Coenoses 8(2):97-104, 1993

COMPUTER ASSISTED PROCEDURES FOR STRUCTURING COMMUNITY DATA

S. Camiz

University of Rome, via S. Francesco a Ripa, 57, I-00153 Rome, Italy.
E-mail: Camiz@Itcaspur.caspur.it

Keywords: Analysis of concentration, Canonical contingency tables analysis, Cluster analysis, Communities, Computer programs, Correspondence analysis, Data analysis, Eigenanalysis, Exploratory analysis, Multivariate analysis, Table structuring, Vegetation.

Abstract. Computer assisted procedures are proposed to serve as a helpful tool for structuring vegetation data tables. Data analysis and typical procedures used for this purpose are discussed in the context of the theoretical steps involved.

Introduction

In a previous paper (Camiz 1988) expert system definitions were introduced and their use considered in the field of community data analysis. The definition of expert system is taken from computer science (Barr and Feigenbaum 1981-82, Hayes-Roth *et al.* 1983, Watermann 1986) as a program having two complementary ways of interacting with a user: knowledge acquisition and consultation modes. Through these the program acts as an interface between expert and non-expert users. The ability of operating in two different modes, which distinguishes the expert system from other computer programs, allows it to perform very complicated and delicate tasks typically done by the human expert. When designed as an expert system, a computer assisted procedure is able to perform automatically different tasks that can be arranged in different ways, to modify action according to intermediate results, and to guide a non-expert user through an optimal path to his objective.

Actually, one can find contiguity from expert system to computer assisted procedures, whose difference is in the program structure. An expert system can handle the knowledge as a set of data, collected in acquisition mode, including new reasoning rules that can be added from time to time without any change in the program itself, and merged with rules already stored, in order to improve the system's experience. In particular, the system is able to store reasoning on successes and failures in the form of data.

An expert system can be conceived as a driver for several computational tasks, that can be executed according to some rules stored in its knowledge base. If the system can test the results of the performed tasks, it can adjust its reasoning by using the stored rules for further tasks. Eventually, new rules may be added to the knowledge base to take into account the method, results of previous cumputations, and theoretical developments. It is likely that the best way is to allow the expert system to create a procedure, a sequence of commands for running programs in batch mode, and enable in to examine some results in order to modify its subsequent action. This is different from a computer assisted procedure which is only a program that gets back some results, but retains in its source all its knowledge, so that no improvement in its reasoning is possible, except by modifying the source code.

Vegetation analysis is a typical field where computer assisted procedures, including expert systems, can find their application. In fact, many attempted to develop expert systems in statistics and data analysis. Gale and Pregibon (1984), many papers in De Antoni *et al.* (1986), and Gale (1987) represent typical examples. Experience suggests that this can be a field in which computer assisted procedures not only can help users to perform better the analysis, but importantly, also to speed-up the interaction with the computer. This means that after the user supplied the computer with the essential information, he concentrates his attention on the results. After careful examination of the results, the user may modify some steps of the procedure to improve the analysis. In practice, a computer assisted procedure would interact with some existing statistical package, with its data archives, and it would allow the user to perform directly the analysis.

The problem is highly relevant since many statistical packages, in particular those of highest quality an richest choices, are often very difficult to handle without continuous consultation of manuals. One should realize that it is not easy to build a complete analytical procedure, and even if one learned the program's control language and can issue the correct sequence of the instructions needed to execute a task,

one must also remember the special options appropriate at every step and be familiar with the method to invoke the right choices. Furthermore, the amount of data output has to be controllable, and ideally it should in the first phase of the analysis be very limited. This is not so in most applications.

The programmers of high quality statistical packages should be ready to ameliorate the code by a friendly interface, since it can reduce user interactions with the package. SPAD, a well-known program for exploratory data analysis from the French School, is an example. The way the user can communicate instructions to the program is typically a file containing several procedure calls, one for each of the required tasks, each one followed by a series of procedure parameters, and other instructions necessary in particular operations. In most cases it is impossible to build such a file without manual in hands. In SPAD.n Intégré (Lebart *et al.* 1991), the latest release for PCs, the user is provided with a very friendly interface, structured with pull-down menus and sub-menus. User choices are then guided by the interface and stored in a data file as instructions for the program. Only for particular purposes, or in order to modify the analytical procedure after a run, one needs to change the procedure calls "à la carte", which is done by editing directly the parameters and instructions.

An old procedure for structuring vegetation tables

In particular handling of vegetation data, the package MULVA-4 (Wildi and Orlóci 1990), is very well designed in that it minimizes the need of external help to perform the desired computations. MULVA-4 is a set of specialized programs, each one devoted to a particular task. These programs communicate through data sets. Each program has a reduced number of choices which are shown to the user as a list of possible options. It is easy to make a choice at every step given user familiarity with the theoretical coherence of selected options and with the correct sequence of executed programs. Considering the broad group of users, it would be desirable to have a driver program which would link and perform automatically at least some of the most common computations. The MULVA-4 structure is such that it is easy to create such a driver program. In fact, based on a previous release of this package (Wildi and Orlóci 1983), a computer assisted procedure has been developed for processing on UNIVAC and VAX mainframes (program SWILDI by Camiz 1988). This version produces a structured vegetation table with very little effort in the way of user-computer interactions. SWILDI performs its tasks in three steps:

A. It requires specifications:

1) whether data are recorded according to presence/absence, the van der Maarel transformation (Westhoff and van der Maarel 1973), or the Braun-Blanquet mean class values (Wildi and Orlóci 1983);

2) if species with low occurrence should be eliminated;

3) if a presence/absence or a quantitative analysis is needed;

4) if standardization is required and if centered or non-centered sum of squares are needed.

After receiving responses under A. the program establishes a procedure by running the Wildi an Orlóci programs: INIT (after prior data matrix transposition, if the parameter loaded for it specifies species as row entities); EDIT to remove rare species; INIT again for reinitialization; RESE for computations of resemblance matrices according to various functions described in Orlóci (1978); PCAB for principal components analysis; ORDB for graphics, limited to the first three axes; CLTR for cluster analysis, using complete linkage for species and minimum variance for relevés. Finally, it displays fusion levels and their first derivative for both relevé and species dendrograms. The latter allow the user to choose among classifications.

B. After these steps, the program is restarted with new requests for information from the user:

5) the number of comparisons between classifications of relevés and species performed through analysis of concentration (Feoli and Orlóci 1979, Orlóci 1991a);

6) the number of relevé and species groups to be used in each of the analyses of concentration. Based on user specifications, the procedure runs programs CLTR for cluster analysis, AOCL for analysis of concentration, and ORDB for 2-dimensional graphics for each requested classification.

C. Finally,

7) the chosen cross-classification is rerun in order to memorize it for further steps;

8) the order is requested in which the groups of relevés and species will appear in the structured table;

9) the re-introduction of species is offered as an option.

Following these, programs are run: INIT on the complete data set, RESE to generate a sum of squares matrix including all species, IDEN for allocation of low occurrence species to existing groups, EDIT to rearrange orderings, and TABS for table printing. The program printout includes the output of all running programs, which could be very large, but since this was saved in a file, it can be manipulated with a text editor.

At the time of SWILDI's implementation, no particular facility was available to run MULVA-4 procedures automatically, but later Wildi (1988) discussed the problem of building procedures for running many steps in batch mode. The latest release is far more comfortable to use in this respect. In the light of this development, it appears timely to call for a discus-

sion of problems involved in the design of a new computer assisted procedure.

Structuring vegetation tables

The table structuring procedures (see Podani and Feoli 1991, and references therein) in vegetation analysis rearrange the relevés and species in such a way that the table takes on the possible sharpest block structure. The blocks represent concentrations of species scores within groups of similar relevés. In order to evaluate the success of table structuring by concentration, Feoli and Orlóci (1979) used canonical contingency table analysis (Orlóci 1978, 1991) which they described as "concentration analysis" to distinguish it from the several variants that already existed, most notably correspondence analysis (Benzécri 1973) and reciprocal averaging (Hill 1973, 1974; Orlóci 1974). Mathematically, structuring is defined as a combinatorial problem, but the task aiming at rearranging relevés and species in all possible ways, and look for an optimum solution, or several optima, that give the best results, would be unmanageable. The matter of fact is that an optimal maneuver would require N! x M! different analyses (N and M being the number of relevés and species respectively) which could require several billion years of computational labor to rearrange a medium size table. The practical approach is therefore to build a structured table according to some a priori model, that, even when leading to a sub-optimal solution, may further the objective of laying open the data and to reveal facts that help the understanding of the vegetation which the data describe. In particular, multivariate techniques claim to extract the most interpretable part of information from the data, specifically that portion which is tied to dominant environmental gradients.

Many different techniques are described in the literature, each deriving its justification from some mathematical reference model (see Dale, 1988, Camiz 1991, Feoli and Orlóci 1991 and references therein), as extensions to classical data analytical techniques, that will be considered in the following sections, and others that are built around uncommon, sometimes bizarre concepts. Lagonegro and Feoli (1979, 1984) based analysis on the intersection of sets of relevés in species terms and Feoli and Zuccarello (1986) defined an ordination based on a previous classification, and a matrix which comes about by multiplication of the data matrix with the transposed matrix of the average values of the original variables in the cluster. As a result, they claim more interpretable ordination. Roberts (1986) ordinated vegetation as fuzzy sets, a concept that also reappeared in other works (e.g., Feoli and Zuccarello 1988, Dale 1988, Marsili-Libelli 1989). Orlóci and Orlóci (1985, Orlóci 1988, 1991d) departed from the vector description of vegetation, relying on score matrix description (relevé) by application of an intersection taxonomy, and developed techniques of analysis to handle score matrix relevés. Pillar and Orlóci (1991) added further to the methodology by considerations of fuzzy community components in score matrix manipulations. Wildi and Orlóci (1987) tried flexible criteria to create a stable ordination, and Orlóci and Pillar (1989, Orlóci 1991b) placed the evaluation of sample size optimality (a stopping rule in sampling) on the structural stability of the sample rather than on a priori manipulations with the idea of the sampling error. Orlóci (1991c) has analyzed the sampling and analytical enviroment in bioenvironmental work and rejected a Fisherian scenario in favour of Poorean approximation.

Clearly, the problem area is extremely broad in a corner of which lies vegetation table structuring and the analytical manipulations that it entails. Wildi (1989) discussed these in some detail and suggested pathways through MULVA-4 that may give optimal results. He used as an example the Ellenberg's grassland data set (Mueller-Dombois and Ellenberg 1974) rearranged by a non-numerical method. His results notwithstanding, further adjustments of his proposals are appropriate, specifically to avoid some superfluous data manipulations and to simplify the theoretical approach. By doing this, a structure is developed which can serve as background for further developments. In fact, the first suggestion is to include an initial exploratory procedure, easily understood by the user, and continue with analyses that involve increasingly more specific and more sophisticated manipulations to reveal deeper insights.

Procedure A: data explorations

A rather simple, yet potentially informative initial analysis can be constructed as follows: 1) compute a suitable resemblance matrix between variables; 2) perform eigenanalysis on the matrix; 3) choose a suitable number of axes, according to some criteria and explain the axes' ecological significance; 4) classify the relevés based on their coordinates on the chosen axes by some suitable classification method; and 5) describe the clusters in terms of their position on the ordination axes and in the ordination planes, and in terms of the original variable values. The mathematical model underlying this procedure entails the original variables as generators of a vector space, sustaining the affine space of individuals (Bourbaki 1962, Godement 1966, Camiz 1991). The resemblance matrix may obey the definition of a scalar product, in order to render analytical space Euclidean, although even weaker conditions allow an Euclidean representation (Gower 1966). The eigenanalysis (such as principal components analysis, Hotelling 1933) aims at building an orthonormal basis for this space, having the property that each reduced p-dimensional subspace, whose basis corresponds to the first p eigenvectors, is the subspace of dimension p that loses minimum information compared

to the whole space. The individuals' coordinates on the axes reflect their position in that space. Consequently, classification is based on the proximity of individuals in the reduced space. In case of correspondence analysis (that can be conceptualized as an eigenanalysis with pre- and post-eigenanalysis adjustments, Orlóci 1978, 1991b), an alternative model is invoked, i.e., the projection of two vector spaces, the row entities and column entities, into a subspace that best approximates their common structure. In vegetation work one advantage of this is that species and relevés both may be represented simultaneously in the reference space of the ordination axes and correspondences between them can be revealed. The eigenvectors of the two sets are functionally linked an the squared roots of the eigenvalues are canonical correlations, i.e., measures of the strength of correspondence of the two sets. By the very nature of the mathematics involved, these correlations measure linear correspondences. To understand the implications of this, it is sufficient to think of a contingency table whose elements are measures of species performances (occupancy counts, total densities) and the columns are categories of an environmental variable, such as pH. If species performance does not have a simple linear relationship to pH, or equivalently, the species with one another in terms of the correlation of their responses ordered by pH, the canonical correlations will be meaningless as phytosociological measures of correspondence. Usually, canonical contingency table analysis is taken one step further, namely to the construction of deviations partitons (Orlóci 1991a) to gain maximum information about correspondences.

The analysis is usually considered successful, if a small number of axes is obtained that represent a large proportion of information in the data set. But the true test of success is not this; the structural and ecological information gained as a consequence of the analysis is wath matters most. Unfortunately, it may happen that some results from these procedures are not interpretable in such terms. Nevertheless, even then the analysis as an exploration of the data can suggest patterns and pattern connections that should be tested in a confirmatory analysis.

Even with very high quality program packages, running a simple procedure like the one just described can be a serious problem even for the experienced user, since each analytical step may involve delicate choices that must be theoretically justified. For this, the underlying theory has to be understood and the package control language (dialect) must be intelligible. In particular, the choise of the resemblance matrix to be computed is an example. This is not a simple matter, since consistency must be maintained as pointed out in Orlóci (1974) through the entire analysis as the computations move from one method to the next. This point was totally missed in many comparative studies from well-known schools dealing with ordinations. The following section describes some characteristics of a stepwise computer assisted procedure:

1) Usually covariance or correlation matrices are used as a based of comparison in the case of ratio-scale data. Correlation is chosen if the variables are measured in different units; Spearman or Kendall rank correlation (Siegel, 1956) may be appropriate for ordinal data. But these may not be ideal. On this point different considerations come into play with different types of data under different objectives. This topic is well described in the literature (e.g., Benzécri 1973, Lebart *et al.* 1977, Orlóci 1978, Legendre and Legendre 1983, Wishart 1978). To compute a suitable matrix, ideally the program should ask the user what kind of data are analyzed and it should suggest or decide itself which resemblance measure is suitable for the data and for the purpose.

2) The program should check the type of resemblance computed and decide which type of ordination is appropriate. The user should be informed and given an opportunity to over-rule the program. The centering or non-centering of the solution is a case in point which may depend on the users' particular notion about the structures and relationships to be investigated. Gower's (1966) Principal Coordinates Analysis is applicable to many association indexes, while Correspondence Analysis (Benzécri 1973; Lebart *et al.* 1977) is based on chi- square distance.

3) The program must provide the user sufficient information about the run, particularly threshold values for the evaluation of the significance of axes, groups, etc. These are needed to decide, for example, the limit under which eigenvalues should be dropped or the level at which clusters should be interpreted. The rapidity of decline in eigenvalues or sums of squares, for instance, is important in this respect. It is evident that an effective follow-up analysis should be contemplated for the further scrutinization of trended (non-random) variation in the analytical space of the initial variables. The program should assist the user with the isolation of random variation. Pragmatic approaches to this use definitions that interpret the lack of trends with regards to the orderings of the reference space. Specifically, variation may be random with regard to one set of axes and not with regard to others.

4) Since the analysis begins with ordination, it may be best to base classification on the ordination coordinates, limited to the number of significant axes as discussed under (3). Since in exploratory analysis the number of groups is not known in advance, a computer assisted procedure should be able to suggest an initial number of clusters based on some criteria in sensible computational terms: a) a resemblance criterion (Orlóci and Kenkel 1985) in which regard the program should not only give options but inform the users about the advantages or disadvantages of the options; The selction of the criterion, such as minimum sum of

squares (Ward 1963, Orlóci 1967, Wishart 1978, Anderberg 1973) is critical, and so is consistency with the clustering method and data structure; b) a most "natural" classification is combinatorial such as in Edwards and Cavalli-Sforza (1965), but this needs enormous computation effort and a suboptimal solution may have to be accepted, preferably still hierarchical (Anderberg, 1973) for general purposes, refined by iterative reallocations (Wishart, 1978, Wildi and Orlóci, 1990; Lebart *et al.*, 1991); c) the rule for deciding where to cut a dendogram to obtain groups in which regard suggestions are found in Sandland and Young (1979), Mojena and Wishart's (1980) and in all the other references already quoted; in general, localization of the longest branches of dendogram can be performed via the second derivatives of the fusion levels (possibly negative values).

5) It is evident that the program should leave the final choice of classification to the user. A good way for helping the users in the analysis and description of the obtained groups is to provide them with the cluster descriptors such as the mean values of the original variables in each cluster, their dispersion or distribution and possibly, the outcome of statistical tests preferably based on randomization procedures such as described in directly comparable terms by Edgington (1987). Yet, the printout should be limited to the essentials needed for comprehension. In this way, the researcher's attention can be directed away from computer interaction to the reading and comprehension of the printouts. Only later, there may be a need for further details from specific steps of the analysis under available options.

Procedure B: structuring a vegetation table

Noy-Meir (1971, 1974) suggested that multivariate methods are useful in vegetation study because they can make a complicated picture simple for comprehension. If the vegetation entities can be considered as spaces of linearly correlated species, the linear model of principal components analysis can produce the simple picture (Orlóci 1978, 1979, 1988, 1991c). Formally, a vegetation table can be considered a contingency table and canonical contingency table analysis is likely to be a more appropriate method than principal components analysis for the porpuse of ordination, notwithstanding the linear correlation constraints that still have to be obeyed. To satisfy this constraint the coenocline width has to be reduced in the sample. Alternatively, the nonlinearity problem manifested by a "horse-shoe effect" will still remain and combine for added complications with the effect of pre- and post-eigenanalysis adjustments. The treatment of the nonlinearity problem is, of course, a task performed best in other data manipulations (see references to early work in Orlóci 1978, 1979, and for latter work in Fewster and Orlóci 1983, Orlóci 1988, 1991c, Kenkel and Orlóci 1986, Bradfield and Kenkel 1987, and Camiz 1991).

Procedure A, while effective in understanding the data structure, is not sufficient to construct a contingency table. For this, additional cluster analysis is needed to derive optimal row groups and column groups, a task in the general procedure of concentration analysis. Once the table is constructed, CONAPACK is available to perform canonical contingency table analysis (Orlóci 1991a). Procedure B is this type of analysis: 1) reduce the raw data table by eliminating the single-occurrence species; 2) compute a resemblance matrix between species and another between relevés. 3) create species and relevé groups in a cluster analysis; 4) perform canonical contingency table analysis on all chosen cross-classifications; 5) consider as an indication of the likely number of gradients the number of canonical correlation coefficients higher than a threshold value; 6) choose as the optimum cross-classification based on the value of the highest canonical correlation; 7) Use the canonical scores for purposes of ordination and the deviations graphs for the purpose of seeing performance variations of species among the relevé groups; 8) repeat the steps with other resemblance measures based on the chosen dimensions under (2) and (3); 9) rearrange the data table according to the final set of canonical scores on the first axis.

The elimination of single-occurrence species helps to reduce both species number and relevé number in the analysis (Wildi 1989, Wildi and Orlóci 1990). Those removed can be reassigned at the end by a producere described in the enlarged version of the manual for CONAPACK. The choice of a preclassification resemblance measure depends strictly on the type of data and the type of cluster structure of interest. Canonical correlation coefficients measure the sharpness of block structure, but in an evaluation of the best cross-classification on this basis one must consider that by increasing the number of groups, the correlation coefficient may raise. It is normal to have a horseshoe type arrangement of relevé and species groups in the ordination. This is actually the consequence of mapping a nonlinear data structure generated by linear correlations among the rows (columns) of the contingency table owing, indirectly, to nonlinear species responses (Orlóci 1991). The use of empirical adjustments for this, as in detrended correspondence analysis (Hill 1979) does not solve the problem for reasons that Feoli and Orlóci (1991) explains, instead adding local distortion as Camiz (1991) shows. Due to its particular model, the use of canonical contingency table analysis is of high interest in ecology, but only with short gradients where monotonicity can be assumed in the relationship of row (column) entities. If the gradient is long, strong nonlinearity takes over, the interpretation of the axes becomes impossible, and one may consider trying other methods of ordination. Ter Braak (1985) suggests a second or higher order polynomial regression of correspondence analysis factors, in order to clarify non-linear trends, a method that overcomes Hill

(1979) in consistency: variations on the Kruskal (1964a, 1964b) method, particularly with external (fixed) configuration defined by Clymo's (1980) or Jancey's (1989) distance, in a manner as Kenkel and Orlóci (1986) and Bradfield and Kenkel (1987) have used resemblance functions are appealing. Classifications based on ordination coordinates, limited to the axes that depict non-random variation, can be expected in practice to give sharper results. The choice of Euclidean distance and minimum variance clustering are consistent with the geometrical model of the Eigentechniques which use products. Canonical contingency table analysis can be used to measure the closeness of classifications to check, for example, the consistency of the arrangement of sampling units in environmental groups and vegetation groups. A close correspondence indicate ecological reliability of the vegetation classification.

Procedure B) is being incorporated into a new version of SWILDI. Compared to MULVA-4 latest release on PC/MS-DOS, only four steps needed special attention to improve the package effectiveness: a) the computation of first and second derivatives of cluster fusion levels (to be included in program CLTR); b) the computation of Euclidean distances based on ordination coordinates (from canonical contingency table analysis), which need complicated file manipulations in MULVA-4 (via EDCO, EDDA, a text editor, CREO and INIT), but with a revised RESE the task may become easy; c) the rearrangement of dendrograms according to coordinate scores that may be incorporated in program CLTR, or in some other program, having stored somewhere the original dendrogram; d) improved printout to turn upside-down some dendrograms as well as some scatter diagrams to be incorporated in programs EDGR and ORDB without difficulties. With these changes the procedure becomes more or less automated, leaving the user only the tasks of selecting among the best dendogram cut-off points, identifying the best cross-classifications, and verifying the number of likely gradients. Indeed, the final choice about the gradients and cross-classifications can only be based on experience in community studies.

Acknowledgements. I am indebted to Professor L. Orlóci for his help to revise and improve my initial manuscript, which I prepared in his laboratory in 1989. I thank the editors of CENSUS for their patience and Dr. O. Wildi for the source code of MULVA-4 which he graciously made available to me.

Bibliography

Anderberg, M. R. 1973. Cluster Analysis for Applications. Academic Press, New York.

Barr, A. and E. A. Feigenbaum. 1981-82. The Handbook of Artificial Intelligence. 3 volumes. William Kaufmann (I), Pitman (II-III).

Benzécri, J. P. 1973. L'Analyse des Données. Tome II. L'Analyse des Correspondances. Dunod, Paris.

Benzécri, J. P. 1979. Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Les Cahiers de l'Analyse des Données 4: 377-378.

Bourbaki, N. 1962. Eléments de Mathématiques. Livre II, Algèbre, chap. 2, Algèbre linéaire. Hermann, Paris, ASI 1236.

Bradfield, G., E. and N. C. Kenkel. 1987. Nonlinear ordination using flexible shortest path adjustment of ecological distance. Ecology 68: 750-753.

Camiz, S. 1988. Expert systems: utility in Community studies and examples. Coenoses 3: 33-39.

Camiz, S. 1990. La procedura STATIS per il confronto fra diverse matrici di similarità: un esempio ecologico. Società Italiana di Statistica, Atti della XXXV Riunione Scientifica, CEDAM, Padova, Vol 2: 461-468.

Camiz, S. 1991. Reflections on spaces and relationships in ecological data analysis: effects, problems, possible solutions. Coenoses 6: 3-13.

Clymo, R. S. 1980. Preliminary survey of peat-bog Hummel Knowe Moss using various numerical methods. Vegetatio 42: 129-148.

Dale, M. 1988. Some fuzzy approaches to phytosociology: ideas and instances. Folia Geobotanica et Phytotaxonomia 23: 239-274.

De Antoni, F., N. Lauro and A. Rizzi (eds.). 1986. COMPSTAT 1986: Proceedings in computational statistics. Physica-Verlag, Vienna.

Edwards, A. W. F. and L. L. Cavalli-Sforza. 1965. A method for cluster analysis. Biomtrics 21: 362-375.

Edgington, E. S. 1987. Randomization Tests. 2nd ed. Marcel Dekker, New York.

Feoli, E. and L. Orlóci. 1979. Analysis of concentration and detection of underlying factors in structured tables. Vegetatio 40: 49-54.

Feoli, E. and L. Orlóci. 1991. The properties and interpretation of observations in vegetation study. In: E. Feoli and L. Orlóci (eds.). Computer Assisted Vegetation Analysis, pp. 3-13, Handbook of Vegetation Science, Vol. 11. Kluwer, The Hague.

Feoli, E. and V. Zuccarello. 1986. Ordination based on classification: yet another solution? Abstracta Botanica 19: 203-219.

Feoli, E. and V. Zuccarello. 1988. Syntaxonomy: a source for useful fuzzy sets for environmental analysis? Coenoses 3: 141-147.

Fewster, P. H. and L. Orlóci. 1983. On choosing a resemblance for non-linear predictive ordination. Vegetatio 54: 27-35.

Gale, W. A. 1987. Student: a tool for constructing consultation systems in data analysis. Bull. Int. Stat. Inst., Proceedings of the 46th session, Tokio, pp. 305- 321.

Gale, W. A. and D. Pregibon. 1982. An expert system for regression analysis. In: Hemer, Sacher and Wilkinson (eds.). Proceedings of the 14th symposium on the Interface, pp. 110-117. Springer Verlag, New York.

Godement, R. 1966. Cours d'algèbre. Hermann, Paris.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53: 325-338.

Guttman, L. 1953. A note on Sir Cyril Burt's factorial analysis of qualitative data. Br. J. of Stat.Psychol. 6: 1-4.

Hand, D. J. 1984. Statistical Expert Systems: Design. The statistician 33: 351-369.

Hayes-Roth, F., D. A. Watermann and D. B. Lenat (eds.). 1983. Building Expert Systems. Addison-Wesley, New York.

Hill, M. O. 1973. Reciprocal averaging: an eigenvector method of ordination. J. of Ecol. 61: 237-249.

Hill, M. O. 1974. Correspondence analysis: a neglected multivariate method. Appl. Statis. 23: 340-354.

Hill, M. O. 1979. DECORANA - a FORTRAN program for detrended correspondence analysis and reciprocal averaging. Ecology and Systematics, Cornell University, Ithaca, New York.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. J. Ed. Psych. 24: 417-441, 498-520.

Jancey, R. C. 1989. The measurement of non-linear association in community descriptor sets. Coenoses 4: 11- 13.

Kenkel, N. C. and L. Orlóci. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology 67: 919-928.

Kenkel, N. C. and T. Booth. 1987. A Comparison of presence-absence resemblance coefficients for use in biogeographical studies. Coenoses 2: 25-30.

Kruskal, J. B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometric 29: 1-27.

Kruskal, J. B. 1964b. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29: 115-129.

Lagonegro, M. and E. Feoli. 1979. Intersection analysis in phytosociology: computer program and application. Vegetatio 40: 55-59.

Lagonegro, M. and E. Feoli. 1984. Three Packages for Classification and Ordination of Multivariate Data. Libreria Goliardica, Trieste.

Lebart, L., A. Morineau, T. Lambert and P. Pleuvret. 1991. SPAD. N. Version 2.0 - Système Portable pour l'Analyse des Données. Manuel de référence. CISIA, Paris.

Lebart, L., A. Morineau and N. Tabard. 1977. Techniques de la Description Statistique. Dunod. Paris.

Legendre, L. and P. Legendre. 1983. Numerical Ecology. Elsevier, Amsterdam.

Mahalanobis, P. C. 1936. On the generalized distance. Nat. Inst. Sci. India Proc. 2: 49-55.

Marsili-Libelli, S. 1989. Fuzzy clustering of ecological data. Coenoses 4: 95-106.

Mojena, R. and D.Wishart. 1980. Stopping rule for Ward's clustering method. Compstat '80 - Proceedings in Computational Statistics, pp. 426-432. Physica Verlag, Wien.

Mueller-Dombois, D. and E.Ellenberg. 1974. Aims and Methods of Vegetation Ecology. Wiley, New York.

Noy-Meir, I. 1971. Multivariate analysis of the semi-arid vegetation in south-eastern Australia: nodal ordination by component analysis. Proc. Ecol. Soc. Aust. 6: 159-193.

Noy-Meir, I. 1974. Catenation: quantitative methods for the definition of coenoclines. Vegetatio 29: 89-99.

Ochiai, A. 1957. Zoogeographic studies on the solenoid fishes found in Japan and its neighboring regions. Bull. Jap. Soc. Sci. Fish. 22: 526-530.

Orlóci, L. 1967. An agglomerative method for classification of plant communities. J. of Ecol. 55: 193-206.

Orlóci, L. 1974. On information flow in ordination. Vegetatio 29: 11-16.

Orlóci, L. 1978. Multivariate Analysis in Vegetation Research. Junk. The Hague.

Orlóci, L. 1979. Non-linear data structures and their description. In: Orlóci, C. R. Rao and W. M. Stiteler (eds.). Multivariate Methods in Ecological work. pp. 191- 202. International Cooperative Publishing House, Fairland, Maryland, U. S. A.

Orlóci, L. 1988. Detecting vegetation patterns. ISI Atlas of Science: Plant and Animals 1: 173-177.

Orlóci, L. 1991a. CONAPACK: Program for Canonical Analysis of Classification Tables. Ecological Computations Series (ECS): Volume 4. SPB Academic Publishing bv, The Hague.

Orlóci, L. 1991b. On character-based community analysis: choice, arrangement, comparison. Coenoses 6: 103-107.

Orlóci, L. 1991c. Poorean approximation and Fisherian inference in bioenvironmental analysis. Research Trends, Advances in-cology 1: 65-71.

Orlóci, L. 1991d. On character-based community analysis: choice, arrangement, comparison. Coenoses 6: 103-107.

Orlóci, L. and N. Kenkel. 1987. Data Analysis in Population and Community Ecology. Department of Plant Sciences. The University of Western Ontario.

Orlóci, L. and M. Orlóci. 1985. Comparison of communities without the use of species: model and example. Annali di Botanica (Roma) 43: 275-285.

Orlóci, L. and V. De Patta Pillar. 1989. On sample size optimality in ecosystem survey. Biom. Praxim. 29: 173- 184.

Pignatti, S., S. Camiz, P. Squartini and V. Squartini 1989. Chorological and ecological information as basis for the syntaxonomy of beech forests in Italy. Ber. d. Reinhold-Tüxen-Ges.,Göttingen 1: 73-82.

Pillar, V. De Patta and L. Orlóci. 1991. Fuzzy components in community level comparisons. In: E. Feoli and L. Orlóci (eds.). Computer Assisted Vegetation Analysis, Handbook of Vegetation Science, Vol. 11. pp. 87-93. Kluwer, London.

Podani, J. and E. Feoli. 1991. A general strategy for simultaneous classification of variables and observations in ecological data tables. In: L. Feoli and L. Orlóci (eds.), Computer Assisted Vegetation Analysis, Handbook of Vegetation Science, Vol. 11, pp. 197-209. Kluwer, Dordrecht.

Roberts, D. W. 1986. Ordination on the basis of fuzzy sets theory. Vegetatio 66: 123-131.

Sandland, R. L. and P. C. Young. 1979. Probabilistic tests and stopping rules associated with hierarchical classification techniques. Aust. J. Ecol. 4: 399-406.

Siegel, S. s. 1956. Non Parametric Statistics for the Behavioral Sciences. McGraw-Hill, New York.

ter Braak, C. J. F. 1985. Canonical Correspondence Analysis: a new Eigenvector Technique for Multivariate Direct Gradient Analysis. IWIS-TNO Wageningen, Report A855T17.

Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. J. Amer, Stat. Ass. 58: 236-244.

Watermann, D. A. 1986. A Guide to Expert Systems. Addison-Wesley. New York.

Westhoff, V., and E. van der Maarel. 19733. The Braun-Blanquet approach. In: R. H. Whittaker (ed.). Ordination and Classification of Vegetation, Handbook of Vegetation Science, Vol. V. Junk, The Hague. pp. 617-725.

Wildi, O. 1988. Chaining multivariate techniques. New approaches to the classification of Vegetation Data. Paper presented at the IAVS Symposium on Progress in Theoretical Vegetation Science, Vienna.

Wildi, O. 1989. A new numerical solution to traditional phytosociological tabular classification. Vegetatio 81: 95-106.

Wildi, O. and L. Orlóci. 1983. Management and Multivariate Analysis of Vegetation Data. Report 215, Eidgenöss. Anst. für das Fortstlische Versuchswesen, Birmensdorf (CH).

Wildi, O. and L. Orlóci. 1987. Flexible gradient analysis: a note on ideas and application. Coenoses 2: 61-65.

Wildi, O. and L. Orlóci. 1990. Numerical Exploration of Community Patterns. SPB Acad. Pub., The Hague.

Wishart, D. 1978. CLUSTAN User Manual. 3rd edition. Inter-Univ. Res. Comm. Ser. Edinburgh University. Report 47.

Manuscript received: May 1992.