Océanis • vol. 24 nº 4 • 1998 • p. 393-416

ADE-4 Software: a tool for multivariate analysis and graphical display*

Monique SIMIER[(1)], Jean THIOULOUSE[(2)] & Jean-Michel OLIVIER[(3)]

(1) Laboratoire halieutique et écosystèmes aquatiques
IRD
BP 5045
34032 Montpellier cedex 1
France
monique.simier@mpl.ird.fr
(2) Laboratoire de biométrie, génétique et biologie des populations
UMR CNRS 5558
Université Lyon 1
69622 Villeurbanne cedex
France
(3) Laboratoire d'écologie des eaux douces et des grands fleuves
URA CNRS 1974
Université Lyon 1
69622 Villeurbanne cedex
France

Keywords: Statistical software, multivariate analysis, principal component analysis, correspondence analysis, instrumental variables, canonical correspondence analysis, partial least squares regression, coinertia analysis, multi-table analysis, graphics, multivariate graphics, Macintosh, HyperCard, Windows 95, Windows NT

Abstract
We present ADE-4, a multivariate analysis and graphical display software. Multivariate analysis methods available in ADE-4 include usual

*This paper is from: Thioulouse J., Chessel D., Dolédec S. & Olivier J.M., 1997. ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, nº 7, p. 75–83, which should be referenced by any publication using this software.

Océanis ISSN 0182-0745

one-table methods like principal component analysis and correspondence analysis, spatial data analysis methods (using a total variance decomposition into local and global components, analogous to Moran and Geary indices), partitioning and classification methods, discriminant analysis and within/between groups analyses, many linear regression methods including lowess and polynomial regression, multiple and PLS (partial least squares) regression and orthogonal regression (principal component regression), projection methods like principal component analysis on instrumental variables, canonical correspondence analysis and many other variants, coinertia analysis and the RLQ method, and several three-way table (*k*-table) analysis methods. Graphical display techniques include an automatic collection of elementary graphics corresponding to groups of rows or to columns in the data table, thus providing a very efficient way for automatic *k*-table graphics and geographical mapping options. A dynamic graphic module allows interactive operations like searching, zooming, selection of points, and display of data values on factor maps. The user interface is simple and homogeneous among all the programs; this help makes the use of ADE-4 very easy for non specialists in statistics, data analysis or computer science.

Logiciel ADE-4, outil pour l'analyse des données multivariées et leur représentation graphique

Mots clés : logiciel statistique, analyse multivariée, analyse en composantes principales, analyse des correspondances, variables instrumentales, analyse canonique des correspondances, régression aux moindres carrés partiels, analyse de co-inertie, analyse multi-tableaux, graphiques, graphiques multivariés, Macintosh, HyperCard, Windows 95, Windows NT

Résumé

Nous présentons ADE-4, un logiciel pour l'analyse et la représentation graphique des données multivariées. Les méthodes d'analyse multivariées disponibles dans ADE-4 sont les méthodes classiques à un seul tableau comme l'analyse en composantes principales et l'analyse des correspondances, des méthodes d'analyse des données spatialisées (décomposant la variance totale en composantes locales et globales, comme les indices de Moran et de Geary), des méthodes de partitionnement et de classification, l'analyse discriminante et les analyses inter- ou intra-classes, plusieurs méthodes de régression linéaire comme les régressions « lowess » et polynomiales, les régressions multiples et PLS (moindres carrés partiels) et les régressions orthogonales (régression en composantes principales), des méthodes de projection comme l'analyse en composantes principales sur variables instrumentales, l'analyse canonique des correspondances et autres variantes, les analyses de co-inertie et la méthode RLQ, et enfin plusieurs méthodes d'analyse multi-tableaux. Les techniques de représentation graphique permettent de réaliser automatiquement des collections de graphiques correspondant à des groupes de lignes ou à des colonnes du tableau de données, offrant ainsi un outil très efficace pour les graphiques multi-tableaux ou les cartes géographiques. Un module graphique dynamique autorise des opérations interactives comme la recherche, l'agrandissement de parties du graphe, la sélection de points, et la représentation de valeurs sur des cartes factorielles. L'interface utilisateur est simple et homogène pour tous les modules ; cela contribue à faire d'ADE-4 un outil accessible aux non-spécialistes en statistiques, analyse de données ou informatique.

1. Introduction

ADE-4 (Thioulouse *et al.*, 1997) is a multivariate analysis and graphical display software for Apple Macintosh and Windows 95 or NT microcomputers. It is made of several stand-alone applications, called modules, that feature a wide range of multivariate analysis methods, from simple one-table analysis to three-way table analysis and two-table coupling methods. It also provides many possibilities for helpful graphical display in the process of analyzing multivariate data sets. It has been developed in the context of environmental data analysis, but can be used in other scientific disciplines where data analysis is frequently used. ADE-4 is developed by J. Thioulouse (computer science), D. Chessel (statistics), S. Dolédec (ecology) and J.-M. Olivier (contact & distribution). The whole software, including documentation and sample data sets are is freely available on the Internet network. Here, we wish to present the main characteristics of ADE-4, from three point of view: (1) data analysis methods, (2) graphical display capabilities, and (3) user interface.

2. Data analysis methods

2.1. Mathematical basis of the computation

The data analysis methods available in ADE-4 are based on the duality diagram (Cailliez & Pagès, 1976), introduced in ecology by Escoufier (1987) in which a raw data table is considered as a statistical triplet composed of a transformed table of the same dimensions and two weight matrices, one for the lines, the other for the columns. Depending on the transformation and the weights, all linear data analysis methods can be defined.

In many modules, Monte-Carlo tests (Good, 1994, chapter 13) are available to study the significance of observed structures.

2.2. One table methods

Three basic multivariate analysis methods can be applied to one-table data sets (figure 1A) (DOLÉDEC & CHESSEL, 1991). The corresponding modules are the PCA (principal components analysis) module for quantitative variables, the COA (correspondence analysis) module for contingency tables (GREENACRE, 1984), and the MCA (multiple correspondence analysis) module for qualitative (discrete) variables (NISHISATO, 1980; TENENHAUS & YOUNG, 1985).

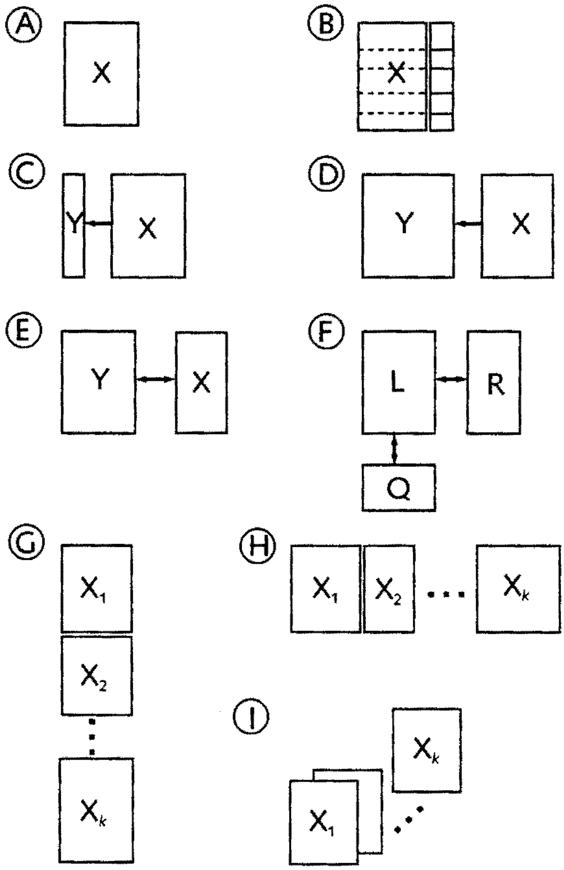

Figure 1: Main data set configurations treated by ADE-4. A: One table X. B: One table X with groups of rows. C: One dependant variable (Y) explained by several independant variables (X). D: Data table Y explained by data table X (dissymmetrical relation). E: Common structure to X and Y (symmetrical relation). F: Common structure to R and Q through a link table (L). G: Collection of k tables with same columns. H: Collection of k tables with same rows. I: Collection of k tables with same rows and columns.

The PCA module offers several options, corresponding to different duality diagrams: correlation matrix PCA, covariance matrix PCA, non centered PCA (NOY-MEIR, 1973), decentered PCA, partially standardized PCA (BOUROCHE, 1975), within-groups standardized PCA (DOLÉDEC & CHESSEL, 1987). See OKAMOTO (1972) for a discussion of different types of PCA.

The COA module offers six options for correspondence analysis (CA): classical CA, reciprocal scaling (THIOULOUSE & CHESSEL, 1992), row weighted CA, internal CA (CAZES *et al.*, 1988), decentered CA (DOLÉDEC *et al.*, 1995).

The MCA module offers two options for the analysis of tables made of qualitative variables: Multiple Correspondence Analysis (TENENHAUS & YOUNG, 1985) and Fuzzy Correspondence Analysis (CHEVENET *et al.*, 1994; CASTELLA & SPEIGHT, 1996).

A fourth module, HTA (homogeneous table analysis) is intended for homogeneous tables, *i.e.*, tables in which all the values come from the same variable (for example, a toxicity table containing the toxicity of some chemical compounds toward several animal species, see DEVILLERS & CHESSEL, 1995).

Lastly, the DDUtil (duality diagram utilities) module provides several interpretation helps that can be used with any of the methods available in the first four modules, namely: biplot representation (GABRIEL, 1971, 1981), inertia analysis for rows and columns (particularly for COA, see GREENACRE, 1984), supplementary rows and/or columns (LEBART *et al.*, 1984), and data reconstitution (LEBART *et al.*, 1984).

2.3. Partitioning and classification methods

The Clusters module provides a partitioning algorithm using moving centers near to the *k*-means clustering method, and five algorithms for computing hierarchies from a distance matrix including average linkage, simple linkage, complete linkage, Ward's method and second order moment divisive algorithm (ROUX, 1985, 1991).

2.4. One table with spatial structures

Environmental data very often include spatial information (*e.g.*, the spatial location of sampling sites), and this information is difficult to introduce in classical multivariate analysis methods. The NGStat module provides a way to achieve this by using a neighbouring relationship between sites. See LEBART (1969) for a presentation of this approach and THIOULOUSE *et al.* (1995) for a general framework based on variance decomposition formulas. Also available in Distances module are the Mantel test (MANTEL, 1967), the principal coordinate analysis (MANLY, 1994), and the minimum spanning tree (KEVIN & WHITNEY, 1972).

2.5. One table with groups of rows

When *a priori* groups of individuals exist in the data table (figure 1B), the Discrimin module can be used to perform a discriminant analysis (DA, also called canonical variate analysis), and between-groups or within-groups analyses (DOLÉDEC & CHESSEL, 1989). These three methods can be performed after a PCA, a COA, or an MCA, leading to a great variety of analyses. For example, in the case of DA, we can obtain after a PCA, the classical DA (MAHALANOBIS, 1936; TOMASSONE *et al.*, 1988), after a COA, the correspondence DA, and after an MCA the DA on qualitative variables (SAPORTA, 1975; PERSAT *et al.*, 1985). Monte-Carlo tests are available to test the significance of the between-groups structure.

2.6. Linear regression

Four modules provide several linear regression methods (figure 1C). These modules are UniVarReg (for univariate regression), OrthoVar (for orthogonal regression), LinearReg (for linear regression) and PLSgen2 (for second generation Partial Least Squares regression). Here also, Monte-Carlo tests are available to test the results of these methods.

The UniVarReg module deals with two regression models: polynomial regression and Lowess method (locally weighted regression and smoothing scatterplots; CLEVELAND, 1979; CLEVELAND & DEVLIN, 1988).

The OrthoReg module performs multiple linear regression in the particular case of orthogonal explanatory variables. This is useful for example in PCR (principal component regression; NÆS, 1984), or in the case of the projection on the subspace spaned by a series of eigenvectors (THIOULOUSE *et al.*, 1995).

The LinearReg module performs the usual multiple linear regression (MLR), and the first generation partial least squares (PLS) regression (LINDGREN, 1994). See also GELADI & KOWALSKI (1986); HÖSKULDSSON (1988) for more details on PLS regression.

The PLSgen2 module performs second generation PLS regression (TENENHAUS, 1997).

2.7. Two-tables coupling methods

The Projectors module is dedicated to two-tables coupling methods (figure 1D) based on projection onto vector subspaces (TAKEUCHI *et al.*, 1982). The first six options allow to build orthonormal bases on which the projections can be made. The following five options provide several two-tables coupling methods, and mainly PCAIV (PCA on Instrumental Variables) methods. The "PCA on Instrumental Variables" option can be used with any statistical triplet from the

PCA, COA and MCA modules, which corresponds for example to methods like CAIV (correspondence analysis on instrumental variables, LEBRETON *et al.*, 1988a,b, 1991) or CCA (canonical correspondence analysis, TER BRAAK, 1987a,b). This last method can also be performed using CCA module, which is especially dedicated to it.

2.8. Coinertia analysis method

There are two modules for coinertia analysis. The Coinertia module, which performs the usual coinertia analysis (figure 1E) (CHESSEL & MERCIER, 1993; DOLÉDEC & CHESSEL, 1994; THIOULOUSE & LOBRY, 1995; CADET *et al.*, 1994), and the RLQ module, which performs a three-table generalization of coinertia analysis (figure 1F) (DOLÉDEC *et al.*, in press).

2.9. *k*-table analysis methods

Collections of tables (three-ways tables, or *k*-tables, figure 1G, H, I) can be analyzed with the STATIS or the KTA-MFA modules. STATIS module features three distinct methods: STATIS (ESCOUFIER, 1980; LAVIT, 1988; LAVIT *et al.*, 1994), the partial triadic analysis (THIOULOUSE & CHESSEL, 1987; BLANC *et al.*, 1998), and the analysis of a series of contingency tables (FOUCART, 1978). Two other *k*-tables methods are provided by the KTA-MFA module: Multiple Factorial Analysis (ESCOFIER & PAGÈS, 1994) and Multiple Co-inertia analysis (CHESSEL & HANAFI, 1996) which is a generalisation of the two-tables co-inertia analysis. The KTabUtil module provides a series of three-ways table manipulation utilities : *k*-table transposition, sorting, centering, standardization, etc. Using KTabUtil and Statis modules, one can also perform extensions of STATIS to COA Analysis (GAERTNER *et al.*, 1998) and to Co-inertia Analysis for the analysis of *k* pairs of tables (SIMIER *et al.*, 1999).

3. Graphical representations

ADE-4 graphical modules fall broadly in five categories: one dimensional graphics, curves, scatters, geographical maps and dendrograms. Most modules have the possibility automatically to draw collections of graphics, corresponding to the columns of the data file (one graphic for each variable), to groups of rows (one graphic for each group), or to both (one graphic for each group and for each variable). This feature is particularly useful in multivariate analysis, where one always deals with many variables and/or groups of samples. Moreover, several modules have two versions, according to the way they treat the collections: elementary graphics can be either simply put side by side, or superimposed. Superimposition is done in the modules with a name ending with the "Class" suffix.

3.1. One dimensional graphics

The Graph1D module is intended for one dimensional data representation, such as the values of one factor score. It has four options: Histograms, Labels, Stars and Bars. The Histograms option computes the distribution of the values into classes and draws the corresponding histogram, with optionally the adjusted Gauss curve superimposed over it (figure 2A). The Labels option simply draws regularly spaced labels (that can be chosen by the user) vertically or horizontally along an axis; these labels are connected by lines to the corresponding coordinates on the axis. The Bars option draws bars along an axis, according to corresponding coordinates (figure 2B). The Stars option is specially designed for circular data such as monthly measurements (figure 2C). For every option, the columns and groups of rows corresponding to each elementary graphic of a collection can be chosen by the user.

The Graph1DClass module is also intended for representing one dimensional data, but, as the "Class" suffix indicates, for groups of rows and with the corresponding graphics superimposed instead of placed side by side. The only available option is GaussCurves, which draws a collection of Gauss curves along an axis (figure 2D), each curve corresponding to a group a rows.

3.2. Curves

The Curves module draws curves, *i.e.* a series of values (ordinates) are plotted along an axis (abscissa). It features five options: Lines, Bars, Steps, Boxes and Eigenvalues. Here also, columns and groups of rows can correspond to the elementary graphics of a collection. Values can be presented as dots joined by a line with the Lines option (figure 3A), or as separate bars with the Bars option (figure 3B), or as steps with the Steps option. The Eigenvalues option is dedicated to drawing of the eigenvalues diagram after a factorial analysis (figure 3C). Boxes are classical "box and whiskers" display, showing the median, quartiles, minimum and maximum.

The CurveClass module acts in the same way as the Curves module, except that the curves defined by the qualitative variable are superimposed in the same elementary graphic instead of being displayed in several graphics. Two options are available: Boxes (figure 3D) or Lines (figure 3E)

The CurveModels module allows fitting models to observed series of values. This module automatically fits a model for each elementary graphic in a collection. Polynomial models (figure 3F) and Lowess models (figure 3G)

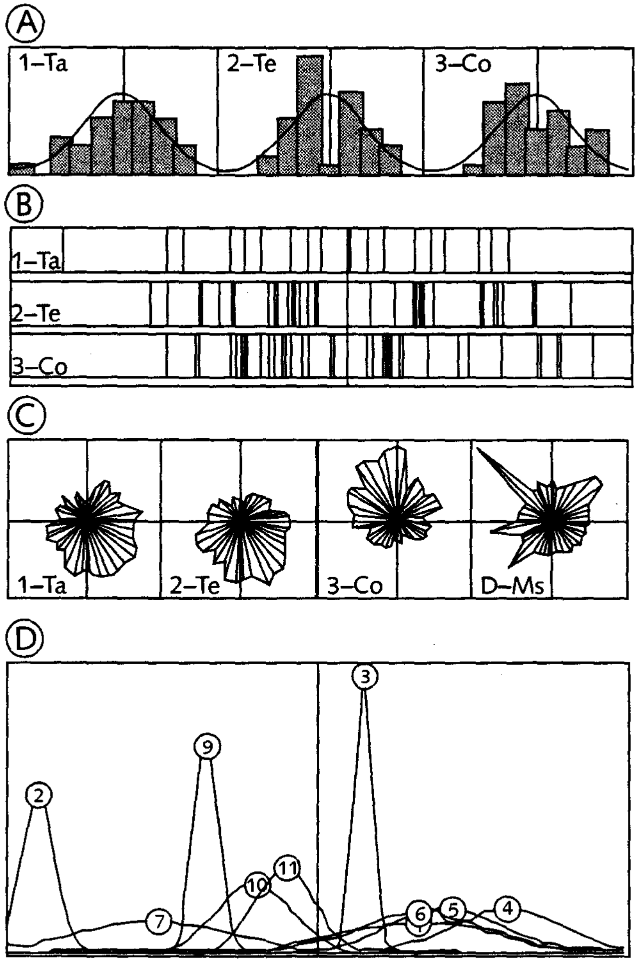

Figure 2: One dimensional graphics drawn by ADE-4. A: Histograms and adjusted Gauss curves (optional) are automatically drawn for each variable with Graph1D, option Histograms. B: Bars position values along an axis, with Graph1D, option Bars. C: Star charts are used for plotting of circular data. D: Gauss curves, corresponding to groups of rows, are superimposed on the same graph with Graph1DClass, option GaussCurves.

can be fitted and superimposed to observed values with the Polynomials and Lowess options.

3.3. Scatters

The most classical graphic in multivariate analysis is the factor map. The Scatters module is designed to draw such graphics, with several options. For

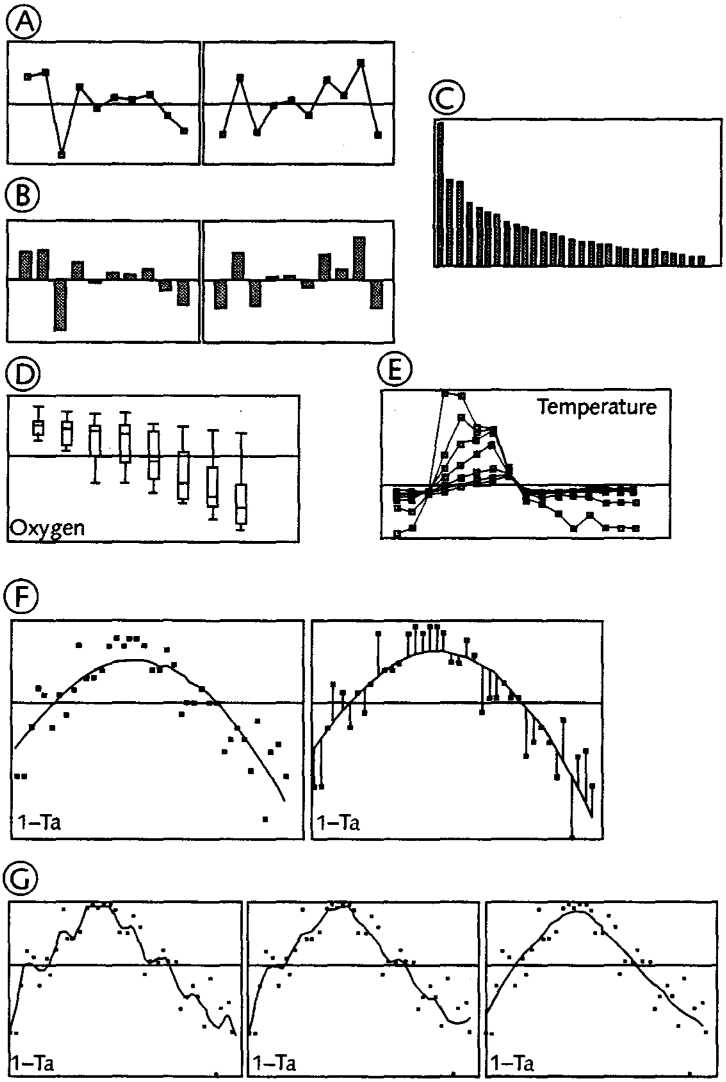

Figure 3: Graphs obtained with the Curves, CurveClass and CurveModels modules: for each of them, a series of values (ordinates) are drawn along an axis (abscissa). A: The dots are joined with Curves, option Lines. B: Values are drawn as separate bars with Curves, option Bars. C: An eigenvalue diagram is obtained by Curves, option EigenValues. D: Box-Plots corresponding to different groups of rows are drawn side by side by CurveClass, option Boxes. E: Lines corresponding to different groups of rows are superimposed on the same axis by CurveClass, option Lines. F: A polynomial model (line) is fitted to an observed series of values (dots) with CurveModels, option Polynomials. Observed values can be optionnally related to the predicted ones (right). G: A Lowess model is fitted to the same series of values as in F. The highest the number of neighbours is, the smoothest the line (from left to right, respectively 4, 6 and 12 neighbours are used for the computations).

all options, the user can interactively select the columns and the groups of rows that will be used to draw the elementary graphics that make a collection. The simplest option is Labels. For each point, it only draws a characters string (label) on the factor map (figure 4A).

The Trajectories option underlines the fact that the elements are ordered (for example in the case of time series) by linking the points with a line.

The Stars option computes the gravity center of each group of points and draws lines connecting each element to its gravity center.

The Ellipses option computes the means, variances and covariance of each group of points on both axes, and draws an ellipse with these parameters: the center of the ellipse is centered on the means, its width and height are given by the variances, and the covariance sets the slope of the main axis of the ellipse.

The Convex hulls option simply draws the convex hull of each set of points. Ellipses and convex hulls are labeled by the number of the group.

The Values option is slightly more complex. For each point on the factor map, it draws a circle or a square which size is proportional to a series of values (circle are for positive values, and squares for negative ones). This technique is particularly useful to represent data values on the factor map (figure 4B).

Lastly, the "Match two scatters" option can be used when two sets of scores are available for the same points (this is frequently the case in co-inertia analysis and other two-table coupling methods). An arrow is drawn connecting the point in the first set with that same point in the second set.

The ScatterClass module incorporates the Labels, Trajectories, Stars (figure 4C), Ellipses (figure 4D) and Convex hulls options. It superimposes the elementary graphics corresponding to a collection.

The last module for scatter diagrams is ADEScatters (THIOULOUSE, 1996). It is a dynamic graphic module, in which the user can perform several actions that help interpreting the factor map: searching, zooming, selection of sets of points, interactive display of data values on the factor map. For the moment, this module is available for Macintosh only.

3.4. Cartography modules

Five cartography modules are available in ADE-4. They can be used to map either the initial (raw or transformed) data, or the factor scores resulting from a multivariate analysis.

The Digit module allows user to digitize points coordinates on a background map.

The Levels module draws contour curves on the map (figure 5A). It can be used with sampling points having any distribution on the map: an interpolated regular grid is computed before drawing. Contour curves are computed by

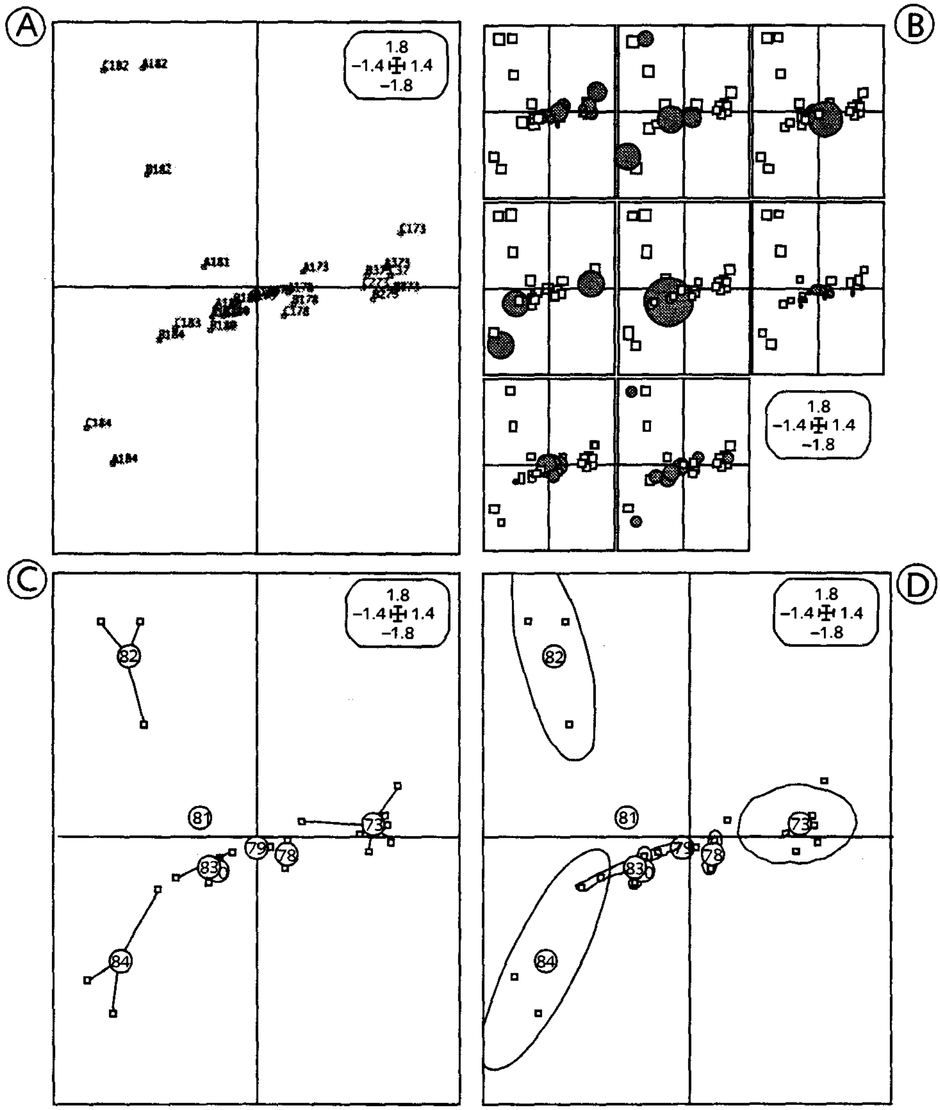

Figure 4: Examples of scatterplots drawn by ADE-4 (data from BOUCHER, 1985). After a COA, rows and columns of the data table (here samples and species) are projected on the factorial axes. Here, the same factorial plane (samples projected on axes 1 and 2) is shown with different points of vue. A: Classical factorial plane of the samples obtained with Scatters, option Labels. B: Projection of an external value on the factorial plane with Scatters, option Values (here relative abundance of 8 species are projected on the samples factorial plane in 8 different windows). Circles stand for positive values and squares for negative values. C: Groups of samples are connected by lines to their gravity centers (circles) in ScatterClass module, option Stars. D: Ellipses are drawn on the same groups of samples with ScatterClass, option Ellipses.

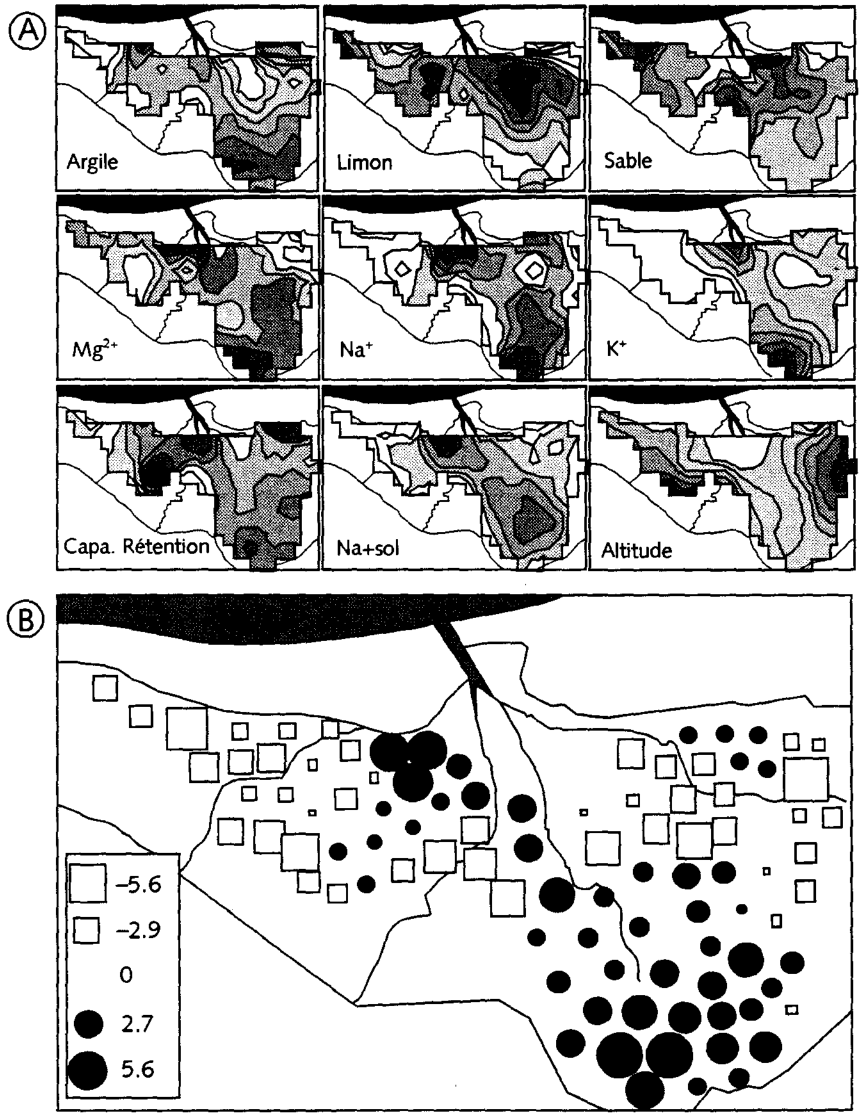

Figure 5: Examples of maps obtained with ADE-4 software (data from BELAIR *et al.*, 1987). A: Grey levels contour curves for 9 environmental variables are automatically displayed on the background map of the studied zone by the Levels module. B: After a normed PCA on the data table, coordinates of the 97 samples on the first factorial axis are projected as squares (negative values) and circles (positive values) on the same background map.

a lowess regression over a number of neighbors chosen by the user see THIOULOUSE *et al.*, 1995 for a description and example of use of this technique in environmental data analysis).

The Maps module has three options: the Labels option simply draws a label on the map at each sampling point, the Values option draws circles (positive values) and squares (negative values) with sizes proportional to the values of the data file (figure 5B), and the Neighbouring graph option draws the edges of a neighbouring relationship between points.

The Areas module draws maps with grey level polygons, starting from a file containing the coordinates of the vertices of each of the areas making the map, and a second file corresponding to the grey levels.

3.5. Dendrograms module

The Dendrograms module can draw hierarchical trees after a classification using Clusters module.

4. The user interface

ADE-4 is made of a series of small independent modules that can be used independently from each other or launched through a HyperCard for the Macintosh version (figure 6) or WinPlus for Windows 95 and NT version interface. There are two categories of modules: computational modules and graphical ones, with a slightly different user interface.

4.1. Computational modules

Computational modules present an "Options" menu that enables the user to choose between the possibilities available in the module. For example, in the PCA (principal component analysis) module, it is possible to choose between PCA on correlation (Normed PCA) or on covariance matrix (Centred PCA) or other variants.

According to the option selected by the user, a dialog window is displayed, showing the parameters required for the execution of the analysis. Values are set using standard dialog windows. At the end of calculations, a text report that contains a description of input and output files and of analysis results is created.

In addition to the statistical modules listed above, several ADE-4 modules deal with file manipulation (import/export of files from and to other softwares, data recoding, categorical variables manipulation...). Other modules provide utilities such as ecological indexes computation (EcolTools module) and radio-tracking (Tracking module).

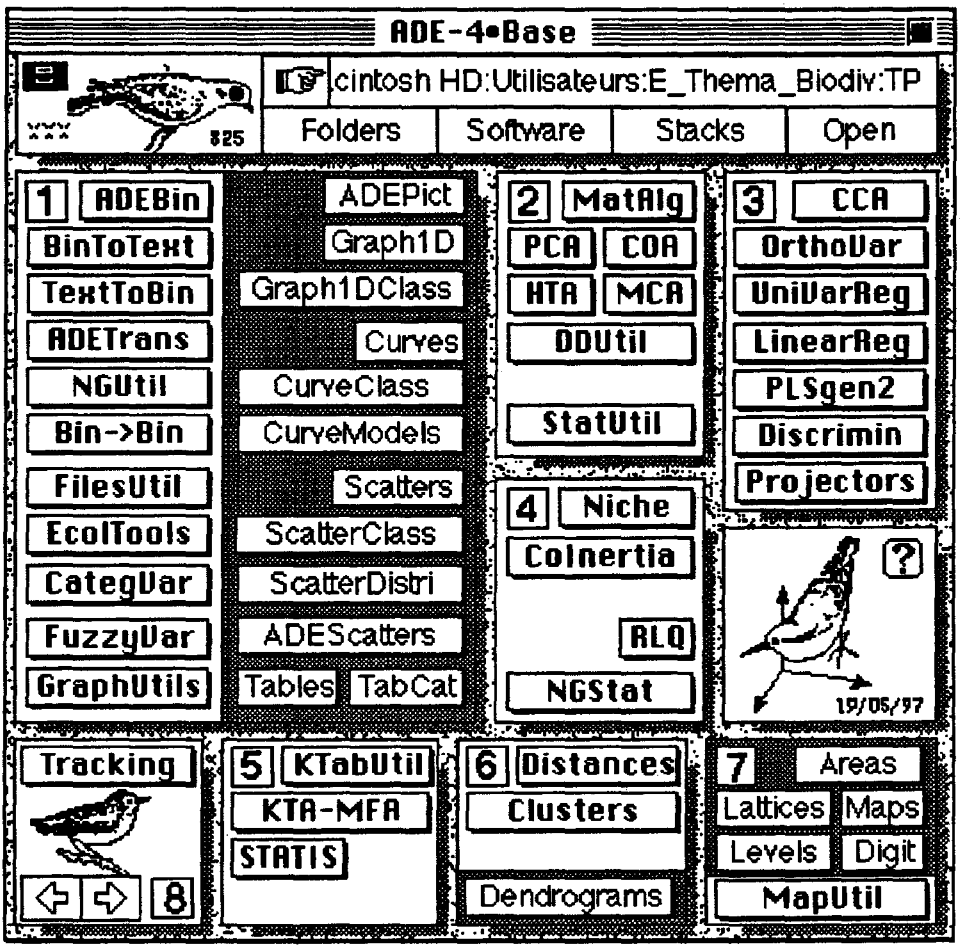

Figure 6: ADE-4 Hypercard (Macintosh) main interface. Computational modules appear in bold font and graphical ones in standard font. 1: File utilities, 2: One-table analyses, 3: Linear regression and two-table coupling methods, 4: Coinertia analysis, 5: *k*-table methods, 6: Distance and partitions, 7: Cartography modules.

4.2. Graphical modules

Graphical modules also present an "Options" menu to choose the type of graphic. They have an additional "Windows" menu. This menu can be used to choose one of the three parameter windows that allow an interactive definition of the graphical parameters. The user can thus freely modify the values of all the parameters and the resulting graphic is displayed in the "Graphics" window.

The main dialog window allows to choose the input file and related parameters. In the "Min/Max" window, the user can set the values of the minimum and maximum of abscissas and ordinates, the number of horizontal and vertical graphics (for graphic collections), the graphical window width and height, legend scale options, etc. In the "Row & Col. selection" window, he can choose the columns and the rows of the data set that will be used to make each elementary graphic of a collection. The rows can be chosen either through a selection file containing a qualitative variable which modalities define the groups corresponding to each graphic, or by typing the number of the rows belonging to each group.

4.3. HyperCard and WinPlus interface

A HyperCard (Macintosh version) or WinPlus (Windows 95 and NT version) stack (ADE-4•Base) can be used to launch the modules. This stack also provides a way to navigate through other softwares and the two other stacks: ADE-4•Data, and ADE-4•Biblio.

ADE•Data (figure 7A) is a library of *ca.* 200 example data sets that can be used for trial runs of data analysis methods (figure 7B). Most of these data sets come from environmental studies. ADE•Biblio is a bibliography stack with more than 1 200 bibliographic references on the statistical methods and data sets available in ADE-4 (figures 7C, D, E).

5. Conclusion

The computing power of micro-computers today is such that the time needed to perform the computations of multivariate analysis methods is no longer a limiting factor. The time needed to compute all the eigenvalues and eigenvectors of a 100 × 100 matrix is only a few seconds. The limiting factor is rather the amount of field work needed to collect the data. This fact has several consequences on multivariate analysis software packages. Monte-Carlo-like methods (permutation tests) can, and should be used much more widely (GOOD, 1994, p. 8). Moreover, it is possible to use an interactive approach to multivariate analysis, trying several methods in just a few minutes. But this implies an easy to use graphical user interface, particularly for graphic programs, allowing to explore many ways of displaying data structures (data values themselves or factor scores). Trend surface analysis and contour curves are valuable tools for this purpose. We also need a graphical software able to display simultaneously all the variables of the dataset, and several groups of samples, corresponding for example to the experimental design (*e.g.*, samples coming from several regions). We have tried to address these points in ADE-4.

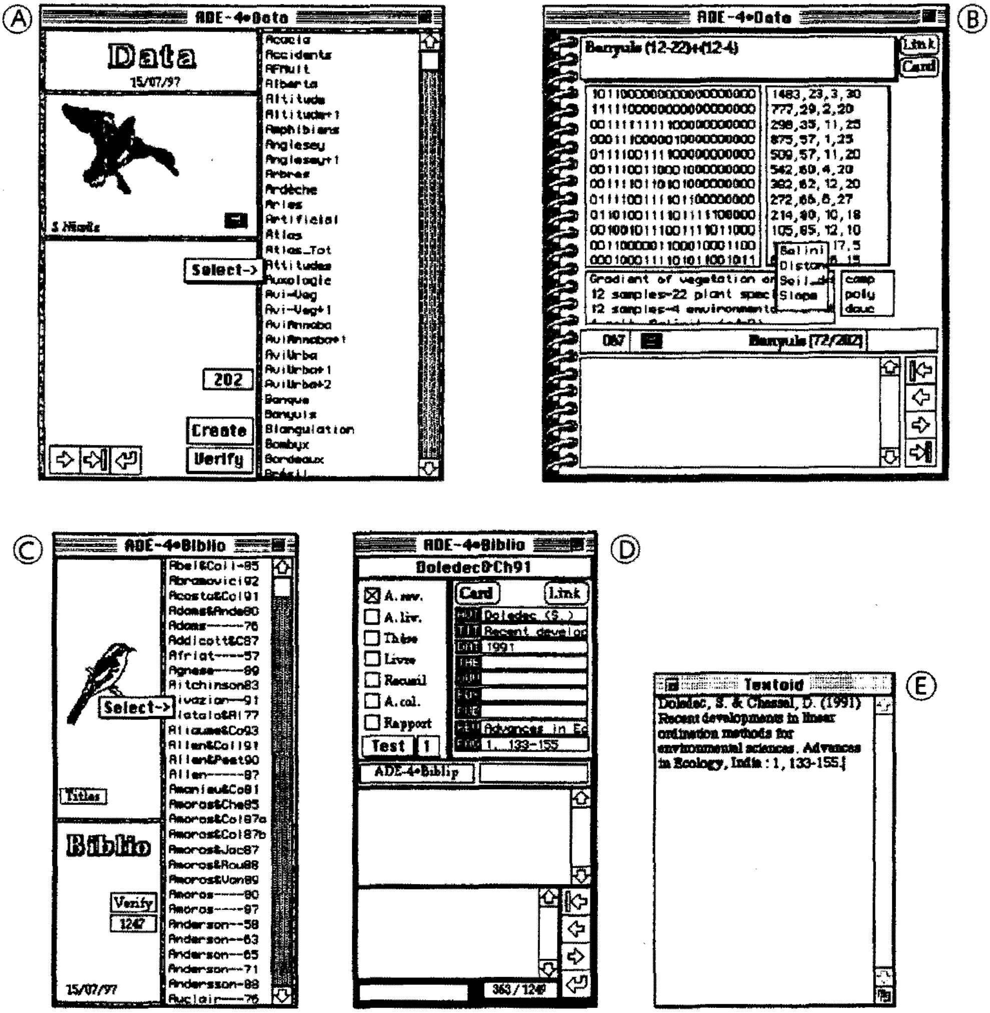

Figure 7: ADE-4 Hypercard Interface Data and Biblio stacks. A: 202 example data sets are available by menu on the Data card; B: User can get data files by clicking on the selected fields of each data card; C: 1247 bibliographic references are available on the Biblio card; D: A biblio card; E: A bibliographic reference can be exported to an external software in text mode.

One of the benefits of having a systematic approach to elementary graphics collection in modules Graph1D, Graph1DClass, Curves, CurveClass, Scatters, and ScatterClass is the possibility to draw automatically all the graphics of a *k*-table analysis (*i.e.*, the graphics corresponding to the analyses of all the elementary tables).

6. Availability

Previous versions of ADE-4 (up to version 3.7) has already been distributed to many research laboratories in France and other countries. Now, the whole software for PC or Macintosh and its documentation (Adobe Acrobat Reader PDF files) can be obtained freely by anonymous FTP to pbil.univ-lyon1.fr, in the /pub/mac/ADE/ADE4 directory, or on the WWW (world-wide web) downloading page, available at: `http://pbil.univ-lyon1.fr/ADE-4/ADE-4.html` (figure 8).

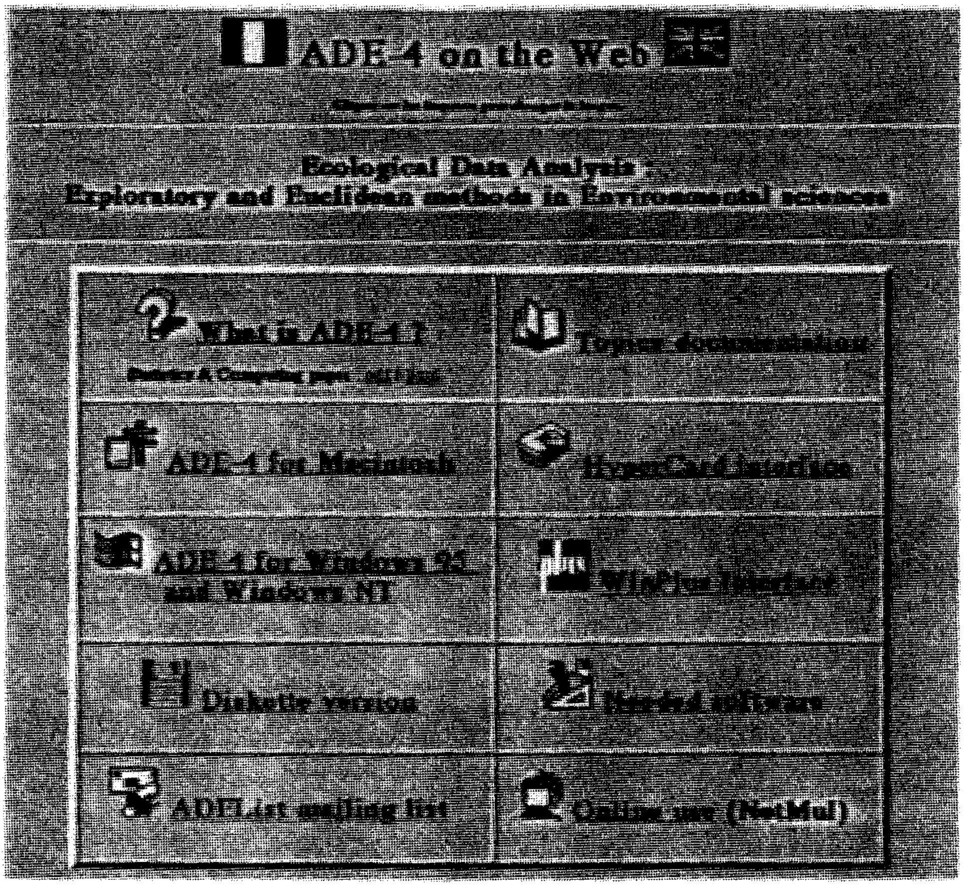

Figure 8: The ADE-4 WWW (world-wide web) homepage.

A sub-set of ADE-4 can also be used on line on the Internet, through a WWW user interface called NetMul (THIOULOUSE & CHEVENET, 1996) at the following address: `http://pbil.univ-lyon1.fr/ADE-4/NetMul.html`. A discussion list has been created on the internet (140 subscribers in December 1997) which is a very useful exchange tool between users and authors of ADE-4 and also provides access to updates and user support.

Acknowledgments

ADE-4 development was supported by the "Programme Environnement" of the French National Center for Scientific Research (CNRS), under the "Méthodes, Modèles et Théories" contract.

References

BELAIR G. de & BENCHEIKH-LEHOCINE M., 1987. — Composition et déterminisme de la végétation d'une plaine côtière marécageuse : La Mafragh (Annaba, Algérie). *Bull. Écol.*, n° 18(4), p. 393–407.

BLANC L., CHESSEL D. & DOLÉDEC S., 1998. — Étude de la stabilité temporelle des structures spatiales par analyses d'une série de tableaux de relevés faunistiques totalement appariés. *Bull. Fr. Pêche Piscic.*, n° 348, p. 1–21.

BOUCHER G., 1985. — Long term monitoring of meiofauna densities after the *Amoco Cadiz* oil spill. *Mar. Pollut. Bull.*, n° 16(8), p. 328–333.

BOUROCHE J.M., 1975. — *Analyse des données ternaires : la double analyse en composantes principales.* Thèse de 3e cycle, université Paris VI.

CADET P., THIOULOUSE J. & ALBRECHT A., 1994. — Relationships between ferrisol properties and the structure of plant parasitic nematode communities on sugarcane in Martinique (French West Indies). *Acta Œcol.*, n° 15, p. 767–780.

CAILLIEZ F. & PAGÈS J.P., 1976. — *Introduction à l'analyse des données.* SMASH, Paris, 616 p.

CASTELLA E. & SPEIGHT M.C.D., 1996. — Knowledge representation using fuzzycoded variables: an example based on the use of Syrphidae (Insecta, Diptera) in the assessment of riverine wetlands. *Ecol. Model.*, n° 85, p. 13–25.

CAZES P., CHESSEL D. & DOLÉDEC, S., 1988. — L'analyse des correspondances internes d'un tableau partitionné: son usage en hydrobiologie. *Rev. Stat. Appl.*, n° 36, p. 39–54.

CHESSEL D. & HANAFI M., 1996. — Analyses de la co-inertie de K nuages de points. *Rev. Stat. Appl.*, n° 44, p. 35–60.

CHESSEL D. & MERCIER P., 1993. — Couplage de triplets statistiques et liaisons espèces-environnement. *In: Biométrie et Environnement* (Lebreton J.D. & Asselain B., eds), p. 15–44. Masson, Paris.

CHEVENET F., DOLÉDEC S. & CHESSEL D., 1994. — A fuzzy coding approach for the analysis of long-term ecological data. *Freshwat. Biol.*, n° 31, p. 295–309.

CLEVELAND W.S., 1979. — Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Ass.*, n° 74, p. 829–836.

CLEVELAND W.S. & DEVLIN S.J., 1988. — Locally weighted repression: an approach to regression analysis by local fitting. *J. Am. Stat. Ass.*, n° 83, p. 596–610.

DEVILLERS J. & CHESSEL D., 1995. — Can the enucleated rabbit eye test be a suitable alternative for the *in vivo* eye test? A chemometrical response. *Toxicol. Model.*, n° 1, p. 21–34.

DOLÉDEC S. & CHESSEL D., 1987. — Rythmes saisonniers et composantes stationnelles en milieu aquatique. I- Description d'un plan d'observations complet par projection de variables. *Acta Œcol./Œcol. Gen.*, n° 8, p. 403–426.

DOLÉDEC S. & CHESSEL D., 1989. — Rythmes saisonniers et composantes stationnelles en milieu aquatique. II- Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Œcol./Œcol. Gen.*, n° 10, p. 207–232.

DOLÉDEC S. & CHESSEL D., 1991. — Recent developments in linear ordination methods for environmental sciences. *Adv. Ecol.*, India, n° 1, p. 133–155.

DOLÉDEC S. & CHESSEL D., 1994. — Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwat. Biol.*, n° 31, p. 277–294.

DOLÉDEC S., CHESSEL D., TER BRAAK C.J.F. & CHAMPELY S., in press. — Matching species traits to environmental variables: a new three-table ordination method. *Environ. Ecol Stat.*

DOLÉDEC S., CHESSEL D. & OLIVIER J.M., 1995. — L'analyse des correspondances décentrée : application aux peuplements ichtyologiques du Haut-Rhône. *Bull. Fr. Pêche Piscic.* (in press).

ESCOFIER B. & PAGÈS J., 1994. — Multiple factor analysis (AFMULT package). *Comput. Stat. Data Anal.*, n° 18, p. 121–140.

Escoufier Y., 1980. — L'analyse conjointe de plusieurs matrices de données. *In: Biométrie et Temps* (Jolivet M., eds), p. 59–76. Société française de biométrie, Paris.

Escoufier Y., 1987. — The duality diagramm: a means of better practical applications. *In: Development in numerical ecology* (Legendre P. & Legendre L., eds), p. 139–156. NATO advanced Institute, Serie G, Springer Verlag, Berlin.

Foucart T., 1978. — Sur les suites de tableaux de contingence indexés par le temps. *Stat. Anal. données*, n° 2, p. 67–84.

Gabriel K.R., 1971. — The biplot graphical display of matrices with application to principal component analysis. *Biometrika*, n° 58, p. 453–467.

Gabriel K.R., 1981. — Biplot display of multivariate matrices for inspection of data and diagnosis. *In: Interpreting multivariate data* (Barnett V., ed), p. 147–174. John Wiley and Sons, New York.

Gaertner J.-C., Chessel D. & Bertrand J., 1998. — Stability of spatial structures of demersal assemblages: a multitable approach. *Aquat. Liv. Res.*, n° 11(2), p. 75–85.

Geladi P. & Kowalski B.R., 1986. — Partial least-squares regression: a tutorial. *Analyt. Chim. Acta*, n° 1, p. 185, 19–32.

Good P., 1994. — *Permutation tests*. Springer-Verlag, New York, 238 p.

Greenacre M., 1984. — *Theory and applications of correspondence analysis*. Academic Press, London.

Höskuldsson A., 1988. — PLS regression methods. *J. Chemomet.*, n° 2, p. 211–228.

Kevin V. & Whitney M., 1972. — Algorithm 422. Minimal Spanning Tree [H]. *Communic. Ass. Comput. Machin.*, n° 15, p. 273–274.

Lavit C., 1988. — *Analyse conjointe de tableaux quantitatifs*. Masson, Paris, 251 p.

Lavit Ch., Escoufier Y., Sabatier R. & Traissac P., 1994. — The ACT (Statis method). *Comput. Stat. Data Anal*, n° 18, p. 97–119.

Lebart L., 1969. — Analyse statistique de la contiguïté. *Pub. Inst. Statist Univ. Paris*, n° 28, p. 81–112.

Lebart L., Morineau L. & Warwick K.M., 1984. — *Multivariate descriptive analysis: correspondence analysis and related techniques for large matrices*. John Wiley and Sons, New York, 231 p.

LEBRETON J.D., CHESSEL D., PRODON R. & YOCCOZ N., 1988a. — L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. I. Variables de milieu quantitatives. *Acta Œcol./Œcol. Gen.*, n° 9, p. 53–67.

LEBRETON J.D., RICHARDOT-COULET M., CHESSEL D. & YOCCOZ N., 1988b. — L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. II Variables de milieu qualitatives. *Acta Œcol./Œcol. Gen.*, n° 9, p. 137–151.

LEBRETON J.D., SABATIER R., BANCO G. & BACOU A.M., 1991. — Principal component and correspondence analyses with respect to instrumental variables: an overview of their role in studies of structure-activity and species-environment relationships. *In: Applied Multivariate Analysis in SAR and Environmental Studies* (Devillers J. & Karcher W. eds), p. 85–114. Kluwer Academic Publishers, Dordrecht.

LINDGREN F., 1994. — *Third generation PLS. Some elements and applications.* Umeå University, Dept of Organic Chemistry, Research Group for Chemometrics, Umeå, p. 1–57.

MAHALANOBIS P.C., 1936. — On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India*, n° 12, p. 49–55.

MANLY B.F., 1994. — *Multivariate Statistical Methods. A primer.* Second edition, Chapman & Hall, London, 215 p.

MANTEL M., 1967. — The detection of disease clustering and a generalized regression approach. *Cancer Res.*, n° 27, p. 209–220.

NÆS T., 1984. — Leverage and influence measures for principal component regression. *Chemomet. Intel. Lab. Syst.*, n° 5, p. 155–168.

NISHISATO S., 1980. — *Analysis of categorical data: dual scaling and its applications.* University of Toronto Press, London, 276 p.

NOY-MEIR I., 1973. — Data transformations in ecological ordination. I. Some advantages of non-centering. *J. Ecol.*, n° 61, p. 329–341.

OKAMOTO M., 1972. — Four techniques of principal component analysis. *J. Jap. Statist. Soc.*, n° 2, p. 63–69.

PERSAT H., NELVA A. & CHESSEL D., 1985. — Approche par l'analyse discriminante sur variables qualitatives d'un milieu lotique le Haut-Rhône français. *Acta Œcol./Œcol. Gen.*, n° 6, p. 365–381.

ROUX M., 1985. — *Algorithmes de classification.* Masson, Paris, 147 p.

ROUX M., 1991. — Basic procedures in hierarchical cluster analysis. *In: Applied Multivariate Analysis in SAR and Environmental Studies* (Devillers J. & Karcher W., eds), p. 115–136. Kluwer Academic Publishers, Dordrecht, The Netherlands.

SAPORTA G., 1975. — *Liaisons entre plusieurs ensembles de variables et codage de données qualitatives*. Thèse de 3e cycle, université Paris VI.

SIMIER M., BLANC L., PELLEGRIN F. & NANDRIS D., 1999. — Approche simultanée de K couples de tableaux : Application à l'étude des relations pathologie végétale-environnement. *Rev. Statist. Appl.*, n° XLVII(1), p. 31–36.

TAKEUCHI K., YANAI H. & MUKHERJEE B.N., 1982. — *The foundations of multivariate analysis. A unified approach by means of projection onto linear subspaces*. John Wiley and Sons, New York, 458 p.

TENENHAUS M. & YOUNG F.W., 1985. — An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, n° 50, p. 91–119.

TENENHAUS M., 1997. — *Association et prédiction en analyse des données. De l'analyse canonique à la régression PLS. Version provisoire*. Groupe HEC, Jouy-en-Josas, France, p. 1–217.

TER BRAAK C.J.F., 1987a. — The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, n° 69, p. 69–77.

TER BRAAK C.J.F., 1987b. — *Unimodal models to relate species to environment*. Agricultural Mathematics Group, Wageningen, 152 p.

THIOULOUSE J. & CHESSEL D., 1987. — Les analyses multi-tableaux en écologie factorielle. I De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Œcol./Œcol. Gen.*, n° 8, p. 463–480.

THIOULOUSE J. & CHESSEL D., 1992. — A method for reciprocal scaling of species tolerance and sample diversity. *Ecology*, n° 73, p. 670–680.

THIOULOUSE J., CHESSEL D. & CHAMPELY S., 1995. — Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environ. Ecol. Statist.*, n° 2, 1–14.

THIOULOUSE J., & LOBRY J.R., 1995. — Co-inertia analysis of amino-acid physico-chemical properties and protein composition with the ADE package. *Comput. Appl. Biosci.*, n° 11, p. 3, 321–329.

THIOULOUSE J. & CHEVENET F., 1996. — NetMul, a word-wide web user interface for multivariate analysis software. *Comput. Statist. Data Anal.*, n° 21, p. 369–372.

THIOULOUSE J., 1996. — Towards better graphics for multivariate analysis: the interactive factor map. *Comput. Statist.*, n° 11, p. 11–21.

THIOULOUSE J., CHESSEL D., DOLÉDEC S. & OLIVIER J.M., 1997. — ADE-4: a multivariate analysis and graphical display software. *Statist. Comput.*, n° 7, p. 75–83.

TOMASSONE R., DANZARD M., DAUDIN J.-J. & MASSON J.P., 1988. – *Discrimination et classement*. Masson, Paris, 173 p.