Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 228 (2004) 523–537
Journal of Theoretical Biology
www.elsevier.com/locate/jtbi

From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis

Sandrine Pavoine*,*, Anne-Béatrice Dufour*, Daniel Chessel*

* *Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Claude Bernard LYON I, 43, boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France*

Received 18 June 2003; received in revised form 28 November 2003; accepted 9 February 2004

Abstract

This paper presents a new ordination method to compare several communities containing species that differ according to their taxonomic, morphological or biological features. The objective is first to find dissimilarities among communities from the knowledge about differences among their species, and second to describe these dissimilarities with regard to the feature diversity within communities. In 1986, Rao initiated a general framework for analysing the extent of the diversity. He defined a diversity coefficient called quadratic entropy and a dissimilarity coefficient and proposed a decomposition of this diversity coefficient in a way similar to ANOVA. Furthermore, Gower and Legendre (1986) built a weighted principal coordinate analysis. Using the previous context, we propose a new method called the double principal coordinate analysis (DPCoA) to analyse the relation between two kinds of data. The first contains differences among species (dissimilarity matrix); the second the species distribution among communities (abundance or presence/absence matrix). A multidimensional space assembling the species points and the community points is built. The species points define the original differences between species and the community points define the deduced differences between communities. Furthermore, this multidimensional space is linked with the diversity decomposition into between-community and within-community diversities. One looks for axes that provide a graphical ordination of the communities and project the species onto them. An illustration is proposed comparing bird communities which live in different areas under mediterranean bioclimates. Compared to some existing methods, the double principal coordinate analysis can provide a typology of communities taking account of an abundance matrix and can include dissimilarities among species. Finally, we show that such an approach generalizes some of these methods and allows us to develop new analyses.

Keywords: Dissimilarity; Diversity; Quadratic entropy; PCoA

1. Introduction

Studying communities for species composition is a central topic in ecology. According to Gittins (1985), when several communities are compared, we can study (1) the relationships between variables characterizing species and variables characterizing communities, (2) the structure, i.e. the way the data set is organized or built. Communities are defined as collections of species found in the same habitat. The study of relationship was well discussed by Dolédec et al. (1996); Legendre et al. (1997) and Legendre and Legendre (1998, p. 565–574). The study of the structure is the aim of this paper and a new method based on Rao's axiomatization is proposed.

The data are composed of two matrices (Fig. 1): the first contains distances or dissimilarities between species; the second contains abundance (or presence/absence) of species (row) in communities (column). Differences among species are assessed, for example, according to their taxonomy (Izsak and Papp, 1995; Warwick and Clarke, 1995), their morphology (Blondel et al., 1984; Cody and Mooney, 1978; Losos, 1992), or their biological traits (Lamouroux et al., 2002). The dissimilarities are evaluated using indices either directly from practical or experimental observations or indirectly from an observed data matrix.

Many studies raise the question of distinguishing differences between communities from differences

*Corresponding Author. Tel.: 33-4-72-43-27-57; Fax: 33-4-72-43-13-88.
E-mail addresses: pavoine@biomserv.univ-lyon1.fr (S. Pavoine), dufour@biomserv.univ-lyon1.fr (A.-B. Dufour), chessel@biomserv.univ-lyon1.fr (D. Chessel).

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.02.014

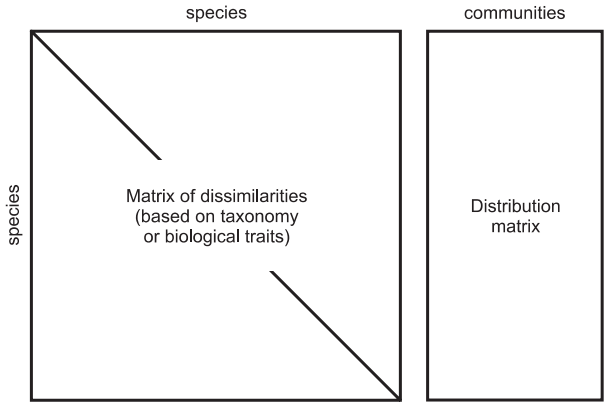

Fig. 1. Original data.

between species (Lande, 1996; Whittaker, 1972). The species diversity indices are usually calculated as the relationship of the number of species (richness) to the number of individuals per species (abundance) for a given community. The most common species indices are Gini-Simpson diversity (Gini, 1912; Simpson, 1949) and Shannon information (Shannon, 1948). However, in using species diversity, they did not calculate the differences between species. We define the term "feature diversity" to denote indices including differences among species in one or several biological traits. While properties and the biological meaning of Shannon information have been argued by many authors (see for instance Hurlbert, 1971; Lande, 1996; Rao, 1982), many developments of Gini–Simpson index have been made (e.g. Hendrickson and Ehrlich, 1971; Rao, 1982; Warwick and Clarke, 1995). These authors make innovations by introducing differences between species in the within-community diversity measures.

A new approach is also possible. Gimaret-Carpentier et al. (1998) and Pélissier et al. (2003) recently showed that ordination techniques and several usual species diversity measurements could be related. They demonstrated that inertia measurements for correspondence analysis and non-symmetric correspondence analysis equate common species diversity indices such as species richness, Gini–Simpson diversity and Shannon information. Our subject matter is in line with these ordination methods.

A new method, called the double principal coordinate analysis (DPCoA), is proposed. Its main objective is to obtain a community typology from the heterogeneity of species identities, but also from differences between species and from the relative abundances of each species. First of all, we present the origin of the (DPCoA) using Rao's axiomatization (1986). We compare this with other ordination methods, revealing its originality. Finally, we give an ecological example: comparison of bird communities in three regions under mediterranean bioclimates and a control region under temperate bioclimate (Blondel et al., 1984).

2. Rao's axiomatization

For years, the question of diversity measurements has been producing an incredible variety of solutions coming from ecology, population biology, genetics, molecular biology, and now molecular ecology. This is why a general framework must be defined. Rao (1986) was the first to begin such research. He characterized the measure of diversity of a distribution, defined a diversity coefficient called quadratic entropy and a dissimilarity coefficient. Finally, he proposed a decomposition of the diversity coefficient in a way similar to ANOVA (Fisher, 1925).

2.1. The measure of diversity of a distribution

The definition of the measure of diversity of a distribution comes from Rao's axiomatization (1986), based on a convex set $\mathbb{P}$ of probability distributions, i.e.

$$\forall P\in\mathbb{P}, \forall Q\in\mathbb{P}, \forall \alpha\in[0,1], \quad (\alpha P+(1-\alpha)Q)\in\mathbb{P}.$$

Let $\mathbb{P}$ be the following convex set of frequency distributions:

$$\mathbb{P}=\left\{P=(P_1,\ldots,P_n), P_k\geqslant 0, \sum_{k=1}^{n} P_k=1\right\}.$$

To be characterized as a measure of diversity of a distribution, H, a real-valued function defined on $\mathbb{P}$, has to verify at least two axioms:

First H must be nonnegative,

$$H(P)\geqslant 0 \quad \forall P\in\mathbb{P}.$$

And secondly H must be concave,

$$\forall P\in\mathbb{P},\ \forall Q\in\mathbb{P},\ \forall \mu_1\in\mathbb{R}^+,\ \forall \mu_2\in\mathbb{R}^+,\ \mu_1+\mu_2=1, \quad H(\mu_1 P+\mu_2 Q)\geqslant \mu_1 H(P)+\mu_2 H(Q).$$

In concrete terms, this last property of H means that diversity increases by mixing.

2.2. The diversity and dissimilarity coefficients

Consider r communities and n different species. The frequency distribution of the species in the community j is denoted by the probability vector $\mathbf{p}_j=(p_{1/j},\ldots,p_{n/j})\left(\mathbf{p}_j\in\mathbb{P}\right)$. Rao (1982) defined a diversity coefficient (DIVC), also called quadratic entropy, by

$$H_{\mathbf{\Delta}_n}(\mathbf{p}_j)=\sum_{k=1}^{n}\sum_{l=1}^{n} p_{k/j}p_{l/j}\delta_{kl}^{SP}. \tag{1}$$

$\mathbf{\Delta}_n=\left[\delta_{kl}^{SP}\right]_{1\leqslant k\leqslant n,1\leqslant l\leqslant n}$ is the $n\times n$ symmetric matrix containing dissimilarities between species, where $\delta_{kk}^{SP}=0$ for all k and $\delta_{kl}^{\mathrm{SP}}>0$ for all k and $l\neq k$. $\delta_{kl}^{\mathrm{SP}}$ is a conditionally negative definite function so that $H_{\mathbf{\Delta}}$ is concave (Rao, 1986).

Rao's DIVC may be considered as an expansion of the Gini-Simpson index (Gini, 1912; Simpson, 1949). In fact, if $\delta_{kl}^{SP} = 1$ for all $l \neq k$, then

$$H_{\Delta_n}(\mathbf{p}_j) = \sum_{k=1}^{n} \sum_{l=1, l\neq k}^{n} p_{k/j} p_{l/j} = 1 - \sum_{k=1}^{n} p_{k/j}^2,$$

is the Gini–Simpson index. In the literature, a great number of diversity indices which have been calculated from a distance or dissimilarity matrix can be translated in DIVC form. For example, we find that the Hendrickson and Ehrlich index (1971) is a non-biased version of Rao's DIVC. In fact, it corresponds to Rao's DIVC multiplied by a constant depending on the total number of species. The Warwick and Clarke index (1995) called taxonomic diversity turns out to be a special use of the Hendrickson and Ehrlich index with an arbitrary taxonomic dissimilarity.

Recently many authors have suggested the use of Rao's DIVC in order to assess the diversity within communities, taking into account differences between species. Izsak and colleagues (Izsak and Papp, 1995, 2000; Izsak and Szeidl, 2002) proposed to use Rao's index by integrating taxonomic dissimilarities between species. These dissimilarities are arbitrarily obtained from a taxonomic tree and are equal to those proposed by Warwick and Clarke (1995). In microbial ecology, Watve and Gangal (1996) suggested that indices should consider taxonomic dissimilarities. From this prospect, they introduced an index derived from Rao's DIVC that includes genetic dissimilarities between pairs of isolates. In ecology, Shimatani (2001) suggested the use of Rao's index by integrating between-species taxonomic dissimilarities and dissimilarities based on the study of amino acids. He outlined the link between Rao's index and Warwick and Clarke's index. By applying this index to tree populations with the goal of observing the effects of thinning operation for promoting survival of specific desirable species, he concluded that it can be expected that biodiversity indices incorporating species differences have more applications in ecology.

Rao (1982) went a step further by suggesting a unifying approach of diversity and dissimilarity measures. He introduced a dissimilarity coefficient (DISC) between two communities i and j with the respective species frequency vectors $\mathbf{p}_i$ and $\mathbf{p}_j$:

$$D_{H_{\Delta_n}}\left(\mathbf{p}_i, \mathbf{p}_j\right) = 2H_{\Delta_n}\left(\frac{\mathbf{p}_i + \mathbf{p}_j}{2}\right) - H_{\Delta_n}(\mathbf{p}_i) - H_{\Delta_n}\left(\mathbf{p}_j\right). \tag{2}$$

2.3. Decomposition of quadratic entropy

Rao (1982, 1984, 1986) then proposed a decomposition of quadratic entropy in a way similar to ANOVA. Let $\boldsymbol{\mu} = (\mu_1, ..., \mu_r) (\boldsymbol{\mu} \in \mathbb{P})$ be the community weight vector. Let $\mathbf{p}_\bullet = (p_{1\bullet}, ..., p_{n\bullet})$ $(\mathbf{p}_\bullet = \sum_{i=1}^{r} \mu_i \mathbf{p}_i,\ (\mathbf{p}_\bullet \in \mathbb{P}))$ be the species frequency vector in mixed communities, i.e. in the whole data set. This decomposition is

$$H_{\Delta_n}(\mathbf{p}_\bullet) = \sum_{i=1}^{r} \mu_i H_{\Delta_n}(\mathbf{p}_i) + \sum_{i=1}^{r} \sum_{j=1}^{r} \mu_i \mu_j D_{H_{\Delta_n}}\left(\mathbf{p}_i, \mathbf{p}_j\right).$$

The total quadratic entropy $H_{\Delta_n}(\mathbf{p}_\bullet)$ is divided into the within-community quadratic entropy (the first term) and the between-community quadratic entropy (second term).

The interest of these two partitions is to assess the part of the total quadratic entropy due to differences among the communities compared to the differences among the species within communities. These measures can be quantified, for example, to identify communities of high conservation value. Such communities have a great internal diversity and are very different from the remaining communities.

3. Description of the DPCoA

3.1. Entries

Recall that the data to be analysed arise in two matrices.

Consider first $\mathbf{A} = \left[a_{kj}\right]_{1\leqslant k\leqslant n, 1\leqslant j\leqslant r}$, in which a_{kj} is the abundance of the species k in the community j and $\mathbf{P} = \left[p_{kj}\right]_{1\leqslant k\leqslant n, 1\leqslant j\leqslant r}$, in which p_{kj} is the frequency of the species k in the community j. These matrices are linked by the following relations:

$$\left.\begin{array}{c} a_{\bullet j} = \sum_{k=1}^{n} a_{kj} \\ a_{k\bullet} = \sum_{j=1}^{r} a_{kj} \\ a_{\bullet\bullet} = \sum_{j=1}^{r} a_{\bullet j} = \sum_{k=1}^{n} a_{k\bullet} \end{array}\right\} \Rightarrow \left\{\begin{array}{l} p_{\bullet j} = a_{\bullet j}/a_{\bullet\bullet}, \\ p_{k\bullet} = a_{k\bullet}/a_{\bullet\bullet}, \\ p_{kj} = a_{kj}/a_{\bullet\bullet}. \end{array}\right.$$

Let $\mathbf{D}_n = diag(p_{1\bullet}, ..., p_{n\bullet})$ and $\mathbf{D}_r = diag(p_{\bullet 1}, ..., p_{\bullet r})$ be the diagonal matrices containing the marginal weighting associated with $\mathbf{P}$. Species and communities each have a natural weighting matrix, $\mathbf{D}_n$ and $\mathbf{D}_r$, respectively.

Consider secondly $\boldsymbol{\Delta}_n = \left[\delta_{kl}^{SP}\right]_{1\leqslant k\leqslant n, 1\leqslant l\leqslant n}$ the $n \times n$ matrix containing dissimilarities between species. We choose a Euclidean matrix because this type of matrix is associated with a typology and because $\left(\delta_{kl}^{SP}\right)^2$ is a conditionally negative definite function. We show the importance of this last property in section 6. $\boldsymbol{\Delta}_n$ is said to be Euclidean if and only if n points M_k $(k = 1, 2, ..., n)$ can be embedded in a Euclidean space such that the Euclidean distance between M_k and M_l is δ_{kl}^{SP}, i.e. $\delta_{kl}^{SP} = |M_k - M_l|$ (Gower and Legendre, 1986). This, of course, implies that δ_{kl}^{SP} must be nonegative. We consider that calling $\boldsymbol{\Delta}_n = \left[\delta_{kl}^{SP}\right]$ Euclidean is synonymous with stating that δ_{kl}^{SP} has Euclidean properties. These two expressions are used in the following process.

3.2. Building a common space

The n points M_k can be obtained by a principal coordinate analysis (PCoA). In PCoA, a projector $\mathbf{Q}$ centers the scatter of points. This projector is usually equal to $\mathbf{Q} = \mathbf{I}_n - (1/n)\mathbf{1}_n\mathbf{1}_n^t$, and gives a uniform weighting to the species. From now on, the weighting inserted in the centering projector is arbitrary (Gower, 1982, 1984; Gower and Legendre, 1986), and leads to a weighted PCoA. The theoretical reasons for this freedom were discussed by d'Aubigny (1989), but do not seem to have provided any concrete result. We decided to use this weighting to DPCoA.

Let us denote $\mathbf{Q} = \mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^t\mathbf{D}_n$ the new projector, and $\mathbf{\Omega}_n = \left[\left(\delta_{kl}^{SP}\right)^2/2\right]_{1\leqslant k\leqslant n, 1\leqslant l\leqslant n}$, where $\mathbf{I}_n$ is the $n\times n$ identity matrix, and $\mathbf{1}_n$ is the unit n-vector. The following matrix $-\mathbf{D}_n^{1/2}\mathbf{Q}\mathbf{\Omega}_n\mathbf{Q}^t\mathbf{D}_n^{1/2}$ is built. Let $\mathbf{\Lambda}$ and $\mathbf{U}$ be the eigenvalues and eigenvectors of this matrix. We can write

$$
\begin{aligned}
&-\mathbf{D}_n^{1/2}\mathbf{Q}\mathbf{\Omega}_n\mathbf{Q}^t\mathbf{D}_n^{1/2} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t\\
&\Rightarrow -\mathbf{Q}\mathbf{\Omega}_n\mathbf{Q}^t = \mathbf{D}_n^{-1/2}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^t\mathbf{D}_n^{-1/2}\\
&= \mathbf{D}_n^{-1/2}\mathbf{U}\mathbf{\Lambda}^{1/2}\left(\mathbf{D}_n^{-1/2}\mathbf{U}\mathbf{\Lambda}^{1/2}\right)^t = \mathbf{X}\mathbf{X}^t.
\end{aligned}
$$

The rows of the obtained matrix $\mathbf{X}$ give the coordinates of the species. The Euclidean distance between rows k and l of $\mathbf{X}$ provides exactly δ_{kl}^{SP}. Therefore, the communities may be represented by points whose coordinates are given by $\mathbf{Y} = \mathbf{D}_r^{-1}\mathbf{P}^t\mathbf{X}$. According to $\mathbf{Y}$, communities are placed on the barycenter of their species points. Furthermore, the coordinates of species are then $\mathbf{D}_n$-centered and coordinates of the communities are $\mathbf{D}_r$-centered (Appendix A). We call this space shared by species and communities the "space of the double principal coordinate analysis".

3.3. Defining a typology

The principal axes of the species points enable us to obtain a typology of the species with a reduced number of dimensions. To obtain a typology of the communities with a reduced number of dimensions, we look for the orthogonal principal axes of the scatter of community points. Let $\mathbf{I}_f$ be the $f\times f$ identity matrix. The generalized singular value decomposition (GSVD) (Greenacre, 1984) of the triplet $(\mathbf{Y}, \mathbf{I}_f, \mathbf{D}_r)$, i.e. the PCA of $\mathbf{Y}$ weighted by $\mathbf{D}_r$, gives the principal axes of community points. These axes are contained in a $f\times g$ matrix $\mathbf{V}$ defined by

$$\mathbf{Y}^t\mathbf{D}_r\mathbf{Y} = \mathbf{X}^t\mathbf{P}\mathbf{D}_r^{-1}\mathbf{P}^t\mathbf{X} = \mathbf{V}\mathbf{\Psi}\mathbf{V}^t,$$

where $\mathbf{\Psi}$ contains the eigenvalues of $\mathbf{Y}^t\mathbf{D}_r\mathbf{Y}$. Each of these axes sequentially explains as much of the variance of the community points as possible and the amount of the variance explained by an axis is given by its associated eigenvalue. We choose two (or more) of these principal axes and project on them the species and community points as well as the principal axes of the scatter of species points, so that their coordinates are given by the rows of $\mathbf{XV}$, $\mathbf{YV}$, and $\mathbf{I}_f\mathbf{V} = \mathbf{V}$, respectively.

This methodology leads to the representations of the dissimilarities between communities on which species are positioned.

4. DPCoA and apportionment of quadratic entropy

4.1. Distances among the community points

Changing $\mathbf{\Delta}_n = \left[\delta_{kl}^{SP}\right]_{1\leqslant k\leqslant n, 1\leqslant l\leqslant n}$ for $\mathbf{\Omega}_n = \left[\left(\delta_{kl}^{SP}\right)^2/2\right]_{1\leqslant k\leqslant n, 1\leqslant l\leqslant n}$ leads to the following diversity index according to Rao's DIVC:

$$H_{\mathbf{\Omega}_n}(\mathbf{p}_j) = \frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n} p_{k/j}p_{l/j}\left(\delta_{kl}^{SP}\right)^2 = \mathbf{p}_j^t\mathbf{\Omega}_n\mathbf{p}_j. \tag{3}$$

Let $\mathbf{\Delta}_r = \left[\delta_{ij}^{CO}\right]_{1\leqslant i\leqslant r, 1\leqslant j\leqslant r}$ be the matrix containing dissimilarities between the communities and $\mathbf{\Omega}_r = \left[\left(\delta_{ij}^{CO}\right)^2/2\right]_{1\leqslant i\leqslant r, 1\leqslant j\leqslant r}$. We define δ_{ij}^{CO} by

$$\delta_{ij}^{CO} = \sqrt{2\left(2H_{\mathbf{\Omega}_n}\left(\frac{\mathbf{p}_i+\mathbf{p}_j}{2}\right) - H_{\mathbf{\Omega}_n}(\mathbf{p}_i) - H_{\mathbf{\Omega}_n}\left(\mathbf{p}_j\right)\right)}. \tag{4}$$

This expression can be rewritten with matrices:

$$\delta_{ij}^{CO} = \sqrt{\left(\mathbf{p}_i-\mathbf{p}_j\right)^t(-\mathbf{\Omega}_n)\left(\mathbf{p}_i-\mathbf{p}_j\right)}. \tag{5}$$

If $\mathbf{\Delta}_n = \left[\delta_{kl}^{SP}\right]_{1\leqslant k\leqslant n, 1\leqslant l\leqslant n}$ is a Euclidean matrix, then $H_{\mathbf{\Omega}_n}$ is concave (Rao and Nayak, 1985) and $\mathbf{\Delta}_r = \left[\delta_{ij}^{CO}\right]_{1\leqslant i\leqslant r, 1\leqslant j\leqslant r}$ is Euclidean (Champely and Chessel, 2002). The concavity assures that $H_{\mathbf{\Omega}_n}$ can be partitioned. Moreover δ_{ij}^{CO} is the Euclidean distance between the point of community i and the point of community j.

4.2. Decomposition of inertia

Let $\mathbf{\mu} = (p_{\bullet 1}, ..., p_{\bullet r})(\mathbf{\mu}\in\mathbb{P})$ be the community weight vector. The quadratic entropy decomposition becomes

$$H_{\mathbf{\Omega}_n}(\mathbf{p}_\bullet) = \sum_{i=1}^{r} p_{\bullet i}H_{\mathbf{\Omega}_n}(\mathbf{p}_i) + H_{\mathbf{\Omega}_r}(\mathbf{\mu}). \tag{6}$$

where $H_{\mathbf{\Omega}_r}(\mathbf{\mu}) = \sum_{i=1}^{r}\sum_{j=1}^{r} p_{\bullet i}p_{\bullet j}\left(\delta_{ij}^{CO}\right)^2/2$. The overall quadratic entropy, $H_{\mathbf{\Omega}_n}(\mathbf{p}_\bullet)$, is divided into within-community entropy, $\sum_{i=1}^{r} p_{\bullet i}H_{\mathbf{\Omega}_n}(\mathbf{p}_i)$, and between-community entropy, $H_{\mathbf{\Omega}_r}(\mathbf{\mu})$. Since $\mathbf{\Delta}_r$ is Euclidean,

H_{Ω_r} is concave. Therefore H_{Ω_r} is actually a measure of diversity.

The diversity $H_{\Omega_n}(\mathbf{p}_j)$ within a community j is the inertia (variance) of the species points weighted by $\mathbf{p}_j$ in the space of the DPCoA. The overall diversity $H_{\Omega_n}(\mathbf{p}_\bullet)$ is the inertia of all the species points weighted by $\mathbf{p}_\bullet$ in the space of the DPCoA. And finally the between-community diversity $H_{\Omega_r}(\boldsymbol{\mu})$ is the inertia of all the community points weighted by $\boldsymbol{\mu}$ in the space of the DPCoA (Appendix B).

By selecting two principal axes of the points of the communities in the DPCoA space, we obtain a two-dimensional typology of the communities and the species are positioned on this typology. These two axes explain a great part of the between-community diversity. We then give an index of the diversity within a community on the typology by an ellipse. Its center is at the community point and its amplitude depends on the positions of the species and the abundance of the species in the community.

5. Relationships between DPCoA and other ordination methods

The multivariate ordinations which are concerned with data structure can maximize the variance among communities, and allow the projection of the species on the typology of the communities. Four methods are studied in this paper: the canonical variate analysis or discriminant analysis (Fisher, 1936; Rao, 1952), the canonical analysis of principal coordinates used with a discriminant analysis (Anderson and Willis, 2003), the canonical correspondence analysis (CCA) (ter Braak, 1986, 1987) and the between-class principal component analysis (BPCA) (Dolédec and Chessel, 1987). We present succinctly these methods, their relationships and the relationships between these methods and our DPCoA.

5.1. Special use of the four methods connected with DPCoA

For a start, the canonical variate analysis or discriminant analysis (CVA) and the BPCA are used in their original meanings. The other two methods are presented in a transformed way. The CCA is usually performed from a relevé-species matrix and a relevé-environmental variables matrix (ter Braak, 1986, 1987). The goal of CCA is to maximize the variance between species by choosing ordination axes that are linear combinations of environmental variables. In this paper, CCA is performed from a species-community matrix and a species-biological variables matrix. The goal is then to maximize the variance between communities by choosing axes that are linear combinations of biological variables. The canonical analysis of principal coordinates (CAP) combined with a discriminant analysis is usually performed from a matrix of distances or dissimilarities among sites and a partition of sites into groups. Its goal is to maximize the variance between groups. Here we perform a specific application from a matrix of dissimilarities among species and a partition of species into communities.

5.2. Relationship between the four methods

The four above methods (CVA, CAP, CCA and BPCA) can be interconnected in three different ways. First, they can be decomposed into two families: the canonical (CVA-DA, CCA, CAP) and the between-class (BPCA) methods. The canonical methods studied here eliminate the redundancy between the variables characterizing species whereas the BPCA does not eliminate this redundancy. We note that BPCA is equal to redundancy analysis which is also called principal component analysis in respect to instrumental variables (Israels, 1984; Johansson, 1981; Rao, 1964; van den Wollenberg, 1977) when the explanatory variables are categorical. Secondly, all these methods differ in the original data form they consider. In CVA, CAP and BPCA, the species are divided into communities. This means that the occurrence rather than the abundance of the species is considered and that one species cannot belong to more than one community. The matrix that tells which community each species belongs to is called the indicator matrix. We distinguish this matrix from the abundance (or presence/absence) matrix. In fact, with the abundance matrix, occurrences as well as abundances of the species may be regarded and any species may belong to more than one community. Abundance matrix is used in CCA. Finally, in CVA, CCA and BPCA, the species are characterized by several variables put in a matrix for which a principal component analysis would be an appropriate separated analysis. Distances between species are implicitly computed from these matrices. Conversely, in CAP, species are distinguished with the help of any distance or dissimilarity matrix.

5.3. Affinities with DPCoA

We now study the relationships between these four methods and DPCoA. Consider that the data on the features of the species are organized in a matrix with species (row) and Gaussian variables (column). We perform an indirect distance matrix from these variables. In fact we compute the principal component analysis of the variable matrix, get the matrix containing the standardized coordinates of the species, and then compute the Euclidean distances between the rows of this matrix. The resulting distance metric eliminates the redundancy between the Gaussian variables. It is a Mahalanobis metric. We then perform a DPCoA on the resulting Mahalanobis distance matrix and the indicator

matrix. This process is exactly equal to CVA (Appendix C1).

Consider now that we have a between-species dissimilarity matrix. Suppose that we perform the PCoA of this matrix and keep some resulting orthonormal axes. If we compute the Euclidean distances between the standardized coordinates of the species on the retained axes (instead of raw species coordinates), we obtain a decorrelated dissimilarity matrix according to Anderson and Willis (2003). Performing DPCoA on the resulting distance matrix and the matrix of indicator variables is exactly similar to the computing of CAP on the raw distance matrix and the indicator matrix (Appendix C2).

Assume that data are organized in a matrix with species (row) and biological variables (column). If we simply compute the Euclidean distances between the rows of this matrix and perform DPCoA on the resulting distance matrix and the indicator matrix, the process is exactly equal to BPCA (Appendix C3).

Finally, the matrix giving species composition within communities contains abundance or presence/absence variables, i.e. a correspondence analysis would be appropriate for an ordination of this matrix. Consider that the data about the features of the species are organized in a matrix with species (row) and biological variables (column). If we compute the Mahalanobis distances between the rows of this matrix, weighting each species by its abundance in the whole data set, and perform DPCoA on the resulting weighted Mahalanobis distance matrix and the matrix of abundance (or presence/absence) variables, the process is exactly equal to CCA. The community coordinates are then linear combinations of the biological variables (Appendix C4).

All the cited ordination analyses are thus special applications of DPCoA. The advantage of DPCoA is to allow the choice of appropriate dissimilarities or distances between the species, not only Euclidean, weighted or unweighted Mahalanobis distances. The appropriate dissimilarity or distance matrix may be decorrelated or not. The main interest of DPCoA with regard to these ordination methods appears when the data lead directly to between-species distances without considering feature variables. Taxonomic or phylogenetic distances are good examples. Furthermore, our method also allows the choice of an appropriate distribution matrix: indicator matrix or abundance matrix (Table 1).

Table 1
The relationships between four ordination methods and the double principal coordinate analysis

	Dissimilarity between species	Distribution matrix of species into communities
Canonical analyses		
CAP	Euclidean properties	Indicator
CVA-DA	Mahalanobis	Indicator
CCA	Weighted Mahalanobis	Abundance
Between-class analyses		
BPCA	Euclidean	Indicator

5.4. Two other methods

Some other methods found in the literature can be linked to DPCoA. For instance, we present the non-symmetric correspondence analysis (NSCA) (Lauro and D'Ambra, 1984) and the distance-based redundancy analysis (dbRDA) (Legendre and Anderson, 1999).

In NSCA, the data are made up of a species × communities matrix. There are no distances between species. Correspondence analysis studies contingency tables or tables containing abundances (or presence/absence). We can analyse the simultaneous discrimination of the communities and the species. The result is a compromise between these two discriminations. NSCA carries out only one of the discriminations. And our interest focuses on the discrimination of the communities. So, NSCA is equal to DPCoA with an artificial matrix containing equal distances between species (Appendix C5).

In dbRDA, the data are made up of sites and groups. Ter Braak (1986) defines a site as a basic sampling unit, separated in space or time from other sites. A group is a set of sites characterized by geography or experimentation. Two matrices are built: a dissimilarity matrix for sites and an indicator matrix giving which site belongs to which group. The objective is different from ours. In fact, this method is focused on the statistical test of the difference between the groups. It performs a decomposition of diversity, whose components are used to build F-like statistics. This decomposition is the partition of the inertia in the space of the redundancy analysis applied on the principal coordinates of the dissimilarity matrix and on the indicator matrix. The decomposition of the diversity given by DPCoA is in this case that provided by the distance-based redundancy analysis (Legendre and Anderson, 1999) when only one factor is of concern (Appendix C6).

6. Illustration: bird communities and mediterranean area

Computations and graphical displays were done with R (Ihaka and Gentleman, 1996). The "dpcoa" function and the data called "ecomor" are available in the ade4 package (http://cran.r-project.org).

In the framework of the evolutionary convergence paradigm which was very popular in the seventies,

Blondel et al. (1984) evaluated the soundness of the concept of ecomorphological convergence by studying bird communities living in different parts of the world but in similar types of environments, namely three regions under mediterranean bioclimates: central Chile, California (United States), and Provence (France). These regions are compared to a control region under temperate bioclimate: Burgundy (France). In each region, these authors determined four habitats corresponding to a vegetation gradient. They tried to find a precise correspondence between equivalent habitats among the four regions in terms of structure, height and physiognomy of vegetation.

6.1. Previous studies

Blondel et al. (1984) took four kinds of information: the species composition for each community (i.e. in each habitat of each region), the foraging sites, the diet habits, and the morphometric characteristics of each species. They concluded that a morphometric convergence was not controversial on the scales of species and guilds. But, using discriminant analysis, on the scale of communities, they found that the phylogenetic relationship between the species of Burgundy and Provence seemed stronger than a hypothetical mediterranean convergence.

Schluter and Ricklefs (1993) reanalysed these data in another way. They studied the number of species found in diet categories in the three mediterranean regions by using a test similar to an analysis of variance in order to find possible effects of categories and of regions. They found a strong level of convergence of the number of species per diet category for the three mediterranean regions. By adding Burgundy to their analysis, we found that the level of convergence for the four regions was still strong (unpublished data) indicating that the nature of the convergence is not climatic.

6.2. Description of the data

We will not reopen here the question of ecomorphological convergence. To illustrate our method, we choose to analyse data structure dealing with the taxonomy, morphology and foraging substrate of the species. We denote **A** the matrix giving the species present in each community and **P** the corresponding frequency matrix. Only French regions share species. But habitats within a region share many species. The data we use here are slightly different from those of Blondel et al. (1984). Blondel informed us that the repartition of the species, shared by the two European regions, among the habitats is different according to the region. We took this fact into account.

To compute taxonomic dissimilarities between species, we assign the value 1 between two species of the same genus, 2 between two species of different genera belonging to the same family, 3 for two species of different families belonging to the same order, 4 for two species belonging to different orders. Pairwise dissimilarities between species build the following matrix $\left[\left(\delta_{kl}^{SP}\right)_{TAX}^{2}/2\right]$ so that $\mathbf{\Delta}_n^{TAX} = \left[\left(\delta_{kl}^{SP}\right)_{TAX}\right]$ is Euclidean and the diversity index given by formula (3) is exactly the index of taxonomic diversity proposed both by Izsak and Papp (1995) and Warwick and Clarke (1995).

To compute substrate dissimilarities between species, we resolve foraging sites into the six following modalities: aerial, foliage feeders, twig feeders, bush feeders, trunk feeders, ground feeders. For each species, a percentage is assigned to each modality, according to its affinities. From these percentages, we calculate the Edwards distance (Edwards, 1971) between species. The resulting distance matrix is denoted $\mathbf{\Delta}_n^{SUB}$.

To compute morphometric dissimilarities between species, we studied the following traits: wing length, tail length, total culmen length, bill height, bill width, tarsus length, length of middle toe and claw, and weight. Pairwise dissimilarities between species are estimated by working out Mahalanobis distances on the mean logarithmic morphometric measures of the species. The resulting distance matrix is denoted $\mathbf{\Delta}_n^{MOR}$.

6.3. Decomposition of diversity

For each biological and ecological trait, we compute a within-community diversity coefficient with formula (3) and the apportionment of the diversity coefficient with formula (6). On the whole, species richness, taxonomic diversity, foraging-substrate diversity and morphometric diversity increase with the complexity of the vegetation. More precisely, there are two exceptions. First, the species richness within Burgundy is roughly constant along the gradient of habitats. Second, the morphometric diversity within California decreases with the complexity of the vegetation. We also note that the within-community morphometric diversity converges in the complex habitats for the four regions (Fig. 2).

Decomposition of the diversity coefficient (Table 2) shows that the differences between communities correspond to 5%, 8% and 9% of the total diversity according to the taxonomy, morphology and foraging substrate, respectively. To test the significance of these differences, we compute the following permutation test. For each iteration, we perform a random permutation of each species occurrence across communities and calculate a statistic corresponding to the division of between-community diversity by within-community diversity. After 999 iterations, we find that only 3% of the random values are superior to the observed value with taxonomy, and less than 1% with morphology and foraging substrate. We conclude that differences

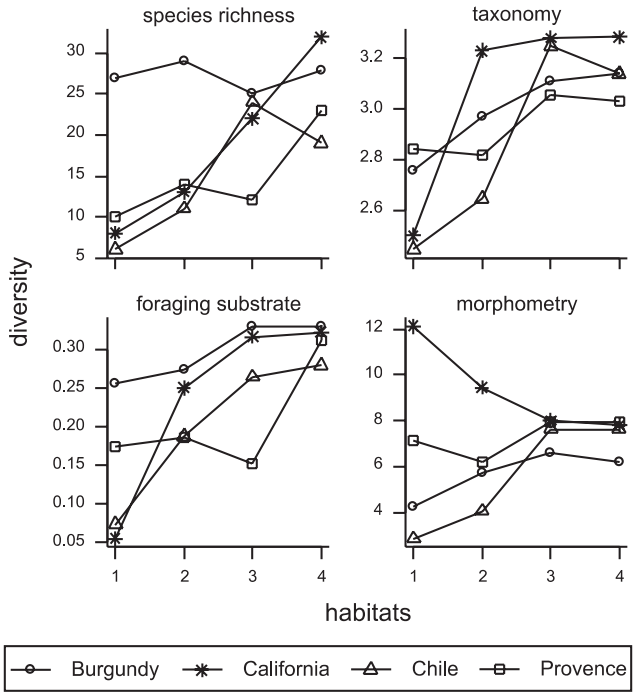

Fig. 2. Diversity patterns along the habitat gradient within regions. Four criteria are concerned: species richness, taxonomic diversity, foraging-substrate diversity and morphometric diversity.

Table 2
Apportionment of quadratic entropy

Diversity source	Taxonomy		Morphology		Foraging substrate	
Among communities	0.176	(5%)	0.587	(8%)	0.027	(9%)
Within communities	3.039	(95%)	6.893	(92%)	0.267	(91%)
Total	3.215		7.480		0.294	

between communities are significant for all dissimilarity criteria.

We also compute a matrix containing the Euclidean Jaccard dissimilarities (Gower and Legendre, 1986; Jaccard, 1901) between the rows of the presence/absence matrix **A** and perform the procruste (Jackson, 1995) and the RV tests (Heo and Gabriel, 1998) on this matrix and $\mathbf{\Delta}_n^{TAX}$, $\mathbf{\Delta}_n^{MOR}$, and $\mathbf{\Delta}_n^{SUB}$, respectively. The aim is to assess the coherence between the species structure given by **A** and the species structure given by $\mathbf{\Delta}_n^{TAX}$, $\mathbf{\Delta}_n^{MOR}$, and $\mathbf{\Delta}_n^{SUB}$, respectively. For both tests and the three dissimilarity criteria, the coherence is highly significant.

6.4. Community typology

The above results reveal differences between the communities and the DPCoA allows us to describe them. We detail the results for the taxonomy. The data are summarized in Fig. 3. The objective is to put the species occurrence matrix **A** (left part of Fig. 3) in order according to the species taxonomy (right part of Fig. 3).

We compute a DPCoA on **P** and $\mathbf{\Delta}_n^{TAX}$ (Fig. 4). The four plots in Fig. 4 are superimposable. They focus either on species points (Fig. 4(a) and (b)), on community points (Fig. 4(d)), or simultaneously on species and community points (Fig. 4(c)). The first two axes of community points explain 36% and 17% of the between-community taxonomic diversity, respectively. Fig. 4(a) and (b) inform us about the role of the species on the typology of the communities. Only three families contribute to it: Sylvidae, Emberizidae, and Picidae. Fig. 4(c) contains both species and community points and shows the distribution ellipses centered on community points. These ellipses indicate the area in which species from a given community are likely to be located. Traits also connect a community point to its species. Fig. 4(d) shows that the first axis highly separates the European communities from the American communities. And the second axis distributes communities according to the gradient of vegetation complexity. It is worth noticing that the gradient of vegetation in Burgundy extends the gradient of vegetation in Provence. So superimposing Fig. 4(b) and (d) indicates that species from Sylvidae are mainly found in the studied European regions, species from Emberizidae in the studied American regions and species from Picidae in the communities with complex vegetation regardless of the region.

We have also performed the DPCoA on P and the dissimilarity matrices: morphometry ($\mathbf{\Delta}_n^{MOR}$), foraging substrate ($\mathbf{\Delta}_n^{SUB}$) and a new matrix of equidistances between species ($\mathbf{\Delta}_n^{EQU}$). For each trait studied, we found a different pattern of differences between communities. The structure given by Fig. 5(a) is trivial. The regions are well separated, except for Burgundy and Provence which have many species in common. The foraging substrate (Fig. 5(b)) shows the stratification of the vegetation. As underlined in Fig. 4(d), this plot suggests that the gradient of vegetation in Burgundy follows the gradient of vegetation in Provence. The morphometry (Fig. 5(c)) separates the four regions and particularly the continents. Finally the taxonomy (Fig. 5(d)) simultaneously separates the continent and shows the vegetation structure.

This illustration shows the importance of integrating the differences between species into diversity indices. The DPCoA expresses the modification of the diversity pattern according to the nature of the differences between the concerned species. This illustration also shows that DPCoA can compare communities with no species in common, as for instance those of California and Chile.

7. Conclusion

In this paper, we propose a new method, the double principal coordinate analysis (DPCoA), based on a

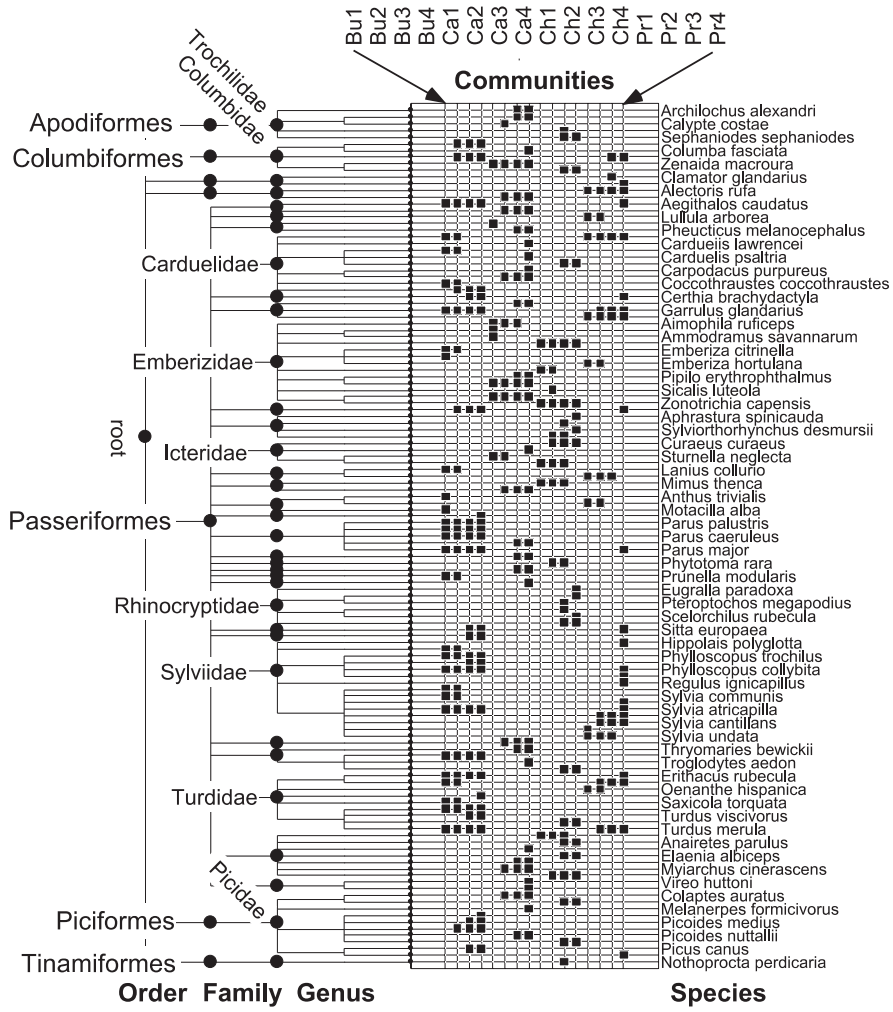

Fig. 3. Summary of the data set. On the right a representation of the species occurrence matrix with species as rows and communities as columns. Each community is labeled by the first two letters of its region name (Bu, Ca, Ch, Pr) and its rank along the gradient of increasing complex vegetation structure (1,2,3,4). For the sake of legibility, one species name out of two is indicated. A black square indicates the presence of a species. On the left: the associated species taxonomic tree.

weighted PCoA, which enables us to compare communities from two kinds of data: a matrix giving the abundance or presence/absence of the species within the communities and a matrix containing distances or dissimilarities between the species. The dissimilarity matrix is obtained either directly from experimental observations or indirectly from an observed data matrix.

We link this new method to four ordination methods. We show that the canonical correspondence analysis (CCA) and the canonical variate analysis (CVA) are particular applications of DPCoA when a Mahalanobis metric is used. This metric is interesting when correlated Gaussian variables are under study. The same is true for the between-class principal component analysis (BPCA), which is under the same pattern as CVA but is based on Euclidean distance. DPCoA allows us to choose the appropriate metric to compute distances. For example, when categorical variables are of concern, dissimilarity indices based on frequencies may be more appropriate (Manly, 1994, formula 5.8 p. 68). When a direct distance or dissimilarity matrix is under study, Anderson and Willis (2003) propose using PCoA as a first step before a CVA. We state that this same process may be used before a CCA with a weighted PCoA. These methods eliminate the redundancy in the knowledge obtained about species. This operation is not always interesting. Anderson and Willis (2003) state that the canonical analysis of principal coordinates (CAP), which uses any distance or dissimilarity matrix, takes into account the correlation structure in the response data cloud; but it provides us no information concerning the overall pattern of dispersion points in the multivariate cloud, or potential differences in multivariate variability, or dispersion among groups. Conversely, the principal component analysis or the principal coordinate analysis for the study of one matrix and BPCA for the study of two matrices is interested in the description of the overall raw data. The DPCoA thus enables us to

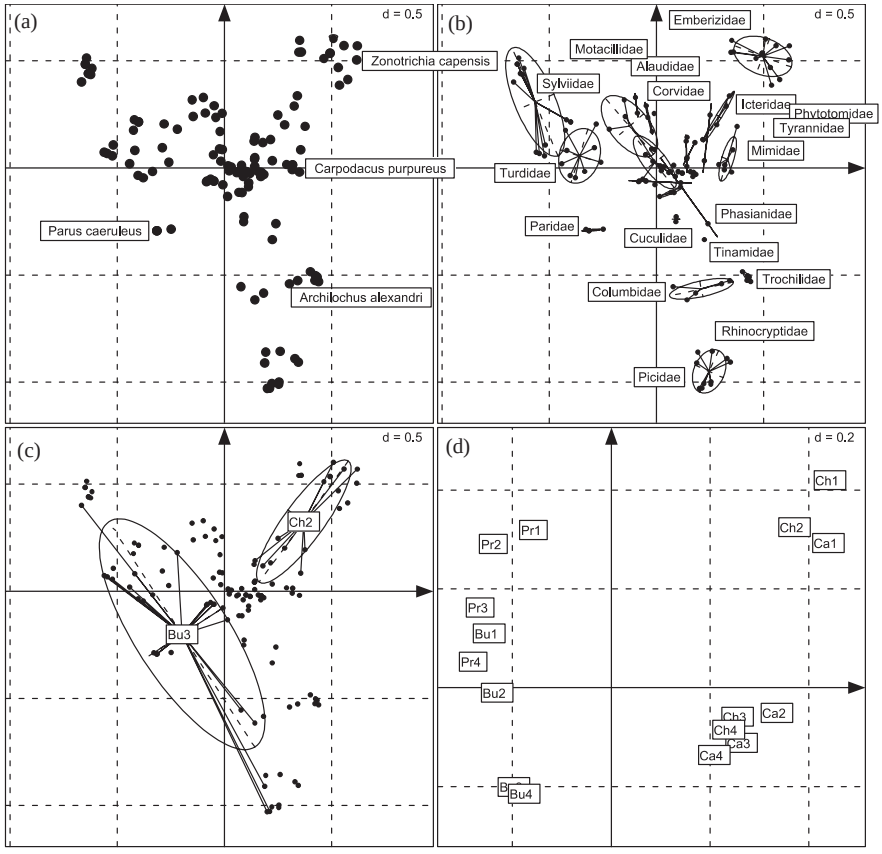

Fig. 4. DPCoA on **P** and Δ_n^{TAX}: (a) species points; (b) species points grouped by families; (c) species points grouped for two communities; (d) community points. In each figure, a grid indicates the scale; the length of a square side is indicated by the d value. Each community is labeled by the first two letters of its region's name (Bu, Ca, Ch, Pr) and its rank along the gradient of increasing complex vegetation structure (1,2,3,4). All the scatters are superimposable (by adjusting the scale if necessary). Distribution ellipses are centered on (b) family points or (c) community points. They give an index of the species distribution for each community or family. Lines connect (b) family or (c) community to their species.

develop the BPCA principle for (1) a matrix giving the abundance (or presence/absence) of the species within communities instead of an indicator matrix and (2) a matrix containing distances or dissimilarities between the species.

The second interest of the developed DPCoA is to generate all its possible extensions. In fact, the organization, showed in Table 1, allows a natural expression of new methods. In the scheme of canonical analyses, taking account of an abundance matrix and a Euclidean matrix for the distance between species leads to a new method: the CCA of weighted principal coordinate. This is equal to the computation of CCA after a weighted PCoA.

Furthermore, in the scheme of between-class analyses, three new methods can be exposed. A between-class analysis of principal coordinates can be built using Euclidean properties for distances between species and an indicator distribution matrix. A between-class correspondence analysis can be built with weighted Euclidean distance between species and an abundance distribution matrix. Likewise, a between-class correspondence analysis of weighted principal coordinates can be built with Euclidean properties for the distance between species and an abundance distribution matrix. This last method can be viewed as an extension of the between-class analysis of principal coordinates for abundance matrices instead of indicator matrices.

In this paper, we perform a descriptive study of the diversity focusing on the graphic display. DPCoA takes into account the dissimilarities between species and describes the diversity of a community (or a site) and the differences between two communities. With the same goal as Legendre, Anderson and McArdle (Anderson, 2001; Legendre and Anderson, 1999; McArdle and Anderson, 2001), we are now able to test the effects of a factor on the differences between communities by focusing on the analytical aspect instead of the graphic aspect. In this case, we have the possibility of taking into account the differences between the species in measurements of differences between communities.

The DPCoA, with related methods, can improve the analyses of diversity by providing further information concerning the overall pattern of dispersion between communities. This means that they describe the differences and the relationships among communities by

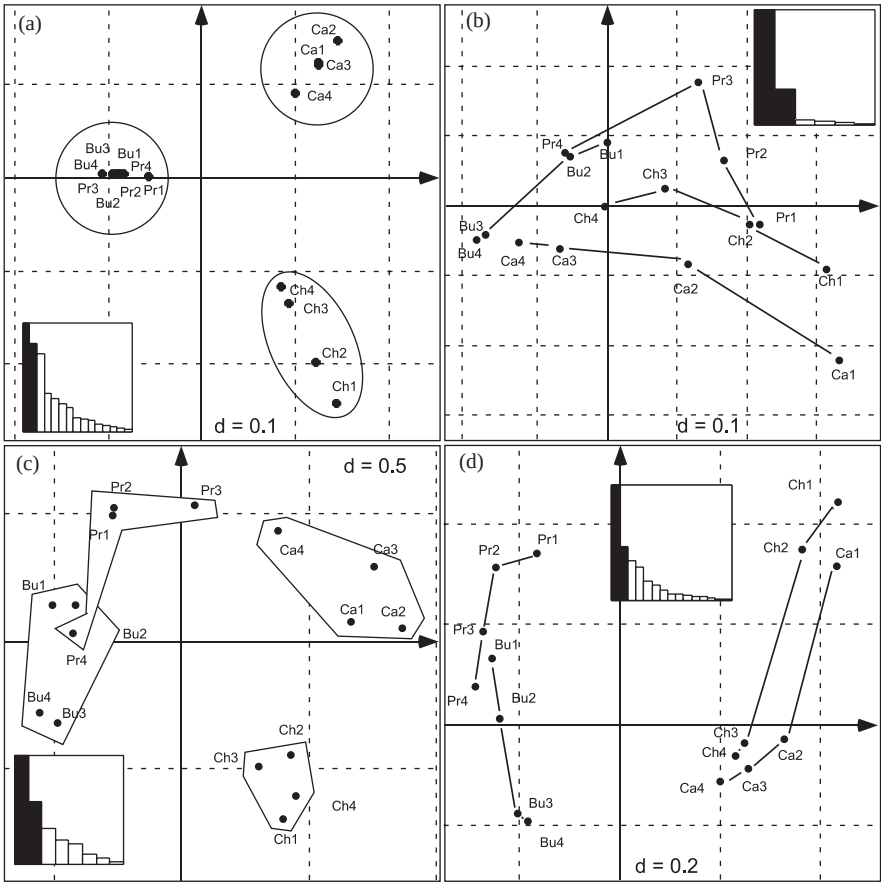

Fig. 5. Community points on the plane given by the first two principal axes of the DPCoA on **P** and (a) Δ_n^{EQU}; (b) Δ_n^{SUB}; (c) Δ_n^{MOR}, (d) Δ_n^{TAX}. In each figure, a grid indicates the scale; the length of a square side is indicated by the d value. Each community is labeled by the first two letters of its region's name (Bu, Ca, Ch, Pr) and its rank along the gradient of increasing complex vegetation structure (1,2,3,4). In Fig. 6(b) and (d), lines connect habitats from the same regions. In Fig. 6(a) and (c), ellipses or polygons surround the communities of a single region. In each figure, a box gives barplot of eigenvalues (in black, the retained axes).

showing the degree of overlap between communities in terms of features of species (taxonomy, morphology or otherwise). They also supply information about the role of each species in the differences among communities. Endemic species with special features are highlighted. Another important point is that the method is flexible enough to allow its use in any scope of study: for examples, biology, ecology, economics, and sociology. In population genetics, a similar concern arises (e.g. Nei, 1987; Weir, 1996; Weir and Cockerham, 1984; Wright, 1951, 1965). Several populations of a species may be compared according to genetic traits of their individuals (DNA sequence, fingerprinting pattern, or microsatellites). The DPCoA can provide interesting concrete results for the future.

Acknowledgements

The authors would like to thank C. Ter Braak and two anonymous referees for their relevant comments, J. Blondel for helpful discussions about data sets and R. Grantham for his councils on editorial quality. They all contributed to improving the presentation of this paper.

Appendix A. Centering points in DPCoA

*A.1. Proof that coordinates of the species (**X**) are $\mathbf{D}_n$-centered*

$$\left|\mathbf{X}^t\mathbf{D}_n\mathbf{1}_n\right|^2 = \mathbf{1}_n^t\mathbf{D}_n\mathbf{X}\mathbf{X}^t\mathbf{D}_n\mathbf{1}_n = -\mathbf{1}_n^t\mathbf{D}_n\mathbf{Q}\boldsymbol{\Omega}_n\mathbf{Q}^t\mathbf{D}_n\mathbf{1}_n$$
$$= trace\left(-\mathbf{1}_n\mathbf{1}_n^t\mathbf{D}_n\mathbf{Q}\boldsymbol{\Omega}\mathbf{Q}^t\mathbf{D}_n\right) = 0.$$

A.2. Proof that coordinates of the communities ($\mathbf{Y} = \mathbf{D}_r^{-1}\mathbf{P}^t\mathbf{X}$) are $\mathbf{D}_r$-centered

$$\mathbf{Y}^t\mathbf{D}_r\mathbf{1}_r = \mathbf{X}^t\mathbf{P}\mathbf{1}_r = \mathbf{X}^t\mathbf{D}_n\mathbf{1}_n.$$

Appendix B. Decomposition of the inertia in DPCoA

The link between this decomposition and the DPCoA can be made in the space we call DPCoA space, where

scatters of species points and community points are centered. In this space, we denote M_k the point of the species k, G_j the point of the community j. The scores of M_k are given by the kth row of $\mathbf{X}$ and denoted by $\mathbf{x}_k$. The scores of G_j are given by the jth row of $\mathbf{Y}$ and denoted by $\mathbf{y}_j$. As previously mentioned, G_j is the barycenter of the species points weighted by $\mathbf{p}_j$: $\mathbf{y}_j = \mathbf{X}^t\mathbf{p}_j$. We denote $G_\bullet$ the null coordinate point. $G_\bullet$ is the barycenter of all the species points weighted by $\mathbf{p}_\bullet$, and the barycenter of all the community points weighted by $\boldsymbol{\mu}$.

As previously mentioned, the Euclidean distance between M_k and M_l ($|M_k - M_l| = \sqrt{(\mathbf{x}_k - \mathbf{x}_l)^t(\mathbf{x}_k - \mathbf{x}_l)}$) equals δ_{kl}^{SP}. As $\mathbf{p}_j^t\mathbf{D}_n\mathbf{1}_n = 1$ for all j, formula (5) can be rewritten as

$$\delta_{ij}^{CO} = \sqrt{\left(\mathbf{p}_i - \mathbf{p}_j\right)^t(-\mathbf{Q}\boldsymbol{\Omega}_n\mathbf{Q})\left(\mathbf{p}_i - \mathbf{p}_j\right)}.$$

This can be rewritten as

$$\begin{aligned}\delta_{ij}^{CO} &= \sqrt{\left(\mathbf{p}_i - \mathbf{p}_j\right)^t(\mathbf{X}\mathbf{X}^t)\left(\mathbf{p}_i - \mathbf{p}_j\right)}\\ &= \sqrt{\left(\mathbf{X}^t\mathbf{p}_i - \mathbf{X}^t\mathbf{p}_j\right)^t\left(\mathbf{X}^t\mathbf{p}_i - \mathbf{X}^t\mathbf{p}_j\right)}\\ &= \sqrt{\left(\mathbf{y}_i - \mathbf{y}_j\right)^t\left(\mathbf{y}_i - \mathbf{y}_j\right)}\\ &= \left|G_i - G_j\right|.\end{aligned}$$

The components of diversity can thus be rewritten as follows:

$$\begin{aligned}H_{\boldsymbol{\Omega}_n}(\mathbf{p}_\bullet) &= \frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}p_{k\bullet}p_{l\bullet}\left(\delta_{kl}^{SP}\right)^2 = \frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}p_{k\bullet}p_{l\bullet}|M_k - M_l|^2\\ &= \sum_{k=1}^{n}p_{k\bullet}|M_k - G_\bullet|^2,\end{aligned}$$

$$\begin{aligned}H_{\boldsymbol{\Omega}_n}(\mathbf{p}_j) &= \frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}p_{kj}p_{lj}\left(\delta_{kl}^{SP}\right)^2 = \frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}p_{kj}p_{lj}|M_k - M_l|^2\\ &= \sum_{k=1}^{n}p_{kj}\left|M_k - G_j\right|^2,\end{aligned}$$

$$\begin{aligned}H_{\boldsymbol{\Omega}_r}(\boldsymbol{\mu}) &= \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}p_{\bullet i}p_{\bullet j}\left(\delta_{ij}^{CO}\right)^2 = \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}p_{\bullet i}p_{\bullet j}\left|G_i - G_j\right|^2\\ &= \sum_{i=1}^{r}p_{\bullet i}|G_i - G_\bullet|^2.\end{aligned}$$

Appendix C. Clarification of the relationship between DPCoA and other methods

Let $\mathbf{L}$ be an indicator matrix and $\mathbf{A}$ be an abundance (or presence/absence) matrix. Let $\mathbf{Z}$ be a matrix of centered variables. Let $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{V}_*$ be three matrices containing eigenvectors and $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ be two matrices containing eigenvalues. Note that when $\mathbf{L}$ is of concern, we have the following relations: $\mathbf{D}_n = (1/n)\mathbf{I}_n$, $\mathbf{D}_n\mathbf{L} = \mathbf{P}$ and $\mathbf{D}_r = \mathbf{L}^t\mathbf{D}_n\mathbf{L}$.

C.1. CVA

Process: Let $\mathbf{T} = \mathbf{Z}^t\mathbf{D}_n\mathbf{Z}$ be the total covariance matrix and $\mathbf{B} = \mathbf{Z}^t\mathbf{D}_n\mathbf{L}(\mathbf{D}_r)^{-1}\mathbf{D}_r(\mathbf{D}_r)^{-1}\mathbf{L}^t\mathbf{D}_n\mathbf{Z}$ the between-community covariance matrix. The CVA on $\mathbf{Z}$ and $\mathbf{L}$ is the eigenanalysis of $\mathbf{T}^{-1}\mathbf{B}$. They are computational advantages in working with the symmetric matrix $\mathbf{T}^{-1/2}\mathbf{B}\mathbf{T}^{-1/2}$, rather than $\mathbf{T}^{-1}\mathbf{B}$. The eigenvalues of $\mathbf{T}^{-1/2}\mathbf{B}\mathbf{T}^{-1/2}$ are identical to those of $\mathbf{T}^{-1}\mathbf{B}$, while its eigenvectors, $\mathbf{V}$, are connected with those of $\mathbf{T}^{-1}\mathbf{B}$, $\mathbf{V}_0$, by the relation $\mathbf{V}_0 = \mathbf{T}^{-1/2}\mathbf{V}$:

$$\begin{aligned}\mathbf{T}^{-1/2}\mathbf{B}\mathbf{T}^{-1/2} &= \left(\mathbf{Z}^t\mathbf{D}_n\mathbf{Z}\right)^{-1/2}\mathbf{Z}^t\mathbf{D}_n\mathbf{L}(\mathbf{D}_r)^{-1}\mathbf{D}_r(\mathbf{D}_r)^{-1}\\ &\quad\mathbf{L}^t\mathbf{D}_n\mathbf{Z}\left(\mathbf{Z}^t\mathbf{D}_n\mathbf{Z}\right)^{-1/2} = \mathbf{V}\boldsymbol{\Psi}\mathbf{V}^t.\end{aligned}$$

Affinity with DPCoA:

(1) Consider the GSVD of $(\mathbf{Z},\mathbf{I}_m,\mathbf{D}_n)$:

$$\mathbf{Z}^t\mathbf{D}_n\mathbf{Z} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^t.$$

Consider $\mathbf{X} = \mathbf{Z}\mathbf{U}\boldsymbol{\Lambda}^{-1/2}$. The Euclidean distances between the rows of $\mathbf{X}$ are the Mahalanobis distances between the rows of $\mathbf{Z}$. Let $\boldsymbol{\Delta}$ contain these distances so that the species coordinates provided by the PCoA of $\boldsymbol{\Delta}$ weighted by $\mathbf{D}_n$ are given by $\mathbf{X}$.

(2) The matrix $\mathbf{T}^{-1/2}\mathbf{B}\mathbf{T}^{-1/2}$ can be rewritten as

$$\begin{aligned}\mathbf{T}^{-1/2}\mathbf{B}\mathbf{T}^{-1/2} &= \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^t\mathbf{Z}^t\mathbf{P}(\mathbf{D}_r)^{-1}\mathbf{D}_r(\mathbf{D}_r)^{-1}\mathbf{P}^t\mathbf{Z}\mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^t\\ &= \mathbf{U}\mathbf{Y}^t\mathbf{D}_r\mathbf{Y}\mathbf{U}^t.\end{aligned}$$

Consider $\mathbf{V}_* = \mathbf{U}^t\mathbf{V}$, then

$$\mathbf{Y}^t\mathbf{D}_r\mathbf{Y} = \mathbf{V}_*\boldsymbol{\Psi}\mathbf{V}_*^t,$$

which is the GSVD of $(\mathbf{Y},\mathbf{I}_k,\mathbf{D}_r)$ and thus the DPCoA of $\mathbf{P}$ and $\boldsymbol{\Delta}$. The coordinates of the communities are in $\mathbf{Y}\mathbf{V}_* = (\mathbf{D}_r)^{-1}\mathbf{P}^t\mathbf{Z}\left(\mathbf{Z}^t\mathbf{D}_n\mathbf{Z}\right)^{-1/2}\mathbf{V}$, and those of the species are in $\mathbf{X}\mathbf{V}_* = \mathbf{Z}\left(\mathbf{Z}^t\mathbf{D}_n\mathbf{Z}\right)^{-1/2}\mathbf{V}$. The DPCoA on $\mathbf{P}$ and $\boldsymbol{\Delta}$ is thus equal to the CVA on $\mathbf{Z}$ and $\mathbf{L}$.

C.2. CAP

CAP computes CVA after having performed PCoA on a dissimilarity matrix. In order to find its link with DPCoA, in the previous part, replace $\mathbf{Z}$ with a matrix containing m principal coordinates of a dissimilarity matrix.

C.3. BPCA

Process: Let $\hat{\mathbf{Z}}$ be defined as

$$\hat{\mathbf{Z}} = \mathbf{L}\left(\mathbf{L}^t\mathbf{D}_n\mathbf{L}\right)^{-1}\mathbf{L}^t\mathbf{D}_n\mathbf{Z}.$$

The BPCA on $\mathbf{Z}$ and $\mathbf{L}$ is the GSVD of $(\hat{\mathbf{Z}},\mathbf{I}_m,\mathbf{D}_n)$, i.e.

$$\begin{aligned}\hat{\mathbf{Z}}^{\mathrm{t}}\mathbf{D}_n\hat{\mathbf{Z}} &= \mathbf{Z}^{\mathrm{t}}\mathbf{D}_n\mathbf{L}\left(\mathbf{L}^{\mathrm{t}}\mathbf{D}_n\mathbf{L}\right)^{-1}\mathbf{L}^{\mathrm{t}}\mathbf{D}_n\mathbf{L}\left(\mathbf{L}^{\mathrm{t}}\mathbf{D}_n\mathbf{L}\right)^{-1}\mathbf{L}^{\mathrm{t}}\mathbf{D}_n\mathbf{Z}\\ &= \mathbf{Z}^{\mathrm{t}}\mathbf{P}(\mathbf{D}_r)^{-1}\mathbf{D}_r(\mathbf{D}_r)^{-1}\mathbf{P}^{\mathrm{t}}\mathbf{Z}\\ &= \mathbf{V}\boldsymbol{\Psi}\mathbf{V}^{\mathrm{t}}.\end{aligned}$$

Affinity with DPCoA:

(1) $\boldsymbol{\Delta}$ is defined by computing the Euclidean distances between the rows of $\mathbf{Z}$.

(2) The PCoA on $\boldsymbol{\Delta}$ weighted by $\mathbf{D}_n$ is equal to the GSVD of $(\mathbf{Z},\mathbf{I}_m,\mathbf{D}_n)$:

$$\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n\mathbf{Z} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{t}}.$$

The species coordinates are thus given by the rows of $\mathbf{X} = \mathbf{Z}\mathbf{U}$.

(3) The covariance matrix $\hat{\mathbf{Z}}^{\mathrm{t}}\mathbf{D}_n\hat{\mathbf{Z}}$ can be rewritten as

$$\begin{aligned}\hat{\mathbf{Z}}^{\mathrm{t}}\mathbf{D}_n\hat{\mathbf{Z}} &= \mathbf{U}\mathbf{X}^{\mathrm{t}}\mathbf{P}(\mathbf{D}_r)^{-1}\mathbf{D}_r(\mathbf{D}_r)^{-1}\mathbf{P}^{\mathrm{t}}\mathbf{X}\mathbf{U}^{\mathrm{t}}\\ &= \mathbf{U}\mathbf{Y}^{\mathrm{t}}\mathbf{D}_r\mathbf{Y}\mathbf{U}^{\mathrm{t}}.\end{aligned}$$

This implies that

$$\mathbf{Y}^{\mathrm{t}}\mathbf{D}_r\mathbf{Y} = \mathbf{U}^{\mathrm{t}}\mathbf{V}\boldsymbol{\Psi}\mathbf{V}^{\mathrm{t}}\mathbf{U},$$

which is the GSVD of $(\mathbf{Y},\mathbf{I}_k,\mathbf{D}_r)$. So the coordinates of the communities are in $\mathbf{Y}\mathbf{U}^{\mathrm{t}}\mathbf{V} = (\mathbf{D}_r)^{-1}\mathbf{P}^{\mathrm{t}}\mathbf{X}\mathbf{U}^{\mathrm{t}}\mathbf{V} = (\mathbf{D}_r)^{-1}\mathbf{P}^{\mathrm{t}}\mathbf{Z}\mathbf{V}$, and those of the species are in $\mathbf{X}\mathbf{U}^{\mathrm{t}}\mathbf{V} = \mathbf{Z}\mathbf{V}$. The DPCoA on $\mathbf{P}$ and $\boldsymbol{\Delta}$ is thus equal to the BPCA on $\mathbf{Z}$ and $\mathbf{L}$.

C.4. CCA

Process: We define

$$\bar{\mathbf{P}} = \mathbf{D}_n^{1/2}\left(\mathbf{D}_n^{-1}\mathbf{P}\mathbf{D}_r^{-1}\right)\mathbf{D}_r^{1/2}$$

and

$$\hat{\mathbf{P}} = \mathbf{D}_n^{1/2}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n^{1/2}\bar{\mathbf{P}}.$$

CCA on $\mathbf{P}$ and $\mathbf{Z}$ is the GSVD of $(\hat{\mathbf{P}}^{\mathrm{t}},\mathbf{I}_n,\mathbf{I}_r)$:

$$\hat{\mathbf{P}}\hat{\mathbf{P}}^{\mathrm{t}} = \mathbf{V}\boldsymbol{\Psi}\mathbf{V}^{\mathrm{t}}.$$

Affinity with DPCoA:

(1) Consider the GSVD of $(\mathbf{Z},\mathbf{I}_m,\mathbf{D}_n)$:

$$\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n\mathbf{Z} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{t}}.$$

Consider $\mathbf{X} = \mathbf{Z}\mathbf{U}\boldsymbol{\Lambda}^{-1/2}$. The Euclidean distances between the rows of $\mathbf{X}$ are Mahalanobis distances between the rows of $\mathbf{Z}$ weighted by $\mathbf{D}_n$. $\boldsymbol{\Delta}$ contains these distances so that the species coordinates provided by the PCoA of $\boldsymbol{\Delta}$ weighted by $\mathbf{D}_n$ are in the rows of $\mathbf{X}$.

(3) Matrix $\hat{\mathbf{P}}$ can be rewritten as

$$\hat{\mathbf{P}} = \mathbf{D}_n^{1/2}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n^{1/2}\mathbf{D}_n^{1/2}\left(\mathbf{D}_n^{-1}\mathbf{P}\mathbf{D}_r^{-1}\right)\mathbf{D}_r^{1/2},$$

$$\Leftrightarrow \hat{\mathbf{P}} = \mathbf{D}_n^{1/2}\mathbf{Z}\mathbf{U}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^{\mathrm{t}}\mathbf{Z}^{\mathrm{t}}\mathbf{P}\mathbf{D}_r^{-1/2},$$

$$\Leftrightarrow \hat{\mathbf{P}} = \mathbf{D}_n^{1/2}\mathbf{X}\mathbf{Y}^{\mathrm{t}}\mathbf{D}_r^{1/2}.$$

The GSVD of $\hat{\mathbf{P}}\hat{\mathbf{P}}^{t}$ is

$$\hat{\mathbf{P}}\hat{\mathbf{P}}^{\mathrm{t}} = \mathbf{D}_n^{1/2}\mathbf{X}\mathbf{Y}^{\mathrm{t}}\mathbf{D}_r^{1/2}\mathbf{D}_r^{1/2}\mathbf{Y}\mathbf{X}^{\mathrm{t}}\mathbf{D}_n^{1/2} = \mathbf{V}\boldsymbol{\Psi}\mathbf{V}^{\mathrm{t}},$$

$$\Leftrightarrow \mathbf{Y}^{\mathrm{t}}\mathbf{D}_r\mathbf{Y} = \mathbf{X}^{\mathrm{t}}\mathbf{D}_n^{1/2}\mathbf{V}\boldsymbol{\Psi}\mathbf{V}^{\mathrm{t}}\mathbf{D}_n^{1/2}\mathbf{X}.$$

Consider $\mathbf{V}_* = \mathbf{X}'\mathbf{D}_n^{1/2}\mathbf{V}$, then

$$\mathbf{Y}^{\mathrm{t}}\mathbf{D}_r\mathbf{Y} = \mathbf{V}_*\boldsymbol{\Psi}\mathbf{V}_*^{\mathrm{t}},$$

which is the GSVD of $(\mathbf{Y},\mathbf{I}_k,\mathbf{D}_r)$ and thus the DPCoA of $\mathbf{P}$ and $\boldsymbol{\Delta}$. The coordinates of the communities are in

$$\begin{aligned}\mathbf{Y}\mathbf{V}_* &= \mathbf{D}_r^{-1}\mathbf{P}\mathbf{X}\mathbf{X}^{\mathrm{t}}\mathbf{D}_n^{1/2}\mathbf{V}\\ &= \mathbf{D}_r^{-1}\mathbf{P}\mathbf{Z}\left(\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n^{1/2}\mathbf{V},\end{aligned}$$

and those of the species are in

$$\begin{aligned}\mathbf{X}\mathbf{V}_* &= \mathbf{X}\mathbf{X}^{\mathrm{t}}\mathbf{D}_n^{1/2}\mathbf{V}\\ &= \mathbf{Z}\left(\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{t}}\mathbf{D}_n^{1/2}\mathbf{V}.\end{aligned}$$

The DPCoA on $\mathbf{P}$ and $\boldsymbol{\Delta}$ is thus equal to the CCA on $\mathbf{Z}$ and $\mathbf{P}$.

C.5. NSCA

Process: We define

$$\bar{\mathbf{P}} = \mathbf{P}\mathbf{D}_r^{-1} - \mathbf{D}_n\mathbf{1}_n\mathbf{1}_r^{\mathrm{t}} = \left(\mathbf{I}_n - \mathbf{D}_n\mathbf{1}_n\mathbf{1}_n^{\mathrm{t}}\right)\mathbf{P}\mathbf{D}_r^{-1}.$$

The NSCA of $\mathbf{P}$ is the GSVD of $(\bar{\mathbf{P}}^{\mathrm{t}},\mathbf{I}_n,\mathbf{D}_r)$:

$$\bar{\mathbf{P}}\mathbf{D}_r\bar{\mathbf{P}}^{\mathrm{t}} = \mathbf{V}\boldsymbol{\Psi}\mathbf{V}^{\mathrm{t}}.$$

Affinity with DPCoA:

(1) We take a matrix of equidistances among the species, say, $\boldsymbol{\Delta} = \left(\mathbf{1}_n\mathbf{1}_n^{\mathrm{t}} - \mathbf{I}_n\right)*\sqrt{2}$. The multiplicative factor $\sqrt{2}$ leads to a total inertia equal to the between-communities Gini-Simpson diversity.

(2) The PCoA on $\boldsymbol{\Delta}$ weighted by $\mathbf{D}_n$ is equal to the centered PCA of $\mathbf{I}_n$ weighted by $\mathbf{D}_n$ i.e.:

$$\left(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\mathrm{t}}\mathbf{D}_n\right)^{\mathrm{t}}\mathbf{D}_n\left(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\mathrm{t}}\mathbf{D}_n\right) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{t}}.$$

The species coordinates are given by $\mathbf{X} = \left(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\mathrm{t}}\mathbf{D}_n\right)\mathbf{U}$.

(3) $\bar{\mathbf{P}}\mathbf{D}_r\bar{\mathbf{P}}^{\mathrm{t}}$ can be rewritten as

$$\begin{aligned}\bar{\mathbf{P}}\mathbf{D}_r\bar{\mathbf{P}}^{\mathrm{t}} &= \left(\mathbf{I}_n - \mathbf{D}_n\mathbf{1}_n\mathbf{1}_n^{\mathrm{t}}\right)\mathbf{P}\mathbf{D}_r^{-1}\mathbf{D}_r\mathbf{D}_r^{-1}\mathbf{P}^{\mathrm{t}}\left(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\mathrm{t}}\mathbf{D}_n\right)\\ &= \mathbf{U}\mathbf{X}^{\mathrm{t}}\mathbf{P}\mathbf{D}_r^{-1}\mathbf{D}_r\mathbf{D}_r^{-1}\mathbf{P}^{\mathrm{t}}\mathbf{X}\mathbf{U}^{\mathrm{t}}\\ &= \mathbf{U}\mathbf{Y}^{\mathrm{t}}\mathbf{D}_r\mathbf{Y}\mathbf{U}^{\mathrm{t}}.\end{aligned}$$

Consider $\mathbf{V}_* = \mathbf{U}^{\mathrm{t}}\mathbf{V}$

$$\mathbf{Y}\mathbf{D}_r\mathbf{Y} = \mathbf{V}_*\boldsymbol{\Psi}\mathbf{V}_*^{\mathrm{t}},$$

which is the GSVD of $(\mathbf{Y},\mathbf{I}_k,\mathbf{D}_r)$ and thus the DPCoA of $\mathbf{P}$ and $\boldsymbol{\Delta}$. The coordinates of the communities are given by

$$\mathbf{Y}\mathbf{V}_* = (\mathbf{D}_r)^{-1}\mathbf{P}^{\mathrm{t}}\mathbf{X}\mathbf{U}^{\mathrm{t}}\mathbf{V} = \bar{\mathbf{P}}^{\mathrm{t}}\mathbf{V},$$

and those of the species are in

$$\mathbf{X}\mathbf{V}_* = \mathbf{X}\mathbf{U}^{\mathrm{t}}\mathbf{V} = \left(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\mathrm{t}}\mathbf{D}_n\right)\mathbf{V}.$$

Then the DPCoA on $\mathbf{P}$ and $\mathbf{\Delta}$ is equal to the NSCA on $\mathbf{P}$.

C.6. dbRDA

Let $\mathbf{\Delta}$ be a matrix of distances among replicates. Let $\mathbf{L}$ be an indicator matrix corresponding to the design of the experiment (only one factor is concerned).

Let $\mathbf{X}$ be the principal coordinates of $\mathbf{\Delta}$. Compute the redundancy analysis (i.e. BPCA) on $\mathbf{X}$ and $\mathbf{L}$. Let $\hat{\mathbf{X}}$ be defined as

$$\hat{\mathbf{X}} = \mathbf{L}\left(\mathbf{L}^{t}\mathbf{D}_{n}\mathbf{L}\right)^{-1}\mathbf{L}^{t}\mathbf{D}_{n}\mathbf{X}.$$

The BPCA on $\mathbf{X}$ and $\mathbf{L}$ is the GSVD of $(\hat{\mathbf{X}},\mathbf{I}_{m},\mathbf{D}_{n})$, i.e.

$$\begin{aligned}\hat{\mathbf{X}}^{t}\mathbf{D}_{n}\hat{\mathbf{X}} &= \mathbf{X}^{t}\mathbf{D}_{n}\mathbf{L}\left(\mathbf{L}^{t}\mathbf{D}_{n}\mathbf{L}\right)^{-1}\mathbf{L}^{t}\mathbf{D}_{n}\mathbf{L}\left(\mathbf{L}^{t}\mathbf{D}_{n}\mathbf{L}\right)^{-1}\mathbf{L}^{t}\mathbf{D}_{n}\mathbf{X}\\ &= \mathbf{X}^{t}\mathbf{P}(\mathbf{D}_{r})^{-1}\mathbf{D}_{r}(\mathbf{D}_{r})^{-1}\mathbf{P}^{t}\mathbf{X}\\ &= \mathbf{Y}^{t}\mathbf{D}_{r}\mathbf{Y}\\ &= \mathbf{V}\mathbf{\Psi}\mathbf{V}^{t}.\end{aligned}$$

Thus, as shown in Appendix C.3, this process is equal to the GSVD of $(\mathbf{Y},\mathbf{I}_{k},\mathbf{D}_{r})$ and so to the DPCoA on $\mathbf{\Delta}$ and $\mathbf{L}$. Matrices $\mathbf{X}$ and $\mathbf{Y}$ provides coordinates in DPCoA space.

The decomposition of the diversity performed in dbRDA uses three components. The first is the inertia of $\mathbf{X}$ (sum of all unconstrained eigenvalues). The second is the inertia of $\mathbf{Y}$ (sum of the canonical eigenvalues). And the last is equal to the inertia of $\mathbf{X}$ minus the inertia of $\mathbf{Y}$. This process is exactly the decomposition of the inertia in the space of the DPCoA applied to $\mathbf{\Delta}$ and $\mathbf{L}$ (cf. Appendix B).

References

Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32–46.

Anderson, M.J., Willis, T., 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. Ecology 84, 511–525.

Blondel, J., Vuilleumier, F., Marcus, L.F., Terouanne, E., 1984. Is there ecomorphological convergence among mediterranean bird communities of Chile, California, and France. In: Hecht, M.K., Wallace, B., MacIntyre, R.J. (Eds.), Evolutionary Biology. Plenum Press, New York, pp. 141–213.

Champely, S., Chessel, D., 2002. Measuring biological diversity using Euclidean metrics. Environ. Ecol. Stat. 9, 167–177.

Cody, M.L., Mooney, H.A., 1978. Convergence versus nonconvergence in mediterranean-climate ecosystems. Annu. Rev. Ecol. Syst. 9, 265–321.

Dolédec, S., Chessel, D., 1987. Rhythmes saisonniers et composantes stationnelles en milieu aquatique I– Description d'un plan d'observation complet par projection de variables. Acta Oecol.—Oecol. Gener. 8, 403–426.

Dolédec, S., Chessel, D., Ter Braak, C.J.F., Champely, S., 1996. Matching species traits to environmental variables: a new three-table ordination method. Environ. Ecol. Stat. 3, 143–166.

Drouet d'Aubigny, G. 1989. L'analyse multidimensionelle des données de dissimilarité. Thèse de Doctorat, Université Grenoble 1.

Edwards, A.W.F., 1971. Distance between populations on the basis of gene frequencies. Biometrics 27, 873–881.

Fisher, R.A., 1925. Statistical Methods for Research Workers. Oliver & Boyd, Edinburgh.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7, 179–188.

Gimaret-Carpentier, C., Chessel, D., Pascal, J.-P., 1998. Non-symmetric correspondence analysis: an alternative for species occurences data. Plant Ecol. 138, 97–112.

Gini, C. 1912. Variabilitá e mutabilitá. Studi economicoaguridici delle facoltta di giurizprudenza dell, Universite di Cagliari III, Parte II.

Gittins, R., 1985. Canonical Analysis: A Review with Applications in Ecology. Springer, Berlin.

Gower, J.C., 1982. Euclidean distance geometry. Math. Scientist 7, 1–14.

Gower, J.C., 1984. Distance matrices and their Euclidean approximation. In: Diday, E., Jambu, M., Lebart, L., Pagès, J., Tomassone, R. (Eds.), Data Analysis and Informatics III. Elsevier, North-Holland, Amsterdam, pp. 3–21.

Gower, J.C., Legendre, P., 1986. Metric and Euclidean properties of dissimilarity coefficients. J. Classif. 3, 5–48.

Greenacre, M.J., 1984. Theory and Applications of Correspondence Analysis. Academic Press, London.

Hendrickson, J.A.J., Ehrlich, P.R., 1971. An expanded concept of "species diversity". Notulae Naturae 439, 1–6.

Heo, M., Gabriel, K.R., 1998. A permutation test of association between configurations by means of the RV coefficient. Commun. Stat. — Simul. Comput 27, 843–856.

Hurlbert, S.H., 1971. The non-concept of species diversity: a critique and alternative parameters. Ecology 52, 577–586.

Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. J. Comput Graph. Stat. 5, 299–314.

Israels, A.Z., 1984. Redundancy analysis for qualitative variables. Psychometrika 49, 346–661.

Izsak, J., Papp, L., 1995. Application of the quadratic entropy indices for diversity studies of drosophilid assemblages. Environ. Ecol. Stat. 2, 213–224.

Izsak, J., Papp, L., 2000. A link between ecological diversity indices and measures of biodiversity. Ecol. Model. 130, 151–156 doi:10.1016/S0304-3800(00)00203-9.

Izsak, J., Szeidl, L., 2002. Quadratic diversity: its maximization can reduce the richness of species. Environ. Ecol. Stat. 9, 423–430.

Jaccard, P., 1901. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques régions voisines. Bull. Soc. Vaudoise Sci. Natur. 37, 241–272.

Jackson, D.A., 1995. Protest: a PROcustean randomization TEST of community environment concordance. Ecosciences 2, 297–303.

Johansson, J.K., 1981. An extension of Wollenberg's redundancy analysis. Psychometrika 46, 93–103.

Lamouroux, N., LeRoy Poff, N., Angermeier, P.L., 2002. Intercontinental convergence of stream fish community traits along geomorphic and hydraulic gradients. Ecology 83, 1792–1807.

Lande, R., 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos 76, 5–13.

Lauro, N., D'Ambra, L., 1984. Non-symmetrical correspondence analysis. In: Tomassone, R. (Ed.), Data Analysis and Informatics, III. Elsevier, North-Holland, Amsterdam, pp. 433–446.

Legendre, P., Anderson, M.J., 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol. Monogr. 69, 1–24.

Legendre, P., Legendre, L., 1998. Numerical ecology. Elsevier Science BV, Amsterdam.

Legendre, P., Galzin, R., Harmelin-Vivien, M.L., 1997. Relating behavior to habitat: solutions to the fourth-corner problem. Ecology 78, 547–562.

Losos, J.B., 1992. The evolution of convergent structure in Caribbean *Anolis* communities. Syst. Biol. 41, 403–420.

Manly, B.F., 1994. Multivariate Statistical Methods. A primer 2nd Edition.. Chapman & Hall, London.

McArdle, B.H., Anderson, M.J., 2001. Fitting multivariate models to community data: comment on distance-based redundancy analysis. Ecology 82, 290–297.

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY, USA.

Pélissier, R., Couteron, P., Dray, S., Sabatier, D., 2003. Consistency between ordination techniques and diversity measurements: two strategies for species occurence data. Ecology 84, 242–251.

Rao, C.R., 1952. Advanced Statistical Methods in Biometric Research. Wiley, New York.

Rao, C.R., 1964. The use and interpretation of principal component analysis in applied research. Sankhya A 26, 329–359.

Rao, C.R., 1982. Diversity and dissimilarity coefficients: a unified approach. Theor. Popul. Biol. 21, 24–43.

Rao, C.R., 1984. Convexity properties of entropy functions and analysis of diversity. Inequalities Statist. Probab. 5, 68–77.

Rao, C.R., 1986. Rao's axiomatization of diversity measures. In: Kotz, S., Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences. Wiley, New York, pp. 614–617.

Rao, C.R., Nayak, T.K., 1985. Cross entropy, dissimilarity measures, and characterizations of quadratic entropy. IEEE Trans. Inf. Theory IT-31, 589–593.

Schluter, D., Ricklefs, R.E., 1993. Convergence and regional component of species diversity. In: Ricklefs, R.E., Schluter, D. (Eds.), Species Diversity in Ecological Communities: Historical and Geographical Perspectives. The University of Chicago Press, Chicago, pp. 230–242.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. 27, 379–423 623–656.

Shimatani, K., 2001. On the measurement of species diversity incorporating species differences. Oikos 93, 135–147.

Simpson, E.H., 1949. Measurement of diversity. Nature 163, 688.

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

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

van den Wollenberg, A.L., 1977. Redundancy analysis, an alternative for canonical analysis. Psychometrika 42, 207–219.

Warwick, R.M., Clarke, K.R., 1995. New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. Mar. Ecol. Prog. Ser. 129, 301–305.

Watve, M.G., Gangal, R.M., 1996. Problems in measuring bacterial diversity and a possible solution. Appl. Environ. Microbiol. 62, 4299–4301.

Weir, B.S., 1996. Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Associates, Inc Publishers., Sunderland, MA.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Whittaker, R.H., 1972. Evolution and measurement of species diversity. Taxon 21, 213–251.

Wright, S., 1951. The genetical structure of populations. Ann. Eugen. 15, 323–354.

Wright, S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19, 395–420.