## HIERARCHICAL SEMI-CLUSTER ANALYSIS (HSCA): A NEW METHOD OF GRADIENT ANALYSIS

### KAREL MATĚJKA

Forestry and Game Management Research Institute, Jfloviště-Strnady, 156 04 Praha 516, The Czech Republic<sup>1</sup>

#### Abstract

Matějka K.: Hierarchical semi-cluster analysis (HSCA): a new method of gradient analysis. Ecology (Bratislava), Vol. 12, No. 2, 131 – 152, 1993.

A novel method of a priori sequentially ordered samples has been developed for vegetation transect evaluation. Some new coefficients of associated similarity (homotoneity) are suggested for its application. The HSCA procedures operate on the principle of agglomeration and can be used to identify typical sample groups on a gradient. The method has been applied to data of four short quadrat transects of pond littoral vegetation and adjacent meadows of the Kratochvíle pond in South Bohemia. Vegetation zonation has been studied by use of this method in dependence on the complex moisture gradient. Results are confronted with those following from the classical cluster analysis and common classification as carried out by the Zurich-Montpellier school.

### 1. Introduction

Recent intensive progress of the method of multivariate analysis resulted in several new methods or in innovated existing ones (e.g. Orlóci, 1978; Gauch, 1982; Legendre, Legendre, 1987; Ludwig, Reynolds, 1988 among others). The hierarchical semi-cluster analysis, described in the present paper, sets out also from the fundamentals of classical cluster analysis, yet it makes use of the a priori information on linear ordering of the samples.

Let us have this exercise: There is a set of n points in a multidimensional space and established a sequence of these points. We should like to classify these points so, that we would obtain some continuous sections regarding to the sequence. It is a problem of constrained clustering. For using in ecology, a few methods were described (cf. Legendre, 1987). The typical plant ecological data are a quantified list of species growing in the respective quadrats (phytosociological relevés) along the transect on some gradient. In this example, the order numbers of the quadrats along the transect determine the sequence

<sup>&</sup>lt;sup>1</sup>This paper was created in Institute of Soil Biology, Czechoslovak Academy of Sciences, Na sádkach 7, 370 05 České Budějovice, The Czech Republic

of the classifying points (relevés). The method makes it possible also a classification of sections and boundaries for other sequentially ordered sets.

The HSCA is a complement of the existing methods of constrained clustering which are possible to divide into groups:

- A. 1. one-dimensional methods,
  - 2. two- or more dimensional methods.
- B. 1. divisive methods (e.g. Gordon, 1973),
  - 2. agglomerative methods:
    - a) routine clustering methods clustering with a constraint (e.g. Margules, Faith, Belbin, 1985; a difference between methods with and without the constraint consists in reduction of number of admissible combination of samples),
    - b) special constrained methods (e.g. HSCA).

It is included under term gradient analysis only the ordination techniques but also an interpretation of community changes along the gradient. The determination of homogeneous areas the gradient is one of conditions of this interpretation. The successful solution of this target is possible to realize by tools of constrained clustering. It is a reason to the including of HSCA into the group of gradient analysis methods.

### 2. Description of the HSCA method

In the following text, P denotes the total number of species on a transect (i.e. the total number of characters of the set treated), N is the total number of records within the set (i.e. of classified individuals), and X is the basic data matrix (for each its element, first index indicates the order of the species and the second one the order of the sample:  $X = [x_{ij}], 0 < i \le P$  and  $0 < j \le N$ ).

Using the basic matrix X we compute the matrix  $S = [s_{ij}]$ ,  $0 < i \le N$ ,  $0 < j \le N$  of the associated similarities (homotoneities) of subsets. The elements of a subset are formed by all relevés on the transect section between quadrats *i* and *j*. Indices of these boundary plots are the indices of the respective element of the matrix S. The choice of suitable coefficients of associated similarity is very important for successful applications of the method.

The HSCA procedure selects clusters (i.e. sections along the transect) having maximum homotoneity and the method ensures simultaneously a hierarchical classification structure in the form of a dichotomic tree (terms have been selected to suit the classical methods of cluster analysis).

The entire process is carried out in the following steps:

a) Computation of the matrix S of associated similarities (independent of the choice of acceptable coefficients).

b) 1. Selection of the sequence of n = |i-j| + 1 samples  $(i \neq j)$  so as to make the respective associated similarity  $S_{ij}$  higher or maximally equal to any other element  $S_{kl}$   $(k \neq 1)$ .

b) 2. Rearrangement of the matrix S as to prevent in subsequent steps an impermissible classification of some sample i to j within the cluster of a higher order (that one may contain just whole clusters of lower orders). Elements  $S_{kl}$  (and  $S_{kl}$ ) of the matrix S are then substituted by an arbitrary negative value for

1. 
$$k = i \text{ to } j$$
-1 &  $l = 1 \text{ to } j$ ,

2. 
$$k = i + 1$$
 to  $j \& l = j$  to  $N$ .

b) 3. The entire selection procedure including the rearrangement of the matrix S is repeated following items b.1 and b.2 until a complete union of all N samples is obtained [it is (N-1)-times].

When carrying out the computer calculation after the program, some steps may be simplified. We may e.g. use just the half of the matrix S due to its symmetry. The computation of single elements is carried out stepwise, as in cluster formation only two neighbouring clusters of a lower order are considered. By selecting a defined cluster, we eliminate in the following course of the algorithm the possibility of including some samples from the series of theoretically feasible clusters and the respective similarity coefficients are useless. From this reason, some elements of matrix S are not computed all. The portion of coefficients necessary out of all coefficients (i.e. (N-1)\*N/2) is dependent on the sample number N and on the average it makes

$$\frac{2}{N} + \frac{4}{N^*(N-1)} \sum_{i=1}^{N-2} \frac{N-i-2}{N-i-1}.$$

For N=5 this is approximately 78 %, for N=10 47 %, and for N=50 just 11 %, respectively.

### Choice of the associated similarity coefficients

The fundamental condition for applying any coefficient of associated similarity (homotoneity) is as follows (under presumption i < j):

$$S_{ij} \le \min_{\substack{i \le k < l \le j}} (S_{kl}) \tag{1}$$

It must be valid also:

$$S_{ii} = 1, \tag{2}$$

$$S_{ij} \in \langle 0; 1 \rangle. \tag{3}$$

The coefficients for application are the following:

1. Average constancy of all participating species in the given subset of records, which may be written as

$$S_{ij} = \sum_{k=1}^{p} \sum_{l=i}^{j} Q_{kl} / \sum_{k=1}^{p} \prod_{l=i}^{j} (1 - Q_{kl}),$$

where  $Q_{kl}$  is the information on presence or absence of k-th species in l-th record, expressed by 1 or 0 respectively. If n is the number of records (n = |i-j|+1), the minimum theoretical value of this coefficient is given by

$$\min\left(S_{ij}\right)=1/n.$$

The behaviour of the average constancy is explainable using the coefficient expressed by

$$Y_{ij} = 1 + \frac{1 - 1/S_{ij}}{|i - j|}.$$

It is identical for n=2 with the Sörensen's index of floristic similarity Cs (Sörensen, 1948). Because the Y coefficient does not however meet condition (1), the Y is not an applicable coefficient for the HSCA method.

2. Next coefficient, using quantitative data (e.g. data on cover, biomass, abundance, etc.), is the variable  $Sq_{ij}$  defined as the supplement of the sum (over all species) of variances along the examined section of the transect containing samples *i* to *j*. We may therefore write

$$1 - Sq_{ij} = \sum_{k=1}^{p} var_{ij} \left( x_{k.} \right).$$

This coefficient requires the data standardization on sum of significance of all species for each record (in other words, for each record we must have  $\Sigma x_{k.} = 1$ ), the latter relation may be re-written to

$$1 - Sq_{ij} = C^{ij} - \tilde{a}^{ij},$$

where  $C^{ij}$  is the average value of the Simpson's index of dominance (Simpson, 1949) for relevés *i* to *j* and  $\bar{a}^{ij}$  is the identical index calculated for relevant average record, i.e. such a record in which the significance of each species is expressed by the average significance (for example the cover of species) for the respective species in records *i* to *j*. Further we have

### $\min(Sq_{ij}) = 1/P_{ij}$

( $P_{ij}$  is the total number of species in records i to j).

3. An entire series of coefficients can be derived from any single similarity index (whose supplement to one is the metric or submetric) by taking for the associated similarity coefficient the lowest value of the single similarity index from those which exist for all pairs of records within the given section of the transect (principle of most distant neighbours). For such a category of measures the already mentioned Sörensen's index Cs had been already chosen. The respective coefficient is given by

$$dC_{ij} = \min_{k,l \in \langle ij \rangle} (Cs_{kl}) \,.$$

4. Due to the fact that the maximum theoretical Euclidean distance of two samples as points in a *P*-dimensional Euclidean space is equal to  $\sqrt{2}$  by standardization as for the coefficient Sq, we may use the supplement  $1-ED/\sqrt{2}$  as the similarity index and treat it analogously to the previous case:

$$dE_{ij} = \min_{k,l \in \langle ij \rangle} (1 - ED_{kl}/\sqrt{2}).$$

Using similarity coefficient (S, Sq, dC, dE, respectively) is in any case a part of the denotation of an actual procedure (HSCA/S, HSCA/Sq, etc.).

Concluding, it is necessary to remark that the Moravec's coefficients of homotoneity bH and cH (Moravec, 1971) do not satisfy the condition (1) and cannot therefore be used in the HSCA procedure.

### 3. Study area

The littoral and meadow vegetation of the Kratochvíle pond (432 to 435 m a.s.l.) in the vicinity of Netolice in South Bohemia has been studied (latitude 49°13'40"N, longitude 14°10'05"E). The main part of the area is covered by a holoceneous floating loam, in contact with loess clays of Würme. Marginally, the neogeneous sediments of Mydlovary strata series are also encountered. The porphyric melanocratic granite is the primary mineral, uncovered only outside the study area. The area is counted among middle warm, humid to moderately humid, hilly climatic region (Vesecký et al.,1958), having a mean yearly temperature about 7°C and annual precipitation of 600 to 650 mm.

Littoral vegetation of the pond is formed by the stands of association of *Caricetum elatae* K o c h 1926, *Caricetum gracilis* A l m q u is t 1929, *Phragnitetum communis* (G a m s 1927) S c h m a l e 1939 and *Glycerietum maximae* H u e c k 1931 on the place with more intensive sedimentations and accumulation of organic matter. The meadow vegetation is formed by an untypically developed association *Angelico-Cirsietum oleracei* T ü x e n 1937 in several variants containing different transition types. Further there exist communities with *Geranium pratense* and other stands belonging to the alliance *Alopecurion* and *Arrhenatherion*. Riverine forest occurring in the region of acidophilous oak forests represents the natural vegetation in this area (Mikyška et al., 1968).

### 4. Data collection

Four transects orientated from the littoral zone towards the higher situated drier parts (consisting of altogether 53 quadrats of 1 m<sup>2</sup> size; the quadrats did not follow one to other closely, their distances were more or less regular) were established in the area. For each quadrat, the cover of all vascular plant species was estimated (Table 1 and 2). New twelve degree modification of the Braun-Blanquet scale for abundance and dominance was used (similar to the ordinal transformation, van der Maarel, 1979). Single values are expressed by integers, 1 to 12. In brackets, original values of the Braun-Blanquet scale are given: 1 (r), 2 (+), 3 (+-1), 4 (1), ..., 11 (4-5), 12 (5), the zero value is used for absence of the species. The nomenclature of vascular plants is according to Rothmaler et al. (1982).

The transect T1 (Table 1; total length 55 m) intersects the strongly limited littoral stands containing *Carex gracilis* and passes over the mowed wet meadow of the *Calthenion*, another transition to drier parts over the communities of the alliance *Alopecurion* to *Arrhenatherion* is recorded. At this point, the strongly ruderalisated stand with *Geranium pratense* and *Urtica dioica* begins and continues to a formerly

Transect No. of quadrat	T1 1		3	4	5	6	7	8	9	10	iı	12	13	14	15	16	17	18	19	72 1		3	4	5	6	7
Filipendula ulmaria	2	3	5																		4					
Glyceria maxima	7	6	1		÷	•	¥.		•			٠	•							1	4				÷	
Phalaris arundinacea	7	8			1R		-																			
Carex gracilis	9				×																7					
Equisetum palustre	2						æ	×			e.										1					
Spirodella polyrhiza	2															æ			i.				×.			
Taraxacum officinale ag	g.																									
population 1		3	5	5	5	4	3	1	1	1						•						1	1	1	3	5
population 2		÷	•		÷								4	6	6	5	7	5								
Carex vesicaria		6	1		•	•	÷		•	×	•	÷	•			•			*		8	•	•		÷	
Galeopsis tetrahit		1					*							X			•					÷			÷	÷
Poa pratensis		-	5	6	6	4	5	7	5	5	5	5	5	4	6	6	4	3				1	5	1	6	5
Alopecurus pratensis			6	3	,					5		3	4	4	4	3	3	3			6	4	4	5	4	1
Festuca pratensis			3	3	4	5	6	5	6	6	5	5		6	3		3					6	5	5	4	
Rumex acetosa			1	1	3	1	3	3	3	3	3	3	4	3		•						•	3		3	
Lathyrus pratensis			3		1	6	5	1	4	3	4	3	5	3					2			3	3	1	1	5
Ranunculus auricomus			1	1	1	3	3	1	3	3			1	1								4	1	3		
Ranunculus acris			3	6	6	3	3	3	3	3	3	3	3					•				1	1	1	1	
Cardamine pratensis			5	4	3	4	3	1			к		1	,									1			
Festuca rubra			5	1	1	2	4	3	5	6	5	7														
Deschampsia cespitosa			3	3	3	2	•			•	1	3												4		
Sanguisorba officinalis																								1		
Myosotis nemorosa																										
Trifolium repens															343									1		
Lychnis flos-cuculi			3	1				1	1												,	5	3			
Ranunculus repens															•									3		
Carex panicea																										
Caltha palustris ssp. laeta																										
Juncus articulatus			1				1.						2				2	141					ļ			
Veronica serpyllifolia			1																							
Trifolium pratense							4								540			4	1	2		5	5	4	3	5
Cerastium holosteoides																								3		

T a b l e 1. Composition of vegetation along transects T1 and T2. Estimated covers of species in values of 12-degree scale in sample plots 1  $m^2$  are given

# Table 1. (Continue.)

Transect No. of quadrat	T1 1		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	T2 1		3	4	5	6	7
Holcus lanatus				4	5	4	5	6	6	6	6	5	5						•			5	5	4	4	
Anthaxanthum odoratum				4	6	6	6	7	5	3	5		1									5	5			
Eleocharis palustris					1							•													•	
Plantago lanceolata						2	3	1	1	1	3	3	3				•							÷.	4	3
Juncus filiformis	•					2								•							٠			÷	ě	÷
Angelica sylvestris						1														4	÷	3			4	
Thalictrum lucidum	•					1									•							1				
Achillea millefolium							3		1	3	5	5	5	3								×		3	5	5
Luzula campestris							3	3			1	3						×			•		•			
Ajuga reptans							1										: •5						•			
Alchemilla subcrenata								1		1	3	3	3	3	2							3	4	3		3
Cynosurus cristatus		,						1	1	3	3		1	•		•					•	3	4	3	2.0	
Cirsium palustre								3	3		-								342						5.0	
Lotus comiculatus								3						(1 <b>4</b> ))				×						×		
Cirsium oleraceum									1			•2	×				. e.)	*							3	
Avenula pubescens	•					•			1	1		3	3	•			•		31			į,		4		÷
Succisa pratensis	2								1	4					÷		÷	÷		36	e	3				
Geranium pratense		÷			4		4			1	3		4	6	5			÷				5	5	5	6	5
Dactylis glomerata											3	1	5	4		a.								5	6	5
Equisetum arvense											1		1	2					240						142	
Trisetum flavescens											3	3	5					×	900 I					3	4	5
Leucanthemum incutionum											3	3							80							1
Saxifraga granulata		÷									1	5					•		•		÷		•			
Veronica chamaedrys					•						1	3											•		2	1
Stellaria graminea											1			•			,					1		14		1
Pimpinella major												3	5	4			•						1	3	3	3
Knautia arvensis												3							•	340					•	
Betonica officinalis												1										3	3			1
Heracleum sphondylium													5	5	4	5	3					2				
Urtica dioica												,		4	4	4	5	6	1		,					
Aegopodium podagraria															5						ļ			4		
Agropyron repens	9						2						226			8			7						. 5	10

### Table 1. (Continue.)

Transect	T1																			Т	2					
No. of quadrat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	1	2	3	4	5	6	7
Stellaria graminea	•				•				3			•			3	3	3	5			,					
Tripleurospennum inodonum			ł	•			÷.	•	÷	•	•		•		3	1	4	1	•					•	٠	
Capsella bursa-pastoris							×.		÷				×.		1			1	٠						•	
Hordeum vulgare								,				•			1	4	2	•						•	•	
Myosotis arvensis	•			•						×					1		1					•				
Brassica napus	•	•							•		•				1	3										
Veronica agrestis		•			•	•				•	÷		•	•	2	÷		•								
Carex hirta	•	•		•			•		•	•			•		1	•		•								
Galium aparine					•											4	4	3					•		•	
Anthriscus sylvestris				•												4	4								4	6
Rumex obtusifolius	•														•		4	1	3							
Poa palustris	•		1			•	•	•	٠	•	•	•	÷		•			6	7		4	5	5	5		
Plantago major	•	•			•		•			•	•			•	٠			1						٠	•	•
Myosoton aquaticum										÷				•	٠			•	5				•			
Ranunculus sceleratus	•								•										3					•		
Symphytum officinale	•					•	×		•										3							
Alopecurus aequalis	•	•																	1							
Stellaria uliginosa																	,		1							
Scirpus sylvaticus	•		•	٠	•	•		•		÷	•	•	·	•	٠	•				i.	4	5	4	•	•	
Galium palustre	•					÷			×			-			•	•					1	1				

abandoned field. Soil is strongly compacted in these parts of the transect and a more intensive wetting manifests itself (most probably as a consequence of the preceding phenomenon), with a prevailing stand of *Glyceria maxima*.

The transect T2 (Table 1; total length 12 m) is an analogy to the first one and depicts the basic sequence of wet mowed meadow vegetation. Its first quadrat catches however the end of the *Glyceria maxima* swamp community.

The transect T3 (Table 2; total length 34 m) leads from shore swamp with accumulation of (prevalent organic) deposits (with *Phragmites communis, Glyceria maxima, Carex gracilis* and *Phalaris arundinacea*) over the periodically mowed stand beginning by a transition zone with *Filipendula ulmaria* and *Carex vesicaria*. The total species

Transect No. of quadrat	T3 1		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	T4 1		3	4	5	6	7	8	
	_	_	_	_	_				-		_						-					-					_
Phragmites australis	6	6	5	2	3	3	4	3	5	4	2	4	·	٠	•	·	•	•	•	٠	÷	•	•	•	4	ą.	
Glyceria maxima	7	7	•	•	•	٠	·	3	•	•	·	٠	•	•	·	•	196		٠		×	•					
Equisetum palustre		3	•	•	•	•	•	•		•	2	2	2	٠	3	3	3	3	1	1	1			1			
Phalaris arundinacea		6	9	9	9	9	10	10	6	•		•	•	•		•			•	•		•	9 <b>7</b>			×	
Carex gracilis		3	7	5	5	6		•				٠	•	•	·		•		4								
Caltha palustris ssp. laeta	•	٠	٠	•	•		4	3	•	4	٠	4	•	٠	•	•	•	÷	•	٠	•	•	•	÷		÷.	
Ranunculus auricomus	•	•				٠	•	3	3	4	4	4	2	2	3	3	4	٠	1	4	4	4	4	1	8	1	
Carex brizoides		•			•	•		•	6	5	5	4	5	6	5	6	4	3		•	×	*	<b>.</b>			j.	
Cirsium palustre				•	•	•			3	2	4	2	5	5	3		4				1	•		÷	*	30	
Filipendula ulmaria					•				5	2	•	2	4	2			•	×		5	5				1	1	
Ranunculus repens		÷			•	•		•	5	4	4	5	4	•			•		3	3		1	•	÷	÷		
Carex vesicaria	•				•	•	•	•	3	2	•	2	•					×	1	1	•				9		
Lathyrus pratensis									3	4	4											•	5			163	
Myosotis nemorosa									3		2								1	3						14	
Galium palustre									3	2							œ		3	4			•	×			
Poa palustris	•					•		•	5						×				6	1			•	*	28	2	
Juncus effusus	•								3				•				•2	*		•	•	3			37		
Festuca rubra					•					4	2	5	6	6	8	7	7	9		6	7	7	6	7	9	9	
Cardamine pratensis								•		2	4	4	4	4	4	3	4	3	4	3	1	1	1	1	1		
Holcus lanatus									•	2	5	5	5	5	,	3	5	3		5	5	6	5	4		45	
Sanguisorba officinalis					•					5	5	5	4	4	3		4	3		5	5	6	5	5	6	7	
Poa pratensis	•									4	ļ,			4	4	6	4	6	1				1	4	1	•	
Carex nigra					•					4	4	4	5		3		3		4	5	5	3		1			
Anthoxanthum odoratum										4	5	5	4			•	4			5	4	•				÷	
Lychnis flos-cuculi										4	4	2	4	2	3												
Carex panicea										4	4	4	4						144		4	1					
Geum rivale				,						2	2	2										1					
Juncus filiformis										6									6	4							
Valeriana dioica										4																	
Carex canescens			,							2									4							18	
Festuca pratensis											2	4	4	4	5	5	4	3		5	5	1	4	5	3		
Rumex acetosa																		3							4		

T a b 1 e 2. Composition of vegetation along transects T3 and T4. Estimated covers of species in values of 12-degree scale in sample plots  $1 \text{ m}^2$  are given

# Table 2. (Continue.)

Transect	T																		T4					÷			
No. of quadrat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	1	2	3	4	5	6	7	8	9
Geranium pratense				,						÷	2		4	4		3	3	3		•			1	3			
Ranunculus acris		×.		÷	÷		÷		٠	٠	2	2	4	2			•	3	1	1	3	4	5	1	1	1	
Deschampsia cespitosa		×.	•			•		•	ŝ.	•	4	,	4	5	5	5	5	÷	•		•	4	•	•		1	•
Agrostis stolonifera			×		×				×			2							4	1							
Cerastium holosteoides									•			2	30		×									1	5		•
Alchemilla subcrenata				21					<b>:</b> •0				2	2	3		3			•		1	5	5	4	•	1
Angelica sylvestris		•					,		٠	•	•		2		3	•	3	•			ŀ		•	ŀ	•	•	
Luzula campestris	4						•						2	2	•	3					3	1					
Carex hirta	~						×		•			a		2	4		3								•		
Selinum carvifolia	*								×					2			•						•				
Avenula pubescens					•										3		3					1	4				
Alopecurus pratensis																6	4	6		•		1	1	5	1	1	1
Ajuga reptans			÷	,	ŝ		÷			ł			,	÷			3					÷					
Saxifraga granulata							,										÷	3					4	4	3	1	1
Trifolium hybridum								÷		÷			÷						1	4	3						
Ranunculus flammula		E.					,			k									4	3							
Trifolium repens						,								×.	,				4	1							
Epilobium sp. div.													÷						1								
Lysimachia nummularia		•			ŕ	÷	÷		÷	•					•		÷	•	1				•	•			
Taraxacum officinale agg."					÷				÷				•							1	4	1		4	5	1	1
Carex leporina							4		4		4										1			4			
Galium uliginosum				240	×				×.	÷											1						
Potentilla erecta							De.								×						1						
Phyteuma nigrum		×			192	×	×.	×														4	1			1	
Galium boreale		,				,																5	4				
Betonica officinalis						÷	4		R													3					
Briza media													1			7.			-			3					
Carex pallescens	100 100	9) 16	े अ							-				-					333 746	•		1					
Campanula patula																							1		1	4	1
Phleum pratense			Ĵ							Ĵ	Ĵ,							÷		Ĵ				1	1		

 $^{*}$ This taxon is identical with population 2 in Table 1.

### Table 2. (Continue.)

Transient	T																		T4	ł							
Transect No. of quadrat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	1	2	3	4	5	6	7	8	9
Veronica arvensis									•									÷						1	•		
Agrostis tenuis				٠										•					*		•	•	٠		1	1	6
Trisetum flavescens																											
Pimpinella major							•													•					•]	1	

diversity and equitability is increased here immediately. The respective stands may be classified within the suballiance *Calthenion*. Next follows the stepwise transition to dryish growths (Alopecurion pratensis) of Festuca pratensis, Geranium pratense, Luzula campestris, Alopecurus pratensis, Saxifraga granulata, etc.

The transect T4 (Table 2; total length 16 m) leads from the edge of the meadow near the coastal willow woods and represents further an analogous sequence of phytocoenoses as presented in the preceding transect, yet on a steeper gradient.

The cluster analysis after Ward's method (sum of squares clustering, SSC, see e.g. Orlóci, 1978; Legendre et Legendre, 1983) using Euclidean distance (ED) has been used for comparison.

### 5. Results of an application

Cluster analysis of all the relevés has been carried out using data on simple presence of species (Fig. 1) or on estimates of cover for single species (Fig. 2). Based on obtained results, seven or ten vegetation types have been discerned, having some specified distribution within the shore plant zonation (Fig. 3). As follows from a comparison of both classification procedures (Table 3), both points of view must be considered when evaluating the communities represented by vegetation records of single quadrats: the presence of a species and its quantitative representation (cover).

All the transects have been evaluated separately by the new method using two associated similarity indices - the qualitative index (HSCA/S procedure, Figs 5, 7, 9a, and 9b), and the quantitative one (HSCA/Sq procedure, Figs 6 and 8). The obtained results correspond to classifications produced by common methods of cluster analysis using presence-absence data or covers.

As may be seen from all the presented classifications, their results differ, sometimes apparently diametrally. Some mutual features can be found despite that fact. Setting out from numerical classifications of all quadrats (Figs 1 and 2), and considering two systems of units differentiated in this way, we may construct their sequence along individual transects (Fig. 3). As may be seen, we have generally distributed sequences

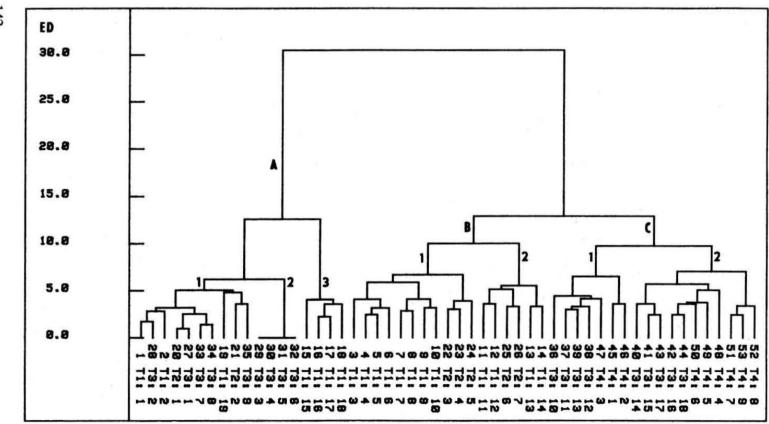


Fig. 1. Numerical classification of all quadrats (transects T1 to T4 in meadows on shore of Kratochvílsky pond, South Bohemia) using the SSC/ED (sum of square clustering with Euclidean distance metric) method and presence/absence data. A<sub>1</sub>, A<sub>2</sub>, ..., C<sub>2</sub> are designation of the classification groups (by the use of this type of data and method).

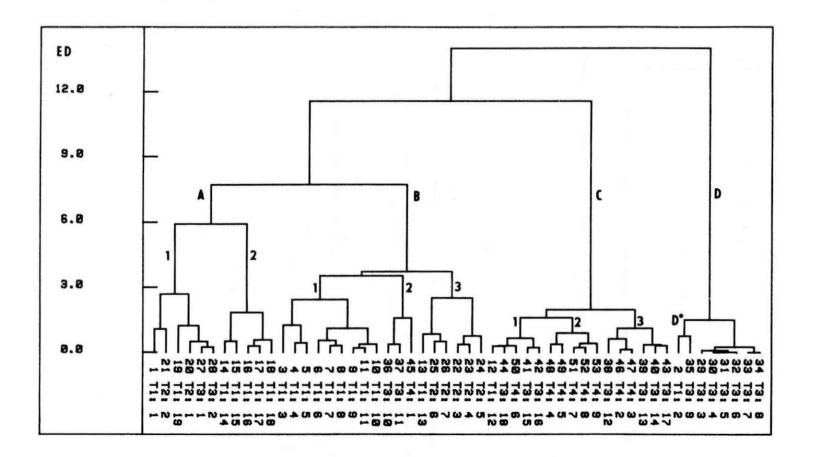


Fig. 2. Numerical classification (similar to Fig. 1) of all quadrats (transects T1 to T4) using the SSC/ED method and estimates of covers of all species as a measure of their representation. Data are normalized. A<sub>1</sub>, A<sub>2</sub>, ..., D, D<sup>\*</sup> are designation of the classification groups.

 T1:
  $A_1$  -  $B_1$  -  $B_2$  -  $A_3$  -  $A_1$  

 T2:
  $A_1$  -  $B_1$  -  $B_2$  -  $A_1$  

 T3:
  $A_1$  -  $A_1$  -  $C_1$  -  $C_2$  

 T4:
  $C_1$  -  $C_2$ 

B T1:  $A_1 - - - - D^* - B_1 - C_1 - B_3 - A_2 - A_1$ T2:  $A_1 - - - - - - B_3$ T3:  $A_1 - D - D^* - B_2 - C_3 - (C_3, C_1)$ T4:  $B_2 - C_3 - (C_2, C_1)$ 

Fig. 3. The sequence of vegetation types along studied transects (T1 to T4) on the basis of simple species attendances (A) or of species cover values (B). Designations of classification groups are given in the Figs 1 and 2, respectively.

of vegetation types on the place. We can also evaluate a relation between classification based on presence or absence (0/1 data), and classification based on species covers,

T a ble 3. The dependence between two classification systems with data of presence/absence (+/-) or with data of species covers (s. c.). The relation is expressed by numbers of quadrats at the separate combinations of vegetation types (see Fig. 1 and 2)

s.c. data	D	D	$\mathbf{A}_1$	A <sub>2</sub>	<b>B</b> <sub>3</sub>	$B_1$	B <sub>2</sub>	C3	C <sub>2</sub>	C1
+/- data:										
A <sub>2</sub>	4				•		٠	ž.	1.	ž
A <sub>1</sub>	2	2	6	•			•			ż
A <sub>3</sub>			*:	4		:•;				
$B_2$				1	3	1		•		1
$\mathbf{B}_1$				•	3	8			.•	
C1	•	( <b>a</b> .)		•	·	(ie)	3	4		
C <sub>2</sub>								2	5	4

A

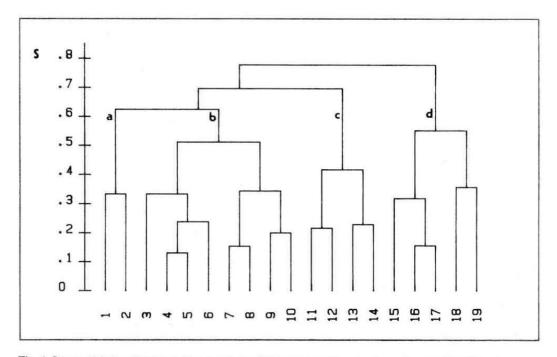


Fig. 4. Sequential classification of the quadrats of the transect T1 using the method HSCA/S (with average constancy of all participating species).

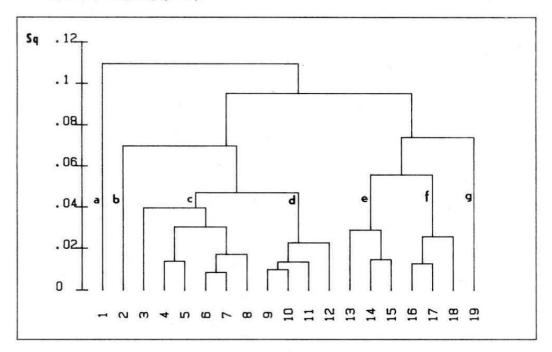


Fig. 5. Sequential classification of the quadrats of the transect T1 using the method HSCA/Sq (with supplement of sum of species variances).

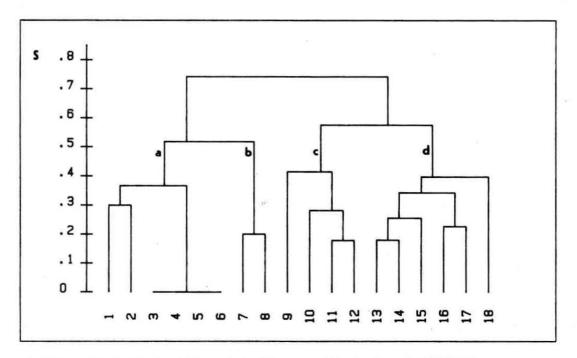


Fig. 6. Sequential classification of the quadrats of the transect T3 using the method HSCA/S.

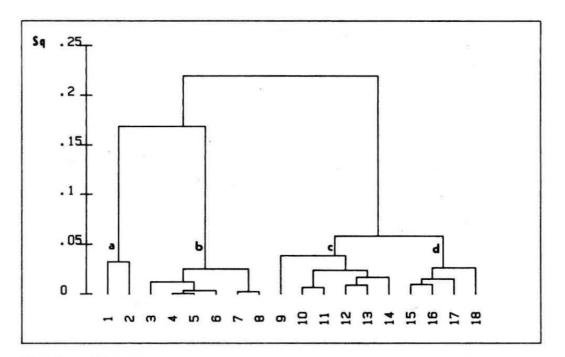


Fig. 7. Sequential classification of the quadrats of the transect T3 using the method HSCA/Sq.

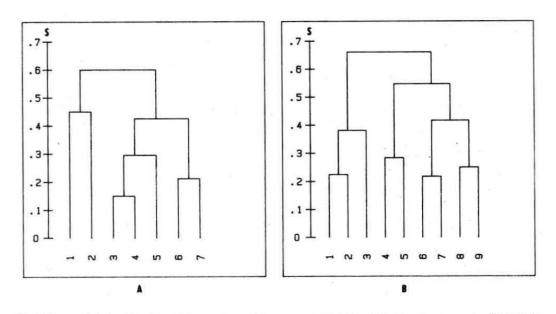
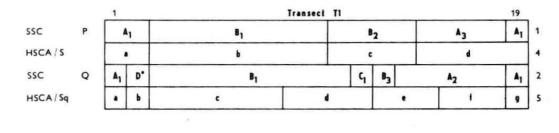


Fig. 8. Sequential classifications of the quadrats of the transects T2 (A) and T4 (B) using the method HSCA/S.

which is strongly significant (Table 3). Based on such a classification and making use of current methods of inductive phytosociological classification, the vegetation sequences along the transects may be valued as follows:



		1			ransec	1 13		18	В
SSC	Ρ	A1	A2	A		۲,		¢2	
HSC4/S			•	b		c		d	
SSC	Q	A <sub>1</sub>	D		D.	B2	¢3	C <sub>1</sub>	
HSCA/Se			b			c		4	

Fig. 9. The comparison of four classification methods of two transects (T1 and T3). Denotation of classification groups after respective figures (SSC P - Fig. 1, SSC Q - Fig. 2, HSCA/S - Figs 4 [transect T1] and 6 [T3], HSCA/Sq - Figs 5 [T1] and 7 [T3].

Transect T1:

1. Quadrats (further only q.) 1 and 2 are moderately mowed littoral stands, strongly limited, containing *Glyceria maxima* and *Phalaris arundinacea*.

2. Q.3 to q.10 in the mowed meadow *Calthenion* grassland stand can be identified as an association *Angelico-Cirsietum oleracei* T ü x e n 1937 (containing species as *Sanguisorba officinalis, Myosotis nemorosa, Caltha palustris, Cirsium oleraceum, C. palustre, Lychnis flos-cuculi*, etc.).

3. Q.11 to q.14 represent a transition zone to drier, higher situated stands (here with Achillea millefolium, Alchemilla subcrenata, Dactylis glomerata, Trisetum flavescens,

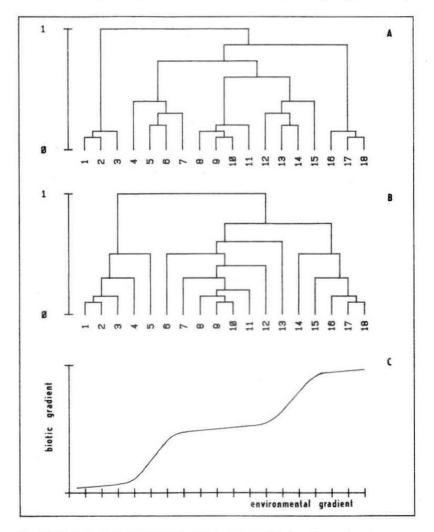


Fig. 10. Two theoretically possible results of classification of samples along some environmental gradient. A - The transition zone is classified as an independent unit having more increased heterogeneity. B - Both halves of the transition zone are ascribed to the neighbouring types. C - Relation between environmental and biotic gradient.

*Pimpinella major* and other species whose ecological coenological optima are known from the alliance *Alopecurion*).

4. Q.15 to q.18 are situated at the edge of an abandoned field, at present mowed together with the lower lying meadow. Growth is indicated by Agropyron repens, Stellaria media, Tripleurospermum inodorum and some other weeds and surviving culture plants.

5. Q.19 is positioned in a small ground depression, strongly wet where Glyceria maxima dominated.

The composition of the transect T2 is as follows:

1. Q.1 and 2 - littoral from time to time mowed swamp,

2. q.3 to q.5 - the Calthenion stand,

3. q.6 and 7 - transition zone of the Alopecurion community.

A relatively simple zonation was found for communities growing along the transect T3:

1. Q.1 to q.9 - reed swamp, which may be sub-classified on a deeper investigation, the q.9 is mowed occasionally,

2. q.10 to q.13 contain growth of the suballiance Calthenion,

3. q.14 to q.18 are distinguished by transition zone vegetation at a gradually decreasing soil moisture.

The transect T4 may be divided into two parts:

1. Q.1 to q.3 with growth of Calthenion, and

2. q.4 to q.9 representing the zone with decreasing moisture.

### 6. Discussion

The new HSCA method gives results analogous to those of hierarchical cluster analysis methods. Most important differences are not accounted by the proper method but by data transformations, as e.g. utilizing presence data (0/1 type) instead of data on quantitative abundance of all species (Fig. 9), in agreement with the work of other authors (see e.g. Kovář et Lepš, 1986). There is also an agreement between numerical and classical classification methods for vegetation on the investigated gradients. Most problems in comparing various classification do not arise by differentiated types of communities but by transition zones sometimes designated as ecotones (cf. Odum, 1971 and others).

Two possible artificial cases are shown in Fig. 10. Different methods may classify as follows:

a) the transition zone is classified as an independent unit showing however an increased heterogeneity when compared with adjacent types,

b) both halves of the transition zone are ascribed to the neighbouring types separated by this transition zone.

The first situation can be observed e.g. for transect T1 for quadrats 11 to 14 when using the HSCA/Sq method, the second situation manifests itself when applying the method HSCA/S. Many analogous examples may be found. We can presuppose that HSCA procedure identifies groups of typical samples on gradients. It ensues from the agglomerative principle of this method.

Classification of the vegetation records on a topographic gradient represented by a transect is far from being the only application of this new method. Sequential ordering of samples may be carried out using any (logically and formally correct) method. It may use for example the results of a direct gradient analysis (in the sense of Whittaker et al., 1973) or of some first coordinates of an ordination space. The method can be therefore used both for direct as well as indirect gradient analysis of communities.

In another sense, time gradient may also be considered. Cattell (1966) has defined altogether six types of analysis, the HSCA method can be used for four of these types: Q - among sequentially ordered objects (records) described by sets of participating species, O - among times in a time series described by all species on a defined area, T - among times in time series using the change of representation of one species on different areas, and S - among sequentially ordered objects (records) described by the change in representation of a certain defined species in time.

Generally, any sequentially ordered set may be classified using the HSCA method, all characterized by a defined number of characters. It could be the results of palynological analysis of peat profiles, or data on the succession or dynamics of some communities. The method can be used also for differentiation and classification of soil layers and microhorizons or for the classification of sediment layers.

The ordering of samples may be provided in one-, two- or more dimensional space. Margules, Faith and Belbin (1985) discussed the two-dimensional methods of constrained classification and their use in geography. The HSCA operates on the simplest possibility - a one-dimensional constraint (in sense Legendre, 1987).

The so-called barrier analysis was described in the past (Gordon, 1973) based on the principle of division and it is being used for automatic evaluation of pollen stratigraphy.

Legendre (Legendre et al., 1985) has used the similar base as the preceding author for evaluating successional changes in communities. Their method is called chronological clustering, and it uses some ideas of the ecological model of succession.

Webster (1973) used a statistical approach to constrained clustering (by comparison of halves of segments along a gradient) and this method has been applied for data of soil properties.

### 7. Summary

For the vegetation classification along transects, a new method - hierarchical semi-cluster analysis (HSCA) involving new similarity indices was developed and applied. As proved, this method meets the demand on a method classifying samples ordered sequentially in advance. The results obtained are comparable with the classification by means of hierarchical clustering. More possibilities for applications of the method are proposed.

Phytocoenological records resulting from quadrats having an area of  $1 \text{ m}^2$  have been classified, situated along four transects on the shore of the Kratochvíle pond near Netolice (South Bohemia, Czechoslovakia). The formation of vegetation zones on the shore, depending on the moisture conditions of the stand and on the way of its cultivation (or general anthropic influence), has been described. The general outline of zonation can be expressed as follows, making use of the Zurich-Montpellier synta-xonomical system: littoral stands (*Phragmitetum communis, Ghycerietum maximae*), *Calthenion* (e.g. *Angelico-Cersietum oleracei*), transition community types, *Alopecurion* (in a certain sense its growths may be taken also as transitional, depending on the steepness of the gradient), *Arrhenatherion*.

The HSCA program was written in FORTRAN and compiled under MS-DOS for the IBM PC and compatible computers (as up-building of the PC-ORD system; McCune, 1987) and in BASIC (for Hewlett-Packard 85, with graphical output). Both ones are available on the address of the author.

Translated by Z. Štěrbáček and the author

### References

- Cattell, R.B., 1966: The data box: its ordering of total resources in terms of possible relational systems. In Cattell R.B. (ed.): Handbook of multivariate experimental psychology. Rand McNally & Co., Chicago, p. 67 – 128.
- Gauch, H.G., 1982: Multivariate analysis in community ecology. Cambridge Univ. Press, 298 pp.
- Gordon, A.D., 1973: Classification in the presence of constraints. Biometrics, 29, p. 821-827.
- Kovář, P., Lepš, J., 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, p. 141 – 163.
- Legendre, P., 1987: Constrained clustering. In Legendre, P., Legendre, L. (eds): Developments in numerical ecology. Springer Verlag, Berlin etc., p. 289 – 307.

Legendre, P., Dallot, S., Legendre, L., 1985: Succession of species within a community: chronological clustering, with applications to marine and freshwater zooplankton. Amer. Natur., 125, p. 257 – 288.

Legendre, L., Legendre, P., 1983: Numerical ecology. In Developments in environmental modelling. Vol. 3, Elsevier, Amsterdam, 419 pp.

Legendre, P., Legendre, L. (eds), 1987: Developments in numerical ecology. Springer Verlag, Berlin etc., 585 pp.

- Ludwig, J.A., Reynolds, J.F., 1988: Statistical ecology. A primer on methods and computing. J. Wiley & Sons, New York etc., 337 pp.
- van der Maarel, E., 1979: Transformation of cover-abundance values in phytosociology and its effects on community similarity. Vegetatio, 38, p. 97 114.

Margules, C., Faith, D., Belbin, L., 1985: An adjacency constraint in agglomerative classification of geographic data. Environment Planning A, 17, p. 397 – 412.

- McCune, B., 1987: Multivariate analysis on the PC-ORD system. A software documentation report (April 1987 version). Holcomb Res. Inst. Rep. No. 75. Bulter Univ., Indiana.
- Mikyška, R. et al., 1968: Geobotanische Karte der Tschechoslowakei. 1. Tschechische Länder. In Vegetace ČSSR, Ser. A, Vol. 2. Academia, Prague, 204 pp.
- Moravec, J., 1971: A simple method for estimating homotoneity of sets of phytosociological relevés. Folia Geobot. Phytotax., 6, p. 147-170.

Odum, E.P., 1971: Fundamentals of ecology. W. B. Saunders, Philadelphia, 574 pp.

Orlóci, L., 1978: Multivariate analysis in vegetation research. W. Junk, The Hague et Boston, 451 pp.

Rothmaler, W. et al., 1982: Exkursionsflora für die Gebiete der DDR und der BRD. Vol. 4 (Kritischer Band). Ed. 5. Volk und Wissen Volkseigener Verlag, Berlin, 811 pp.

Simpson, E.H., 1949: Measurement of diversity. Nature, 163, p. 688.

Sörensen, T., 1948: A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. Biol. Skr., K. Danske Vidensk. Selsk., 5, p. 1 – 34.

Vesecký, A. et al., 1958: Climatological atlas of Czechoslovakia (in Czech). Prague, 11 pp, 100 sheets.

Webster, R., 1973: Automatic soil-boundary location from transect data. J. Int. Assoc. Math. Geology, 5, p. 27 – 37.

Whittaker, R.R. (ed.), 1973: Ordination and classification of vegetation. In Handbook of vegetation science, Vol. 5, W. Junk, The Hague, 737 pp.

Received 15. 4. 1991

### Matějka K.: Hierarchická semishluková analýza - nová metoda gradientní analýzy.

Nová metoda klasifikace předem sekvenčně uspořádaných vzorků byla vyvinuta a aplikována při hodnocení vegetačních transektů. Současně byly navrženy některé koeficienty vnitřní podobnosti (homotoneity) skupin vzorků a tyto byly zařazeny do a analyzovány v systému již existujících koeficientů. Nová metoda HSCA pracuje na aglomerativním principu. Může být použita k identifikování typických skupin vzorků na gradientech. Metoda byla aplikována při hodnocení čtyř krátkých kvadrátových transektů na pobřeží Kratochvílského rybníka u Netolic v jižních Čechách. Byla studována vegetační zonace na pobřeží v souvislosti s komplexním vlhkostním gradientem. Výsledky metody HSCA byly srovnány s výsledky klasických metod shlukové analýzy a s hodnocením provedeným na základě postupů Curyšsko-Montpellierské fytocenologické školy.