COENOSES 5 (2): 69-77, 1990

CLASSIFICATION OF SEQUENCES IN COMMUNITY ECOLOGY - THREE CASE STUDIES

J. Lepš and M. Straškraba, Department of Biomathematics, Biological Research Centre, Czechoslovak Academy of Sciences, Branišovská 31, 370 05 České Budějovice, Czechoslovakia
Š. Husák, Department of Hydrobotany, Institute of Botany, Czechoslovak Academy of Sciences, Dukelská 145, Třeboň, Czechoslovakia
and
J. Boháč, Institute of Landscape Ecology, Czechoslovak Academy of Sciences, Na sádkách 7, 370 05 České Budějovice, Czechoslovaka

Keywords: Sequences, Classification, Numerical, Plant communities, Phytoplankton, Soil macro-fauna

Abstract. Classification of sequences in space or time was used to search for repeatable patterns in zonation on an environmental gradient and in seasonal dynamics. Three data sets were used which capture the zonation of plant communities on an emerged fishpond bottom along a moisture gradient, seasonal dynamics of phytoplankton communities in different years, and seasonal dynamics of soil macrofauna communities in different sites. A two step procedure was used: all sampling units were classified and assigned to particular states in the first step; sequences of states were classified in the second step. The procedure permits many methodical choices (type of primary classification, similarity between sequences, etc.) and examination of their influence on the resulting classification. The procedure was successful in revealing repeatable patterns in ecological data sets.

Introduction

In many cases, the ecological reality is characterized by a sequence of connected events, which cannot be represented by a single observation. Sequences of observations in time or in space are needed. Classification of sequences was apparently first used in community ecology by Williams et al. (1969) for classification of successional series in tropical forest. Despite the apparent usefulness of this approach, it has been seldom applied in subsequent community studies. Communities have to be characterized by a sequence of observations to reveal the seasonal dynamics or the successional sequence. Repeatable sequences have to be sought on environmental gradients to show zonation on a mountain side or on a shore. In each case, the classification of sequences will help to search for distributional and temporal patterns. For classification of sequences, a two step procedure is appealing. In the first step, all the particular sampling units from all of the sequences are classified and assigned to particular states. Sequences of states defined by the first classification are classified in the second step. In the present study, we applied this method to three data sets. One of the sets describes spatial sequences, and the other two temporal sequences. The aim of the study is to reveal repeatable patterns in the data sets and to test the influence of particular methodical choices on the resulting classification.

Materials

Plant communities on the emerged fishpond bottom

The data set is from the Lednice fishponds in south Moravia, Czechoslovakia; one transect from the Svět fishpond in Třeboň basin in south Bohemia is added for comparison. In each transect three to twenty equidistantly located 1m^2 sample plots were analysed. A few very short transects were included to demonstrate the influence of unequal sequence length on the resulting classification. Species were characterized by their dry weight (biomass) and by density. The results of the same procedure based on biomass and densities are compared. Exceptionally (in very dense growth), smaller plots were used and values were expressed on a 1 m^2 plot basis. Transects are numbered consecutively 1 to 26. Transects 1 to 25 were taken from the Lednice region, transect 26 from the Třeboň site. Transects 1 and 2 were laid in the littoral synantropic vegetation on emerged bottom and transects 6 to 9 in synantropic vegetation on deposits of fishpond bottom material after fishpond cleaning. Transects 3 to 5, 10 to 12, 17 to 22 and 24 to 26 represent typical vegetation of emerged fishpond bottom. Transects 13 to 16 form a group of marginal transects with halophytic vegetation. Transect 23 contains only submerged vegetation. The aim of the analysis was to compare particular transects with respect to their zonation pattern on the moisture gradient.

Phytoplankton in the reservoir

The second data set characterizes the seasonal dynamics of phytoplankton communities in Slapy Reservoir (Vltava River, central Bohemia). Slapy Reservoir is characterized by a riverine shape, approximately 42 km long with maximum depth of 55 m. The total volume is $270.10^6 m^3$ with a retention time of 18-40 days (mean river flow is $84 m^3 sec^{-1}$).

The aim of the Slapy Reservoir Study (performed by the Hydrobiological Laboratory, Czechoslovak Academy of Sciences, Prague) was to identify the relationships between physical, chemical and biological processes in aquatic ecosystems in general, and in man-made lakes in particular. Slapy Reservoir is a appropriate for such studies because years with different hydrometeorological conditions result in very different stratification patterns in the water column. Wet years (1965, 1966) are characterized by riverine, nonstratified water masses. In dry years (1964, 1967) very pronounced temperature differences between the surface and bottom exist, with a very sharp temperature decline at a depth of 2 to 3 m. Phytoplankton samples were collected from one point near the dam at 3-week intervals during 1964-1967. A tube sampler provided an integrated column from the epilimnion. In the last year, some autumn samples are missing and therefore the last year is not comparable unequivocally. More detailed information on phytoplankton, chemistry and other variables including sampling methods is presented in Hrbáček and Straškraba (1973). Phytoplankton data treated in the present paper were obtained by Javornický and Komárková (1973). The biomass of particular species estimated from microscope counting and sizing was used to provide an 'importance value'. The aim of classification was to compare temporal variation in phytoplankton community dynamcis.

Soil macrofauna

The soil macrofauna data come from a study in the 'Centralnochernozemnoy zapovednik' (The Chernozem Biosphere Nature Reserve, near the village of Neremushki, Kurskaya oblast district, USSR). Detailed information on the study is presented in Bogach [= Boháč] et al. (1984). Soil macrofauna of six natural, seminatural and man-made habitats was compared: unmowed steppe, mowed steppe, oak forest, pasture, wheat field and experimental field, ploughed twice a year, without any crop. Samples were taken from 625 cm^2 quadrats to a depth of 30 cm. Material was subsequently sorted by hand in the laboratory. Originally, three layers (10 cm each) were distinguished; for the purpose of the present study, samples from particular layers were pooled. In 1982, eight samples were taken from each plot, approximately in monthly intervals (March-October). Altogether, 55 macrofaunal species were found - the most common groups were staphylinid beetles (Coleoptera, Staphylinidae) and millipedes (Diplopoda, Chilo-

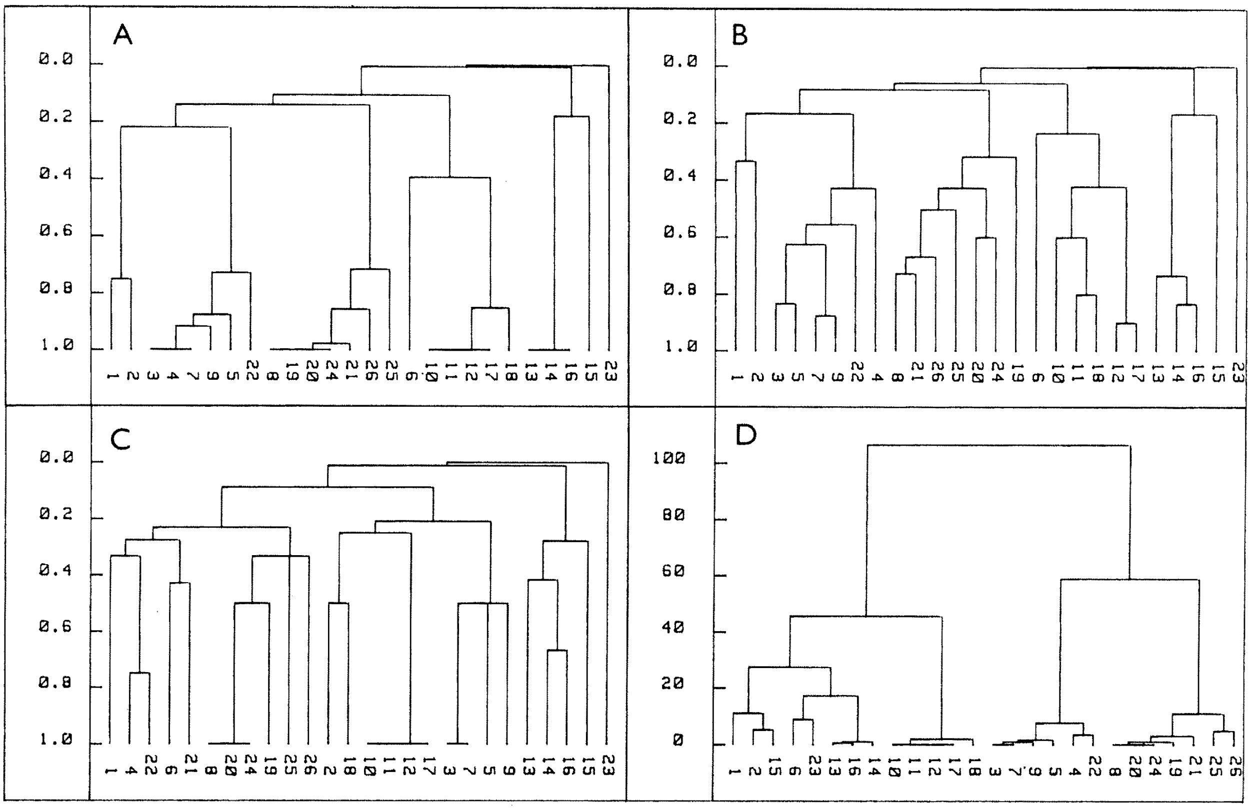

Fig. 1. Classification of fishpond bottom transects based on 11-group primary classification of biomass data. Similarity: A - LCS/min (length); B - LCS/max (length); C - LCS/max (length) after string collapsing; D - information gain.

poda). Counts of individuals were taken as an 'importance value' of each species. The classification was performed for all macrofaunal species and for separate taxonomical groups - staphylinid beetles and millipedes.

Methods

Primary classification

We used the two-step procedure for classifying communities on a gradient, (cf. Williams et al. 1969, Dale et al. 1970, Dale et al. 1988). All the sampling units were classified by a numerical technique in the first step and assigned to a state. Sequences of states were classified in the second step. The methods for the primary classification were chosen according to the nature of primary data. In the fishpond bottom vegetation data, with a larger number of primary sampling units (216), the divisive hierarchical method (TWINSPAN, Hill 1979) was used. The standardized Euclidean distance (chord distance, Orlóci 1978) was used as a dissimilarity; in the phytoplankton study, the non-standardized Euclidean distance was also applied as an alternative. Ward's method (minimum sum of squares, Orlóci 1978) was used for hierarchical clustering. In all cases, the appropriate cut-off level in the hierarchy was chosen subjectively, according to the interpretability of resulting groups. In the fishpond bottom data set, two different levels were applied.

Classification of sequences

Some methodological decisions had to be made during the classification of sequences:

1) Should the sequences be considered 'with repetition'? In some cases, only the order of particular zones is of interest, not the length of these zones (a zone is a part of a string composed of samples belonging to the same state). In such a case, both the sequences AAABBCAA and ABBCCCCA may be collapsed, expressed as ABCA and considered as identical. This approach was applied, for example, by Little and Ross (1985) when classifying soil samples. We shall call this string collapsing. Both possibilities were used.

2) How to measure the similarity of two sequences? Two measures were used. The first is called information gain and is based on transition matrices (Dale et al. 1970). The second is the longest common subsequence (LCS) of two strings (Dale, pers. comm., Coggins 1983), and is a special case of Levenshtein metrics (Dale et al. 1988). Computing the value of the first measure, the sequences are first converted into transition ma-

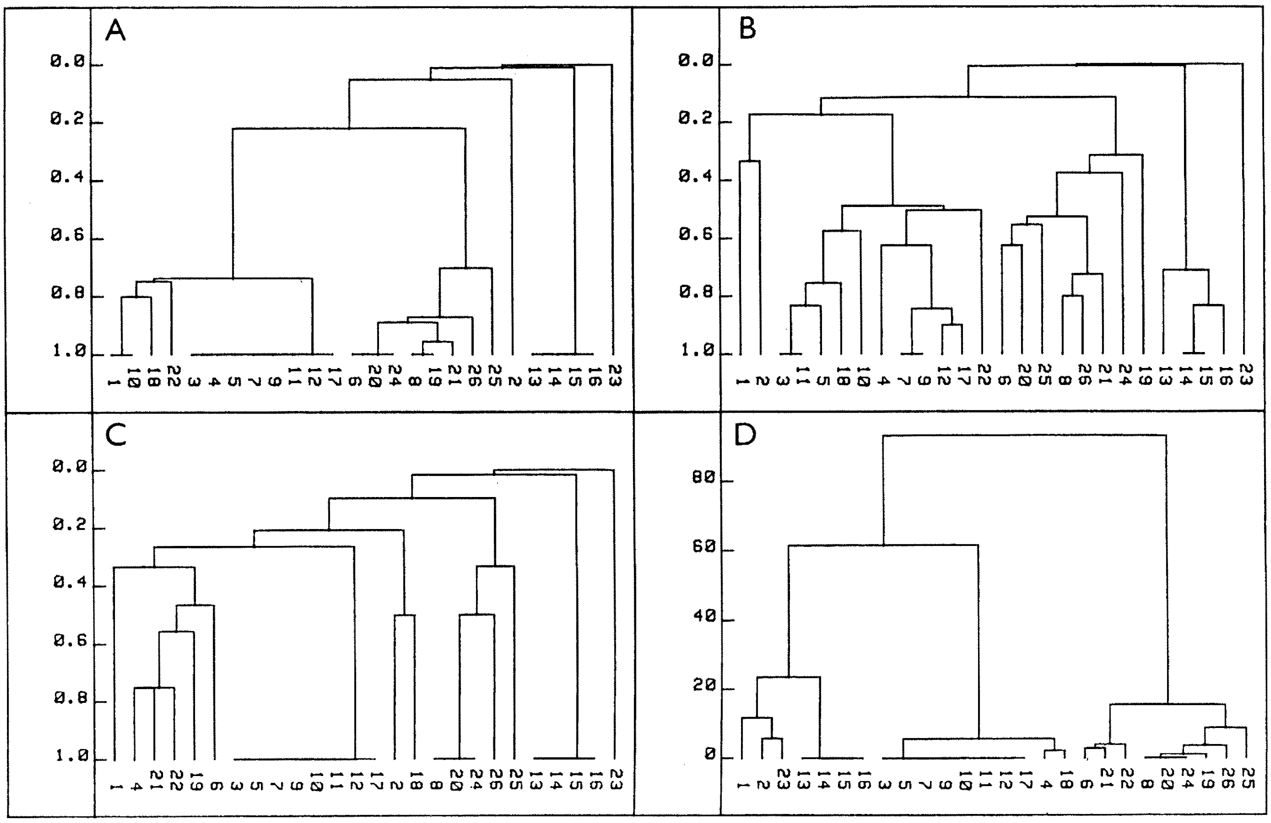

Fig. 2. Classification of fishpond bottom transects based on 8-group primary classification of biomass data. Similarity: A - LCS/min (length); B - LCS/max (length); C - LCS/max (length) after string collapsing; D - information gain.

trices (each entry in the matrix $x_{i,j}$ is the number of appearance of 'i matched by j' in the sequence). The dissimilarity of two matrices A and B is expressed as the information gain

$$DI_{A,B} = I_{A+B} - I_A - I_B,$$

where $I = g \ln g - \Sigma_i \Sigma_j x_{i,j} \ln x_{i,j}$ is a measure of information content in a matrix and $g = \Sigma_i \Sigma_j x_{i,j}$ ($x_{i,j} \ln x_{i,j}$ is considered 0 for $x_{i,j} = 0$) I_{A+B}, I_A and I_B correspond to information content in matrices A+B, A and B respectively. As the similarity of A+B with C is defined in the same way, i.e.

$$DI_{A+B,C} = I_{A+B+C} - I_{A+B} - I_C,$$

hierarchical clustering is directly defined.

The longest common subsequence (LCS) is the longest sequence contained in the compared sequences. For example

Sequence 1:	AABBCCCCDD
Sequence 2:	ABBBBCDDDA
LCS:	ABBCDD

As a similarity measure, LCS should be standardized to the length of the sequences under comparison. In the case of sequences of unequal length, there are (at least) three possibilities: standardization to the length of the longer sequence, i.e., max (length), the shorter one - min (length), or to the mean. We used both standardization to the minimum and to the maximum length. In this way, we obtained a matrix of similarities between sequences. The standard average linkage hierarchical procedure was then applied.

Results

Plant communities on emerged fishpond bottom

Only the classification based on biomass data will be treated in detail. The results based on density are briefly presented for comparison. The use of TWINSPAN yielded 11 ecologically interpretable groups. As both the minimal group size and hierarchical level of division were used as a stopping rule, the number of groups need not be an integer power of 2. The groups may be characterized as follows:

A and B) halophytic vegetation typically with *Juncus gerardi* and *Plantago maritima*; higher representation of *Tripolium pannonicum* and *Spergularia maritima* in the group A;

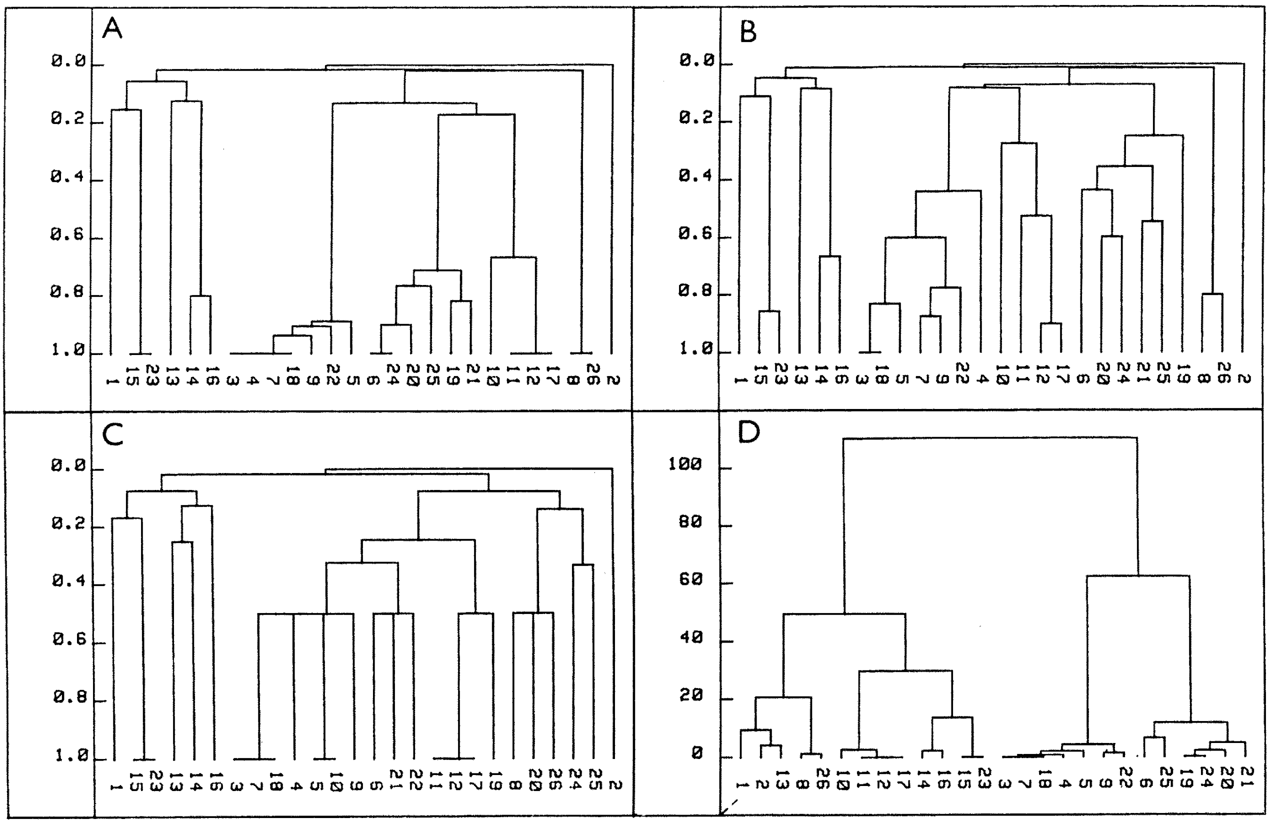

Fig. 3. Classification of fishpond bottom transects based on 13-group primary classification of density data. Similarity: A - LCS/min (length); B - LCS/max (length); C - LCS/max (length) after string collapsing; D - information gain.

C) dominated by *Carex riparia;*

D) dominated by *Carex otrubae;*

E and F) large groups typically with *Atriplex prostrata, Myosoton aquaticum*; in E also *Persicaria lapathifolia* and *Chenopodium ficifolium*, in F *Ranunculus sceleratus, Oenanthe aquatica* and *Phragmites communis* (often as seedlings);

G and H) characterized by *Typha angustifolia* (often as seedlings), *Juncus articulatus, Rumex maritimus,* in G also *Alopecurus aequalis;*

I) characterized by *Calystegia sepium, Mentha aquatica, Malva pusilla* and *Lycopus europeus;*

J) submerged vegetation dominated by *Myriophyllum spicatum;*

K) stands of *Chara coronata* on emerged bottom.

For comparison, the 8 group level also was applied; in this, group A is pooled with B, E with F and G with H. As an alternative, primary classification based on individuals yielded 13 groups.

The resulting sequences of states on particular transects are presented in Table 1. Four classifications were applied; three based on LCS (similarity measures: LCS/max (length), LCS/min (length), LCS/max (length) after string collapsing, Fig. 1a, b, c) and the fourth on information gain (Fig. 1d). All the LCS based classifications clearly separate transect 23 (submerged vegetation) and then the group fo transects 13, 14, 16 and 15 with halophytic vegetation; in classifications without collapsing strings also the synantropic vegetation on emerged bottom (transects 1 and 2) was distinguished. The transition matrix classification distinguished only partly the group of transects with halophytic vegetation. In this case, the LCS based classification appeared to provide the most, and transition matrix the least useful information; the last one is most sensitive to differences in length of sequences. String collapsing partially obscures the groups. The position of extremely short transects differs considerably among particular classifications. The classification based on the 8 groups (Fig. 2) is very similar to the previous one. The major groups (e.g., the halophytic vegetation) are better differentiated, but the structure within groups is less clear. The results of classification based on the density data (Fig. 3) are difficult to interpret. Density appears to reflect ecological conditions other than does biomass; differences in the size of individuals are not considered.

In all classifications, the transect from the Třeboň basin is classified in a group of transects from the Lednice fishponds with similar vegetation from the Lednice fishponds. There is some level of similarity in vegetation of fishpond bottoms in the two regions.

The plankton community data

Results of classifications of particular samples based on standardized data and non-standardized data differ considerably. Various primary classifications and use

Table 1. Sequences of states defined by the TWINSPAN classification of fishpond bottom data. Left: based on biomass data. Right, based on densities.

1:	EIGEDEICI	1:	DHIHHHHCA
2:	IIII	2:	BBBB
3:	EEEE	3:	DDDD
4:	EEEEEEEEEEEGEG	4:	DDDDDDDDDDDDDF
5:	EEEEEF	5:	DDDDDE
6:	GFHFHHEH	6:	FEDEGFFF
7:	EEEEEEEE	7:	DDDDDDDD
8:	GGGGGGGG	8:	GGGGGGGG
9:	FEEEEEEE	9:	EDDDDDDD
10:	FFF	10:	DEE
11:	FFFFF	11:	EEEEE
12:	FFFFFFFFFF	12:	EEEEEEEEEE
13:	BBBB	13:	LLLL
14:	ABBBBA	14:	JJLJJI
15:	AAAAAA	15:	IIIII
16:	ABBBB	16:	JJJJJ
17:	FFFFFFFFF	17:	FFFFFFFFF
18:	IFFFF	18:	DDDDD
19:	GGGGGGGGGGGGGGGGGGFF	19:	FFFFFFFFFFFFFFFFFFEE
20:	GGGGG	20:	FFFFG
21:	GGGGGGGGFGF	21:	FFFFFFFFDEC
22:	EEEGGGGEE	22:	DDDFEDDDD
23:	JJJJJJ	23:	IIIIII
24:	GGG	24:	FFF
25:	GGGGGCCCA	25:	FFFFFCCCK
26:	HGGGGGGGKK	26:	GGGGGGGGMM

Table 2. Sequences of states defined by the classification of phytoplankton data based on standardized (left) and non-standardized (right) classification.

1964:	AEAABBBBCDDDDCCEEE	1964:	AAABBBBBBCCEEEAAAAA
1965:	CBBBABCACDDDAEEA	1965:	AABAABCACECCAAAA
1966:	CABBBBCCDDDCAEEE	1966:	AABBDDCCEECACAAA
1967:	EBBBBBCCCDE	1967:	ABBBBDCCCEA

of other multivariate techniques in analysis of this data set is described in Lepš et al. (1990). The standardization results in a pattern which seems to be more concise and more resembles the classical limnological classification. In both case, a five-group cut-off level was accepted. In classification based on standardized data, the groups are:

A) mainly winter (October to January) samples dominated by species of the genera *Closterium* and *Scenedesmus;*

B) late winter to spring (February to May) samples with dominant *Asterionella formosa* and *Cyclotella comta;*

C) mainly summer samples, often with blue-greens, but also some winter samples;

D) a homogeneous group of full summer samples with high bloom of *Aphanizomenon flos-aquae;*

E) winter samples with *Closterium.*

The groups in the classification based on nonstandardized data are:

A) mainly winter samples dominated by *Closterium polymorphum* and *C. limneticum* and large and medium size Chlorophyta;

B) spring to early summer samples with medium biomass represented by *Cryptomonas* (particularly *C. reflexa, C. curvata)* and also *Peridinium inconspicuum;*

C) mainly summer samples with high biomass, dominated by *Cryptomonas reflexa, C. marsonii, Asterionella formosa* and *Chlamydomonas species;*

D) short spring phytoplankton peaks with similar species as group B, but much higher biomass;

E) short summer period of maximum biomass with large species like *Ceratium hirundinella* and the colonies of *Fragilaria crotonensis.*

In some cases, a few summer samples were classified as belonging to groups of winter samples. This is caused by very low total phytoplankton density comparable with winter values produced by rapid throughflows (short retention times) in the reservoir. The years 1964 and 1967 are characterized by low to medium discharges, which is also reflected in highly developed summer water mass stratification. In high-flow years

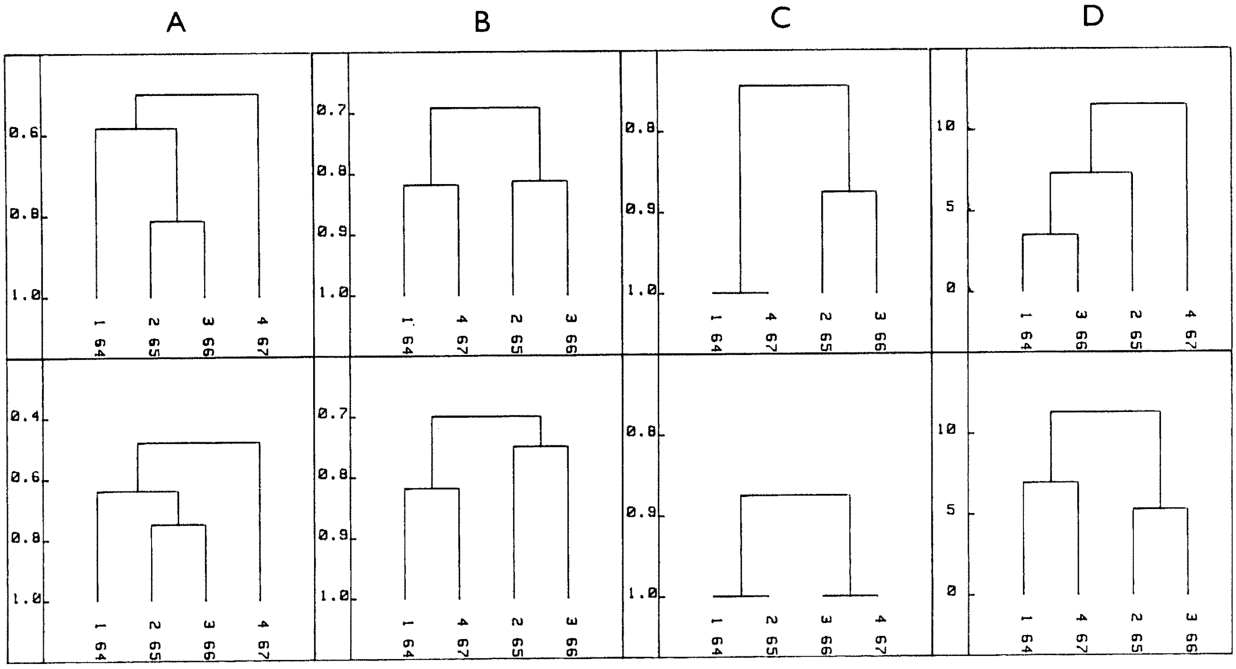

Fig. 4. Classification of the years 1964 to 1967 according to sequences of phytoplankton states. Upper row: primary classification with standardized Euclidean distance. Bottom row: primary classification with non-standardized Euclidean distance. Columns correspond to similarity measures in sequence classification: A - LCS/min (length); B - LCS/max (length); C - LCS/max (length) after string collapsing; D - information gain.

Table 3. Sequences of states defined by the classification of soil macrofauna data. Particular habitats are: 1 US - unmowed steppe; 2 MS - mowed steppe; 3 OF - oak forest; 4 PA - pasture; 5 FI - wheat field; 6 PF - ploughed field (without crop).

all organisms			millipeds			staphylinids		
1	US:	AAAAAABA	1	US:	ABCBBBBB	1	US:	ACAACAAA
2	MS:	AAAAAAAA	2	MS:	CBBBBCCC	2	MS:	AABBBDCC
3	OF:	CCCCCCCB	3	OF:	DDDDDDDD	3	OF:	DCCCCDDC
4	PA:	DCCCCCCC	4	PA:	ABBCBBBB	4	PA:	DDCDBADB
5	FI:	BCDDDBDC	5	FI:	CCAAAAAC	5	FI:	CBAAAAAA
6	PF:	BBDDCCBB	6	PF:	CAAAACCC	6	PF:	ACAACAAA

1965 and 1966 summer stratification was disrupted. The maximum surface-bottom temperature difference in the first group of years is 12-18 °C, whereas in the second group it is only 4-8 °C. The differences in stratification are accompanied by differences in the phytoplankton dynamics. In dry years, the representation of blue-greens is stronger. Sequences of states based on both primary classifications are shown in Table 2. The resulting classifications are in Fig. 4. Most classifications show two pairs of similar years: 1964 and 1967 and 1965 and 1966. The classification based on LCS standardized to the maximum length separated 1967 as different from other years. The reason is its different length. String collapsing resulted in hardly interpretable results. Clearly, the information on the duration of period, typical for particular types of years is important. Similarly, the results of information gain classification based on standardized data do not reflect the basic hydrolo-

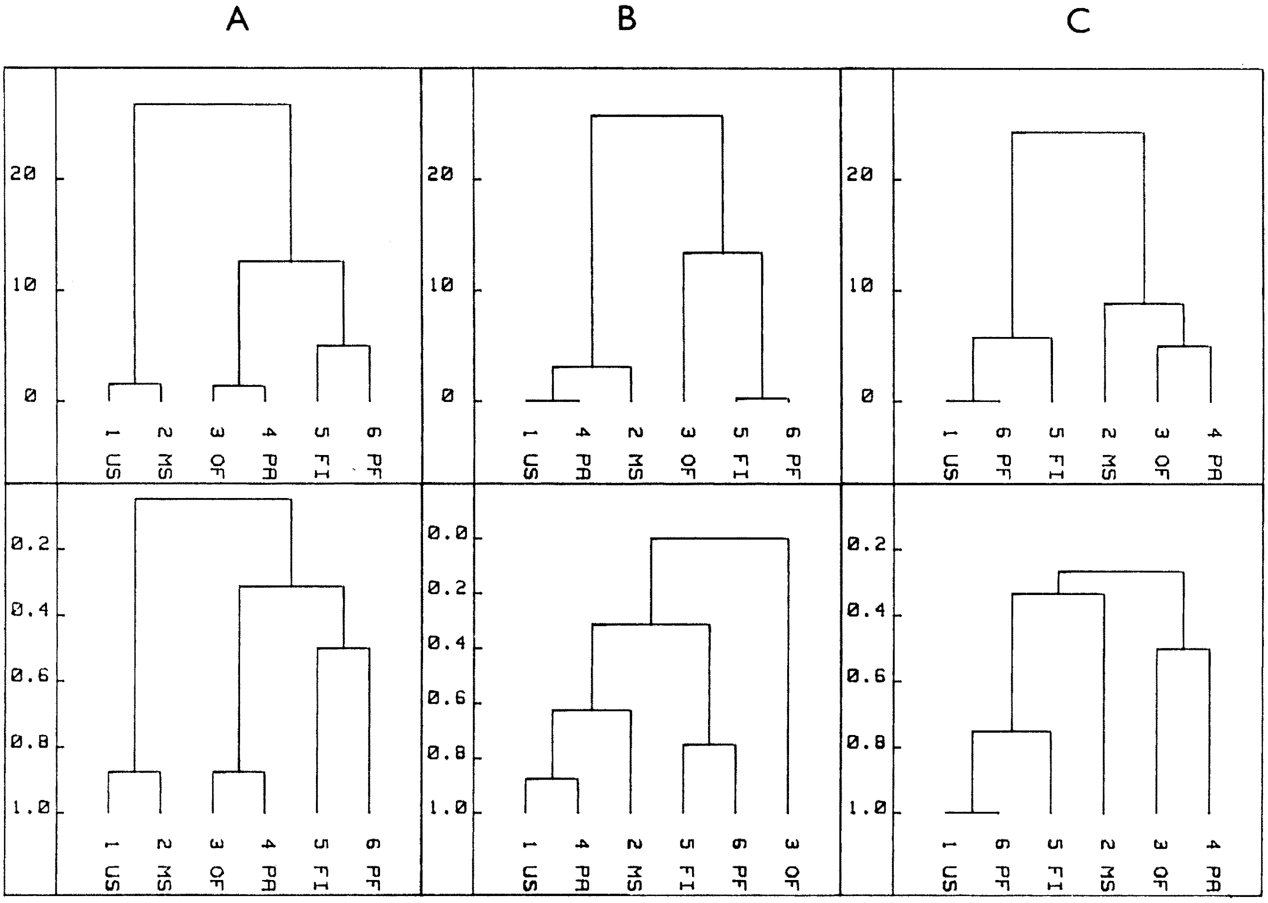

Fig. 5. Classification of soil macrofauna time sequences based on all organisms (column A), on millipeds only (B) and on staphylinids only (C). Similarity measure: LCS/length (upper row), and information gain (bottom row). Particular habitats are: 1 US - unmowed steppe; 2 MS - mowed steppe; 3 OF - oak forest; 4 PA - pasture; 5 FI - wheat field; 6 PF - ploughed field (without any crop).

gical and stratification differences between years.

Soil macrofauna

The primary classifications based on the whole set and on taxonomically defined subsets differ considerably, and consequently, the sequences are also different (Table 3). With respect to the low total number of samples, the four group classification level was accepted. The classification of sequences based on millipedes (Fig. 5b) corresponds best to different habitats: grasslands, fields and forest. Classifications based on all taxonomic groups (Fig. 5a) and particularly on staphylinids (Fig. 5c) are ecologically less interpretable. There are virtually no differences between the two methods used.

Discussion and conclusions

Despite the pioneering works of Williams et al. (1969) and Dale et al. (1970) the classification of sequences has been seldom used in community ecology (Dale et al. 1988). It is used more often in molecular biology (for classification of sequences of amino acids, nucleotides, etc., such as in Davison 1985); other applications are in soil science (Norris and Dale 1971, Little and Ross 1985). Advantages of similar approaches in environmental studies are discussed by Faith et al. (1985). However, many problems are connected with this method. The dependence of results of simple classification on the methodological options used were stressed many times (e.g., van der Maarel 1979, Kovář and Lepš 1986). As the two-step procedure was applied here, the number of possible combinations of options increases considerably. The application of divisive methods for primary classification leading to clearly defined groups may be practical (Little and Ross 1985), particularly for large data sets. In our case, the TWINSPAN method appeared to be useful.

Another question to be solved is the choice of the optimal level in hierarchical classification for the definition of groups. We have used the subjective criterion of ecological interpretability of results. Use of high hierarchical level (i.e., large groups) results in information losses (very different samples are treated as identicals). Conversely, use of a very detailed classification results in very dissimilar sequences (sequences contain very few common types). Moreover, as the hierarchy is not taken into account, the closed (i.e., similar) groups are treated in the same way as the different groups. It seems that use of higher hierarchical levels in primary classification leads to better determined major groups, but to less clear structure within them. String collapsing does not appear useful in our data sets. Predictably, as in all of our examples, the length of particular zones or duration of particular aspects is important.

In classification of sequences, the most difficult problems appear when classifying sequences of unequal length. Particularly, in the case of information gain as a measure, the sequences might be considered by the procedure as similar just because they are very short. In LCS based similarities, the standardization to the maximum length leads to low (and the minimum to very high) similarity of sequences of unequal length. It seems useful to consider for classification strings of similar length only. In all cases, the classifications based on LCS appeared to be superior to or the same as those based on information gain. The differences were greater in classifications of sequences of unequal length.

Despite apparent problems, the classification of sequences seems to be a promising method for solving some problems of community ecology. Nevertheless, its aim is only to compare sequences. Other methods should be used to analyse sequences. Whereas the transition matrices are not very useful for comparison, they help the analysis of time sequences (Usher 1981, Lepš 1987, Orlóci and Orlóci 1988). Various methods of gradient analysis are available to study the dependence of community composition on the position on a gradient (e.g., Orlóci 1988, Ter Braak 1986). The classification of sequences is only one method of data analysis, which should be complemented by others. It seems that the wider use of classification of sequences is limited by the lack of available programs. Our programs are written in extended BASIC for HP-9845 and are available from the first author.

Acknowledgements. The paper was inspired by a lecture by M.B. Dale presented in Budapest in 1985. We thank Dr. Russ Schiel, Albury, Australia for corrections of our English.

REFERENCES

Bogach, J., S.I. Golovach, A.D. Pokarzhevski and L.P. Titova. 1984. Diplopods (Diplopoda, Chilopoda) in soils of the Centralnochernozemnoy zapovednik nature reserve. In: Anonymus (ed.), Ecological-faunistic Investigations in Central Forest Steppe of the European Part of the USSR, pp. 75-90. Moscow, Lesnaya Promyshlenost. (In Russian).

Coggins, J.M. 1983. Dissimilarity measures for clustering strings. In: D. Sankoff and J.B. Kruskal (eds.), Time Warps, String Edits and Macromolecules: the Theory and Practice of Sequence Comparison, pp. 311-321. Addison Wesley, London.

Dale, M.B., P. MacNaughton-Smith, W.T. Williams and G.N. Lance. 1970. Numerical classification of sequences. Austr. Computer J. 2: 9-13.

Dale, M.B. and P.E.R. Dale. 1986. Similarity and structured attributes in ecological classification. Abstr. Bot. 10: 17-34.

Dale, M.B., P.E.R. Dale and R. Coutts. 1988. Classification of vegetation sequences in Toohey Forest, Queensland. Vegetatio 76: 113-129.

Davison, D. 1985. Sequence similarity ('homology') searching for molecular biologists. Bull. Math. Biol. 47: 437-474.

Faith, D.P., M.P. Austin, L. Belbin and C.R. Margules. 1985. Numerical classifications of profile attributes in environmental studies. J. Environ. Manage. 20: 73-85.

HILL, MO. 1979. TWINSPAN - a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Ecology and Systematics, Cornell University, Ithaca.

HRBÁČEK, J. and M. STRAŠKRABA (ed.). 1973. Hydrobiological Studies. Vol. 2. Academia, Praha.

JAVORNICKÝ, P. and J. KOMÁRKOVÁ. 1973. The changes in several parameters of plankton primary productivity in Slapy Reservoir 1960-1967, their mutual correlations and correlations with the main ecological factors. In: J. Hrbáček and M. Straškraba (ed.) 1973. Hydrobiological Studies, Vol. 2., pp. 155-212. Academia, Praha.

KOVÁŘ, P. and LEPŠ. 1986. Ruderal communities of the railway station Česká Třebová (Eastern Bohemia, Czechoslovakia) - remarks on the application of classical and numerical methods of classification. Preslia 58: 141-163.

LEPŠ, J., M. STRAŠKRABA, B. DESORTOVÁ and L. PROCHÁZKOVÁ. 1990. Annual cycles of plankton species composition and physical chemical conditions in Slapy Reservoir detected by multivariate statistics. Arch. Hydrobiol. Beih., Erg. Limnol. 33: 933-945

LEPŠ, J. 1987. Vegetation dynamics in early old field succession: a quantitative approach. Vegetatio 72: 95-102.

LITTLE, I.P. and D.R. ROSS. 1985. The Levenshtein metric, a new means for soil classification tested by data from a sand-podzol chronosequence and evaluated by discriminant analysis. Aust. J. Soil Res. 23: 115-130.

MAAREL, E. VAN DER. 1979. Multivariate methods in phytosociology with reference to the Netherlands. In: M.J.A. Werger (ed.), The Study of Vegetation, pp. 163-225. Dr. W. Junk, The Hague.

NORRIS, J.M. and M.B. DALE. 1971. Transition matrix approach to numerical classification of soil profiles. Soil Sci. Soc. Amer. Proc. 35: 487-491.

ORLÓCI, L. 1978. Multivariate Analysis in Vegetation Research. 2nd ed. Dr. W. Junk, The Hague.

ORLÓCI, L. 1988. Detecting vegetation patterns. ISI Atlas of Science, Plants and Animals, Vol. I, pp. 173-177.

ORLÓCI, L. and N.C. KENKEL. 1985. Introduction to Data Analysis with Applications from Population and Community Ecology. Statistical Ecology Monographs, Vol. 1. International Co-operative Publishing House, Fairland, Maryland.

ORLÓCI, L. and M. ORLÓCI. 1988. On recovery, Markov chains, and canonical analysis. Ecology 69: 1260-1265.

TER BRAAK, C.J.F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67: 1167-1179.

USHER, M.B. 1981. Modelling ecological succession with particular reference to Markov models. Vegetatio 46: 11-18.

WILLIAMS, W.T., G.N. LANCE, L.J. WEBB, J.G. TRACEY and M.B. DALE. 1969. Studies in the numerical analysis of complex rain-forest communities III. The analysis of successional data. J. Ecol. 57: 635-654.

Manuscript received: July 1989