

disk using the option `ei.MD.bayes(..., ret.beta = "s")`, or discard the unit-level draws entirely using `ei.MD.bayes(..., ret.beta = "d")`. To reconstruct the chains, users can select the row marginals, column marginals, and units of interest, without reconstructing the entire matrix of unit-level draws:

```
> read.betas(rows = c("black", "white"),
+ columns = "dem", units = 1:150,
+ dir = getwd())
```

If users are interested in some function of the unit-level parameters, the implementation of the MD model allows them to define a function in R that will be called from within the C sampling algorithm, in which case the unit-level parameters need not be saved for post-processing.

Acknowledgments

eiPack was developed with the support of the Institute for Quantitative Social Science at Harvard University. Thanks to John Fox, Gary King, Kevin Quinn, D. James Greiner, and an anonymous referee for suggestions and Matt Cox and Bob Kinney for technical advice. For further information, see <http://www.olivialau.org/software>.

Bibliography

- W. T. Cho and A. H. Yoon. Strange bedfellows: Politics, courts and statistics: Statistical expert testimony in voting rights cases. *Cornell Journal of Law and Public Policy*, 10:237–264, 2001.
- O. D. Duncan and B. Davis. An alternative to ecological correlation. *American Sociological Review*, 18: 665–666, 1953.

- L. Goodman. Ecological regressions and the behavior of individuals. *American Sociological Review*, 18: 663–664, 1953.
- B. Grofman. A primer on racial bloc voting analysis. In N. Persily, editor, *The Real Y2K Problem: Census 2000 Data and Redistricting Technology*. Brennan Center for Justice, New York, 2000.
- K. Imai and Y. Lu. *eco: R Package for Fitting Bayesian Models of Ecological cov c Inference in 2x2 Tables*, 2005. URL <http://imai.princeton.edu/research/eco.html>.
- A. D. Martin and K. M. Quinn. Applied Bayesian inference in R using MCMCpack. *R News*, 6:2–7, 2006.
- M. Plummer, N. Best, K. Cowles, and K. Vines. CODA: Convergence diagnostics and output analysis for MCMC. *R News*, 6:7–11, 2006.
- O. Rosen, W. Jiang, G. King, and M. A. Tanner. Bayesian and frequentist inference for ecological inference: The $R \times C$ case. *Statistica Neerlandica*, 55(2):134–156, 2001.
- J. Wakefield. Ecological inference for 2×2 tables (with discussion). *Journal of the Royal Statistical Society*, 167:385–445, 2004.

Olivia Lau, Ryan T. Moore, Michael Kellermann
 Institute for Quantitative Social Science
 Harvard University, Cambridge, MA
olivia.lau@post.harvard.edu
ryantmoore@post.harvard.edu
kellerm@fas.harvard.edu

The ade4 Package — II: Two-table and K-table Methods

by Stéphane Dray, Anne B. Dufour and Daniel Chessel

Introduction

The **ade4** package proposes a great variety of explanatory methods to analyse multivariate datasets. As suggested by the acronym **ade4** (Data Analysis functions to analyse Ecological and Environmental data in the framework of Euclidean Exploratory methods), the package is devoted to ecologists but it could be useful in many other fields (e.g., [Goecke, 2005](#)). Methods available in the package are particular cases of the duality diagram ([Escoufier, 1987](#);

[Holmes, 2006](#); [Dray and Dufour, 2007](#)) and the implementation of the functions follows the description of this unifying mathematical tool (class `dudi`). The main functions of the package for one-table analysis methods have been presented in [Chessel et al. \(2004\)](#). This new paper presents a short summary of two-table and K -table methods available in the package.

Ecological illustration

In order to illustrate the methods, we used the dataset `jv73` ([Verneaux, 1973](#)) which is available in the package. This dataset concerns 12 rivers. For

each river, a number of sites have been sampled. The number of sites per river is not constant. `jv73$poi` is a `data.frame` and contains presence/absence data for 19 fish species (columns) in 92 sites (rows). `jv73$fac.riv` is a factor indicating the river corresponding to each site. `jv73$morpho` contains the measurements of six environmental variables (altitude (*m*), distance between the site and the source (*km*), slope (*per thousand*), wetted cross section (*m²*), average flow (*m³/s*) and average speed (*m/s*)) for the same sites. Several ecological questions are related to these data:

1. Are the groups of fish species living together (i.e. species communities)?
2. Is there a relation between the composition of fish communities and the environmental variations?
3. Does the composition of fish communities vary (or not) among rivers?
4. Do the species-environment relationships vary (or not) among rivers?

Multivariate analyses help to answer these different questions: one-table methods for the first question, two-table methods for the second one and *K*-table methods for the last two.

Matching two tables

The main purpose of ecological data analysis is the matching of two data tables: a sites-by-environmental variables table and a sites-by-species table, to study the relationships between the composition of species communities and their environment. The `ade4` package contains the main variants of these methods (procrustean rotation, co-inertia analysis and principal component analyses with respect to instrumental variables).

The first approach is procrustean rotation (Gower, 1971), introduced in ecology by Digby and Kempton (1987, p. 116).

```
data(jv73)
pca1 <- dudi.pca(jv73$morpho, scannf = FALSE)
pca2 <- dudi.pca(jv73$poi, scale = FALSE,
  scannf = FALSE)
plot(procuste(pca1$tab, pca2$tab,
  nf = 2))
```

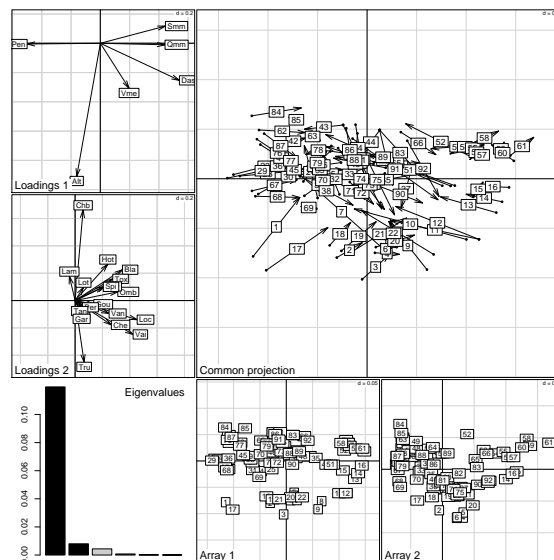


Figure 1: Plot of a Procrustes analysis: loadings for environmental variables and species, eigenvalues screeplot, scores of sites for the two data sets, and projection of the two sets of sites after rotation (arrows link environment site score to the species site score) (Dray et al., 2003a).

Two randomization procedures are available to test the association between two tables: PROTEST (Jackson, 1995) and RV (Heo and Gabriel, 1998).

```
plot(procuste.randtest(pca1$tab,
  pca2$tab), main = "PROTEST")
```

```
plot(RV.rtest(pca1$tab, pca2$tab),
  main = "RV")
```

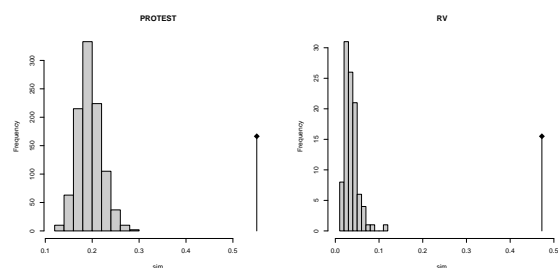


Figure 2: Plots of PROTEST and RV tests: histograms of simulated values and observed value (vertical line).

Co-inertia analysis (Dolédéc and Chessel, 1994; Dray et al., 2003b) is a general approach that can be applied to any pair of duality diagrams having the same row weights. This method is symmetric and seeks for a common structure between two datasets. It extends psychometricians inter-battery analysis (Tucker, 1958), canonical analysis on qualitative variables (Cazes, 1980), and ecological profiles analysis (Montaña and Greig-Smith, 1990; Mercier et al., 1992). Co-inertia analysis of the pair

of triplets (X_1, Q_1, D) and (X_2, Q_2, D) leads to the triplet $(X_2^t D X_1, Q_1, Q_2)$. Note that the two triplets must have the same row weights. For a comprehensive definition of the statistical triplet of matrices X, Q, D , the reader could consult [Chessel et al. \(2004\)](#).

```
coal <- dudi.coa(jv73$poi, scannf = FALSE)
pca3 <- dudi.pca(jv73$morpho,
  row.w = coal$lw, scannf = F)
plot(coinertia(coal, pca3, scannf = FALSE))
```

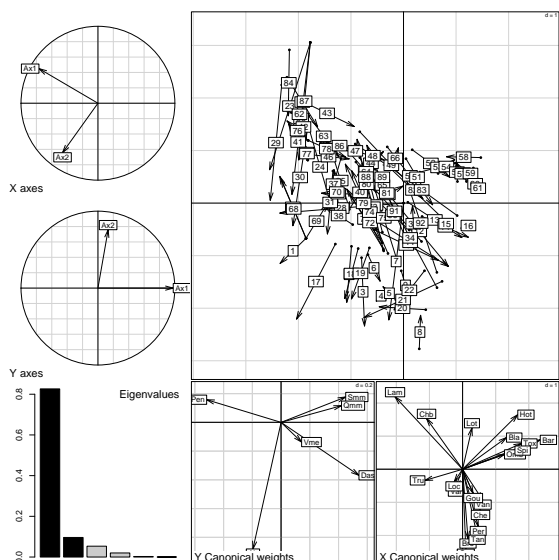


Figure 3: Plot of a co-inertia analysis: projection of the principal axes of the two tables (species and environment) on co-inertia axes, eigenvalues screeplot, canonical weights of species and environmental variables, and joint display of the sites.

For each coupling method, a generic plot function presents the various elements required to interpret the results. However, the quality of graphs could vary according to the data set. It is consequently impossible to manage relevant graphical outputs for all cases. That is why these generic plot use graphical functions of **ade4** which can be directly called by the user. A brief description of some of these functions is given in Table 1.

Another two-table matching strategy is principal component analyses with respect to instrumental variables (pcaiv, [Rao, 1964](#)). This approach consists in explaining a triplet (X_2, Q_2, D) by a table of independent variables X_1 and leads to triplet $(P_{X_1} X_2, Q_2, D)$ where $P_{X_1} = X_1(X_1^t D X_1)^{-1} X_1^t D$. This family of methods are constrained ordinations, among which redundancy analysis ([van den Wollenberg, 1977](#)) and canonical correspondence analysis ([Ter Braak, 1986](#)) are the most frequently used in ecology. Note that canonical correspondence analysis can also be performed using the `cca` wrapper function which takes two tables as arguments. The example given below is then exactly equivalent to

`plot(cca(jv73$poi, jv73$morpho, scannf=FALSE))`. While the `cca` function of **ade4** is a particular case of `pcaiv`, the `cca` function of the package **vegan** is a more traditional implementation of the method which could be preferred by ecologists.

Function	Objective
<code>s.arrow</code>	cloud of points with vectors
<code>s.chull</code>	cloud of points with groups by convex hulls
<code>s.class</code>	cloud of points with groups by stars or ellipses
<code>s.corcircle</code>	correlation circle
<code>s.distri</code>	cloud of points with frequency distribution by stars and ellipses
<code>s.hist</code>	cloud of points with two marginal histograms
<code>s.image</code>	grid of gray-scale rectangles with contour lines
<code>s.kde2d</code>	cloud of points with kernel density estimation
<code>s.label</code>	cloud of points with labels
<code>s.logo</code>	cloud of points with pictures
<code>s.match</code>	matching two clouds of points with vectors
<code>s.traject</code>	cloud of points with trajectories
<code>s.value</code>	cloud of points with numerical variable

Table 1: Objectives of some graphical functions.

```
plot(pcaiv(coal, jv73$morpho, scannf = FALSE))
```

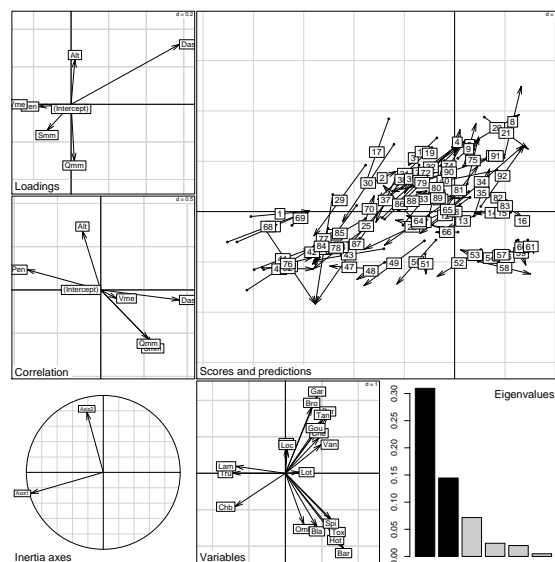


Figure 4: Plot of a CCA seen as a particular case of PCAIV: environmental variables loadings and correlations with CCA axes, projection of principal axes on CCA axes, species scores, eigenvalues screeplot, and joint display of the rows of the two tables (position of the sites by averaging (points) and by regression (arrow tips)).

Orthogonal analysis (`pcaivortho`) removes the effect of independent variables and corresponds to the triplet $(\mathbf{P}_{\perp X_1}, \mathbf{X}_2, \mathbf{Q}_2, \mathbf{D})$ where $\mathbf{P}_{\perp X_1} = \mathbf{I} - \mathbf{P}_{X_1}$. Between-class (between) and within-class (within) analyses (see [Chessel et al., 2004](#), for details) are particular cases of PCAIV and orthogonal PCAIV when there is only one categorical variable (i.e. factor) in X_1 . Within-class analyses take into account a partition of individuals into groups and focus on structures which are common to all groups. It can be seen as a first step to *K*-table methods.

```
wit1 <- within(coa1, fac = jv73$fac.riv,
              scannf = FALSE)
plot(wit1)
```

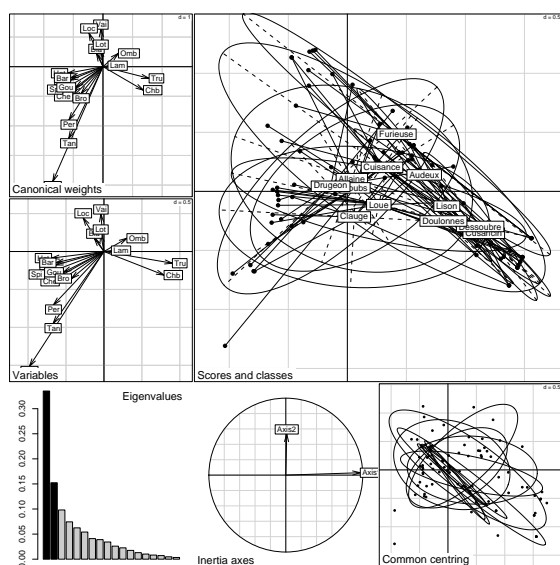


Figure 5: Plot of a within-class analysis: species loadings, species scores, eigenvalues screeplot, projection of principal axes on within-class axes, sites scores (common centring), projections of sites and groups (i.e. rivers in this example) on within-class axes.

The *K*-table class

Class `ktab` corresponds to collections of more than two duality diagrams, for which the internal structures are to be compared. Three formats of these collections can be considered:

- $(\mathbf{X}_1, \mathbf{Q}_1, \mathbf{D}), (\mathbf{X}_2, \mathbf{Q}_2, \mathbf{D}), \dots, (\mathbf{X}_K, \mathbf{Q}_K, \mathbf{D})$
- $(\mathbf{X}_1, \mathbf{Q}, \mathbf{D}_1), (\mathbf{X}_2, \mathbf{Q}, \mathbf{D}_2), \dots, (\mathbf{X}_K, \mathbf{Q}, \mathbf{D}_K)$ stored in the form of $(\mathbf{X}_1^t, \mathbf{D}_1, \mathbf{Q}), (\mathbf{X}_2^t, \mathbf{D}_2, \mathbf{Q}), \dots, (\mathbf{X}_K^t, \mathbf{D}_K, \mathbf{Q})$
- $(\mathbf{X}_1, \mathbf{Q}, \mathbf{D}), (\mathbf{X}_2, \mathbf{Q}, \mathbf{D}), \dots, (\mathbf{X}_K, \mathbf{Q}, \mathbf{D})$ which can also be stored in the form of $(\mathbf{X}_1^t, \mathbf{D}, \mathbf{Q}), (\mathbf{X}_2^t, \mathbf{D}, \mathbf{Q}), \dots, (\mathbf{X}_K^t, \mathbf{D}, \mathbf{Q})$

Each statistical triplet corresponds to a separate analysis (e.g., principal component analysis, correspondence analysis ...). The common dimension of the *K* statistical triplets are the rows of tables which can represent individuals (samples, statistical units) or variables. Utilities for building and manipulating `ktab` objects are available. *K*-table can be constructed from a list of tables (`ktab.list.df`), a list of `dudi` objects (`ktab.list.dudi`), a within-class analysis (`ktab.within`) or by splitting a table (`ktab.data.frame`). Generic functions to transpose (`t. ktab`), combine (`c. ktab`) or extract elements (`[. ktab`) are also available. The `sepan` function can be used to compute automatically the *K* separate analyses.

```
kt1 <- ktab.within(wit1)
sep1 <- sepan(kt1)
kplot.sepan.coa(sep1, permute.row.col = TRUE)
```

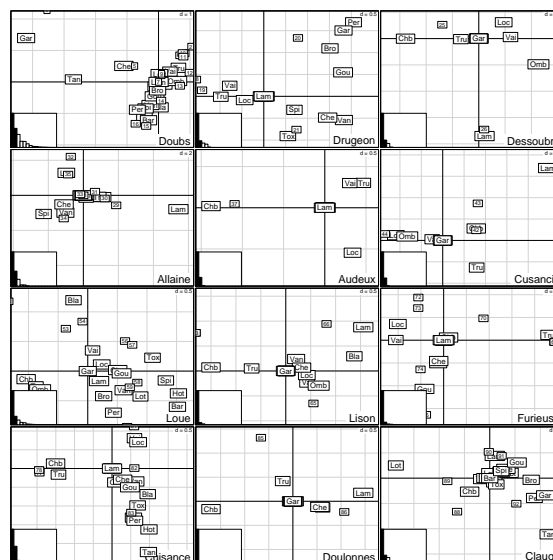


Figure 6: *K*plot of 12 separate correspondence analyses (same species, different sites).

When the `ktab` object is built, various statistical methods can be used to analyse it. The `foucart` function can be used to analyse *K* tables of positive number having the same rows and the same columns and that can be analysed by a CA ([Foucart, 1984](#); [Pavoine et al., 2007](#)). Partial triadic analysis ([Tucker, 1966](#)) is a first step toward three modes principal component analysis ([Kroonenberg, 1989](#)) and can be computed with the `pta` function. It must be used on *K* triplets having the same row and column weights. The `pta` function can be used to perform the STATICO method ([Simier et al., 1999](#); [Thioulouse et al., 2004](#)). This makes it possible to analyse a pair of `ktab` objects which have been combined by the `ktab.match2ktabs` function.

Multiple factor analysis (`mfa`, [Escofier and Pagès, 1994](#)), multiple co-inertia analysis (`mcoa`, [Chessel and](#)

Hanafi, 1996) and the STATIS method (statis, Lavit et al., 1994) can be used to compare K triplets having the same row weights. The STATIS method can also be used to compare K triplets having the same column weights, which is a first step toward Common PCA (Flury, 1988).

```
sta1 <- statis(kt1, scannf = F)
plot(sta1)
```

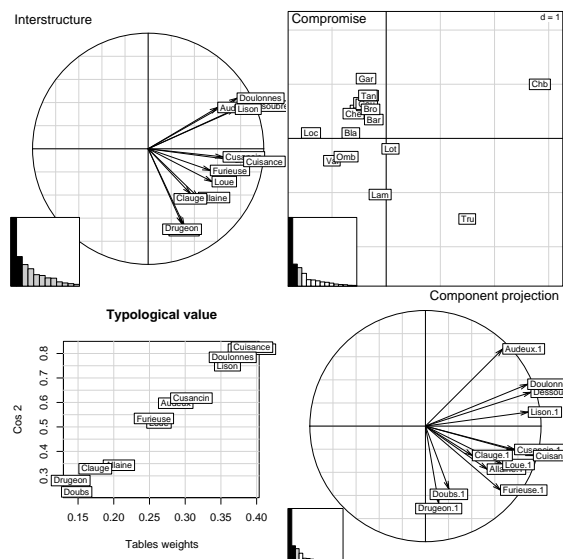


Figure 7: Plot of STATIS analysis: interstructure, typological value of each table, compromise and projection of principal axes of separate analyses onto STATIS axes.

The `kplot` generic function is associated to the `foucart`, `mcoa`, `mca`, `pta`, `sepan`, `sepan.coa` and `statis` methods, giving adapted collections of graphics.

```
kplot(sta1, traj = TRUE, arrow = FALSE,
      unique = TRUE, clab = 0)
```

Conclusion

The `ade4` package provides many methods to analyse multivariate ecological data sets. This diversity of tools is a methodological answer to the great variety of questions and data structures associated to biological questions. Specific methods dedicated to the analysis of biodiversity, spatial, genetic or phylogenetic data are also available in the package. The `adehabitat` brother-package contains tools to analyse habitat selection by animals while the `ade4TkGUI` package provides a graphical interface to `ade4`. More resources can be found on the `ade4` website (<http://pbil.univ-lyon1.fr/ADE-4/>).

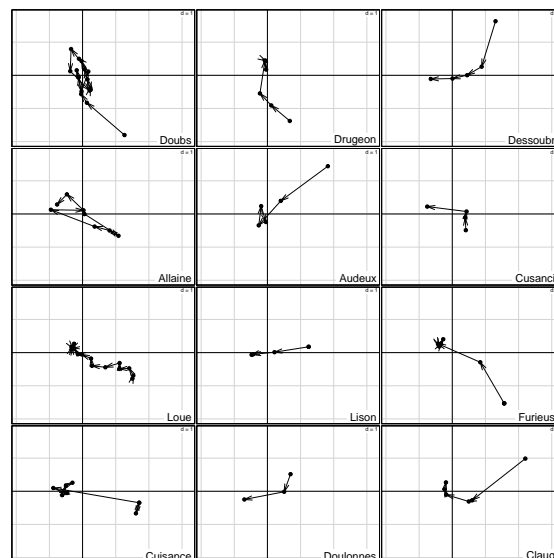


Figure 8: `kplot` of the projection of the sites of each table on the principal axes of the compromise of STATIS analysis.

Bibliography

- P. Cazes. L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. I. Définitions et applications à l'analyse canonique des variables qualitatives. *Les Cahiers de l'Analyse des Données*, 5:145–161, 1980.
- D. Chessel and M. Hanafi. Analyse de la co-inertie de K nuages de points. *Revue de Statistique Appliquée*, 44(2):35–60, 1996.
- D. Chessel, A.-B. Dufour, and J. Thioulouse. The `ade4` package-I- One-table methods. *R News*, 4:5–10, 2004.
- P. G. N. Digby and R. A. . Kempton. *Multivariate Analysis of Ecological Communities*. Chapman and Hall, Population and Community Biology Series, London, 1987.
- S. Dolédec and D. Chessel. Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology*, 31: 277–294, 1994.
- S. Dray, D. Chessel, and J. Thioulouse. Procrustean co-inertia analysis for the linking of multivariate datasets. *Ecoscience*, 10:110–119, 2003a.
- S. Dray, D. Chessel, and J. Thioulouse. Co-inertia analysis and the linking of ecological tables. *Ecology*, 84(11):3078–3089, 2003b.
- S. Dray and A. Dufour. The `ade4` package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4):1–20, 2007.

- B. Escoufier and J. Pagès. Multiple factor analysis (AF-MULT package). *Computational Statistics and Data Analysis*, 18:121–140, 1994.
- Y. Escoufier. The duality diagram : a means of better practical applications. In P. Legendre and L. Legendre, editors, *Development in numerical ecology*, pages 139–156. NATO advanced Institute , Serie G .Springer Verlag, Berlin, 1987.
- B. Flury. *Common Principal Components and Related Multivariate. models*. Wiley and Sons, New-York, 1988.
- T. Foucart. *Analyse factorielle de tableaux multiples*. Masson, Paris, 1984.
- R. Goecke. 3D lip tracking and co-inertia analysis for improved robustness of audio-video automatic speech recognition. In *Proceedings of the Auditory-Visual Speech Processing Workshop AVSP 2005*, pages 109–114, 2005.
- J. Gower. Statistical methods of comparing different multivariate analyses of the same data. In F. Hodson, D. Kendall, and P. Tautu, editors, *Mathematics in the archaeological and historical sciences*, pages 138–149. University Press, Edinburgh, 1971.
- M. Heo and K. Gabriel. A permutation test of association between configurations by means of the RV coefficient. *Communications in Statistics - Simulation and Computation*, 27:843–856, 1998.
- S. Holmes. Multivariate analysis: The French way. In N. D. and S. T., editors, *Festschrift for David Freedman*. IMS, Beachwood, OH, 2006.
- D. Jackson. PROTEST: a PROcustean randomization TEST of community environment concordance. *Ecosciences*, 2:297–303, 1995.
- P. Kroonenberg. The analysis of multiple tables in factorial ecology. iii three-mode principal component analysis:"analyse triadique complète". *Acta OEcologica, OEcologia Generalis*, 10:245–256, 1989.
- C. Lavit, Y. Escoufier, R. Sabatier, and P. Traissac. The ACT (STATIS method). *Computational Statistics and Data Analysis*, 18:97–119, 1994.
- P. Mercier, D. Chessel, and S. Dolédec. Complete correspondence analysis of an ecological profile data table: a central ordination method. *Acta OEcologica*, 13:25–44, 1992.
- C. Montaña and P. Greig-Smith. Correspondence analysis of species by environmental variable matrices. *Journal of Vegetation Science*, 1:453–460, 1990.
- S. Pavoine, J. Blondel, M. Baguette, and D. Chessel. A new technique for ordering asymmetrical three-dimensional data sets in ecology. *Ecology*, 88:512–523, 2007.
- C. Rao. The use and interpretation of principal component analysis in applied research. *Sankhya A*, 26: 329–359, 1964.
- M. Simier, L. Blanc, F. Pellegrin, and D. Nandris. Approche simultanée de K couples de tableaux: application à l'étude des relations pathologie végétale-environnement. *Revue de Statistique Appliquée*, 47:31–46, 1999.
- C. Ter Braak. Canonical correspondence analysis : a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67:1167–1179, 1986.
- J. Thioulouse, M. Simier, and D. Chessel. Simultaneous analysis of a sequence of pairs of ecological tables with the STATICO method. *Ecology*, 85:272–283, 2004.
- L. . Tucker. An inter-battery method of factor analysis. *Psychometrika*, 23:111–136, 1958.
- L. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.
- A. van den Wollenberg. Redundancy analysis, an alternative for canonical analysis. *Psychometrika*, 42 (2):207–219, 1977.
- J. Verneaux. *Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie*. Thèse de doctorat, Université de Besançon, Besançon, 1973.

Stéphane Dray, Anne-Béatrice Dufour, Daniel Chessel
 Laboratoire de Biométrie et
 Biologie Evolutive (UMR 5558) ; CNRS
 Université de Lyon ; université Lyon 1
 43, Boulevard du 11 Novembre 1918
 69622 Villeurbanne Cedex, France
 dray@biomserv.univ-lyon1.fr
 dufour@biomserv.univ-lyon1.fr
 chessel@biomserv.univ-lyon1.fr