

Algorithms and biplots for double constrained correspondence analysis

Cajo J. F. ter Braak¹  · Petr Šmilauer²  ·
Stéphane Dray³ 

Received: 10 July 2017 / Revised: 13 November 2017 / Published online: 16 January 2018
© The Author(s) 2018. This article is an open access publication

Abstract Correspondence analysis with linear external constraints on both the rows and the columns has been mentioned in the ecological literature, but lacks full mathematical treatment and easily available algorithms and software. This paper fills this gap by defining the method as maximizing the fourth-corner correlation between linear combinations, by providing novel algorithms, which demonstrate relationships with related methods, and by making a detailed study of possible biplots and associated approximations. The method is illustrated using ecological data on the abundances of species in sites and where the species are characterized by traits and sites by environ-

Handling Editor: Bryan F. J. Manly.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10651-017-0395-x>) contains supplementary material, which is available to authorized users.

✉ Cajo J. F. ter Braak
Cajo.terBraak@wur.nl
<https://www.vcard.wur.nl/Views/Profile/View.aspx?id=138&ln=eng>

Petr Šmilauer
petrsm@jcu.cz
<http://www.canoco5.com>

Stéphane Dray
stephane.dray@univ-lyon1.fr
<http://pbil.univ-lyon1.fr/members/dray/>

- ¹ Biometris, Wageningen University and Research, Box 16, 6700 AA Wageningen, The Netherlands
- ² Faculty of Science, University of South Bohemia, Branišovská 1760, České Budějovice, Czech Republic
- ³ Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Claude Bernard Lyon I, Villeurbanne, France

mental variables. The trait data and environment data form the external constraints and the question is which traits and environmental variables are associated, how these associations drive species abundances and how they can be displayed in biplots. With microbiome data becoming widely available, these and related multivariate methods deserve more study as they might be routinely used in the future.

Keywords Biplot · Canonical correlation analysis · Canonical correspondence analysis · Community ecology · Fourth-corner correlation · Multivariate analysis · Trait-environment relations

1 Introduction

Double constrained correspondence analysis (dc-CA) was developed by Jean-Dominique Lebreton, Robert Sabatier and co-workers as a natural extension of canonical correspondence analysis (Lebreton et al. 1988a, b; ter Braak 1986, 1987) and applied in studies relating species attributes (traits) and environmental variables via the central sites-by-species abundance table (Lavelle et al. 1999, 1998). Canonical correspondence analysis (CCA) constrains the row scores of a correspondence analysis (CA) of the central table by linear combinations of environmental variables. In addition to a single constraint on the rows, dc-CA uses also a constraint on the columns. In applications, the column (species) scores are constrained by linear combinations of species traits. Lavelle et al. (1999) nicknamed the method therefore double CCA and Kleyer et al. (2012) also use this name. Here the method is abbreviated to dc-CA. Its inputs are three data tables: the central site-by-species table **Y** containing the abundances (≥ 0) of the species in the sites (for short the community table), the site-by-environment table **E**, and the species-by-traits table **T**. To give another application domain, the central table may contain preference scores of persons on different products and the tables **E** and **T** person and product characteristics, respectively.

Curiously, dc-CA has never been explicitly defined mathematically in a scientific paper, presumably because its mathematics was crystal clear to its inventors. An exception is perhaps Böckenholt and Böckenholt (1990) but they defined the constraints in a different way, namely by the null space method (Takane 2013). As some applications of this method did appear, its novelty was gone and no thorough statistical description of the method was published. For Takane (2013), dc-CA is only a special case of constrained principal component analysis, but there is more to say than that, as Hill (1974) said for CA or Tenenhaus and Young (1985) for multiple CA. Meanwhile, Kleyer et al. (2012) evaluated several methods for analysing trait-environment relationships, including dc-CA, and did not detect any advantage of dc-CA over rival methods such as RLQ analysis (Dolédec et al. 1996), another three-table ordination method, and redundancy analysis on community weighted means (CWM-RDA) that consists in combining tables **T** and **Y** to build a new table of community weighted trait means (CWMs) that is related to **E**, in a second step, by a two-table method. All these methods are closely linked but differ notably in whether they take account of correlations among traits and/or among environmental variables by built-in multiple regressions. RLQ is based on PLS-like regressions and thus does not take account of

any of these correlations, CWM-RDA takes account of the correlations among environmental variables, and dc-CA takes account of both. As a consequence, RLQ is more robust and can analyse any number of traits and environmental variables, whereas in CWM-RDA the number of environmental variables must not be large compared to the number of sites, and in dc-CA also the number of trait variables must not be large compared to the number of species. But provided care has been taken of these limitations, for example by prior dimension reduction (Lavorel et al. 1999) or variable selection (ter Braak and Verdonschot 1995), dc-CA can reveal trait and environment associations that cannot be identified in RLQ as we demonstrate in Sect. 4.

Reasons for the renewed interest in dc-CA are its relation with the fourth-corner approach (ter Braak 2017), the fact that its inertia is a Rao score test statistic for testing trait-environment association in a log-linear model (ter Braak 2017), and also the desirability of regression-based methods for trait-environment relations (Peres-Neto et al. 2017; Warton et al. 2015) where abundances or community weighted means are considered as the response variables. In this paper the focus is on algorithms for dc-CA. The first one is based on a singular value decomposition (SVD), the second one is an iteration algorithm based on the transition formulae, and the third is based on a combination of CCA and redundancy analysis (RDA) which are widely available together with associated methods for statistical testing and selection of variables (Dray and Dufour 2007; Oksanen et al. 2013; ter Braak and Šmilauer 2012).

These novel algorithms come in addition to three existing algorithms. The first is based on the observation of Lavorel et al. (1999) that dc-CA can be seen as a canonical correlation analysis of traits and environment weighted by the central table. The input data would be inflated trait and environment data tables as exemplified in Dray and Legendre (2008); see also ter Braak (2017). Although canonical correlation analysis is widely available in statistical packages, a weighted version does not exist to our knowledge; the unweighted version could be used only for integer abundance data after ‘super’ inflation of the data so that every individual has a separate row.

In practice, Lavorel et al. (1999) used another algorithm developed by Bacou et al. (1989). This second algorithm consisted of two CCA’s: a traditional CCA of the abundance table with respect to the environment data from which a table of fitted abundances is computed by the reconstitution formula known from CA, and then another CCA of the transposed fitted table with respect to the trait data. A problem with this approach is that the fitted table may contain negative values and most programs for CCA do not allow negative values in the abundance table. Lavorel et al. (1999) do not mention this problem or how they dealt with it. Presumably their software allowed negative abundance values in a CCA.

The third algorithm is in the Appendices S2 and S4 of Kleyer et al. (2012). It starts from the results of a correspondence analysis of the central table, in line with how a canonical correspondence analysis is computed in the R package *ade4* (Dray and Dufour 2007; R Core Team 2015). In this algorithm, the CA is foremost used to create row and column weights and a transformed abundance table that is projected on the environmental variables. The result is projected onto the traits, the result of which is finally rotated to principal axes. The projections are obtained by weighted regression. This approach, implemented in R function *doublerda* in Kleyer et al. (2012), is generic in that it can also be used for a double constrained principal component analysis (dc-

PCA) when starting from a principal components analysis of the central table. See Takane (2013) for a related generic framework. The second and third algorithms share the feature that they work on fitted site-by-species tables.

Our new algorithms do not create such tables. The iteration algorithm works directly on the original data tables \mathbf{Y} , \mathbf{E} and \mathbf{T} , the SVD on the matrix product of \mathbf{Y} and orthogonalized \mathbf{E} and \mathbf{T} (say \mathbf{E}^* and \mathbf{T}^*) and the combination of CCA and RDA can be seen as a variant of CWM-RDA in which the community-weighted trait means (CWMs), usually obtained by combining \mathbf{Y} and \mathbf{T} , are computed with respect to \mathbf{T}^* instead of to \mathbf{T} . Reversely, we also described an alternative SNC-RDA in which the species-niche centroids (SNCs), usually computed by combining \mathbf{Y} and \mathbf{E} (see Peres-Neto et al. 2017 for more details), are computed with respect to \mathbf{E}^* instead of to \mathbf{E} . The CWMs and SNCs are of smaller dimensions than the original sites-by-species table \mathbf{Y} , when the number of environmental variables is smaller than the number of sites and the number of traits is smaller than the number of species, respectively.

The paper is structured as follows. Section 2 defines dc-CA, presents the new algorithms, and defines some additional quantities that are useful for summarizing and plotting the result in biplots. Section 3 discusses the available biplots in detail and Sect. 4 presents a data example and simulated example comparing dc-CA with RLQ. Most derivations are given in the Appendices; all equations are illustrated on real data in an R-script in Supplementary “Appendix S1”.

2 Theory and method

2.1 Notation

The three data tables (Environment, Community table and Trait data), denoted here by \mathbf{E} , \mathbf{Y} and \mathbf{T} , respectively Dolédec et al. (1996) used \mathbf{R} , \mathbf{L} and \mathbf{Q} ; ter Braak (1986) used \mathbf{Z} for \mathbf{E}), can be arranged as

$$\begin{bmatrix} \mathbf{Y}^T & \mathbf{T} \\ \mathbf{E}^T & \mathbf{F} \end{bmatrix}$$

where \mathbf{F} is the missing matrix (i.e., fourth-corner table), representing relations between the environment and species traits, that are of interest and should be estimated. With n sites, m species, p environmental variables and q traits, the dimensions of tables \mathbf{Y} , \mathbf{E} and \mathbf{T} are $n \times m$, $n \times p$ and $m \times q$, respectively, so that \mathbf{F} is $p \times q$. In the original method proposed by Legendre et al. (1997) the link table \mathbf{Y} contained presence-absence of the m species in the n sites but this was later generalized to abundance or count data by Dray and Legendre (2008). The community table is denoted here by \mathbf{Y} as it will be treated as a response matrix in CCA and dc-CA.

In the paper, bold lower case is used for column vectors, e.g. \mathbf{x} is a vector with elements $\{x_i\}$, $i = 1, \dots, n$. Bold upper case is used for matrices. Elements of \mathbf{Y} will be denoted by y_{ij} ; subscript i denotes a site (row of \mathbf{Y} or \mathbf{E}) and subscript j denotes a species (column of \mathbf{Y} or row of \mathbf{T}). A symbol “+” replacing an index means the sum over the index, e.g. $y_{i+} = \sum_j y_{ij}$. Further, \mathbf{R} and \mathbf{K} are diagonal matrices with the weights $\{y_{i+}\}$ and $\{y_{+j}\}$, respectively, on their diagonal.

2.2 Definition and corresponding equations

There are several distinct derivations of correspondence analysis and therefore also of CCA and dc-CA, all leading to equivalent eigen-equations. Two model-based derivations are presented in ‘‘Appendix A1’’. But, perhaps it is most natural to define dc-CA in the same spirit as in canonical correlation analysis, namely as maximizing a correlation. In dc-CA, this is the fourth-corner correlation f between trait \mathbf{t} and environmental variable \mathbf{e} (Legendre and Legendre 2012; Legendre et al. 1997; Peres-Neto et al. 2017), defined as

$$f = \text{cor}_{\mathbf{Y}}^2(\mathbf{t}, \mathbf{e}) = \frac{\sum_{i,j} y_{ij} \tilde{t}_j \tilde{e}_i}{\left\{ \sum_j y_{+j} \tilde{t}_j^2 \sum_i y_{i+} \tilde{e}_i^2 \right\}^{1/2}} \tag{1}$$

with \tilde{t}_j and \tilde{e}_i being centred versions of the trait and environmental variable, with weighted-means computed using weights derived from the abundance table \mathbf{Y} :

$$\tilde{t}_j = t_j - \sum_j y_{+j} t_j / y_{++} \text{ and } \tilde{e}_i = e_i - \sum_i y_{i+} e_i / y_{++} \tag{2}$$

The first axis of dc-CA can now be defined as follows.

Definition dc-CA is a method that finds linear combinations of the traits and of the environmental variables such that the fourth corner correlation between these linear combinations is maximized

The definition leads to an eigen-equation of which the first (non-trivial) eigen-vector is the solution. Later axes are then subsequent eigen-vectors which have the interpretation that they also maximize the fourth corner correlation but subject to the additional constraint of their orthogonality (in a particular metric) to the previous axes. The definition is in line with CA as a method of optimal scaling of the row and column categories of a contingency table (Gifi 1990; Hill 1974).

For simplicity and without loss of generality, it is assumed from now onwards that the traits and environmental variables are centred as in Eq. (2) so that $\mathbf{1}_n^T \mathbf{R} \mathbf{E} = \mathbf{0}_p$ and $\mathbf{1}_m^T \mathbf{K} \mathbf{T} = \mathbf{0}_q$. This centring eliminates the need for the intercept terms and avoids the trivial axis of (dc-)CA. The definition of dc-CA above leads to the following maximization problem:

$$\max_{\mathbf{b}, \mathbf{c}} \mathbf{x}^T \mathbf{Y} \mathbf{u} \text{ with } \mathbf{x} = \mathbf{E} \mathbf{b}, \mathbf{u} = \mathbf{T} \mathbf{c}, \mathbf{x}^T \mathbf{R} \mathbf{x} = 1 \text{ and } \mathbf{u}^T \mathbf{K} \mathbf{u} = 1. \tag{3}$$

This definition lacks explicit centring of \mathbf{x} (site scores) and \mathbf{u} (species scores), but the linear combinations $\mathbf{x} = \mathbf{E} \mathbf{b}$ and $\mathbf{u} = \mathbf{T} \mathbf{c}$ are centred, because \mathbf{E} and \mathbf{T} are centred. Insertion of $\mathbf{x} = \mathbf{E} \mathbf{b}$ and $\mathbf{u} = \mathbf{T} \mathbf{c}$ in Eq. (3) leads to

$$\max_{\mathbf{b}, \mathbf{c}} \mathbf{b}^T \mathbf{E}^T \mathbf{Y} \mathbf{T} \mathbf{c} \text{ subject to } \mathbf{b}^T \mathbf{E}^T \mathbf{R} \mathbf{E} \mathbf{b} = 1 \text{ and } \mathbf{c}^T \mathbf{T}^T \mathbf{K} \mathbf{T} \mathbf{c} = 1. \tag{4}$$

Applying the Lagrange multiplier method (Magnus and Neudecker 1988) leads to the necessary conditions for the maximizer of Eq. (4):

$$\mathbf{E}^T \mathbf{Y} \mathbf{T} \mathbf{c} - \lambda_b \mathbf{E}^T \mathbf{R} \mathbf{E} \mathbf{b} = \mathbf{0}_p \text{ and } \mathbf{T}^T \mathbf{Y}^T \mathbf{E} \mathbf{b} - \lambda_c \mathbf{T}^T \mathbf{K} \mathbf{T} \mathbf{c} = \mathbf{0}_q, \tag{5}$$

where λ_b and λ_c are Lagrange multipliers, one for each constraint. Re-expression gives the transition formulae of dc-CA between \mathbf{b} and \mathbf{c} :

$$\lambda_b \mathbf{b} = (\mathbf{E}^T \mathbf{R} \mathbf{E})^{-1} \mathbf{E}^T \mathbf{Y} \mathbf{T} \mathbf{c} \tag{6}$$

$$\lambda_c \mathbf{c} = (\mathbf{T}^T \mathbf{K} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y}^T \mathbf{E} \mathbf{b} \tag{7}$$

Insertion of Eq. (7) in Eq. (6) gives the eigen-equation of dc-CA for \mathbf{b}

$$\lambda \mathbf{b} = (\mathbf{E}^T \mathbf{R} \mathbf{E})^{-1} \mathbf{E}^T \mathbf{Y} \mathbf{T} (\mathbf{T}^T \mathbf{K} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y}^T \mathbf{E} \mathbf{b}, \tag{8}$$

where $\lambda = \lambda_b \lambda_c$ is the first eigenvalue of dc-CA and is equal to the square of the maximized fourth-corner correlation between the linear combinations of traits and environmental variables. Equation (8) shows a great similarity with the eigen-equation of canonical correlation analysis (Gittins 1985; Mardia et al. 1980), therewith confirming the observation of Lavorel et al. (1999) that dc-CA is a weighted canonical correlation analysis of traits and environment with the central table acting as weights.

For viewing the relationship of dc-CA with CA and CCA, it is instructive to rewrite the transition formulae (6) and (7) so that these include site and species scores as in ter Braak (1986) for CCA. There are two sets of each: scores that are linear combinations (LC scores) and scores that are weighted averages (WA scores). The latter are indicated with an asterisk in the following **transition formulae of dc-CA**. A redundant constant α ($0 \leq \alpha \leq 1$) is inserted so as to allow various scalings of ordination diagrams based on scores from these transition formulae for different dimensions, as in CA (Greenacre 1984). Example scalings are row-metric preserving ($\alpha = 1$), symmetric ($\alpha = 1/2$) or column-metric preserving ($\alpha = 0$). In Eqs. (3) and (4) the LC site and species scores were \mathbf{R} - and \mathbf{K} -normalized so as to define a correlation, but for a correlation the scaling does not matter and from now onwards, we set $\mathbf{x}^T \mathbf{R} \mathbf{x} = \lambda^\alpha$ and the scaling of all other scores then follows from the transition formulae (for example, $\mathbf{u}^T \mathbf{K} \mathbf{u} = \lambda^{1-\alpha}$, see next section):

$$\lambda^\alpha u_k^* = \sum_i y_{ik} x_i / y_{+k} \text{ or in matrix notation, } \lambda^\alpha \mathbf{u}^* = \mathbf{K}^{-1} \mathbf{Y}^T \mathbf{x} \tag{9}$$

$$\mathbf{c} = (\mathbf{T}^T \mathbf{K} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{K} \mathbf{u}^* \tag{10}$$

$$\mathbf{u} = \mathbf{T} \mathbf{c} \tag{11}$$

$$\lambda^{1-\alpha} x_i^* = \sum_k y_{ik} u_k / y_{i+} \text{ or in matrix notation, } \lambda^{1-\alpha} \mathbf{x}^* = \mathbf{R}^{-1} \mathbf{Y} \mathbf{u} \tag{12}$$

$$\mathbf{b} = (\mathbf{E}^T \mathbf{R} \mathbf{E})^{-1} \mathbf{E}^T \mathbf{R} \mathbf{x}^* \tag{13}$$

$$\mathbf{x} = \mathbf{E} \mathbf{b}. \tag{14}$$

In words, the transition formulae read as follows. Starting with Eq. (9):

- The WA species scores (u_k^*) are (proportional to) weighted averages of LC site scores (i.e. the constrained site scores, \mathbf{x}).

- Canonical weights for the traits (**c**) are coefficients of the regression of the WA species scores (u_k^*) on the traits (**T**) with the species total abundances (y_{+k}) as weights.
- LC species scores (i.e. the constrained species scores, **u**) are a linear combination of the traits.
- WA site scores (x_i^*) are (proportional to) weighted averages of LC species scores (**u**).
- Canonical weights for the environmental variables (**b**) are coefficients of the regression of the WA site scores (x_i^*) on the environmental variables (**E**) with the site total abundances (y_{i+}) as weights.
- LC site scores (constrained site scores, **x**) are a linear combination of the environmental variables (**E**).

There are two extra equations for dc-CA compared to CCA [Eqs. (10) and (11)] which define the constraints on the species scores based on trait values. In CCA the only species scores are WA scores, and therefore these are used in Eq. (12) for the WA site scores. In dc-CA the LC species scores are used to define the WA site scores [Eq. (12)] and reversely the LC site scores are used to define the WA species scores [Eq. (9)]. In CA, there are only WA site scores and the transition formulae are reduced to Eqs. (9) and (12). Note that dc-CA has two kinds of canonical weights, one for environmental variables and one for traits. The transition formulae show that dc-CA is constrained reciprocal averaging (Hill 1973).

Special cases of dc-CA are of course CCA if $\mathbf{T} = \mathbf{I}_m$ or $q \geq m$ so that there are effectively no constraints and CA if also $\mathbf{E} = \mathbf{I}_n$ or $p \geq n$.

2.3 Algorithm based on a SVD

The eigenproblem posed in Eq. (8) could be solved by an asymmetric eigenproblem solver or it could be first transformed to a two-sided eigenvalue problem. However, it is more convenient to obtain the solution of dc-CA via a singular value decomposition (SVD) (Golub and Reinch 1970; Golub and Loan 1989). Let

$$\mathbf{D} = (\mathbf{E}^T \mathbf{R} \mathbf{E})^{-1/2} \mathbf{E}^T \mathbf{Y} \mathbf{T} (\mathbf{T}^T \mathbf{K} \mathbf{T})^{-1/2} \tag{15}$$

and the SVD of **D** be

$$\mathbf{D} = \mathbf{P} \mathbf{\Delta} \mathbf{Q}^T \tag{16}$$

with **P** and **Q** orthonormal matrices and **Δ** a diagonal matrix with singular values in decreasing order. Then the singular values are the maximized fourth corner correlations of the dc-CA axes and the columns of

$$\mathbf{B} = (\mathbf{E}^T \mathbf{R} \mathbf{E})^{-1/2} \mathbf{P} \mathbf{\Delta}^\alpha \text{ and } \mathbf{C} = (\mathbf{T}^T \mathbf{K} \mathbf{T})^{-1/2} \mathbf{Q} \mathbf{\Delta}^{\alpha-1} \tag{17}$$

satisfy the transition formulae (6) and (7). This can be seen as follows: from Eq. (16), $\mathbf{P} \mathbf{\Delta}^\alpha = \mathbf{D} \mathbf{Q} \mathbf{\Delta}^{\alpha-1}$; insert **P** and **Q** from Eq. (17) and **D** from Eq. (15); simplify

and rearrange to obtain Eq. (6). By consequence, dc-CA eigenvalues are the squared singular values.

This approach immediately solves for all dc-CA eigen-vectors. The LC site scores and LC species scores of all dimensions, $\mathbf{X} = \mathbf{E}\mathbf{B}$ and $\mathbf{U} = \mathbf{T}\mathbf{C}$, are \mathbf{R} - and \mathbf{K} -orthogonal and the scaling factor Δ^α in Eq. (17) ensures that $\mathbf{X}^T\mathbf{R}\mathbf{X} = \mathbf{\Lambda}^\alpha$ and $\mathbf{U}^T\mathbf{K}\mathbf{U} = \mathbf{\Lambda}^{1-\alpha}$, where $\mathbf{\Lambda} = \mathbf{\Delta}^2$, respectively, as in the transition formulae.

Note that $tr(\mathbf{D}^T\mathbf{D})$ is equal to the sum of all eigenvalues satisfying Eq. (8), also known as the total inertia of the dc-CA. ter Braak (2017) showed that $y_{++}tr(\mathbf{D}^T\mathbf{D})$ is the Rao score test statistic for testing the trait-environment interaction in a Poisson log-linear model with saturated main effects. The first eigenvalue is the one-dimensional replacement thereof that could be also useful as test statistic in permutation tests if the alternative hypothesis is likely to be one-dimensional.

2.4 Algorithm based on the transition formulae

If only the first or only a few eigen-vectors need to be calculated, iterative methods can be used that repeatedly cycle through the transition formulae starting from arbitrary, non-constant starting values for either \mathbf{b} or \mathbf{c} , when using Eqs. (6) and (7), and for either \mathbf{b} , \mathbf{c} , \mathbf{x}^* or \mathbf{u}^* , when using Eqs. (9)–(14). The latter is a constrained reciprocal averaging algorithm (Hill 1973). Such an algorithm for CCA is described in the ‘‘Appendix’’ of ter Braak and Prentice (1988) and the extension to dc-CA is trivial. This iterative algorithm is related to the well-known power algorithm for solving eigenproblems (Good 1969; Gourlay and Watson 1973). Power algorithms tend to be slow, but acceleration methods make them practical.

2.5 Algorithm based on combining CCA and a weighted RDA

This subsection presents an algorithm based on a combination of CCA and RDA. The algorithm also gives insight into the relation between dc-CA and CWM-RDA.

CWM-RDA starts with computing a table \mathbf{M} of community weighted mean trait values (CWM)

$$\mathbf{M} = \mathbf{R}^{-1}\mathbf{Y}\mathbf{T} \tag{18}$$

and then applies an RDA of \mathbf{M} with respect to the environmental data \mathbf{E} . This analysis is essentially an SVD of

$$\mathbf{D}_{\text{cwm-rda}} = (\mathbf{E}^T\mathbf{E})^{-1/2}\mathbf{E}^T\mathbf{M} = (\mathbf{E}^T\mathbf{E})^{-1/2}\mathbf{E}^T(\mathbf{R}^{-1}\mathbf{Y}\mathbf{T}). \tag{19}$$

In comparison with (15), this equation lacks the weighing of the sites with weights \mathbf{R} and lacks a term involving the covariances among the traits ($\mathbf{T}^T\mathbf{K}\mathbf{T}$). The latter can be included implicitly by transforming the traits, prior to analysis, to orthonormal ones by:

$$\mathbf{T}^* = \mathbf{T}(\mathbf{T}^T\mathbf{K}\mathbf{T})^{-1/2} \tag{20}$$

so that $\mathbf{T}^{*T} \mathbf{K} \mathbf{1}_m = \mathbf{0}_q$ and $\mathbf{T}^{*T} \mathbf{K} \mathbf{T}^* = \mathbf{I}_q$. By replacing the unweighted RDA by a weighted RDA with site weights \mathbf{R} , the equivalent SVD is

$$\mathbf{D}_{\text{cwm-wrda}} = \left(\mathbf{E}^T \mathbf{R} \mathbf{E}\right)^{-1/2} \mathbf{E}^T \mathbf{R} (\mathbf{R}^{-1} \mathbf{Y} \mathbf{T}^*) = \left(\mathbf{E}^T \mathbf{R} \mathbf{E}\right)^{-1/2} \mathbf{E}^T \mathbf{Y} \mathbf{T}^* = \mathbf{D}. \quad (21)$$

As a consequence, dc-CA can be obtained by the following steps:

1. Transforming the traits to \mathbf{K} -orthonormal ones using species weights \mathbf{K} ,
2. Computing community weighted means of the orthonormal traits \mathbf{T}^* and then
3. Applying a weighted RDA of these community weighted means with respect to the environmental data \mathbf{E} using site weights \mathbf{R} .

The difference between dc-CA and CWM-RDA is thus that dc-CA uses differential weights for sites and accounts for the covariances among the traits, whereas CWM-RDA uses equal weights for the sites and accounts for differences in variance of the traits only. As a side note, in Kleyer et al. (2012) and the Traits example in Canoco 5.0 (ter Braak and Šmilauer 2012), the RDA implicitly standardized the response data, which are the community weighted means. However, it makes much more sense to standardize the traits from which they are computed (McCune 2015) and to leave the obtained community weighted means as they are (i.e. **not** to standardize them in RDA). The reason is that the variances of community weighted means of standardized traits carry important information on the relative importance of the traits (Peres-Neto et al. 2017). For an important trait, the variance will be relatively high and for a trait that is unrelated to the abundance table \mathbf{Y} , the variance will be very low. This also applies to dc-CA where traits are made orthogonal and also normalized (orthonormalized) and the subsequent community weighted means are analyzed untransformed (non-standardized) by an RDA so as to obtain a dc-CA.

The above Steps 1 and 2 can be combined in a single CCA of the transposed abundance table with respect to the traits, i.e. $\text{CCA}(\mathbf{Y}^T \sim \mathbf{T})$. The row scores of this CCA are linear combinations of the traits which are \mathbf{K} -orthogonal and, depending on the scaling of the axes, also \mathbf{K} -normalized (the scaling is column-metric preserving) (ter Braak 1986, 2014). The column scores of this analysis (response variable scores, in this case representing rows of \mathbf{Y}) are weighted averages of the row scores and thus community weighted means of \mathbf{K} -orthonormal traits [cf. Eq. (9)]. This way of making the traits orthonormal has an advantage for trait data that are (near) singular: the CCA ranks the dimensions in order of their importance for \mathbf{Y}^T so that it is unlikely that an important dimension is dropped due to an unlucky cut-off in the decision for rank-deficiency.

The resulting algorithm is thus

1. Perform $\text{CCA}(\mathbf{Y}^T \sim \mathbf{T})$: a CCA of the transposed community table onto the traits
2. Obtain the column scores (\mathbf{M}^* , say) from this analysis in site-metric preserving scaling, an $n \times q^*$ table of scores with q^* the rank of the trait data, which are community weighted means of orthonormalized traits.
3. Perform a weighted $\text{RDA}_{\mathbf{R}}(\mathbf{M}^* \sim \mathbf{E})$: an RDA of \mathbf{M}^* on the environmental variables using row weights \mathbf{R} .

The canonical coefficients of this RDA are the dc-CA coefficients for the environmental variables [Eqs. (13) or (17)] and the resulting linear combinations of this RDA are the dc-CA linear combinations of environmental variables [Eq. (14)], both up to sign changes of axes. In “Appendix A2” it is shown that the unconstrained row scores of the RDA are identical to the WA site scores.

Because of the symmetry in dc-CA between rows and columns (sites and species) it is also possible to start with a CCA of \mathbf{Y} with respect to \mathbf{E} :

1. Perform CCA($\mathbf{Y} \sim \mathbf{E}$): a CCA of the community table on to the environmental variables
2. Obtain the column scores (\mathbf{S}^* , say) from this analysis in species-metric preserving scaling, an $m \times p^*$ table of scores with p^* the rank of the environmental data, which are species niche centroids (Peres-Neto et al. 2017) of orthonormalized environmental variables.
3. Perform weighted RDA $_{\mathbf{K}}$ ($\mathbf{S}^* \sim \mathbf{T}$): an RDA of \mathbf{S}^* on the traits using row weights \mathbf{K} .

The canonical coefficients of this RDA are the dc-CA coefficients for the traits [Eqs. (10) or (17)], the resulting linear combinations of this RDA are the dc-CA linear combinations of traits [Eq. (11)] and the unconstrained row scores of the RDA are identical to the scores in Eq. (9), which can be shown analogously to the proof in “Appendix A2”.

The computer program Canoco 5.10 (ter Braak and Šmilauer 2012) implements dc-CA via these combinations of CCA and weighted RDA where the first combination is used for selection and significance testing of environmental variables and the second for selection and significance testing of traits.

2.6 Derived scores

Most scores that are useful for interpretation or plotting in ordination diagrams are already defined and available from the transition formulae of Sect. 2.2. As in CCA and canonical correlation analysis (ter Braak 1990), there are a few more scores that are useful for interpretation and plotting:

- Intra-set correlations of the traits with the constrained dc-CA axes, $cor_{\mathbf{K}}(\mathbf{T}, \mathbf{U})$, and similarly for the environmental variables, $cor_{\mathbf{K}}(\mathbf{E}, \mathbf{X})$.
- Inter-set correlations of the traits with the WA dc-CA axes, $cor_{\mathbf{K}}(\mathbf{T}, \mathbf{U}^*)$, and similarly for the environmental variables, $cor_{\mathbf{K}}(\mathbf{E}, \mathbf{X}^*)$. These are inter-set correlations in the setting of CCA as, for example, \mathbf{X}^* is a linear combination of \mathbf{Y} and not of \mathbf{E} .
- Fourth-corner correlations of the traits with the constrained dc-CA axes, $cor_{\mathbf{Y}}(\mathbf{T}, \mathbf{X})$, and similarly for the environmental variables, $cor_{\mathbf{Y}}(\mathbf{E}, \mathbf{U})$. When dc-CA is interpreted as a canonical correlation analysis on super inflated data, \mathbf{Y} seemingly disappears and these fourth-corner correlations are then in fact the inter-set correlations of the canonical correlation analysis. Unless noted explicitly otherwise, inter-set correlations of dc-CA refer to their definition in the setting of CCA and RDA, that is, to $cor_{\mathbf{K}}(\mathbf{E}, \mathbf{X}^*)$ and $cor_{\mathbf{K}}(\mathbf{T}, \mathbf{U}^*)$.

- Centroids of scores for categories of nominal traits and environmental variables. Such centroids can be viewed as scores for ‘super species’ and ‘super sites’ as they average scores of species or of sites belonging to the same category. The centroids are all weighted averages using the weights $\{y_{i+}\}$ for sites and $\{y_{+k}\}$ for species.

3 Biplots

Biplots serve to visualize the main pattern in the analyzed data, by plotting the scores on, typically, the first two axes of the analysis, so that their inner product approximates a matrix, typically a data table or table with statistics such as correlation or regression coefficients (Gabriel 1982). With three kinds of items in the plot (sites, species, environmental variables) such plots are often called triplots, in which pairs of items have an inner product (biplot) interpretation. In dc-CA there is a fourth kind of items: traits. This section proposes quadriplots with all four kinds of items in which almost all pairs of items have a biplot interpretation.

As in CCA, RDA and canonical correlation analysis (ter Braak 1990) there is a choice to visualize in the biplots regression coefficients or correlations. Biplots visualizing regression coefficients are treated in “Appendix A5”. For example, a biplot of both sets of canonical weights (\mathbf{B} and \mathbf{C}) approximates the regression coefficients associated with the bilinear interaction between (i.e. products of) environmental variables and traits. The other biplots in “Appendix A5” essentially follow from considering dc-CA as a canonical correlation analysis on inflated trait and environment data. In this section, the focus is on biplots of fourth-corner correlations between traits and environmental variables.

For \mathbf{K} - and \mathbf{R} -normalized trait and environmental variables, the $p \times q$ matrix of fourth-corner correlations between environmental variables and traits is simply

$$\mathbf{F} = \mathbf{E}^T \mathbf{Y} \mathbf{T} \tag{22}$$

Following ter Braak (1990), a biplot of \mathbf{F} can be based on a “rank r weighted least-squares approximation” of the form $\mathbf{F} \approx \mathbf{B}_f \mathbf{C}_f^T$ with \mathbf{B}_f and \mathbf{C}_f matrices of order $p \times r$ and $q \times r$, respectively. For producing biplots, a convenient choice is r is 2 or 3. Whether such an r is adequate can be judged by permutation tests (see example section). With as weight matrices the inverses of $\mathbf{E}^T \mathbf{R} \mathbf{E}$ and $\mathbf{T}^T \mathbf{K} \mathbf{T}$, the approximation is invariant to linear transformations of \mathbf{E} and \mathbf{T} and can be obtained from dc-CA as follows. We seek the minimum over \mathbf{B}_f and \mathbf{C}_f of

$$\begin{aligned} & \left\| \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1/2} \left(\mathbf{F} - \mathbf{B}_f \mathbf{C}_f^T \right) \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-1/2} \right\|^2 \\ &= \left\| \mathbf{D} - \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1/2} \mathbf{B}_f \mathbf{C}_f^T \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-1/2} \right\|^2 \end{aligned} \tag{23}$$

As follows from the Eckhart–Young theorem (Greenacre 1984) the minimum is obtained from the singular value decomposition of \mathbf{D} . By consequence, the minimum of (23) is $\lambda_{r+1} + \dots + \lambda_{\min(p,q)}$ and is obtained by

$$\mathbf{B}_f = \left[\left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{1/2} \mathbf{P} \Delta^\alpha \right]_r \quad \text{and} \quad \mathbf{C}_f = \left[\left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{1/2} \mathbf{Q} \Delta^{1-\alpha} \right]_r. \quad (24)$$

For $\alpha = 1$, \mathbf{B}_f contains the fourth-corner correlations of the environmental variables with the axes \mathbf{U} , $cor_Y(\mathbf{E}, \mathbf{U})$, and \mathbf{C}_f the intra-set correlations for the traits, $cor_K(\mathbf{T}, \mathbf{U})$ (“Appendix A3”). For $\alpha = 0$, the role of traits and environmental variables is reversed (“Appendix A3”) so that \mathbf{C}_f contains the fourth-corner correlations of the traits with the axes \mathbf{X} , $cor_Y(\mathbf{T}, \mathbf{X})$, and \mathbf{B}_f the intra-set correlations for the environmental variables, $cor_R(\mathbf{E}, \mathbf{X})$. The coefficients \mathbf{B}_f and \mathbf{C}_f are termed biplot scores as in CCA (Oksanen et al. 2013; ter Braak and Šmilauer 2012); they vary with α unlike the intra, inter and fourth corner correlations. In the symmetric scaling ($\alpha = 0.5$), each biplot score is the geometric mean of its intra-set correlation and its fourth-corner correlation with the axis.

“Appendix A4” shows that when plotting \mathbf{B}_f , \mathbf{C}_f , \mathbf{X} and \mathbf{U} together the pairs \mathbf{B}_f – \mathbf{U} and \mathbf{C}_f – \mathbf{X} form a weighted least-squares biplot of the species niche centroids ($\mathbf{K}^{-1} \mathbf{Y}^T \mathbf{E}$) and CWMs ($\mathbf{R}^{-1} \mathbf{Y} \mathbf{T}$), respectively. Also, \mathbf{X} and \mathbf{U} form a weighted least-squares biplot of the fitted contingency ratios, which are the contingency ratios projected on both \mathbf{E} and \mathbf{T} , analogously to the situation in CCA (ter Braak 2014).

The above biplot options are least-squares for all values of α . For $\alpha = 1$, the plot is row-metric preserving, and thus approximates the chi-square distance between sites based on fitted values, and \mathbf{C}_f contains the intra-set correlations for the traits and the biplot thus weakly approximates the correlations among the traits. For $\alpha = 0$, the plot is column-metric preserving, and thus approximates the chi-square distance between species based on fitted values, and \mathbf{B}_f contains the intra-set correlations for the environmental variables and the biplot thus weakly approximates the correlations among the environmental variables. In conclusion, when plotting \mathbf{B}_f , \mathbf{C}_f , \mathbf{X} and \mathbf{U} together with $\alpha = 0$ or 1, five of the six pairs of items have a biplot interpretation (the pairs \mathbf{B}_f – \mathbf{X} and \mathbf{C}_f – \mathbf{U} have no known useful biplot interpretation for $\alpha = 1$ and 0, respectively).

4 Real data and simulation example

We illustrate dc-CA by analysing the example dataset of Jamil et al. (2013). It represents 20 sites from Dutch dune meadows, where the plant community composition was recorded together with five environmental variables from which two important ones are used here for illustration (abbreviations between parentheses): moisture content of the soil (Moisture) and manure quantity applied (Manure). For each of the 28 plant species, five functional traits were available from which two important ones are used here for illustration: specific leaf area (SLA) and seed mass. With these variables in the model, further traits and environmental variables do not contribute much to explain the community table as judged by species-based and site-based permutation tests, respectively, using the total dc-CA inertia as the test statistic.

Figure 1 shows the row-metric preserving dc-CA biplot ($\alpha = 1$) with arrows for traits and environmental variables (\mathbf{C}_f and \mathbf{B}_f), which display by their inner product

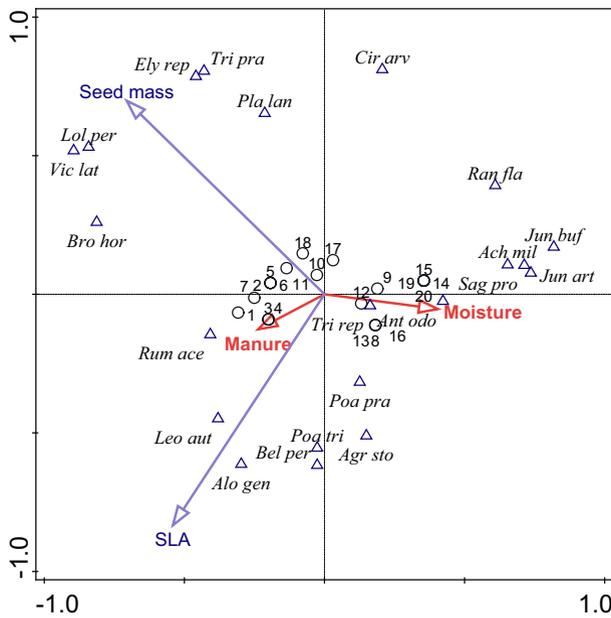


Fig. 1 Quadriplot of dc-CA relating two selected traits (blue) and two selected environmental variables (red) in the Dutch dune meadow data with 20 sites (circular points) and 29 plant species (triangles) ($\alpha = 1$, $\lambda_1 = 0.43$ and $\lambda_2 = 0.15$). Abbreviations follow ter Braak (1987). For interpretation see text

their fourth-corner correlation, together with points for species and sites. The strongest correlations are those between Moisture and Seed mass (-0.32) and Manure and SLA (0.23) as indicated by projecting the arrows for Seed mass and SLA on the arrows for Moisture and Manure; alternatively consider their obtuse and sharp angles, respectively, and the lengths of the arrows. The maximized fourth-corner correlations along the first (horizontal) and second (vertical) axes are 0.43 and 0.15 , respectively. Because $\alpha = 1$, the configuration of site points and environmental arrows shows the importance of the first axis compared to the second; the environmental arrows are fourth-corner correlations and the trait arrows intra-set correlations.

Beyond the pair $\mathbf{B}_f\text{--}\mathbf{C}_f$, four other pairs of items have a biplot interpretation. Example interpretations are as follows:

Biplot of $\mathbf{B}_f\text{--}\mathbf{U}$: the species points on the left in Fig. 1 have low SNC of Moisture (i.e. occur more at drier conditions) and the species on the right have high SNC with respect to Moisture (i.e. occur more at wetter conditions).

Biplot of $\mathbf{C}_f\text{--}\mathbf{X}$: the site points when projected onto the arrow for SLA represent the CWM of SLA (their mean SLA), so that, for example, site 1 is inferred to have the highest mean SLA and sites 14 and 15 the lowest mean SLA.

Biplot of $\mathbf{U}\text{--}\mathbf{X}$: the species and site points form a biplot of the contingency ratios, for example, the share of species *Alo gen* is high in site 1 compared to sites 14 and 15.

Biplot of $\mathbf{C}_f\text{--}\mathbf{U}$: the species points when projected on to Seed mass represent their Seed mass, so that the species *Vic lat* and *Lol per* are inferred to have high Seed mass and *Jun art* and *Agr Sto* low Seed mass.

Table 1 Percentiles of the fourth-correlations of the first two axes and their squared ratio of dc-CA and RLQ in 10,000 simulated data sets with $n = m = 100$

	ρ_1		ρ_2		λ_1/λ_2	
	dc-CA	RLQ	dc-CA	RLQ	dc-CA	RLQ
2.50%	0.15	0.03	0.04	0.02	8.0	0.2
50%	0.21	0.08	0.05	0.03	18.2	7.4
97.50%	0.28	0.15	0.06	0.08	39.2	45.6

In this simple example, there is little difference between dc-CA and RLQ: although dc-CA maximized the fourth-corner correlation and RLQ maximized the fourth-corner covariance (subject to standardized axes), the resulting fourth-correlations are almost identical. That is not always the case as will be shown in a simulation example. The R code for the simulation is in Supplementary “Appendix S2”.

In the simulation example, six traits and nine environmental variables are generated from a multivariate normal distribution with variance matrix with entries $\{\rho^{i-j}\}$, so that the pairwise correlation is ρ for neighbouring columns and decreases with the column number distance between variables. The difference between the first two traits was taken as a hidden trait dimension (\mathbf{t}) and, similarly, the difference between the first two environmental variables was taken as a hidden environment dimension (\mathbf{e}) and the product \mathbf{te} was used as one of the predictors in a log-linear model from which negatively binomial counts (y_{ij}) were generated. The other traits and environmental variables had no effect on the counts, but to make the data a little more realistic two more latent effects were added to the log-linear model: one product between \mathbf{e} and an independent latent standard normal trait \mathbf{z} and one product between a latent standard normal environmental variable and the latent trait \mathbf{z} . The count data are thus in essence two dimensional but only one dimension can usefully be correlated with the measured traits and environmental variables. The intercept in the log-linear model was $\log(10)$, all other coefficients were 0.2 and $\rho = 0.7$, so that neighbouring variables share about 50% of their variance.

Table 1 shows results of the simulation. The maximized fourth-corner correlation of dc-CA of the first axis was much higher than the RLQ fourth-corner correlation of the first axis, with the 2.5% percentile of dc-CA being even bigger than the 97.5% percentile of RLQ. The fourth-corner correlations of the second axes (which were zero in the model) were comparable. The ratio of the first over the second eigenvalue (squared correlation), which is a measure of the dominance of the first axis over the second in each simulated data set, is on average much higher in dc-CA than in RLQ (last two columns in Table 1). Compared to RLQ, dc-CA thus much better indicates that only one dimension is important for describing the association between the observed trait and environmental variables.

Figure 2 shows the distribution of the canonical coefficients of traits and environmental variables on the first axis in the 10,000 simulations. The contrast between the first and second variable is clear in almost all data sets in dc-CA whereas it is blurred or even absent in RLQ. The remaining noise variables have larger coefficients in RLQ than in dc-CA and also show negative bias.

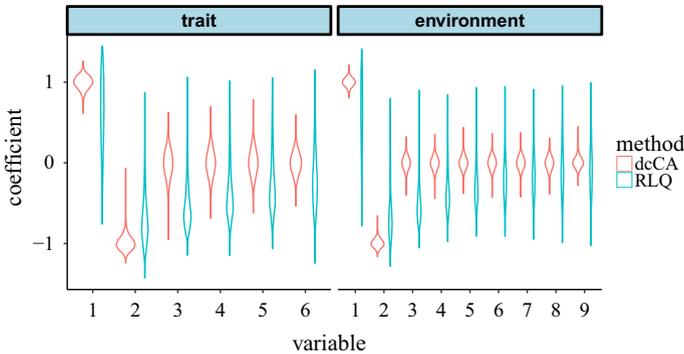


Fig. 2 Violin plots of the canonical coefficients of traits and environmental variables on the first axis in the 10,000 simulated data set in the simulation example

5 Discussion

This paper fills a gap by giving a mathematical description of double constrained correspondence analysis (dc-CA) starting from the idea that it maximizes a correlation, in particular the fourth-corner correlation between linear combinations of traits and of environmental variables. It was known from the start that dc-CA is identical with canonical correlation analysis of super-inflated trait and environment data. But dc-CA deserves special treatment as the units of sampling are not the individuals that are counted but the sites with individuals belonging to different species. Our mathematical development shows the precise role of community (site-) weighted trait means (CWM) and, its reverse, species niche centroids (species-weighted mean environment, SNC) in dc-CA. In “Appendix A6” it is reiterated why CWMs and SNCs are key statistics in trait-environment studies and that the within-site trait variance and the within-species environmental variance (niche breadth) may deserve separate study in relation to the environment and traits, respectively. The novel algorithm that combines a singly constrained correspondence (i.e. CCA) with a weighted singly constrained principal component analysis (i.e. redundancy analysis) shows the relation with CWM-RDA, an ad-hoc method that is commonly used to relate traits to environment. CWM-RDA uses regression onto the environmental variables, whereas dc-CA also uses regression onto the traits. By contrast, RLQ, one of the oldest multivariate methods for trait-environment analysis, is based on covariance without using regression at all. Our small simulation example demonstrated that, by combining correlated observed variables, dc-CA can detect trait and environment relationships that remain hidden in RLQ.

RLQ is based on coinertia analysis (Dray et al. 2003) while dc-CA is based on CCA. Therefore the comparison between coinertia analysis and CCA by Dray et al. (2003) is of interest. They showed that CCA deteriorates in detecting the hidden gradients when many highly correlated environmental variables that have no real effect are included in the analysis. In such an extreme situation coinertia performs better than CCA and the same is expected for RLQ, and dc-CA. However, with moderate correlations or when multicollinearity problems are taken care of, for example by

variable selection or by changing regression to ridge regression, we expect dc-CA to outperform RLQ.

Many researchers see CA as a special form of principal component analysis and thus dc-CA as a special form of double constrained principal component analysis (dc-PCA) (Takane 2013), also known as two-way CANDELINC (Douglas Carroll et al. 1980). The methods are perhaps even more similar in the double constrained case than in the unconstrained case. The similarity (and difference) is best seen from the matrix that must be subjected to SVD to obtain a weighted dc-PCA with weight matrices \mathbf{R} and \mathbf{K} . Whereas dc-CA uses a SVD of \mathbf{D} defined in Eq. (15), dc-PCA can be obtained from an SVD of

$$\mathbf{D}_{\text{dc-pca}} = \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1/2} \mathbf{E}^T \mathbf{R} \mathbf{Y} \mathbf{K} \mathbf{T} \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-1/2} \quad (25)$$

This equation shows immediately that dc-CA is a weighted dc-PCA of the contingency ratios $y_{++} \mathbf{R}^{-1} \mathbf{Y} \mathbf{K}^{-1}$ where the weight matrices \mathbf{R} and \mathbf{K} are diagonal with the row- and column sums of \mathbf{Y} on the diagonal. The post-processing to obtain the canonical weights and the row and column scores is identical to that in dc-CA [e.g. Eq. (17)]. Viewed in this framework, there is no interpretation of dc-CA in terms of fourth-corner correlations. The link of the total inertia of dc-CA with the Rao score test statistic for testing linear-by-linear interaction in a contingency table (ter Braak 2017) shows explicitly that dc-CA is a natural method for count-like data.

Aitchison's log-ratio analysis is essentially the analysis of double-centred log-transformed \mathbf{Y} (see discussion by Dawid of Aitchison (1982) which leads to the centred log-ratio transformation). Microbiome data are sometimes analyzed by Aitchison's log-ratio PCA (Gloor et al. 2016), despite the fact that they contain many zeros. Using weighted (double) (constrained) log-ratio analysis with row- and column sums of \mathbf{Y} as weights (Greenacre and Lewi 2009) will decrease the adverse effect of rows and columns with many zeroes, at least if the number of zeroes is reflected in the weights (alternatively the row-wise and column-wise numbers of non-zeroes could be used as weights). A natural alternative is in our view (double) (constrained) CA which does not need tricks for handling zeroes in the data.

The computer program Canoco 5.10 (ter Braak and Šmilauer 2012) implements dc-CA as a combination of CCA and a weighted RDA. Weighted dc-PCA is implemented by changing the initial CCA ($\mathbf{Y}^T \sim \mathbf{T}$) by a weighted RDA $\mathbf{K}(\mathbf{Y}^T \sim \mathbf{T})$. For this purpose, both RDAs must be centred by rows and by columns in the metrics \mathbf{R} and \mathbf{K} , respectively, which is in agreement with the idea that the trait-environment association is an interaction and should not involve main effects. By prior log-transformation, a (weighted or unweighted) double constrained log-ratio analysis is obtained. For statistical inference about the trait-environment relation in dc-CA see ter Braak (2017) and ter Braak et al. (2017) and, in the log-ratio context, Cormont et al. (2011). These methods can quickly provide an overview of which variables appear important. We believe that they deserve more consideration, evaluation and use.

Acknowledgements We thank John Birks and Richard Furnas for inspiration and comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

6 Appendices

6.1 Appendix A1: model-based derivations

Model-based derivations of CA, CCA and thus of dc-CA start from the Gaussian response model (ter Braak 1985, 1986) or from a log-linear model with interaction terms (Goodman 1981). They all lead to the transition formulae in Sect. 2.2.

The Gaussian response model is

$$E(y_{ij}) = s_i h_j \exp \left[-\frac{(x_i - u_j)^2}{2\sigma_j^2} \right], \tag{26}$$

where s_i is a site-specific parameter (Ihm and Groenewoud 1975; ter Braak 1988) or $s_i = 1$, so that there are no site specific parameters (ter Braak 1985, 1986), h_j is a species-specific parameter denoting the maximum expected abundance, x_i is a site score, often interpreted as a latent environmental variable, u_j is a species score, often interpreted as a latent trait and, in this model, the optimum for the j th species and σ_j is the niche width of the j th species. The link with the correspondence analysis methods is strongest when $\sigma_j = \sigma$ and when this constant σ is small compared to the range of the site scores (ter Braak 1985), i.e. when there is strong and uniform niche differentiation among the species.

The alternative model-based start is from the log-linear model with saturated main effects for rows (sites) and columns (species) and one or more linear-by-linear interaction terms (Goodman 1981; Ihm and Groenewoud 1984; ter Braak 2017):

$$\log (E (y_{ij})) = r_i^* + c_j^* + b_{te} u_j x_i, \tag{27}$$

where r_i^* and c_j^* are row and column saturated main effects, x_i and u_j latent row and column scores and b_{te} a scalar coefficient which is set to 1 unless $\{x_i\}$ and $\{u_j\}$ are both standardized. When $\sigma_j = \sigma$ in model (26), the models (26) and (27) are re-parametrization of one-another.

The link with correspondence analysis (CA) is easiest to see in model (27). Indeed, rewriting model (27) in an exponential form and using a first order Taylor expansion in terms of $b_{te} u_j x_i$ yields the reconstitution formula of correspondence analysis (Greenacre 1984; Ihm and Groenewoud 1975):

$$\mu_{ij} = R_i^* C_j^* \exp (b_{te} u_j x_i) \approx R_i^* C_j^* (1 + b_{te} u_j x_i), \tag{28}$$

where $R_i^* = \exp(r_i^*)$ and $C_j^* = \exp(c_j^*)$. So, for small b_{te} and standardized $\{x_i\}$ and $\{u_j\}$, both models can be expected to be very similar. Goodman (1981) showed that their estimation equations are then also very similar. For standardized \mathbf{u} and \mathbf{x} , b_{te} is the square-root of the first eigenvalue of correspondence analysis on \mathbf{Y} .

With a linear constraint on the site scores, $\mathbf{x} = \mathbf{E}\mathbf{b}$, where \mathbf{b} is a vector of unknown coefficients, one for each environmental variable, and the same approximation as in obtaining CA from models (26) and (27), CCA is obtained (ter Braak 1986, 1988). With an additional linear constraint on the species scores, $\mathbf{u} = \mathbf{T}\mathbf{c}$, where \mathbf{c} is a vector of unknown coefficients, one for each trait, dc-CA is obtained in a similar way. The resulting transition formulae are presented in Sect. 2.2. Böckenholt and Böckenholt (1990) showed an example where the estimates by dc-CA and maximum likelihood of the log-linear model are indeed very close.

6.2 Appendix A2: WA scores of dc-CA from CCA/RDA algorithm

This appendix shows that the unconstrained scores of weighted RDA $\mathbf{R}(\mathbf{M}^* \sim \mathbf{E})$ are the WA site scores of dc-CA and how to obtain the other set of canonical coefficients of dc-CA from the CCA and RDA pair.

The unconstrained scores of RDA $\mathbf{R}(\mathbf{M}^* \sim \mathbf{E})$ are a linear combination of the response data, i.e.

$$\mathbf{X}_{\text{rda}}^* = \mathbf{M}^*\mathbf{B}^* \quad (29)$$

where \mathbf{B}^* are the response variable scores of the RDA and

$$\mathbf{M}^* = \mathbf{R}^{-1}\mathbf{Y}\mathbf{T}\mathbf{B}_1 \quad (30)$$

the community weighted means of the constrained axes of the first CCA, CCA($\mathbf{Y}^T \sim \mathbf{T}$), with \mathbf{B}_1 the matrix of canonical coefficients of this first CCA. Insertion of Eq. (30) into Eq. (29) gives

$$\mathbf{X}_{\text{rda}}^* = \mathbf{R}^{-1}\mathbf{Y}\mathbf{T}\mathbf{B}_1\mathbf{B}^* = \mathbf{R}^{-1}\mathbf{Y}\mathbf{U} \quad (31)$$

so that $\mathbf{X}_{\text{rda}}^*$ are WA site scores derived from constrained species scores as in Eq. (12). These are the desired WA sites scores because their projection onto \mathbf{E} using weights \mathbf{R} are the constrained site scores of the RDA and thus also of the dc-CA. The regression coefficients of this projection are the canonical weights for the environmental variables and satisfy Eq. (13). The dc-CA canonical coefficients for the traits \mathbf{C} follow from Eq. (31) with Eq. (11): $\mathbf{C} = \mathbf{B}_1\mathbf{B}^*$.

6.3 Appendix A3: biplot of fourth-corner correlations

This appendix shows that, for $\alpha = 1$ in Eq. (24), the intra-set correlations of the traits plotted with the fourth-corner correlations of the environmental variables with the axes, together form a weighted least-squares biplot of the fourth-corner correlations between traits and environmental variables. Vice versa, for $\alpha = 0$ in Eq. (24), the intra-set correlations of the environmental variables can be plotted with the fourth-

correlations of the traits with the axes. This appendix also explores how intra- and inter-set correlations relate to the fourth-corner correlations of the traits and environmental variables with the axes.

The first equation in (24) can be rewritten as

$$\begin{aligned} \mathbf{B}_f &= \left[\left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{\frac{1}{2}} \mathbf{P} \mathbf{\Delta}^\alpha \right]_r = \left[\left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{\frac{1}{2}} \mathbf{D} \mathbf{Q} \mathbf{\Delta}^{\alpha-1} \right]_r \\ &= \left[\mathbf{E}^T \mathbf{Y} \mathbf{T} \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-\frac{1}{2}} \mathbf{Q} \mathbf{\Delta}^{\alpha-1} \right]_r \end{aligned} \tag{32}$$

by inserting, from Eq. (16), $\mathbf{P} \mathbf{\Delta}^\alpha = \mathbf{D} \mathbf{Q} \mathbf{\Delta}^{\alpha-1}$ and inserting \mathbf{D} from Eq. (15). With the second equation in (17), this gives, with $\mathbf{\Lambda} = \mathbf{\Delta}^2$,

$$\mathbf{B}_f = \left[\mathbf{E}^T \mathbf{Y} \mathbf{T} \mathbf{C} \mathbf{\Delta}^{2(\alpha-1)} \right]_r = \left[\mathbf{E}^T \mathbf{Y} \mathbf{U} \mathbf{\Lambda}^{\alpha-1} \right]_r = \text{cor}_Y(\mathbf{E}, \mathbf{U}) \mathbf{\Delta}^{\alpha-1} \tag{33}$$

so that \mathbf{B}_f consists of fourth-corner correlations between \mathbf{E} and $[\mathbf{U}]_r$ for \mathbf{K} -normalized \mathbf{U} ($\alpha = 1$). For $\alpha = 1$ in Eq. (24) and \mathbf{K} -normalized \mathbf{U} ,

$$\mathbf{C}_f = \left[\left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{1/2} \mathbf{Q} \right]_r = \left[\left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right) \mathbf{C} \right]_r = \left[\mathbf{T}^T \mathbf{K} \mathbf{U} \right]_r. \tag{34}$$

so that \mathbf{C}_f consists of intra-set correlations between \mathbf{T} and $[\mathbf{U}]_r$.

We now explore how intra- and inter-set correlations of the environmental variables relate to the fourth-corner correlations of the traits and with the axes.

By using Eq. (17), \mathbf{B}_f in equation in (24) can be rewritten as

$$\mathbf{B}_f = \left[\left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-\frac{1}{2}} \mathbf{P} \mathbf{\Delta}^\alpha \right]_r = \left[\left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right) \mathbf{B} \right]_r = \left[\mathbf{E}^T \mathbf{R} \mathbf{X} \right]_r = \text{cor}_R(\mathbf{E}, \mathbf{X}) \mathbf{\Delta}^\alpha \tag{35}$$

so that, for \mathbf{R} -normalized \mathbf{X} ($\alpha = 0$), \mathbf{B}_f consists of intra-set correlations between \mathbf{E} and $[\mathbf{X}]_r$. From Eqs. (33) and (35), and analogously for \mathbf{T} and \mathbf{X} ,

$$\text{cor}_Y(\mathbf{E}, \mathbf{U}) = \text{cor}_R(\mathbf{E}, \mathbf{X}) \mathbf{\Delta} \text{ and } \text{cor}_Y(\mathbf{T}, \mathbf{X}) = \text{cor}_R(\mathbf{T}, \mathbf{U}) \mathbf{\Delta} \tag{36}$$

so that fourth-corner correlations with the axes are a factor $\sqrt{\lambda}$ smaller than the intra-set correlations.

An expression for \mathbf{B}_f in terms of inter-set correlations can be obtained from Eq. (35), using the matrix version of Eq. (13),

$$\mathbf{B}_f = \left[\left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right) \mathbf{B} \right]_r = \left[\left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right) \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1} \mathbf{E}^T \mathbf{R} \mathbf{X}^* \right]_r = \left[\mathbf{E}^T \mathbf{R} \mathbf{X}^* \right]_r \tag{37}$$

so that, for any α , \mathbf{B}_f consists of the inter-set correlation of the environmental variables times the standard deviation of \mathbf{X}^* , i.e.

$$\mathbf{B}_f = \text{cor}_{\mathbf{R}}(\mathbf{E}, \mathbf{X}^*) \text{diag}(sd(\mathbf{X}^*)) \tag{38}$$

In CCA and RDA implementations in the Canoco software (ter Braak and Šmilauer 2012), this equation is used for so-called biplot scores of environmental variables, so that \mathbf{B}_f is such a score, which depends on α . In dc-CA, effectively, Eq. (35) is used for \mathbf{B}_f which together with $\mathbf{C}_f = \text{cor}_{\mathbf{K}}(\mathbf{T}, \mathbf{U}) \mathbf{\Lambda}^{1-\alpha}$ forms a biplot as follows from Eq. (24) using Eq. (35) for \mathbf{B}_f and, for \mathbf{C}_f , the analogous version of this equation and Eq. (36).

We end this appendix with a, perhaps, simpler derivation of equation (33). With the matrix version of Eq. (12) in Eq. (37), we obtain Eq. (33):

$$\mathbf{B}_f = [\mathbf{E}^T \mathbf{R} \mathbf{X}^*]_r = [\mathbf{E}^T \mathbf{R} (\mathbf{R}^{-1} \mathbf{Y} \mathbf{U} \mathbf{\Lambda}^{\alpha-1})]_r = [\mathbf{E}^T \mathbf{Y} \mathbf{U} \mathbf{\Lambda}^{\alpha-1}]_r \tag{39}$$

6.4 Appendix A4: biplot of species niche centroids (SNCs) and CWMs

This appendix shows that an ordination diagram with the species scores $[\mathbf{U}]_r$ supplemented with environmental arrows based on \mathbf{B}_f form a least-squares biplot of the species niche centroids (SNCs),

$$\mathbf{N} = \mathbf{K}^{-1} \mathbf{Y}^T \mathbf{E} \tag{40}$$

which is an $m \times p$ matrix. Analogously, an ordination diagram with the site scores $[\mathbf{X}]_r$ supplemented with trait arrows based on \mathbf{C}_f form a least-squares biplot of the community weighted means (CWMs) of Eq. (18). Moreover, if the species scores and sites scores \mathbf{U} and \mathbf{X} satisfy the transition formulae and thus form a biplot pair for the (fitted) \mathbf{Y} via the reconstitution formulae (as in CA or CCA), then the environmental arrows \mathbf{B}_f not only form a least-squares biplot of the fourth-corner correlation with trait arrows \mathbf{C}_f , but with $[\mathbf{U}]_r$ forms also a biplot of the SNCs and \mathbf{C}_f simultaneously forms with $[\mathbf{X}]_r$ a biplot of the CWMs.

For notational convenience we drop the $[\cdot]_r$ notation and write \mathbf{U} where it is clear that only r columns are being used. When \mathbf{U} is given and scaled such that $\mathbf{U}^T \mathbf{K} \mathbf{U} = \mathbf{\Lambda}^{1-\alpha}$ and \mathbf{N} is to be approximated by a biplot, the optimal scores for the environmental variables are obtained by fitting the model $\mathbf{N} = \mathbf{U} \mathbf{B}_0 + \text{error}$ by a weighted regression of \mathbf{N} onto \mathbf{U} with, as standard in dc-CA, species weights \mathbf{K} . The estimated regression coefficients are

$$\hat{\mathbf{B}}_0 = (\mathbf{U}^T \mathbf{K} \mathbf{U})^{-1} \mathbf{U}^T \mathbf{K} \mathbf{N} = \mathbf{\Lambda}^{\alpha-1} \mathbf{U}^T \mathbf{Y}^T \mathbf{E} = \mathbf{B}_f^T \tag{41}$$

where $\mathbf{\Lambda} = \mathbf{\Delta}^2$, and the last equality follows from Eq. (33). $\hat{\mathbf{B}}_0$ is thus the transpose of \mathbf{B}_f . For $r = \min(p, q)$, $\mathbf{U} \mathbf{B}_f^T$ is equal to the full rank fitted SNC values (see also Sect. 6.5).

Analogously, suppose \mathbf{X} is given and scaled such that $\mathbf{X}^T \mathbf{R} \mathbf{X} = \mathbf{\Lambda}^\alpha$ and the community weighted mean matrix \mathbf{M} [Eq. (18)] is to be approximated by a biplot, the optimal scores for the environmental variables are obtained by fitting the model $\mathbf{M} = \mathbf{X} \mathbf{C}_0 + \text{error}$ by a weighted regression of \mathbf{M} onto \mathbf{X} with, as standard in dc-CA, site weights \mathbf{R} . The estimated regression coefficients are

$$\hat{\mathbf{C}}_0 = (\mathbf{X}^T \mathbf{R} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{R} \mathbf{M} = \mathbf{\Lambda}^{-\alpha} \mathbf{X}^T \mathbf{Y} \mathbf{T} \tag{42}$$

The last equality can be shown by the route followed in Eqs. (32) and (33):

$$\begin{aligned} \mathbf{C}_f &= \left[(\mathbf{T}^T \mathbf{K} \mathbf{T})^{\frac{1}{2}} \mathbf{Q} \mathbf{\Delta}^{1-\alpha} \right]_r = \left[(\mathbf{T}^T \mathbf{K} \mathbf{T})^{\frac{1}{2}} \mathbf{D}^T \mathbf{P} \mathbf{\Delta}^{-\alpha} \right]_r \\ &= [(\mathbf{T}^T \mathbf{Y}^T \mathbf{E}) (\mathbf{E}^T \mathbf{R} \mathbf{E})^{-\frac{1}{2}} \mathbf{P} \mathbf{\Delta}^{-\alpha}]_r \end{aligned} \tag{43}$$

$$\mathbf{C}_f = [(\mathbf{T}^T \mathbf{Y}^T \mathbf{E}) \mathbf{B} \mathbf{\Lambda}^{-2\alpha}]_r = (\mathbf{T}^T \mathbf{Y}^T \mathbf{X}) \mathbf{\Lambda}^{-\alpha} \tag{44}$$

$\hat{\mathbf{C}}_0$ is thus the transpose of \mathbf{C}_f . For $r = \min(p, q)$, $\mathbf{X} \mathbf{C}_f^T$ is equal to the full rank fitted CWM values (see also Sect.6.5).

6.5 Appendix A5: biplots involving canonical weights

This appendix describes biplots involving canonical weights: $\mathbf{B}-\mathbf{C}$, $\mathbf{B}-\mathbf{C}_f$, $\mathbf{B}-\mathbf{X}$ and the biplots obtained by symmetry: $\mathbf{B}_f-\mathbf{C}$ and $\mathbf{C}-\mathbf{U}$. For completeness, the reconstitution formula for the fitted community matrix \mathbf{Y} is given with its biplot based on \mathbf{X} and \mathbf{U} as a corollary of the $\mathbf{B}-\mathbf{C}$ biplot.

The biplot of \mathbf{B} and \mathbf{C}

A weighted regression of the contingency ratios $y_{++} \mathbf{R}^{-1} \mathbf{Y} \mathbf{K}^{-1}$ on the traits and environmental variables, with weights \mathbf{R} and \mathbf{K} , results (ignoring y_{++}) in the regression coefficients (Gabriel 1998)

$$\begin{aligned} \mathbf{F}_{reg} &= (\mathbf{E}^T \mathbf{R} \mathbf{E})^{-1} \mathbf{E}^T \mathbf{R} (\mathbf{R}^{-1} \mathbf{Y} \mathbf{K}^{-1}) \mathbf{K} \mathbf{T} (\mathbf{T}^T \mathbf{K} \mathbf{T})^{-1} \\ &= (\mathbf{E}^T \mathbf{R} \mathbf{E})^{-1} \mathbf{E}^T \mathbf{Y} \mathbf{T} (\mathbf{T}^T \mathbf{K} \mathbf{T})^{-1} \end{aligned} \tag{45}$$

Following ter Braak (1990) and Sect. 3, a biplot of \mathbf{F}_{reg} can be based on a “rank r weighted least-squares approximation” of the form $\mathbf{F}_{reg} \approx \mathbf{B} \mathbf{C}^T$ with \mathbf{B} and \mathbf{C} matrices of order $p \times r$ and $q \times r$, respectively. It is shown below that the optimal \mathbf{B} and \mathbf{C} are the first r columns of the canonical weights of dc-CA. For simplicity of notation these matrices are already indicated by the symbols for canonical coefficients with the $[\cdot]_r$ notation dropped as well, as in Sect. 6.4. When dc-CA is a good approximation of the models in Sect. 6.1 [e.g. Eq. (28)], the \mathbf{F}_{reg} is likely close to the regression coefficients associated with the interactions between traits and environment in a log-linear model (Brown et al. 2014; ter Braak 2017).

As the regression coefficients have a variance that is proportional to the tensor product of the inverses of the matrices $\mathbf{E}^T \mathbf{R} \mathbf{E}$ and $\mathbf{T}^T \mathbf{K} \mathbf{T}$, it is natural to use $\mathbf{E}^T \mathbf{R} \mathbf{E}$ and $\mathbf{T}^T \mathbf{K} \mathbf{T}$ as weights. The weighted approximation can be obtained from dc-CA as follows. We seek the minimum over \mathbf{B} and \mathbf{C} (free matrices, not yet equal to the canonical coefficients) of

$$\begin{aligned} & \left\| \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{1/2} \left(\mathbf{F}_{reg} - \mathbf{B} \mathbf{C}^T \right) \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{1/2} \right\|^2 \\ & = \left\| \mathbf{D} - \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1/2} \mathbf{B} \mathbf{C}^T \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-1/2} \right\|^2 \end{aligned} \tag{46}$$

As follows from the Eckhart–Young theorem (Greenacre 1984) the minimum is obtained from the singular value decomposition of \mathbf{D} . By consequence, the minimum of (46) is $\lambda_{r+1} + \dots + \lambda_{\min(p,q)}$ and is obtained by setting \mathbf{B} and \mathbf{C} equal to the first r columns of the canonical weights of Eq. (17). The scores of \mathbf{X} and \mathbf{U} thus form a biplot of the fitted contingency ratios. The biplot is weighted least-squares with weights \mathbf{R} and \mathbf{K} .

The regression of the contingency ratios on the traits and environmental variables leads to fitted values and thus also to fitted values of \mathbf{Y} itself. The fitted values,

$$\hat{\mathbf{Y}} = y_{++}^{-1} \mathbf{R} \left(\mathbf{1}_n \mathbf{1}_m^T + \mathbf{X} \mathbf{U}^T \right) \mathbf{K} = y_{++}^{-1} \mathbf{R} \left(\mathbf{1}_n \mathbf{1}_m^T + \mathbf{X} \mathbf{B} \mathbf{C}^T \mathbf{U}^T \right) \mathbf{K} \tag{47}$$

have the form of the usual reconstitution formula for \mathbf{Y} but with constrained instead of unconstrained scores as in CA.

The biplot of \mathbf{B} and \mathbf{C}_f

The other biplots essentially follow from considering dc-CA as a canonical correlation analysis on inflated trait and environment data and noting that canonical correlation analysis can be seen as reduced-rank regression fitted by maximum likelihood (ter Braak 1990; Tso 1981).

In the super inflated data of integer-valued \mathbf{Y} , each row represents an individual. When predicting traits from environmental variables, the predicted values of the individuals of the same site are all identical and are thus equal to the community weighted mean of the predicted trait values. This suggests to consider the regression of the community weighted means \mathbf{M} onto the environment \mathbf{E} . With weights \mathbf{R} , the estimated regression coefficients are

$$\hat{\mathbf{B}}_{\mathbf{T} \sim \mathbf{E}} = \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1} \mathbf{E}^T \mathbf{R} \mathbf{M} = \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1} \mathbf{E}^T \mathbf{Y} \mathbf{T} \tag{48}$$

A biplot of $\hat{\mathbf{B}}_{\mathbf{T} \sim \mathbf{E}}$ can be based on a “rank r weighted least-squares approximation” of the form $\hat{\mathbf{B}}_{\mathbf{T} \sim \mathbf{E}} \approx \mathbf{B} \mathbf{C}_f^T$ with \mathbf{B} and \mathbf{C}_f matrices of order $p \times r$ and $q \times r$, respectively. It is shown below that the optimal \mathbf{B} and \mathbf{C}_f are the first r columns of the canonical weights for the environmental variables and the biplot scores of the traits of the dc-CA. For simplicity of notation these matrices are already indicated by their symbols in the main text and the $[\cdot]_r$ notation is dropped as well, as in Sect. 6.4.

The weighted approximation is motivated as follows. Because the regression coefficients for each column of \mathbf{M} have a variance that is proportional to the inverse of the matrix $\mathbf{E}^T \mathbf{R} \mathbf{E}$, it is natural to use $\mathbf{E}^T \mathbf{R} \mathbf{E}$ as weights. To make the approximation invariant to linear transformation of \mathbf{T} , the inverse of $\mathbf{T}^T \mathbf{K} \mathbf{T}$ forms the other set of weights, as in Eq. (23). We seek thus the minimum over \mathbf{B} and \mathbf{C}_f (free matrices for now) of

$$\begin{aligned} & \left\| \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{1/2} \left(\hat{\mathbf{B}}_{\mathbf{T} \sim \mathbf{E}} - \mathbf{B} \mathbf{C}_f^T \right) \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-1/2} \right\|^2 \\ & = \left\| \mathbf{D} - \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{1/2} \mathbf{B} \mathbf{C}_f^T \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-1/2} \right\|^2 \end{aligned} \tag{49}$$

As follows from the Eckhart–Young theorem (Greenacre 1984) the minimum is obtained from the singular value decomposition of \mathbf{D} . By consequence, the minimum of (46) is $\lambda_{r+1} + \dots + \lambda_{\min(p,q)}$ and is obtained by setting \mathbf{B} and \mathbf{C}_f equal to the first r columns of the canonical weights of the environmental variables in Eq. (17) and to the biplot scores of the traits in Eq. (24). This result (and the version with traits and environmental variables interchanged) shows that dc-CA is both a reduced rank regression of CWMs on the environment \mathbf{E} and a reduced rank regression of SNCs on the traits \mathbf{T} .

The result can now be linked to a canonical correlation analysis of super inflated data. Such a canonical correlation analysis is simultaneously a multivariate regression of the traits on the environment and a multivariate regression of the environment on the traits with all data in super inflated form. The resulting optimal biplots are precisely the same as those obtained above for the regression of CWMs and, by symmetry, SNCs and the regression coefficients of these two regressions satisfy the transition formula in Eq. (6) and (7). In fact, these equations can be written explicitly in terms of CWMs and SNCs:

$$\lambda_b \mathbf{b} = \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1} \mathbf{E}^T \mathbf{R} \left(\mathbf{R}^{-1} \mathbf{Y} \mathbf{T} \right) \mathbf{c} = \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1} \mathbf{E}^T \mathbf{R} \mathbf{M} \mathbf{c} \tag{50}$$

$$\lambda_c \mathbf{c} = \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-1} \mathbf{T}^T \mathbf{K} \left(\mathbf{K}^{-1} \mathbf{Y}^T \mathbf{E} \right) \mathbf{b} = \left(\mathbf{T}^T \mathbf{K} \mathbf{T} \right)^{-1} \mathbf{T}^T \mathbf{K} \mathbf{N}^T \mathbf{b} \tag{51}$$

In conclusion, the biplot of canonical weights \mathbf{B} and biplot scores \mathbf{C}_f gives a weighted least-squares approximation of the regression coefficients of the regression of CWMs on the environment, or equivalently of the regression of the traits on the environmental variables (in super inflated form). Conversely, the biplot of canonical weights \mathbf{C} and biplot scores \mathbf{B}_f gives a weighted least-squares approximation of the regression coefficients of the regression of SNCs on the environment, or equivalently of the regression of the environment on the traits. The fitted traits and environmental values in these equivalent regression are equal to the fitted CWMs and fitted SNCs and their biplot representation is covered in Sect. 6.4, where CWM and SNC were regressed on \mathbf{X} and \mathbf{U} , respectively, and thus implicitly on \mathbf{E} and \mathbf{T} .

The interpolative biplot of \mathbf{B} and \mathbf{X}

Gower and Hand (1996) distinguish predictive and interpolative biplots. All biplots so far are predictive biplots in the sense that the two sets of items approximate a matrix by inner products (Gower and Hand 1996). Such biplots use loadings or biplot scores. Interpolative biplots use regression coefficient-like quantities. As their name suggests, they are useful for interpolation or adding a new site or species to the plot (Rui Alves and Beatriz Oliveira 2004). It is thus clear that \mathbf{B} and \mathbf{X} form an interpolative biplot.

6.6 Appendix A6: Why CWMs and SNCs are key in analyzing trait-environment relationships

Community weighted means (CWMs) and species niche centroids (SNC) appear several times in this paper. This appendix shows their importance in analyzing trait-environment relationships. Three models are considered, a log-linear model for \mathbf{Y} and two related linear models: a model for the trait data \mathbf{T} and one for the environment data \mathbf{E} .

When \mathbf{Y} consists of count data that are Poisson distributed and is modelled by a log-linear model with saturated main effects and interactions between all traits and all environmental variables, then the minimal sufficient statistics are $\mathbf{E}^T \mathbf{Y} \mathbf{T}$ together with \mathbf{R} and \mathbf{K} (ter Braak 2017). The CWMs $\mathbf{M} = \mathbf{R}^{-1} \mathbf{Y} \mathbf{T}$ and SNCs $\mathbf{N} = \mathbf{K}^{-1} \mathbf{Y}^T \mathbf{E}$ with \mathbf{R} and \mathbf{K} are thus sufficient statistics.

When \mathbf{Y} consists counts of individuals, it is natural to consider the super-inflated data \mathbf{T}_{infl} and \mathbf{E}_{infl} in which each individual is represented by a row: a row of \mathbf{T}_{infl} consisting of the trait values of the individual of the particular species it belongs to and a row of \mathbf{E}_{infl} consisting of the environmental values that the individual may experience because it occurs in a particular site. Associated with \mathbf{T}_{infl} and \mathbf{E}_{infl} are also the factors species and sites coding for which species and site each row belongs to. We now consider the linear model for \mathbf{T}_{infl} as a function g of \mathbf{E}_{infl} and the L_2 norm for the residuals.

$$\|\mathbf{T}_{infl} - g(\mathbf{E}_{infl})\|^2 = \|\Pi_s(\mathbf{T}_{infl} - g(\mathbf{E}_{infl})) + (1 - \Pi_s)(\mathbf{T}_{infl} - g(\mathbf{E}_{infl}))\|^2 \tag{52}$$

where Π_s is the projector onto the factor site. Because the two added terms are orthogonal, the square of their sum is the sum of their squares. Also $\Pi_s g(\mathbf{E}_{infl}) = g(\mathbf{E}_{infl})$ so that $(1 - \Pi_s)(\mathbf{T}_{infl} - g(\mathbf{E}_{infl})) = (1 - \Pi_s) \mathbf{T}_{infl}$ and Eq. (52) becomes

$$\|\mathbf{T}_{infl} - g(\mathbf{E}_{infl})\|^2 = \|\Pi_s \mathbf{T}_{infl} - g(\mathbf{E}_{infl})\|^2 + \|(1 - \Pi_s) \mathbf{T}_{infl}\|^2 \tag{53}$$

Such ANOVA-like decomposition was also given in Peres-Neto et al. (2017). The regression of \mathbf{T}_{infl} as a function g of \mathbf{E}_{infl} thus depends only on the first part which can be further simplified to a weighted regression of CWM \mathbf{M} with weights \mathbf{R}

$$\|\Pi_s \mathbf{T}_{infl} - g(\mathbf{E}_{infl})\|^2 = \|\mathbf{R}^{1/2}(\mathbf{M} - g(\mathbf{E}))\|^2 \tag{54}$$

because, being a projection on sites, $\Pi_s \mathbf{T}_{infl}$ consists of trait means per site, i.e. CWMs, and each site is replicated y_{i+} times, leading to the weights \mathbf{R} . The least-squares regression of \mathbf{T}_{infl} as a function g of \mathbf{E}_{infl} can thus be carried out as a weighted regression of the CWMs on the environmental data with weights \mathbf{R} . Similarly it can be shown that by projection of \mathbf{E}_{infl} on the factor species the least-squares regression of \mathbf{E}_{infl} as a function h of \mathbf{T}_{infl} is a weighted regression of the SNCs on the trait data with weights \mathbf{K} .

The variance of the residuals $(1 - \Pi_s) \mathbf{T}_{infl}$ per site represents the within-site trait variance that may deserve separate study in relation to the environment. Similarly, the variance of the residuals $(1 - \Pi_{species}) \mathbf{E}_{infl}$ per species represents the within-species environmental variance (niche breadth) that may deserve separate study in relation to the traits.

In the above derivations, the Poisson assumption is either explicit or implicit, but can also be overcome by choosing a transformation that makes the result Poisson-like, with variance proportional to the mean. For example, if the data follow a Poisson log-normal distribution it may make sense to analyse $\log(\mathbf{Y} + 1)$ instead of \mathbf{Y} .

References

- Aitchison J (1982) The statistical analysis of compositional data. *J R Stat Soc B* 44:139–177. <http://www.jstor.org/stable/2345821>
- Bacou AM, Sabatier R, Lespinasse P (1989) Analyses des correspondance avec une ou deux contraintes avec le logiciel Biomeco, manuel de l'utilisateur. CEFE, CNRS, Montpellier
- Böckenholt U, Böckenholt I (1990) Canonical analysis of contingency tables with linear constraints. *Psychometrika* 55:633–639. <https://doi.org/10.1007/bf02294612>
- Brown AM, Warton DI, Andrew NR, Binns M, Cassis G, Gibb H (2014) The fourth-corner solution—using predictive models to understand how species traits interact with the environment. *Methods Ecol Evol* 5:344–352. <https://doi.org/10.1111/2041-210x.12163>
- Cormont A, Vos CC, van Turnhout CAM, Foppen RPB, ter Braak CJF (2011) Using life-history traits to explain bird population responses to changing weather variability. *Clim Res* 49:59–71. <https://doi.org/10.3354/cr01007>
- Dolédéc S, Chessel D, ter Braak CJF, Champely S (1996) Matching species traits to environmental variables: a new three-table ordination method. *Environ Ecol Stat* 3:143–166. <https://doi.org/10.1007/BF02427859>
- Douglas Carroll J, Pruzansky S, Kruskal JB (1980) Candelinc: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika* 45:3–24. <https://doi.org/10.1007/bf02293596>
- Dray S, Chessel D, Thioulouse J (2003) Co-inertia analysis and the linking of ecological tables. *Ecology* 84:3078–3089. <http://www.jstor.org/stable/3449976>
- Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* 22:1–20. <https://doi.org/10.18637/jss.v022.i04>
- Dray S, Legendre P (2008) Testing the species traits environment relationships: the fourth-corner problem revisited. *Ecology* 89:3400–3412. <https://doi.org/10.1890/08-0349.1>
- Gabriel KR (1982) Biplot. In: Kotz S, Johnson NL (eds) *Encyclopedia of statistical sciences*, vol 1. Wiley, New York, pp 263–271
- Gabriel KR (1998) Generalised bilinear regression. *Biometrika* 85:689–700. <https://doi.org/10.1093/biomet/85.3.689>
- Gifi A (1990) *Nonlinear multivariate analysis*. Wiley, New York. ISBN 9780471926207
- Gittins R (1985) *Canonical analysis. A review with applications in ecology*. Springer, Berlin. ISBN 978-3-642-69878-1
- Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ (2016) It's all relative: analyzing microbiome data as compositions. *Ann Epidemiol* 26:322–329. <https://doi.org/10.1016/j.annepidem.2016.03.003>

- Golub GH, Reinch C (1970) Singular value decomposition and least squares solutions. *Numerische Mathematik* 14:403–420. <https://doi.org/10.1007/BF02163027>
- Golub GH, van Loan CF (1989) *Matrix computations*. The John Hopkins University Press, Baltimore. ISBN 9780801854149
- Good IJ (1969) Some applications of singular decomposition of a matrix. *Technometrics* 11:823–831. <https://doi.org/10.2307/1266902>
- Goodman LA (1981) Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J Am Stat Assoc* 76:320–334. <https://doi.org/10.2307/2287833>
- Gourlay AR, Watson GA (1973) *Computational methods for matrix eigenproblems*. Wiley, New York. ISBN 0471319155
- Gower JC, Hand DJ (1996) *Biplots*. Chapman, London. ISBN 9780412716300
- Greenacre M, Lewi P (2009) Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio-scale measurements. *J Classif* 26:29–54. <https://doi.org/10.1007/s00357-009-9027-y>
- Greenacre MJ (1984) *Theory and applications of correspondence analysis*. Academic Press, London. ISBN 978-0122990502
- Hill MO (1973) Reciprocal averaging: an eigenvector method of ordination. *J Ecol* 61:237–249. <https://doi.org/10.2307/2258931>
- Hill MO (1974) Correspondence analysis: a neglected multivariate method. *Appl Stat* 23:340–354. <https://doi.org/10.2307/2347127>
- Ihm P, van Groenewoud H (1975) A multivariate ordering of vegetation data based on Gaussian type gradient response curves. *J Ecol* 63:767–777. <https://doi.org/10.2307/2258600>
- Ihm P, van Groenewoud H (1984) Correspondence analysis and Gaussian ordination. *Compstat Lect* 3:5–60
- Jamil T, Ozinga WA, Kleyer M, ter Braak CJF (2013) Selecting traits that explain species–environment relationships: a generalized linear mixed model approach. *J Veg Sci* 24:988–1000. <https://doi.org/10.1111/j.1654-1103.2012.12036.x>
- Kleyer M et al (2012) Assessing species and community functional responses to environmental gradients: which multivariate methods? *J Veg Sci* 23:805–821. <https://doi.org/10.1111/j.1654-1103.2012.01402.x>
- Lavorel S, Rochette C, Lebreton J-D (1999) Functional groups for response to disturbance in mediterranean old fields. *Oikos* 84:480–498. <https://doi.org/10.2307/3546427>
- Lavorel S, Touzard B, Lebreton J-D, Clément B (1998) Identifying functional groups for response to disturbance in an abandoned pasture. *Acta Oecologica* 19:227–240. [https://doi.org/10.1016/S1146-609X\(98\)80027-1](https://doi.org/10.1016/S1146-609X(98)80027-1)
- Lebreton JD, Chessel D, Prodon R, Yoccoz N (1988a) L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. I. Variables de milieu quantitatives. *Acta Oecologia Generalis* 9:53–67
- Lebreton JD, Chessel D, Richardot-Coulet M, Yoccoz N (1988b) L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. II. Variables de milieu qualitatives. *Acta Oecologia Generalis* 9:137–151
- Legendre L, Legendre P (2012) *Numerical ecology*. Elsevier, Amsterdam. ISBN 9780444538680
- Legendre P, Galzin RG, Harmelin-Vivien ML (1997) Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology* 78:547–562. <https://doi.org/10.2307/2266029>
- Magnus JR, Neudecker H (1988) *Matrix differential calculus with applications in statistics and econometrics*. Wiley, New York. ISBN 978-0471986331
- Mardia KV, Kent JT, Bibby JM (1980) *Multivariate analysis*. Academic Press, London. ISBN 9780124712522
- McCune B (2015) The front door to the fourth corner: variations on the sample unit \times trait matrix in community ecology. *Commun Ecol* 16:267–271. <https://doi.org/10.1556/168.2015.16.2.14>
- Oksanen J et al. (2013) *vegan: Community Ecology Package*. R Package. version 20-9. <https://CRAN.R-project.org/package=vegan>
- Peres-Neto PR, Dray S, ter Braak CJF (2017) Linking trait variation to the environment: critical issues with community-weighted mean correlation resolved by the fourth-corner approach. *Ecography* 40:806–816. <https://doi.org/10.1111/ecog.02302>
- R Core Team (2015) *R: a language and environment for statistical computing*, version 3.0. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org

- Rui Alves M, Beatriz Oliveira M (2004) Predictive and interpolative biplots applied to canonical variate analysis in the discrimination of vegetable oils by their fatty acid composition. *J Chemom* 18:393–401. <https://doi.org/10.1002/cem.884>
- Takane Y (2013) *Constrained principal component analysis and related techniques*. Chapman and Hall/CRC, London. ISBN 9781466556669
- Tenenhaus M, Young FW (1985) An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika* 50:91–119. <https://doi.org/10.1007/bf02294151>
- ter Braak CJF (1985) Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics* 41:859–873. <https://doi.org/10.2307/1938672>
- ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167–1179. <https://doi.org/10.2307/1938672>
- ter Braak CJF (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69:69–77. <https://doi.org/10.1007/BF00038688>
- ter Braak CJF (1988) Partial canonical correspondence analysis. In: Bock HH (ed) *Classification and related methods of data analysis*. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, pp 551–558. <http://edepot.wur.nl/241165>
- ter Braak CJF (1990) Interpreting canonical correlation analysis through biplots of structural correlations and weights. *Psychometrika* 55:519–531. <https://doi.org/10.1007/BF02294765>
- ter Braak CJF (2014) History of canonical correspondence analysis. In: Blasius J, Greenacre M (eds) *Visualization and verbalization of Data*. Chapman and Hall, London, pp 61–75. <http://edepot.wur.nl/302963>
- ter Braak CJF (2017) Fourth-corner correlation is a score test statistic in a log-linear trait-environment model that is useful in permutation testing. *Environ Ecol Stat* 24:219–242. <https://doi.org/10.1007/s10651-017-0368-0>
- ter Braak CJF, Peres-Neto P, Dray S (2017) A critical issue in model-based inference for studying trait-based community assembly and a solution. *PeerJ* 5:e2885. <https://doi.org/10.7717/peerj.2885>
- ter Braak CJF, Prentice IC (1988) A theory of gradient analysis. *Adv Ecol Res* 18:271–317. [https://doi.org/10.1016/S0065-2504\(08\)60183-X](https://doi.org/10.1016/S0065-2504(08)60183-X)
- ter Braak CJF, Šmilauer P (2012) *Canoco reference manual and user's guide: software for ordination, version 5.0*. Microcomputer Power, Ithaca, USA
- ter Braak CJF, Verdonschot PFM (1995) Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquat Sci* 57:255–289. <https://doi.org/10.1007/BF00877430>
- Tso MK-S (1981) Reduced-rank regression and canonical analysis. *J R Statist Soc B* 43:183–189. <http://www.jstor.org/stable/2984847>
- Warton DI, Shipley B, Hastie T (2015) CATS regression: a model-based approach to studying trait-based community assembly. *Methods Ecol Evol* 6:389–398. <https://doi.org/10.1111/2041-210x.12280>

Cajo J. F. ter Braak is personal professor at Biometris, Wageningen University and Research with interest in multivariate analysis and Bayesian computing.

Petr Šmilauer teaches students of biological sciences at the University of South Bohemia in Ceske Budejovice how to use statistical methods. Cajo ter Braak and Petr Šmilauer are the authors of the software package Canoco for visualization of multivariate ecological and environmental data.

Stéphane Dray is senior researcher at CNRS, Biometry and Evolutionary Biology Lab, University Lyon 1. He is an author of the ade4 package for multivariate analysis and his research focuses on the development of new tools devoted to the analysis of ecological data.