Journal of Vegetation Science 17: 113-117, 2006
© IAVS; Opulus Press Uppsala.

FORUM

Braun-Blanquet's legacy and data analysis in vegetation science

Podani, János

Department of Plant Taxonomy and Ecology, Eötvös University, Pázmány P. s. 1/c, H-1117 Budapest, Hungary; Fax: +36 13812188; E-mail: podani@ludens.elte.hu

Abstract. This article investigates whether the Braun-Blanquet abundance/dominance (AD) scores that commonly appear in phytosociological tables can properly be analysed by conventional multivariate analysis methods such as Principal Components Analysis and Correspondence Analysis. The answer is a definite NO. The source of problems is that the AD values express species performance on a scale, namely the *ordinal* scale, on which differences are not interpretable. There are several arguments suggesting that no matter which methods have been preferred in contemporary numerical syntaxonomy and why, ordinal data should be treated in an ordinal way. In addition to the inadmissibility of arithmetic operations with the AD scores, these arguments include interpretability of dissimilarities derived from ordinal data, consistency of all steps throughout the analysis and universality of the method which enables simultaneous treatment of various measurement scales. All the ordination methods that are commonly used, for example, Principal Components Analysis and all variants of Correspondence Analysis as well as standard cluster analyses such as Ward's method and group average clustering, are inappropriate when using AD data. Therefore, the application of ordinal clustering and scaling methods to traditional phytosociological data is advocated. Dissimilarities between relevés should be calculated using ordinal measures of resemblance, and ordination and clustering algorithms should also be ordinal in nature. A good ordination example is Nonmetric Multidimensional Scaling (NMDS) as long as it is calculated from an ordinal dissimilarity measure such as the Goodman & Kruskal γ coefficient, and for clustering the new OrdClAn-H and OrdClAn-N methods.

Keywords: Abundance/dominance data; Clustering; Nonmetric Multidimensional Scaling; OrdClAn methods; Ordinal scale; Ordination; Phytosociology.

Abbreviations: AD = Abundance/dominance; NMDS = Nonmetric Multidimensional Scaling.

Introduction

J. Braun-Blanquet has been recognized as a leader in the history of vegetation science. His seminal books (Braun-Blanquet 1928, 1932) greatly influenced the field practice and scientific thinking of many vegetation ecologists throughout continental Europe, and later all over the extra-tropical part of the world (Mueller-Dombois & Ellenberg 1974; Fujiwara 1987). The descriptive and classificatory phase of the discipline cannot be evaluated properly without reference to him, his colleagues and followers, i.e. to the Zürich-Montpellier school of phytosociology. General plant ecology has benefited much from the phytosociological approach (Ewald 2003), and the tradition lives on as demonstrated by new books re-iterating, re-formulating and updating the fundamental concepts (e.g. Dierßen 1990; Dierschke 1994; Dengler 2003). Whereas the positive influence of phytosociology is obvious in several respects, for instance in the accumulation of our immense knowledge on the vegetation of the Earth, many authors have warned that the approach poses problems. I would like to emphasize an aspect that has as yet received relatively little attention, even in critical reviews. This concerns the manner in which Braun-Blanquetian vegetation data are treated by numerical methods.

Relevés: the source of most information – and of most problems

The standard method for recording field observations in phytosociology is sampling by relevés. A relevé is a list of species observed in a quadrat together with estimates of their abundance/dominance or cover. Braun-Blanquet restricted the use of relevés to 'homogeneous' and 'typical' stands of communities, presumably because he was strongly influenced by the traditional attitude of taxonomists towards Nature: the description and classification of organisms should largely be based on the most typical-looking plant or animal individuals. Furthermore, when Braun-Blanquet developed his ideas computers were not yet invented, sampling theory was in an initial stage and the use of statistical methods in biology was exceptional. As a result, the relevé method is burdened by much bias, subjectivity, inconsistency, arbitrariness, circular argumentation and large sampling error (cf. Podani 1984 and references therein; Feoli 1984; Lepš & Hadincová 1992; Chytrý 2001; Chytrý & Otypková 2003) – a peculiar combination of flaws probably unprecedented in the natural sciences. We cannot blame Braun-Blanquet, of course, that he was unable to envisage all difficulties that were revealed only many decades after the approach was launched. The problems associated with sampling can be resolved by modernizing and standardizing field methods, and there are attempts in the contemporary literature to achieve this goal (Bruelheide & Chytrý 2000; Mucina et al. 2000; Chytrý & Otypková 2003; Grabherr et al. 2003). These authors promote the use of various objective sampling designs and standardized quadrat sizes which allow generalizations to be made from relevé data.

Numerical syntaxonomy

Computers were introduced into the processing of phytosociological data as early as the 1960s (cf. van der Maarel 1975, 1982). Transition into the numerical phase was greatly stimulated by the increased availability of computer program packages and personal computers, and the appearance of important books (e.g. Whittaker 1973; Orlóci 1978). Classification and ordination of relevé data became an everyday practice of vegetation scientists. The paradigm shift towards numerical analysis was relatively straightforward, mostly because subjectivity, inconsistency, and arbitrariness in selecting the objects of the study play no direct role in multivariate data exploration. Cluster analysis and ordination impose no restrictions on sampling conditions: any set of objects can be analysed provided that these are characterized by an appropriate set of descriptors. Furthermore, any dissimilarity matrix can be subjected to complete linkage clustering or principal coordinates analysis, for example, without violating any mathematical rule. Of course, sampling methods need to be consistent with the objectives of an investigation (see Kenkel et al. 1989 for a review of this topic). Classification and ordination methods do have their own limitations, of which compatibility of a procedure with the type of data is of primary concern.

The issue of measurement scale

The basic source of computational problems is the scale on which species performance is measured within each relevé. Braun-Blanquet's well-known abundance/dominance scale is *r*, +, 1, 2, 3, 4 and 5, quite often diluted with intermediate scores such as + - 1 or 1 - 2. Obviously, the presence of non-numbers *r* and + in the data immediately excludes the possibility of calculations, therefore several procedures have been suggested to convert the values to some other scale containing only numerals (van der Maarel 1979).

These transformations overcome the non-number problem, but the newly defined 'ordinal scale' preserves a less apparent property: certain operations with the values are strictly inadmissible (Anderberg 1973). Differences, sums and ratios of possible values are not interpretable on the ordinal scale, only the relations = and < are meaningful, so that only the ordering of values conveys information. In other words, the difference between 1 and 2 is not the same as that between 3 and 4, 1+2 is not equal to 3 and 1/2 is not the same as 2/4. The essence of the problem is that conventional dissimilarity functions, clustering and ordination algorithms operate via subtraction, addition and division and consequently their application to relevé data is inappropriate.

It is striking that numerical syntaxonomic surveys analysing Braun-Blanquet-type data completely disregard this incompatibility, although there have been a few reports recognizing this problem. Dale (1989), for example, discussed the possibilities of calculating dissimilarity based on ordinal abundance data, but his work has been overlooked almost completely. The ISI database includes only two references to this important paper, while Google Scholar Search finds only one more document – a painful indication of general ignorance.

At first glance, there appear to be several remedies. At the level of sampling, one could record percentage cover, biomass or counts of individuals, i.e., 'quantitative data', so as to avoid all problems mentioned above. But vegetation databases already include hundreds of thousands of relevés most of which described in terms of ordinal variables, and the joint evaluation of such traditional data and quantitative data would be impossible. Another solution is simplification of data to presence/absence scores whose analysis poses no computational problems. However, this implies information loss because the ordinal scale variables tell us more about interspecific relationships than simple presence and absence data.

A further possibility is conversion of Braun-Blanquet scores to mean values of percentage cover classes, but the arbitrariness and the non-systematic distortion implied in this operation are obvious. For example, if the

AD value of 5 is replaced by 87.5%, as commonly suggested, then the new value will designate all actual cover values from 75 to 100%, thus increasing uncertainty in the data considerably.

Ordinal data analysis

The best and mathematically correct solution of the scale problem is the use of ordination and classification procedures that are compatible with variables measured on the ordinal scale, that is, the application of ordinal methods of data analysis. For those not convinced yet by the above reasoning on inadmissible arithmetic operations, I have three further arguments supporting the point that ordinal phytosociological data should be treated in a way other than they usually are.

Argument 1: Meaningful dissimilarity

The most critical step in selecting the appropriate method is the choice of a dissimilarity coefficient which is meaningful phytosociologically and, at the same time, compatible with ordinal data. As an example, let us consider the following artificial phytosociological table for three relevés and two species:

	Relevés		
	h	*i*	*j*
Bromus erectus	1	2	4
Poa bulbosa	2	1	2

If we formally use the well-known Euclidean distance to measure dissimilarity, we find that relevés *h* and *i* are closest to each other, and then follow the pairs *ij* and *hj*, more precisely, $d_{hi} < d_{ij} < d_{hj}$ (= 1.41 < 2.23 < 3). It catches the eye, however, that in the most similar relevés we find a reverse relationship or 'negative correlation' between the two species. Namely, *P. bulbosa* is more 'important' than *B. erectus* in relevé *h*, whereas it is just the opposite in relevé *i*. Furthermore, $d_{hi} < d_{ij}$ even though the two species occur in the same proportion in relevés *i* and *j*. The conclusion is that Euclidean distance is misleading, and not only because differences are calculated between ordinal values but also due to its diminished ecological interpretability! One could say that standardization by relevé totals provides a more interpretable distance matrix, but this involves addition and division, which is not permissible with such data. We need to find a coefficient that is compatible with ordinal scores and at the same time gives ecologically interpretable results. The Goodman & Kruskal (1954) γ coefficient offers the simplest possibility. It counts the number of species pairs (*a*) that are identically ordered by the two relevés being compared and those that are reversely ordered (*b*). Then, similarity is defined as the ratio $\gamma = (a-b)/(a+b)$ which yields 1 if all species pairs are ordered in the same way by the two relevés, and –1 if the ordering is different for all species pairs. The complement of this coefficient, $1-\gamma$, provides dissimilarities which are ordered as $d_{ij} < d_{hi} = d_{hj}$ for the artificial example above. This coefficient, however, does not use the information present in situations in which AD values are equal and therefore ordering of species is not possible for one or both relevés, for example,

	Relevés	
	h	*i*
Carex humilis	2	1
Campanula sibirica	2	0

For an ecologist, this configuration is still informative in its presence/absence information, because these two species differ between the relevés. As an expansion of γ to incorporate this information, I suggested the use of a hybrid discordance measure (Podani 1997) for phytosociological data so that presence/absence still plays a role in calculating ordinal dissimilarity even if ordering is not possible. According to this function, the pair of *C. humilis* and *C. sibirica* increases the dissimilarity of the two relevés. For further examples illustrating the behaviour of this coefficient under circumstances not examined above, see App. A in Podani (2005).

Argument 2: Overall consistency

There is a definite direction of information flow in every numerical vegetation study, from sampling through various steps of data analysis to displaying the final diagrams. In this sequence, the quality of the very first step greatly influences the subsequent steps, which cannot give results that are of higher quality than the start (Gill & Tipper 1978). In phytosociology, this means that once ordinal data have been recorded, all steps that follow should also be ordinal in nature. The coefficient chosen should accept ordinal data, and then clustering and ordination procedures also should consider only the ordering relationships among the coefficients (Podani 2005). Application of Ward's clustering strategy (incremental sum of squares agglomeration), for example, is inappropriate for ordinal data because the existence of a Euclidean space is implied and thus the validity of all arithmetic operations is assumed. Moreover, the method is much more precise than the precision with which the data were collected. Instead of Ward's method, a clustering procedure is needed which only considers the ordering of ordinal dissimilarities. To achieve this goal, I have suggested a pair of methods, OrdClAn-H for hierarchical and OrdClAn-N for non-hierarchical classifications, and made them available through the SYNTAX 2000 program package (Podani 2001). Similar arguments hold true for ordination procedures: computing the eigenvalues in a metric ordination is a much more precise operation than relevé sampling. The most logical choice in this case is therefore Non-metric Multidimensional Scaling (NMDS) in which only the ordering relationships of dissimilarities are influential, rather than their actual values, in determining the configuration of objects in the ordination (Gordon 1999).

Argument 3: Universality

Imagine a situation in which relevés were made by 'Peter' in 1966 and by 'Paul' in 1990 in the same study area. Assume further that different AD scales were used, Braun-Blanquet scores by 'Peter' and the Domin scale by 'Paul'. If 'Mary' wishes to do a comparative study later, the two data sets cannot be combined and then analysed by conventional multivariate analysis methods. This is not a hypothetical possibility: Chytrý & Rafajova (2003) mention, for example, that the 54 310 relevés in the Czech phytosociological database were collected by 332 authors between 1922 and 2002, so that there is a very high chance for such data to be heterogeneous. In this case, the problem is not only the inadmissibility of subtraction and addition, but also that the values have different meanings on the two scales. For example, 2 implies different abundances on the Braun-Blanquet and Domin scales, therefore agreement of two relevés (one made by 'Peter' and the other by 'Paul') in having the same value of 2 for a given species is a misleading indication of their similarity. Comparable information in the data obtained by the two scientists is conveyed only by the ordering relationships among species, even though the ordering may be finer on one scale than on the other. From the arguments above it is clear that only ordinal procedures can correctly handle data that are expressed on different ordinal scales, and are therefore the only admissible solution in such synthetic studies. These methods allow comparisons based on a mixture of ordinal and quantitative data as well, further expanding the utility of the ordinal approach.

Conclusions

In reviewing the pioneering age of numerical syntaxonomy, Mucina & van der Maarel (1989) concluded that pragmatic criteria, i.e. the ability to reproduce intuitive classifications, seem to have been more important for vegetation scientists in making choices among available techniques than theoretical aspects have. Undoubtedly, a human factor is always present in numerical community studies and, as Gauch (1982) put it, "sampling method or analysis in the first place is an art". I agree with these authors that interpretability, intuitive appeal, custom, availability, simplicity and the like are relevant when judging the relative merits of methods, but I doubt whether the human factor deserves priority in any scientific discipline. This article emphasizes that mathematical admissibility, meaningfulness of dissimilarities, consistency and universality should override any other, mostly subjective criteria. Conventional clustering and ordination procedures do not satisfy these requirements if applied to Braun-Blanquet-type data types. I cannot recommend the use of methods that rely upon data standardization or the calculation of product moment correlation, covariance, Euclidean distance and similar functions because these imply arithmetic operations which are invalid for such data. To mention only a few examples, classification via Ward's method, group average clustering or indicator species analysis (as in TWINSPAN), and ordination by any variants of correspondence analysis (e.g. DCA) and principal components analysis are no proper choices of the investigator. Users of computer packages should be careful with the combinations of methods they choose: a non-metric method such as NMDS is still inappropriate if based on Euclidean distances. Resemblance coefficients developed for ordinal data, such as the Goodman & Kruskal γ, and subsequent analyses by NMDS and ordinal clustering methods provide mathematically correct and phytosociologically meaningful solutions. Podani (2005) presents further information on background theory and provides artificial and actual examples. The electronic Appendix includes an application of ordinal clustering to data matrix rearrangement via separate classifications of species and relevés, demonstrating that reconciliation between traditional tabular sorting and the new ordinal approach is straightforward.

Acknowledgements. I am grateful to my colleague, P. Csontos for discussions on contents and style. Comments from S. Díaz, J.B. Wilson and an anonymous referee contributed much to the improvement of the first version of the manuscript. Financial support came from the Hungarian National Scientific Grant (OTKA) no. 43732.

References

Anderberg, M.R. 1973. *Cluster analysis for applications*. Wiley, New York, NY, US.

Braun-Blanquet, J. 1928. *Pflanzensoziologie. Grundzüge der Vegetationskunde*. Springer, Wien, AT.

Braun-Blanquet, J. 1932. *Plant sociology: the study of plant communities*. McGraw-Hill, New York, NY, US.

Bruelheide, H. & Chytrý, M. 2000. Towards unification of national vegetation classifications: A comparison of two methods for analysis of large data sets. *J. Veg. Sci*. 11: 295-306.

Chytrý, M. 2001. Phytosociological data give biased estimates of species richness. *J. Veg. Sci*. 12: 439-444.

Chytrý, M. & Otypková, Z. 2003. Plot sizes used for phytosociological sampling of European vegetation. *J. Veg. Sci*. 14: 563-570.

Chytrý, M. & Rafajová, M. 2003. Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. *Preslia* 75: 1-15.

Dale, M.B. 1989. Dissimilarity for partially ranked data and its application to cover-abundance data. *Vegetatio* 82: 1–11.

Dengler, J. 2003. *Entwicklung und Bewertung neuer Ansätze in der Pflanzensoziologie unter besonderer Berücksichtigung der Vegetationsklassifikation*. Archiv naturwissenschaftlicher Dissertationen 14, Martina Galunder-Verlag, Nümbrecht, DE.

Dierschke, H. 1994. *Pflanzensoziologie – Grundlagen und Methoden*. Ulmer, Stuttgart, DE.

Dierßen, K. 1990. *Einführung in die Pflanzensoziologie*. Wissenschaftliche Buchgesellschaft, Darmstadt, DE.

Ewald, J. 2003. A critique for phytosociology. *J. Veg. Sci*. 14: 291-296.

Feoli, E. 1984. Some aspects of classification and ordination of vegetation data in perspective. *Studia Geobot*. 4: 7-21.

Fujiwara, K. 1987. *Aims and methods of phytosociology or 'vegetation science'. Plant ecology and taxonomy*. The Kobe Geobotanical Society, Kobe, JP.

Gauch, H.G. Jr. 1982. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge, UK.

Gill, D. & Tipper, J.C. 1978. The adequacy of non-metric data in geology: tests using a divisive omnithetic clustering technique. *J. Geol*. 86: 241-259.

Goodman, L.A. & Kruskal, W.H. 1954. Measures of association for cross classifications. *J. Am. Stat. Ass*. 49: 732-764.

Gordon, A.D. 1999. *Classification*. 2nd. ed. Chapman and Hall, London, UK.

Grabherr, G., Reiter, K. & Willner, W. 2003. Towards objectivity in vegetation classification: the example of the Austrian forests. *Plant Ecol*. 169: 21-34.

Kenkel, N.C., Juhász-Nagy, P. & Podani, J. 1989. On sampling procedures in population and community ecology. *Vegetatio* 83: 195-207.

Lepš, J. & Hadincová, V. 1992. How reliable are our vegetation analyses? *J. Veg. Sci*. 3: 119-124.

Mucina, L. & van der Maarel, E. 1989. Twenty years of numerical syntaxonomy. *Vegetatio* 81: 1-15.

Mucina, L., Schaminée, J.H.J. & Rodwell, J.S. 2000. Common data standards for recording relevés in field survey for vegetation classification. *J. Veg. Sci*. 11: 769-772.

Mueller-Dombois, D. & Ellenberg, H. 1974. *Aims and methods of vegetation ecology*. Wiley, New York, NY, US.

Orlóci, L. 1978. *Multivariate analysis in vegetation research*. 2nd. ed. Junk, The Hague, NL.

Podani, J. 1984. Spatial processes in the analysis of vegetation: theory and review. *Acta Bot. Hung*. 30: 75-118.

Podani, J. 1997. A measure of discordance for partially ranked data when presence/absence is also meaningful. *Coenoses* 12: 127-130.

Podani, J. 2001. SYN-TAX 2000. *Computer programs for data analysis in ecology and systematics*. User's manual. Scientia, Budapest, HU.

Podani, J. 2005. Multivariate exploratory analysis of ordinal data in ecology: pitfalls, problems and solutions. *J. Veg. Sci*. 16: 497-510.

van der Maarel, E. 1975. The Braun-Blanquet approach in perspective. *Vegetatio* 30: 213-219.

van der Maarel, E. 1979. Transformation of cover-abundance values in phytosociology and its effect on community similarity. *Vegetatio* 39: 97-114.

van der Maarel, E. 1982. On the manipulation and editing of phytosociological and ecological data. *Vegetatio* 50: 71-76.

Whittaker, R.H. (ed.) 1973. *Ordination and classification of communities*. Junk, The Hague, NL.

Received 22 September 2005;
Accepted 20 October 2005.
Co-ordinating Editor: S. Díaz.