

NEW PROGRAMS FOR PRELIMINARY SEQUENCING OF RELEVÉS AND SPECIES IN PHYTOSOCIOLOGICAL DATA SETS

The primary aim of phytosociological classifications is to group relevés into vegetation units with these units arranged according to the closest possible floristic relationships. The secondary aim of classifications is to group species in a manner which will clarify these relationships. If the spaces between the first and last occurrence of a species in a phytosociological matrix are

regarded as separation units then the primary aim is achieved when the sum of the total separation units for each species is a minimum. Even currently available mainframe computers preclude use of this method of classification for all but the smallest data sets, because the calculations required are a factorial of the number of relevés.

The programs TWINSpan (Hill 1979), a polythetic divisive classification and PHYTO 20 (Westfall *et al.* 1982), a polythetic agglomerative classification, for preliminary sequencing, attempt to achieve the aims of classification using algorithms suited to available computers. It is unlikely that the same solution could be achieved with the two programs on the same data set and a considerable variation in separation units could be expected. Relevé and species sequences with both programs often require considerable refinement before a pattern approximating that with the minimum number of summed separation units is achieved. However, user refinement of sequences can be time-consuming mainly because pattern detection is hampered by species whose distribution is little affected by environmental factors within a given study area. These are the so-called non-diagnostic species. A new approach was therefore attempted to overcome the problem of non-diagnostic species recognition and to facilitate user sequencing for pattern refinement.

The new program, PHYTO 21, is written in PL/1 for the Burroughs B7900 and uses a simple polythetic divisive technique. It is an addition to the PHYTOTAB program package (Westfall *et al.* 1982). An assumption that final relevé sequence is primarily dependent on species with an occurrence of less than 50 % in the relevés of the data set, is made.

The relevés of the first species with 50 % occurrence are concatenated as are the relevés of the species with the highest occurrence that do not occur in the relevés of the first species. The matrix is thus divided into two extremes and can have intermediate relevés. The total number of species occurring only within the matrix range of both the concatenated species is determined. The process is repeated using the species with the next highest occurrence to the first species. It is assumed that the highest total represents division of the matrix into the greatest extremes. In practice it was found that the number of iterations could be limited.

Species are then sequenced to form seven groups: 1, those occurring only in the left part of the matrix; 2, those in the right; 3, those in the middle; 4, those in the left and middle; 5, those in the right and middle; 6, those in the left, middle and right; and 7, those with 100 % or single occurrences. The matrix, with relevés arranged according to extremes and species according to relative distribution in the matrix, is consequently easy to refine.

Printed output of the new program is a list of absolute and percentage frequency of occurrence for each species as well as the number of species per relevé and the number of species in each relevé as a percentage of the total number of species in the data set. The program also creates species and relevé sequence files. A matrix is obtained using PHYTO 30 (Westfall *et al.* 1982).

PHYTO 30 (Westfall *et al.* 1982) has been amended to give separation units for each species as well as the total number of separation units for the matrix. This facilitates pattern formation and allows an objective comparison of classifications for the same data set.

Classification, especially of large data sets, can often be facilitated by re-running PHYTO 21 on subsets of the data set. This can be achieved with the new program PHYTO 15, which also forms part of the PHYTOTAB program package (Westfall *et al.* 1982). Input for this program consists of species numbers that are required to be omitted. The program then creates a new data set file consisting of a subset of the original data set. Other programs including PHYTO 20 and TWINSpan can also be run on the subsets.

The basic input program, PHYTO 10 (Westfall *et al.* 1982), must first be run on each new subset before other programs can be run. PHYTO 10 has also been amended to check the cover-abundance codes used. This requires an additional input line, namely, the codes used for a particular data set.

The addition of the new programs to the package should greatly facilitate classifications, especially of large data sets (up to 2 000 relevés by 2 000 species).

ACKNOWLEDGEMENTS

The authors thank Dr J. C. Scheepers for comments and suggestions.

REFERENCES

- HILL, M. O. 1979. *TWINSpan—a FORTRAN program for arranging multivariate data in an ordered two way table by classification of individuals and attributes*. Ecology and Systematics, Cornell University, Ithaca, New York.
- WESTFALL, R. H., DEDNAM, G., VAN ROOYEN, N. & THE-
RON, G. K. 1982. PHYTOTAB. A program package for Braun-
Blanquet tables. *Vegetatio* 49: 35–37.

R. H. WESTFALL and B. C. DE WET

MS. received: 1987.09.29.