

ADE SOFTWARE: MULTIVARIATE ANALYSIS AND GRAPHICAL DISPLAY OF ENVIRONMENTAL DATA

J. Thioulouse

*Laboratoire de Biométrie, Génétique et Biologie des Populations
URA CNRS 243, Université Lyon 1
69622 Villeurbanne Cedex, France.*

S. Dolédec, D. Chessel, and J.M. Olivier

*Laboratoire d'Ecologie des Eaux Douces et des Grands Fleuves
URA CNRS 1451, Université Lyon 1
69622 Villeurbanne Cedex, France.*

PROGRAM ABSTRACT

Name: ADE software 4.0. Date of release: 3/95. Developers: Jean Thioulouse, Daniel Chessel, Sylvain Dolédec. E-mail: chessel@biomserv.univ-lyon1.fr - Fax: (33) 72 43 11 41 - Ph: (33) 72 44 82 77. Contact for information: Jean Michel Olivier, Laboratoire d'Ecologie des Eaux Douces et des Grands Fleuves, URA CNRS 1451, Université Lyon 1, 69622 Villeurbanne Cedex, France. Fax: (33) 72 43 11 41 - Ph: (33) 72 43 13 19.

Case studies: Ecology of the Upper Rhône River: a test of habitat templet theories (see references).

SW requirements: MacOS 7.0 or greater and HyperCard 2.2 to use the navigation stacks (optional). Computer type: Macintosh. RAM used: user definable, minimum 800 Kb. Mass storage requirements: 1-10 Mb, according to installed options. Estimated CPU time for standard applications: 1-5 seconds. Availability: ADE is freely available on the Internet by anonymous FTP to biom3.univ-lyon1.fr. It is also available on diskettes.

1. General remarks

ADE (Analysis of Environmental Data) software deals with the multivariate analysis of environmental data sets. ADE 4.0 comes from the combination of ADE 3.7 (Chessel and Dolédec, 1993), and MacMul and GraphMu (Thioulouse, 1989, 1990). In ADE software, a series of multivariate techniques permits to analyze several types of data (e.g., quantitative, qualitative, presence/absence). The examples incorporated in ADE for trial runs mainly deal with ecological questions. However, people from other disciplines (e.g., chemometry, sociology) and interested in the use of linear ordination methods should find in ADE software the most recent developments. They includes one-table, two-table, and K-table ($k \geq 3$) techniques.

ADE software benefits of the user-friendliness of the Macintosh computer (scrolling menu, graphical performance). Furthermore we have designed ADE for non specialists in multivariate analysis and computer sciences. People having a basic statistical background and routinely using Macintosh computers should be able to use ADE software without any difficulty.

2. Multivariate analyses

2.1. One-table ordination

Linear ordination techniques start with the analysis of data sets presented under the form of one array (Dolédec and Chessel, 1991). The ADE software performs three basic multivariate methods. Principal components analysis (PCA) is devoted to quantitative

measurements (Hötteling, 1933). Correspondence analysis (COA) enables the study of contingency tables (counts of the joint occurrences of rows and columns of a table, Hirschfeld, 1935). Multiple correspondence analysis (MCA) permits to analyze qualitative (or categorical) variables (Nishisato, 1980; Tenenhaus and Young, 1985). Furthermore, if the table incorporates *a priori* groups of individuals, the user can process a discriminant analysis - also known as canonical variate analysis - (Mahalanobis, 1936; Tomassone et al., 1988), and between and within class analyses (Dolédec and Chessel, 1987, 1989).

2.2. Two-table ordination

When two arrays are simultaneously analyzed (e.g., a floro-faunistic array and an environmental array to investigate a species-environment relationships), ADE can perform the following methods: (1) analyses with respect to instrumental variables (e.g., CANOCO method of ter Brak, 1986, 1988; Lebreton et al., 1991), and (2) co-inertia analysis (Chessel and Mercier, 1993; Dolédec and Chessel, 1994) and derived methods (such as the PLS regression, Höskuldsson, 1988).

2.3. K-table ordination

The simultaneous ordination of three arrays (e.g., environmental array, faunistic array and species trait array) is available under the name of RLQ analysis (Dolédec et al., in preparation). Collections of arrays (e.g., ternary matrices such as sites x dates x variables) can be analyzed with the STATIS method ("Structuration des Tableaux à Trois indices de la Statistique" method of L'Hermier des Plantes, 1976; Escoufier, 1980; Lavit et al., 1994).

Non parametric tests based on the results of random permutations processed on tables are included to several multivariate analyses (COA, discriminant analysis, co-inertia analysis, and RLQ analysis).

3. User interface

The ADE software incorporates a collection (program library) of computational and graphical modules that perform one or several operations (multivariate analyses or graphical display). The user can launch these modules directly from the Macintosh Finder (stand-alone use). All the modules of the same type have the same user interface, with just an **Options** menu that varies. For example, the menu of the PCA module is as follows:

Options
<input checked="" type="checkbox"/> Correlation matrix PCA
<input type="checkbox"/> Covariance matrix PCA
<input type="checkbox"/> After row % transformation PCA
<input type="checkbox"/> Non centred PCA
<input type="checkbox"/> Decentring $H[i,j] - Model[j]$
<input type="checkbox"/> Decentring $H[i,j] - Model[i,j]$
<input type="checkbox"/> Partial normed PCA

For a more comfortable use of the program library, the user of ADE may take advantage of the HyperCard interface composed of four HyperCard stacks (this requires HyperCard 2.2): (1) *ADE•Base* is the main selection stack. It is used to launch the modules (Fig. 1). In that case, this stack is essential for operating the program library. It

also gives access to the other stacks. (2) *ADE•Biblio* contains up to 870 bibliographic references. (3) *ADE•Data* includes up to 140 numerical examples. (4) *ADE•Doc* (Documentation) contains the information about the function and the use of modules.

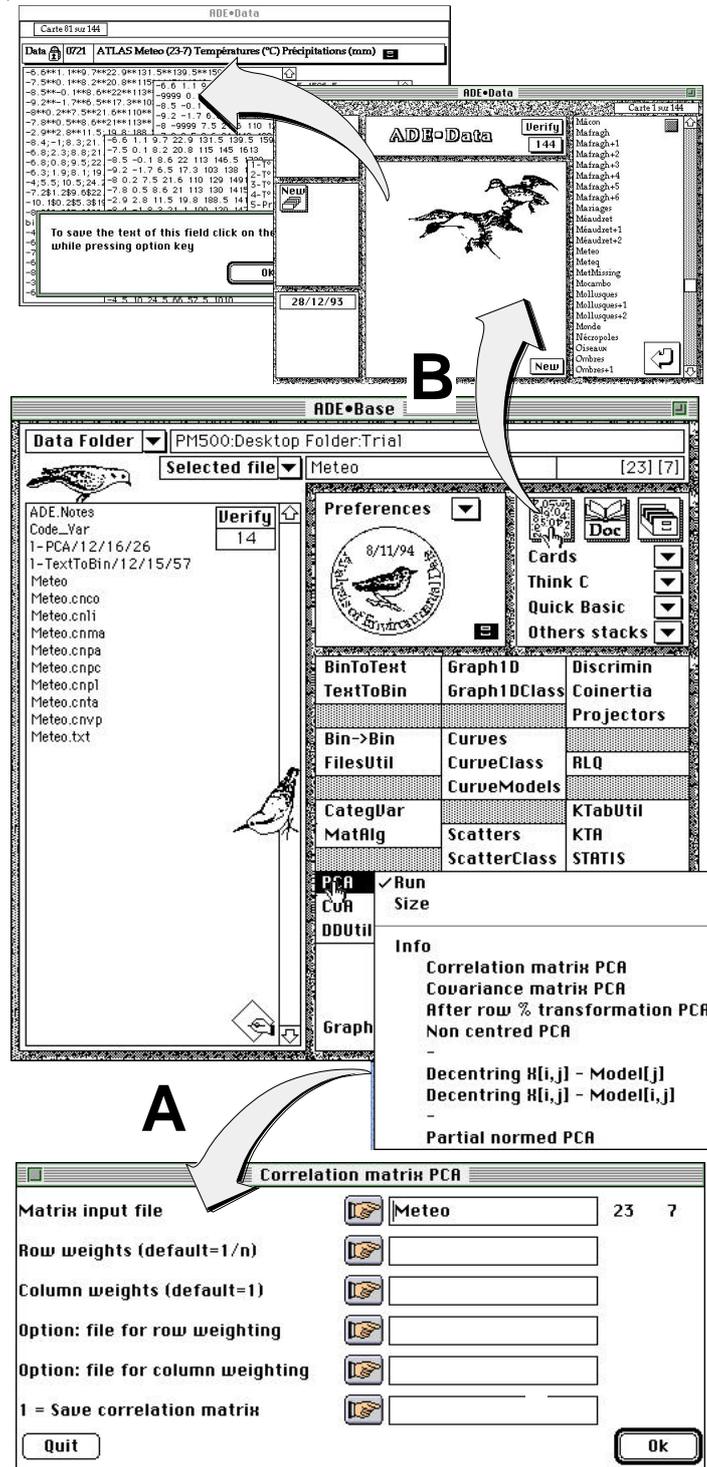


Figure 1. Operating the *ADE•Base* selection card. The left hand field of this card contains a list of the files included into the user data folder. (A) The **PCA** module and the *Correlation matrix PCA* option were selected. Information (Info) about the various options of the module is available from this menu. The *Run* option launches the corresponding module, which main dialog window allows the user to select the input

files and other parameters. (B) Other stacks such as the *ADE•Data* stack can be launched from *ADE•Base*. We present here an example of *ADE•Data* card. These cards include a series of data fields that can be stored into the user data folder.

Each stack incorporates a number of cards. The cards included into *ADE•Data* are of special interest because the data incorporated to a given card may be copied and then processed with modules.

All the analyses and graphics operate in a given directory. This directory is selected or created with the *Data Folder* menu (Fig. 1).

4. Graphical display

The ADE software puts the emphasis on graphical methods according to the theoretical background developed by Thioulouse et al. (1991). Numerous graphical programs are available in ADE software (see example in Statzner et al., 1994). The user can draw graphics according to multivariate analyses (e.g., factor maps), cartography (e.g., isolevel curves (Fig. 2), gray level maps, maps with circles and squares), and several other kinds of graphical display (e.g., histograms, Gauss curves, ellipses, convex hulls). Automatic collections or superimpositions of graphics are available to represent either the variables (columns) of a data set or groups of individuals (rows).

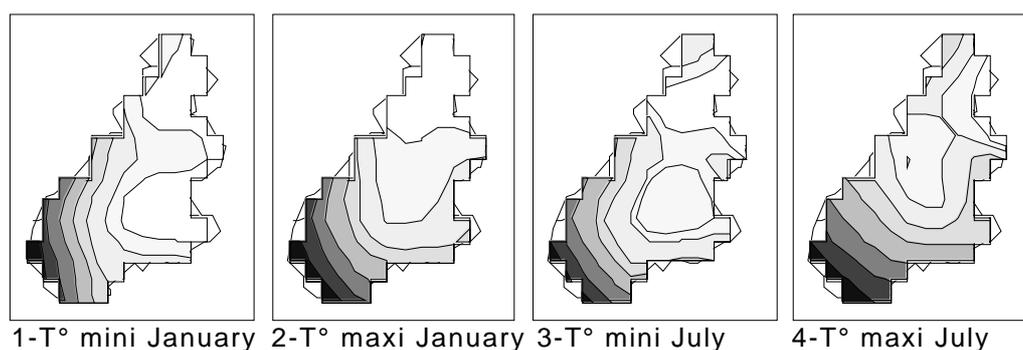


Figure 2. Example of isolevel curves using 9 neighbours. A number of graphic modules (e.g. lattice grid) are dedicated to the handling of spatial arrangement of data. A multiwindowed option is available for that kind of graphics. In this example, four meteorological variables were recorded into 23 districts.

As for computational modules, the user selects an item in the **Options** menu (Fig. 3A). Graphical modules have an additional **Windows** menu (Fig. 3B), from which the user can navigate among the various windows that can be used to set the graphic parameters (Fig. 4).

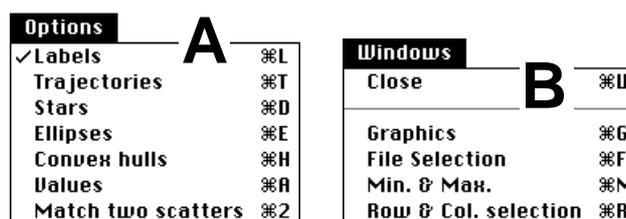


Figure 3. **Options** (A) and **Windows** (A) menus of a graphical module.

The **Min/Max** window (Fig. 5A) allows the user to set the values of several parameters (e.g., minimum and maximum values, number of horizontal and vertical graphics in the case of a collection of graphics, graphical window height and width). The **Row & col.** selection window (Fig. 5B) enables the management of a collection of graphics (each column and each group of rows will make one graphic).

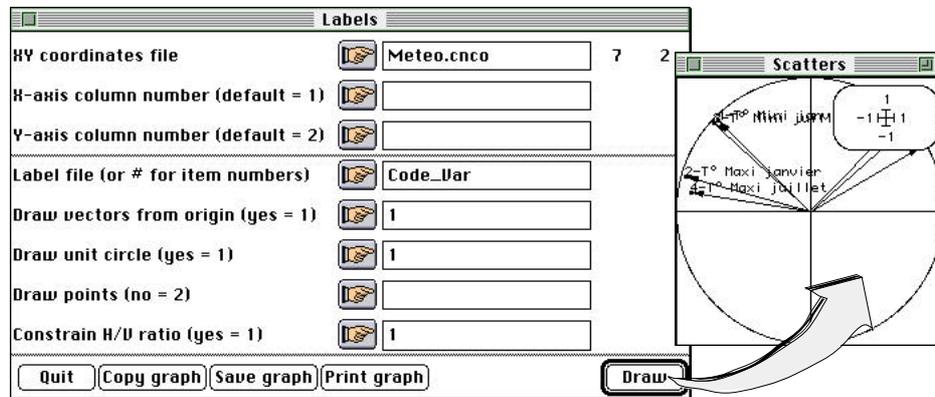


Figure 4. Initiating a correlation circle with the label item of the Options menu.

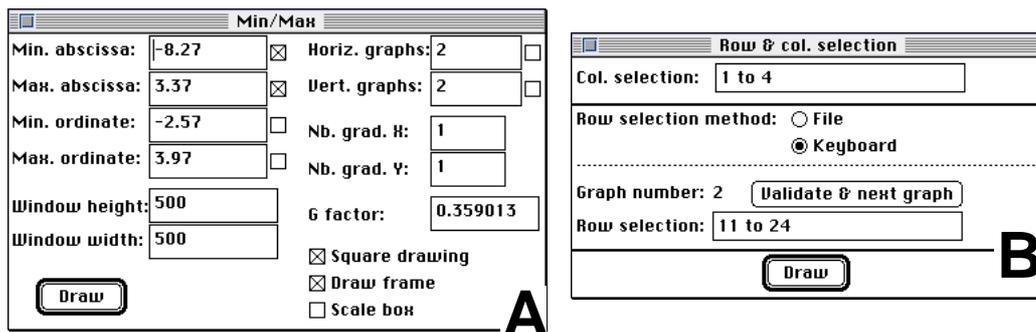


Figure 5. (A) **Min/Max** dialog window for changing the parameters of the set of graphics. (B) **Row&Col.** selection dialog window for separating the entire graphic into sub-graphics using a selection of rows and columns.

ADE software 4.0 is compatible with any software able to export data arrays as plain text files (e.g., Excel, StatView, Word). Graphics can be stored as PICT files that can be handled with commercial software compatible with this format (MacDraw, SuperPaint, ClarisDraw, and many others).

References

- Chessel, D. & Dolédec, S. (1993). *ADE Version 3.7: HyperCard © Stacks and QuickBasic Microsoft © program library for the Analysis of Environmental Data*. URA CNRS 1451, Université Lyon 1, Villeurbanne, France.
- Chessel, D. and Mercier, P. (1993). Couplage de triplets statistiques et liaisons espèces-environnement. In: *Biométrie et Environnement*, Lebreton, J.D. and Asselain, B. (Eds.), pp. 15-44. Masson, Paris.

- Dolédec, S. and Chessel, D. (1987). Rythmes saisonniers et composantes stationnelles en milieu aquatique I- Description d'un plan d'observations complet par projection de variables. *Acta Oecologica, Oecologia Generalis*, vol. 8, pp. 403-426.
- Dolédec, S. and Chessel, D. (1989) Rythmes saisonniers et composantes stationnelles en milieu aquatique II- Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Oecologica, Oecologia Generalis*, vol. 10, pp. 207-232.
- Dolédec, S. and Chessel, D. (1991). Recent developments in linear ordination methods for environmental sciences. *Advances in Ecology, India*, vol. 1, pp. 133-155.
- Dolédec, S. and Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology*, vol. 31, pp. 277-294.
- Dolédec, S., Chessel, D. and Champély, S. (in preparation). A new three-table ordination method for matching species traits to environmental variables.
- Escoufier, Y. (1980). L'analyse conjointe de plusieurs matrices de données. In: *Biométrie et Temps*. Jolivet, M. (Ed.), pp. 59-76. Société Française de Biométrie, Paris.
- Hirschfeld, H.O. (1935). A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences*, vol. 31, pp. 520-524.
- Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, vol. 2, pp. 211-228.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, vol. 24, pp. 417-441.
- L'Hermier des Plantes, H. (1976). *Structuration des tableaux à trois indices de la statistique. Théorie et applications d'une méthode d'analyse conjointe*, Thesis, USTL, Montpellier.
- Lavit, Ch., Escoufier, Y., Sabatier, R. and Traissac, P. (1994) The ACT (Statis method). *Computational Statistics and Data Analysis*, vol. 18, pp. 97-119.
- Lebreton, J.D., Sabatier, R., Banco, G. and 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. and Karcher, W. (Eds.), pp. 85-114. Kluwer, Dordrecht.
- Mahalanobis, P.C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, vol. 12, pp. 49-55.
- Nishisato, S. (1980). *Analysis of caregorical data: dual scaling and its applications*. University of Toronto Press, London.
- Statzner, B., Resh, V.H. and Dolédec, S. (1994). Ecology of the Upper Rhône River: a test of habitat templet theories. *Freshwater Biology*, vol. 31, pp. 253-554.
- Tenenhaus, M. and 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*, vol. 50, pp. 91-119.
- ter Braak, C.J.F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, vol. 69, pp. 69-77.
- ter Braak, C.J.F. (1988). Partial Canonical Analysis. In: *Classification an related methods of data analysis*. Bock, H.H. (Ed.), pp. 551-558. North Holland.
- Thioulouse, J. (1989). Statistical analysis and graphical display of multivariate data on the MacIntosh. *Computer Applications in the BioSciences*, vol. 5, pp. 287-292.
- Thioulouse, J. (1990). MacMul and GraphMu: two Macintosh programs for the display and analysis of multivariate data. *Computers and Geosciences*, vol. 8, pp. 1235-1240.
- Thioulouse, J., Devillers, J., Chessel, D. and Auda, Y. (1991). Graphical techniques for multidimensional data analysis. In: *Applied Multivariate Analysis in SAR and*

Environmental Studies. Devillers, J. and Karcher, W. (Eds.), 153-205. Kluwer, Dordrecht.

Tomassone, R., Danzard, M., Daudin, J.J., and Masson, J.P. (1988). *Discrimination et classement*. Masson, Paris.