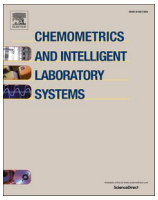
Chemometrics and Intelligent Laboratory Systems 240 (2023) 104898

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



Redundancy analysis includes analysis of variance-simultaneous component analysis (ASCA) and outperforms its extensions

Cajo J.F. ter Braak

Biometris, Wageningen University & Research, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Redundancy analysis
Reduced-rank regression
Analysis of variance-simultaneous component analysis
Weighted effect coding
Factorial experiment
Unbalanced design

ABSTRACT

Chemometrics and statistical ecology share interest in the analysis of multivariate data. In ecology, unconstrained and constrained ordination are popular methods to analyze and visualize multivariate data, with principal component analysis (PCA) and redundancy analysis (RDA) as prototype methods. Constraints give more insight and power by focusing on the response of the variables to particular external predictors or experimental factors, after optional adjustment for covariates. In chemometrics, analysis of variance - simultaneous component analysis (ASCA) was proposed decades later, with particular emphasis on the multivariate main and interaction effects in factorial experiments. This paper shows the similarities and differences between ASCA, its extensions, and (partial) RDA, alias reduced-rank regression. ASCA and RDA (understood as a sequence of partial RDAs, just as ASCA uses a sequence of PCAs) are shown to be mathematically identical for equireplicated designed experiments. Differences appear with unequal replication. As a corollary we show that, with equal replication, a particularly attractive form of ASCA, which displays a main effect together with an interaction, is a special case of principal response curve analysis. RDA is a least-squares method and uses the optimal weights in the dimension reduction of the treatment effects, whereas ASCA extensions for unbalanced data use alternative, suboptimal weights.

1. Introduction

Chemometrics brought partial least-squares (PLS) to statistics [1,2], it also brought analysis of variance (ANOVA)-simultaneous component analysis (ASCA) [3,4]. Both methods extract components; PLS for prediction, ASCA for the analysis and visualization of effects of treatments on multivariate response in designed experiments. In terms of linear models, the treatments and other factors, such as (discrete) time, of such experiments can be viewed as predictors and the treatment effects as regression coefficients. Whereas PLS regression was designed for prediction with correlated predictors, ASCA started from ANOVA in complete (=equireplicated) designed experiments in which predictors are uncorrelated or, in mathematical terms, orthogonal [5] and was later extended under the name ASCA + to unbalanced experiments [6].

In statistical ecology, related methods were, and continue to be, popular, in particular constrained and unconstrained ordination [7,8]. These methods extract components from multivariate response for the analysis and visualization of both observational and experimental data [9,10]. The prototype method of unconstrained ordination is principal component analysis (PCA) with constrained form redundancy analysis (RDA), known in chemometrics as two-block 'mode C' PLS [11]. For a recent review see Ref. [12]. RDA and multivariate PLS regression are compared in Ref. [13]. RDA allows for variance decomposition [14] and for testing by permutation of specific null hypotheses [15–17]. Canoco [18] and the R-package vegan [19] contain user-friendly and versatile software implementations of these methods. In statistics, RDA is known as reduced-rank regression [20].

Whereas the original version of ASCA was based on traditional fixed-effects ANOVA followed by PCA of particular terms and later extended to multivariate linear regression in ASCA+ [6] and weighted effect ASCA (WE-ASCA) [21] for application in unbalanced designs, RDA was based from the start on PCA constrained by a linear model for analysis of both observational data and (incomplete) designed experiments [7]. In this paper we compare ASCA, its extensions and RDA and show their similarities and differences. We show mathematically that these methods are identical for complete data, but that differences appear with unequal replication. As a corollary we show that, for complete data, a particularly attractive form of ASCA, which displays a main effect together with an interaction, is a special case of principal response curve analysis (PRC) [22–24]. The fact that ASCA+ and WE-ASCA, which

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

https://doi.org/10.1016/j.chemolab.2023.104898
Received 15 February 2023; Received in revised form 15 May 2023; Accepted 11 June 2023
Available online 20 June 2023

differ in factor coding, are different versions of ASCA, shows that ASCA is coding dependent, whereas RDA is not. We show that RDA is a least-squares method whereas ASCA+ or WE-ASCA are not.

2. Theory and methods

2.1. Design, model and design matrices

The similarity and differences between ASCA+ and RDA can be shown in models that decompose the variance into two sources. The easiest such model ignores interaction and decomposes the variance of a factorial design with two factors A and B in the variances due to the two main effects. The model formula of this decomposition is A + B. With interaction, and of practical importance, is the decomposition based on the model formula A + (B + A:B). This is the focal model of this paper, giving two sources of variance.

a) the variance due to the main effect of factor A, and
b) the variance due to the main effect of B and its interaction with A.

Source b) thus includes the possible dependence of the effect of B on the levels of A. A common example is a repeated measurement design (with common time points) in which the factor A represents Time and factor B represents a Treatment, and our interest focusses on how the treatment differences vary across time or, said more formally, how the effect of Treatment depends on Time. This model is used in PRC, which is a method of analysis based on RDA [22,23] and in the applications in papers on ASCA [3,4,25,26]. Another way to specify the model is A + A:B. This is the usual specification for nested designs, where the main effect of B has little or no meaning.

For multivariate response, these ANOVA models can be phrased as multivariate multiple regression models of the form

$$\mathbf{Y} = \mathbf{1}_n\boldsymbol{\mu}' + \mathbf{ZA} + \mathbf{XB} + \mathbf{E}, \quad (1)$$

where $\mathbf{Y}$ is a $n \times m$ matrix with measurements of m response variables in n samples, $\mathbf{Z}$ is a design matrix coding for factor A, $\mathbf{X}$ a design matrix coding for the factor B (and its interaction with A, if included), $\mathbf{1}_n$ is a n column vector of ones, $\boldsymbol{\mu}$ is a m column vector of unknown intercepts, and $\mathbf{A}$ and $\mathbf{B}$ are matrices of unknown regression coefficients with column k in each applying to response variable k, and $\mathbf{E}$ is an $n \times m$ error matrix. The unknowns need to be estimated for which we use least-squares. Implicitly we thus assume, as in many ASCA papers, that the errors are independent and identically distributed. This assumption can be relaxed [20,27–30].

The simplest coding of factors is perhaps indicator coding, in which each level of the factor is represented by a dummy (1/0) variable (column in the design matrix of the regression model). For factor A, for example, and focusing on rows, each row of $\mathbf{Z}$ then contain a single 1 in the column of the level of A of the corresponding sample and value 0 in the columns for the other levels. With intercept, one column (i.e. one dummy variable) is redundant (collinear) and either the first or the last column is often deleted, leading to the default treatment coding and to SAS coding, respectively, in R [31]. The deleted column determines the reference of this type of coding. ASCA+ [6] uses 'sum coding', also called 'effect coding' [32]. In sum coding, all 0-entries in the indicator-coded $\mathbf{Z}$ corresponding to cases with the last level of A are changed to −1 and the last column is deleted. With sum coding in a complete design, the intercepts estimate the overall means of the response variables. For this to hold true in an incomplete design, we need weighted effect (WE) coding [32,33] which is slightly more complicated as it depends on the frequency of individual levels of the factor in the design.

Coding for interaction is obtained by multiplying each column for factor A in $\mathbf{Z}$ with each similarly constructed column for factor B. Without deletion of redundant columns, encoding for factors A and for B using either treatment or sum coding, the "raw" interaction so consists of $n_A \times n_B$ columns, with n_A and n_B the number of levels of factor A and B, respectively. Of these columns, the n_A columns involving the last level of B are redundant if the focal model is specified as A + A:B and are not included in matrix $\mathbf{X}$. If the model is specified as A + (B + A:B), the first $(n_B - 1)$ columns of $\mathbf{X}$ encode for B and $(n_A - 1) \times (n_B - 1)$ columns encode for the interaction A:B, with the interaction columns obtained by deleting all columns from the raw interaction that involve the last level of either A or B. Either way, $\mathbf{X}$ contains $n_A \times (n_B - 1)$ columns, which corresponds to the number of A-dependent effects of treatment B. In WE-coding interactions are not simple multiplications of the main effect matrices; they are coded in a such way that interactions are orthogonal to (=uncorrelated with) the main effects [32]. Examples of these ways of coding are given in Supplement S3. See also chapter 6 in Ref. [34], where treatment and sum coding are treated in the sections "set-to-zero restrictions" and "sum-to-zero restrictions", respectively. The effect of unbalance on the data analysis is discussed in Refs. [34,35].

Equation (1) is fitted to data by least-squares giving the estimates $\widehat{\boldsymbol{\mu}}$, $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ of the unknowns μ, $\mathbf{A}$ and $\mathbf{B}$, respectively and residuals $\widehat{\mathbf{E}}$. These estimates are used in both ASCA+ and RDA. While the estimates differ between different ways of encoding the design matrices, the fitted values $\widehat{\mathbf{Y}}$ do not.

2.2. ASCA+

ASCA and ASCA+ are essentially two step algorithms. ASCA uses traditional ANOVA computations for complete designs, essentially averaging per level or combination of levels [3], and ASCA+ uses multiple regression with sum coding of the model so as to extend ASCA to incomplete designs [6]. We focus on ASCA+. The first step is to fit equation (1) by least-squares. In a second step, a (column-centered) principal component analysis (PCA) is applied to $\mathbf{Z}\widehat{\mathbf{A}}$ and a separate PCA to $\mathbf{X}\widehat{\mathbf{B}}$, yielding two sets of row scores ($\mathbf{T}^a_\mathbf{Z}$ and $\mathbf{T}^a_\mathbf{X}$, say, with the superscript a for ASCA to distinguish them from similar scores in RDA) and of loadings ($\mathbf{P}^a_\mathbf{Z}$ and $\mathbf{P}^a_\mathbf{X}$) for the response variables on a number of PCA components (axes or dimensions). These PCAs lead to dimension reduction when only the first few components are used. The scores and loadings on the first few (usually 2) axes can be plotted as in a normal PCA to yield a biplot [36] of the form $\mathbf{TP}'$ of the fitted values ($\widehat{\mathbf{Y}}_\mathbf{A} = \mathbf{Z}\widehat{\mathbf{A}}$ for the first PCA and $\widehat{\mathbf{Y}}_\mathbf{B} = \mathbf{X}\widehat{\mathbf{B}}$ for the second PCA). In PCA, the scores can be derived from the loadings, so that, applied to ASCA+, $\mathbf{T}^a_\mathbf{Z} = \mathbf{Z}\widehat{\mathbf{A}}\mathbf{P}^a_\mathbf{Z}$ and $\mathbf{T}^a_\mathbf{X} = \mathbf{X}\widehat{\mathbf{B}}\mathbf{P}^a_\mathbf{X}$ with an additional centering by columns of the resulting scores (as the PCA was, tacitly in Thiel et al. [6], centered by columns). The scores for A and B are constant within their levels. Additional scores, $\mathbf{T}^{Ea}_\mathbf{Z} = \mathbf{T}^a_\mathbf{Z} + \widehat{\mathbf{E}}\mathbf{P}^a_\mathbf{Z}$ and $\mathbf{T}^{Ea}_\mathbf{X} = \mathbf{T}^a_\mathbf{X} + \widehat{\mathbf{E}}\mathbf{P}^a_\mathbf{X}$, show variability among replicates within the levels of A and B [37]. The equation for such scores in Ref. [6] is inaccurate as it lacks the column-centering of $\mathbf{T}^a_\mathbf{Z}$ and $\mathbf{T}^a_\mathbf{X}$.

2.3. RDA

The name RDA started with Ref. [38], but the method RDA dates back at least till Ref. [39] (Supplement S1). RDA is best known for its use with quantitative predictors and covariates in observational studies [8,40]. An early application to an experimental design, with an RDA-based analysis of variance, appeared in Ref. [10], re-analyzed in Ref. [41] and another is PRC [22,23] (see section 2.6 and Supplement 1.3 for further details).

It is of interest to note that RDA has been developed further in the framework of generalized linear (mixed) models [42–45], but for the comparison with ASCA it is sufficient to consider fixed effects models fitted by least-squares.

There are two ways of defining RDA [7,23](see also Appendix A1.3): 1) RDA is a PCA constrained by a linear model; 2) RDA is a multivariate

regression with a rank-restriction on the matrix of regression coefficients. The second way is more general as it allows for explicit specification of the variances and covariances of the errors [28,46]. For easy comparison with ASCA, we use the version in which the errors are independent and of equal variance and the version of RDA with covariates, also called covariables [7,8] or concomitant variables [20,47].

Let $\mathbf{Z}$ in equation (1) contain the covariates, in our focal model (section 2.1) coding for the factor A, and $\mathbf{X}$ the design matrix that codes for A:B or B + A:B, giving the same fitted values $\widehat{\mathbf{Y}}$ in either way of model specification. While ASCA requires sum coding, any coding system can be used in RDA as is verified in the numerical comparisons for sum and treatment coding (section 4.1). Equation (1) is then fitted to the data $\mathbf{Y}$ by least-squares with the restriction that the matrix $\mathbf{B}$ has reduced rank r (with r smaller than the rank of $\mathbf{X}$), whereas no restrictions are applied to $\mathbf{A}$. Covariates thus differ from the other predictors (here $\mathbf{X}$) in that they are fitted without dimension reduction. If $\mathbf{B}$ is of reduced rank r, it is of the form $\mathbf{B} = \mathbf{C}^{[r]}\,\mathbf{P}^{[r]'}$, where the superscript $[r]$ indicates that only the first r columns are used. Recall that PCA with two axes retained or plotted in a biplot [36] provides a rank 2 approximation of the matrix to which it was applied. Similarly, $\mathbf{X}\widehat{\mathbf{B}}$ is approximated in ASCA+ with r axes by $\mathbf{T}_{\mathbf{X}}^{\mathrm{a}[r]}\mathbf{P}_{\mathbf{X}}^{\mathrm{a}[r]'}$.

For the comparison of ASCA+ and RDA it is useful to show how the least-squares solution is obtained in RDA. For a full derivation see Appendix A1.1. For ease of exposition, the intercept ($\mathbf{1}_n$) is added to $\mathbf{Z}$ as an extra column (as is standard in multiple regression using matrix notation), so that also $\mathbf{A}$ has an extra column, but that does not matter for now as the focus is on the reduced-rank estimation of $\mathbf{B}$. How RDA is obtained can be understood by first projecting $\mathbf{X}$ on to $\mathbf{Z}$, subtracting and adding the fit (the projected $\mathbf{X}$) to equation (1) and then rearranging the equation [47]

$$\mathbf{Y} = \mathbf{Z}\mathbf{A} + \Pi_{\mathbf{Z}}\mathbf{X}\mathbf{B} + \mathbf{X}\mathbf{B} - \Pi_{\mathbf{Z}}\mathbf{X}\mathbf{B} + \mathbf{E} = \mathbf{Z}\mathbf{A}^{*} + \widetilde{\mathbf{X}}\mathbf{B} + \mathbf{E} \tag{2}$$

with $\Pi_{\mathbf{Z}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, the projection operator on to $\mathbf{Z}$, $\mathbf{A}^{*} = \mathbf{A} + (\mathbf{Z}'\mathbf{Z})^{-}\mathbf{Z}'\mathbf{X}\mathbf{B}$, and $\widetilde{\mathbf{X}} = (\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{X}$. Note that $\widetilde{\mathbf{X}}$ is orthogonal to $\mathbf{Z}$ so that least-squares reduced-rank estimate of $\mathbf{B}$ can be obtained from the formal model $\mathbf{Y} \sim \widetilde{\mathbf{X}}$, a reduced rank model without covariates $\mathbf{Z}$. Similarly, multiply the left- and right-hand side of equation (1) by $(\mathbf{I}_n - \Pi_{\mathbf{Z}})$ and obtain

$$\widetilde{\mathbf{Y}} = (\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{Y} = (\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{X}\mathbf{B} + \mathbf{E} = \widetilde{\mathbf{X}}\mathbf{B} + \mathbf{E}. \tag{3}$$

For the same reason, the full-rank least-squares estimate of $\mathbf{B}$ in equation (3) is equal to $\widehat{\mathbf{B}}$, the least-squares estimate of $\mathbf{B}$ in equation (1). Note that $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{X}}$ are the residuals of the multivariate regression of $\mathbf{Y}$ and $\mathbf{X}$ onto $\mathbf{Z}$, respectively.

In the simplest case of RDA, $\mathbf{Z}$ consists of the intercept only and $\widetilde{\mathbf{X}}$ is simply $\mathbf{X}$ after column-centering, a step that is included as a preprocessing step in Ref. [47]. For this reason, we do not distinguish between "simple" RDA and partial RDA (or between reduced-rank regression and reduced-rank regression with concomitant variables), in line with the fact that all software for RDA does both by a single function, particularly the rda function in the R package vegan [19] and the f_rda.m function in the Phantom toolbox for Matlab [48].

The least-squares reduced-rank fit of $\mathbf{B}$ in equation (3) is obtained in two steps [20,47] 1) obtain the (full rank) least-squares-fit $\widehat{\mathbf{Y}}_{\mathbf{B}}$ by multivariate regression of $\widetilde{\mathbf{Y}}$ on to $\widetilde{\mathbf{X}}$, giving fit $\widetilde{\mathbf{X}}\widehat{\mathbf{B}} = \widetilde{\mathbf{X}}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{Y}} = \Pi_{\widetilde{\mathbf{X}}}\widetilde{\mathbf{Y}}$, and 2) perform a PCA on $\widetilde{\mathbf{X}}\widehat{\mathbf{B}}$ giving scores and loading matrices $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$ and $\mathbf{P}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$, respectively, which are related by $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{r}} = \widetilde{\mathbf{X}}\widehat{\mathbf{B}}\mathbf{P}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$ (the superscript r stands for RDA to distinguish them from similar scores in ASCA). The scores $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$ are the standard scores in partial redundancy analysis [7, 20] and are a function of both $\mathbf{X}$ and $\mathbf{Z}$. As $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{r}} = \widetilde{\mathbf{X}}(\widehat{\mathbf{B}}\mathbf{P}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}) = \widetilde{\mathbf{X}}\mathbf{C}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$, the canonical coefficients are $\mathbf{C}_{\widetilde{\mathbf{X}}}^{\mathrm{r}} = \widehat{\mathbf{B}}\mathbf{P}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$ [7]. The least-squares reduced-rank estimate of $\mathbf{B}$ is thus $\mathbf{C}_{\mathbf{X}}^{\mathrm{r}[r]}\mathbf{P}_{\widetilde{\mathbf{X}}}^{\mathrm{r}[r]'}$. Key in the above is projection, so that any inverse can be replaced by a generalized inverse in case of singularity, for example, if indicator coding is used.

The additional scores that show the variability among replicates around the scores $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$ are defined as, $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{Er}} = (\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{Y}\mathbf{P}_{\mathbf{X}}^{\mathrm{r}}$ [7,8]. Note that $(\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{Y} = (\mathbf{I}_n - \Pi_{\mathbf{Z}})(\mathbf{Z}\widehat{\mathbf{A}^{*}} + \widetilde{\mathbf{X}}\widehat{\mathbf{B}} + \widehat{\mathbf{E}}) = \widetilde{\mathbf{X}}\widehat{\mathbf{B}} + \widehat{\mathbf{E}}$, so that $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{Er}} = \mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{r}} + \widehat{\mathbf{E}}\mathbf{P}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$. The RDA scores that show variability among replicates are thus derived in the same way as in ASCA from their PCA scores and loadings in conjunction with the error matrix (which is the same for ASCA and RDA). The scores $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{r}}$ go under a variety of names: constrained scores, X-space scores, CaseE scores in Canoco 5 [derived from the explanatory variables] and LC-scores [linear combination scores] in R package vegan, whereas names for $\mathbf{T}_{\widetilde{\mathbf{X}}}^{\mathrm{Er}}$ are unconstrained scores, Y-space scores, CaseR scores in Canoco 5 [derived from the response] and WA-scores [weighted averaging scores] in R package vegan. The name WA-scores is inherited from canonical correspondence analysis, which relates to correspondence analysis as redundancy analysis relates to PCA [7,12]. In (canonical) correspondence analysis the response values in $\mathbf{Y}$ are non-negative and the unconstrained scores are weighted averages of the column scores ('loadings') using the response values of the sample (a row of $\mathbf{Y}$) as weights [7].

To obtain a reduced-rank decomposition of $\mathbf{A}$ as well, simply interchange the role of $\mathbf{Z}$ and $\mathbf{X}$ in equation (1), except that it may not be prudent to adjust a main effect for its interaction [49].

It is by now easy to formulate in which sense RDA is a constrained PCA. RDA is a PCA with linear restrictions on the scores $\mathbf{T}$: $\mathbf{T} = \widetilde{\mathbf{X}}\mathbf{C}$, *i.e.* the scores must be a linear combination of $\widetilde{\mathbf{X}}$ or, equivalently, of both $\mathbf{X}$ and $\mathbf{Z}$ but orthogonal to the covariates $\mathbf{Z}$ ($\mathbf{Z}'\mathbf{T} = \mathbf{0}$). This view on RDA is exactly what is known as partial redundancy analysis in ter Braak & Prentice [7].

It is important to remark that the steps to obtain the least-squares reduced-rank fit of $\mathbf{B}$ are just one algorithm for RDA (Supplement S2). An iterative algorithm is provided in Ref. [7], implemented in the software Canoco [18] with sparse matrix operations, in which neither the fitted values $\widehat{\mathbf{Y}}$ nor $\widehat{\mathbf{Y}}_{\mathbf{B}}$ nor PLS-type deflation [13] appear. This algorithm is based on simple steps that alternate between scores, loadings and coefficients and that together define the eigen equations of RDA [7]. See Ref. [50] for examples of such algorithms. The key point is that the algorithm solves the least-squares RDA problem: it minimizes the sum of squared residuals ($\widehat{\mathbf{E}}$) under the rank constraint on $\mathbf{B}$.

2.4. Mathematical comparison

ASCA(+) and RDA perform PCA on matrices $\mathbf{X}\widehat{\mathbf{B}}$ and $(\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{X}\widehat{\mathbf{B}} = \mathbf{X}\widehat{\mathbf{B}} - \Pi_{\mathbf{Z}}\mathbf{X}\widehat{\mathbf{B}} = \widetilde{\mathbf{X}}\widehat{\mathbf{B}}$, respectively, and are thus exactly identical if $\Pi_{\mathbf{Z}}\mathbf{X}$ vanishes, *i.e.* if $\mathbf{X}$ is orthogonal to $\mathbf{Z}$, *i.e.* $\mathbf{Z}'\mathbf{X} = \mathbf{0}$. This happens for complete designs using sum coding, in which case RDA does not even need covariates for this equivalence. ASCA(+) and RDA are also the same if the full rank of $\widehat{\mathbf{B}}$ is 1 so that no further rank- or dimension reduction is possible, e.g. in the model A + B with factor B having two levels only or in any (incomplete versions) of a 2^p-factorial design, nevertheless used as example in Thiel et al. [6]. In the remainder of this paper we focus on (dimension reduction in) incomplete designs. If $\mathbf{X}$ is not orthogonal to $\mathbf{Z}$, the methods are not identical; the difference is $\Pi_{\mathbf{Z}}\mathbf{X}\widehat{\mathbf{B}}$. Appendix A1.2 shows that RDA uses the optimal weights in the reduced-rank approximation of $\widehat{\mathbf{B}}$, whereas ASCA+ uses alternative, sub-optimal weights.

For our focal model, with $\mathbf{Z}$ coding for the factor A and $\mathbf{X}$ coding for the A-dependent effects of B, the difference between ASCA+ and RDA can be expressed in simple words. ASCA+ performs a PCA on the part of the fitted values corresponding to the A-dependent B-effects, whereas

RDA first subtracts from these values the means per level of A and then performs a PCA. ASCA+ thus applies PCA to a matrix (table) that has means that vary across levels of A in incomplete designs, whereas RDA applies PCA to a table that has, per level of A, zero mean for each variable. The table used in ASCA+ does not even have zero mean overall in incomplete designs (but the PCA removes these non-zero means as it is column-centered by default). To link to the formulas, the fitted values of the term (A:B or B + A:B) are $\mathbf{X}\widehat{\mathbf{B}}$ and the means per level of A, when assigned to the corresponding samples of each level, are $\Pi_{\mathbf{Z}}\mathbf{X}\widehat{\mathbf{B}}$. In the formulas, the formation of means per level of A and the subtraction is indicated by $(\mathbf{I}_n - \Pi_{\mathbf{Z}})$.

Consequently, the RDA-scores $\mathbf{T}^{\mathrm{r}}_{\widehat{\mathbf{X}}}$ and $\mathbf{T}^{\mathrm{Er}}_{\widehat{\mathbf{X}}}$ have zero mean per level of A, whereas the ASCA-scores $\mathbf{T}^{\mathrm{a}}_{\mathbf{X}}$ and $\mathbf{T}^{\mathrm{Ea}}_{\mathbf{X}}$ may have non-zero means for levels of A, as we illustrate in Fig. 1.

ASCA+ performs one multivariate regression and several PCAs which are all least-squares methods, but ASCA+ is not least-squares overall. RDA is least-squares for any prescribed reduced-rank for the focal term (**XB** in equations (1) and (2), and B + A:B in PRC) and full rank of the other terms (**ZA** in equations (1) and (2), and A in PRC), but is not least-squares either for prescribed reduced-ranks for all terms. Algorithms to provide the overall least-squares solution for prescribed reduced-ranks of all terms are discussed in Velu [27] and Velu et al. [51] with an algorithm for a more general case in Takane et al. [52].

ASCA+ and RDA provide exactly the same decomposition of the variance, whatever the design. This equality is due to the choice in ASCA+ [6] to base explained variance on type III sum of squares. This choice introduces a slight inconsistency in ASCA+ in that the matrix analyzed by PCA in ASCA+ differs from the matrix from which the explained variance is computed [29]. For **X**, for example, the PCA is applied to $\mathbf{X}\widehat{\mathbf{B}}$ and the explained variance is computed from $(\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{X}\widehat{\mathbf{B}}$ (Supplement S1). A corollary is that the sum of the eigenvalues of the PCA in ASCA+ will be bigger than (or equal to) the reported explained sum of squares. In RDA, the PCA and sum of squares are both computed from $(\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{X}\widehat{\mathbf{B}}$.

ASCA+ needs sum coding of factors and interactions, whereas in RDA any coding that preserves the spaces spanned by the terms gives the same eigenvalues and loadings. Consider the model A + B + A:B and its decomposition in A and B + A:B (the PRC model). In terms of fitted response, the model can be specified as A + A:B as well, but would ASCA+ always provide the same results for the two specifications and, if not, does any of these specifications perform as well as RDA? We investigate this question, and what the differences noted in this section can mean in practice, using simulation and real data.

2.5. Remarks

For the equivalence of RDA and ASCA in complete designs, RDA must be understood as a sequence of RDAs, just as ASCA uses a sequence of PCAs. In the theory of RDA the emphasis is on the decomposition of the focal term (*i.e.* the term that is rank-reduced; in this paper **XB**), but in the function rda of the R software library vegan [19] all terms in equation (1) are decomposed sequentially as in type I sums of squares, including a PCA of the residual $\widehat{\mathbf{E}}$. In the Canoco software [18], the **Z** term is not decomposed, but the **X** and $\widehat{\mathbf{E}}$ terms are. Also, for variance decomposition, ASCA+ and RDA require the same sequence of analyses.

The typical display of a factor with two levels in ASCA and RDA is embarrassingly uninteresting when the constrained scores are displayed as these contain two values only. ASCA+ and RDA are identical in this case, but rarely so in terms of what is displayed. In the standard RDA display, the unconstrained scores are added and the second dimension is used to display the first component of the residual matrix $\widehat{\mathbf{E}}$.

More, generally, there are two ways in which error and natural variability can be displayed in ASCA and RDA: 1) by the unconstrained scores and 2) by decomposition of the matrix of residuals $\widehat{\mathbf{E}}$. The displayed rank of **X** can be limited to a single dimension in what is called 'hybrid ordination' (namely a combination of constrained ordination and unconstrained ordination of the residuals), so that the second dimension is free to show the first component of the residuals $\widehat{\mathbf{E}}$. Both Canoco and vegan allow the user to display either the constrained or the unconstrained scores of any term that is not a residual. Displays with unconstrained scores can also be enhanced with ellipsoids summarizing their dispersion [18,53]

Note also that the highest order interaction in factorial design may be unreplicated and, thus, be indistinguishable from error. An example is the guinea pig example in Smilde et al. [4].

ASCA typically decomposes both main effects and interactions. However, interaction is a statistical concept that is perhaps easy to grasp theoretically, but nevertheless difficult to interpret in practice for two reasons. One reason is that an interaction, at least in a complete design, models the deviation from the main effects. So, one needs to keep the main effects in mind when interpreting the interactions. Moreover, the loadings differ between the PCA plots of main effects and interaction. The second reason becomes clear when the model is rephrased as a regression model, as it is in ASCA+. Regression coefficients have a simple interpretation when associated with quantitative variables. The regression coefficient gives the change in the response per unit change in a quantitative predictor, with the values of the other predictors held constant. But a regression coefficient associated with an interaction cannot have this interpretation. A change in the value of an interaction necessarily implies a change in at least one value of the constituting predictors. Analogously, the regression coefficient associated with a factor level estimates the change in response when changing from the reference level to the factor level, assuming treatment coding and constant values of everything else in the model. In other words, it is the difference in response between the associated level and the reference level. A regression coefficient associated with an interaction is not a change in response due to changing levels. It is a difference of a difference, thus involving four levels (two of A and two of B).

A simpler way of looking at interactions is to define interaction asymmetrically as the effect of B that changes with the level of A (or reversely). This simpler point of view is exploited in PRC [22,23]. It applies PCA to the model B + A.B, with covariate factor A, so that the main effect of B and its interaction with A can be interpreted simultaneously in terms of a single set of loadings as discussed further in the next subsection.

2.6. Special cases

2.6.1. Weighted effect ASCA (WE-ASCA)

WE-ASCA [21] is ASCA+ with sum coding replaced by weighted effect coding [32]. Weighted effect coding is an ingenious way of coding a design matrix such that the main effects represent deviations from the overall sample mean (column-means of **Y**) and do not change when interactions are added to the model, properties that do not apply to sum coding in unbalanced designs. In WE coding, main effects and interactions are orthogonal to the intercept and the interactions are also orthogonal to the main effects. The decomposition of the models A + B and A + B + A.B, may therefore yield different results for A and B in ASCA+, but these parts yield the same result with WE-coding. Main effects are not necessarily orthogonal to one another in WE-ASCA. Therefore, WE-ASCA is identical to RDA for the model decomposition (A + B) + A.B, but not for A + B or A + (B + A.B) as A is not necessarily orthogonal to B in incomplete designs. However, we discovered that the current implementation of WE coding [33] fails to fulfill these properties if some combinations of A and B levels are missing in the data. For complete designs, WE-ASCA is identical to ASCA and RDA.

2.6.2. Principal response curves (PRC)

PRC was introduced as a model to summarize the multivariate time-

dependent response of aquatic species to a toxicant in a designed experiment and was estimated by RDA with covariates [22,23]. With A representing the factor time and B the factor representing the different levels of application of the toxicant, PRC is based on the model A + A.B or, equivalently A + (B + A.B), with **Z** coding for A and **X** coding for the second term. The regression coefficients of the second term are assumed to be of reduced rank and are estimated by RDA as in section 2.3. The full-rank regression coefficients are thereby approximated by their reduced-rank counterparts. PRC is thus precisely RDA applied to our focal model.

A PRC graph is a plot of the canonical coefficients of the first axis, *i.e.* $\mathbf{C}_{\mathbf{X}}^{\mathrm{r}[1]}$, against time, with points of the same treatment level connected by lines, and with a plot of the loadings $\mathbf{P}_{\mathbf{X}}^{\mathrm{r}[1]}$ on the (right-hand) vertical axis (Fig. S1). Canonical coefficients depend on the coding of matrix **X**. At its introduction, PRC [23] used treatment coding with the control treatment as reference, so that the control is represented by the horizontal x-axis at y = 0 and the lines or curves of the other treatments are deviations from the control. Because RDA is invariant to the coding system, the coding can be adopted post-RDA. The canonical coefficients for the new coding system can be obtained by regression of the score vector ($\mathbf{T}_{\tilde{\mathbf{X}}}^{\mathrm{r}}$ or $\mathbf{T}_{\tilde{\mathbf{X}}}^{\mathrm{Er}}$) on newly coded **X** and **Z**. The coefficients can be computed even manually as they are the deviations, per time point, of the constrained scores from those of the reference. The variability of the PRC curves can be displayed using the unconstrained scores ($\mathbf{T}_{\tilde{\mathbf{X}}}^{\mathrm{Er}}$), optionally shifted, *i.e.* using the $\mathbf{T}_{\tilde{\mathbf{X}}}^{\mathrm{r}}$ - scores of the reference as offset (so that the reference samples vary around the horizontal axis at y = 0). In other words, the PRCs do not need the canonical coefficients per se; they can equally well be obtained from the usual constrained RDA scores. The innovation of this paper is that the PRCs can be supplemented with the unconstrained RDA scores so as to show replicate variability. Figs. 1 and 3 show examples.

PRC treats the factor interaction asymmetrically, while statistical interaction is a symmetric concept. There are a number of reasons why the asymmetric treatment in PRC is beneficial. First, the factors are often asymmetric in that one factor is of more scientific interest than the other. In the prototype example of PRC where A is the factor time and B is a treatment, the scientific question 'how the treatment effect changes over time' is already asymmetric. Second, while the statistical interaction may be statistically significant, it may be small or may not change the ranking of the treatments between different levels of the other factor(s). Ranking is often of greater practical interest than precise quantities. If the factors are of equal scientific importance, it seems sensible to construct both a PRC diagram based on the model A + A:B and one based on the model B + A:B.

As ASCA+ and RDA are identical for complete designs, ASCA and PRC are identical for our focal model, herein after referred to as the PRC model. Opposite claims in the literature [3,25] were refuted in Vendrig et al. [54]. Fig. 4 in Timmerman et al. [26] is thus a regular PRC - particularly as the displayed scores are deviations from the reference - with loadings in a bar plot, a layout independently chosen in Verdonschot et al. [55].

3. Numerical comparison

3.1. Simulation

Three simulation series were performed, all based on the PRC model. We compare the methods and the two ways to specify the PRC model: A + A.B and A + (B + A.B) for factors A and B, with dimension reduction of the terms A.B and B + A.B, respectively. The first series illustrates when ASCA+, WE-ASCA and RDA differ in results and how the differences depend on the design of the study and the way the model is specified and coded. The differences turn out to be largest in the design with one or more factorial combinations missing ('empty cell'), which is investigated further in a second series of analyses. In the third series hundreds of unbalanced data sets are simulated according to scenarios that differed in numbers of response variables and levels of factors, in size of the error variance and in balance of the design.

Four types of designs were generated. Design 'Complete' is an equireplicated design with all combinations of two factors A and B replicated twice. Sometimes one treatment level is replicated more (or less) often [23], and this is illustrated in design 'Proportional' in which the level B1 has four replications instead of two (Table 1). The levels of the two factors were assigned equi-spaced numbers, then scaled to zero mean and unit variance, resulting in two quantitative predictors $\mathbf{x}_1$ and $\mathbf{x}_2$, which were then used to generate *m* response variables from the model

$$y_{ik} = a_0 + a_k x_{1i} + b_k(x_{2i} + 0.5x_{1i}x_{2i}) + \sigma\varepsilon_{ik} \quad (i = 1,\ldots,n; k = 1,\ldots,m) \tag{4}$$

with $a_0 = 10$, a_k and b_k uniform random between −1 and 1 and ε_{ik} independent standard normal error with standard deviation σ (1, 2 or 3). The model has reduced-rank 1 because the coefficient of the interaction is a multiple (half) of the effect of $\mathbf{x}_2$. The model gives treatment differences for levels of B that increase with the level of factor A without changing sign (Fig. S2).

Table 1
Design types: percentage squared difference between the fitted X-terms (Δ) and correlation (ρ) of the loadings of the first axis of ASCA versions with those of RDA[a], or (in the last two rows) between those of the two ASCA versions, for the terms A.B and B + A.B in the models A + A.B and A + (B + A.B), respectively, for four examples with four response variables in designs with factors A and B (sum: sum coding; WE: weighted effect coding; λ_k: RDA eigenvalue of axis *k*).

Design type		Complete		Proportional		Unbalanced		Empty cell	
Counts in 4 × 3 design		B1 B2 B3 A1 2 2 2 A2 2 2 2 A3 2 2 2 A4 2 2 2		B1 B2 B3 A1 4 2 2 A2 4 2 2 A3 4 2 2 A4 4 2 2		B1 B2 B3 A1 8 2 3 A2 2 2 2 A3 3 2 1 A4 3 2 2		B1 B2 B3 A1 8 2 3 A2 2 2 2 A3 3 2 0 A4 3 2 2	
λ_1/λ_2		1.6		1.4		2.7		2.4	
ASCA+	term	Δ	ρ	Δ	ρ	Δ	ρ	Δ	ρ
Comparisons of ASCA extensions with RDA									
sum	A.B	0%	1.00	4%	0.83	6%	0.66	16%	0.76
sum	B + A.B	0%	1.00	4%	0.83	6%	0.66	29%	0.69
WE	B + A.B	0%	1.00	0%	1.00	2%	0.98	25%	0.17
Comparisons between ASCA extensions									
WE[b]	B + A.B	0%	1.00	4%	0.83	3%	0.69	3%	0.60
A.B[c]	B + A.B	0%	1.00	0%	1.00	0%	1.00	7%	0.98

[a] RDA is invariant to coding and term and so would give Δ = 0 and ρ = 1 in all cases.
[b] This row compares two versions of ASCA (WE-ASCA and ASCA+).
[c] This row compares the two model specifications for ASCA+ using sum coding.

The remaining two designs are 'Unbalanced' and 'Unbalanced with an empty cell'. These designs were generated by drawing the predictors $\mathbf{x}_1$ and $\mathbf{x}_2$ independently from the standard normal distribution and by subsequent discretization of $\mathbf{x}_1$ and $\mathbf{x}_2$ into factors A and B with equal numbers of observations per level. The number of observations for the first level of factor B could be doubled by starting with one category more and merging the first two levels.

In the first series we simulated 4 × 3 factorial designs of each of the four types of designs with four response variables (Table 1). We report two statistics. The first is the percentage squared difference between the fitted X-terms (Δ), defined as 100 times the sum of squared differences between $\mathbf{X}\widehat{\mathbf{B}}$ for ASCA+ and WE-ASCA and $\widetilde{\mathbf{X}}\widehat{\mathbf{B}}$ for RDA, divided by the sum of squares of $\widetilde{\mathbf{X}}\widehat{\mathbf{B}}$. The second statistic is the correlation (ρ) of the loadings of the first axis of the X-term in ASCA+ and WE-ASCA with those in RDA. RDA was chosen as a reference because its results do not change with the way the model and design matrices are specified.

In the second series we simulated a complete 10 × 5 design in two replications with 7 response variables. The first eigenvalue is 2.7 times the second, so that the first axis is well determined. From the 100 samples in this design we deleted the two observations on the (A10,B5) combination, resulting in one empty cell in the A × B contingency table. The data set so obtained was re-analyzed repeatedly under random re-ordering of the levels of factors A and of B. We report the correlation between the loadings of the first axis and the loadings of the complete design as the latter do neither depend on the method nor on the way the model is specified nor on the order of the factor levels. We also report the variance explained of the first axis after deletion of the two observation, expressed as percentage of the variance explained in the complete data. The variance is adjusted for the effects of factor A.

In the third series, the methods are compared on unbalanced data across all combinations of: 1) numbers of response variables (m = 10, 100 or 1000), 2) numbers of levels of factors A (4 or 10) and B (3 or 5), 3) size of the error variance (σ = 1 or 4) and 4) balance of the design (doubling or no doubling of the first level of factor B). For each of these 48 combinations (scenarios), 200 data sets were generated, many of which contained one or more empty cells. The comparison is in terms of the variance explained by the first axis, expressed as percentage of the variance explained by the first axis in a sum-coded RDA of the term A:B with model specification A + A:B.

R-code [31] of the simulations and analyses can be found in Supplements S2–S6. All numerical implementations of the methods used the same $\widehat{\mathbf{B}}$, if their coding and model specification matched (Supplement S2.2).

3.2. Real data: the Ossenkampen experiment

The methods do not only differ in theory but may also differ in practice. This is demonstrated with an ecological, publicly available data set [9] which comes from the Ossenkampen experiment, a long-term fertilizer experiment (1958–2007) in grassland with 98 response variables (abundances of the plant species that grew in the Ossenkampen plots, measured as counts in 100 subsamples). The analysis in Ref. [9] focused on the (non-linear) time trends in the species abundances in the plots and how these relate to nitrogen deposition. Here we focus on the time-dependent treatment effects (see also the supplementary information in Ref. [9]). The experiment started in 1958 with 12 plots arranged in two randomized complete blocks with six treatments: four types of fertilizer (K, P, PK or NPK), liming and no fertilizer. Eight years later (in 1966) each block was extended with two plots, one with N and another with NPK fertilization. The two limed plots behaved rather differently from the rest and are omitted here. We combined the K and P plots with the no fertilizer plots so that the control treatment (Cntrl) consisted of 6 plots, with the aim to make the variability among replicates more interesting. Three plots of the control and one PK plot were not sampled in 1967 and two NPK samples were not sampled in 1984; these missing samples were not imputed in this paper; they were imputed in Ref. [9] so as to enable design-based permutation testing of plots (instead of individual samples) [23]. The data analyzed here consist of 512 samples from 14 plots with four treatments (Cntrl, PK, N and NPK) measured in 39 different (not necessarily consecutive) years, taken as unrelated levels in the analysis and plotted as number of years after the start of the experiment. Factor A is thus year (with 39 levels) and factor B fertilizer treatment (with four levels) and the missing (A,B) combinations are due to the late start of the N treatment. The counts were log(y+1)-transformed.

4. Results

4.1. Simulation

For a complete design, all methods are equal and do not depend on the model specification ($\Delta = 0$ and $\rho = 1$ in Table 1). In the Proportional design, in which the levels of B differ in the number of replications (Table 1), WE-ASCA equals RDA (because the main effects are still orthogonal), but ASCA+ differs slightly (Δ = 4%; ρ = 0.83). The two versions of ASCA thus differ. The Proportional design allows an orthogonal ANOVA [56] and thus a unique ASCA solution, which is identical to the WE-ASCA and RDA solution. This example shows that ASCA+ does not include ASCA. The two ways to specify the model yield the same result with ASCA+ (last row in Table 1). With unbalanced data, WE-ASCA is no longer equal to RDA (because the main effects are not orthogonal any more), but the difference is small ($\Delta = 3\%$; $\rho = 0.98$). By removing the single (A3,B3) observation from the unbalanced data, a data set arises with an empty cell and all methods produce different results (Table 1, last column). This is investigated further in Table 2.

In the example of Table 2, an empty cell was created by deleting two observations of the 100 observations in a complete twice replicated 10 × 5 design. Applied to the complete data, all methods give the same result and the first eigenvalue is twice the second (so that the first axis is well determined). But after the deletion of two observations, the results of the ASCA extensions depend on the order of the factor levels. The dependence on factor level order is particularly large when the model is specified as A + (B + A.B) instead of as A + A.B (Table 2). A removal of 2% of the data may give loadings that are almost uncorrelated ($\rho = 0.01$) to the loadings of the complete data and a reduction of about three-quarter of the variance explained (100–23 = 77% or, when compared to the best factor order, (87–23)/.87 = 74%). The deterioration in the performance of ASCA+ and WE-ASCA depends on method, model type

Table 2
Example, based on a 10 × 5 design in two replications with one empty cell and 7 response variables, of the dependence of ASCA+ and WE-ASCA, and the independence of RDA, to factor level order and model specification in designs with a single empty cell. Dependence is measured by the range of correlations (ρ) between the (first axis) loadings of the complete design (100 samples) and those of the design with one cell missing (98 samples) across 10^4 random reordering's of the levels of the two factors and the corresponding percentage variance explained (Explained) compared to the fit by the first axis in the complete data. RDA is not dependent on factor level order, here visible in that minimum and maximum are equal. Two ways to specify the model are used, A + A.B and A+(B + A.B), with focal terms A.B and B + A.B, respectively. (Min.: minimum; Max.: maximum; -: impossible specification; in bold the value that occurs in more than 80% of the 50 possible positions of the empty cell).

method	term							
	ρ				Explained			
	A.B		B+A.B		A.B		B+A.B	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
ASCA+	**0.75**	0.98	**0.01**	0.98	**84%**	87%	**23%**	87%
WE-ASCA	-	-	0.06	**0.98**	-	-	23%	**87%**
RDA	**0.98**	**0.98**	**0.98**	**0.98**	**87%**	**87%**	**87%**	**87%**

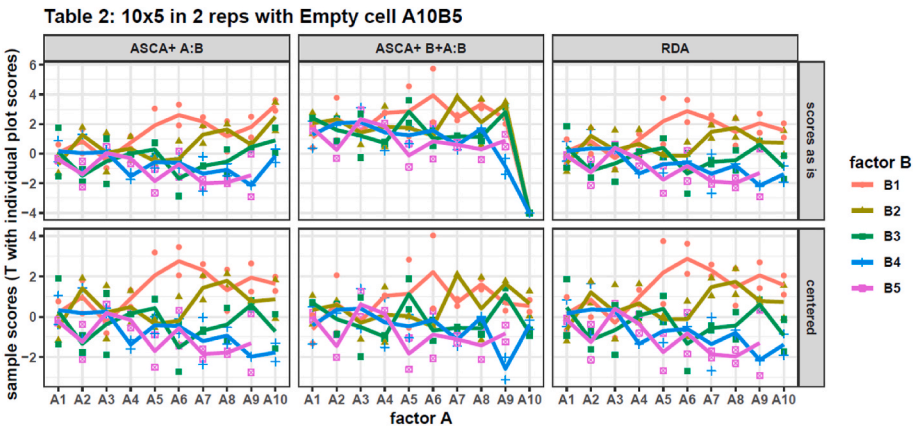


Fig. 1. ASCA+ and RDA scores of principal response (first axis) to factor B in dependence of the levels of factor A (horizontal) for the data of Table 2 (empty cell A10B5). Top row: ASCA+ scores ($\mathbf{T}^{a}_{\mathbf{X}}$ and $\mathbf{T}^{Ea}_{\mathbf{X}}$) and the constrained and unconstrained scores of RDA (lines and points, respectively). The bottom row re-expresses the scores in the top row as deviation from mean per A category (which does not change the RDA scores). WE-ASCA gave the same result as ASCA with model specification A + (B + A:B) and the natural order of factor levels (A1-A10 and B1–B5); for other orders of levels their results may differ (Table 2). Extreme negative scores in the top row are truncated to −4.

and position of the empty cell in the 10 × 5 cross-table (Supplement S4). WE-ASCA does somewhat better than ASCA+, in that it deteriorates in far fewer factor level orders. For RDA, minima and maxima are equal (Table 2), showing that RDA is neither dependent on factor level order nor on the way the model is specified. Table S1 shows qualitatively similar results for the Complete 4 × 3 design of Table 1, from which two observations of a single AB-combination are deleted.

Fig. 1 show score plots of the data analyzed in Table 2. The graphs (except one) show that the treatment differences increase in size with the level of A (in agreement with the truth in Fig. S2). Because the loadings of all variables are positive (not shown), the values of all response variables are shown to decrease with the increase of the level of B, and the decrease become larger for increasing level of A. Note that the purple line for B5 in each graph ends at A9 as the (A10, B5)-combination is missing and that, at A10 in the top row, the lines differ between ASCA+ and RDA. RDA gives the same result for the two ways to specify the model, but ASCA+ gives different results with extreme scores at level A10 for model A+(B + A:B). For the complete data, the plots are identical across methods and model specifications and almost indistinguishable from those of RDA in Fig. 1 (not shown).

Fig. 1 also shows the variation in the samples around the treatment lines. The points of the two individual replicates per (A,B) combination has not been connected as the simulated error was independent. The treatment lines are means of two points and thus lie precisely in middle of the two points per (A,B)- combination.

In the bottom row of Fig. 1, the scores in the top row are re-expressed, for each level of A, as deviation from the mean. In other words, bottom row shows the residuals of the regression of the original scores onto factor A, *i.e.* using $(\mathbf{I}_n - \Pi_Z)\mathbf{X}\widehat{\mathbf{B}}\mathbf{P} = \widetilde{\mathbf{X}}\widehat{\mathbf{B}}\mathbf{P}$ instead of $\mathbf{X}\widehat{\mathbf{B}}\mathbf{P}$ with additional column-centering and $\mathbf{P}$ the loadings of the method. The re-expression removes any effect of A (compare the top and bottom rows), thereby removing the defects of the original scores in the top row. Whereas RDA removes the effect of A before applying PCA, the re-expression of the ASCA scores removes the effect of A after PCA is applied. The explained variance reported in Table 2 and Fig. 2 are based on the re-expressed scores, and are thus not influenced by the extreme scores such as those in Fig. 1 for ASCA B + A:B.

The third simulation series gives an impression how the ASCA extensions perform compared to RDA across 14400 data sets when the design is unbalanced with or without empty cells (Fig. 2). Without empty cells, the ASCA extensions are slightly inferior to RDA in terms of variance explained, and never better. Without empty cells WE-ASCA is slightly better than ASCA+, but slightly worse than RDA. With empty cells, RDA outperforms the ASCA extensions by large, and the different ways to specify and encode the model yield wildly different results in the ASCA extensions, but identical results in RDA (Fig. 2).

4.2. *Real data: the Ossenkampen experiment*

Fig. 3 compares ASCA+ and RDA for the Ossenkampen experiment. The differences are largest for the original scores in the top row, in particular for the first ten years of the experiment when the N treatment was not yet included. When the scores are re-expressed as differences from the mean per year or, as in the bottom row of Fig. 3, from the

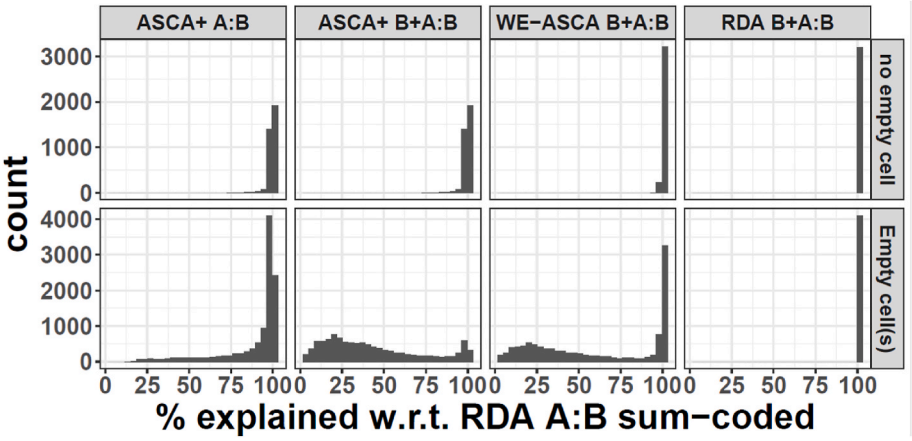


Fig. 2. Histograms of percentage variance explained by the first axis of the PRC term (A:B or B + A:B, both given A) compared to that of a sum-coded redundancy analysis (RDA) of the term A:B given A, showing that RDA with model B + A:B (sum or treatment coded) given A gives the same results (relative variance explained 100%) and the ASCA extensions (ASCA+ and WE-ASCA) yield results which depend on model specification and factor coding with less than or equal to 100% relative variance explained. The data were simulated from the one-dimensional PRC model of equation (4) according to 72 scenarios from each of which 200 data sets were generated. The scenarios were all combinations of numbers of response variables (10, 100 or 1000), numbers of levels of factor A (4 or 10), and factor B (3 or 5), σ (1, 2 or 3), and doubling (yes or no) of the number of observations for the first level of factor B. The height of the bars at 100% for RDA has been reduced to the maximum per row, so as to allow maximum visual detail in the histograms of the other

methods (empty cell: a factor combination without observations).

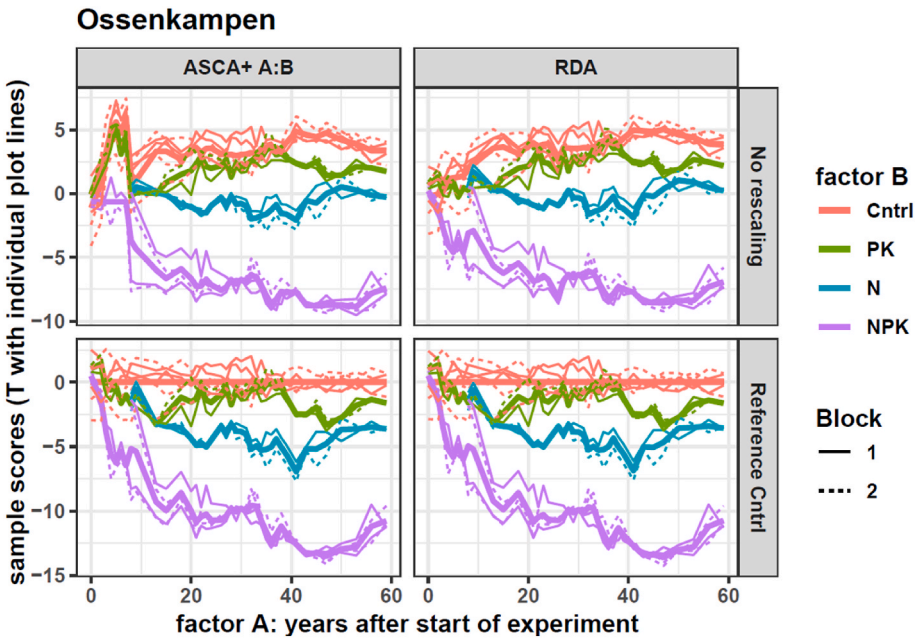


Fig. 3. Ossenkampen experiment: principal time-dependent (factor A) response of 98 grassland plant species to fertilizers (factor B) as analyzed by ASCA+ and RDA with the original scores (top row) and re-expressed scores (bottom row). The RDA plot in the bottom right is identical to a traditional PRC, without loadings for the individual variables, but with lines that show the natural variation around the treatment lines. The ratio of first two RDA eigenvalues is 6.3.

control treatment in each year, the differences are rather small.

The factor block was not used in the ASCA and RDA analyses, but is visible in Fig. 3 in the form of the individual plot lines of the 14 plots (solid versus dotted lines). The block effects, if any, appear small on the first axis of the analysis shown in Fig. 3, as do the differences between the six plots classified here as control but that in the experiment were different treatments (reference, P and K). The factor block could have been added to **Z** as it was in Ref. [9].

The individual plot lines in Fig. 3 of the two N and two NPK plots that started after ten years, are initially close to the lines for the control but move later on in the direction of the two NPK plots that were included from the start of the experiment. The individual plot lines suggest that the NPK plots that started later might have been better classified as control for the first few years. Because the plant species composition lags behind the fertilization, the immediate classification of these plots as NPK potentially underestimated the NPK effect in these years in Fig. 3.

Fig. 3 shows just one of the many ASCA solutions. Fig. 4 shows some other solutions, obtained by reordering the levels of the treatment factor (three orders: in order of the legend of Fig. 4, alphabetical order, alphabetical but with the N treatment moved to last). Fig. 4 (first column) shows that the ASCA solution shown in Fig. 3 is not the one that is most similar to the RDA solution, but after re-expression of the scores to remove any effect of years (A) all solutions with model A + A:B shown are very similar to RDA. However, with model A + (B + A:B) there are more solutions that differ appreciably from RDA and among one-another. WE-ASCA (last column) gave about the same solution as RDA

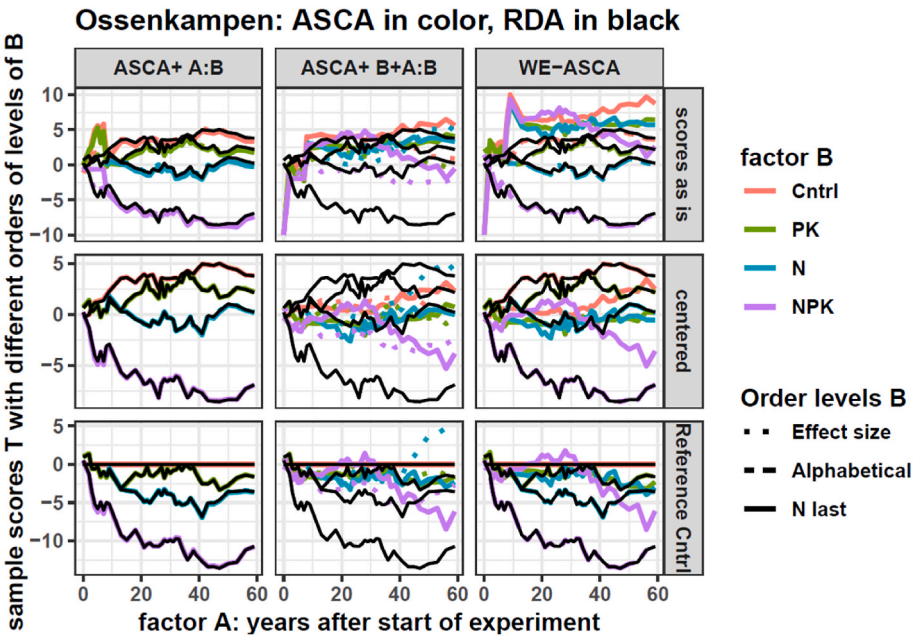


Fig. 4. Ossenkampen experiment: Dependence of ASCA+ and WE-ASCA on the order levels of factor B (fertilizer: Cntrl, PK, N, NPK) and the model specification (A + (A:B) or A+ (B + A:B)), of which the second term is analyzed here. Shown are three type of scores: (scores as is: the standard scores of section 2 ($\mathbf{T}^{a}_{\mathbf{X}}$ for ASCA versions and $\mathbf{T}^{r}_{\mathbf{X}}$ for RDA); centered: scores that are centered per level of A; Reference Cntrl: the standard PRC scores, *i.e.* scores expressed as deviation of Cntrl (the reference level)). Excessive negative scores in the top row are truncated to −10. The black lines in each panel are the corresponding scores of RDA.

for some orders (as is visible by the colored lines with a thinner black line close or overlying), but not for all, whereas ASCA+ with the same model (middle column) resulted in completely different treatment lines throughout (the black lines are separate from the colored lines).

5. Discussion and conclusions

This research highlights issues in the ASCA extensions ASCA+ and WE-ASCA, particularly when applied to incomplete experimental designs with missing factor combinations, and a slight inconsistency that went unnoticed so far (the variance decomposed by PCA is not equal to that used in the variance decomposition) [29].The good news is two-fold: 1) in slightly incomplete designs the ASCA extensions perform reasonably well (Fig. 2) so that almost no ASCA application needs to be revised, and 2) an alternative exists, RDA, which has a solid mathematical basis and for which user-friendly, versatile software implementations exists [18,19].

The model A + (B + A:B), the PRC model, was taken as an example to show the similarities and differences between RDA and ASCA extensions, but the theory in section 2 is quite general and, in particular, is neither limited to models that combine a main effect and an interaction nor limited to a particular experimental design. With minor disbalance, the differences appear small. The biggest issues of the ASCA extensions appear with empty cells; models with main effects only are expected to show small differences because marginal cells of main effects are rarely completely empty.

RDA and the ASCA extensions use the same multivariate regression fit $\widehat{\mathbf{Y}}$ (which can be obtained by a sequence of univariate fits), and RDA can use the same coding for factors and interactions as ASCA+ and WE-ASCA (but also indicator and treatment coding). The key difference is that RDA first makes the focal term orthogonal to the covariates and then performs PCA, whereas the ASCA extensions do it the other way round, as in the bottom and middle rows of Figs. 1 and 4, respectively, or PCA on $\mathbf{X}\widehat{\mathbf{B}}$ only as in the top rows of Figs. 1, 3 and 4. Consequently, RDA and the ASCA extension perform PCA on different matrices. In the PRC model, orthogonalization means: take out the means per level of the covariate factor A. This is presumably the first paper that considers post-processing of the ASCA+ and WE-ASCA scores by projection in order to circumvent the issue of extreme scores that may arise in designs with empty cells (Figs. 1 and 4). The extreme scores are taken care of by this post-processing/projection step and improved the summary statistics on the basis of which we compared the methods quantitatively (Table 2 and Fig. 2). Without the post-processing step, ASCA+ and WE-ASCA performed even worse.

The multivariate regression step in ASCA+ takes care of unbiased estimation of the parameters. However, with rank-reduction of an ASCA term, bias is introduced again as terms (submatrices) are not orthogonal in unbalanced designs (see the issues at A10 in Fig. 1). This is why, in RDA, the focal submatrix is made orthogonal with respect to the other submatrices before PCA is applied. Such an orthogonalization might also be of interest to improve performance of ASCA extensions for random effects [57,58]. Specifically, the suggestion is to make each focal submatrix of a fixed effect orthogonal to submatrices of other fixed effects, before PCA is applied.

By-products of this work are three-fold.

1) The proportional design allows an orthogonal decomposition, which WE-ASCA reveals but ASCA+ does not (Table 1). Consequently, ASCA+ does not include ASCA.
2) Principal response curves are traditionally based on the canonical coefficients. This paper shows that they can be equally well derived from the constrained RDA scores. In consequence, replicate variability can be shown by also plotting the unconstrained scores (Figs. 1 and 3).
3) An error was detected in the function summary.prc of vegan version 2.6–4, which affects the results of PRC on unbalanced data and which will likely be repaired in the next version. This shows the importance of continued scrutiny and maintenance of software.

RDA is typically introduced with dimension reduction of the entire model or a single part of it, as in equation (2). In contrast, ASCA typically starts with a complete decomposition of the model into main effects and interactions and PCA is applied to each term of the decomposition. It is difficult to interpret all PCA plots simultaneously, particularly because they have different loadings. For the decomposition of one term (or a combination of terms) RDA is optimal. We see little reason to aim for joint optimality of the decompositions, although technically possible [27,52]. It is wiser to go beyond ordinary least-squares and to introduce explicit weighting with the inverse of the error variance [20,28] and covariance [46,59].

The paper introducing PRC [23] using RDA had a proportional design and therefore would have allowed a similar analysis using WE-ASCA, but not ASCA+ (Table 1). A recent PRC application with a complete design is Ref. [60].

The response variables in real data example consisted of counts, which we analyzed after the logarithmic transformation log(count+1). A log-transformation generally avoids that variables with large variation ranges will overly influence the dimension reduction of a term of interest. After log-transformation, large variation is of natural interest, as it reflects large percentage change on the original scale [50]. For data with zeroes, addition of a pseudo count is needed to avoid problems with the zeroes. This is a simple approach that we find often effective. However, it is good to mention canonical correspondence analysis [7] which is particularly attractive for count-like strictly compositional data with many zeros, *i.e.* count-like data, the sample sum of which is a technical artefact, such as microbiome data [16,18], and to also mention novel approaches that extend RDA to the framework of generalized linear modeling with [44,45] and without [42,43] additional random effects, with software implementations in Stata [45] and R [43,61]. These are the types of method the ASCA extension for random effects [57] has to be compared with. Also, reduced-rank regression is an active research area in statistics [62].

Statistical inference in RDA, ASCA and its extensions must proceed via resampling methods (permutation testing, bootstrapping and/or cross-validation). The possibilities, issues and pitfalls are many, but do not differ between RDA and ASCA variants. Nevertheless, we list some important points to consider: 1) restrictions on permutations or bootstraps [15,18,63,64] (e.g. restrictions to permute samples within blocks only, to keep samples of the same experimental unit together, the balanced bootstrap [65]), 2) what aspect of the data must be resampled (e.g. the raw data value, a response or predictor residual) [17,66,67] and 3) what statistic is used (mean square or F-statistic in testing [68], naïve bootstrap or t-value bootstrap [65]). Specifically, each individual term essentially requires its own permutation test. It is tempting to save computer time by performing a single set of permutations only (rows of **Y**, or levels within each factor of the design). However, this does not leads to good significance tests for individual terms [68,69].

RDA obviates the necessity to impute missing samples or to otherwise rebalance a data set. Nevertheless, the statistical inference can sometimes be simplified or be made more robust by rebalancing the data. An example in case is the repeated measurement design in the Ossenkampen experiment. For statistical inference only, the missing values (after 1965) were imputed in Ref. [9], so that samples of the same plot could be kept together and the consecutive samples could be permuted cyclic so as to allow for autocorrelation in each sequence of samples of the same plot [18,63,64].

In the Ossenkampen analysis, year is treated as a factor with 39 levels, giving irregular treatment curves. These curves can be make smoother by replacing the factor year by a set of B-splines [70] and including their interaction with the treatments. This presents no

problem to RDA as it is based on regression resulting in a regression spline fit. Smoothing splines or P-splines [71] would be even nicer, but require a modification of the RDA algorithm because they require a ridge regression version of RDA (and, likely, a cross-validation approach to choose the smoothing parameter).

With resampling and Bayesian computing, repeated analysis of a single data set may result in slightly different results, but otherwise one might expect that a statistical method gives reproducible results. If one or more factor combinations of a factorial design are missing, the results of ASCA extensions are not reproducible without explicit specification of the factor order used in the analysis and the precise way the model is specified and encoded. Some ways to specify and encode the PRC model lead to inferior results in both ASCA+ and WE-ASCA. These issues can be prevented by using RDA instead.

The results of this paper can be summarized by the highlights.

1. A sequence of redundancy analyses (RDA) is more general in theory and practice than ASCA.
2. The ASCA extensions for unbalanced data, ASCA+ and WE-ASCA, are unstable in designs with a missing factor combination.
3. RDA outperforms ASCA+ and WE-ASCA.

Consequently, extensions should be based on RDA's statistical model rather than ASCA-related algorithms and this should be clear in the name of new extensions.

Author statement

Cajo J.F. ter Braak: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing- Original Draf, Review & Editing, Visualization.

Declaration of competing interest

Cajo ter Braak is the first author of the software package Canoco for visualization of multivariate ecological and environmental data and a co-author of vegan function prc. I have no commercial links to the above and no further conflicts of interest.

Data availability

Open science: all is on https://doi.org/10.6084/m9.figshare.22099844.

Acknowledgements

I thank Pietro Franceschi for providing an ASCA+ implementation that kick-started this project, and Petr Šmilauer, Pietro Franceschi and the anonymous reviewers for comments on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2023.104898.

Appendix A1: theory

A1.1. Derivation and algorithm of RDA and relation to PCA and multivariate regression

This section provides an original and complete derivation of RDA. We start with the model in which the intercept is included in the matrix $\mathbf{Z}$

$$\mathbf{Y} = \mathbf{ZA} + \mathbf{XB} + \mathbf{E} \tag{A1}$$

and estimate the unknown parameter matrices $\mathbf{A}$ and $\mathbf{B}$ by least-squares by minimizing the sum of squares of residuals

$$\|\mathbf{Y} - (\mathbf{ZA} + \mathbf{XB})\|^2, \tag{A2}$$

subject to the restriction that the rank of $\mathbf{B}$ is r, where $\|\mathbf{F}\|^2 = \mathrm{tr}(\mathbf{F}'\mathbf{F})$, the Frobenius norm of $\mathbf{F}$. Let $\Pi_{\mathbf{Z}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-}\mathbf{Z}'$, the projection operator on to $\mathbf{Z}$ with $(\mathbf{Z}'\mathbf{Z})^{-}$ a generalized inverse. For brevity, we define $\mathbf{E} = \mathbf{Y} - (\mathbf{ZA} + \mathbf{XB})$, the matrix of residuals, so that we must minimize

$$\|\mathbf{E}\|^2 = \|\Pi_{\mathbf{Z}}\mathbf{E} + (\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{E}\|^2 = \|\Pi_{\mathbf{Z}}\mathbf{E}\|^2 + \|(\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{E}\|^2, \tag{A3}$$

because $\Pi_{\mathbf{Z}}\mathbf{E}$ is orthogonal to $(\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{E}$. The last two terms in equation (A3) are developed further. The argument of the first term

$$\Pi_{\mathbf{Z}}\mathbf{E} = \Pi_{\mathbf{Z}}(\mathbf{Y} - (\mathbf{ZA} + \mathbf{XB})) = \Pi_{\mathbf{Z}}\mathbf{Y} - (\mathbf{ZA} + \Pi_{\mathbf{Z}}\mathbf{XB}) = \Pi_{\mathbf{Z}}\mathbf{Y} - \mathbf{ZA}^{*} \tag{A4}$$

with $\mathbf{A}^{*} = \mathbf{A} + (\mathbf{Z}'\mathbf{Z})^{-}\mathbf{Z}'\mathbf{XB}$. Because $\mathbf{A}$ is without constraint, $\mathbf{A}^{*}$ is without constraint so that the minimum of $\|\Pi_{\mathbf{Z}}\mathbf{E}\|^2$ is 0. Note that the least-squares estimate of $\mathbf{A}^{*}$ consists of the regression coefficients of the formal model $\mathbf{Y} \sim \mathbf{Z}$.

With $\widetilde{\mathbf{Y}} = (\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{Y}$ and $\widetilde{\mathbf{X}} = (\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{X}$, the argument of the second term in equation (A3)

$$(\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{E} = (\mathbf{I}_n - \Pi_{\mathbf{Z}})(\mathbf{Y} - (\mathbf{ZA} + \mathbf{XB})) = \widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{B}. \tag{A5}$$

The minimum of $\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{B}\|^2$ is derived, starting as in equation (A3) with $\Pi_{\widetilde{\mathbf{X}}} = \widetilde{\mathbf{X}}\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-}\widetilde{\mathbf{X}}'$ replacing $\Pi_{\mathbf{Z}}$, $\widetilde{\mathbf{Y}}$ replacing $\mathbf{Y}$ and $\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{B}$ replacing $\mathbf{E}$ giving

$$\left\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{B}\right\|^2 = \left\|\Pi_{\widetilde{\mathbf{X}}}\left(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{B}\right)\right\|^2 + \left\|\left(\mathbf{I}_n - \Pi_{\widetilde{\mathbf{X}}}\right)\widetilde{\mathbf{Y}}\right\|^2 = \left\|\Pi_{\widetilde{\mathbf{X}}}\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{B}\right\|^2 + \|\widehat{\mathbf{E}}\|^2 \tag{A6}$$

with $\widehat{\mathbf{E}}$ the residuals of the (full-rank) fit to equation (A1), which are independent of the differences in the preceding term. The reduced-rank estimate

of $\mathbf{B}$ is thus obtained by minimizing $\|\Pi_{\widetilde{\mathbf{X}}}\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{B}\|^2$. Because $\mathbf{B}$ is of reduced rank r, it can be factorized as the product of two matrices $\mathbf{C}^{[r]}$ and $\mathbf{P}^{[r]}$**:** $\mathbf{B} = \mathbf{C}^{[r]}\mathbf{P}^{[r]'}$, so that $\widetilde{\mathbf{X}}\mathbf{B} = \widetilde{\mathbf{X}}\mathbf{C}^{[r]}\mathbf{P}^{[r]'} = \mathbf{T}^{[r]}\mathbf{P}^{[r]'}$, with scores $\mathbf{T}^{[r]} = \widetilde{\mathbf{X}}\mathbf{C}^{[r]}$. We seek thus the minimum of $\left\|\Pi_{\widetilde{\mathbf{X}}}\widetilde{\mathbf{Y}} - \mathbf{T}^{[r]}\mathbf{P}^{[r]'}\right\|^2$. By the Eckhart-Young theorem, the minimum follows from the singular value decomposition [43] which, because $\Pi_{\widetilde{\mathbf{X}}}\widetilde{\mathbf{Y}}$ has zero column-means, is the same as a PCA [12] of the fitted values $\Pi_{\widetilde{\mathbf{X}}}\widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}}\widehat{\mathbf{B}}$, with $\widehat{\mathbf{B}}$ the least-squares estimate without rank restriction. Let $\mathbf{T}^{\mathrm{r}[r]}$ and $\mathbf{P}^{\mathrm{r}[r]}$ be the scores and loadings of the first r axes of this PCA. Then, because of the properties of PCA, $\mathbf{T}^{\mathrm{r}[r]} = \widetilde{\mathbf{X}}\widehat{\mathbf{B}}\mathbf{P}^{\mathrm{r}[r]}$, but we also have for any constrained score matrix $\mathbf{T}^{[r]} = \widetilde{\mathbf{X}}\mathbf{C}^{[r]}$, so that the canonical coefficients of RDA [7] are $\mathbf{C}^{\mathrm{r}[r]} = \widehat{\mathbf{B}}\,\mathbf{T}^{\mathrm{r}[r]}$ and the least-squares reduced rank estimate of $\mathbf{B}$ is $\mathbf{C}^{\mathrm{r}[r]}\mathbf{P}^{\mathrm{r}[r]'}$.

Remark

The derivation of RDA hinges on $\mathbf{A}$ being unconstrained. If $\mathbf{A}$ is constrained in some way, alternative methods are needed [27,46,52]. It is argued in the Discussion that these methods are of little practical interest.

Number of parameters

A rank r RDA has $(m\text{-}r)(p\text{-}r)$ parameters less than multivariate regression [43]. The number of parameters in RDA given in Ref. [7] did not account for the rotational freedom in the loadings and canonical weights.

A stable algorithm for RDA

Step 1. Perform a multivariate regression of $\mathbf{Y}$ onto covariate matrix $\mathbf{Z}$ and predictor matrix $\mathbf{X}$ (with intercept included in $\mathbf{Z}$). Retain the fitted values $\widehat{\mathbf{Y}}$ and residuals $\widehat{\mathbf{E}}$.

Step 2. Perform a multivariate regression of the fitted values $\widehat{\mathbf{Y}}$ from Step 1 on covariate matrix $\mathbf{Z}$**.** Retain the residuals and name them $\widehat{\mathbf{R}}$.

Step 3. Perform a PCA on the residuals $\widehat{\mathbf{R}}$ from Step 2. Retain the loadings $\mathbf{P}$ and scores $\mathbf{T}$, which are constrained scores.

Step 4. Calculate the unconstrained scores $\mathbf{T}^{\mathbf{E}} = \mathbf{T} + \widehat{\mathbf{E}}\mathbf{P}$.

Step 5. Perform a multivariate regression of $\mathbf{T}^{\mathbf{E}}$ onto $\mathbf{Z}$ and $\mathbf{X}$**.** Retain regression coefficients with their standard errors. The subset of coefficients corresponding to $\mathbf{X}$ are the canonical coefficients $\mathbf{C}^{\mathrm{r}}$.

Remarks

Step 1 gives $\widehat{\mathbf{Y}} = \mathbf{Z}\widehat{\mathbf{A}} + \mathbf{X}\widehat{\mathbf{B}}$ so that $\widehat{\mathbf{R}}$ in step 2 is equal to $(\mathbf{I}_n - \Pi_{\mathbf{Z}})(\mathbf{Z}\widehat{\mathbf{A}} + \mathbf{X}\widehat{\mathbf{B}}) = (\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{X}\widehat{\mathbf{B}} = \widetilde{\mathbf{X}}\widehat{\mathbf{B}}$. The potential advantage of Steps 1 and 2 over alternatives is that the steps do not involve regression coefficients, which can be numerically unstable. This algorithm is essentially the one used since decades in the rda function of vegan [8,19] except that $\mathbf{Y}$ is column-centered and $\mathbf{Z}$ and $\mathbf{X}$ are scaled to zero mean and unit variance in vegan for increased numerical stability of Step 1. With these preprocessing steps, the intercept does not need to be included, but the intercept is still needed in weighted RDA [17].

The standard errors of the canonical coefficients in Step 5 are an underestimate of the true standard errors, as they do not account for the uncertainty in the loadings $\mathbf{P}$.

The scores $\mathbf{T}$ are a function of both $\mathbf{Z}$ and $\mathbf{X}$. In an early stage of this work, we also defined alternative scores $\mathbf{X}\mathbf{C}^{\mathrm{r}}$ (with $\mathbf{X}$ column-centered) which are a function of $\mathbf{X}$ only, but we could not find substantive advantages over the usual constrained scores of RDA. See also Figure 1 and 3 that show this type of scores for ASCA+ and WE-ASCA in the top row.

For the computations in this paper, we took $\widehat{\mathbf{B}}$ from equation (A1) and calculated $(\mathbf{I}_n - \Pi_{\mathbf{Z}})\mathbf{X}\widehat{\mathbf{B}}$, so that the differences between RDA, ASCA+ and WE-ASCA cannot be explained by the use of a potentially more stable algorithm for RDA (Supplement S2).

A1.2. Dimension-reduction of the treatment effects in ASCA+ and RDA

The scores and loadings receive all attention in ASCA and its extensions. The scores contain the treatments effects, but there is little attention for the rank-reduction (dimension-reduction) of the treatment effects themselves. In this section we show that ASCA+ and RDA yield a weighted least-squares approximation to the treatment effects, whereby the weighting matrix differs between the methods. In RDA the weighting matrix is the inverse of the variance of the treatment effects (as is optimal), in the ASCA extensions it is slightly different.

In sections 2.2 and 2.3 RDA and the ASCA extensions are obtained via a PCA of a particular matrix, yielding loadings and scores. From this PCA an explicit decomposition of the treatment effects $\widehat{\mathbf{B}}$ can be derived as in ter Braak and Looman [20] for RDA. We start with ASCA.

ASCA+ and WE-ASCA perform a PCA on $\mathbf{X}\widehat{\mathbf{B}}$ which we column center if it is not already column-centered, giving the PCA decomposition

$$\mathbf{X}\widehat{\mathbf{B}} = \mathbf{T}^{\mathrm{a}}\mathbf{P}^{\mathrm{a}'} \text{ with } \mathbf{T}^{\mathrm{a}'}\mathbf{T}^{\mathrm{a}} = \mathbf{\Lambda}^{\mathrm{a}} \text{ and } \mathbf{P}^{\mathrm{a}'}\mathbf{P}^{\mathrm{a}} = \mathbf{I} \tag{A7}$$

with $\mathbf{\Lambda}^{\mathrm{a}}$ a diagonal matrix with the eigen values in decreasing order on the diagonal, $\mathbf{T}^{\mathrm{a}}$ the score matrix, $\mathbf{P}^{\mathrm{a}}$ the loading matrix and $\mathbf{I}$ the identity matrix, so that

$$(\mathbf{X}\widehat{\mathbf{B}})'\mathbf{X}\widehat{\mathbf{B}} = \widehat{\mathbf{B}}'\mathbf{X}'\mathbf{X}\widehat{\mathbf{B}} = \mathbf{P}^{\mathrm{a}}\,\mathbf{\Lambda}^{\mathrm{a}}\mathbf{P}^{\mathrm{a}'}. \tag{A8}$$

The loading matrix $\mathbf{P}^{\mathrm{a}}$ and the diagonal matrix with eigenvalues, $\mathbf{\Lambda}^{\mathrm{a}}$, also appear in the singular value decomposition

$$(\mathbf{X}'\mathbf{X})^{1/2}\widehat{\mathbf{B}} = \mathbf{R}^{\mathrm{a}}\,(\mathbf{\Lambda}^{\mathrm{a}})^{1/2}\mathbf{P}^{\mathrm{a}'} \text{with } \mathbf{R}^{\mathrm{a}'}\mathbf{R}^{\mathrm{a}} = \mathbf{I}. \tag{A9}$$

This can be checked by noting that equation (A8) follows from equation (A9) because $\mathbf{R}^{\mathrm{a}'}\mathbf{R}^{\mathrm{a}} = \mathbf{I}$. By equation (A9), ASCA+ provides the solution to the least-squares problem of minimizing

$$\left\|(\mathbf{X}'\mathbf{X})^{1/2}(\widehat{\mathbf{B}}-\mathbf{C}\mathbf{P}')\right\|^2 \tag{A10}$$

with respect to the matrices $\mathbf{C}$ and $\mathbf{P}$, both having r columns. Here $\|\mathbf{W}^{1/2}\mathbf{M}\|^2 = tr(\mathbf{M}'\mathbf{W}\mathbf{M}) = \sum_{ij} w_{ij} m_{ij}^2$ is the Frobenius norm in a notation (with $\mathbf{W} = \mathbf{X}'\mathbf{X}$ and $\mathbf{M} = \widehat{\mathbf{B}} - \mathbf{C}\mathbf{P}'$) that shows that this gives the weighted least-squares approximation to the matrix $\widehat{\mathbf{B}}$. The solution to equation (A10) gives the rank r weighted least-squares approximation to $\widehat{\mathbf{B}}$ using weight matrix $\mathbf{X}'\mathbf{X}$. By the Eckhart-Young theorem, the solution follows from the singular value decomposition in equation (A9), which gives the canonical coefficients $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{R}^{a[r]}(\Lambda^{a[r]})^{1/2}$ and loadings $\mathbf{P} = \mathbf{P}^{a[r]}$.

RDA performs a PCA on $\widetilde{\mathbf{X}}\widehat{\mathbf{B}}$. By a similar route as above for ASCA+, starting from the PCA decomposition $\widetilde{\mathbf{X}}\widehat{\mathbf{B}} = \mathbf{T}^{r}\mathbf{P}^{r'}$, it can be shown that RDA minimizes

$$\left\|\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{1/2}(\widehat{\mathbf{B}}-\mathbf{C}\mathbf{P}')\right\|^2, \tag{A11}$$

with the solution obtained by the singular value decomposition

$$\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{1/2}\widehat{\mathbf{B}} = \mathbf{R}^{r}\,(\mathbf{\Lambda}^{r})^{1/2}\,\mathbf{P}^{r'} \text{with } \mathbf{R}^{r'}\mathbf{R}^{r} = \mathbf{I}, \tag{A12}$$

so that, for RDA, the minimum is obtained with canonical coefficients $\mathbf{C} = (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1/2}\mathbf{R}^{r[r]}\,(\Lambda^{r[r]})^{1/2}$ and loadings $\mathbf{P} = \mathbf{P}^{r[r]}$.

In conclusion, ASCA+ and RDA both provide a weighted least-squares approximation to the matrix of regression coefficients $\widehat{\mathbf{B}}$. The weights used in RDA are proportional to the inverse of the variance of $\widehat{\mathbf{B}}$ [20] under the simplifying assumption that the error covariance matrix is $\sigma\mathbf{I}_m$, whereas as the weights used in ASCA+ and WE-ASCA are non-standard and sub-optimal.

These results can easily be generalized to the case where the error variances and covariances $[s_{ij}]$ are of a different form. Let $\mathbf{S} = [s_{ij}]$. Before $\mathbf{S}$ can be used in the reduced-rank approximation; $\mathbf{S}$ or its inverse may need to be regularized, particularly with many response variables ($m > n$) [30,59]. With $\mathbf{\Phi}$ the (regularized) inverse of the (regularized) error variance-covariance matrix, the optimal weighted least-squares approximation of the treatment effects is obtained from

$$\left\|\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{1/2}(\widehat{\mathbf{B}}-\mathbf{C}\mathbf{P}')\mathbf{\Phi}^{1/2}\right\|^2 \tag{A13}$$

with the solution obtained by the singular value decomposition

$$\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{1/2}\widehat{\mathbf{B}}\,\mathbf{\Phi}^{1/2} = \mathbf{R}^{rs}\,(\mathbf{\Lambda}^{rs})^{1/2}\,\mathbf{P}^{rs'} \text{with } \mathbf{R}^{rs'}\mathbf{R}^{rs} = \mathbf{I}, \tag{A14}$$

so that, for RDA, the minimum is obtained with canonical coefficients $\mathbf{C} = \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1/2}\mathbf{R}^{rs[r]}\left(\mathbf{\Lambda}^{rs[r]}\right)^{1/2}$ and loadings $\mathbf{P} = \mathbf{\Phi}^{-1/2}\mathbf{P}^{rs[r]}$. The superscript s is added to distinguish the matrices from those of Eqn A12.

A simple case of particularly interest (available in Canoco [18]) is that in which the variables are assumed independent given $\mathbf{X}$ and $\mathbf{Z}$ but have different variances s_{ii} so that $\mathbf{\Phi} = \mathrm{diag}(s_{11}^{-1}, \ldots, s_{mm}^{-1})$. This weights the response variables in dependence of how well they can be predicted without dimension reduction [28]. This simple case is obtained by starting from the PCA of $\widetilde{\mathbf{X}}\widehat{\mathbf{B}}\mathbf{\Phi}^{1/2}$ instead of from $\widetilde{\mathbf{X}}\widehat{\mathbf{B}}$ [20].[1]

Equation (A13) covers the scaling options in Timmerman et al. [26], as $\mathbf{\Phi}$ is free to choose.

A1.3. Summary of RDA

RDA (Section 2.2) can be viewed in three complementary ways [7,23], which are here summarized first in formulas and thereafter in words:

Start with a (partial) PCA, with the intercept included in the covariate matrix $\mathbf{Z}$:

$$\mathbf{Y} = \mathbf{Z}\mathbf{A}^{*} + \mathbf{T}\mathbf{P}' + \mathbf{E} \tag{A15}$$

with $\mathbf{T}$ and $\mathbf{P}$ with r columns each. In the least-squares solution for $\mathbf{T}$ and $\mathbf{P}$, we have $\mathbf{Z}'\mathbf{T} = \mathbf{0}$. By requiring that $\mathbf{T} = \widetilde{\mathbf{X}}\mathbf{C}$ we obtain the constrained form of PCA, *i.e.* RDA

$$\mathbf{Y} = \mathbf{Z}\mathbf{A}^{*} + (\widetilde{\mathbf{X}}\mathbf{C})\mathbf{P}' + \mathbf{E} \tag{A16}$$

which can be rewritten to the reduced-rank form

$$\mathbf{Y} = \mathbf{Z}\mathbf{A}^{*} + \widetilde{\mathbf{X}}(\mathbf{C}\mathbf{P}') + \mathbf{E} \tag{A17}$$

or, with $\mathbf{B} = \mathbf{C}\mathbf{P}'$ (also $\mathbf{C}$ has necessarily r columns),

$$\mathbf{Y} = \mathbf{Z}\mathbf{A}^{*} + \widetilde{\mathbf{X}}\mathbf{B} + \mathbf{E}. \tag{A18}$$

[1] Note that ter Braak and Looman [19] used a different notation, in which $\mathbf{\Gamma} = \mathbf{\Phi}^{1/2}$.

With the constraints on **T**, we have

$$(\mathbf{I}_n - \Pi_{\mathbf{Z}})\hat{\mathbf{Y}} = \mathbf{TP}'. \tag{A19}$$

In words:

Ad (A16):It is a partial PCA in which the components (axes) are constrained, compared to (A15), to be linear combinations of all predictor variables and to be uncorrelated to any covariates.

Ad (A17): It is multivariate regression with, compared to (A18), a restriction on the rank of the matrix of regression coefficients [27,47]. With covariates, the restriction applies to a subset of this matrix. The technical term "reduced rank" simply means that the matrix is a product of two matrices with, for rank 2, two columns each, just as in PCA. The difference is that PCA decomposes the original response data, whereas RDA decompose the fitted values and the matrix of regression coefficients.

Ad (A19): RDA is a form of PCA applied to the fitted values of the regression [47] after adjustment for any covariates. This last view follows from the previous as shown in section 2.3 of the main text, and in more detail in Appendix 1.1.

References

[1] P. Geladi, Notes on the history and nature of partial least squares (PLS) modelling, J. Chemometr. 2 (1988) 231–246, https://doi.org/10.1002/cem.1180020403.

[2] H. Chun, S. Keleş, Sparse partial least squares regression for simultaneous dimension reduction and variable selection, J. Roy. Stat. Soc. B 72 (2010) 3–25. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2009.00723.x.

[3] J.J. Jansen, H.C.J. Hoefsloot, J. van der Greef, M.E. Timmerman, J.A. Westerhuis, A.K. Smilde, ASCA: analysis of multivariate data obtained from an experimental design, J. Chemometr. 19 (2005) 469–481, https://doi.org/10.1002/cem.952.

[4] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.-J.A.N. Lamers, J. van der Greef, M. E. Timmerman, ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data, Bioinformatics 21 (2005) 3043–3048, https://doi.org/10.1093/bioinformatics/bti476.

[5] A.K. Smilde, H.C.J. Hoefsloot, J.A. Westerhuis, The geometry of ASCA, J. Chemometr. 22 (2008) 464–471, https://doi.org/10.1002/cem.1175.

[6] M. Thiel, B. Féraud, B. Govaerts, ASCA+ and APCA+: extensions of ASCA and APCA in the analysis of unbalanced multifactorial designs, J. Chemometr. 31 (2017) e2895, https://doi.org/10.1002/cem.2895.

[7] C.J.F. ter Braak, I.C. Prentice, A theory of gradient analysis, Adv. Ecol. Res. 18 (1988) 271–317, https://doi.org/10.1016/S0065-2504(08)60183-X.

[8] L. Legendre, P. Legendre, Numerical Ecology, Elsevier, Amsterdam, 2012.

[9] F. Berendse, R.H.E.M. Geerts, W.T. Elberse, T.M. Bezemer, P.W. Goedhart, W. Xue, E. Noordijk, C.J.F. ter Braak, H. Korevaar, A matter of time: recovery of plant species diversity in wild plant communities at declining nitrogen deposition, Divers. Distrib. 27 (2021) 1180–1193, https://doi.org/10.1111/ddi.13266.

[10] J. Baar, C.J.F. ter Braak, Ectomycorrhizal sporocarp occurrence as affected by manipulation of litter and humus layers in Scots pine stands of different age, Appl. Soil Ecol. 4 (1996) 61–73, https://doi.org/10.1016/0929-1393(96)00097-2.

[11] H. Wold, Soft modeling: the basic design and some extensions, in: K.G. Joreskog, H. Wold (Eds.), Systems under Indirect Observations II, North-Holland, Amsterdam, 1982, pp. 1–54.

[12] M. Greenacre, P.J.F. Groenen, T. Hastie, A.I. D'Enza, A. Markos, E. Tuzhilina, Principal component analysis, Nature Reviews Methods Primers 2 (2022) 100, https://doi.org/10.1038/s43586-022-00184-w.

[13] C.J.F. ter Braak, S. de Jong, The objective function of partial least squares regression, J. Chemometr. 12 (1998) 41–54, doi:10.1002/(SICI)1099-128X (199801/02)12:1<41::AID-CEM500>3.0.CO;2-F.

[14] P.R. Peres-Neto, P. Legendre, S. Dray, D. Borcard, Variation partitioning of species data matrices: estimation and comparison of fractions, Ecology 87 (2006) 2614–2625, https://doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2.

[15] M.J. Anderson, C.J.F. ter Braak, Permutation tests for multi-factorial analysis of variance, J. Stat. Comput. Simulat. 73 (2003) 85–113, https://doi.org/10.1080/00949650215733.

[16] C.J.F. ter Braak, D.E. te Beest, Testing environmental effects on taxonomic composition with canonical correspondence analysis: alternative permutation tests are not equal, Environ. Ecol. Stat. 29 (2022) 849–868, https://doi.org/10.1007/s10651-022-00545-4.

[17] C.J.F. ter Braak, Predictor versus response permutation for significance testing in weighted regression and redundancy analysis, J. Stat. Comput. Simulat. 92 (2022) 2041–2059, https://doi.org/10.1080/00949655.2021.2019256.

[18] C.J.F. ter Braak, P. Šmilauer, Canoco Reference Manual and User's Guide: Software for Ordination (Version 5.10), Microcomputer Power, Ithaca, USA, 2018.

[19] J. Oksanen, G.L. Simpson, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, P. Solymos, M.H.H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M. De Caceres, S. Dur, H. B.A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M.O. Hill , L. Lahti, D. McGlinn, M.-H. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C.J.F. ter Braak, J. Weedon, vegan: Community Ecology Package. R package version 2.6-4, http://CRAN.R-project.org/package=vegan, 2022.

[20] C.J.F. ter Braak, C.W.N. Looman, Biplots in reduced-rank regression, Biom. J. 36 (1994) 983–1003, https://doi.org/10.1002/bimj.4710360812.

[21] N. Ali, J. Jansen, A. van den Doel, G.H. Tinnevelt, T. Bocklitz, WE-ASCA: the weighted-effect ASCA for analyzing unbalanced multifactorial designs—a Raman spectra-based example, Molecules 26 (2021) 66. https://www.mdpi.com/1420-3049/26/1/66.

[22] P.J. van den Brink, C.J.F. ter Braak, Multivariate analysis of stress in experimental ecosystems by Principal Response Curves and similarity analysis, Aquat. Ecol. 32 (1998) 163–178, doi:10.1023/A:1009944004756.

[23] P.J. van den Brink, C.J.F. ter Braak, Principal Response Curves: analysis of time-dependent multivariate responses of a biological community to stress, Environ. Toxicol. Chem. 18 (1999) 138–148, https://doi.org/10.1002/etc.5620180207.

[24] P.J. van den Brink, P.J. den Besten, A. bij de Vaate, C.J.F. ter Braak, Principal response curves technique for the analysis of multivariate biomonitoring time series, Environ. Monit. Assess. 152 (2009) 271–281, doi:10.1007/s10661-008-0314-6.

[25] A.K. Smilde, M.E. Timmerman, M.M.W.B. Hendriks, J.J. Jansen, H.C.J. Hoefsloot, Generic framework for high-dimensional fixed-effects ANOVA, Briefings Bioinf. 13 (2012) 524–535, https://doi.org/10.1093/bib/bbr071.

[26] M.E. Timmerman, H.C.J. Hoefsloot, A.K. Smilde, E. Ceulemans, Scaling in ANOVA-simultaneous component analysis, Metabolomics 11 (2015) 1265–1276, https://doi.org/10.1007/s11306-015-0785-8.

[27] R.P. Velu, Reduced rank models with two sets of regressors, Journal of the Royal Statistical Society. Series C (Applied Statistics) 40 (1991) 159–170, https://doi.org/10.2307/2347914.

[28] C.J.F. ter Braak, Interpreting canonical correlation analysis through biplots of structural correlations and weights, Psychometrika 55 (1990) 519–531, https://doi.org/10.1007/BF02294765.

[29] C. Bertinetto, J. Engel, J. Jansen, ANOVA simultaneous component analysis: a tutorial review, Anal. Chim. Acta X 6 (2020), 100061, https://doi.org/10.1016/j.acax.2020.100061.

[30] J. Engel, L. Blanchet, B. Bloemen, L.P. van den Heuvel, U.H.F. Engelke, R. A. Wevers, L.M.C. Buydens, Regularized MANOVA (rMANOVA) in untargeted metabolomics, Anal. Chim. Acta 899 (2015) 1–12, https://doi.org/10.1016/j.aca.2015.06.042.

[31] R Core Team, R: A Language and Environment for Statistical Computing, version 4.1, R Foundation for Statistical Computing, Vienna, Austria, 2022, www.R-project.org.

[32] M. te Grotenhuis, B. Pelzer, R. Eisinga, R. Nieuwenhuis, A. Schmidt-Catran, R. Konig, A novel method for modelling interaction between categorical variables, Int. J. Publ. Health 62 (2017) 427–431, https://doi.org/10.1007/s00038-016-0902-0.

[33] R. Nieuwenhuis, H. te Grotenhuis, B. Pelzer, Weighted effect coding for observational data with wec, The R Journal 9 (2017) 477–485, https://doi.org/10.32614/RJ-2017-017.

[34] G.A. Milliken, D.E. Johnson, Analysis of Messy Data, second ed. (second ed., vol. 1, Designed Experiments, 2009. Chapman and Hall/CRC.

[35] S.R. Searle, Linear Models for Unbalanced Data, Wiley, New York, 1987.

[36] K.R. Gabriel, The biplot graphic display of matrices with application to principal component analysis, Biometrika 58 (1971) 453–467. https://www.jstor.org/stable/2334381.

[37] G. Zwanenburg, H.C.J. Hoefsloot, J.A. Westerhuis, J.J. Jansen, A.K. Smilde, ANOVA–principal component analysis and ANOVA–simultaneous component analysis: a comparison, J. Chemometr. 25 (2011) 561–567, https://doi.org/10.1002/cem.1400.

[38] A.L. van den Wollenberg, Redundancy analysis. An alternative for canonical correlation analysis, Psychometrika 42 (1977) 207–219, https://doi.org/10.1007/BF02294050.

[39] C.R. Rao, The use and interpretation of principal component analysis in applied research, Sankhya 26 (1964) 329–358. http://www.jstor.org/stable/25049339.

[40] D. Borcard, F. Gillet, P. Legendre, Numerical Ecology with R, Springer, New York, 2011.

[41] C.J.F. ter Braak, P. Šmilauer, Topics in constrained and unconstrained ordination, Plant Ecol. 216 (2015) 683–696, https://doi.org/10.1007/s11258-014-0356-5.

[42] K.R. Gabriel, Generalised bilinear regression, Biometrika 85 (1998) 689–700. http://biomet.oxfordjournals.org/content/85/3/689.abstract.

[43] T.W. Yee, Vector Generalized Linear and Additive Models with an Implementation in R, Springer, New York, 2015.

[44] B. van der Veen, F.K.C. Hui, K.A. Hovstad, R.B. O'Hara, Concurrent ordination: simultaneous unconstrained and constrained latent variable modelling, Methods Ecol. Evol. (2023), https://doi.org/10.1111/2041-210X.14035 n/a.
[45] S. Rabe-Hesketh, A. Skrondal, A. Pickles, Generalized multilevel structural equation modeling 69 (2004) 167–190, https://doi.org/10.1007/BF02295939.
[46] R. Velu, G.C. Reinsel, Multivariate Reduced-Rank Regression: Theory and Applications, Springer Science & Business Media, 2013.
[47] P.T. Davies, M.K.-S. Tso, Procedures for reduced-rank regression, Applied Statistics 31 (1982) 244–255, https://doi.org/10.2307/2347998.
[48] D.L. Jones, Fathom Toolbox for MATLAB: Software for Multivariate Ecological and Oceanographic Data Analysis, College of Marine Science, University of South Florida, St. Petersburg, FL, USA, 2017. https://www.usf.edu/marine-science/research/matlab-resources/fathom-toolbox-for-matlab.aspx.
[49] Ø. Langsrud, ANOVA for unbalanced data: use Type II instead of Type III sums of squares, Stat. Comput. 13 (2003) 163–167, https://doi.org/10.1023/A:1023260610025.
[50] R.H.G. Jongman, C.J.F. ter Braak, O.F.R. van Tongeren, Data Analysis in Community and Landscape Ecology, Cambridge University Press, Cambridge, 1995.
[51] R.P. Velu, G.C. Reinsel, D.W. Wichern, Reduced rank models for multiple time series, Biometrika 73 (1986) 105–118.
[52] Y. Takane, H.A.L. Kiers, J. de Leeuw, Component analysis with different sets of constraints on different dimensions, Psychometrika 60 (1995) 259–280, https://doi.org/10.1007/BF02301416.
[53] K.H. Liland, A. Smilde, F. Marini, T. Næs, Confidence ellipsoids for ASCA models based on multivariate regression theory, J. Chemometr. 32 (2018), e2990, https://doi.org/10.1002/cem.2990.
[54] N.J. Vendrig, L. Hemerik, C.J.F. ter Braak, Response variable selection in principal response curves using permutation testing, Aquat. Ecol. 51 (2017) 131–143, https://doi.org/10.1007/s10452-016-9604-1.
[55] R.C.M. Verdonschot, A.M. van Oosten-Siedlecka, C.J.F. ter Braak, P.F. M. Verdonschot, Macroinvertebrate survival during cessation of flow and streambed drying in a lowland stream, Freshw. Biol. 60 (2015) 282–296, https://doi.org/10.1111/fwb.12479.
[56] R.R. Sokal, F.J. Rohlf, Biometry, Freeman, New York, 1981.
[57] M. Martin, B. Govaerts, LiMM-PCA: combining ASCA+ and linear mixed models to analyse high-dimensional designed data, J. Chemometr. 34 (2020), https://doi.org/10.1002/cem.3232 e3232.
[58] A.H. Jarmund, T.S. Madssen, G.F. Giskeødegård, ALASCA: an R package for longitudinal and cross-sectional analysis of multivariate data by ASCA-based methods, Front. Mol. Biosci. 9 (2022), https://doi.org/10.3389/fmolb.2022.962431.
[59] J. Engel, L. Buydens, L. Blanchet, An overview of large-dimensional covariance and precision matrix estimators with applications in chemometrics, J. Chemometr. 31 (2017), e2880, https://doi.org/10.1002/cem.2880.
[60] L.B. Merga, P.J. Van den Brink, Ecological effects of imidacloprid on a tropical freshwater ecosystem and subsequent recovery dynamics, Sci. Total Environ. 784 (2021), 147167, https://doi.org/10.1016/j.scitotenv.2021.147167.
[61] J. Niku, F.K.C. Hui, S. Taskinen, D.I. Warton, gllvm: fast analysis of multivariate abundance data with generalized linear latent variable models in r, Methods Ecol. Evol. 10 (2019) 2173–2182, https://doi.org/10.1111/2041-210X.13303.
[62] K. Chen, W. Wang, Rrpack: Reduced-Rank Regression, 2022 version 0.1-13, https://CRAN.R-project.org/package=rrpack.
[63] G.L. Simpson, Permute: functions for generating restricted permutations of data, R package version 0.9-5, https://CRAN.R-project.org/package=permute, 2019.
[64] C.J.F. ter Braak, Update notes: CANOCO version 3.1, Agricultural Mathematics Group. http://edepot.wur.nl/250652, 1990. Wageningen.
[65] A.C. Davison, D.V. Hinkley, Bootstrap Methods and Their Applications, Cambridge University Press, Cambridge, 1997.
[66] M.J. Anderson, J. Robinson, Permutation tests for linear models, Aust. N. Z. J. Stat. 43 (2001) 75–88, https://doi.org/10.1111/1467-842X.00156.
[67] A.M. Winkler, G.R. Ridgway, M.A. Webster, S.M. Smith, T.E. Nichols, Permutation inference for the general linear model, Neuroimage 92 (2014) 381–397, https://doi.org/10.1016/j.neuroimage.2014.01.060.
[68] M. Stapel, C.J.F. ter Braak, Randomization and Bootstrap Tests in Factorial Experiments: Does Analysis Follow from Design? Dutch-German Biometrics Meeting, Münster, 1994. https://edepot.wur.nl/584838.
[69] P. Legendre, J. Oksanen, C.J.F. ter Braak, Testing the significance of canonical axes in redundancy analysis, Methods Ecol. Evol. 2 (2011) 269–277, https://doi.org/10.1111/j.2041-210X.2010.00078.x.
[70] M.E. Timmerman, C.J.F. ter Braak, Bootstrap confidence intervals for principal response curves, Comput. Stat. Data Anal. 52 (2008) 1837–1849, https://doi.org/10.1016/j.csda.2007.05.032.
[71] P.H. Eilers, B.D. Marx, Practical Smoothing: the Joys of P-Splines, Cambridge University Press, 2021.