SOOSIANA, 18:11-16, 1990.

MOLNÁR, A., SÜMEGI, P.:

Classification and ordination methods in the division of the Pleistocene malacological zones of Debrecen I. profile - Klasszifikációs és ordinációs módszerek pleisztocén malakológiai zónák lehatárolására, Debrecen I. szelvény

ABSTRACT: The use of multivariate statistical methods in the interpretation of mollusca data of stratigraphical profiles can promote the objective division of palaeoecological zones. This paper presents two method - Cluster analysis and Principal Component Analysis (PCA) - for the zone division of stratigraphical profiles.

INTRODUCTION

The most important purpose of stratigraphical studies is the division of the different zones in stratigraphical sequences, the estimation of their age and the determination of their evolution ordes on the basis of their taxonomical composition and species dominance.

The above problem can be solved by the study of mollusca matter of the stratigraphical sequences with statistical methods, which offers an objective determination of zones (groups).

The application of statistical methods is obvious in this case since the data from stratigraphical profiles are complex, multivariate and quantitative. From the abundance data and composition of species in samples from different deposit, the division of stratigraphical sequences into zones, the changes of species between the zones and within the entire profile can be described. The changes in climate and vegetation on the bases of prferential parameters characteristic of species and communites as well as the interactions between populations can also be determined using the appropriate analysis. In this way the evolution model for the areas in present time can be established by the reconstruction of the past conditions.

As a first step, a mollusca diagram should be set up by abundance and frequency data of species obtained from the samples of the stratigraphical profiles. These units (samples) provide only the scale for the data processing; the evaluation of the samples requires the knowledge of the palaeoecological zones. This can be performed by the classification of the sample units and the determination of their similarity degree on the basis of the composition of the species and number of the individuals.

CLUSTER ANALYSIS

Cluster analysis is a widespread numerical method for classification, numerous method of which are well known and widely used. The most fundamental role of this method is not in the classification of the sample units into groups, but the procedure itself shows the group structure of the profile, on the basis of the parameters of the sample (variables; i.e. the mollusca species in this case). The cluster analysis is not a single procedure but series of procedures in which after the transformation of data and determination of similarity or dissimilarity; in a final stage of which a dendrogram can be constructed using a certain linking procedure (EVERITT 1974, BIRKS 1986). The transformation of data simple in this case since our data are given in percentage thus they not need to be undertaken any standardization procedure (no differences in their magnitude, and unit of measurement).

In the next step the similarity or dissimilarity of the samples are studied. Our experiences suggest that the modified the modified Sorensen index (SOUTHWOOD 1978) is of high applicability for malacological purposes:

$$C_N = 2_{jN}/(aN + bN) \qquad (1)$$

where a_N, b_N are the number of individuals in samples "a" and "b", respectively, j_N is the sum of the smaller abundancies of the common species in both samples. This index refers to the similarity of the composition of species in the two habitats (-diversity). The next step is the representation of cluster groups of stratigraphically adjacent samples (similar mollusca composition) in terms of the single link criterion.

The well distinguishable groups on the dendrogram (Fig.1.) well agree with the mollusca matter. The matter of the sample 16, 15, 14, 13, 11, 10, 8, 9 settled under the same palaeoecological circumstances (cold, moderately moist). The ecostatistical zones (B_1 and B_2) are separated because of the mild climatic period of the sample 12 where the cryophillic species disappeared and the dominance of the mesophillic species raised (SÜMEGI 1989). The fauna of the samples in group C refers to a mild, moist climate. The differentation of this group from the zones B_1 and B_2 is very characteristic. The other samples contain species groups referring to transitions of the different ecological circumstances.

The advantage of this method is the relatively simple calculation procedure, though it does not take into account species which are low in quantity but significant as a factor affecting the zone formation (BIRKS 1986). This problem can be solved by the following ordination procedure.

PRINCIPAL COMPONENT ANALYSIS

The methods of ordination represent the sample units as points in a two ore more dimensional space (according to number of variables) so that the distances between the points express the correlation between units. Thus the ordination is simply a multidimensional scaling.

The most frequently encountered ordination methods is the Principal Component Analysis (PCA) (BIRKS 1986), which is also suitable for the division of stratigraphical sequences. The multidimensional space representation mentioned above can be derived by taking this method as a geometrical model (SEAL 1964). The axes of the multidimensional coordinate system represents the variables of the sample and points are the observation units. PCA projects the sample points perpendicularly onto a plane chosen so that the sum of squared distances from the points to the plane is minimal. The first principal component is the best fitting line along points; the second is at right angles to the first, and together with the first defines the plane of best fit etc. Axes corresponding to the given variables are the principal scores.

In the evaluation of data the variance-covariance (S) or the correlation (R) matrix of the sample is calculated from the raw-data matrix. After the calculation of eigenvalues (characteristic roots), and their cumulative $_{j\%}$ values; eigenvectors (u_{ij} principal component loadings, which means the contribution of each species in a given sample. C_j principal component variables can be obtained using the principle component coefficients (3).

$$C_j = u_{ij}X_1 + u_{ij}X_2 + \dots + u_{ij}X_i + \dots + u_{pj}X_p = \sum_{i=1}^{p} u_{ij}X_i \quad (3)$$

where C_j is the j-th principal component variable, u_{ij} coefficients are the elements of u_j eigenvector pertaining to the jth eigenvalue (in the case of X_i element). C_I-C_{II}, C_I-C_{III}, etc. variables can be represented in a rectangular coordinate system.

As Fig.2. shows the groupping results of the two method are very similar however the plot obtained by PCA is more characteristic. Results in Fig.2a, extracted from the variance-covariance matrix gives a better approach because of the higher contribution of the firs two principal component coordinates (89.63% vs 58.78%).

CONCLUSIONS

The use of multivariate statistical methods in the interpretation of mollusca data of stratigraphical profiles cab promote the objective division of palaeoecological zones.

The application of cluster analysis a solution for the classification of zones, while PCA is a suitable method for the elucidation of the inherent relations of the entire profile which makes possible the division of the different palaeoecological groups referring to different factors.

The procedures presented here constitutes only a small part of the classification and ordination methods. The improvement of other methods for malacological data is also in progress, results are planned to be published in case of their applicability.

ÖSSZEFOGLALÁS

A klasszifikáció és ordináció módszerei a sztratigráfiai szelvények paleoökológiai lehatárolására kinálnak objektiv lehetősé et. Cikkünk két olyan módszert mutat be, amelyet a quartermalakológia területén még nem alkalmaztak. A Cluster Analizis által kapott csoportositást a Főkomponens Analizis eredményeivel vetettük össze. Az eredmények jó egyezést mutattak. Ez alapján e két módszert (amelyek az általunk elkészitett "MŠ" programcsomag részei) alkalmasnak találjuk malakostatisztikai elemzésekre.

REFERENCES

BIRKS, H.J.B. (1986): Numerical zonation, comparison and correlation of Quarternary pollen-stratigraphical data. In: Handbook of Holocene Palaeoecology and Palaeohydrology (Ed: B.E. BERGLUND), -J. Wiley, Chichester, p: 743-774. - EVERITT, B. (1974): Cluster analysis Heinemann, London, pp.: 1-122. - PRENTICE, I.C. (1986): Multivariate methods for data analysis. In: Handbook of Holocene Palaeoecology and Palaeohydrology (Ed: B.K. BERGLUND), J. Wiley, Chichester, p: 775-797. - SEAL, H.L. (1964): Multivariate Statistical Analysis for Biologists, Methuen and Co., London, p: 101-122. - SOUTHWOOD, T.R.E. (1978): Ecological Methods with particular reference to the study of insect populations (2nd ed.), Chapmann and Hall, London. - SÜMEGI, P. (1989): The Upper-Pleistocene evolution history of the Hajduság region, on the basis of stratigraphical, palaeontological, sedimentological, and geochemical investigations. B.S. dissertation-KLTE University, Debrecen.

MOLNÁR ANDOR DR. SÜMEGI PÁL

Kossuth Lajos Tudományegyetem
Ásvány-és Földtani Tanszék

Debrecen
Egyetem tér 1.
H-4010 Hungary

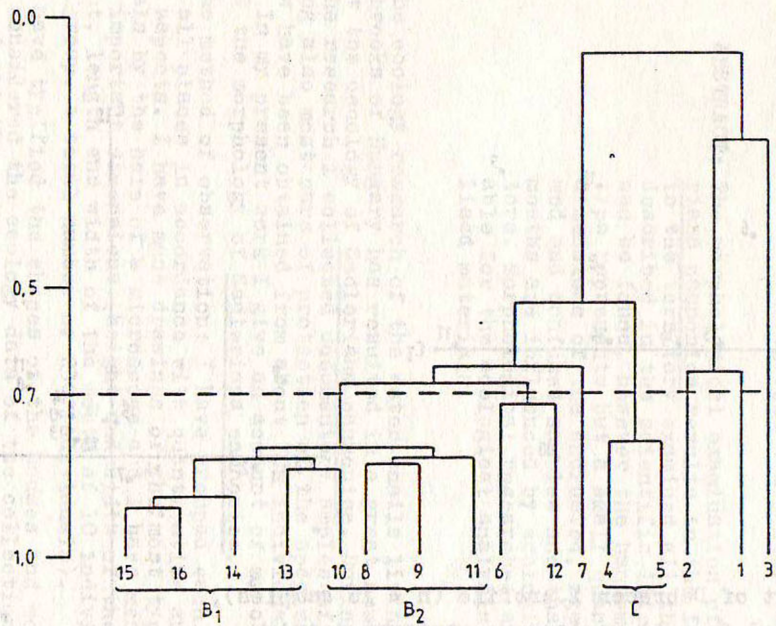
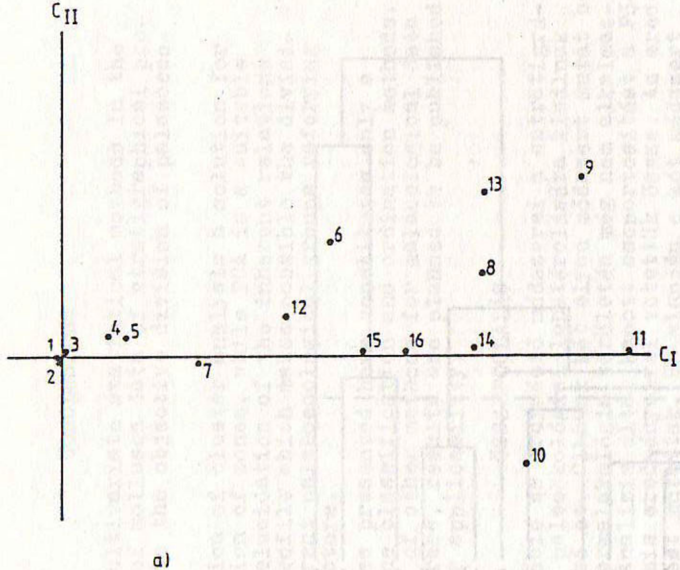


Fig.1. Dendrogram of Debrecen I.profile (n = 16 samples) using the modified Sorensen index for the measure of similarity
1.ábra. A Debrecen I.szelvény dendrogramja (n = 16 minta), szimilaritás értékek módositott Sorensen index-szel számitva



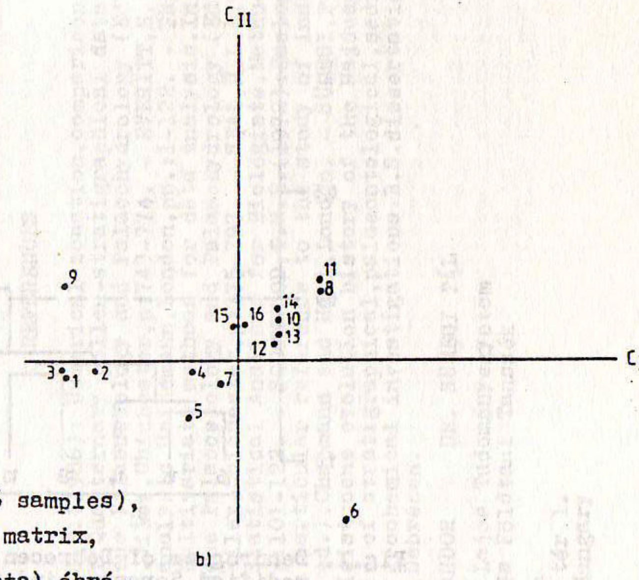


Fig.2. C_I - C_{II} plot of Debrecen I.profile (n = 16 samples), data extracted from a) variance-covariance matrix, b) correlation matrix

2.ábra.A Debrecen I.szelvény mintáinak (n = 16 minta) ábrázolása a C_I - C_{II} főkomponensváltozókkal, a)variancia-kovariancia, b) korrelációs mátrix adataiból