Annals of Arid Zone 25(2) : 111-131, 1986

A TAXON-FREE NUMERICAL APPROACH TO THE STUDY OF PLANT COMMUNITIES

L. ORLOCI AND S.L. STOFELLA

Department of Plant Sciences, University of Western Ontario
London, Canada N6A 5B7

ABSTRACT

In case of unknown flora, unshared species or clinal variation within species population, the conventional taxon-based schemes for community comparisons are not useful. Yet, comparisons are needed to order the communities on axes or to arrange them into groups. Taxon-free numerical methods to achieve these objectives are illustrated on the basis of data from forest, savanna and desert shrub vegetation.

DESCRIPTIVE SCHEMES

The usual methods of community analysis are species based. A species list is compiled for each stand and species performance (often cover/abundance) is estimated. The most often used scales are approximative, such as the Domin-Krajina scale and the Braun-Blanquet scale which are described by Becking (1957).

The main difficulty in dealing with species-based descriptions is the potential for indeterminacy when comparing distant communities. This indeterminacy is a consequence of communities not having a completely shared species list. To show the significance of this in an actual comparison, we consider two communities A, B compared by the function

$$S_{AB} = [X_{1A} X_{1B} + ... + X_{pA} X_{pB}]/(L_A L_B)$$

In this L denotes vector length, X represents an estimate of performance and p is the number of species in the combined list. The relationship measured is that of the cosine of the angle subtending vectors X_A and X_B. With s zero/non-zero matches..... i.e s zero products in the expression....an s-order indeterminacy will result. This implies that if one of the elements in the *i*th product $X_{iA} X_{iB}$ is zero, meaning that species *i* is absent from at least one of the lists, its presence in the second list will have no effect on the comparison. In other words, OX=O is satisfied for any X. When no species are shared (s=p), indeterminacy is complete.

Since indeterminacy is likely in species-based community comparisons, the question arises: "At what level will indeterminacies weaken a comparison to the point of uninterpretability of relationships?" We refer to discussion by Lambert and Dale

(1964) and Swan (1970) which considered the interpretation and the effect of zeros in the data. Experience suggests that the critical level may be reached at a relatively low incidence of abscences. Should this be concluded in a specific case, an alternative scheme will be required upon which to base community descriptions.

The use of higher taxonomic levels, such as genus, family and order, is one solution (van der Maarel, 1972). However, while the higher taxonomic levels increase coincidences, finer ecological information may be obscured. Furthermore, communities appearing to be similar on higher taxonomic level may differ significantly in ecological or evolutionary characteristics.

Humboldt's (1806) 'Hauptformen', Kerner von Marilaun's (1864) 'Grund form', Warming's (1884) and Raunkiar's (1907, 1937) 'Livsformer' or life-forms are classical examples of methods for character based, taxon-free community descriptions. Dansereau (1957), Mueller-Dombois and Ellenberg (1974), Ellenberg and Mueller-Dombois (1967), Barkman (1979) and Orloci and Kenkel (1985) described comprehensive schemes while Lacza and Fekete (1969) reviewed the early literature. The schemes differ in the arrangement of the character set.

THE CHARACTER SET

Selection

The character set has implications when attempting to explain biological phenomena. To provide information about community adaptation, suitable characters are those which reflect species adaptation to the enviroment. These characters are likely to be those of the non-reproductive structures (Lausi and Nimis, 1985; Barkman, 1979).

As regards the selection of character states, non-overlap and completeness are required. Non-overlap implies that the states must be recognizable without ambiguity. Completeess means that all cases are covered by the states. Some states may be left unspecified under the collective term 'else'. This implies a unique state when the characters are dichotomous. When the characters are multistate, 'else' will include all states not covered. 'Succulent stem' is an example for which 'else' is the unique alternative 'not succulent'. Should 'else' be the alternative to 'deciduous' and 'evergreen', it would not be unique, since it could mean either 'semideciduous' or 'withering', if the species is not 'leafless', which is another alternative.

It is customary to enumerate as many species as possible in a vegetation survey. Taxon-free surveys designed on this basis would require that as many characters as possible be enumerated. There are three points to be made to show that by increasing the number of characters, the benefits may not increase :

(1) Characters tend to be correlated (Dale 1968, Jardine and Sibson 1971). On this basis one may expect a small set of characters to convey the same information about the vegetation as a large set.

(2) Variation in some characters will be more related to the factors determining the distribution of the vegetation than variation in some other variables. If such a distinction can be made *a priori* character selection will be aided.

(3) The logistic difficulties of community surveys may disproportionately increase if the character set is large. This implies a need for careful selection of characters.

Critical character selection is not unique to the taxon-free scheme; it is sufficient just to consider the problem of constructing taxonomic keys for species identification. Although ecologists using these keys have the characters selected for them by others, the problem still exists.

Arrangement

Specific applications use different character arrangements. Knight and Loucks' (1969) work is a case example for a serially arranged character set. This is evident in their releve

Character	...	Th	Ch	Cr	H	...	V	A
Frequency	...	O	14	28	58	...	61	37

where life-form (Th, Ch, Cr, H) and blooming season (V, A) are the characters. The states overlap and the frequencies involve multiple scoring of the same plant type. For example, a plant type which contributes to frequency under 'Life-form' also contributes to frequency under 'Blooming season'. Clearly, a serial arrangement can be expanded only at the cost of multiple scoring.

The alternative is a nested arrangement (Orloci and Kenkel, 1985) in which an analogy to an inverted tree is appropriate. In this, levels represent characters and noda the character states. The number of branches radiating from a particular nodum at one level is equal to the number of states of the character at the next level.

Table 1 and Figure 1 give an example. In this there are three levels with twelve noda at the first, six at the second, and two at the third. The third level corresponds to a character with two states, the second to one with three states, and the last to another with three states. The number of noda on the first level is equal to 2x3x2. On the second level it is 2x3. In general, given m characters, the number of noda on the *i*th level is the product of the states of the characters from 1 to $m-i+1$.

The dendrogram (Fig. 1) has an associated character-score-matrix (CSM, Table 1). This CSM is a vegetation releve involving 4 character-set types (CSTs). The

CSTs may be viewed as OTUs (Sokal and Sneath, 1963) if we wished to force a taxonomic analogy. They represent distinct plant types as determined by the chosen character set. Not all CSTs need occur in the course of a vegetation survey and some may, in fact, be impossible. Still, all must be retained to serve as reference points in the comparisons.

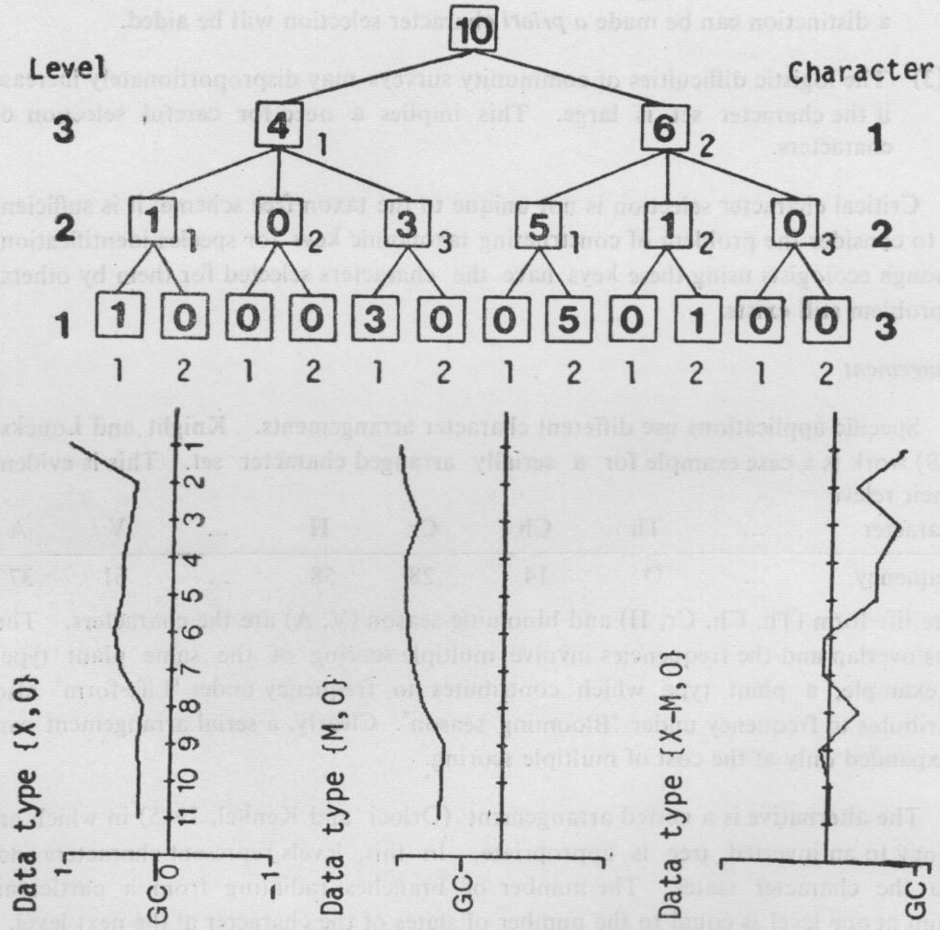


Fig. 1. Dendrogarm showing the relationship of levels, charaeters and nesting of character states based on data in table 1. Numbers inside boxes are cover/abundance estimates of the character set types. Numbers outside are character state labels.

Table 1. Score matrix for stem characters in a community. Numbers across to top identify character set types (CSTs), the numbers at the base are cover/abundance estimates (CAEs) of the CSTs, and the entries in the body of the table identify character states. The arrangement of the characters is nested combining two information sources : one distinguishing CSTs, and the other structuring the character set

* Character	CST 1	2	3	4
1. Green (1), not green (2)	1	2	2	1
2. Tall (1), medium (2), low (3)	1	1	2	3
3. Fleshy (1), not fleshy (2)	1	2	2	1
CAE	1	5	1	3

A dendrogram is a graphical representation of a CSM. The CSTs are mapped into the dendrogram as runs through stems from top down. A non-zero value at the base indicates that the associated CST has, in fact, occurred in the sample. For example, CST 3 in Table 1 defines the run selecting the second nodum on level 3, the fifth (3+2) on level 2, and the tenth ((3+1) x 2+2) on level 1. This simple arithmetic permits locating the relevant nodum for any CST on any specific level. Therefore, the handling of CSTs in comparisons does not require the drawing of the actual dendrogram.

How are CSMs constructed in the field ? An individual plant is selected and a CST determined for it by scoring character states (entries in the body of a CSM). Following this, an estimate of the cover/abundance of all individuals within the same CST is made. The CSM construction continues in like manner until all plant types are identified.

Sequence

Given p characters, the number of possible permutations is p! which can be a very large number, beyond the possibilities of actual enumeration. Yet if the arrangement is nested a single sequence must be chosen in order to present the characters for analysis. Can this in fact be logically done? What criteria should be used? A sequence may be tried based on the logical connectedness of the characters, but it is doubtful that this could be sustained through a large number of levels. Recognizing this, we see utility of basing the sequence on an appropriate quantitative criterion. Mutual information is one possibility (Feoli *et al.*, 1985).

Considering the effect of the sequence in a nested model, the anticipated consequences are similar to those of other partitioning schemes which use, for instance, orthogonal functions (*e.g.* Rao, 1952; Orloci, 1978). The idea is that each character is replaceable by a function specific to the character, but independent from the functions already derived for the preceding characters. Similar partitioning is also used in some pattern analyses (Greig-Smith, 1952; Noy-Meir and Anderson, 1971), but for these the sequence is not a problem because it is fixed by the model.

CAE ARRAYS AND THE WILLIAMS-DALE PARTITION

Considering Fig. 1, the cover/abundance estimates (CAEs) on the first level are the numerical values in

$$(X,O)_1=(1\ 0\ 0\ 0\ 3\ 0\ 0\ 5\ 0\ 1\ 0\ 0)$$

By pooling adjacent scores we obtain the arrays on successive levels.

$$(X,O)_2=(1\ 0\ 3\ 5\ 1\ 0)$$

$$(X,O)_3=(4\ 6)$$

One may elect to apply a data partition (Williams and Dale, 1962) to isolate two data components on level 1,

$$(X,O)_1=(M,O)_1+(X-M,O)_1$$

(M,O) forms the purely 'qualitative' component and (X-M,O) the purely 'quantitative' component of the data. M represents the mean of the non-zero values in $(X,O)_1$. The values in X-M are deviations from the common mean. For example, given

$$(M,O)_1=(2.5\ 0\ 0\ 0\ 2.5\ 0\ 0\ 2.5\ 0\ 2.5\ 0\ 0)$$
$$(M,O)_2=(2.5\ 0\ 2.5\ 2.5\ 2.5\ 0)$$
$$(M,O)_3=(5\ 5)$$

we have

$$(X-M,O)_1=(-1.5\ 0\ 0\ 0\ 0.5\ 0\ 0\ 2.5\ 0\ 1.5\ 0\ 0)$$
$$(X-M,O)_2=(-1.5\ 0\ 0.5 2.5\ -1.5\ 0)$$
$$(X-M,0)_3=(-1\ +1)$$

Methods suitable for the analysis of data of this kind have been described (Orloci and Kenkel, 1985; Orloci, *et al.*, 1986; Orloci and Orloci, 1986) and will be discussed next.

THE ANALYSIS

The two schemes of character arrangement raise different analytical problems. In a serial arrangement, characters become 'quasi taxa' and the analysis proceeds along conventional lines (Orloci, 1978; Knight and Loucks, 1969). In the nested arrangement, a different methodology is required since the characters are parts in two dimensional reference scheme where the CSTs are 'quasi taxa' with actual states depending on the hierarchical level.

Pattern similarity of CSMs

We have found that

$$r_{AB} = Q_{AB} / [Q_{AA} Q_{BB}]^{\frac{1}{2}}$$

for data type (X,O) and

$$r_{AB/d} = Q_{AB/d} / [Q_{AA/d} Q_{BB/d}]^{\frac{1}{2}}$$

$$r_{iAB/d} = Q_{iAB/d} / [Q_{iAA/d} Q_{iBB/d}]^{\frac{1}{2}}$$

involving the data (d) arrays (M,O) and (X-M,O) are suitable functions to measure pattern similarity. Symbols A,B designate CSMs and d is the data type. $r_{AB}/_d$ is a global correlation and $r_{iAB}/_d$ is a partial one specific to level i and data type d. The products are additive over levels such that

$$Q_{AB/d} = Q_{1AB/d} + Q_{2AB/d} + ... + Q_{mAB/d}$$

where m is the number of levels (characters) and the Qs are the products

$$Q_{AB}/_d = \Sigma X_{1jA}/_d X_{1iB/d} - (1/k_1) \Sigma X_{1jA}/_d \Sigma X_{1jB}/_d$$

and

$$Q_{iAB}/_d = (1/n_i) \Sigma X_{ijA}/_d X_{ijB}/_d - (1/n_{(i+1)} \Sigma X_{(i+1)jA}/_d X_{(i+1)jB}/_d$$

The summations are taken over all stems from $j=1$ to k_i or k_{i+1} at levels i and i+1. Since the $Q_{iAB}/_d$ are additive, a partial correlation is defined for any combination of levels using the pooled Qs (but not the rs). Should $Q_{iAB}/_d$ be defined as

$$Q_{iAB}/_d = (1/n_i) \Sigma X_{ijA}/_d X_{ijB}/_d - (1/k_1) \Sigma X_{1jA}/_d \Sigma X_{1jB}/_d$$

$r_{iAB}/_d$ would be a measure of the nominal correlation of CSMs A,B for character $m-i+1$. The following definitions apply :

$X_{ijA}/_d$ -- CAE of CST j in CSM A, data type d

k_i -- number of noda on level i

n_t -- number of stems fused in a nodum on level i

$Q_{AB}/_d$ -- Produce for CSMs A and B, data type d

$Q_{iAB}/_d$ -- Product specific to level i, CSMs A and B, data type d

Profiles of partial correlations

The profile [$r_{1AB}/_d$ $r_{2AB}/_d$ ... $r_{mAB}/_d$] is characteristic for the A,B comparison for a given data type d. A polygon through the m points – called a 'partial correlation profile' – enables the examination of the contribution of characters individually to pattern similarity regarding sense, intensity, and type. The sense may be negative or positive if not zero. A zero value implies pattern indifference.

The correlation values may be deemed significant or trivial, depending upon whether they lie outside or inside given confidence limits. These limits can be determined experimentally in Monte Carlo simulation (Orloci and Kenkel, 1985). Significance is not conditional on a correlation's value being close to the limit -1 or $+1$. In this sense, the absolute magnitude of the correlation is also important.

Depending on the data type, the correlations depict either the purely qualitative or quantitative aspect of the relationships. In ths case of data type (M,O), correlations respond to presence/absence, irrespective of quantity. In the case of data type (X–M,O), correlations measure quantitative relatilonships which are superimposed on joint presence.

Displaying correlations in ordination

We are using profiles to display partial correlations in paired comparisons of CSMs. To display an entire correlation matrix, we fall back on ordination. Any of the matrices below may serve as the basis of an ordination :

(i) $R = [\, r_{AB};\ A,B = 1, \ldots, s \,]$
(ii) $R_d = [\, r_{AB/d};\ A,B = 1, \ldots ; s;\ d = (M,O), (X\text{–}M,O) \,]$
(iii) $R_{i/d} = [\, r_{iAB/d};\ i = 1, \ldots, M;\ A,B = 1, \ldots, s; d = (M,O), (X\text{–}M,O) \,]$

These are s x s correlation matrices : global, irrespective of data type (R), global, specific to data type (R_d), partial, specific to hierarchical level and dependent on data type ($R_{i/d}$).

To perform an ordination of the CSMs, eigenanalysis is one option (Orloci, 1978). Noting that the correlations are centered for axes, representing the CSMs, within which the points are the CSTs, and further noting that axes are preferred which order the CSMs as points, we base eigenanalysis on a distance matrix derived from the correlation matrix. The distances are defined by either of the following formulae :

(i) $d_{AB} = [\, 2\,(1-r_{AB}) \,]^{1/2}$
(ii) $d_{AB/d} = [\, 2\,(1-r_{AB/d}) \,]^{1/2}$
(iii) $d_{iAB/d} = [\, 2\,(1-r_{iAB/d}) \,]^{1/2}$

These are centreed chord distances with limits 0 to 2. Considering further that sums of squares and products are partitioned, the distances are defined for pooled correlations according to

$$d(_{U+\cdots+UK})_{AB/d} = \sqrt{[2(1 - r(_{U1+\cdots+Uk})_{AB/d}]}$$

in which r is a pooled partial correlation,

$$r(_{U1}+\ldots+_{Uk})_{AB/d} = [Q_{U1AB/d}+\ldots+Q_{Uk1AB/d}] / \sqrt{[Q_{U1AA}+\ldots+Q_{UkAA})(Q_{U1BB}+\ldots+Q_{UkBB})]}$$

The k levels may be contiguous, but this is not a requirement. To obtain ordination coordinates comparable to the component scores in PCA (Orloci 1978), we analyze the Q matrix derived from the D matrix by one of the transformations

(i) $q_{AB} = -0.5[d^2_{AB} - d^2_A - d^2_B + d^2]$
(ii) $q_{AB/d} = -0.5[d^2_{AB/d} - d^2_{A/d} - d^2_{B/d} + d^2_{/d}]$
(iii) $q_{iAB/d} = -0.5[d^2_{iAB/d} - d^2_{iA/d} - d^2_{iB/d} + d^2_{i/d}]$

The d^2 terms represent respectively the squared distance of CSMs A,B, mean squared distance of CSM A from the other CSMs, mean squared distance of CSM B from the other CSMs, and the mean of all squared distance.

Displaying correlations in cluster analysis

Considering two groups of CSMs, U and V, the squared centroid distance of the two groups is

$$d^2_{UV} = [(S_U + S_V)/(S_U S_V)] Q_{UV}$$

in which S_U and S_V are the number of CSMs and the quantity

$$Q_{UV} = Q_{U+V} - Q_U - Q_V$$

is the amount of increase in the sum of squares upon fusion of groups U and V. The quantity

$$Q_U = (1/s_U)\Sigma d^2_{gh/U}; g = 1, \ldots, s_U - 1; h = g, \ldots, s_U$$

is the sum of squareds distances to S_U points from the centroid of group U. Symbols g and h identify CSMs. Clustering algorithms are available for d_{UV} as well as for Q_{UV} (see Orloci and Kenkel, 1985 and references therein) which derive groups and display their hierarchical relationships.

Having derived q groups, the clustering efficiency achieved is the ratio E :

Source	Q	E
Between q groups	$Q_{1 \ldots q}$	$Q_{1 \ldots k}/Q_{1 + \ldots + q}$
Within q groups	$Q_1 + \ldots + Q_q$	
Total	$Q_{1 + \ldots + q}$	

Environmental connections

The CSMs presumably form environmental series. To test that such a series actually exists and to reveal its characteristics graphical or analytical methods may be used. The graphical methods present plots of the environmental data within ordination graphs. The analytical methods include canonical analysis, concentration analysis, regression analysis and possibly other techniques (Williams 1952, Orloci 1978, 1981. Feoli and Orloci 1979, 1985, Gittins 1985).

RESULTS -- AN EXAMPLE

Field data

The data set includes 18 releves from Texas, and one from each of Alabama, Mississippi and Louisiana. The Texas sites are shown in Fig. 2 with temperature and precipitation isolines superimposed. Localities and environmental data are given in table 2. The character set is listed in table 3. The releves describe an early spring aspect and it is, therefore, likely that the results are dominated by the characters of woody and graminoid forms. We do not present the entire data set here, but have

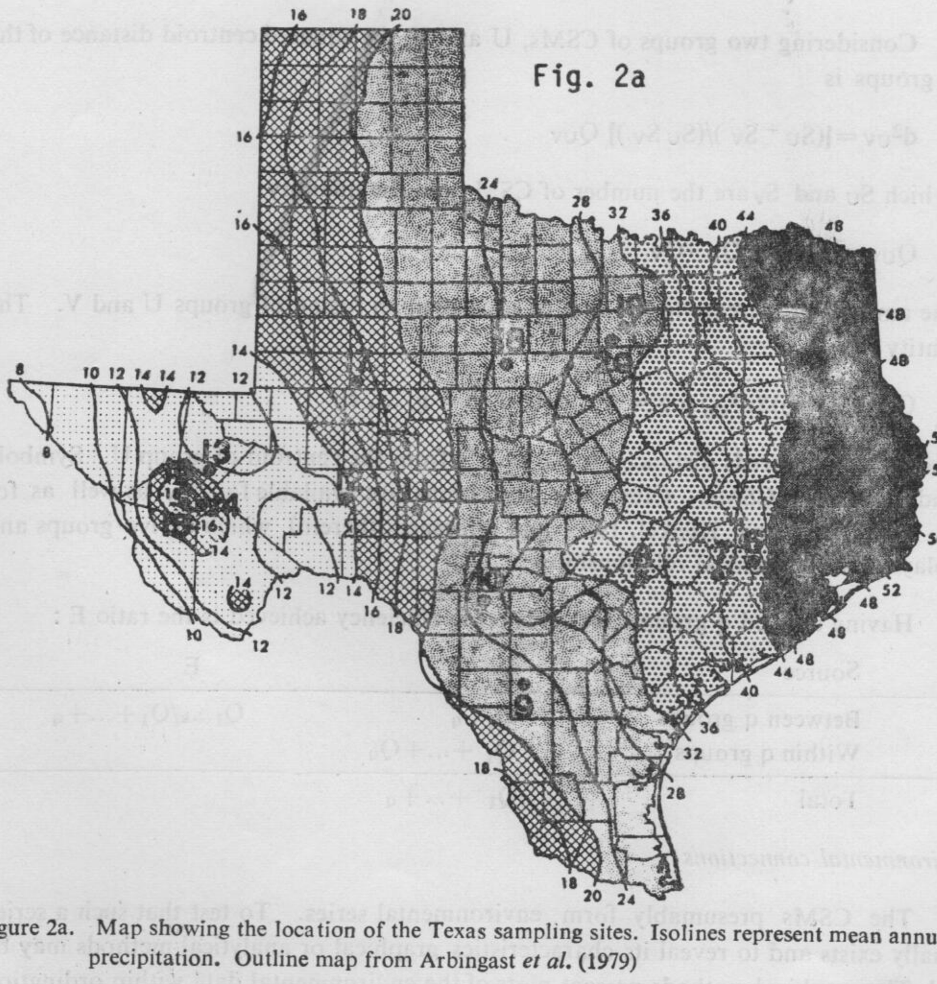


Figure 2a. Map showing the location of the Texas sampling sites. Isolines represent mean annual precipitation. Outline map from Arbingast *et al.* (1979)

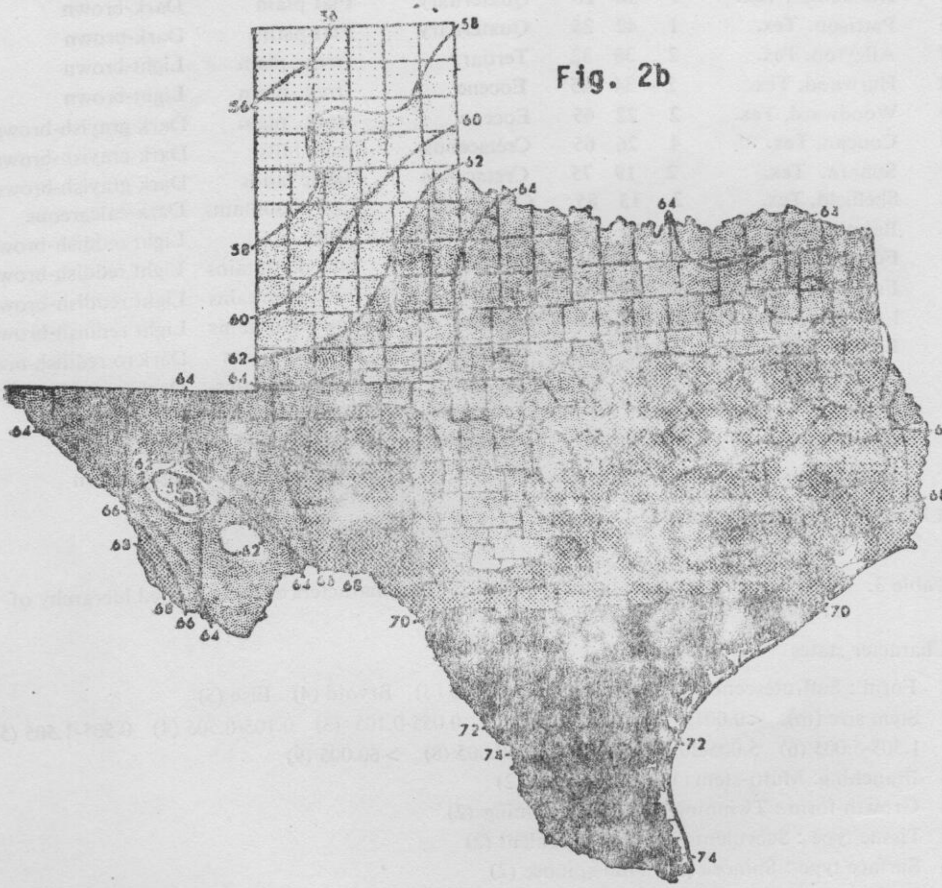


Figure 2b. Map showing the location of the Texas sampling sites. Isolines represent mean annual temperature. Outline map from Arbingast *et al.* (1979)

Table 2. Regional and zonal features of the sampling sites. Terminology is similar to that used by Arbingast *et al.* (1979)

Plot	Location	E	P	D	Surf. geology	Land form	Zonal soil
1	Bon Secour, Al.	1	64	5	Quaternary	Flat plain	Reddish-brown
2	Kiln, Miss.	1	64	10	Quaternary	Flat plain	Dark-brown
3	Jennings, Lous.	1	57	15	Quaternary	Flat plain	Reddish-brown
4	Orange, Tex.	1	54	15	Quaternary	Flat plain	Dark-brown
5	Hankamer, Tex.	1	50	20	Quaternary	Flat plain	Dark-brown
6	Pattison, Tex.	1	42	25	Quaternary	Flat plain	Dark-brown
7	Alleyton, Tex.	2	38	32	Tertiary	Irreg. plain	Light-brown
8	Harwood, Tex.	2	34	45	Eocene	Irreg. plain	Light-brown
9	Woodward, Tex.	2	22	65	Eocene	Irreg. plain	Dark grayish-brown
10	Concan, Tex.	4	26	65	Cretaceous	High hills	Dark grayish-brown
11	Sonora, Tex.	2	19	75	Cretaceous	Table lands	Dark grayish-brown
12	Sheffield, Tex.	2	13	85	Cretaceous	Low mountains	Dark-calcareous
13	Balmorhea, Tex.	2	14	85	Quaternary	Plain	Light reddish-brown
14	Fort Davis, Tex.	4	16	85	Igneous	Low mountains	Light reddish-brown
15	Fort Davis, Tex.	5	18	85	Igneous	Low mountains	Light reddish-brown
16	Fort Davis, Tex.	5	18	85	Igneous	Low mountains	Light reddish-brown
17	Lomax, Tex.	2	17	75	Tertiary	Smooth plains	Dark to reddish-brown
18	Clyde, Tex.	3	26	65	Permian	Irregular plains	Dark to reddish-brown
19	Aledo, Tex.	3	34	55	Cretaceous	Table lands	Dark
20	Weatherford, Tex.	3	30	55	Cretaceous	Table lands	Dark-brown
21	Texarkana, Tex.	2	48	15	Eocene	Table lands	Light-brown

E-Elevation P-Precipitation D-Drought

Table 3. The character set used in the example. The characters define a nested hierarchy of 11 levels

* Character states

1 Form : Suffrutescent (1) Graminoid (2) Forb (3) Bryoid (4) Else (5)
2 Stem size (m): <0.0015 (1) 0.015-0.055 (2) 0.055-0.105 (3) 0.105-0.505 (4) 0.505-1.505 (5) 1.505-5.005 (6) 5.005-20.005 (7) 20.005 60.005 (8) >60.005 (9)
3 Branching: Multi-stem (1) Mono-stem (2)
4 Growth-form : Twinning (1) Non-twinning (2)
5 Tissue type : Succulent (1) Non-succulent (2)
6 Surface type : Spinose (1) Non-spinose (2)
7 Leaf length (cm) : <1.55 (1) 1.55-5.55 (2) 5.55-10.55 (3) 10.55-50.55 (4) 50.55-100.55 (5)
8 Leaf width (mm): <1.55 (1) 1.55-5.55 (2) 5.55-10.55 (3) 10.55-50.55 (4) 50.55-100.55 (5)
9 Leaf type : Simple (1) Simple cleaved (2) Compound leaf (3)
10 Leaf quality : Filmy (1) Leathery (2) Fleshy (3) Else (4)
11 Leaf surface : Glaucous (1) Hairy (2) Prickly (3)

included CSM # 13 as an example (Table 4). The total data set contains 21 such matrices with a varying number of CSTs (Table 5).

Table 4. Releve 13 given by its character score matrix. This matrix illustrates the data upon which the example is based. The eight columns are CSTs and the 11 rows are characters. The entries in the table are character states, except in the last row where cover/abundance values are given. CST 1 is typified by *Larrea tridentata* DC. The other CSTs depict yucca, cacti, suffrutescent shrubs, and grasses.

CSM 13								
CHAR 1	5	5	5	1	3	2	2	2
CHAR 2	6	6	6	6	2	2	2	2
CHAR 3	1	1	1	1	2	1	2	2
CHAR 4	2	2	2	2	2	2	2	2
CHAR 5	2	2	1	2	2	2	2	2
CHAR 6	2	2	1	2	2	2	2	2
CHAR 7	1	2	2	1	1	2	2	1
CHAR 8	2	1	2	2	1	1	1	1
CHAR 9	1	1	1	1	1	1	1	1
CHAR 10	2	4	4	4	4	4	4	4
CHAR 11	1	1	1	1	1	1	1	1
CAE	3.000	0.100	0.100	2.000	2.000	1.000	2.000	1.000

Table 5. Distribution of sampling sites among the vegetation regions as described by Arbingast *et al.* (1979), except sites 1, 2, 3. The number of character set types (CSTs) per site is also specified. The community in any given site may floristically differ from that of the zonal vegetation owing to local variations in climate, soil, and topography.

Sampling site	Vegetation zone	Number of CSTs
1	Oak-Pine-Magnolia Forest	9
2	Oak-Pine-Magnolia Forest	8
3	Oak-Hickory-Pine Forest	10
4	Oak-Hickory-Pine Forest	10
5	Oak-Hickory-Pine Forest	11
6	Coastal Prairie	5
7	Oak-Hickory Forest	10
8	Oak Savanna	13
9	Mesquite-Chaparral Savanna	12
10	Juniper-Oak-Mesquite Savanna	9
11	Juniper-Oak-Mesquite Savanna	13
12	Desert Shrub Savanna	11
13	Desert Shrub Savanna	8
14	Desert Shrub Savanna	8
15	Desert Shrub Savanna	10
16	Desert Shrub Savanna	10
17	Mesquite Savanna	8
18	Mesquite Savanna	9
19	Blackland Prairie	4
20	Oak Forest and Prairies	12
21	Oak-Hickory-Pine Forest	13

The sampled sites fall along a major climatic gradient from east to west. Vegetational variation, however, has added complexity attributable to responses to elevation, soils, topography, and other local influences such as past forestry, agricultural and range use. The vegetation regions described from Texas (Arbingast *et al.*, 1979) range from the Oak-Pine forests of the east, through the Oak - Shrub savanna of the central regions, and into the Desert Shrub types of the west. Where soil conditions are suitable, grasslands occur under a moderate precipitation regime. The two easternmost releves are from Oak-Pine-Magnolia forests and the third (outside Texas) from an Oak-Hickory-Pine type. Table 5 shows the distribution of the sampling sites among the vegetation regions.

The analyses which we describe in this section produced large volumes of results for three data types explained in the preceding sections. We present here only selected cases as examples, but are willing to make available the entire set of results for inspection to enquirers.

Partial, nominal and global correlations, chord distance

A product moment is computed for each pair of CSTs on each character level for each data rype. This gives a total of (21 x 20/2) x 11x3 = 6930 product moment values. We reproduce only some partial correlation profiles as examples (Fig. 3) and give in table 6 the distance matrix from global correlations for data type (M-0).

Ordination of the global chord distance

Twenty eigenvalues were obtained for data type (M,0) and (X,M,0). The first three eigenvectors are displayed in stereograms (Figure 4). The total variation accounted for by the three eigenvectors is about 28 % for each data type. The 'horseshoe' in Figure 4a seemed to be a response to an East-West precipitation gradient and in Figure 4b to an elevation gradient from highlands to lowlands.

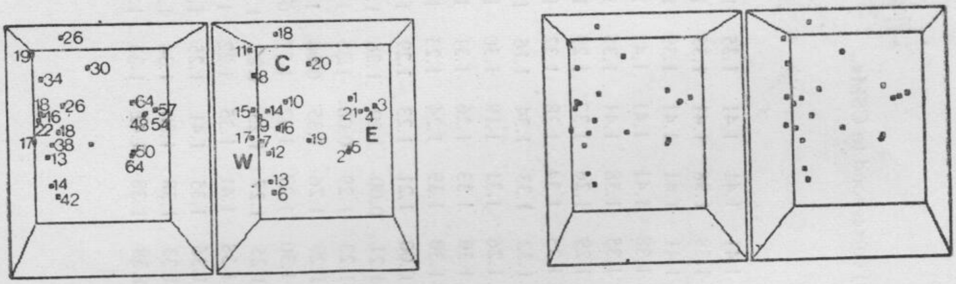
Cluster analysis of the global chord distance

The dendrograms and the analysis of sum-of-squares table for three groups are presented in Figure 5 and Table 7. Differences in group contents among the data types, and the similarity of the efficiency values (E), are noted. The groups recognized on the basis of the (M, 0) data indicate precipitation classes (western, central and eastern; Figure 5a). The groups from the (X-M,0) data are related to elevation classes (hills, uplands and lowlands; Figure 5b).

Computations

The methods are programmed. See the Appendix for brief descriptions.

(a)



(b)

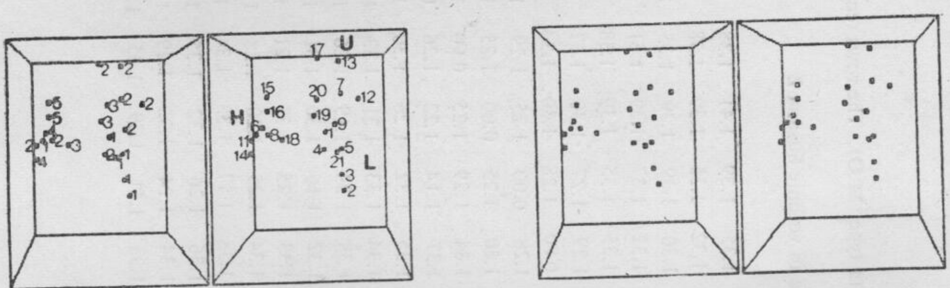


Figure 4. Stereograms for data type (M.O) (a) and (X-M.O) (b). Small squares in the diagram indicate releves. Letters W,C,E, indicate the precipitation gradient through the western, central and eastern regions. Letters H,U,L identify the major orographic features (Hills, Uplands, Lowlands). The numbers in the right stereograms are releve labels. In the lcft stereograms the numbers are precipitation (a) and elevation values (b).

Table 6. Chord distance matrix for data type (M.O). The rows (columns) correspond to CSMs.
Data Type 2
Pooled chord distances of this data type in volume Susana 2
File Dist/112

0.00	1.32	1.26	1.32	1.35	1.41	1.35	1.39	1.41	1.38	1.41	1.41	1.41	1.41	1.35	1.39	1.41	1.35	1.41	1.30	1.19
1.32	0.00	1.24	1.29	1.18	1.33	1.32	1.35	1.39	1.38	1.39	1.38	1.38	1.41	1.35	1.36	1.38	1.38	1.36	1.39	1.26
1.26	1.24	0.00	1.11	1.23	1.41	1.30	1.39	1.39	1.35	1.41	1.41	1.41	1.41	1.39	1.41	1.41	1.36	1.41	1.26	1.19
1.32	1.29	1.11	0.00	1.25	1.41	1.32	1.37	1.39	1.31	1.41	1.38	1.41	1.41	1.41	1.39	1.41	1.38	1.41	1.28	1.20
1.35	1.18	1.23	1.25	0.00	1.33	1.35	1.35	1.39	1.38	1.39	1.35	1.38	1.41	1.35	1.36	1.34	1.35	1.36	1.39	1.19
1.41	1.33	1.41	1.41	1.33	0.00	1.29	1.27	1.30	1.37	1.39	1.29	1.26	1.37	1.29	1.30	1.26	1.41	1.27	1.38	1.34
1.35	1.32	1.30	1.32	1.35	1.29	0.00	1.28	1.30	1.34	1.37	1.35	1.34	1.38	1.32	1.33	1.34	1.38	1.36	1.34	1.33
1.39	1.35	1.39	1.37	1.35	1.27	1.28	0.00	1.25	1.29	1.14	1.32	1.33	1.34	1.16	1.25	1.25	1.17	1.26	1.24	1.31
1.41	1.39	1.39	1.39	1.39	1.30	1.30	1.25	0.00	1.25	1.22	1.26	1.31	1.19	1.30	1.31	1.20	1.33	1.32	1.32	1.39
1.38	1.38	1.35	1.31	1.38	1.37	1.34	1.29	1.25	0.00	1.26	1.30	1.33	1.26	1.31	1.21	1.28	1.34	1.35	1.29	1.35
1.41	1.39	1.41	1.41	1.39	1.39	1.37	1.14	1.22	1.26	0.00	1.30	1.39	1.24	1.23	1.33	1.25	1.17	1.34	1.26	1.37
1.41	1.38	1.41	1.38	1.35	1.29	1.35	1.32	1.26	1.30	1.30	0.00	1.21	1.23	1.29	1.30	1.25	1.38	1.36	1.33	1.38
1.41	1.38	1.41	1.41	1.38	1.26	1.34	1.33	1.31	1.33	1.39	1.21	0.00	1.29	1.26	1.27	1.27	1.41	1.35	1.38	1.38
1.41	1.41	1.41	1.41	1.41	1.37	1.38	1.34	1.19	1.26	1.24	1.23	1.29	0.00	1.35	1.32	1.25	1.35	1.41	1.30	1.41
1.35	1.35	1.39	1.41	1.35	1.29	1.32	1.16	1.30	1.31	1.23	1.29	1.26	1.35	0.00	1.27	1.22	1.22	1.25	1.34	1.33
1.39	1.36	1.41	1.39	1.36	1.30	1.33	1.25	1.31	1.21	1.33	1.30	1.27	1.32	1.72	0.00	1.35	1.33	1.37	1.32	1.33
1.41	1.38	1.41	1.41	1.34	1.26	1.34	1.25	1.20	1.28	1.25	1.25	1.27	1.25	1.22	1.35	0.00	1.34	1.28	1.35	1.38
1.35	1.38	1.36	1.38	1.35	1.41	1.38	1.17	1.33	1.34	1.17	1.38	1.41	1.35	1.22	1.33	1.34	0.00	1.30	1.27	1.32
1.41	1.36	1.41	1.41	1.36	1.27	1.36	1.26	1.32	1.35	1.34	1.36	1.35	1.41	1.25	1.37	1.28	1.30	0.00	1.41	1.36
1.30	1.39	1.26	1.28	1.39	1.38	1.34	1.24	1.32	1.29	1.26	1.33	1.38	1.30	1.34	1.32	1.35	1.27	1.41	0.00	1.29
1.19	1.26	1.19	1.20	1.19	1.34	1.33	1.31	1.39	1.35	1.37	1.38	1.38	1.41	1.33	1.33	1.38	1.32	1.36	1.29	0.00

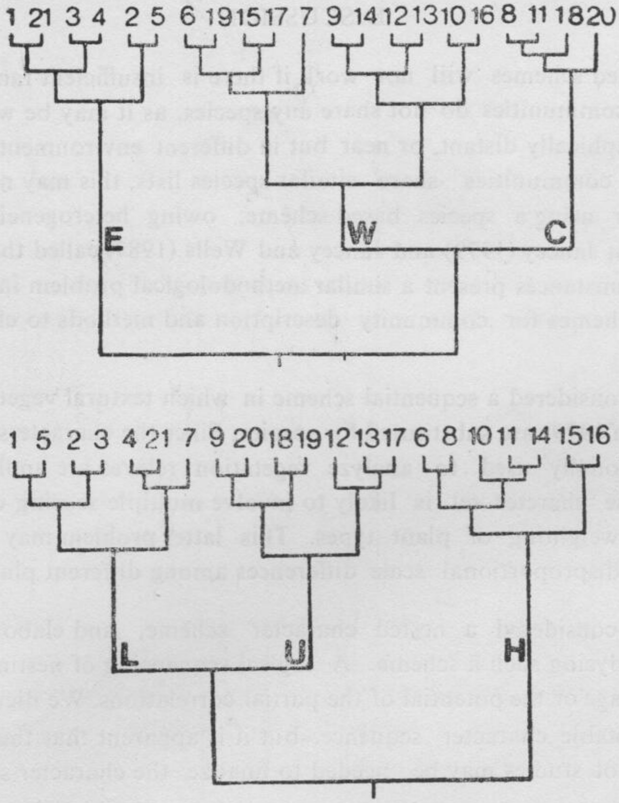


Figure 5. Dendrograms showing the hierarchical relationship of releves for two data types. Numbers across the top identify releves. Clustering efficiency values at the three-group level are given in Table 7. The groups are correlated with increasing precipitation (W, C,E in dendrogram a) and increasing elevation (L,U,H in dendrogram b).

a. Data type (M,O) b. Data type (X-M,O)

Table 7. Determination of clustering efficiency for two data types as function of the sums of squares for the three groups in Figures 5a, 5b. Variable E has potential values from zero to one. The values obtained indicate a relatively high level of heterogeneity within the groups.

a. Data type (M, O)

Source	Sum of Squares	Clustering efficiency
Between 3 groups	$Q_B = 3.19$	$E = Q_B / Q_T = 0.18$
Within 3 groups	$Q_W = 3.84 + 8.44 + 2.19 = 14.47$	
Total	$Q_T = 17.66$	

b. Data type (X-M, O)

Source	Sum of squares	Clustering efficiency
Between 3 groups	$Q_B = 3.34$	$E = Q_B / Q_T = 0.23$
Within 3 groups	$Q_W = 5.424 + 5.851 + 4.72 = 15.99$	
Total	$Q_T = 19.43$	

DISCUSSION

Taxon based schemes will not work if there is insufficient familiarity with the flora, or if two communities do not share any species, as it may be when the communities are geographically distant, or near but in different environments. Furthermore, even when two communities share similar species lists, this may not be a sufficient justification for using a species based scheme; owing heterogeneity within species populations what Jancey (1979) and Jancey and Wells (1987) called the 'locality' problem. These circumstances present a similar methodological problem in finding alternative character schemes for community description and methods to effectively analyze these schemes.

We have considered a sequential scheme in which textural vegetation characters (*sensu*, Barkman 1979) are substituted for species. Since the characters are 'quasi taxa', methods traditionally used to analyze vegetation releves are applicable; however, expansion of the charcter set is likely to involve multiple scoring which implies *de facto* unequal weighting of plant types. This latter problem may also occur as a consequence of disproportional scale differences among different plant types.

We also considered a nested character scheme, and elaborated on suitable methods for analysing such a scheme. A logical sequencing of nestings is required to take full advantage of the potential of the partial correlations. We dicussed criteria for finding an acceptable character sequence, but it is apparent that the problem is very complex and pilot studies may be needed to finalize the character set (cf. Lausi and Nimis, 1985).

Having decided on a nested character scheme, the data at the base of this scheme describe the character set types as a set of states togetherwith estimates of the cover/ abundance. However, the cover/ abundance values confound qualitative and quantitative information. To isolate these two sources as different data types, the Williams-Dale data partition is used. Since information lies in the mapping of the observed character set types in the scheme with either of the data types, the level of nesting is important. The information most relevant to us here is the comparative pattern of these mappings among the CSMs. defined in terms of partial and global correlations and distances. The sense of the partial correlations determine whether similarity or dissimilarity is expressed, and their magnitude describes the strength of the relationship. Note that these properties may change with data type.

We mentioned at the beginning of the paper the problem of indeterminacy in taxon based schemes. With the nested character scheme this problem can be resolved. This is an advantage which has to be contrasted with the diffculties related to character sequence and the potential of excessive random variation in the data. However, just how heavily should these weigh in the evaluation of the taxon-free approach has yet to be seen.

ACKNOWLEDGEMENTS

The first author (L. Orloci) has been recipient of a N.S.E.R.C of Canada operating grant during tenure of this study.

REFERENCES

Arbingast. S. A., Kennamer, L. G. Ryan, R.H. Buchanan, J.R. Hezlep, W.L. Ellis. L.T. Jordan, T.G. Granger C.T. and Zlatkovich C.P. 1979. *Atlas of Texas.* The University of Texas, Austin, Texas, U.S.A 78712

Barkman, J.J, 1979. The investigation of vegetation texture and structure. p. 129-160. *In* J.A. Werger (ed.) *The Study of Vegetation.* Dr. W. Junk, The Hague.

Becking, R.W. 1957. The Zurich-Montpellier School of Phytosociology. *Bot. Rev.* 23: 411-488.

Dale, M.B. 1968. On property structure, numerical taxonomy and data handling. p. 185-197. *In* V.H. Heywood (ed.), *Modern methods in Plant Taxonomy.* Academic Press, London.

Dansereau, P. 1957. *Biogeography. An Ecological Perspective.* The Ronald Press, New York.

Ellenberg, H. and Mueller-Dombois, D. 1967. A key to Raunkiaer plant life forms with revised subdivisions. *Ber. Geobot. Inst. ETH, Stift. Rubel, Zurich* 37:56-73.

Feoli, E. and L. Orloci, 1985. Species dispersion profiles of anthropogenic grasslands in the Italian Eastern Pre-Alps. *Vegetatio* 60:113-118.

Feoli, E., Lagonegro, M. and Orloci, L. 1985. *Information Analysis of Vegetation Data* Dr. W. Junk, The Hague.

Gittins, R. 1985. *Canonical Analysis. A review with Applications in Ecology.* Spring-Verlag, New York.

Greig-Smith, P. 1952. The use of random and contiguous quadrats in the study of the structure of plant communities *Ann. Bot. Lond.* 16:293-316.

Humboldt, A. van. 1806. *Ideen zu einer Physiognomik der Gewachsen* Cotta, Stuttgart.

Jancey, R. C. 1979. Species ordering on a variance criterion. *Vegetatio* 39:59-63.

Jancey, R.C. and Wells, T.C. 1987. Locality theory-the phenomenon and its significance. *Coenoses* 2: (in press).

Jardine, N. and R. Sibson 1971. *Mathematical Taxonomy.* Wiley, New York.

Kerner von Marilaun, A. 1864. *Das Pflanzenlaben der Donaulander.* (2nd ed. by F. Vierhapper, 1929.Univ.-Verlag Wagner, Innbruck).

Knight. D.H. and O. L. Loucks. 1969. A quantitative analysis of Wisconsin forest vegetation on the basis of plant function and gross morphology *Ecology* 50:219-234.

Lacza, J. Sz. and G. Fekete. 1969. A survey of plant life-form systems and the respective research approaches. I. *Annales Historico-Naturales Musei Nationalis Hungarici, Pars Botanica* 61:129-139.

Lambert. J. M. and Dale, M. B. 1964. The use of statistics in phytosociology. *Adv. Ecol. Res.* 2:59-29.

Lausi, D. and Nimis, P.L. 1985. The study of convergent evolution in plants and plant communities. A quantitative approach. *Abstracta Botanica* 9:67-77.

Maarel, E. van der. 1972. Ordination of plant communities on the basis of their plant genus, family and other relationships. p, 183-206. *In* E. van der Maarel and R. Tuxen (eds.) *Grundfragen und Methoden in der Pflanzensociologie*. Dr. W. Junk, The Hague.

Mueller-Dombois, D. and Ellenberg, H. 1974. *Aims and Methods of* vegetation *Ecology*. Wiley, New York.

Noy-Meir, I. and Anderson, D.J. 1971. Mulltiple pattern analysis, or multiscale ordination: towards a vegetation hologram ? p. 207-231. *In* G. P. Patil, E.C. Pielou and W.E. Waters (eds.) *Statistical Ecology*. Vol. 3. *Many Species Populations, Ecosytems and Systems Ecology*. Pennsylvania State University Press, University Park, Pennsylvania, U.S.A.

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

Orloci, L. 1981. Probing time series vegetation data for evidence of succession. *Vegetatio* 46:31-35.

Orloci, L., Feoli, E. Lusi, D. and Nimis, P. 1986. Estimation of character structure convergence (divergence) in plant communities: a nested hierarchical model. *Coenoses* 1:11-20,

Orloci, L. and Kenkel, 1985. *Introductian to Data Analysis with Examples from Population and Community Ecology*. ICPC, Fairland, MD., U.S.A.

Orloci, L. and Orloci, 1986. Comparison of communities without the use of species: moddel and example. *Proceedings of the 1985 International Symposium on Climate and life Forms in Mediterranean Type Vegetation*. Unversity of Rome. (In press).

Rao, C.R. 1952. *Advanced statistical Methods in Biometroc research*. Wiley, New York.

Raunkiaer, C. 1907. *Planterigets Livsformer og deres Betydning for Geografien*. Munksgaar, Kobenhavn. 1937. *Plant Life Forms*. Clarendon, Oxford.

Sokal, R.R. and Sneath, P.H.A. 1963. *Principles of Numerical Taxonomy*. Freeman, San Francisco.

Swan, J.M.A. 1970. An examination of some ordination problems by use of simulated vegetational data. *Ecology* 51:19-102.

Warming, E. 1884. *Om Skudbygning, Overvintring og Foryngelse*. Festskr. Naturh. Foren. Kjobenhaven.

Williams, E.J. 1952. Use of scores for the analysis of association in contingency tables. *Biometrika* 39:274-289.

Williams, W. T. and Dale, 1962. Partition correlation matrices for heterogeneous quantitative data *Nature* 196:602.

APPENDIX

Computational Implications

The analyses are performed by automatic programes available in IBM PC BASIS Macintosh MS BASIC and Applesoft BASIC. Programe MAKE TEXT creates the text file which contains the CSMs and CAEs for CSTs. This file is input for programes CSA/1 which sets up the data for analysis by program CSA/2. This computes nominal and partial correlations between pairs of CSMs, draws partial correlation profiles, and performs Monte Carlo simulation in which confidence limits are set. The same program creates files GRAPH, PROD and COR which hold respectively the confidence limits, specific product moments (level-by-level) and specific correlations for the three data types. The confidence limits are drawn in program CSA/SORT. The limits are not included in the present example.

Levels are specified among which specific product moments are pooled whan running programme CSA/SELECTOR. This computes a matrix of pooled products and converts them into chord distances which are stored in a file. When m levels are specified, the pooled matrix holds the global product moments from which global correlations and global chord distances are computed.

The analysis branches out at this point. Three programs are particularly relevant: SSA, EIGENAN and MDSCAL. SSA performs cluster analysis based on the chord distances and EIGENAN or MDSCAL performs ordination. EIGENAN uses an eigenanalysis route (chained program EIGEN) and by calling PCAR/PART-II prints the eigenvalues and eigenvectors of a Q matrix derived from the chord distances. The output from SSA is a tree file from which a dendrogram is drawn by programme TREE. The output from EIGENAN and MDSCAL include sets of metric coordinates for the CSMs. Scattergrams and stereograms are drawn by program PLOS and STEREO.