A proposed procedure for the analysis of large phytosociological data sets in the classification of South African grasslands

G.J. BREDENKAMP and H. BEZUIDENHOUT

Bredenkamp, G.J. and H. Bezuidenhout. 1995. A proposed procedure for the analysis of large phytosociological data sets in the classification of South African grasslands. *Koedoe* 38(1): 33-39. Pretoria. ISSN 0075-6458.

A procedure for the effective classification of large phytosociological data sets, and the combination of many data sets from various parts of the South African grasslands is demonstrated. The procedure suggests a region by region or project by project treatment of the data. The analyses are performed step by step to effectively bring together all relevés of similar or related plant communities. The first step involves a separate numerical classification of each subset (region), and subsequent refinement by Braun-Blanquet procedures. The resulting plant communities are summarised in a single synoptic table, by calculating a synoptic value for each species in each community. In the second step all communities in the synoptic table are classified by numerical analysis, to bring related communities from different regions or studies together in a single cluster. After refinement of these clusters by Braun-Blanquet procedures, broad vegetation types are identified. As a third step phytosociological tables are compiled for each identified broad vegetation type, and a comprehensive abstract hierarchy constructed.

Key words: Classification, grassland, large data set, new procedures, South Africa.

G.J. Bredenkamp, Department of Botany, University of Pretoria, Pretoria, 0002 Republic of South Africa; H. Bezuidenhout, Scientific Services: Kimberley Office, PO Box 110040, Hadison Park, Kimberley, 8306 Republic of South Africa.

Introduction

Although the results of a number of research projects in southern African grassland vegetation have been published since the turn of the century, very few of these offer comparable results and reconciliation is virtually impossible (Scheepers 1987). Since the introduction of the Braun-Blanquet method to South Africa (Werger 1973) and the decision to standardise on this method for the analysis and description of South African vegetation, several Braun-Blanquet type surveys have been completed within grasslands or related vegetation (eg. Bredenkamp & Theron 1978, 1980; Coetzee 1974, 1975; Van Wyk & Bredenkamp 1986; Behr & Bredenkamp 1988; Bredenkamp *et al.* 1989; Kooij *et al.* 1990). Unfortunately most of these studies were confined to relatively small areas, such as nature reserves or other limited regions. For this reason, formal fixing of syntaxonomical ranks and names has been avoided (Bredenkamp *et al.* 1989), as information from a wider area is desirable to ensure correct ranking and naming of syntaxa (Tüxen 1970; Werger *et al.* 1972). The fragmentary information and the lack of a comprehensive formal hierarchical syntaxonomical system for South African vegetation in general, and the grasslands in particular, have led to the description and naming of many plant communities, without considering other studies or earlier names for possibly similar syntaxa. After the initiation of the South African Grassland Biome Project (Mentis & Huntley 1982) the urgent need for and necessity of an ecologically sound, comprehensive classification, description and mapping of South African grasslands was expressed (Scheepers 1987).

Due to the lack of data, especially from more extensive areas, the necessity of fieldwork as well as the incorporation of all existing compatible data was emphasised (Scheepers 1987). Consequently a phytosociological research program was encouraged by the founding of a Vegetation Classification and Mapping Working Group under the auspices of the Grassland Biome Project (Scheepers 1987). Within this working group the ecology groups of the University of Pretoria, National Botanical Institute (Department of Agricultural Development) and Potchefstroom University for Christian Higher Education initiated various phytosociological projects. The large number of relevés compiled in these studies facilitated the opportunity for a comprehensive formal syntaxonomical and synecological synthesis of the grasslands concerned. The questions were: i) how to treat the large data set available for grassland vegetation, and ii) how should the available information ideally be used in the envisaged formal syntaxonomical hierarchy. Although the grassland gives the impression of being floristically and even ecologically relatively homogeneous, this is not the case. The data represent many different plant communities and encompass an ecologically heterogeneous area.

A number of clustering programs can treat very large data sets e.g. CLUSLA (Louppen & Van der Maarel 1979), TWINSPAN (Hill 1979) and COMPCLUS (Gauch 1979). Furthermore, very large data sets have been treated together, using TABORD (Van der Maarel *et al.* 1978) or PHYTOPAK (Huntley *et al.* 1981).

In our comprehensive study an attempt was made to consolidate data of related plant communities from various previous studies and to incorporate the relevant data generated in current surveys.

The goals of this study were:

* to develop a procedure to process large data sets, which include relevés, and communities, from previous studies as well as newly compiled relevés,
* to create a suitable data base to manage all relevant data sets efficiently,
* to combine relevés representative of the same or related plant communities, but compiled by various workers in various projects, which may include various geographical regions, in a single phytosociological table or synoptic table, and
* to develop an efficient computer program to accomplish all necessary procedures accurately and efficiently.

However many problems were experienced regarding the treatment of this large data set (see also Van der Maarel *et al.* 1987):

* It is difficult or virtually impossible to handle phytosociological tables of this dimension by standard Braun-Blanquet procedures, especially where character species are absent and syntaxa are characterised by differential taxa or specific combinations of species or species groups.
* The identification and definition of certain ecologically interpretable communities may be difficult, as specific species combinations may be obscured in the large, heterogeneous data set.
* The objective demarcation of the data into various subsets, by using numerical classification methods, was also ineffective. This was mainly due to the heterogeneity of the data and also the presence of a large number of species with very limited occurrences in the total data set.

This leads to difficulty in choosing various options in many of the multivariate clustering methods (Van der Maarel *et al.* 1987).

* The particular algorithm used in many multivariate methods may result in unrealistic major divisions.

Van der Maarel *et al.* (1987) questioned the desirability and effectivity (see also Van der Maarel 1982) of simultaneous treatment of large heterogeneous data sets and discussed the various problems—concerning both methodology and ecology—of such treatments. Stratification of the data seems to be a realistic solution. A region by region treatment of geographically heterogeneous data is also preferred by Orlóci & Stanek (1979) and Jensén & Van der Maarel (1980).

Recently, a two-step procedure was found to be effective in the treatment of large data sets (Van der Maarel *et al.* 1987). By combining ideas on this two-step vegetation analysis method and ideas on the refinement of numerical classifications by Braun-Blanquet procedures (Bezuidenhout *et al.* 1988; Behr & Bredenkamp 1988), a procedure was developed for the treatment of large phytosociological data sets from South African grasslands.

This report aims to give an outline of the proposed procedures. The principles of the procedure were tested in a pilot study and the results of this study are presented.

Methods

Data collection

The data set includes 820 relevés with > 1 000 species, representing many different vegetation types and encompasses an ecologically heterogeneous area, covering approximately 22 000 km^2 in the north-western part of the grassland biome of South Africa (Fig. 1). Data were obtained from four previous studies and four current studies. The data were stratified by region in the following way (Fig. 1):

Previous studies:

Data set 1, from the Witwatersrand Geological Super Group (quartzite mountains), Heidelberg area (Bredenkamp & Theron 1978);

Data set 2, from the Ventersdorp Geological Super Group (lavas), Heidelberg area (Bredenkamp & Theron 1980);

Data set 3, from the Jack Scott Nature Reserve, Magaliesburg area (Coetzee 1974);

Data set 4, from the plains in the Potchefstroom-Fochville-Parys area (Bredenkamp *et al.* 1989);

Current studies:

Data set 5 from the Dolomitic region, in the Potchefstroom-Ventersdorp-Randfontein area (Bezuidenhout & Bredenkamp 1990);

Data set 6 from the quartzite hills and ridges in the Potchefstroom area (Bezuidenhout & Bredenkamp, new survey);

Data set 7 from the Roodepoort area (Bredenkamp & Bezuidenhout, new survey);

Data set 8 from the extensive bottomland areas in the Kroonstad area (Kooij & Bredenkamp, new survey).

Relevé data include total floristic composition with Braun-Blanquet type cover/abundance values (Westhoff & Van der Maarel 1978). Habitat data at least include geology, topography, soil type, soil depth, and rockiness of the soil surface while other soil properties may also be available in certain studies. Climatic data from the regions represented by the relevés are also available.

Data analysis

An elaboration of a recently proposed two-step procedure (Van der Maarel *et al.* 1987), was applied in this study. The procedure includes three steps and is as follows:

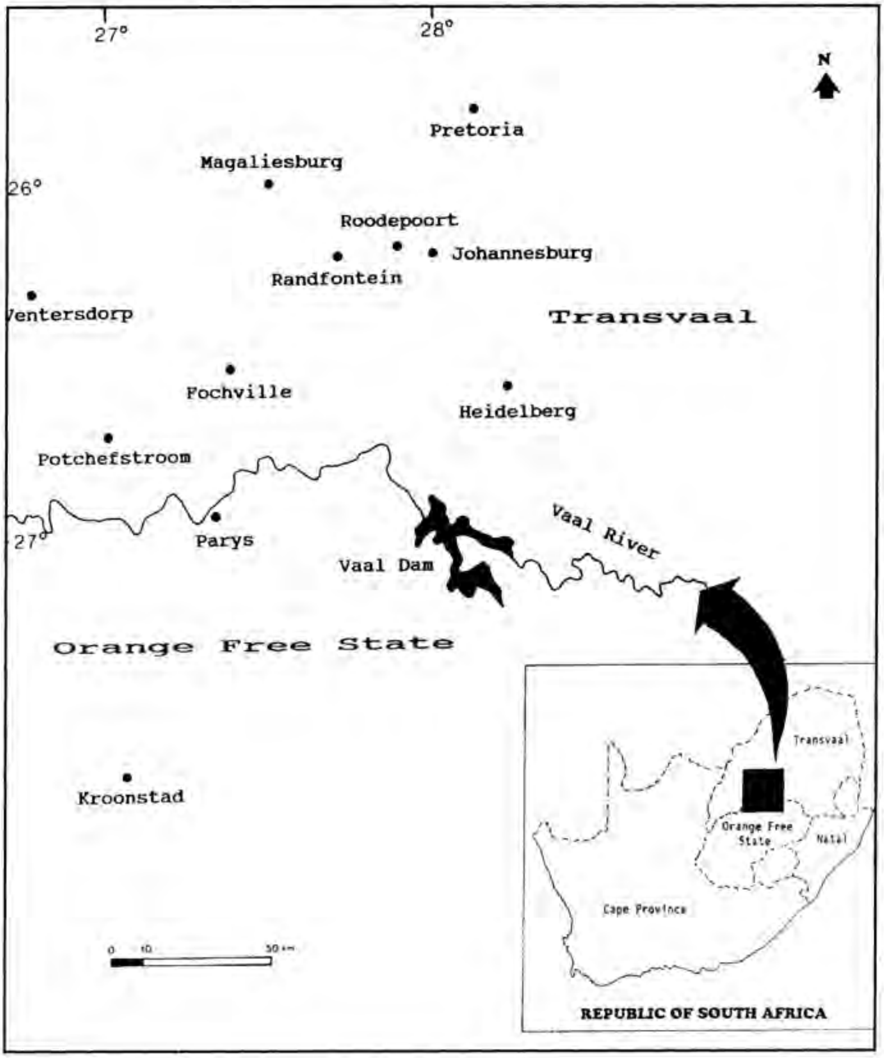

Fig. 1: The location of the study area.

Step 1.

a. Stratify the entire data set either by area or project, or if possible, by vegetation type. This stratification enables a region by region treatment of possibly geographical heterogeneous material.
b. Subject each subset to any suitable numerical classification. This analysis should result in a number fairly homogeneous clusters, representing a first approximation of possible community types. In our case we applied TWINSPAN (Hill 1979), as it was found that this algorithm results in a fair first approximation of the vegetation types and produces a fairly ordered two-way table (Bezuidenhout *et al.* 1988; Behr & Bredenkamp 1988; Bredenkamp *et al.* 1989).
c. Interpret the clusters obtained ecologically.
d. Refine, if necessary, by application of Braun-

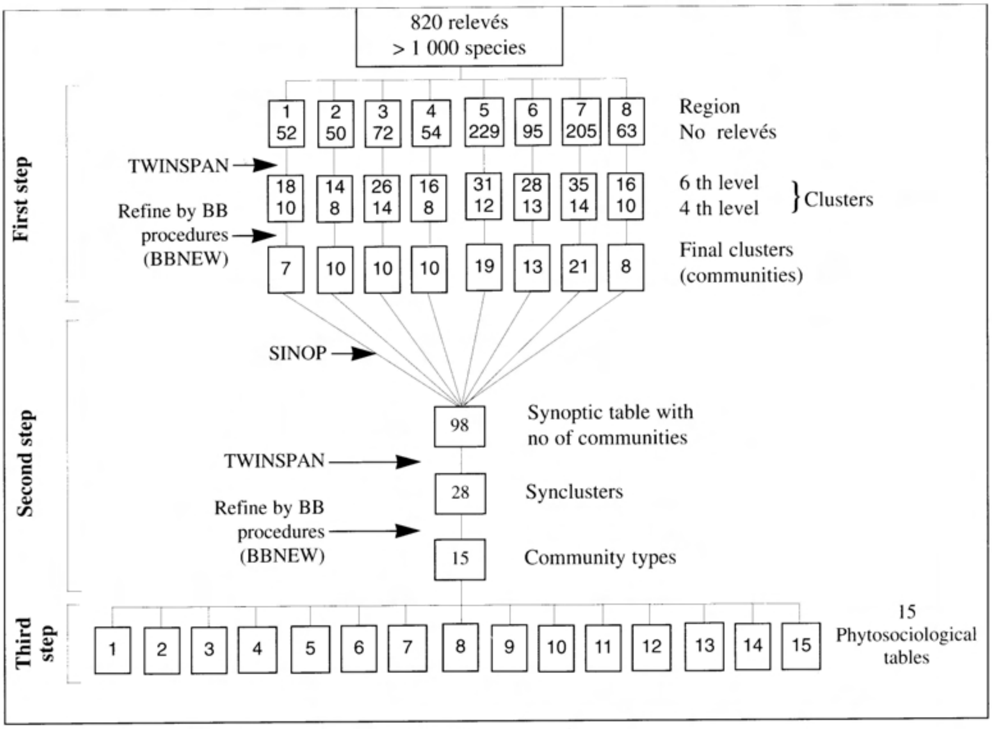

Fig. 2: Diagram showing the three step procedure and some results.

Blanquet procedures to establish ecologically interpretable plant communities.

e. Construct a synoptic relevé (composite sample) for each community, in each subset, using constancy/frequency classes with 20 % intervals, resulting in constancy/frequency values of 1 to 5 for the species. In this way each community is summarised to a single column in a synoptic table. The synoptic cover abundance value proposed by Van der Maarel *et al.* (1987) was not used in this study, but can easily be incorporated.

Step 2

a. Re-enter each community, identified from all subsets, as a synoptic relevé in a new synoptic data set and reclassify using the suitable numerical classification technique (TWINSPAN). Here we have the option to ignore all constancy/frequency values of < 20 %.
b. Refine the resulting classification, if necessary, by using Braun-Blanquet procedures. In this way all similar or related plant communities are brought together in a single cluster.
c. Identify from these clusters in the re-organised synoptic table the major relevant vegetation types probably representing higher syntaxa, for example classes.

Step 3

a. Construct different phytosociological tables to include all original relevés of a particular higher syntaxon.
b. Compile a hierarchical classification.
c. Identify and describe the relevant syntaxa.

Results and discussion

The results of the procedure are shown in Fig. 2. In the separate numerical analyses (TWINSPAN) of the eight subsets (Step 1) a total of 89 and 184 clusters were obtained at the fourth and sixth levels of division respectively. Refinement by Braun-Blanquet procedures produced a total of 98 plant communities. By applying Step 2, the TWINSPAN revealed 28 clusters (synclusters) which were reduced to 15 community types in the refinement using Braun-Blanquet procedures. The results of Step 3, that is the construction of 15 separate phytosociological tables, each containing relevés from one or more regions or surveys, and the description of the syntaxa, are not reported on in this paper.

Table 1
The regions in each of the distinguished community types

Community type number	Region number	Total number/ regions
1	1, 8	2
2	8	1
3	8	1
4	1, 3	2
5	1, 2, 3, 4, 6	5
6	2	1
7	2, 3	2
8	7	1
9	6, 7	2
10	4, 5	2
11	3, 7	2
12	7	1
13	4, 6, 7	3
14	6	1
15	4, 6	2

The figures given in Table 1 indicate that six of the community types contain relevés from only one region, whereas the remaining nine community types contain relevés from 2–5 different regions.

By using the above procedure all related relevés, and related communities, from different regions, compiled during different surveys by different workers, are effectively brought together in a single phytosociological table. Simultaneously the classification of the synoptic data set provides the basis for a comprehensive hierarchical classification of all the syntaxa identified during the study.

Another advantage is that data from new regions can easily be added to the analysis, thus enabling the compilation of a comprehensive data base of relevés and distinguished syntaxa, of the area concerned.

Acknowledgement

The Foundation for Research Development gave substantial financial support. Proff. S. Pignatti and E. van der Maarel are thanked for their valuable comments.

References

BEHR, C.M. AND G.J. BREDENKAMP. 1988. A phytosociological classification of the Witwatersrand National Botanical Garden. *South African Journal of Botany* 54: 525-533.

BEZUIDENHOUT, H. AND G.J. BREDENKAMP. 1990. A reconnaissance survey of the vegetation of the dolomitic region in the Potchefstroom-Ventersdorp-Randfontein area, South Africa. *Phytocoenologia* 18: 387-403.

BEZUIDENHOUT, H., G.J. BREDENKAMP AND J.H. ELSENBROEK. 1988. Die plantegroei van die alkaligraniet en aangrensende kwartsiet in die Vredefortkoepel noordwes van Parys. *Die Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* 7: 4-9.

BREDENKAMP, G.J., A.F. JOUBERT AND H. BEZUIDENHOUT. 1989. A reconnaissance survey of the vegetation of the plains in the Potchefstroom-Fochville-Parys area. *South African Journal of Botany* 55: 199-206.

BREDENKAMP, G.J. AND G.K. THERON. 1978. A synecological account of the Suikerbosrand Nature Reserve 1. The phytosociology of the Witwatersrand geological system. *Bothalia* 12: 513-529.

BREDENKAMP, G.J. AND G.K. THERON. 1980. A synecological account of the Suikerbosrand Nature Reserve 2. The phytosociology of the Ventersdorp geological system. *Bothalia* 13: 199-216.

COETZEE, B.J. 1974. A phytosociological classification of the vegetation of the Jack Scott Nature Reserve. *Bothalia* 11: 329-347.

COETZEE, B.J. 1975. A phytosociological classification of the Rustenburg Nature Reserve. *Bothalia* 11: 561-580.

GAUCH, H.G. 1979. *COMPCLUS: A fortran program for rapid initial clustering of large data sets.* Cornell University, Ithaca, New York.

HILL, M.O. 1979. *TWINSPAN: A fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes.* Cornell University, Ithaca, New York.

HUNTLEY, B., J.P. HUNTLEY AND H.J.B. BIRKS. 1981. PHYTOPAK: A suite of computer programs designed for the handling and analysis of phytosociological data. *Vegetatio* 45: 85-95.

JENSÉN, S. AND E. VAN DER MAAREL. 1980. Numerical approaches to lake classification with special reference to macrophyte communities. *Vegetatio* 42: 117-128.

KOOIJ, M.S., G.J. BREDENKAMP AND G.K. THERON. 1990. Classification of the vegetation of the B land type in the north-western Orange Free State. *South African Journal of Botany* 56: 309-318.

LOUPPEN, M.J.W. AND E. VAN DER MAAREL. 1979. CLUSCLA: A computer program for the clustering of large phytosociological data sets. *Vegetatio* 40: 107-114.

MENTIS, M.T. AND B.J. HUNTLEY. 1982. A description of the Grassland Biome Project. *South African National Scientific Programmes* 62: 1-29.

ORLÓCI, L. AND STANEK, W. 1979. Vegetation survey of the Alaska highway, Yukon territory: Types and gradients. *Vegetatio* 41: 1-56.

SCHEEPERS, J.C. 1987. Grassland Biome Project: Proceedings of the workshop on classification and mapping. *South African Ecosystem Programmes Occasional Report Series* 16: 1-31.

TÜXEN, R. 1970. Pflanzensoziologie als synthetische Wissenschaft. *Mededeelingen van het Botanish Tuinen Belmonte Arboretum Landbouwhogischool Wageningen* 12: 141-159.

VAN DER MAAREL, E. 1982. On the manipulation and editing of phytosociological and ecological data. *Vegetatio* 50: 71-76.

VAN DER MAAREL, E., I. ESPEJEL AND P. MORENO-CASASOLA. 1987. Two-step vegetation analysis based on very large data sets. *Vegetatio* 68: 139-143.

VAN DER MAAREL, E., J.G.M. JANSEN AND J.M.W. LOUPPEN. 1978. TABORD: A program for structuring phytosociological tables. *Vegetatio* 38: 143-156.

VAN WYK, S. AND G.J. BREDENKAMP. 1986. A Braun-Blanquet classification of the Abe Bailey Nature Reserve. *South African Journal of Botany* 52: 321-331.

WERGER, M.J.A. 1973. *Phytosociology of the upper Orange River Valley: A syntaxonomical and synecological study.* DSc thesis, University of Nijmegen, Nijmegen.

WERGER, M.J.A., F.J. KRUGER AND H.C. TAYLOR. 1972. A phytosociological study of the Cape fynbos and other vegetation at Jonkershoek, Stellenbosch. *Bothalia* 10: 599-614.

WESTHOFF, V. AND E. VAN DER MAAREL. 1978. The Braun-Blanquet approach. Pp. 287-399. *In*: WHITTAKER, R.H. (ed.). *Classification of plant communities.* The Hague: W Junk.