition is small, as judged from the small distances between the C- and A-centroids. In this example, the same conclusions would have been reached by displaying the age-treatment centroids in an RDA without interaction. The overall interpretation is that many taxa benefit from sod-cutting as indicated by position of the supplementary response variables *Nspec* (number of species) and *Ncarp* (number of sporocarps) in Fig. 4, and also that there is a (cross-sectional) successional sequence.

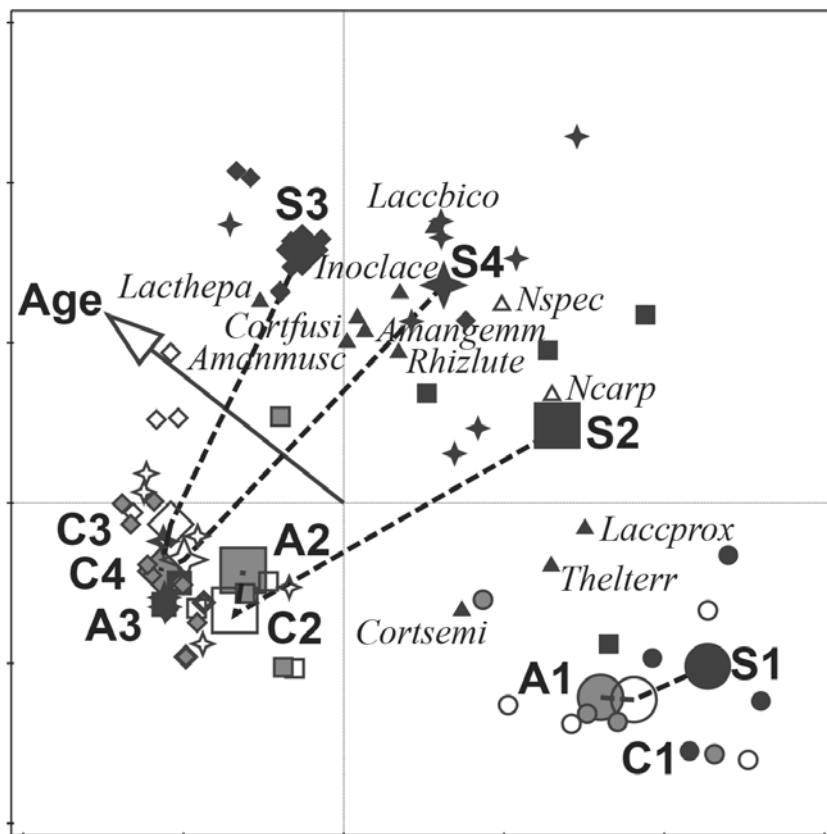


Fig. 4 Sod data set: RDA of fungi with respect to age, treatment and their interaction showing 64% of the total variation in fungal log-abundance. The treatment (C, S and A) by age class (1-4) combinations (big symbols) are located at centroids of the plots (small symbols, CaseR scores) of these combinations (age classes 1 to 4 have symbols circle, square, diamond and star, respectively, and the C, S and A treatments are open black, closed black and closed grey, respectively). Combinations of the same age class are connected by the thick dashed lines. The effect of age is shown as a biplot arrow. Taxa are shown by closed uptriangles, supplementary “taxa” (*Nspec*, *Ncarp*) by open triangles.

Fig. 4 also suggests that the effect of age is non-linear, even non-monotone, as the order of age classes along the direction of the arrow for age has classes 3 and 4 in reverse order, and also that the interaction effect is non-linear, with a big change in effect size between age classes 1 and 2 but little changes between age classes 2-4. In a similar analysis using CCA, classes 3 and 4 nearly coincide. We could perhaps improve the ordination by allowing for non-linear effects by treating age as a factor or as a fuzzy factor (Greenacre 2012) or by replacing log-age by another set of basis functions such as P-splines (Eilers and Marx 1996).

The first three eigenvalues of the RDA (expressed as fractions of the total variation) are 0.26, 0.20 and 0.02, showing that two dimensions suffice to show the modelled effects. The pseudo-canonical correlation (the correlation between the CaseR and CaseE scores) is 0.83 for both these axes. If Fig. 4 would have had shown the CaseE scores, it would explain 46% of the total variance. But it shows the CaseR scores which increases the explained variance by a factor $1/0.83^2 = 1.4$, giving 64%, and after adjustment for degrees of freedom 58%. The general formula for this factor is $(r_1^2 e_1 + r_2^2 e_2)/(e_1 + e_2)$ with r_s and e_s the correlation and eigenvalue for axis s (Graffelman and Tuft 2004).

Ordination diagrams with interactions like Fig. 4 have the disadvantage that they become cluttered and difficult to read, and Fig. 4 is not even an extreme example of this (see van den Brink and ter Braak 1999). It would be advantageous to directly focus on the interactions. A partial ordination that show the pure interaction (with both main effects removed) is fine for testing its significance, but its biplot is difficult to interpret. Even for a single response variable, pure interaction effects are difficult to interpret, whereas the meaning of “age-dependent treatment effect” is clear. In statistical terms, this phrase takes one main effect together with the interaction with one other factor, in this example “treatment + age.treatment | age” where the “|” means the condition “while accounting for”. The Principal Response Curves method (van den Brink and ter Braak 1999) was developed for showing time-dependent treatment effects in longitudinal studies, but time can be replaced by any factor of an experimental design and surely by the factor Age in the Sod data set.

The PRC diagram (Fig. 5) expresses the treatment effects as differences with the control treatment, which is taken as the reference treatment and represented by the horizontal line at 0. From Fig. 5, we see immediately that the effect of sod-cutting is smaller in age class 1 than in the later age classes and the effect of sod addition is very small and varies in sign across the three available age-classes. The main taxa contributing to this type of response pattern are displayed on the vertical scale at the right. All taxa shown have positive weight, which, in combination with the positive values of sod-cutting in the left-hand subplot, shows that these taxa have higher abundance due to sod-cutting. A more quantitative interpretation is also possible (van den Brink and ter Braak 1999; ter Braak and Šmilauer 2012). We can add that Canoco uses a scaling in which the mean-squared taxon score is 1 so that the magnitude of a treatment score (at the left) represents the mean-squared effect per age class. As all is on the log-scale, an effect of 1 as in Fig. 5 is a large effect representing a multiplicative effect of $\exp(1) = 2.7$ on the original abundance scale. Fig. 5 displays 78% of all age-dependent variation among treatments.

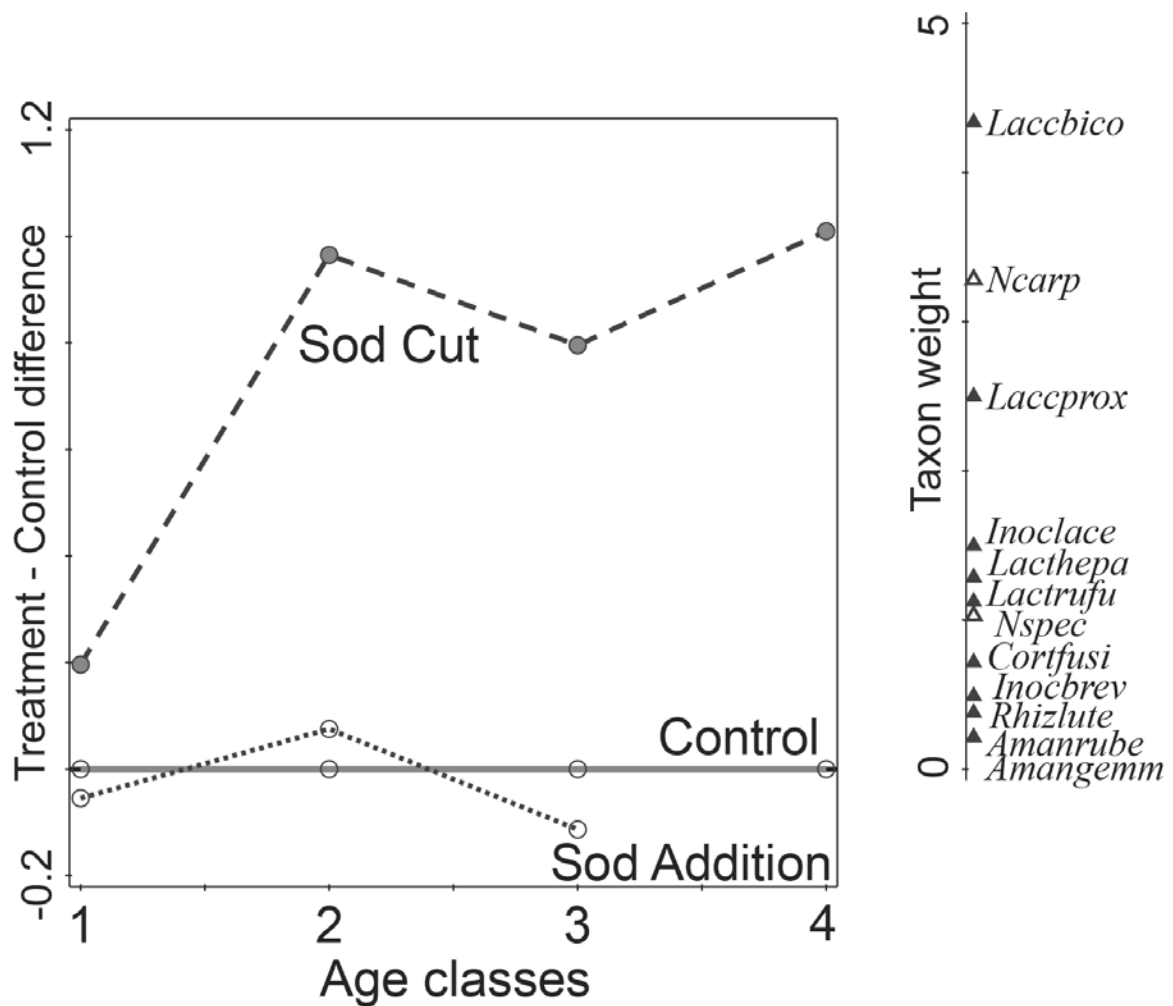


Fig. 5 Sod data set: PRC of the age-dependent treatment effects showing 78% of all age-dependent variation in fungal log-abundance among treatments. For explanation see text.

Discussion

We discussed various topics in unconstrained and unconstrained ordination. We see that the field has matured since its inception in 1954. With software such as Canoco 5 and vegan (Borcard et al. 2011), among others, methods are easily accessible to plant and vegetation ecologists. The Canoco Adviser, an expert system based assistant in Canoco 5, can help preventing common errors.

Except NMDS, all methods we discussed are eigenvalue methods (methods that can be obtained via stable numerical methods with a, nearly always, unique solution), but our working model is in terms of underlying ecological gradients, that is, in terms of latent variables (Walker and Jackson 2011). Latent variable models have their own 'best' solution method, in particular, using (marginal) maximum likelihood. We presented DCA and CCA as approximate methods to the unconstrained and constrained Gaussian model, respectively. We presented PCA whereas we could have advocated probabilistic PCA (Walker and Jackson 2011). Why use approximations, when there exists also methods and computer packages like VGAM (Yee 2004; Yee 2006) that fit the latent variable model exactly? We advocated the use of approximate

methods for ordination because they are quick and robust with noisy and big data sets in providing an overview and allow computer intensive methods such as permutation and bootstrap for statistical inference. The approximation is, in our view, good enough for what ecologists typically do with the results, and it works nearly always in practice, unlike the methods being approximated. Overcoming their technical issues seems a difficult task (given how many times people tried). To move beyond the horizons, to the next-generation ordination methods, we should surely skip maximum likelihood Gaussian ordination and improve the GAM-like approach of Yee (2006), however difficult it seems to get it to produce stable results for the type of data sets that plant ecologists want to analyse. An independent attempt for unconstrained ordination using P-splines (Eilers and Marx 1996) is by Schnabel et al. (2012).

How could David Goodall have envisioned the rise of ordination methods over the past 60 years! And in another 100 years, the field of ordination might have been completely reshaped.

Acknowledgements

This paper benefitted from comments by David Warton and Pierre Legendre. The contribution of PS was funded by the Center of Excellence PLADIAS, 14-36079G, Czech Science Foundation.

Electronic supplementary material is available at the Springer website. It may also be available at www.canoco5.com.

Supplementary material 1 (ZIP 732 kb)

Canoco 5 projects detailing

how analyses are carried out; these details can be inspected by any reader by asking for an one-month trial copy of Canoco 5 at

trial@microcomputerpower.com.

Supplementary material 2 (ZIP 8 kb)

Data tables of the examples

Tikus Island, Minchin87_2b and Sod in comma separated format (two tables per example, one for the community data and one for the explanatory data).

Supplementary material 3 (PDF 410 kb)

Readme file for the zip files with screenshots of the example Canoco 5 projects.

References

- Baar J, ter Braak CJF (1996) Ectomycorrhizal sporocarp occurrence as affected by manipulation of litter and humus layers in Scots pine stands of different age. *Applied Soil Ecology* 4:61-73.
- Birks HJB, Line JM, Juggins S, Stevenson AC, ter Braak CJF (1990) Diatoms and pH reconstruction. *Philosophical Transactions of the Royal Society London, Series B* 327:263-278.

- Borcard D, Gillet F, Legendre P (2011) Numerical ecology with R. Use R! Springer, New York.
- Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88:9-25.
- Davies PT, Tso MK-S (1982) Procedures for reduced-rank regression. *Applied Statistics* 31:244-255.
- Eilers PHC, Marx BD (1996) Flexible smoothing with B-splines and penalties. *Statistical Science* 11:89-121.
- Engel B, Keen A (1994) A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica* 48:1-22.
- Faith DP, Minchin PR, Belbin L (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69:57-68.
- Frontier S (1976) Étude de la décroissance des valeurs propres dans une analyse en composantes principales: comparaison avec le modèle du bâton brisé. *Journal of Experimental Marine Biology and Ecology* 25:67-75.
- Goodall DW (1954) Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Australian Journal of Botany* 1:39-63.
- Goodall DW, Johnson RW (1982) Non-linear ordination in several dimensions. A maximum likelihood approach. *Vegetatio* 48:197-208.
- Goodall DW, Johnson RW (1987) Maximum-likelihood ordination - some improvements and further tests. *Vegetatio* 73:3-12.
- Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53:325-338.
- Graffelman J (2001) Quality statistics in canonical correspondence analysis. *Environmetrics* 12 (5):485-497.
- Graffelman J, Tuft R (2004) Site scores and conditional biplots in canonical correspondence analysis. *Environmetrics* 15 (1):67-80. doi:10.1002/env.629
- Greenacre M (2012) Fuzzy coding in constrained ordinations (*Ecology*). *Ecology* 94 (2):280-286. doi:10.1890/12-0981.1
- Greenacre MJ (1984) Theory and applications of correspondence analysis. Academic Press, London.
- Greenacre MJ (2013) The contributions of rare objects in correspondence analysis. *Ecology* 94 (1):241-249. doi:10.1890/11-1730.1
- Guisan A, Weiss S, Weiss A (1999) GLM versus CCA spatial modeling of plant species distribution (*Plant Ecology*). 143 (1):107-122. doi:10.1023/a:1009841519580
- Heiser WJ (1987) Joint ordination of species and sites: the unfolding technique. In: Legendre P, Legendre L (eds) *Developments in numerical ecology*. Springer-Verlag, Berlin, pp 189-224
- Hill MO (1973) Reciprocal averaging: an eigenvector method of ordination. *Journal of Ecology* 61:237-249.
- Hill MO (1979) TWINSpan - a FORTRAN program for detrended correspondence analysis and reciprocal averaging. Cornell University, Ithaca.
- Hill MO, Gauch HG (1980) Detrended correspondence analysis, an improved ordination technique. *Vegetatio* 42:47-58.
- Ihm P, van Groenewoud H (1984) Correspondence analysis and Gaussian ordination. *Compstat Lectures* 3:5-60.
- Jackson DA, Somers KM (1991) The ups and downs of detrended correspondence analysis. *The American Naturalist* 137:704-712.

- Jamil T, ter Braak CJF (2013) Generalized linear mixed models can detect unimodal species-environment relationships (PeerJ). PeerJ 1:e95. doi:10.7717/peerj.95
- Johnson KW, Altman NS (1999) Canonical correspondence analysis as an approximation to Gaussian ordination. *Environmetrics* 10 (1):39-52. doi:10.1002/(sici)1099-095x(199901/02)10:1<39::aid-env334>3.0.co;2-3
- Johnson RW, Goodall DW (1979) A maximum likelihood approach to non-linear ordination. *Vegetatio* 41:133-142.
- Jongman RHG, ter Braak CJF, van Tongeren OFR (1995) *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge. doi:10.2277/0521475740
- Juggins S, Birks HJB (2012) Quantitative environmental reconstructions from biological data. In: Birks HJB, Lotter AF, Juggins S, Smol JP (eds) *Tracking environmental change using lake sediments*, vol 5. Springer, New York, pp 431-494.
- Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1-27.
- Legendre L, Legendre P (2012) *Numerical ecology*. Elsevier, Amsterdam
- Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multi-species responses in multi-factorial ecological experiments. *Ecological Monographs* 69:1-24.
- Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271-280.
- Legendre P, Oksanen J, ter Braak CJF (2011) Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution* 2 (3):269–277. doi:10.1111/j.2041-210X.2010.00078.x
- Martins TG, Simpson D, Lindgren F, Rue H (2013) Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis* 67:68-83
- McCullagh P, Nelder JA (1989) *Generalized linear models* (second edition). Chapman and Hall, London.
- McCune B (1997) Influence of noisy environmental data on canonical correspondence analysis. *Ecology* 78:2617-2623.
- Minchin PR (1987) An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 67:1167-1179.
- Montgomery DC, Peck EA (1982) *Introduction to linear regression analysis*. Wiley, New York.
- O'Hara RB, Kotze DJ (2010) Do not log-transform count data. *Methods in Ecology and Evolution* 1 (2):118-122. doi:10.1111/j.2041-210X.2010.00021.x
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2013) *vegan: Community Ecology Package*. R package version 2.0-9 <http://cranr-project.org/>
- Oksanen J, Tonteri T (1995) Rate of Compositional Turnover along Gradients and Total Gradient Length. *Journal of Vegetation Science* 6 (6):815-824. doi:10.2307/3236395
- Palmer JA (2014) A glossary of ordination-related terms. <http://ordination.okstate.edu/glossary.htm> (last accessed 30 January 2014).
- Palmer MW (1993) Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74:2215-2230.
- Peet RK, Knox RG, Case JS, Allen RB (1988) Putting Things in Order: The Advantages of Detrended Correspondence Analysis. *The American Naturalist* 131 (6):924-934. doi:10.2307/2461824

- Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology* 87 (10):2614-2625.
- Rao CR (1964) The use and interpretation of principal component analysis in applied research. *Sankhya A* 26:329-358.
- Schnabel SK, van Eeuwijk FA, Eilers PHC Modeling latent curves for genotype by environment interaction. In: Komárek A, Nagy S (eds) *Proceedings of the 27th International Workshop on Statistical Modelling, Prague, 2012*.
- Šmilauer P, Lepš J (2014) *Multivariate Analysis of Ecological Data using Canoco 5*, 2nd edition. Cambridge University Press, Cambridge.
- ter Braak CJF (1985) Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics* 41:859-873.
- ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167-1179.
- ter Braak CJF (1987a) The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69:69-77.
- ter Braak CJF (1987b) Unimodal models to relate species to environment. Agricultural mathematics Group, Wageningen.
- ter Braak CJF (1994) Canonical community ordination. Part I: Basic theory and linear methods. *Écoscience* 1:127-140.
- ter Braak CJF (2014) History of canonical correspondence analysis. In: Blasius J, Greenacre M (eds) *Visualization and verbalization of Data*. Chapman and Hall/CRC, London, pp 61-75.
- ter Braak CJF, Barendregt LG (1986) Weighted averaging of species indicator values: its efficiency in environmental calibration. *Mathematical Biosciences* 78:57-72.
- ter Braak CJF, Looman CWN (1986) Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65:3-11.
- ter Braak CJF, Šmilauer P (2012) *Canoco reference manual and user's guide: software for ordination, version 5.0*. Microcomputer Power, Ithaca, USA.
- ter Braak CJF, van Dam H (1989) Inferring pH from diatoms: a comparison of old and new calibration methods. *Hydrobiologia* 178:209-223.
- van den Brink P, ter Braak C (1998) Multivariate analysis of stress in experimental ecosystems by Principal Response Curves and similarity analysis. *Aquatic Ecology* 32 (2):163-178. doi:10.1023/a:1009944004756
- van den Brink PJ, ter Braak CJF (1999) Principal Response Curves: Analysis of time-dependent multivariate responses of a biological community to stress. *Environmental Toxicology and Chemistry* 18: 138-148.
- van den Wollenberg AL (1977) Redundancy analysis. An alternative for canonical correlation analysis. *Psychometrika* 42:207-219.
- Walker SC, Jackson DA (2011) Random-effects ordination: describing and predicting multivariate correlations and co-occurrences (*Ecological Monographs*). *Ecological Monographs* 81 (4):635-663. doi:10.1890/11-0886.1
- Wang Y, Naumann U, Wright ST, Warton DI (2012) mvabund– an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution* 3 (3):471-474. doi:10.1111/j.2041-210X.2012.00190.x
- Warton DI (2005) Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics* 16 (3):275-289. doi:10.1002/env.702

- Warton DI, Wright ST, Wang Y (2012) Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution* 3 (1):89-101. doi:10.1111/j.2041-210X.2011.00127.x
- Warwick RM, Clarke KR, Suharsono (1990) A statistical analysis of coral community responses to the 1982/3 El Niño in the Thousand Islands, Indonesia, 171-179. *Coral Reefs* 8:171-179.
- Yee TW (2004) A new technique for maximum-likelihood canonical Gaussian ordination. *Ecological Monographs* 74 (4):685-701.
- Yee TW (2006) Constrained additive ordination (Ecology). *Ecology* 87 (1):203-213. doi:10.1890/05-0283.
- Zhang Y, Thas O (2012) Constrained ordination analysis in the presence of zero inflation. *Statistical Modelling* 12 (6):463-485. doi:10.1177/1471082x12460129
- Zhu M, Hastie TJ, Walther G (2005) Constrained ordination analysis with flexible response functions. *Ecological Modelling* 187 (4):524-536.
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2008) *Mixed effects models and extensions in ecology with R* Springer, New York.