ON CHARACTER-BASED PLANT COMMUNITY ANALYSIS: CHOICE, ARRANGEMENT, COMPARISON[1]

L. Orlóci, Department of Biology, New Mexico State University, Las Cruces, New Mexico, U.S.A.[2]

Abstract. 'Character-based' implies 'species-free' in that the analysis is unconstrained by an external species-based taxonomy. The taxonomy is intrinsic and connected with the states of chosen characters. The objective is maximal local relevance. With these in mind, character choice, character arrangement, and community comparison are discussed.

Introduction

Character-based descriptions were central in the early works on vegetation (e.g., Humboldt 1806, Kerner von Marilaun 1864, Warming 1884, Raunkiaer 1907) which had keen interest in the environmental fitness of plants and plant survival. With advances in plant taxonomic thought, and knowledge of the continental flora, a species-based, syntaxonomy-oriented approach emerged. Epitomized by J. Braun-Blanquet's work (Braun-Blanquet 1928), this approach became known as the 'Zürich-Montpellier School' of phytosociology (c.f., Becking 1957, Westhoff and van der Maarel 1978). Admittedly, on the plus side, the species-based approach gave coherence to vegetation studies of many dialects (Whittaker 1962), but it was prone to deficiencies, owing to uniformity in heredity taking precedence over unity in survival characteristics in the delineation of plant taxa. On the analytical side - considering that phytosociology has always been a largely comparative science - the species-based comparisons faced indeterminacy problems (Orlóci and Orlóci 1985, Orlóci and Stofella 1986). Inefficiency in ecological terms, and indeterminacies in analytical terms, are aspects in direct contradiction with the assumptions from which the species-based phytosociological approach derived its justification.

Character-based vegetation analysis has continued sporadically through the years as a parallel approach in plant community studies (e.g., Dansereau 1957, Knight and Loucks 1969, Mooney et al. 1970, Mooney and Dunn 1970, Karr and James 1975, Orians and Solbrig 1977, Chabot and Mooney 1985, Werger et al. 1988, Orshan 1988). Although logically mandated and apt to alleviate the indeterminacy problem, the technical aspects of character choice, arrangement and comparison in this approach are heavy with unsolved problems. It is in these connections that I offer thoughts on character-based community analysis.

Character choice

With focus on the community, the character choice is wide open. Historically, characters such as Humboldt's (1806) 'Hauptformen', the 'Grundformen' of Kerner von Marilaun (1864), and Warming's (1884) and Raunkiaer's (1907) 'livsformer' were the preferred types. Lacza and Fekete (1969; Fekete and Lacza 1970, 1971, 1972), Mueller-Dombois and Ellenberg (1974), and Barkman (1988) give reviews. The historical character types are not alien in modern work, but reappear in various admixtures of selected physiognomic, morphological, and anatomical plant properties (Dansereau 1957, Knight and Loucks 1969, Feoli and Scimone 1984, Lausi and Nimis 1985, Orlóci and Orlóci 1985, 1989, Orlóci and Stofella 1986, Orlóci 1988a, b).

Considering that the number of potential characters is practically limitless, character selection must be constrained. Criteria of relevance, which spring naturally from the objectives, and the rules of specificity and communality may be exercised to fine tune a character set:

The specificity rule. The preferred characters have high unshared and low shared information. The decision relies on reasoning from basic knowledge and information from reconnaissance and preliminary analyses. The character 'functional leaves present/absent' is a typical case whose selection would be justified if the specificity rule were applied for character selection in some desert community. The two states (present/absent) would occur with similar frequency; therefore, the information content should be high. The information shared with other characters could be low, since in such a community the presence and absence of functional leaves would be expected to be independent from other characteristics that would likely be chosen.

The communality rule. The characters preferred have high shared information. This is typical when cha-

---

racters are broadly defined, such as 'plant function' (evergreen, deciduous, etc.), 'biological type' (moss, lichen, fern, conifer, etc.), etc.

Users often prefer selection by the communality rule. This is not surprising, since a parsimonious community description is achieved, which is a common objective. Complete reliance on the communality rule may, however, be unwise; the characters selected may lack the power of qualitative discrimination.

Character arrangement

Contrasting schemes

The scheme in Knight and Loucks (1969) is typical for one type of character arrangement. It involves a mixture of very different characters. In their scheme the character arrangement is sequential and the character states, as used, represent quasi taxa. The plant individuals that carry the same state of the character, say, life-form, are measured for joint performance (frequency, density, yield, etc.).

The character set is more focused in other schemes which use a nested, hierarchical arrangement (Orlóci and Orlóci 1985, Orlóci et al. 1986, Orlóci and Stofella 1986). This arrangement is connected with the definition of a taxon as a Character Set Type (CST) which is a population whose individuals are identical with respect to the states of the chosen characters.

Information potential

By definition, the *sequential scheme* has an information potential proportional to

$$\sum_{h=1}^{m} \log_2 u_h$$

In this expression, symbol u_h signifies the number of states of the hth among m characters. A noted drawback of the sequential schemes, besides the low information content compared to other more complex schemes, is the difficulty of character set expansion. New characters can be added only at the cost of multiple scoring and uncontrolled character weighting.

The information potential is much higher in the *nested, hierarchical scheme*. This is proportional to

$$\sum_{i=1}^{m} \log_2 \left(\prod_{h=1}^{m-i+1} u_h \right)$$

where symbol i defines the hierarchical level. Note that the ith level corresponds to character m-i+1. Character set expansion is not a problem in this scheme, once a character order is established.

Character ordering

The order in which characters are presented for analysis does not matter in the sequential scheme, but it matters very much in the hierarchical scheme. The character order thus is a potential source for difficulties, similar as in other exercises where orthogonal functions replace the variables (Rao 1952). Different, often conflicting requirements have to be balanced when stipulating a character order. For example:

Logical ordering. A logical arrangement may be looked for in evolutionary or developmental terms. Since a unidirectional ordering is sought, a strictly logical arrangement may not be sustainable in these terms in large character sets.

Minimum 'noise'. Ambiguities in the scoring introduce errors. Since the errors are amplified on the lower levels in the hierarchical pathway of CSTs, the least ambiguous characters should be placed at the head of the list.

Non-trivial partial correlations. This requirement should be fulfilled on all hierarchical levels. For this, the high-level characters should have several states.

Maximum partial correlations on all hierarchical levels. This could be achieved simply through a recursive analysis involving different permutations of the characters. The approach faces serious difficulties when the character set is large, or more than two communities are compared, in which case an 'average' correlation has to be maximized.

More on the hierarchical scheme

Every ordered character set defines a unique hierarchy. In this, the mappings of the CSTs are runs through the stems across the hierarchical levels. Where stems meet, a node is recognized. The characteristic functions that describe a nested hierarchy, level i, include -

Number of nodes: $k_i = u_1...u_{(m-i+1)}$. u_h signifies the number of states in character h. $k_i = u_1...u_m$, the potential number of CSTs, that is, the width of the hierarchy at the tip of the branches.

Number of units per node, *i.e.*, block size: $n_i = u_m...u_{(m-i+1)} + 1$.

Data cumulants: X_i of k_i elements. The jth element X_{ij} (a cumulant) is the sum of original performance estimates of n_i CSTs on level 1.

Pattern of CST mappings: the manner in which the X_{ij} cumulants are arranged on level i. It is important to note that the pattern on level i is dependent upon the cumulants for all characters from the first level (last character) to the m-i+1st character (ith level).

Community level comparison

The comparison uses the hierarchical mappings of the CSTs in the different communities. To show how similar these patterns are in communities A and B, I sug-

gested the coefficient

$$r_{iAB} = \frac{W_{iAB}}{[W_{iAA}\ W_{iBB}]^{0.5}}$$

where

$$W_{iAB} = \sum_{j=1}^{k_i} \frac{X_{ijA}\ X_{ijB}}{n_i} - \sum_{j=1}^{k_i} \frac{X_{(i+1)jA}\ X_{(i+1)jB}}{n_{(i+1)}}$$

(Orlóci and Orlóci 1985). The r quantity is a *partial* correlation coefficient, which is regarded here as a mathematical index, rather than a statistic. r can be rendered a probabilistic estimator, but for this, my character sets are likely to preclude the use of standard statistical distributions. Heuristic methods, which derive empirical distributions from the data, are available (*e.g.*, Goodall 1966, Orlóci et al. 1988). The partial correlation coefficient detects divergent (negative r) and convergent (positive r) patterns, or the lack of either (zero r). The presentation of pattern correlations use line graphs, or alternatively, plexus diagrams, scattergrams, or dendrograms. The construction of these is usually preceded by suitable transformations, such as in ordinations and cluster analysis (Orlóci *et al.* 1986, Orlóci and Stofella 1986). Extensions to different data types (mixed, purely qualitative, purely quantitative) as in Williams and Dale (1962) are defined. Averages are also defined.

An example

Community description based upon the nested, hierarchical scheme produces a score matrix for each community (Table 1). I refer to this matrix as a 'relevé'. The characters are row categories and the CSTs are the columns. Each character has two or more states, and all states are categorical, or categorized if the character is a continuous variable. The entries in the cells inside the matrix are character scores. The cells in the last row contain estimates of performance.

This type of data is at the base of analyses for which partial results are presented in Fig. 1. A comparison of two communities is involved, both from the same transect which M. Orlóci and I surveyed at White Sands National Monument in New Mexico's Tula Rosa Basin. The transect runs through alkali flat from playa through fore dune to interdune flat. The locality is on the northern fringe of the Chihuahua Desert. It is interesting to note that the dune sand is pure gypsum. Dune height may reach 6m. The general terrain is at 1219m above sea level, and the mean annual precipitation is about 200mm.

The first community (relevé 2 in my records) is on the playa fringe (Table 1). The second is a playa grassland (relevé 6, in my records). Figure 1 gives the partial correlation profiles. The mixed data profile indicates

Table 1. Sample score sheet. A playa fringe community is described from White Sands National Monument, New Mexico. Plant names typify character set types. Characters 1,2 are general; 3-10 pertain to the photosynthetic organs (leaf, stem, or like structures). C/A code is the Braun-Blanquet (1951) cover/abundance code for plant populations.

Characters/Character set types	Lichen	Sporobolus	Atriplex	Sueda	Allenrolfea
1. Biological type 1:bryoid, 2:lichen, 3:pteridophyte, 4:conifer, 5:graminoid, 6:cactoid, 7:other	2	5	7	7	7
2. Groth form (code from key)	4 (24)	12	8	9	8
3. Spatial continuity 1:contiguous, 2:contagious	1	1	2	2	2
4. Duration 1:persistent, 2:deciduous, 3:withering	1	3	1	3	3
5. Architecture 1:simple, 2:compound	2	2	1	1	1
6. Shape 1:linear/cylindrical, 2:oval/ovoid	2	1	1	1	2
7. Cross section 1:straight, 2:curved, 3:involute, 4:circular	2	2	2	4	2
8. Longitudinal section 1:straight, 2:other	3	3	2	1	2
9. Tissue type 1:filmy, 2:fleshy, 3:fibroid, 4:other	3	3	3	2	2
10. Surface quality 1:coarse, 2:smooth	2	1	2	2	2
C/A code	4	1	2	2	2

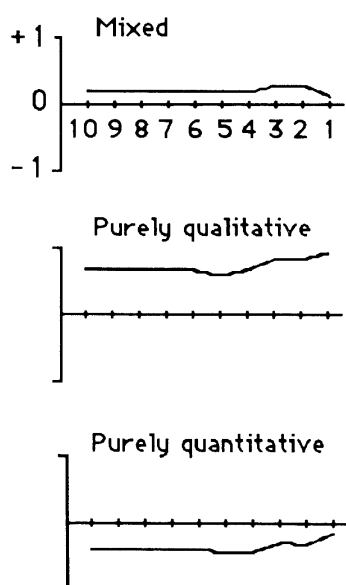weak convergence in compositional terms, or no con-

Fig. 1. Profiles of partial correlations for playa fringe and playa grassland at White Sands National Monument, New Mexico. Vertical scale measures correlation. Numbers from 1 to 10 on horizontal axis identify the variables in Table 1. The individual profiles correspond to data types as labeled. Global convergence is indicated by the first and second graph, and divergence by the third.

vergence (character 1: biological type). In qualitative data terms, the convergence is strong on all characters. In purely quantitative terms, there is evidence for moderately strong divergence, excepting biological type. In other words, while playa fringe and playa grassland are practically indistinguishable in qualitative terms (the sharing of common CSTs), the performance of the same plant forms tend to differ considerably.

Summary

Character-based community analysis is considered in comparison with species-based community analysis. It is in this comparative context that the general rules of character selection, the schemes of character arrangement, and specific data analytical considerations are discussed. It is mentioned that information potential is highest if the character arrangement is nested hierarchical. Different criteria are identified and discussed for ordering characters. These stress logical connectedness, minimum noise, nontrivial correlations, and maximal partial correlations on all hierarchical levels. Concrete cases are considered and community comparisons are described on the basis of a nested character hierarchy. An example from the White Sands National Monument, New Mexico, illustrates the general points.

Through the results and discussions, the paper sketches the broad outlines of a species-free methodology capable of unifying vegetation analysis in a survey or an experimental context. The general conclusion is that

the species-free scheme alleviates the chronic indeterminacy problem, which hinders the effectiveness of numerical methods in comparative community analysis, and permits the selection of ecologically meaningful plant or vegetation characteristics for the basis of the study. The nested character arrangement enhances the utility of the scheme even further through avoiding the character set expansion problem. It permits the partition of a community comparison measure, such as the multiple correlation, into components, such as the partial correlations, specific to hierarchical levels and individual characters, or levels of community convergence and divergence in character set terms.

REFERENCES

BARKMAN, J.J. 1988. New systems of plant growth forms and phenological plant types. In: M.J.A. Werger, P.J.M. van der Aart, H.J. During and J.T.A. Verhoeven (eds.). Plant Form and Vegetation Structure. Adaptation, Plasticity and Relation to Herbivory, pp. 9-44. S.P.B. Academic Publishing, The Hague.

BECKING, R.W. 1957. The Zürich-Montpellier School of phytosociology. Bot. Rev. 23: 411-488.

BRAUN-BLANQUET, J. 1928. Pflanzensoziologie. Berlin, Springer-Verlag.

CHABOT, B.F. and H.A. MOONEY (eds.). 1985. Physiological Ecology of North American plant communities. Chapman and Hall, New York.

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

GOODALL, D.W. 1966. A new similarity index based on probability. Biometrics 22: 883-907.

HUMBOLDT, A. VON. 1806. Indeen zu einer Physiognomik der Gewächse. Cotta, Stuttgart.

KARR, J.R. and F.C. JAMES. 1975. Eco-morphological configurations and convergent evolution in species and communities. In: M.L. Cody and J.M. Diamond (eds.). Ecology and Evolution of Communities, pp. 258-291. Harward University Press, Cambridge, Massachusetts.

KERNER, A. VON MARILAUN. 1864. Das Pflanzenleben der Donauländer. 2nd ed. by F. Vierhapper, 1929. Univ.-Verlag Wagner, Innsbruck.

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 Annals Historico-Naturales Musei Nationalis Hungarici, Pars Botanica 61: 129-139.

FEKETE, G. and J. SZ. LACZA. 1970. A survey of plant life-form systems and the respective research approaches. I Annals Historico-Naturales Musei Nationalis Hungarici, Pars Botanica 62: 115-127; part III (1971) 63: 37-50; part IV (1972) 64: 53-62.

FEOLI, E. and M. SCIMONE. 1984. A quantitative view of tex-

tural analysis of vegetation and examples of application of some methods. Arc. Bt. e Biogeografico Italiano 60: 73-94.

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

MOONEY, H. and E.L. DUNN. 1970. Convergent evolution in mediterranean-climate evergreen schlerophyll shrubs. Evolution 24: 292-303.

MOONEY, H. E.L. DUNN, F. SHROPSHIRE and L. SONG. 1970. Vegetation comparisons between the mediterranean climatic areas of California and Chile. Flora 159: 480-496.

MUELLER-DOMBOIS, D. and H. ELLENBERG. 1974. *Aims and Methods of Vegetation Ecology*. Wiley, New York.

ORIANS, G.H. and O.T. SOLBRIG (eds.). 1977. *Convergent Evolution in Warm Deserts*. Dowden, Hutchison and Ross, Stroutdsburg, Pennsylvania.

ORLÓCI, L. 1988a. Detecting vegetation patterns. ISI Atlas of Science, Animal and Plant Science, Vol. 1, 173-177.

ORLÓCI, L. 1988b. Community organization: recent advances in numerical methods. Can. J. Bot. 66: 2626-2633.

ORLÓCI, L., E. FEOLI, D. LAUSI and P. NIMIS. 1986. Estimation of character structure convergence (divergence) in plant communities: a nested hierarchical model. COENOSES 1: 11-20.

ORLÓCI, L., N.C. KENKEL and M. ORLÓCI. 1988. *Data Analysis in Population and Community Ecology. Vol. 1. Static Systems*. Prepublication edition. Department of Plant Sciences, The University of Western Ontario, London, Canada.

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

ORLÓCI, L. and M. ORLÓCI. 1989. Edge detection in vegetation: Jornada revisited. Mimeographed. Department of Plant Sciences, The University of Western Ontario, London, Canada.

ORLÓCI, L. and S.L. STOFELLA. 1986. The taxon-free approach to the study of plant communities. Annals of Arid Zone 25: 111-131.

ORSHAN, D.G. (ed.). 1988. *Plant Phenomorphological Studies in Mediterranean Type Ecosystems*. Springer-Verlag, New York.

RAO, C.R. 1952. *Advanced Statistical Methods in Biometric Research*. Wiley, New York.

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

WARMING, E. 1884. *On Skudbygning, Overvintring og Foryngelse*. Festskr. Naturh. Foren. Kjöbenhaven.

WERGER, M.J.A., P.J.M. VAN DER AART, H.J. DURING and J.T.A. VERHOEVEN (eds.). 1988. *Plant Form and Vegetation Structure. Adaptation, Plasticity and Relation to Herbivory*. S.P.B. Academic Publishing, The Hague.

WESTHOFF, V. and E. VAN DER MAAREL. 1978. The Braun-Blanquet approach. In: R.H. Whittaker (ed.). *Classification of Plant Communities*. pp. 287-399. Dr. W. Junk, The Hague.

WHITTAKER, R.H. 1962. Classification of natural communities. Bot. Rev. 28: 1-239.

WHITTAKER, R.H. (ed.). 1978. *Classification of Plant Communities*. Dr. W. Junk, The Hague.

WILLIAMS, W.T. and M.B. DALE. 1962. Partition correlation matrices for heterogeneous quantitative data. Nature 196: 602.