



CHAPTER 3. METHODS

In order to achieve the stated aims, the Braun-Blanquet method was viewed from a basic perspective, in terms of the components of the processes of stratification, sampling and classification, rather than to view only selected components. The methods that follow describe, in application sequence, techniques that build on the basic elements of the Braun-Blanquet method, within the flexibility of the Braun-Blanquet method (Coetzee 1975). The techniques were developed and refined over a ten year period and improvements have, therefore, been added to published versions where applicable.

The basic elements of the Braun-Blanquet method are vegetation familiarization leading to a formal or an informal stratification, species presence, sampled in quadrats, cover estimation and classification of species (rows) and samples (columns) in matrix form with cover-abundance values forming the matrix. Accordingly, the two 10 x 10 m plots in the second study area were each divided into 25 contiguous 2 x 2 m sampling units, wherein species presence only were recorded.

The components of the stratification and sampling processes can, for convenience, be grouped according to preparatory work and field sampling.



3.1 PREPARATORY WORK

Preparatory work is that work required before field sampling can commence. This includes a clearly stated, well-defined aim which is probably the most important aspect of a study because selection of appropriate methods is dependent on the aim. Results and the interpretation thereof, will be dependent on the methods selected, to achieve the aim. In this study, preparatory work also included determining linear regressions of rainfall data, obtained from the Weather Bureau, and a combination of altitude and latitude, so that interpolations of mean annual rainfall could be estimated for each sampling site within the study area.

3.1.1 Scale

Scale determines the maximum detail that can be obtained with a particular study and is dependent on the aims of the study.

3.1.1.1 Background

The plant community is a "group of plants sharing a common environment and is distinguished by a particular floristic composition" (Mueller-Dombois & Ellenberg 1974) at any scale smaller than that of the individual plant.

It is generally agreed that plant communities form hierarchies (Poore 1956; Mueller-Dombois & Ellenberg 1974) in that smaller plant communities can be included in larger plant communities.



However, this inclusion is often partial rather than whole. Du Reitz (in Poore 1956) states that "a stand may have parts of other ecological groups or vegetation units". This can often complicate plant community border recognition and the level of heterogeneity to include in a plant community. Which plant community to recognize depends on scale. Scale, in turn, is determined by the purpose of the study. In other words, the amount of detail required in a study determines the scale. Scale should, therefore, have an explicit role in the stratification, sampling and classification of vegetation.

According to Rutherford & Westfall (1986) the detail required in a study determines the working or associated sampling scale which in turn determines the smallest mappable unit area (SMUA). The SMUA is also related to the practical considerations of cartographic accuracy and user requirements of a linear precision seldom less than a millimetre (Rutherford & Westfall 1986). To these may be added the humidity-related, stretch and shrinkage of printed maps. At a given working scale, the area under study is subdivided by lines separating stratified units each containing a number of SMUA's which can include potential sampling sites (Rutherford & Westfall 1986). Each SMUA at working scale should, according to Rutherford & Westfall (1986):

- i. not be closer than 1 mm to a stratified unit's border; and
- ii. not be closer than 2 mm to another sampling site.

It follows that the width of an unmappable ecotone or transition is scale-dependent, being less than 2 mm on a map at a given work-



ing scale. A unit equal to or broader than 2 mm could, therefore, have community status at the relevant scale. The relationship of working scale to the smallest mappable unit area (SMUA) is derived from the distance from sampling unit area to stratified unit border and is given by:

$$y = e^{(-12,67 - \ln x)}$$

where y = SMUA in m^2 , and x = scale, as a representative fraction. In their account, Rutherford & Westfall (1986) give examples of MSDB (minimum sampling distance from a stratified unit border) and SCD (shortest cross distance of a mappable unit). Because working scale corresponds with detail, any map enlargement beyond the working scale will not improve the detail.

3.1.1.2 *Methods applied*

If the smallest mappable unit area (SMUA) is taken as a circle then the relationship between SMUA and working scale can be simplified, as follows:

$$r = x/1\ 000$$

where r = radius of SMUA in metres, and x = denominator of the scale as a representative fraction. For example, the radius, in metres of the SMUA for a working scale of 1:50 000, is 50 m. This is also the minimum distance on the ground between a stratified unit border and the SMUA, at this working scale. The radius of the SMUA and minimum distance between SMUA and stratified unit border is, therefore, always 1 mm on a map at any working scale.



The scale used in this study is 1: 250 000 which corresponds with the detail required for regional planning by the Department of Agricultural Development (see section 4.1.1).

3.1.2 Stand area

The New Collins Concise Dictionary of the English Language (1985) defines a stand as "a growth of plants in a particular area, especially trees in a forest or a crop in a field". The effect of taking scale into account in sampling at 1:250 000 scale (small scale sampling) is that it becomes increasingly more difficult to recognize vegetation stands because of increasing heterogeneity with decreasing scale.

3.1.2.1 *Background*

Gabriel & Talbot (1984) define a stand as follows: "A concrete (vs. abstract) aggregation of plants of more or less similar uniformity in physiognomy, species composition, spatial arrangement and condition to distinguish it from adjacent communities. Concrete stands which we sample or measure are aggregated into abstract communities and communities are further abstracted into a general vegetation". Werger (1974) states that "Stands for sampling should be selected in such a manner that each is representative of the vegetation of which it is part and that each plant sampled therein should yield a more or less typical description of that vegetation in terms of both floristic composition and structure". Werger (1974) further states that "Stands which are obviously heteroge-



neous in habitat, structure or floristic composition, and might, therefore, logically be expected to represent parts of two or more units or associations, should be avoided because they do not contribute information which can be used to describe the two or more community types that they represent". From these definitions it can be inferred that a stand is generally a relatively homogeneous unit of vegetation, smaller than the community of which it is a part, but not smaller than the sampling plot with which it is sampled. Furthermore, a stand should be representative of a community and not of an ecotone.

3.1.2.2 *Methods applied*

Stand area is not explicit in the stand definitions given in the previous section but a stand should be sufficiently large to be representative of the community it is to represent. On the other hand, a stand should not be smaller than the smallest mappable unit area (SMUA) because of scale limitations, where it is intended to map communities.

Werger (1972) says that a vegetation stand can fully manifest itself, in terms of species composition and structure, in an area of about one half to one hectare but recommends a much smaller plot area. However, if stand area were to be equated with the SMUA then the following would apply:

- 1: 5 000 scale represents a stand area of 78 m²;
- 1: 8 000 scale represents a stand area of 201 m²;
- 1: 10 000 scale represents a stand area of 314 m²;



1: 20 000 scale represents a stand area of 1257 m²;

1: 50 000 scale represents a stand area of 7854 m²; and

1: 250 000 scale represents a stand area of 20 ha.

Equating stand area with the SMUA permits scale to be taken into account when sampling vegetation, providing that the sampling plot is then representative of the stand. Furthermore, the criteria defined in section 3.1.2.1 can still be applicable.

The stand is, therefore, defined as a circular area within a stratified unit with radius equal to the denominator of the scale, as a representative fraction. Thus, the minimum area for a community is a function of scale and stand heterogeneity is likely to increase with decreasing scale. The stand radius for the main study area is 250 m, representing a stand area of about 20 ha.

3.1.3 Reconnaissance

A reconnaissance is the preliminary inspection of the study area prior to sampling and can have various objectives.

3.1.3.1 Background

According to Werger (1974) "The area of investigation must be well known in all its variety before the study is started". This viewpoint is supported by Mueller-Dombois & Ellenberg (1974) who go even further in stating that "Once the entitation or subdivisioning of the vegetation cover has been clarified, the communities are essentially established". In other words, tentative communities



should be mapped prior to field sampling. The principle of separate mapping and classification processes, in which the classification confirms the mapping units, has been adhered to in this study. However, what is required for the inexperienced researcher are more explicit details of the actions necessary during reconnaissance. Although plant empathy (Tinley 1977) can give added perspective, systematically recorded observations can only increase the scientific value of a reconnaissance.

3.1.3.2 *Methods applied*

The main purpose of reconnoitring the study area is terrain familiarization. This entails travelling throughout the study area to estimate floristic and environmental variation, familiarization with the flora and to obtain permission from landowners for later work and camping sites, if required. Routes followed during the reconnaissance should correspond to the main environmental gradients such as rainfall, topography, geology and soil, where possible, using available roads, in order to detect the range of floristic variation over these gradients. Data recorded for the main study area include:

- i. estimation of the number of structural/floristic units as an indication of the variation in plant communities in the study area. This is required for stratification and need not be precise, but an overestimate is preferable to an underestimate for colour allocation (section 3.1.4). The number of structural/floristic units, can be estimated by counting the different structural/floristic units which intercept the



reconnaissance route. The structural component is according to Edwards (1983) and the floristic component is the dominant or co-dominant species of each structural stratum. Borders of these units can be indicated on a route map, where they intercept the route for stratification validation, if necessary. Scale can be taken into account by only recognizing those structural/floristic units that exceed 4 times the stand radius on the route. This permits a minimum of one stand, with a buffer zone equal to the stand radius, on either side of the stand, to be included in the structural/floristic unit (see section 3.1.4). Aerial photographs could also be used for these estimations;

- ii. estimation of species richness, in terms of species per m^2 , in each structural/floristic unit. This is required for determination of sampling unit area (section 3.1.6). This is done by counting the number of plant species in an area commensurate with plant height and spacing such that any area in the structural/floristic unit would result in a similar count;
- iii. voucher specimen collection of dominant and other common plant species. These are the plants most likely to be encountered during field sampling. The ability to identify these plant species effectively in the field reduces the time required for field sampling (see section 3.2.2); and,
- iv. photographic records and notes on any features commensurate with the aims of the study can also be taken.



3.1.4 Stratification

Stratification is the process of preliminary vegetation mapping prior to vegetation sampling and classification. Generally, the sampling intensity in a community, using the Braun-Blanquet approach, is far too low for the community borders to be defined according to the sampling units, without some form of stratification.

3.1.4.1 *Background*

It can be argued that to map plant communities using the Braun-Blanquet approach is not necessary because each plant community can be recognized in the field using the community diagnosis. This, however, ignores the spatial relationships between communities, is difficult for those not familiar with the techniques to apply, and some sort of vegetation unit map is required, in any case, for sampling the vegetation.

In stratified sampling the vegetation is divided into preliminary vegetation units to which are allocated specific portions of the total sample. Such sampling can actually improve the accuracy of estimates since it is ensured that the various vegetation units that make up the total vegetation are represented in the sample (Freund & Williams 1958).

Such divisions prior to sampling also reduce the problem of redundancy at vegetation unit borders (Werger 1974), and ensure an even



distribution of sampling units in each vegetation unit, irrespective of vegetation unit area.

Stratification for random or systematic sampling can be based on aerial photo interpretation of a study area, as is often the case, as well as mappable non-vegetation factors, such as geology, geomorphology and soil types. Visual integration of different factors, to detect discontinuities in vegetation, is subject to the observer's experience, for repeatability. Furthermore, correspondence between vegetation and non-vegetation (environmental) factors can usually only be shown after the classification process, although correspondence may be implicit, which can limit the usefulness of aerial photo interpretation for stratification purposes. Non-vegetation factors can only be reliably used for stratification, where their limits correspond with those of the vegetation units.

Stratification of vegetation can also be based on satellite imagery. The advantages of satellite imagery over aerial photography include improved geographic fidelity, data in a digitized form, suitable for processing, and suitability for small scale work. The following article describes the methods applied in this study.

A method for vegetation stratification using scale-related, vegetation-enhanced satellite imagery

R. H. WESTFALL* and O. G. MALAN**

Keywords: Landsat, Munsell parameters, South Africa, stratification, Transvaal, vegetation

ABSTRACT

A method for visual vegetation stratification and pattern refinement, using scale-related, vegetation-enhanced satellite imagery, is described. The method simplifies colour assignment, facilitates accurate vegetation mapping and could lead to balanced floristic classifications.

UITTREKSEL

'n Metode vir visuele plantegroei-stratifikasie en patroon verbetering wat van skaalverwante, plantegroei-versterkte satellietbeelde gebruik maak, word beskryf. Die metode vereenvoudig kleurtoekenning en vergemaklik akkurate plantegroei-kartering en kan tot meer gebalanseerde floristiese klassifikasies lei.

INTRODUCTION

In vegetation sampling (Werger 1974) the subjective selection of sample sites, based on vegetation homogeneity, can only be done effectively by an operator with considerable experience. In addition, there is often a lack of repeatability in the methods and a tendency to ignore vegetation dynamics by only sampling those areas representative of 'good' vegetation.

Stratified random sampling overcomes these problems and increases sampling efficiency by ensuring adequate representation of subdivisions (Elliott 1983). Furthermore, in contrast to random or systematic sampling, the heterogeneity of vegetation, in terms of possible number of communities, can be related to the number of stratified units. Stratification also facilitates the avoidance of transitions which generally do not contribute more information than the adjacent communities (Werger 1974).

Stratification of vegetation prior to floristic sampling entails primarily the categorization of vegetation according to structural characteristics. The categories can be further refined according to the factors apparently responsible for differentiating the strata. The most important factor is usually taken to be topography but others such as geology, pedology, climate or combinations of the four may also be decisive.

Problems encountered with vegetation stratification when using aerial photographs include radial distortions, altitude-related scale differences and often inconvenient scales, which do not facilitate precise vegetation mapping. The excessive detail present in aerial photographs can be potentially confusing and time consuming for stratification, especially

for small-scale work. The use of small-scale, almost orthographic satellite imagery for stratification overcomes these problems but introduces the problems of pattern interference by factors such as soil, and colour assignment where patterns are formed by unfamiliar colours and textures. This paper describes a method for vegetation stratification, using satellite imagery that can overcome the problems of pattern interference and colour assignment and facilitates later pattern refinement.

CONVENTIONAL USE OF LANDSAT DATA IN VEGETATION STRATIFICATION

False colour images

The Landsat multi-spectral scanner (MSS) records radiance from the earth's surface in four spectral bands: 500-600, 600-700, 700-800 and 800-1100 nm, usually referred to respectively as bands 4, 5, 6 and 7. These data are obtainable in image or digital form. The ground resolution of the system is nominally 79×56 m which corresponds to a picture element of about $0,30 \times 0,22$ mm at a scale of 1:250 000.

The MSS collects six lines of data simultaneously using a set of six detectors for each spectral interval. The mismatch in these sets is one of the major causes of noise in MSS data, which is apparent as striping with a periodic cycle of six lines at extreme radiometric enhancements.

Traditionally Landsat MSS data have been used in mapping at scales between 1:1 000 000 and 1:250 000 (under special circumstances up to 1:100 000 or even 1:50 000 scale) in the form of false colour images, i.e. bands 4, 5 and 7 displayed respectively as the colour primaries, blue, green and red.

This type of display suffers from two disadvantages: firstly only three of the four bands can be displayed, with a possible loss of crucial information in band 6. Secondly, because of the high degree of correlation between bands (Table 1), a very limited re-

* Botanical Research Institute, Department of Agriculture and Water Supply, Private Bag X101, Pretoria 0001.

** National Physical Research Laboratories, CSIR, P.O. Box 395, Pretoria 0001.

gion of the available three-dimensional colour space is utilized: live vegetation appears exclusively in various shades of red, depending on its structure and vigour.

TABLE 1. — Correlation matrix of the radiance values of the Landsat MSS data, in the four Landsat MSS bands, for the study area

Band	4	5	6	7
4	1,0			
5	0,88	1,00		
6	0,71	0,62	1,00	
7	0,59	0,47	0,94	1,00

Digital multispectral classification

Digital multispectral classification methods have been used successfully in crop mapping. The methods have proved to be a problem in the stratification of natural vegetation, particularly under southern African conditions. The reasons are mainly: rugged topography (crops are normally grown on level fields), heterogeneity of stands with the consequent problems of selecting 'typical' training sites of sufficient size (i.e. several hectares in size) for the extraction of spectral signatures and the interference of soil reflectance caused by incomplete canopy cover.

Furthermore, this method relies exclusively on spectral characteristics and cannot make use of the contextual or contextual information consciously or subconsciously available to the human interpreter.

DIGITAL ENHANCEMENT OF LANDSAT DATA

In the alternative approach of enhancement of the Landsat MSS data by digital image processing followed by visual interpretation, the versatility of the human analyst is assisted by means of quantitative enhancement of vegetation differences in stratification but particularly in pattern refinement.

Principal Component Analysis

Principal component analysis (PCA) provides a convenient method of data compression and removal of redundant correlation between bands (Lasserre *et al.* 1983). Experience has shown that typically more than 99% of the total variance in the data is retained in the first three principal components (Table 2), the maximum which can be accommodated in any colour display.

TABLE 2. — Variance in terms of proportional eigenvalues for the first four principal components of the Landsat MSS data for the study area

Principal Component	First	Second	Third	Fourth
	0,872	0,106	0,016	0,005

The first component represents shadow-enhanced topography and overall terrain brightness differences, whereas the second and third reflect the spec-

tral differences in surface cover. Although the third component normally contains considerably less variance than the second, this may be crucial information for stratification. The fourth component contains predominantly noise.

For optimum results PCA must be based on the statistics of a (composite) subscene, approximately equally representative of the relevant floristic subdivisions, of the area to be stratified. The three principal components may be displayed in any combination of the colour primaries. However, particularly in regions of large variations in overall brightness, such as in areas of rough topography, this results in multicoloured imagery which is difficult to interpret.

A much more informative product can be obtained by displaying the first three components respectively as the Munsell colour parameters, brightness, hue and saturation.

Display in the Munsell colour space

In order to generate a practicable colour display, the components representing the Munsell colour parameters must be converted into the colour primaries, blue, green and red by computation, special care being taken that visual hue differences in the display truly reflect numerical differences in the data (Malan & Lamb 1985). The first and third components are first contrast-stretched to about 1–2% of the data in maximum and minimum values. The distribution of the first component is approximately Gaussian (Figure 1) while a histogram equalization stretch is applied to the second component (Figure 2). These stretch lookup tables must be based on the statistics of the representative subscene used for PCA.

In the final product the overall impression of terrain brightness of the original image is retained, thus facilitating registration with overlays and later pattern refinement. The effect of the histogram equalization stretch of the hue component is to spread the spectral differences of the cover over the complete hue gamut (as modified in saturation by the third component) as opposed to the limited range of hues in the conventional false colour representation (compare Figures 3 & 4).

Filtering

When vegetation mapping is to be done at a scale where the shortest cross distance of a mappable unit is greater than a ground resolution of 79×56 m, the original resolution of the Landsat MSS data could lead to a product with too much small detail. Prior smoothing of the data produces a scale-related product which facilitates stratification and pattern refinement.

The minimum colour-stratified area is given by $12,56 n \text{ mm}^2$ with a shortest cross distance of 4 mm (Rutherford & Westfall 1986). The value of n is determined by the minimum number of samples required for possible floristic subdivision of a colour-stratified area which is dependent on scale as related to sample area and spacing as well as the area of the colour-stratified unit. For example, with $n=4$, full resolution Landsat MSS data could therefore theo-

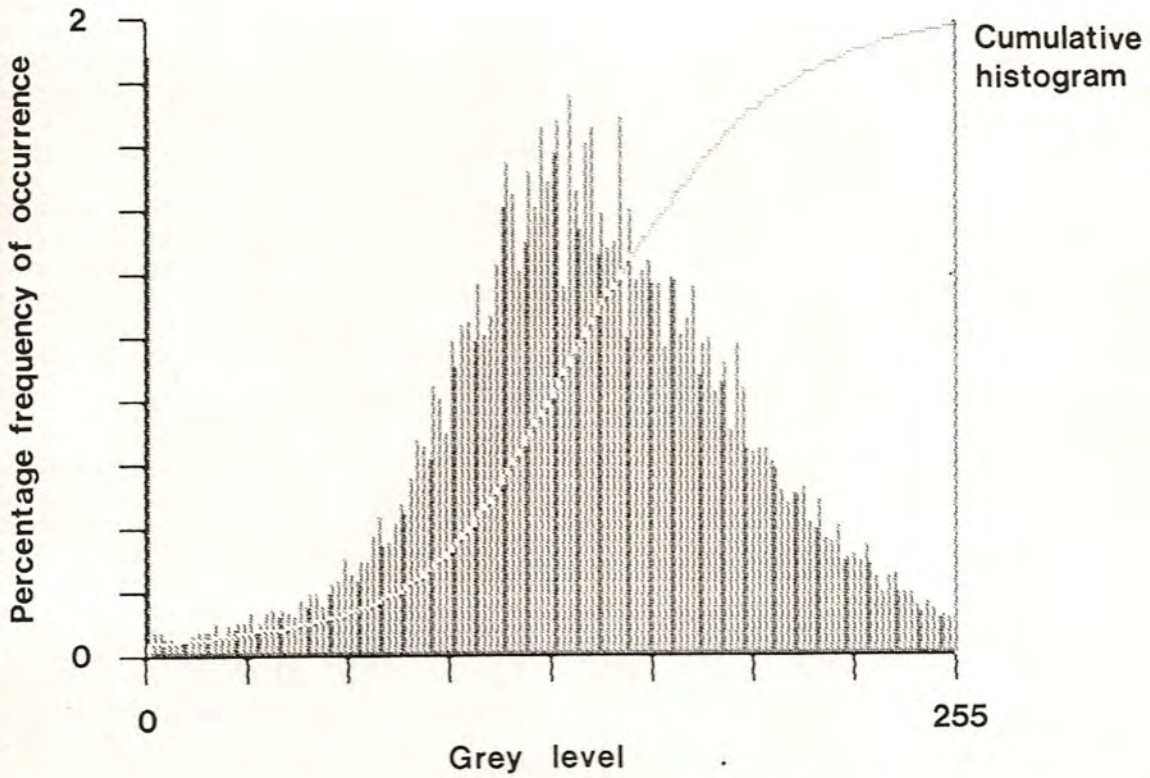


FIGURE 1. — Histogram showing the approximately Gaussian distribution of the first principal component, displayed as the Munsell brightness parameter.

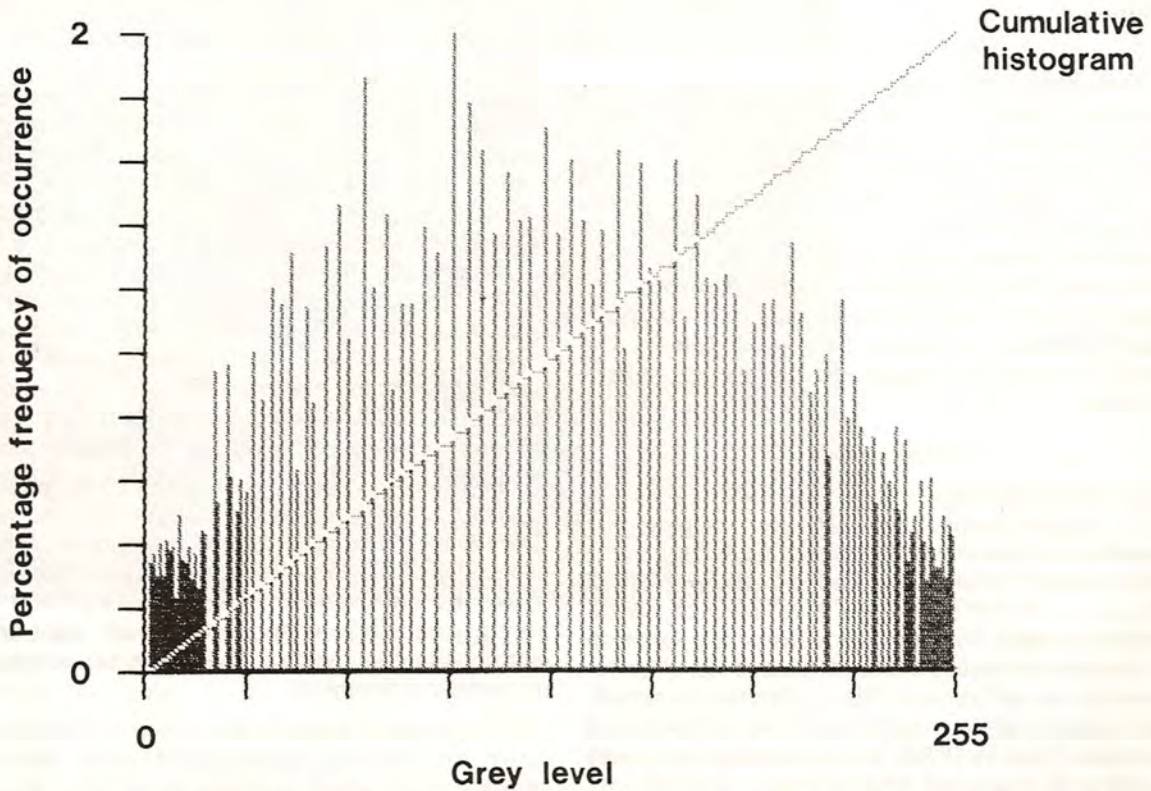


FIGURE 2. — Equalized histogram of the second principal component, displayed as the Munsell hue parameter.

retically be used for vegetation stratification up to a scale of 1:10 000. However, PCA dramatically enhances noise in the data in the higher principal components, particularly the striping in Landsat MSS data. The application of a median filter with a kernel at least six lines wide smoothes the data and removes striping effectively. Consequently, in practice, a median filter with a minimum kernel size of 6×9 picture elements corresponding to a square ground resolution of about 24 ha, is used. This corresponds to a maximum useful working scale of almost 1:50 000.

The (first) brightness component is excluded from the filtering process, to retain fine detail in the representation of topography, which assists in pattern refinement as well as in the identification of ground control points for registration with map overlays.

PATTERN REFINEMENT

Pattern refinement refers to the process of modifying the stratified units, usually after sampling and classification of the vegetation and prior to mapping floristic units. This process includes the grouping together of similar, smaller, discreet areas and subdivision of larger, uniform colour-stratified areas by contextual comparison if required, with suitable, simplified topographical, geological, pedological or meteorological overlays and sample-set classification at the given working scale. The choice of any or all of these overlays is determined by the range of variation exhibited by the overlays that can relate to the structural variation. For example, smaller units of differing colour can be grouped together on the basis of floristics and soils or topography while larger, uniform units could be subdivided on the basis of floristics and geology or climate in order to refine the patterns.

Minor inaccuracies can occur in the registration of Landsat hard copy images and other maps at the same scale used for overlays which are attributed to differential stretch and shrinkage caused by fluctuating humidity as well as some obvious inaccuracies in the maps. These errors can be compensated for by shifting local fit to achieve the maximum number of registration points rather than be compounded by maintaining fixed registration points. However, the effect of stretch and shrinkage can be further reduced by the use of dimensionally stable transparencies of both satellite images and other maps, where available.

RESULTS

The results of colour-stratification using scale-related vegetation-enhanced satellite imagery are given for a portion of the Transvaal Waterberg in the north-western Transvaal (Rutherford & Westfall 1984) at 1:250 000 scale. The area has a highly diverse topography with a consequently high variation in vegetation structure which is inferred from the variation in colour pattern. The vegetation is mainly representative of Sour and Sourish Mixed Bushveld veld types (Acocks 1975). Structural heterogeneities related to topographic diversity were indicative of potentially small stratification units and the value of $n=4$ was accordingly assigned. This corresponds to a

map area of 50 mm² which is equivalent to 320 ha. The appropriate filter kernel size was therefore 32×22 picture elements. A contrast-stretched false-colour MSS image (acquisition date March 1981) is shown in Figure 3. A vegetation-enhanced image, at the original resolution of the same scene, is shown in Figure 4, which illustrates the complexities of colour assignment exacerbated by noise, particularly six line striping. A scale-related, vegetation-enhanced image at the hard copy scale of 1:250 000 of the same scene is shown in Figure 5 with resultant simplification of colour pattern to form stratified units.

INTERPRETATION

A critical prerequisite for successful stratification is the choice of the optimum data acquisition date for maximum differentiation between the structural subdivisions. Because the four wettest months in the study area are from November to February, the data acquisition date of March ensured high vegetation cover with minimal cloud interference.

The inputs required from the user are location of study area, working scale and data acquisition dates of scenes required. Primary colour-stratification is automated after these inputs and is, therefore, objective and time and labour saving. Many researchers use topographic, geologic, pedologic or meteorologic maps for comparison with floristic units or simply to show the variation in these factors. Unless working scale is standardized at the outset of a project, comparisons are difficult. Overlays at a standardized scale are used for pattern refinement and can also be used for later comparisons with floristic units in addition to the enhanced Landsat MSS image. They can, therefore, serve dual purposes and their effective use is increased.

Training sites would often be required for unsimplified images (Westfall & Malan in press), whereas the use of scale-simplified images largely overcomes the need for this training. Furthermore, scale-related stratified units should improve floristic classifications by providing a balanced distribution of sample sites commensurate with vegetation heterogeneity and the amount of detail required for a given working scale. It should be pointed out that, like PCA, the colour stratification by vegetation enhancement is highly scene-dependent. In practice, however, this is not a great disadvantage because one Landsat scene covers 34 000 km². Also multiples of this size can be treated identically if they are contiguous images on one north-south Landsat swath.

The use of satellite images, which have better geometric fidelity than aerial photographs, also simplifies the process of accurate vegetation mapping, especially where first-order stereo-restitution instruments are not available. This ensures greater mapping precision when compared to base maps and facilitates the effective use of overlays for comparison or pattern refinement.

The proposed methods also ensure objectivity because the primary colour-stratification process is computerized, which produces repeatable stratification units. It is doubtful that subjectively stratified units could be repeated by different workers. This

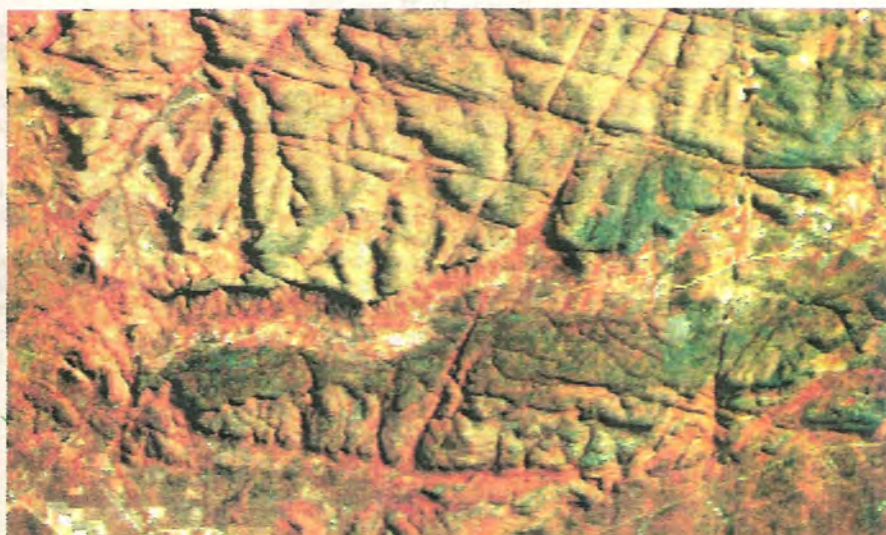


FIGURE 3. — A contrast-stretched, false-colour, multispectral-scanner (MSS) image of a portion of the Transvaal Waterberg with limited range of hues. Scale 1:250 000.

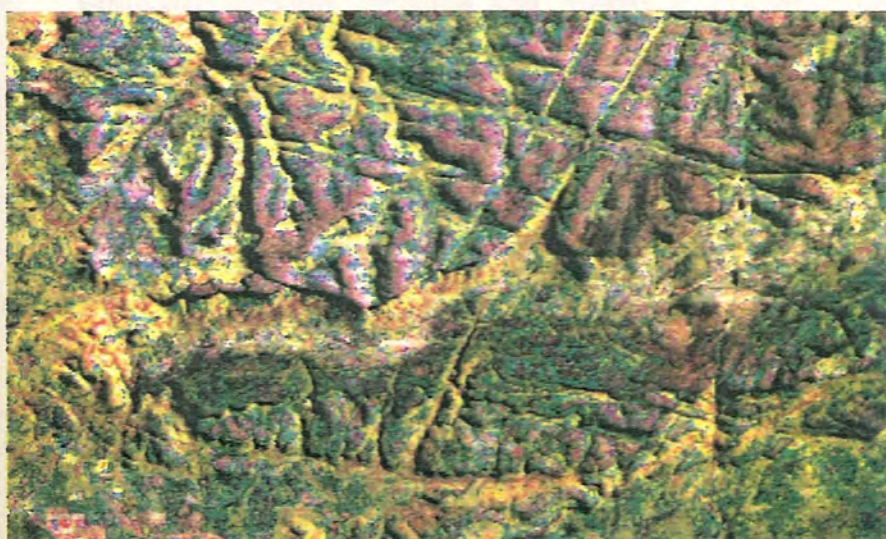


FIGURE 4. — A vegetation-enhanced image at the original resolution of the same scene as shown in Figure 1 using the complete hue gamut.

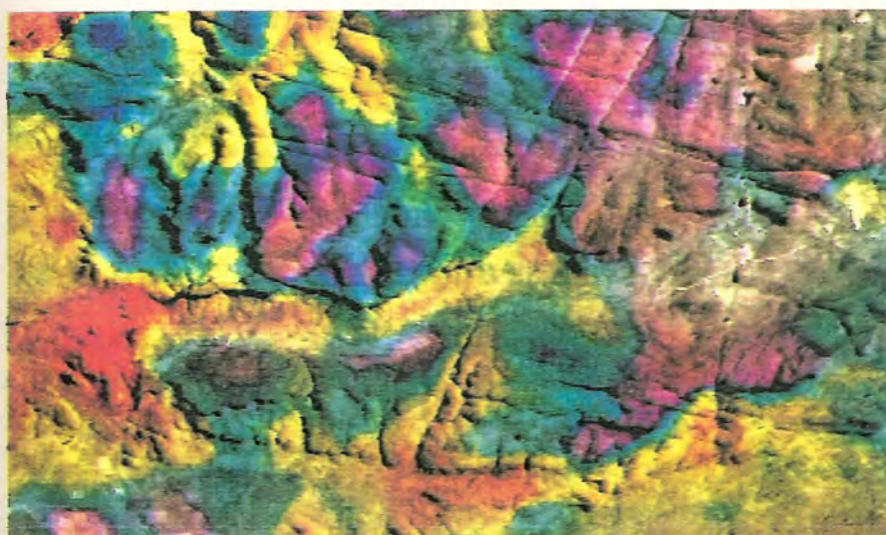


FIGURE 5. — A scale-related vegetation-enhanced image of the same scene as shown in Figure 2 showing simplification of colour pattern to form primary stratified units.

could affect the balance of the resultant floristic classifications and could be especially significant in comparisons over time.

CONCLUSIONS

The proposed methods of vegetation stratification prior to floristic sampling are objective, and time- and labour-saving. The process of colour assignment

to stratified units is simplified. The stratified units are related to working scale and the structural heterogeneity present in the vegetation. Overlay comparison and accurate vegetation mapping are facilitated and more balanced floristic classifications can be expected. Landsat data are generally more detailed than required for vegetation stratification at scales smaller than 1:50 000 and hence filtering is necessary.



ACKNOWLEDGEMENTS

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

REFERENCES

ACOCKS, J. P. H. 1975. Veld types of South Africa. *Memoirs of the Botanical Survey of South Africa* No. 40.

ELLIOTT, J. M. 1983. *Some methods for the statistical analysis of samples of benthic invertebrates*. Kendall, Wilson.

LASSERRE, M., MALAN, O. G. & TURNER, B. 1983. *The application of principal component analysis to Landsat MSS data*. Proceedings of seminar on Principal component analysis in the atmospheric and earth sciences, Pretoria 7-8 February 1983.

MALAN, O. G. & LAMB, A. D. 1985. *Display of digital image data; quantitative and optimal*. Proceedings of the third

South African symposium on Digital image processing, Durban 22-23 July 1985.

RUTHERFORD, M. C. & WESTFALL, R. H. 1984. Sectors of the Transvaal Province of South Africa. *Bothalia* 15: 294-295.

RUTHERFORD, M. C. & WESTFALL, R. H. 1986. Biomes of southern Africa — an objective categorization. *Memoirs of the Botanical Survey of South Africa* No. 54.

WERGER, M. J. A. 1974. On concepts and techniques applied in the Zürich-Montpellier method of vegetation survey. *Bothalia* 11: 309-323.

WESTFALL, R. H. & MALAN, O. G. in press. *A comparison of vegetation units derived from vegetation-enhanced satellite imagery with the vegetation units derived from the floristic classification of the farm Groothoek, Thabazimbi District*. Proceedings of the symposium on Pattern recognition in remote sensing and geophysics.



3.1.5 Stand location

The main disadvantage of stratified systematic sampling is the possibility of hidden periodicities or recurring phenomena in the data, which could influence results (Freund & Williams 1958) so stratified random sampling is preferred. In stratified random sampling the location of stands within each stratified vegetation unit is random, to reduce observer bias. Randomization of stand localities does not necessarily infer statistical validity on the sampling process because the number of repetitions in vegetation sampling within a vegetation unit is generally too few for statistical tests that are based on normal distributions.

3.1.5.1 *Background*

Randomization of stand localities is generally achieved by a grid co-ordinate system in which co-ordinates are selected using random tables or a random number generator. Whatever system of co-ordinate selection is used, the appropriate minimum spacing between stand sites, and hence grid size, should be known, because Werger (1974) states that contiguous stands are not to be recommended as much unnecessary work would be done. Apart from non-contiguous sampling, the literature does not appear to address the problem of minimum spacing of stands.

3.1.5.2 *Methods applied*

From the geometric considerations given in section 3.1.1, a grid

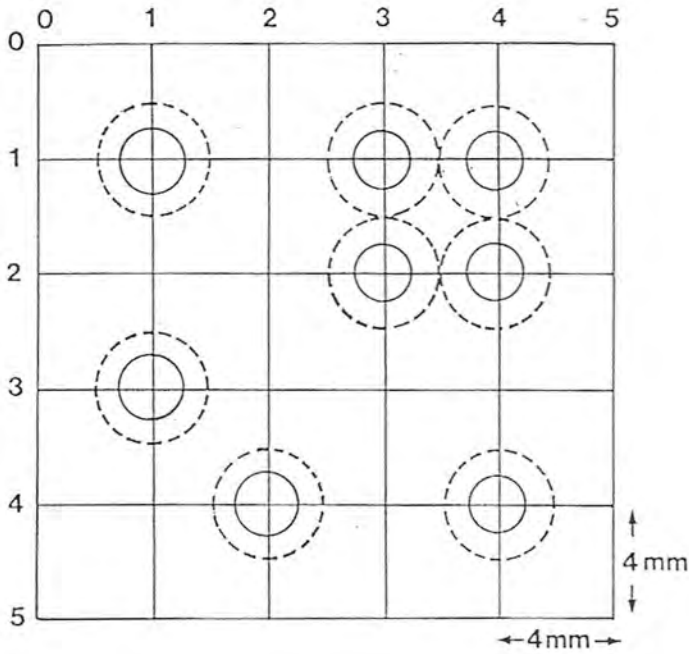


FIGURE 3.1. - A 4 x 4 mm numbered grid for map overlay to determine possible stand sites at the grid intersections. Solid circles represent stands; and dotted circles represent buffer zones around stands. Buffer zone radius = 2x stand radius = maximum ecotone width = minimum distance between stands (from Rutherford & Westfall 1986).



size of 4 x 4 mm should be used, at any given working scale, as a map overlay to locate stand sites from random co-ordinates. This ensures a minimum spacing of one stand diameter between stands. Stand area in relation to the grid is illustrated in Figure 3.1. The use of a standard grid, for any working scale, can simplify computer generation of random numbers for selecting grid co-ordinates and their conversion to the co-ordinates of degrees, minutes and seconds, to facilitate stand location in the field. In the field, if a randomly selected stand includes obvious heterogeneity such as disturbance not commensurate with the working scale, then it is ignored and the following stand is selected. Stands can also be ignored in the field if they are inaccessible. Furthermore, stands may not be used if the circumference is less than one stand diameter from a stratified unit border. It is for these reasons that more stand localities are generated than are actually required. Observer bias is, therefore, reduced by decision-making regarding stand location, although some decisions may be necessary regarding stand suitability. The following article describes a program for the generation of random stand co-ordinates to be used with a 4 mm grid and grid co-ordinate conversion to degrees, minutes and seconds for field location of stands.



PHYTOLOC — A RANDOM-NUMBER GENERATOR AND SAMPLE-SET LOCATION PROGRAM FOR STRATIFIED RANDOM VEGETATION SAMPLING

The stratified random method of vegetation sampling is objective and efficient, in terms of sample-set distributions, for floristic classifications (Westfall & Malan 1986). However, the commonly used random number tables and calculator-generated random numbers often require number abbreviation and manual recording of the numbers, which can be time-consuming. But these inconveniences are insignificant when compared with the time taken to measure the location of random sample sets and express their location in terms of the latitude and longitude co-ordinates of degrees, minutes and seconds. The PHYTOLOC program was developed to generate random numbers for sample set location in terms of random co-ordinates and, in addition, to express these co-ordinates as latitudes and longitudes in degrees, minutes and seconds, thereby saving considerable time and effort.

The program is written in Basic and runs on a Sharp PC 1500 computer. Use is made of a consecutively numbered 4 mm transparent grid map overlay which is related to any working scale, in terms of sample set spacing and size (Rutherford & Westfall 1986). Grid overlay registration with a base map is according to the zero co-ordinates of the overlay with the intersection of minimum latitude and longitude of the study area on the base map, as well as with the zero x-axis of the overlay with the minimum latitude of the base map. Inputs required for the program are:

- (i) maximum latitude of the study area in decimal degrees;
- (ii) minimum latitude of the study area in decimal degrees;
- (iii) minimum longitude of the study area in decimal degrees;
- (iv) difference in millimetres, between minimum and maximum latitudes, at the given working scale;
- (v) mean distance in millimetres, between 1 minute longitudes at the minimum latitude;
- (vi) mean distance in millimetres, between 1 minute longitudes at the maximum latitude;
- (vii) number of sample sets required, estimated by $10 \text{ SU} + (0,25 \times 10 \text{ SU})$ where SU is the number of stratified units. This should generally allow for omissions due to transitions and proportionality;
- (viii) upper limit (integer), within the study area, of the x-axis of the grid overlay, and
- (ix) upper limit (integer), within the study area, of the y-axis of the grid overlay.

The program generates and prints random numbers for the x- and y-axes of the grid overlay and computes and prints the equivalent values in degrees, minutes, seconds and decimal fractions of seconds for longitude and latitude respectively. Convergence of longitude is also taken into account. Each set of co-ordinates, representing a potential sample site, is numbered consecutively. In addition the means and standard deviations of the x- and y-arrays are computed to show the statistical distribution of potential sample sites. Co-ordinates are transferred to the base map using the grid overlay and the print-out can be used for field allocation of latitude and longitude to the field data sheets. These co-ordinates

can also be used on larger-scale maps for more precise field location of sample sites.

For valid categorization and analysis of floristic units, based on multivariate data, a minimum of four sample sets are required, although a single sample set is mappable at the given working scale. Consequently for a single floristic division of a stratified unit a minimum of eight sample sets would be required. However, for statistical comparisons of univariate data such as biomass or number of taxa, sample-set number should be proportional to area (Elliott 1983). It is, therefore, suggested that a sampling intensity of 2,5% of the potential sampling sites (i.e. total number of co-ordinate interceptions) within a stratified unit should be maintained to ensure proportionality. This is approximately commensurate with the relationship of study area to sample number (Rutherford & Westfall 1986) but modified by vegetation heterogeneity in terms of number of stratified units.

A non-random set of sample sites could, therefore, be required to fulfil the categorization and analysis requirements for stratified units with less than a total of 320 interception points. The additional non-random sample sites can be selected objectively by:

- (i) using best fit of additional sample sets for areas equal to those of eight or less interception points (i.e. 100% sampling intensity), and
- (ii) using additional random sample sets, to ensure representation of vegetation variation, for areas equal to those with between 8 and 320 interception points. (i.e. > 2,5% but < 100% sampling intensity).

Additional random sample sets can be obtained together with the relevant co-ordinates, if required, by the same procedure, but with each relevant stratified unit registered separately on the 4 mm grid.

ACKNOWLEDGEMENTS

The author thanks Drs J. C. Scheepers and H. van Ark for comments and suggestions.

REFERENCES

- ELLIOT, J. M. 1983. *Some methods for the statistical analysis of samples of benthic invertebrates*. Kendall, Wilson.
- RUTHERFORD, M. C. & WESTFALL, R. H. 1986. Biomes of southern Africa — an objective categorization. *Memoirs of the Botanical Survey of South Africa* No. 54.
- WESTFALL, R. H. & MALAN, O. G. 1986. A method for vegetation stratification using scale-related, vegetation-enhanced satellite imagery. *Bothalia* 16: 263-268.

R. H. WESTFALL



3.1.6 Sampling unit area

The sampling unit is the vegetation sample plot used to sample a vegetation stand and in the classification phase is referred to as a relevè (Mueller-Dombois & Ellenberg 1974). Where more than one plot is used to sample a vegetation stand then the sampling unit is a sub-plot.

3.1.6.1 Background

Werger (1972), in his review of species-area relationships, concludes that no convincing minimal-area for sampling unit size has been formulated and regards the concept unsuitable for ascertaining a suitable sampling unit size for sampling vegetation. He further suggests a sampling unit size much smaller than half a hectare which depends on species richness and structure of stands. This may be an extreme statement but in southern Africa, sampling unit area can vary considerably. For example Coetzee (1975) used 10 x 10 m; Coetzee *et al.* (1976) used 10 x 20 m for herbaceous plants and variable sized units for woody plants; Bredenkamp & Theron (1978) used 16 m² for grasslands and 100 m² for woodlands; Van der Meulen (1979) used 10 x 20 m; Van Rooyen *et al.* (1981) used 10 x 20 m and variable size for the woody component; Westfall *et al.* (1985) used 10 x 20 m for all plants and Le Roux *et al.* (1988) used 10 x 20 m for herbaceous plants and 100 x 100 m for trees and shrubs.

Such variation in sampling unit area can be expected to complicate



comparisons within and between communities from different studies. Furthermore, difficulty was experienced in selecting a single 10 x 20 m sampling plot to represent a 20 ha stand. A sub-sampling approach was, therefore, adopted to sample stands. It must be emphasized that the sub-samples are based on minimum area for recording species presence only and although the total sample area for a stand is less than 200 m² the species number recorded was often more than that which could be expected from a 10 x 20 m plot.

The following article describes the methods applied in this study to determine minimum sub-sample area.

Predictive species – area relations and determination of subsample size for vegetation sampling in the Transvaal Waterberg

R.H. Westfall*, J.M. van Staden and M.D. Panagos

Botanical Research Institute, Department of Agriculture and Water Supply, Private Bag X101, Pretoria, 0001 Republic of South Africa

Accepted 20 August 1986

An expression for predicting the number of species in a given area is described. Derivatives of this expression, to increase sampling efficiency in a vegetation stand, include a minimum of four separate subsamples; a maximum number of subsamples when less than 10% increment in new species is achieved; and subsample size. It is also suggested that species diversity in terms of species per unit area can be more consistent when derived from this expression.

'n Uitdrukking om die aantal spesies vir 'n bepaalde gebied te voorspel word beskryf. Afleidings van hierdie uitdrukking, om doeltreffendheid by die monsterneming van 'n plantegroei stand te verbeter, sluit die volgende in: 'n minimum van vier afsonderlike submonsterpersele; 'n maksimum aantal submonsterpersele wanneer die nuwe spesiesaanwas minder as 10% is; en submonsterperseelgrootte. Dit word ook voorgestel dat spesies diversiteit, ooreenkomstig spesies per eenheidoppervlakte, meer konsekwent kan wees wanneer dit van hierdie uitdrukking afgelei word.

Keywords: sample size, savanna, species number, Transvaal, vegetation sampling

*To whom correspondence should be addressed

Introduction

The vegetation ecology of the Transvaal Waterberg is currently being investigated at a scale of 1:250 000 (Westfall, in prep.). The smallest mappable unit area or vegetation stand, of which a sample should be representative, is determined by scale (Rutherford & Westfall 1986) and is 20 ha for this study. A single sample of this dimension would be prohibitive both in terms of cost and time. Random subsamples within the stand offer an objective method of obtaining a representative sample. The efficiency of the subsamples, in terms of plant species recorded, can be determined if the total number of plant species in the stand can be estimated.

The object of this study is to determine the optimum subsample size, in terms of reduced effort and improved efficiency, for sampling the vegetation of the Transvaal Waterberg.

Methods

Vegetation analysis

Two observers, A and B, counted species in eight samples representing four very different vegetation types, namely closed grassland (sample 1); open woodland (sample 2); closed woodland (samples 3 and 4) and unidominant forest (samples 5 and 6). In samples 1 and 2 each observer repeated the countings of the other observer for the same area, each without knowledge of the other's results. In each sample cumulative plant species totals were recorded in 14 nested rectangular subsamples, each with a width-to-length ratio of 1:2 (Figure 1). Subsamples size increased from 62,5 mm × 125 mm (0,0078125 m²) to 10 m × 20 m (200 m²) and the length of each subsample was perpendicular to the contour.

In addition, both observers jointly counted the observable species present in a 20-ha stand of which sample 2 was representative. In this stand nested circular subsamples were also used with radii of 6,3 m (128 m²), 12,7 m (512 m²), 25,5 m (2 048 m²), 51 m (8 192 m²), 102 m (32 768 m²), 204 m (131 070 m²) and 252 m (200 000 m²).

Synthesis

The data in the form of species number for a given area were tested for best fit with various curves (Parton & Innis 1972) as well as linear relationships. The curve which best fitted the data is described by the function:

$$f(x, a, b, c, d) = \frac{a}{(1 + b/x)} + \frac{c}{(1 + d/x)}$$

where a and c = the parameters which control the maximum value of the function and b and d = the parameters which control the rate at which the function approaches its maximum value (Parton & Innis 1972).

No parameters could be found with actual or derived values similar to the computed values for parameters a, b, c & d. This reduced the application potential of the function and it was accordingly not further applied.

The linear regression which best fitted the data is described by:

$$y = e^{(m \ln x + \ln c)}$$

where y = number of species for a given area (x), m = slope (rate of species increase for increasing area), x = area (m²), c = number of species in 1 m².

This linear regression was then used to derive the number of species in 1 m² from a given area and to synthesize data for illustrating species – area relationships.

The validity of the relationships was tested by using data pertaining to the first three relevés recorded in the Transvaal Waterberg (Westfall, in prep.) which utilized these relationships.

Results

The number of species recorded by observers A and B for the nested samples 1 to 6, with predicted values, are given in Table 1 and Figure 2. The linear regressions of these samples are illustrated in Figure 3. With the exception of sample A6, the slopes are generally similar. Sample A6 represents a small stand of unidominant *Podocarpus latifolius*

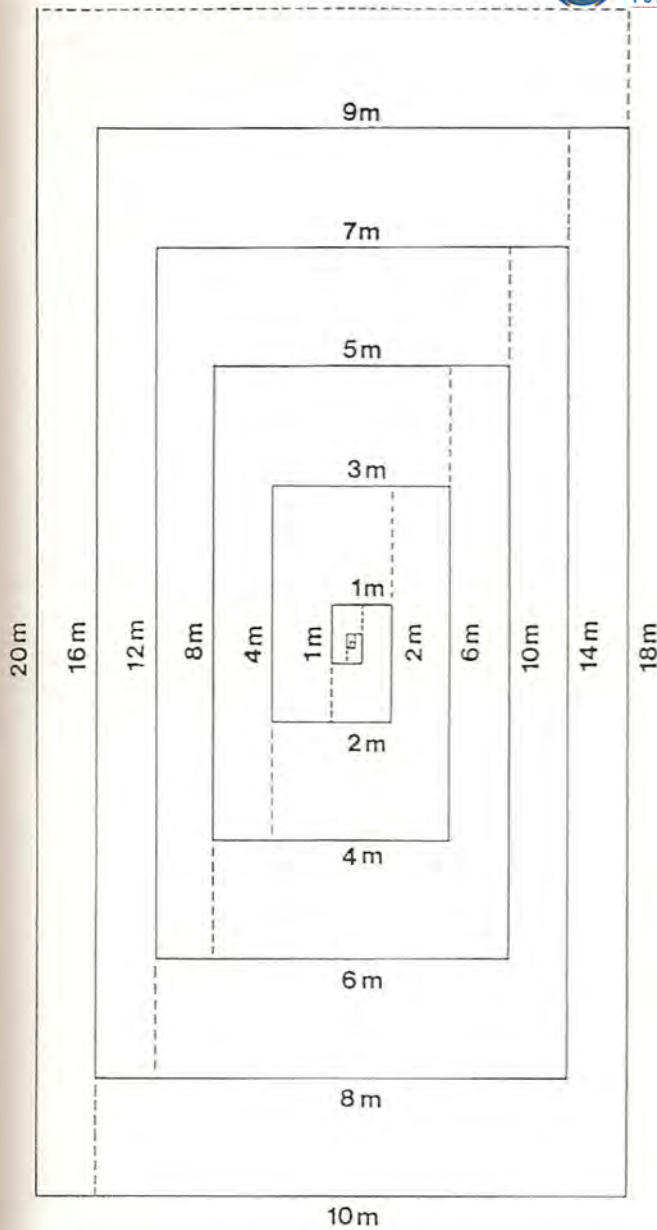


Figure 1 Layout of rectangular nested subsamples. Subsample sizes are: 62,5 mm \times 125 mm; 125 mm \times 250 mm; 250 mm \times 500 mm; 500 mm \times 1000 mm; 1 m \times 2 m; 2 m \times 4 m; 3 m \times 6 m; 4 m \times 8 m; 5 m \times 10 m; 6 m \times 12 m; 7 m \times 14 m; 8 m \times 16 m; 9 m \times 18 m; and 10 m \times 20 m.

(Thunb.) R. Br. Ex Mirb. in which forest-margin species were recorded before the maximum sample size of 200 m² was reached and can therefore be considered atypical. The mean slope of the regressions, excluding sample A6, is 0,34 but is taken as 0,3 which was used for calculating the predicted values in Tables 1 and 2. The mean correlation co-efficient for the recorded and predicted values for the seven samples (excluding sample A6) is $R = 0,9860$, whereas the mean correlation co-efficient for the corresponding recorded and predicted values using the curve function is $R = 0,9819$. The difference between these correlation co-efficients is not significant. The correlation co-efficient for recorded and predicted values in the 20-ha sample is 0,9725 (Table 2).

The value of c is derived from $y = e^{(mlnx + lny)}$ and is given by

$$c = e^{(-mlnx + lny)}$$

which is dependent on the species diversity of the vegetation

type concerned. The values of c , derived from different areas and corresponding derived species number are given in Figure 4.

The derived relationship between number of subsamples and percentage new species increment is given in Figure 5 and the relationship between subsample number and percentage accumulated new species total is given in Figure 6.

The results of the 20-ha sample together with predicted values is given in Table 2.

Discussion

The prediction of number of species for a given area has been tested for a wide range of vegetation types with satisfactory results. The larger difference between recorded and predicted number of species for areas greater than 2 048 m² (Table 2) can be attributed to the difficulty in observing all species in these areas. Destructive sampling would be necessary to ensure the recording of all species. Both observers encountered increasing difficulty in observing all species with increasing sample size where sample size was greater than 20 m². With nested subsamples, however, species omitted in a particular subsample were often recorded in a larger subsample. This is illustrated by the similarity in totals obtained by observers A and B for 200 m² and the differences in totals for 18 m² to 128 m² in samples 1 and 2 (Table 1).

The effect of sampling two highly distinct vegetation types in a small sample is shown in Table 1 (A6) where a greater proportion of new species was recorded from 8 m² than in the other samples. This is also illustrated in Figure 3 where the slopes (m) of all samples except A6 are similar regardless of vegetation type. Local heterogeneity in the 20-ha sample did not influence the results as much as with the smaller sample. This can be attributed to the logarithmic increase in area used in the prediction expression where slope is not significantly affected by a greater proportion of new species in a large area. It is, therefore, essential for predictive purposes that the vegetation should be relatively homogeneous for a sampling area of at least 200 m².

For a given vegetation type, the value of c in the expression should be more or less constant when derived from different areas and a horizontal line could, therefore, be expected when these values are plotted against the area from which they are derived. However, in Figure 4 the values of c increase to an area equivalent to 2 m² before remaining more or less constant. This increase can be attributed to the edge effect of small subsample size where the proportion of species intercepted by the subsample border to species within the subsample is greater than for larger subsamples. It is difficult to record fractions of species present so that in practice intercepted species are often ignored. Subsample size should, therefore, be greater than or equal to 2 m² in the vegetation concerned to reduce this edge effect. Similarly the value of c should also be derived from an area greater than or equal to 2 m² in the vegetation concerned which will also ensure greater precision in comparability of species diversity in terms of species per unit area.

The predicted increment of new species as a percentage of the total number of species for increasing number of subsamples (Figure 5) has maximum inflexion for four subsamples or less or an increment of 8,3%, or greater for given c values and subsample sizes. This indicates an optimum efficiency of four subsamples for vegetation sampling or an increment of 8,3%. With additional subsamples only relatively small increments decreasing from 6,5% could be expected (Figure 5). In practice this could be taken as an increment

Table 1 The number of species recorded by observers A and B for nested quadrats 1 to 6 with the predicted values* in brackets and correlations with observed and predicted values

Area (m ²)	A1	B1	A2	B2	B3	A4	B5	A6
0,00781	2 (4)	2 (3)	1 (3)	1 (3)	0 (2)	0 (3)	0 (0)	0 (0)
0,03125	4 (5)	3 (4)	1 (4)	3 (4)	0 (3)	1 (5)	0 (1)	0 (1)
0,12500	5 (8)	4 (6)	4 (6)	5 (6)	3 (5)	2 (8)	0 (1)	0 (1)
0,50000	9 (12)	8 (10)	8 (10)	9 (9)	7 (7)	8 (12)	1 (1)	1 (1)
2,0	17 (18)	15 (15)	13 (15)	14 (13)	10 (11)	14 (18)	1 (2)	2 (2)
8,0	28 (28)	22 (22)	22 (22)	20 (20)	17 (17)	27 (27)	3 (3)	3 (3)
18,0	37 (36)	22 (28)	26 (28)	24 (26)	19 (22)	34 (34)	3 (4)	8 (4)
32,0	38 (42)	27 (33)	34 (33)	26 (30)	25 (26)	36 (41)	3 (5)	8 (5)
50,0	44 (49)	34 (38)	38 (38)	35 (35)	28 (29)	40 (47)	4 (5)	12 (5)
72,0	46 (54)	35 (43)	46 (43)	39 (39)	35 (33)	49 (52)	4 (6)	16 (6)
98,0	50 (59)	40 (47)	49 (47)	41 (42)	39 (36)	52 (57)	6 (6)	24 (6)
128,0	53 (64)	48 (51)	52 (51)	43 (46)	47 (39)	57 (62)	8 (7)	26 (7)
162,0	55 (69)	55 (54)	54 (54)	49 (49)	51 (42)	60 (67)	9 (7)	30 (7)
200,0	57 (74)	57 (58)	55 (58)	52 (53)	55 (45)	61 (71)	9 (8)	33 (8)
Slope	1,199	1,022	0,945	1,002	0,791	1,044	0,773	0,207
Correlation co-efficient (R)	0,9891	0,9894	0,9970	0,9971	0,9910	0,9948	0,9434	0,9365
Value of lnC**	2,7085	2,4673	2,4673	2,3720	2,2095	2,6721	0,4749	0,4749

*Prediction according to: $y = e^{(m \ln x + \ln C)}$ where y = predicted number of species, $m = 0,3$ (slope), x = area (m²)

**C = species number for 1 m² derived from 8 m²

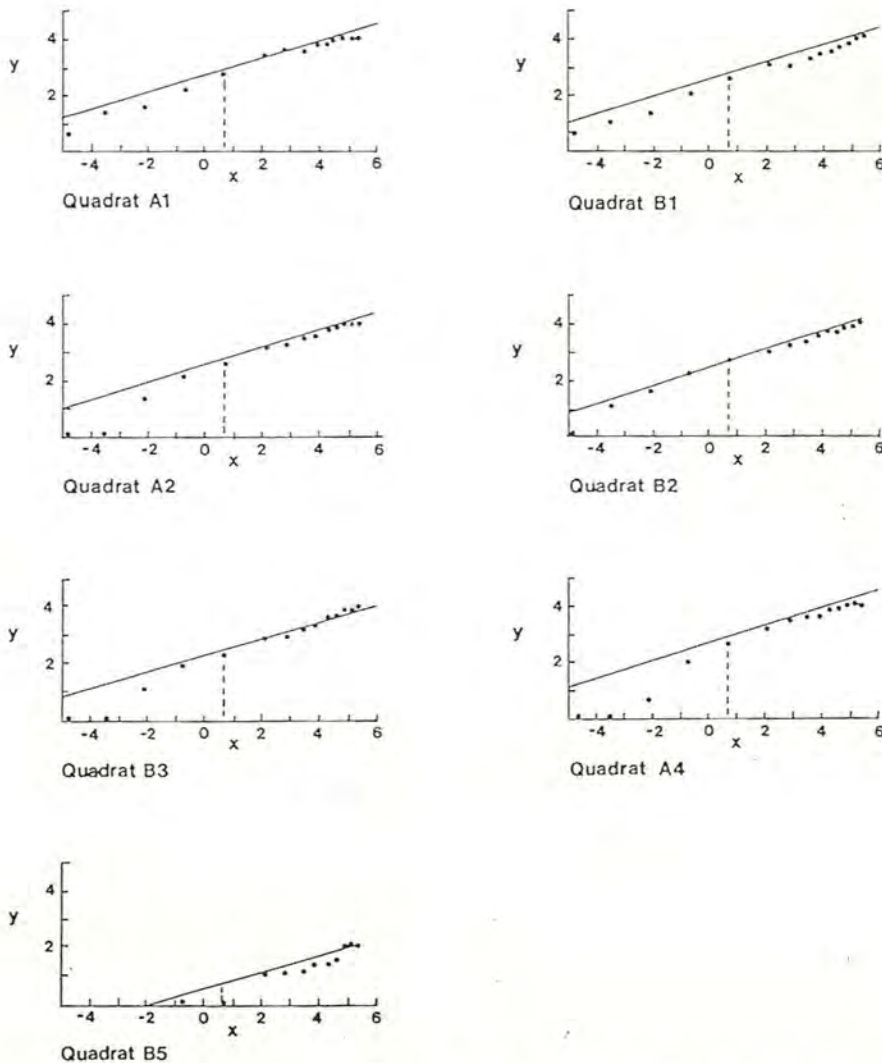


Figure 2 The relationship between ln area (m²) (x-axis) and ln number of species (y-axis) showing observed numbers (dots) and predicted values (regression lines). Generally, fewer species are observed in quadrats smaller than 2 m² (broken lines) than predicted, which is attributed to edge effect. Observer difficulty in recording all species is indicated by the trend of fewer observed species than predicted for the larger quadrat sizes (greater than 20 m²).

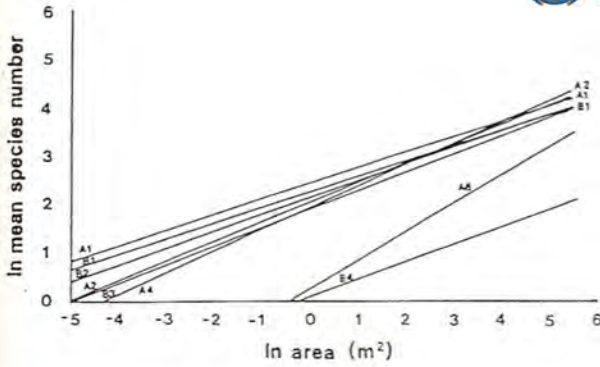


Figure 3 The relationship between ln species number and ln area for eight samples, showing the similarity in slope (*m*). Sample A6 is atypical.

Table 2 The number of species recorded in nested subsamples with predicted values in brackets ($C = 2,5147$ derived from 128 m^2)

Area (m^2)	Number of species
128	53 (53)
512	82 (80)
2 048	120 (122)
8 192	165 (184)
32 768	199 (280)
131 072	213 (424)
200 000	274 (481)

Correlation co-efficient (*R*) between recorded and predicted values: $R = 0,9725$

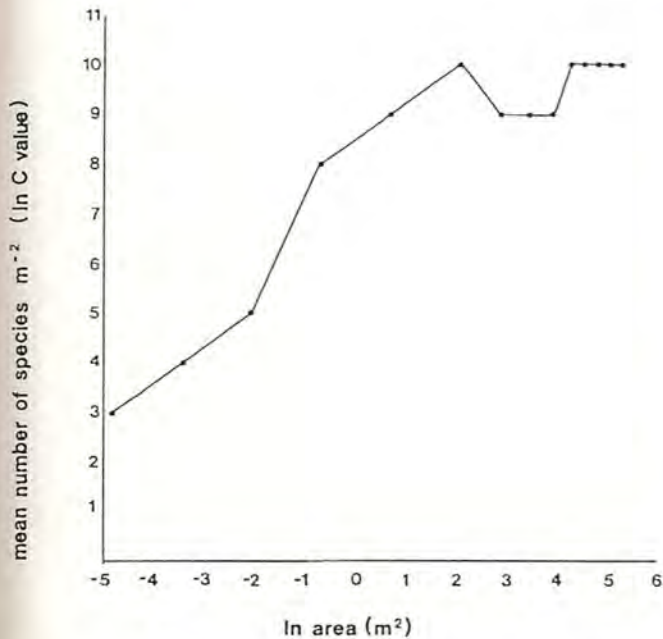


Figure 4 The relationship between ln *c* values, derived from different areas and corresponding ln area showing the increase in *c* values to 2 m^2 ($\ln 2 = 0,693$) representing sample edge effect.

of less than 10% to allow for local heterogeneity. A minimum of four subsamples is recommended because if an increment of less than 10% is achieved with fewer subsamples local disturbance could be indicated.

Species distribution in a stand varies from closely spaced species (high frequency) to 1 species in the stand (low

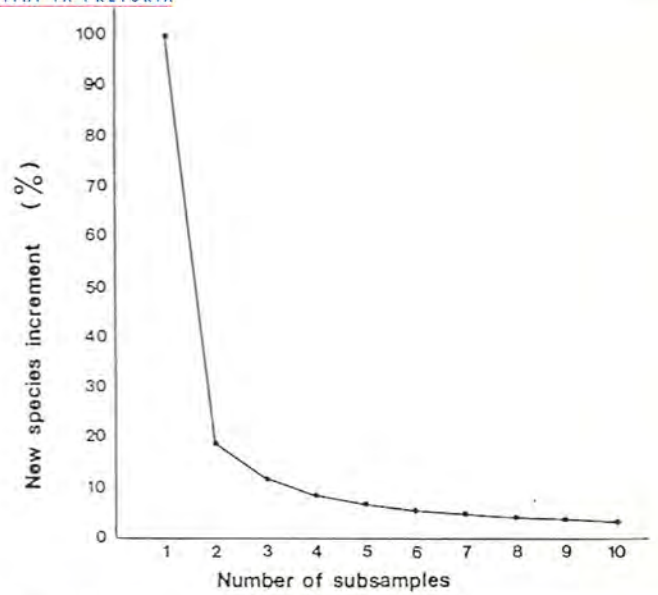


Figure 5 The relationship between number of subsamples and new species increment expressed as a percentage of the accumulated total. Subsample increments: 1 = 100%; 2 = 18,8%; 3 = 11,5%; 4 = 8,3%; 5 = 6,5%; 6 = 5,3%; 7 = 4,5%; 8 = 3,9%; 9 = 3,5% & 10 = 3,1%.

frequency). Too small a subsample size would not include sufficient high frequency species for the relationship in Figure 5 to be valid as each subsample would include only a fraction of the high frequency species resulting in a greater number of new species for successive subsamples. In contrast, too large a subsample size would include too many of the less frequent species with the same result. The percentage difference in the increment between subsamples 3 and 4 is $11,5 - 8,3 = 3,2\%$. This difference should be reflected in a limited number of new species. If the number is only a fraction, then the subsample size is too small and conversely, if too large then the subsample is too large. For convenience and to allow for some local vegetation variation 3,2% difference can be taken to represent 3 new species. This difference can also be expressed as percentage accumulated totals between subsamples 4 and 3 where the first subsample is taken as 100%: $151,6 - 139 = 12,6\%$ (Figure 6). The number of species (*y*) in the first subsample with difference of 12,6% = 3 new species between subsamples 3 and 4 is, therefore:

$$\frac{139}{100} y + 3 = \frac{151,6}{100} y$$

$$y = \frac{3}{0,126}$$

$$= 23,81 \text{ species}$$

If $c = 12 \text{ species m}^{-2}$ then the area (*x*) corresponding to 23,81 species is:

$$x = e^{\frac{(\ln y - \ln c)}{m}}$$

$$= 9,82 \text{ m}^2$$

Subsample size is, therefore, $9,82 \text{ m}^2$. This relationship also takes the diversity of the vegetation into account because subsample size increases as diversity (*c* value) decreases. The subsample size is also within the range of: greater or equal to 2 m^2 , for reduced edge effect; and less than 20 m^2 for observer efficiency, in the vegetation concerned in this study. For less diverse vegetation, maximum subsample size could be increased if required. To ensure a width to length ratio

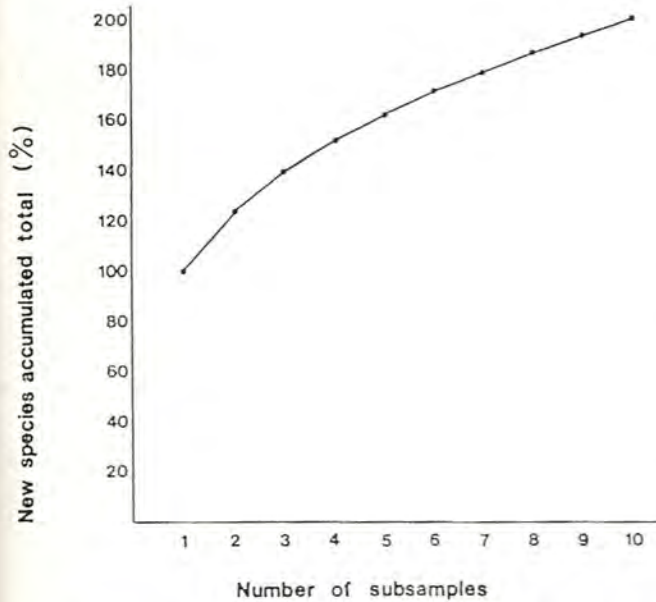


Figure 6 The relationship between number of subsamples and accumulated total of new species expressed as a percentage. Subsample accumulated totals: 1 = 100,0%; 2 = 123,2%; 3 = 139,0%; 4 = 151,6%; 5 = 162,1%; 6 = 171,1%; 7 = 179,2%; 8 = 186,5%; 9 = 193,3% & 10 = 199,5%.

of 1:2 for rectangularity of subsamples (Mueller-Dombois & Ellenberg 1974) the following convenient sizes can be used:

- 1 m × 2 m (2 m²)
- 1,5 m × 3 m (4,5 m²)
- 2 m × 4 m (8 m²)
- 2,5 m × 5 m (12,5 m²)
- 3 m × 6 m (18 m²)
- 3,5 m × 7 m (24,5 m²)

The nearest convenient size which is not less than 9,82 m² is, therefore, 2,5 m × 5 m (12,5 m²). In the first three relevés recorded in the investigation of the vegetation ecology of the Transvaal Waterberg (Westfall, in prep.) using this subsample, size increments of less than 10% were reached with four subsamples in two cases and with five subsamples in one case. These results substantiate the validity of the subsample number and area relationships. Species recorded in each relevé varied from 38 to 60 and sample areas covered varied from 50 to 62,5 m². This is a considerably higher number of species than that recorded by Westfall (1981) in vegetation with similar diversity using single 200 m² sample areas. Separate random subsamples, therefore, appear to increase efficiency. The smaller area also facilitates cover estimation of the herbaceous stratum. It is noteworthy that in all three relevés (Westfall, in prep.) only approximately 10% of the total number of

species estimated to be present in each 20-ha stand was recorded. Recording of all species within the stand would have been prohibitive in terms of time and effort. It can, therefore, be assumed that the remaining 90% of the species have a very low frequency and that the so-called 'infrequent species' in phytosociological classifications are often not as infrequent as assumed. The method proposed here is also objective in distinguishing species, which on a frequency basis, should be more relevant in classifications. In randomly selected subsamples each species chance for inclusion in the subsamples is proportionate to its frequency.

The value of 0,3 for slope (*m*) can be refined with more data for improved predictions especially for areas greater than 20 ha.

The expressions described in this paper, as well as those describing area, sample number and scale relations (Rutherford & Westfall 1986) have, for convenience, been programmed for the SHARP PC 1500 computer. This program, titled VEGFORM, is available from the first author on request.

Conclusions

The advantages of expressing species diversity in terms of the *c* value include:

- (1) overcoming the edge effects of small quadrat size in certain vegetation types; thereby
- (2) improving comparability of species diversity when expressed as species per m²;
- (3) associating subsample size with species diversity;
- (4) determining number of subsamples required for sampling a vegetation stand; and
- (5) predicting the approximate number of plant species which could occur within a vegetation stand.

Acknowledgements

The authors thank Dr J.C. Scheepers, Dr M.C. Rutherford, Dr H. van Ark and Mrs J. Schaap for assistance.

References

- MUELLER-DOMBOIS, D. & ELLENBERG, H. 1974. Aims and methods of vegetation ecology. Wiley, New York.
- PARTON, W.J. & INNIS, G.S. 1972. Some graphs and their functional forms. U.S. International Biological Program, Grassland Biome Project Technical Report No. 153: 1-41.
- RUTHERFORD, M.C. & WESTFALL, R.H. 1986. The Biomes of southern Africa — An objective categorization. *Mem. bot. Surv. S. Afr.* No. 54.
- WESTFALL, R.H. 1981. The plant ecology of the farm Groothoek, Thabazimbi District. M. Sc. thesis, Univ. of Pretoria.
- WESTFALL, R.H. (in prep.) The vegetation ecology of Sour Bushveld in the Transvaal Waterberg.



3.2 FIELD SAMPLING

The purpose of field sampling is to acquire data, that with processing and logical interpretation will achieve the aims of the study. Because field sampling often incurs additional transport and subsistence costs, savings can be effected by reducing time spent in the field, through improvements in methodology.

3.2.1 Sampling unit location and number

This refers to sampling unit location (sample plot) within a vegetation stand and not stand location, as described in section 3.1.5.

3.2.1.1 Background

Where working scale is such that the scale-related vegetation stand can be represented by a single quadrat then selection of quadrat location, based on representativeness, is probably the most efficient method of quadrat location. However, where the area of the scale-related vegetation stand is far larger than the area of a single quadrat used for sampling the stand, then randomization of sub-quadrat locations will reduce the number of observer decisions to be made. This is confirmed by Werger (1974) who states that "sampling plots (sampling units) should be representative of the stands they represent unless random or systematic sampling is carried out when the question of representativeness is by-passed."

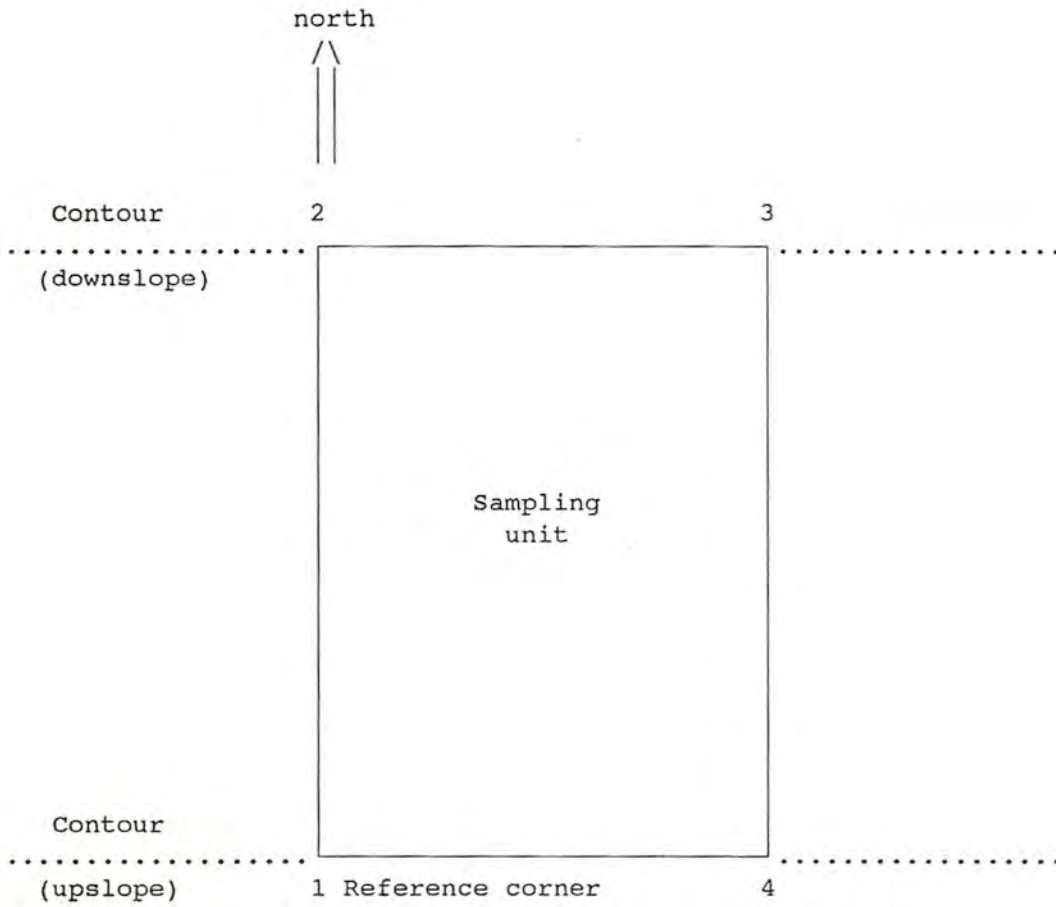


FIGURE 3.2. - Reference corner (1) of sampling unit in terms of direction, for level terrain; and slope, for sloping terrain.



In this study, where 2,5 x 5 m sampling units are used (section 3.1.6) to sample a 20 ha stand, representativeness is extremely difficult to determine and random sampling unit placement is required.

3.2.1.2 *Methods applied*

The stand centre is a point in the field represented by grid intercepts and co-ordinates (section 3.1.5). Placement of sampling units within the stand can be simplest with the stand centre as reference because the stand is circular. A co-ordinate system based on compass bearing and distance from the stand centre is a convenient method of locating a sampling unit.

The program PHYTOLOC (section 3.1.5.3) can be used to generate random co-ordinates for this purpose, where the x co-ordinate is a compass bearing and the y co-ordinate is the distance from the stand centre in metres, or *vice versa*. A list of these co-ordinates can be printed prior to field sampling, to save time.

Decision-making for sampling unit orientation can be further reduced by standardizing on the sampling unit reference corner. This can also facilitate vegetation monitoring in terms of sampling unit re-location. The reference corner is that point on the ground represented by the sampling unit co-ordinates and is the left upslope corner of the sampling unit, when facing downslope, with the longer sides orientated downslope (Figure 3.2). If the ground is level, then the reference corner is the south west corner, when facing



north, with the longer sides orientated due north.

3.2.2 Plant identification and verification

Plant identification aids are often used by researchers in the field. These can be portable field herbaria, photographic records, photostats of herbarium specimens and notes on identifying characteristics.

3.2.2.1 *Background*

Although it is common practice to collect voucher specimens of specific and infra-specific taxa recorded in a study area for reference purposes, the criteria for field identification of these plants are rarely made known. Comparison with a voucher specimen is insufficient if the criteria used for comparison are not given. The problem is exacerbated by the reliance of most identification aids, such as floras, on fertile material, whereas, field identifications are often on sterile material. This can complicate efforts to repeat a given study, for example, for monitoring purposes, and inhibit the transfer of species identification knowledge, which is unique to each field researcher. A lack of standardization is also evident in the character and character states used for identifying plants in published works. Can vegetation studies be repeatable if the criteria used by researchers for plant identifications in the field, are not made known?



A new identification aid combining features of a polyclave and an analytical key

ROBERT H. WESTFALL, HUGH F. GLEN AND MICHAEL D. PANAGOS

Botanical Research Institute, Department of Agriculture and Water Supply, Private Bag X101, Pretoria 0001, South Africa

Received December 1984, accepted for publication February 1985

WESTFALL, R. H., GLEN, H. F. & PANAGOS, M. D., 1986. **A new identification aid combining features of a polyclave and an analytical key.** A new identification aid combining features of a polyclave and an analytical key is described. It is based on presence-or-absence characters and is presented in the form of a matrix, with the characters in rows and the taxa in columns. The PHYTOTAB program package is used to order the matrix, in order to facilitate identification. The method was used to construct an identification aid for vegetative material and has wider taxonomic and teaching implications.

ADDITIONAL KEY WORDS:—Computer-aided-identification – matrix – phytosociology – PHYTOTAB

CONTENTS

Introduction	65
Existing identification methods	66
Analytical keys	66
Polyclaves	67
Methods	68
Results	70
Discussion	70
Conclusions	72
Acknowledgements	72
References	73

INTRODUCTION

During the course of work on a pilot study of the Transvaal Waterberg (Westfall, 1981) the need for an objective method of site identification of the numerous plant species arose. The use of field herbaria proved time-consuming and cumbersome. Existing keys would have been of limited assistance because, first, they often use evanescent characters at an early stage (e.g. requiring flowers out of season), and secondly, they are only available for a minority of the

taxa present in this study area. When a specimen is collected for later identification by a reputable herbarium, the onus of identifying correctly the same taxon at other sites rests entirely on the knowledge, memory and experience of the field ecologist. During phytosociological work, vegetative characters are necessarily used for identification, whereas vegetative keys are rarely available.

Bearing these problems in mind, we decided to investigate alternative methods of identification, a subject which had been pursued by one of us in some detail (Glen, 1974). Traditionally, taxonomists have used one or another form of dichotomous key for identification, and the alternative methods have been given very little practical use. A selection of identification strategies is discussed below.

EXISTING IDENTIFICATION METHODS

A wide variety of different manually operated identification systems has been proposed for use in biology (Leenhouts, 1966), as well as a small but growing number of computer-aided systems (Morse, 1971; Pankhurst, 1978). These include such apparently diverse systems as the 17th-century precursors of the indented dichotomous key described by Voss (1952) and the computer-aided system described by Morse (1971), in which the computer interrogates the operator. All these techniques belong to one or other of two groups: analytical keys, which have only one entry point, and polyclaves (this term, coined by Duke (1969), is used here in the wider sense given to it by Morse, 1974), which have many entry points. The method described later in this paper is unique in being usable as both an analytical key and as a polyclave.

Analytical keys

Analytical keys are easily reproduced, being printed documents, and relatively easily optimized, the theory of optimization having been first worked out by Lamarck (1778). Possibly the most accessible review of these is by Osborne (1963). A dichotomous key can be used with maximum efficiency if each step splits the remaining group of taxa into two equal subgroups. In general, a polychotomous step with N branches at any point in an analytical key operates at maximum efficiency if it splits the group into N equal subgroups. This means that the mean path length from the start of the key to the level at which an identification is achieved is minimized.

Attempts have been made (Morse, 1971, 1974) to weight different taxa according to their relative commonness or rarity, so that common taxa are keyed out with the shortest average path length, and to weight characters so that the easiest to observe are used preferentially. Taxon weighting would tend to minimize the number of errors per unit time on the assumption that the key will be used more often to identify common taxa than rarities. Character weighting tends to minimize the chance of using a difficult character near the start of the process, and so it minimizes the possibility of using the wrong major section of the key, thus minimizing the number of errors per identification.

Analytical keys do have practical disadvantages, despite their clear theory and apparent simplicity. These are largely related to their single entry point and



limited number of paths to the correct identification. The most frequently encountered problem is that it is very difficult to construct a key that will enable one to identify a fragmentary (sterile) specimen unambiguously if it belongs to a large group, for example the Mesembryanthemaceae. Keys to the genera of this family have been constructed both by Bolus (1958) and by Herre & Volk in Herre (1971). Both of these require flowers and ripe fruit for certainty in identification, yet both parts are seldom available at the same time.

Variable taxa present a problem in the construction of analytical keys. If one allows for the variation by allowing the taxon in question to key out in more than one place, then one degrades the overall efficiency of the key by lengthening it. If one allows for variation by describing it where it occurs in the key, then efficiency is degraded by basing a dichotomy on a property which is less than desirably clear, or by lowering the ease of observation.

The addition of taxa to an existing analytical key is a relatively severe problem involving the rewriting of at least part of the key. If the additions are more than minor, it is probably most efficient to rewrite the key from the beginning, if optimization is to be retained. Computer-constructed keys can achieve this end more easily, for example those of Morse (1974) and Pankhurst (1971).

Ease of observation optimization requires that a minimum of characters be used at each step. In fact, the use of more than one character in a step implies that either the characters used are not fully independent, or that one or more of those used are variable. Rypka *et al.* (1967) have shown that it is theoretically possible to construct a key to any number (T) of taxa using not more than N characters, where $N = \log_2 T$, rounded up to the nearest integer. In fact, considerably more characters than this minimum number are usually used, but the number of characters rarely exceeds the number of taxa. Therefore, it may happen that a highly distinctive and potentially useful character, for example flower colour or an accurate locality label, is present on the specimen but is not used in the key.

Polyclaves

A polyclave has as many entry points as there are characters, and an effectively infinite number of pathways from the start to identification of the specimen. Very little skill is necessary to identify a plant using a polyclave, but the speed and accuracy of identifications improve significantly with increasing skill. In general, it appears that the most efficient strategy for using a polyclave is at each step to use the rarest character that is available and still unused.

Sneath & Sokal (1973) divide polyclave algorithms into simultaneous and sequential. The basic strategy of a simultaneous polyclave is to calculate the value of one or another similarity function between the unknown specimen and all taxa in the identification matrix. The value of this function should be high with only one taxon, to which the unknown may be assigned if the value is above a predetermined threshold. Alternatively, taxa are considered to be regions in character hyperspace, and the locality of the unknown in this space may be calculated. The unknown is assigned to a particular taxon if it falls into that region of the space. Typical simultaneous methods are those of Gyllenberg (1965) and Lapage *et al.* (1970, 1973).

In sequential polyclaves, characters are considered one at a time. In each step, a subset of taxa having the character in question in the same state as the unknown is extracted from the complete set, and compared with the subset resulting from the previous step. Those taxa common to both subsets form the set of possible identifications for the next step. When this step contains only one taxon, the unknown has been identified. A wide variety of such polyclaves has been described by Leenhouts (1966). Essentially, they fall into three groups: printed, card- and computer-operated, although the same polyclave may be converted from one form to another.

Printed polyclaves may be even cheaper to reproduce than the equivalent dichotomous key, because they can be compressed into less space. For example, the key of Meyer (1969) to the families of flowering plants of SW Africa/Namibia, a polyclave, occupies about 220 lines (excluding explanatory figures), while the dichotomous key by Merxmüller (1972) to the same taxa plus 15 families of Pteridophyta occupies about 1500 lines. However, a polyclave is frequently more expensive to reproduce.

A form of polyclave on plastic cards has been described (Leenhouts, 1966), but polyclaves on paper cards are far commoner. Most punch-card polyclaves fall into two groups: centre-punched one card per character state, and edge-punched one card per taxon. The former kind is exemplified by the keys of Bianchi (1931) (the first punch-card polyclave), Hansen & Rahn (1969) and Weber & Nelson (1972); the latter by Baker (1970), a key to the species of the genus *Erica*.

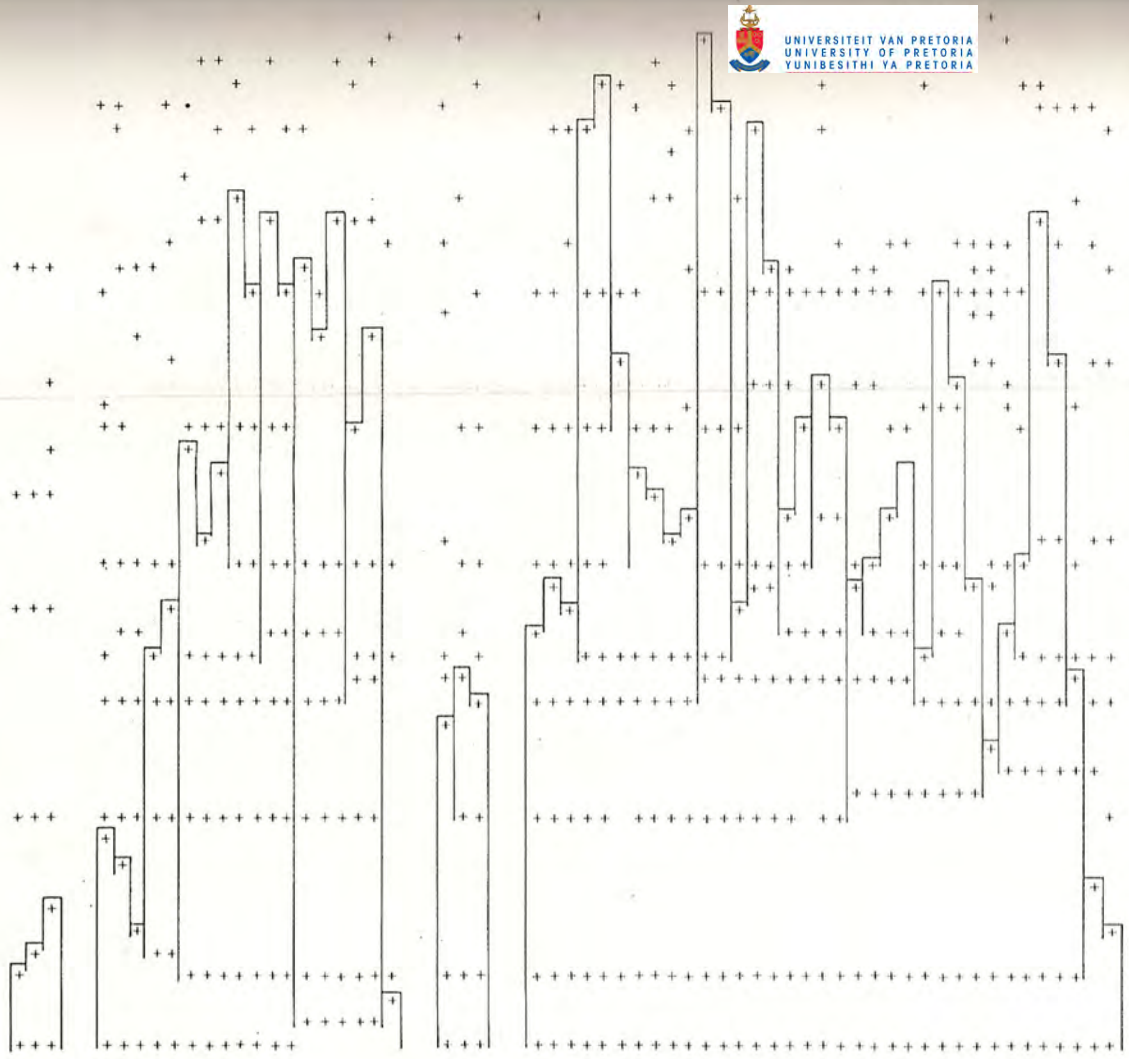
With the possible exception of computer-generated punch-card polyclaves, this kind of key is very expensive to reproduce, as accurate punching dies must be made for each card in the key. With computer-generated keys, standard 80-column cards are punched with a standard card punch, and the key can be produced for a realistic price. With the almost total extinction of the computer card reader-punch as a result of improved magnetic storage media, it is not certain whether computer-generated punch-card keys will remain viable for any length of time.

Edge-punched card keys can be operated with a mechanical card sorter or with a skewer, but are more subject to damage than centre-punched cards. On the other hand, the latter do not usually have any space for annotation.

Computer-aided identification systems (automated polyclaves) are many, for example those of Boughey, Bridges & Ikeda (1968), Goodall (1968), Morse (1974), Pankhurst (1978) and the simultaneous methods mentioned above. Each has its own advantages and disadvantages, but at present none can be used in the field. It is possible that in the future one or more may be made to work on a portable personal computer. Morse (1975) has given a particularly lucid account of the computer-aided systems available at that time. We plan to add to the system described here, routines enabling it to be used in the field on a micro-computer.

METHODS

Eighty plant specimens were collected in the Transvaal Waterberg during March 1983. The specimens were numbered and sent to the National Herbarium, Pretoria (PRE) for identification.



Cladodromous			Venation
Semicraspedodromous			
Reticulodromous			
Hyphodromous			
Brochidromous			
Craspedodromous			
Eucamptodromous			
Asymmetrical base			
Asymmetrical lamina			
Oblong			
Narrow /oblong	Leaf form		
Obovate			
Ovate			
Elliptic			
Cuneate			
Lobate			
Decurrent	Leaf base		
Cordate			
Obtuse			
Acute			
Even pinnate			
Multifoliate			
Twice pinnate	Leaf type		
Odd pinnate			
Digitately trifoliate			
Simple			
Opposite			
Whorled	Leaf arrangement		
Decussate			
Alternate/scattered			
Very different	Colour	Leaf surfaces	
Similar	Texture		
Truncate			
Rounded			
Mucronate	Leaf apex		
Obtuse			
Acute			
Erose			
Serrate			
Dentate	Leaf margins		
Revolute			
Crenate			
Lobed			
Entire			
Amplexicaul			
Sessile	Leaf attachment		
Petiolate			

al Waterberg: Trees, shrubs, climbers, creepers, forbs, ferns

Characters relevant to each specimen were observed in the field and recorded on a field data sheet on which provision is made for recording the collector's number, provisional name and locality or relév . If more than 60 characters are recorded additional page(s) could be used.

The morphological characters used in the preliminary key were taken from various sources such as Henkel (1934), Metcalfe & Chalk (1979) and Tainton, Bransby & Booysen (1976) and condensed into an illustrated photocopied booklet for field use. This proved useful in maintaining consistency in the allocation of characters.

The PHYTOTAB program package (Westfall *et al.*, 1982) was used to construct the matrix. The data from the field data sheets were transferred to encoding forms using the PHYTOTAB input format where each character was regarded as a rel v . Taxa were grouped into trees and shrubs, climbers and creepers, forbs, ferns and graminoids for convenience. This would also allow for each group to be shown on a separate table when more data are collected. Characters can also form groups such as leaf-margin patterns and venation patterns. The character groups are sequenced in descending order of character (group) frequency from right to left. Characters not in groups are on the left of the table in frequency order (Tables 1 & 2). Spaces were inserted between characters to facilitate vertical reading.

Taxa are sequenced within groups so that dichotomies may be formed. That is, taxa with similar characters are grouped together. The dichotomies are continued until they consist of a single line for each taxon, and are ended with a vertical blocking line after the first unique character for each taxon when compared with the two adjacent characters. Blocking was done by hand. If a taxon has two or more character states that express possible variation within a character group falling within the blocked area, the species may be repeated to facilitate use of the key. This is illustrated in our example by *Combretum zeyheri* Sond., in which the leaf apices may be either obtuse or mucronate, and by *Syzygium cordatum* Hochst., in which the leaf variability is so great as to require that the same taxon appears four times (Table 1).

To use the table, the specimen to be identified is compared with the characters given for the group to which the specimen belongs, starting from the right-hand side of the table. The user determines whether the character is present or absent in the specimen. If the character is present in the state described, then the user follows the 'present' pathway indicated by +. If the character is sometimes absent it is still shown as +. If it is always absent then the 'absent' pathway, indicated by blanks, is followed. The next character is found to the left of the newly completed one. This procedure is repeated until the end of a block, indicated by a vertical blocking line, is reached. The user should now be able to look across to only one name on the left-hand side of the table. This will give him the desired identification. The blocked pathways indicate the minimum characters needed for identification. Additional characters outside the blocks may be used for corroboration.

Two other possible methods of using the table suggest themselves. However, these methods should only be necessary as a teaching aid, or if the specimen lacks important characters or in a preliminary version of the key. It must be stressed that these methods are by no means obligatory.

First, it is possible to use the table as a form of tabular key (Newell, 1970).



A profile of the unknown specimen is obtained by noting the characters present, in the order of the table, and this profile is compared with those of the taxa in the table, one by one, until a matching profile is found. This would be the name of the unknown specimen. This method is probably too cumbersome to be of practical value.

Another method of using the table would be as a classical polyclave. The unknown specimen is examined for any conspicuous (or unusual, if one is familiar with the table) character, and the taxa which display this character in this state are determined. Another character is searched for, to see which of the taxa found in the previous step display this character in the required state. The specimen has been identified when only one taxon remains.

RESULTS

The problems discussed in the introduction were overcome by devising a method for constructing vegetative keys during the reconnaissance phase of phytosociological work. The method had to allow for increasing the number of taxa as the work progressed. It had also to be simple to apply in the field. The key presented here will operate satisfactorily only in the geographical area concerned; it is used in this paper solely as an example.

In Table 2, *Schizachyrium sanguineum* (Retz.) Alst. in the Graminoid group has an open pathway indicating that this species does not key out with the available characters. However, its identity can be inferred from the absence of flattened stem bases which are present in *Hyparrhenia filipendula* (Hochst.) Stapf var. *pilosa* (Hochst.) Stapf.

DISCUSSION

The matrix method of key construction described in this paper allows for biological variation where more than one character state, such as different leaf forms, may be present in the same taxon, without detracting from the effectiveness of the key. Characters which are periodically available for which the population is variable may be used, as well as corroborative characters. The right-to-left method of use described above in the Methods may seem at first sight to be idiosyncratic. This method is used because the phytosociological example makes use of the species name file for both vegetation and taxonomic classification. The PHYTOTAB program package allows the user to print species names to the right of the table as well as printing the matrix in mirror image, thereby reading from left to right. Thus, in ordinary taxonomic usage the table would be rearranged to read 'correctly'. A further advantage of this package is that in the printout form, both horizontal and vertical lines may be employed to improve legibility.

This form of key is unique in that it can be used as both a polyclave and an analytical key. It combines the flexible strategy of the former with the ease of reproduction of the latter to produce a very efficient identification system. One advantage of polyclaves is that either taxa or characters can be added at will, with minimal effort required, although the addition of taxa may result in ambiguities in identifications, which may only become apparent after long use. By using PHYTOTAB, both taxa and characters can be added to existing keys

at will, and ambiguities will be revealed when the tables are blocked. Correcting a key is also a straightforward task.

It will be noted that the horizontal length of the block for each taxon is roughly proportional to the path length that would be required in an analytical key to identify that taxon. Two uses for this proportionality (which, it should be stressed, is no more than a rough approximation) suggest themselves. First, it offers a check on the discriminatory efficiency of the characters chosen. If all the blocks are long, then redundant characters which do not contribute to any identifications are present. These can be removed, or placed where they are only used as corroborative characters. Secondly, the proportionality can be used to demonstrate Lamarck's principles of efficient key construction to students elegantly and graphically (Lamarck, 1778; Osborne, 1963). This form of key seems to be the only one available which demonstrates its effectiveness with field data.

A further teaching application of the 'PHYTOTAB-key' may be mentioned. With the falling cost of computer usage (at least in terms of constant units), and with the relative ease and accessibility of use of the PHYTOTAB package, a method is presented of allowing students to learn the mysteries of good key construction in the most effective possible manner, by making their own mistakes in the course of constructing their own keys to taxa they think they know.

The PHYTOTAB program package (Westfall *et al.*, 1982) allows the user to construct a key to which taxa and characters may be added as and when more specimens are collected. The keys can be subdivided easily to form tables of convenient size for field use. An advantage with computer printout is that keys may be duplicated conveniently and cheaply.

CONCLUSIONS

Our identification aid has peculiar and significant advantages in three areas: in teaching, in reconnaissance and in the field. Its suitability as a teaching tool stems from the way in which the printed table, while being usable as a polyclave, displays explicitly the structure of an analytical key. Because the PHYTOTAB program package is relatively simple, quick and inexpensive to use, students can be given practical experience of key construction using this package. For the same reasons, 'PHYTOTAB-keys' can be produced repeatedly during the course of a taxonomic investigation. Ambiguities in the key at each stage may be taken to indicate areas requiring further research, and the keys produced will help in field-work while the group is under revision. In the field, use of keys made by this method will mean that characters present only in fresh material can be used to improve the accuracy of determinations. It may be expected that the facility with which both taxa and characters may be added to the table will be found useful in both taxonomic and ecological work.

ACKNOWLEDGEMENTS

We thank A. V. Hall, O. A. Leistner, R. A. Lubke, E. J. Pankhurst and J. C. Scheepers for helpful suggestions. Miss A. P. Backer and Mr J. F. van Blerk are thanked for assistance in the early stages of this work. The staff of the National Herbarium, Pretoria identified the specimens collected.



REFERENCES

- BAKER, H. A., 1970. A key for the genus *Erica* using edge-punched cards. *Journal of South African Botany*, 36: 151-156.
- BIANCHI, A. T. J., 1931. Een nieuwe determinatie-methode. *Tectona* 24: 884-893.
- BOLUS, H. M. L., 1958. Key to the genera of Mesembryanthemaceae. *Notes on Mesembryanthemum and allied Genera*, 2: 392-403.
- BOUGHEY, A. S., BRIDGES, K. W. & IKEDA, A. G., 1968. An automated biological identification key. *Museum of Systematic Biology, University of California, Irvine, Research Series*, 2: i-xix, 1-35.
- DUKE, J. A., 1969. On tropical tree seedlings. I. Seeds, seedling systems and systematics. *Annals of the Missouri Botanical Gardens*, 56: 125-161.
- GLEN, H. F., 1974. 'A taxonomic revision of the subtribe Gibbaeinae (Mesembryanthemaceae)'. Unpublished M.Sc. Thesis, University of Cape Town, Cape Town.
- GOODALL, D. W., 1968. Identification by computer. *BioScience*, 18: 485-488.
- GYLLENBERG, H. G., 1965. A model for the computer identification of micro-organisms. *Journal of General Microbiology*, 39: 401-405.
- HANSEN, B. & RAHN, K., 1969. Determination of angiosperm families by means of a punched-card system. *Dansk botanisk Archiv*, 26: 1-46 & 172 punched cards.
- HENKEL, J. S., 1934. *A field book of the woody plants of Natal and Zululand*. Durban: Robinson.
- HERRE, A. G. J., 1971. *Genera of Mesembryanthemaceae*: Cape Town: Tafelberg.
- LAMARCK, J. B. A. P. M. de, 1778. *Flore Francaise*, 3 vols. Paris: Imprimerie Royale.
- LAPAGE, S. P., BASCOMB, S., WILCOX, W. R. & CURTIS, M. A., 1970. Computer identification of bacteria. In A. Baillie & R. J. Gilbert (Eds), *Automation, Mechanisation and Data Handling in Microbiology*: 1-22. London: Academic Press.
- LAPAGE, S. P., BASCOMB, S., WILCOX, W. R. & CURTIS, M. A., 1973. Identification of bacteria by computer: general aspects and perspectives. *Journal of General Microbiology*, 77: 273-290.
- LEENHOUTS, P. W., 1966. Keys in biology: a survey and a proposal of a new kind. *Proceedings of the Koninklijke Nederlandsche Akademie van Wetenschappen Amsterdam, Series C*, 69: 571-596.
- MERXMÜLLER, H., 1972. *Prodromus einer Flora von Suedwest Afrika*. Lehre: Cramer.
- METCALFE, C. R. & CHALK, L., 1979. *Anatomy of the Dicotyledons*, Vol. 1, 2nd edition. Oxford: Oxford University Press.
- MEYER, P. G., 1969. *Eintuehrung in die Pflanzenwelt Suedwestafrikas*. Windhoek: S.W.A. Wissenschaftliches Cesellschaft.
- MORSE, L. E., 1971. Specimen identification and key construction with time-sharing computers. *Taxon*, 20: 269-282.
- MORSE, L. E., 1974. Computer programs for specimen identification, key construction and description printing using taxonomic data matrices. *Publications of the Museum, Michigan State University, Biological Series*, 5: 1-128.
- MORSE, L. E., 1975. Recent advances in the theory and practice of specimen identification. In R. J. Pankhurst (Ed.), *Biological Identification with Computers*: 11-52. London: Academic Press (Systematics Association special volume, 7).
- NEWELL, I. M., 1970. Construction and use of tabular keys. *Pacific Insects*, 12: 25-37.
- OSBORNE, D. V., 1963. Some aspects of the theory of dichotomous keys. *New Phytologist*, 62: 144-160.
- PANKHURST, R. J., 1971. Botanical keys generated by computer. *Watsonia*, 8: 357-368.
- PANKHURST, R. J., 1978. *Biological Identification*. London: Edward Arnold.
- RYPKA, B. W., CLAPPER, W. E., BOWEN, I. G. & BABB, R., 1967. A model for the identification of bacteria. *Journal of General Microbiology*, 46: 407-424.
- SNEATH, P. H. A. & SOKAL, R. R., 1973. *Numerical Taxonomy*. San Francisco: Freeman.
- TAINTON, N. M., BRANSBY, D. I. & BOOYSEN, P. de V., 1976. *Common Veld and Pasture Grasses of Natal*. Pietermaritzburg: Shuter & Shooter.
- VOSS, L. G., 1952. The history of keys and phylogenetic trees in biology. *Journal of the Science Laboratory, Denison University*, 43: 1-25.
- WEBER, W. A. & NELSON, P. P., 1972. *Random access key to the genera of Colorado mosses*. Boulder: University of Colorado Museum.
- WESTFALL, R. H., 1981. 'The plant ecology of the farm Groothoek, Thabazimbi district'. Unpublished M.Sc. Thesis, University of Pretoria, Pretoria.
- WESTFALL, R. H., DEDNAM, G., VAN ROOYEN, N. & THERON, G. K., 1982. PHYTOTAB—a program package for Braun-Blanquet tables. *Vegetatio*, 49: 35-37.



3.2.2.2.1 Improvements

Subsequent improvements to the techniques of plant identification are:

- i. an illustrated booklet of plant characters used for identification in this study (Appendix I);
- ii. a program was written, as an option for the PHYTOTAB-PC program package (section 3.6), for sequencing species for plant identification purposes. This program groups species according to the required dichotomies for key production. It was found that the character sequence is easier to apply if characters are standardized in their sequence, namely, from large to small attributes. This sequence is also used in the illustrated booklet and encoding forms. Using the same characters it is also possible to produce a family key; and,
- iii. visual verification of plant identification by means of photographic slide/plant comparisons. A photographic colour slide is taken of the plant when the voucher specimen is collected. The slide incorporates the specimen number. After development of the slide film, the slides are not mounted but returned to the film cassette.

After a plant is identified by means of the matrix key, the cassette with the relevant specimen number is inserted into a field viewer for comparison. The field viewer used here is an old motor-drive Konica FS 1 camera with a) a translucent perspex back for light transmission and b) a magnifier lens for viewing which replaces the standard lens. The mirror system is also removed. The motor-drive facilitates slide location.



3.2.3 Species cover

In this study species cover refers to total projected canopy cover for a plant species and cover estimations are required with a precision suitable for monitoring cover change and determining relative species composition.

Generally species cover is visually estimated and not sampled. However, where greater precision than a visual estimate is required, as in this study, and cover estimation is achieved through sampling, then the sampling unit area for cover should relate to cover. Therefore, the sample attributes of presence and cover can require different sampling unit areas.

3.2.3.1 Background

Mueller-Dombois & Ellenberg (1974) list several scales for estimating species cover-abundance such as the Domin-Krajina scale with eleven classes, the Braun-Blanquet scale with seven classes and the Daubenmire scale with six classes. Werger (1974) advocates a nine class scale.

Londo (1976), however, states that a cover scale should meet the following requirements:

- i. the scale should be sufficiently detailed;
- ii. its points should be related to the actual cover values so that these can be treated arithmetically;
- iii. cover and abundance should be strictly separated. It



- is not logical to combine such dissimilar features in a single quantitative scale; and,
- iv. the symbols should be as simple as possible.

The criterion of precision, commensurate with repeatability, could also be added to these requirements for monitoring cover change. However, the aim of cover estimations should first be considered. If the aim is to detect cover pattern in a matrix then, possibly, an eleven class scale, such as the Domin-Krajina scale, could confuse pattern and the ideal might be a simplified five class scale. On the other hand if arithmetic processing of cover data is required, such as for monitoring cover change or to determine relative species composition, then even an eleven class scale is likely to be too crude. Precision in cover determination is inversely proportional to cover class size and directly proportional to effort required for such determinations. A balance should, therefore, be achieved between class size and effort required, for cover determinations, to meet for example, monitoring or species composition requirements.



Miscellaneous notes

VARIOUS AUTHORS

THE PLANT NUMBER SCALE — AN IMPROVED METHOD OF COVER ESTIMATION USING VARIABLE-SIZED BELT TRANSECTS

INTRODUCTION

The vegetation ecology of the Transvaal Waterberg is currently being investigated at a scale of 1:250 000 (Westfall in prep.). Vegetation structure is being analysed according to Edwards (1983) using the cover meter (Westfall & Panagos 1984) for cover determinations in each height class. In the floristic analysis, individual species cover is estimated by using the Domin-Krajina cover-abundance scale (Mueller-Dombois & Ellenberg 1974).

A comparison of recorded species cover, being the sum of the class midpoints according to the Domin-Krajina scale (Mueller-Dombois & Ellenberg 1974), with the structural cover should result in the summed species cover for a stand being: 1, greater than the cover of the height class with the greatest cover, and 2, less than the summed cover of all the height classes, provided that 3, quadrat size is such as to include those species contributing significantly to the total cover of the stand.

In the vegetation being investigated, quadrat size is generally commensurate with species richness (Westfall *et al.* 1987). However, the summed estimated species cover was often considerably less than the mean cover for the height classes with the greatest cover for the same stands. Overcompensation for the underestimation of species cover often led to the summed estimated species cover being considerably greater than the summed cover of all the height classes for the same stands (Westfall in prep.). Clearly, improved species cover estimations are required for species cover to have any relevance other than an approximate indication of relative abundance.

In estimating species cover, according to Edwards (1983), the observer is often inclined to ignore grasses without inflorescences and to estimate from a static position without taking plant size and distribution into account. For example, a larger plant should require a larger area to be observed than a smaller plant. Furthermore, although mean canopy diameter can be readily estimated it is often far more difficult to estimate mean distance apart in terms of mean canopy diameter because of often highly irregular plant distribution.

To overcome these problems, the approach suggested here is based on a simple estimate of area and a count of the individuals of a species within the area.

METHODS

The cover of a species is given by Edwards (1983) for hexagonal packing by

$$c = \frac{90,7}{(n+1)^2}$$

where c = percentage crown cover and n = the mean number of crown diameters by which the plant crowns are separated.

Assuming hexagonal packing, the transect area, of which the percentage crown cover is a proportion, is given by:

$$\sin 60^\circ (n+1) 30D$$

where D = mean crown diameter and 30 = the value for obtaining a minimum of $0,1\%$ cover. Cover of less than $0,1\%$ is not considered significant. Transect length is, therefore, $30D$ and transect width is $\sin 60^\circ (n+1)$. In practice, transect width was taken as slightly less than the average gap between plants within or nearest to the sample quadrat plus the mean crown diameter. The number of individuals of a species was then counted within the transect. Only species occurring within the sample quadrat were recorded and for each a count of individuals within a transect commensurate with each species spacing and size was made. Counts of individuals did not include the first individual as the transect was started adjacent to the first individual. This permitted a cover of less than $0,1\%$ where no individuals were counted. Transect width was never greater than the length as this could have resulted in actual cover values of less than $0,1\%$ being given higher cover values.

The mean number of crown diameters (n) by which the plant crowns are separated within each transect is given by $n = \frac{30-I}{I}$ where I = number of individuals counted. Percentage crown cover can then be calculated according to Edwards (1983). Table 1 shows number of individuals counted, representation by a single character symbol and percentage cover for recording purposes in the field.

Vegetation structure was analysed using the cover meter and the summed species cover was estimated with both the Domin-Krajina scale and the plant number scale as outlined above for five vegetation stands represented by 21 quadrats. The mean of the shortest and longest cross distances of each crown was taken as the crown diameter for each species and these distances were noted in four categories (Edwards 1983) namely, forbs (herbs), grasses, shrubs and trees. Class intervals were selected on a basis of trial and error to give an approximately normal distribution of occurrences within crown diameter class intervals. All estimations were done by an independent observer.

RESULTS

The results of the crown cover determinations are given in Table 2.



TABLE 1.—Number of plant individuals counted with single character symbol and percentage crown cover

No.	Symbol	% Crown cover
0	+	0,00
1	1	0,10
2	2	0,40
3	3	0,91
4	4	1,61
5	5	2,52
6	6	3,63
7	7	4,94
8	8	6,45
9	9	8,18
10	A	10,08
11	B	12,20
12	C	14,51
13	D	17,03
14	E	19,75
15	F	22,68
16	G	25,80
17	H	29,12
18	I	32,65
19	J	36,38
20	K	40,31
21	L	44,44
22	M	48,78
23	N	53,31
24	O	58,05
25	P	62,99
26	Q	68,13
27	R	73,47
28	S	79,10
29	T	84,76
30	U	90,70
31	V	96,85
>31	W	100,00

The Fibonacci sequence, where each number is the sum of the preceding two numbers, provided the closest resemblance to a normal distribution of occurrences within crown diameter class intervals. This is illustrated in Figure 1 with a frequency polygon with the class intervals on a natural logarithmic scale to reduce the effect of increasingly larger class intervals.

The class intervals used for mean crown diameters according to the Fibonacci sequence, are shown in Table 3. Transect lengths were determined by the midpoints of each class interval.

INTERPRETATION

In Table 2 estimations according to the Domin-Krajina cover-abundance scale are considerably lower than those for the single height classes with the most cover. The plant number scale, in contrast, yielded higher values than that of the single height class with the most cover and lower values than the cover of the combined height classes for each relevé, except relevé 37. This indicates a greater precision in estimating cover when using the plant number scale as opposed to the Domin-Krajina scale.

According to Westfall *et al.* (1987), quadrat size should have been larger for the vegetation type represented by relevé 37, but this was not apparent using the Domin-Krajina scale at the time of sampling. However, simple summation of the values obtained by the plant number scale in the field indicated inadequate quadrat size. It is far too time-consuming to verify quadrat size

for each quadrat according to Westfall *et al.* (1987). The plant number scale together with a structural analysis of the vegetation provides a simple means of verifying adequacy of quadrat size.

In the frequency polygons (Figure 1) the peaks to the left of the central troughs for forbs, grasses and shrubs are caused by a relatively higher proportion of 0,2 m diameter crowns. This can be attributed to the observer rounding off crown diameters to 0,2 m some of which should have fallen into the 0,211 to 0,34 class. If class intervals had been known at the time of recording, it can be expected that greater care would have been exercised in measurements where crown diameters were close to class borders. The troughs mentioned are, therefore, considered to be a result of measurement inaccuracies which could be overcome by using class intervals.

The use of standard transect lengths as illustrated in Table 3 should simplify transect length determination and provide for variability in crown size. A further advantage could be the simultaneous counting of individuals of different species with similar crown diameters and spacing to save time. It is also suggested that a simple counter be used for recording number of individuals for each species as marking paper for this purpose requires stopping at each individual recorded.

It must be emphasized that the parameter determined here is projected crown cover and not projected foliage cover which is more species and age-dependent.

The class 'r' on the Braun-Blanquet scale and '+' on the Domin-Krajina scale (Mueller-Dombois & Ellenberg 1974) both with 'solitary, insignificant cover' are difficult to determine. Species with a single occurrence in a sample quadrat often have significant cover outside the quadrat. If a stand is defined, as in this study, as 20 ha (Westfall *et al.* 1987), it is impracticable to determine whether a species is 'solitary' within that area. The concept of 'solitary' is relative to the area defined. In the plant number scale used here the lowest cover class is less than 0,1%, which seems better defined than 'solitary'.

In contrast to the Domin-Krajina scale, the plant number scale has proportionately finer subdivisions at the lower cover values of the scale. This is of significance in the South African context with often high species richness characterized by many dominant species with generally lower cover in contrast to the few dominant species with higher cover often found in the relatively impoverished European vegetation.

TABLE 2.—Percentage crown cover in five vegetation stands represented by relevés 33 to 37

Relevé no.	Single height class	Percentage crown cover		
		Combined height classes	Domin-Krajina scale	Plant number scale
33	55	100	27	60
34	31	104	25	74
35	44	103	19	79
36	41	86	21	53
37	63	146	17	35

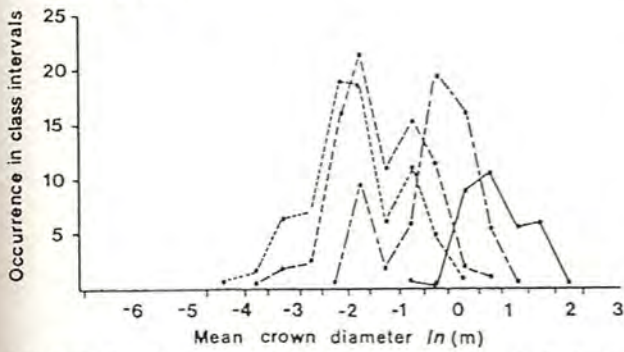


FIGURE 1.—Frequency polygons of occurrence of plants in mean crown diameter class intervals for forbs (....), grasses (---), shrubs (-·-·-) and trees (—). Class intervals are according to the Fibonacci sequence on a ln scale.

CONCLUSIONS

The projected crown cover determinations based on area estimations and counting of individuals shows improved precision compared to the Domin-Krajina cover-abundance scale. Although this method is more time-consuming than a purely visual estimation of cover, the use of standard class intervals and a counter should decrease the time required for cover determinations. The method appears more suitable for the species-rich South African vegetation than the traditional European cover-abundance estimation scales. The method also provides a means of verifying quadrat size adequacy.

ACKNOWLEDGEMENTS

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

REFERENCES

EDWARDS, D. 1983. A broad-scale structural classification of vegetation for practical purposes. *Bothalia* 14: 705–712.

TABLE 3.—Class intervals of crown diameters according to the Fibonacci sequence for determining standard transect lengths

Crown diameter class interval (m)	Transect length (m)
0,001–0,01	0,15
0,011–0,02	0,45
0,021–0,03	0,75
0,031–0,05	1,20
0,051–0,08	1,95
0,081–0,13	3,15
0,131–0,21	5,10
0,211–0,34	8,25
0,341–0,55	13,35
0,551–0,89	21,60
0,891–1,44	34,95
1,441–2,33	56,55
2,331–3,77	91,50
3,771–6,10	148,05
6,101–9,87	239,55

MUELLER-DOMBOIS, D. & ELLENBERG, H. 1974. *Aims and methods of vegetation ecology*. Wiley, London.

WESTFALL, R.H. & PANAGOS, M.D. 1984. A cover meter for canopy and basal cover estimations. *Bothalia* 15: 241–244.

WESTFALL, R.H., VAN STADEN, J.M. & PANAGOS, M.D. 1987. Predictive species area relations and determination of subsample size for vegetation sampling in the Transvaal Waterberg. *South African Journal of Botany* 53: 44–48.

WESTFALL, R.H. in prep. *The vegetation ecology of Sour Bushveld in the Transvaal Waterberg.**

R.H. WESTFALL and M.D. PANAGOS

MS. received: 1987.07.31.



TABLE 3.1. - Symbols used for crown diameter classes

Crown diameter (m)	Symbol
0,001 - 0,010	A
0,011 - 0,020	B
0,021 - 0,030	C
0,031 - 0,050	D
0,051 - 0,080	E
0,081 - 0,130	F
0,131 - 0,210	G
0,211 - 0,340	H
0,341 - 0,550	I
0,551 - 0,890	J
0,891 - 1,440	K
1,441 - 2,330	L
2,331 - 3,770	M
3,771 - 6,100	N
6,101 - 9,870	O



3.2.3.2.1. Improvements

Subsequent improvements to the methods described in section 3.2.3.2 include the allocation of symbols (Table 3.1) to the crown diameter classes and recording of these class symbols together with the cover symbols for each species. The following data can be derived from the relationship between mean cover and mean canopy diameter:

- i . mean cover (%);
- ii . mean crown diameter (m);
- iii. individuals per hectare;
- iv. square metres per individual;
- v. mean spacing, plant centre to centre (m);
- vi. mean canopy radius (m); and,
- vii. mean canopy to canopy gap (m).

It was found, during field testing of these methods, that the precision of species cover estimations could be improved considerably, for the higher strata, if rooting within a sampling unit was substituted for canopy overhang in the sampling unit.

The following is, therefore, applied:

- i. forbs, grasses and dwarf shrubs: species presence recorded if half or more than half of the stem or tuft base occurs within the sampling unit; and,
- ii. shrubs and trees: species presence recorded if any part of the canopy overhangs the sampling unit.

In this way vegetation structure is taken into account for sampling presence and cover.



3.2.4 Floristic data recording

Floristic data is taken here to mean the data relating to terrestrial plant life and its distribution, in a given area. The recording of plant identification data is described in section 3.2.2. Taxonomic nomenclature used in this work is according to Gibbs Russell *et al.* (1985, 1987).

3.2.4.1 *Background*

Mueller-Dombois & Ellenberg (1974) suggest that apart from species presence and cover-abundance other aspects are often worth recording such as species quantity, dispersion, morphological structure or layering. To this is added vitality or vigour and sociability. Although the purpose of the study will often dictate what should be recorded, the minimum requirements for optimum results should be considered, taking into account the aspects that can be derived from those that are recorded.

3.2.4.2 *Methods applied*

It follows from the previous sections that the minimum floristic data which should be recorded in the field, for this study, are the following:

- i. species name abbreviation or voucher specimen reference number. The number is preferred because it requires fewer keystrokes than a name abbreviation if direct computer data capture is used; the plant identification key can serve as a num-



ber control whereas abbreviations would require a separate control to ensure that the same abbreviation is not used for a different species; the number necessitates at least some reference to the plant identification keys which can improve identification accuracy; and programmatic inclusion of voucher specimen reference numbers in a checklist is then possible. Although use of name abbreviations can be quicker than numbers, initially, if the species are known, it is suggested that overall time can be saved by using numbers, especially if data correction, where for example, the same abbreviation is used for different species, is taken into account;

- ii. percentage canopy cover class symbol (section 3.2.3); and,
- iii. crown diameter class symbol (section 3.2.3).

In addition, the growth form symbol (from Edwards 1983 and Rutherford & Westfall 1986) is also recorded, as follows:

- T: Trees are rooted, woody, single-stemmed plants over 2 m high or multi-stemmed over 5 m high;
- S: Shrubs are rooted, woody, multi-stemmed plants from 1 to 5 m high, or single-stemmed when less than or equal to 2 m high;
- D: Dwarf shrubs are rooted, woody or partly woody, plants less than 1 m high;
- G: Grasses are rooted, herbaceous plants belonging to the family Poaceae or graminoid plants such as Cyperaceae and Restionaceae which resemble grasses; and
- F: Forbs are rooted, non-graminoid herbaceous plants.

Not all plant species occurring in a sampling unit can be recorded. Those occurring in seed form and those occurring underground at



the time of sampling, for example, are difficult to identify. Similarly, sterile plant material with a low cover is also often difficult to identify. Such plants would generally have a low ecological significance within the plant community concerned although their taxonomic significance as, for example, endemic species, might be high. Furthermore, unidentifiable plants with a low cover and occurrence, are of doubtful value as diagnostic species. It is seldom likely that identifiable material for plants with high cover and occurrence cannot be obtained. The criterion for the recording of a plant species occurrence within a sampling unit is, therefore, the availability of identifiable material at the time of sampling. It is suggested that this is a more realistic criterion than the often-quoted "total floristic recording" which seldom can be true.

The minimum floristic data requirements for classification, however, are initially, relevè number, species name abbreviation or voucher specimen reference number and cover percentage symbol. The following article describes a field data capture program for these data.



PHYTOCAP. A FIELD-DATA CAPTURE PROGRAM FOR THE PHYTOTAB PROGRAM PACKAGE

Manual field recording of floristic data, for phytosociological studies, entails re-encoding data on computer encoding forms, transfer to magnetic tape and finally loading onto the mainframe computer for multi-variate analyses. The time taken from re-encoding to access on the mainframe computer can be from one to six weeks. Classification of data while fieldwork is in progress, therefore, becomes impracticable in the summer-rainfall areas because fieldwork generally takes place in the relatively short growth period and the delays in computer access would limit fieldwork considerably if data were classified during this period. A serious disadvantage of classifying vegetation only after completion of fieldwork is that vegetation units are often either under-sampled, resulting in invalid syntaxa, or over-sampled resulting in wasted labour and expense. Furthermore, the potential for errors is increased by re-encoding the data.

These problems were overcome by using computerized field-data capture where data is recorded directly onto a hand-held computer and loaded onto the mainframe computer on return from the field. Multivariate analysis techniques used are the PHYTOTAB program (Westfall *et al.*, 1982), which are compatible with both the DECORANA (Hill, 1979a) and TWINSPAN (Hill, 1979b) programs. Preliminary classification using either PHYTO20 or TWINSPAN can then be available within a day of returning from fieldwork.

The system used for field-data capture is the Sharp PC 1 500 computer with an additional 8K expansion memory module, printer/cassette interface, cassette recorder and programmable RS232C interface. The program PHYTOCAP, written in BASIC, is used for recording floristic data in the PHYTOTAB (Westfall *et al.*, 1982) format. The program features include automatic line number allocation, sample and/or subsample numbers, alphanumeric species codes, cover-abundance values, data pertinent to individual species, data display, data printing in two formats, halting and continuing program execution, line editing, saving data to tape and loading data from tape. Furthermore, data input is verified for errors such as sample number length, species omission and cover-abundance omission. The user is also informed when five lines of memory are left. The capacity of the computer is 70 lines which is approximately 14 relevés or samples with 40 species per sample, which is generally more than adequate for the floristic data recorded in one day.

Field procedure includes quadrat location, quadrat demarcation, floristic sampling, voucher specimen collection and environmental parameter sampling. The Sharp PC 1 500 computer is used for floristic sampling. Species for which voucher specimens exist are input as a four-letter genus code and a three-letter species code. Species for which voucher specimens are required are tagged with pre-numbered, specimen number, adhesive address labels

and input as a left-justified specimen number. Specimen collection is effected after completion of floristic data recording. This process ensures a smooth flow of data input and reduces the possibility of species being overlooked. Environmental parameters are recorded directly on a field data sheet. It is envisaged that a second Sharp PC 1 500 be used for recording environmental parameters and as a standby machine. Memory capacity currently precludes the use of a single machine for both floristic and environmental data.

After a day's recording the computer is attached to the printer/cassette interface, which remains in the vehicle. A printout of the floristic data is obtained of each sample for stapling to the field data sheets which form the hardcopy for eventual permanent safe-keeping at the Botanical Research Institute, Pretoria. The data are then transferred to C 15 cassettes, when the computer can be cleared for the following data set.

Loading data to the Burroughs B 7 900 mainframe is effected by means of a Burroughs ET 1 100 terminal, RS232C interface and a transfer program called DATATRAN written by S. J. Crafford. Data are read from the cassette and simultaneously transmitted at 300 Baud.

The advantages of this system of data recording include the cost-saving production of preliminary classifications for optimum sampling as well as the labour- and cost-saving of not having to re-encode data. The potential for errors is also reduced by the reduction in data handling. Documentation and taped copies of the program are available from the author. Please forward a blank C 15 cassette for copying.

ACKNOWLEDGEMENTS

The author thanks Dr J. C. Scheepers for comments and suggestions and Mr S. J. Crafford for writing the program DATATRAN.

REFERENCES

- HILL, M. O., 1979a. *DECORANA*. A FORTRAN program for detrended correspondence analysis and reciprocal averaging. Unpublished report. Ecology and Systematics, Cornell University, Ithaca, New York. pp. 1-30.
- HILL, M. O., 1979b. *TWINSPAN*. A FORTRAN program for arranging multi-variate data in an ordered two-way table by classification of the individuals and attributes. Unpublished report. Ecology and Systematics, Cornell University, Ithaca, New York. pp. 1-48.
- WESTFALL, R. H., DEDNAM, G., VAN ROOYEN, N. & THERON, G. K., 1982. PHYTOTAB. A program package for Braun-Blanquet tables. *Vegetatio* 49: 35-37.

R. H. WESTFALL*

* Botanical Research Institute, Department of Agriculture & Water Supply, Private Bag X101, Pretoria 0001.



3.2.4.3.1 Improvements

A program PHYTOFORM written for the Sharp PC 1500 in BASIC converts the sampling unit data (sub-plots), for each stand to a single relevè representing each stand. The conversion is based on the mean cover for each species recorded for the stand, thereby reducing the sampling unit species lists to a single species list for each stand. Transfer to the host computer is the same as for the original data. Thus the host computer can have two data sets, namely, one with relevè data and the other with sampling unit data.

3.2.5 Habitat data

Habitat data are the environmental data, both biotic and abiotic, which describe the environment of a plant community and which can explain the floristic differences between plant communities. These data are usually recorded in the field.

3.2.5.1 Background

Werger (1974) states that "In the Zürich-Montpellier approach it is empirically determined that patterns in floristic composition correspond with patterns in the environment". The National Working Group for Vegetation Ecology in South Africa, for example, suggested minimum environmental parameters for vegetation and environment correlation and prediction of floristic potential. These are:

- i. sample number;



- ii. sample co-ordinates (latitude, longitude);
- iii. date;
- iv. altitude (above sea level) in metres;
- v. aspect of sample in degrees;
- vi. slope inclination in degrees;
- vii. lithostratigraphy;
- viii. percentage rock and outcrop cover;
- ix. geomorphology;
- x. vegetation structure according to Edwards (1983); and
- xi. land use and management.

The first three items are extremely important, especially if the site is to be re-sampled to monitor change. However, certain questions relating to the aims of a study need clarification. If an environmental description of the vegetation is all that is required then the environment could be described in descriptive terms. A descriptive account of the environment could be improved by standardization of terms to facilitate comparisons with other studies. However, if plant communities are to be correlated with environmental factors, to determine differentiating causes, then sampling of environmental factors is necessary.

Is sample size the same for both vegetation sampling and environmental factor sampling? If the sample size for vegetation sampling does not meet statistical criteria then how can a sample of, for example, soil depth be adequate using the same criteria? Mappable habitat factors, such as lithostratigraphy and soil type should not require statistical sampling if these have already been



mapped. However, factors such as aspect, slope and soil depth are sampled by what are essentially point samples, which cannot be equated with the samples used for vegetation. Cognizance should also be taken of scale when considering environmental factors. A hill could be regarded as a mountain when scale is increased. What environmental factors should be sampled and what can be derived from other sources?

These questions could affect possible community and environment correlations and hence influence the value of a classification.

3.2.5.2 *Methods applied*

Habitat data were determined from two sources, namely, direct measurements in the field at the time of sampling and indirectly, as derived data from topographic and geological maps.

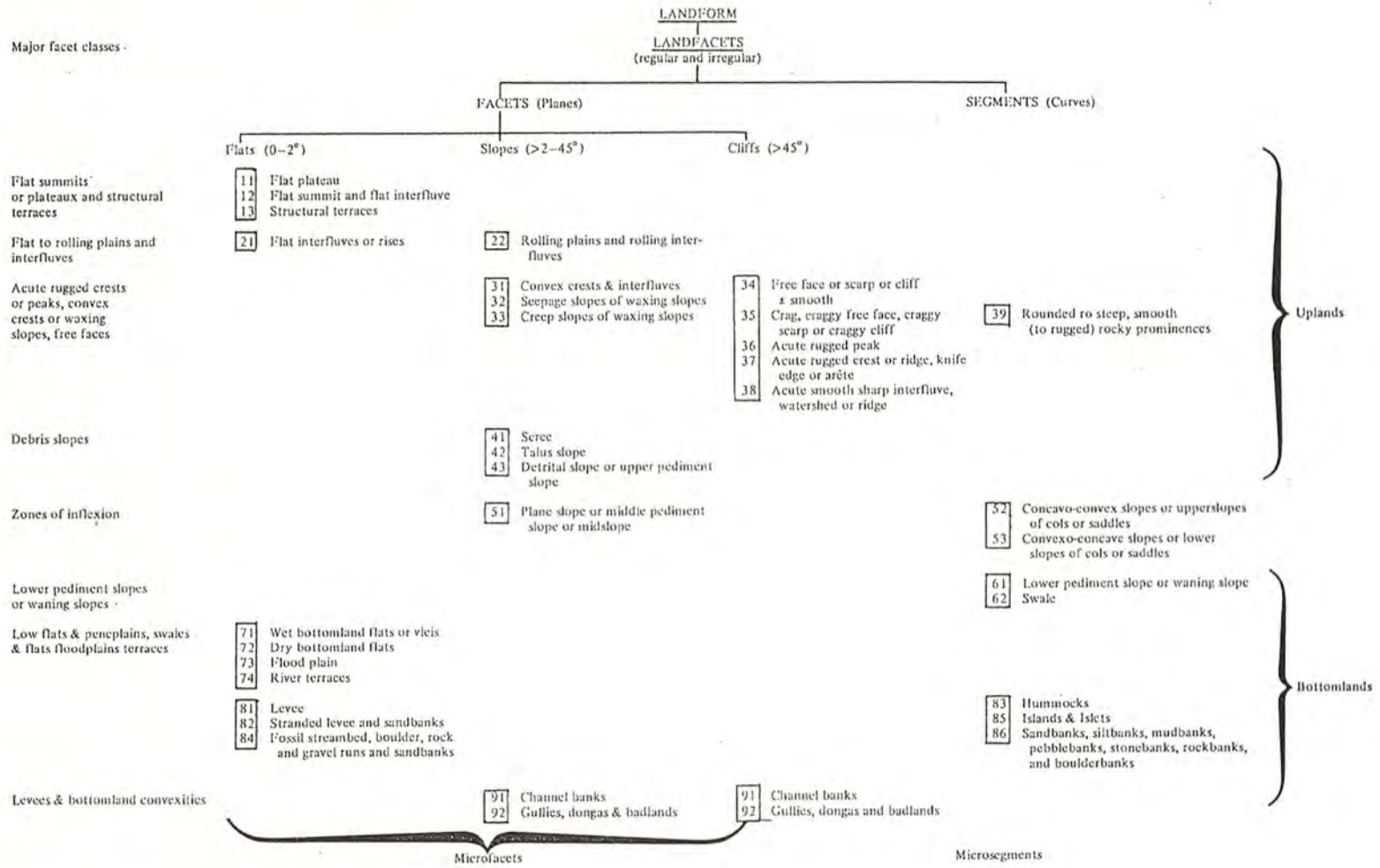
3.2.5.2 A) *Field data*

The disparity in area between stand and sampling unit for this study, the extent of a particular variable's influence on the plant community and sampling considerations necessitated the categorization of habitat data into two groups for recording purposes, namely, those suitable for stand characterization and those suitable for sampling unit characterization.

a) Stand data

- i. Lithostratigraphy, being formation and surface rock type

TABLE 3.2. - Geomorphology classes in terms of facets and segments from Scheepers¹ (unpubl.)



¹ J.C. Scheepers, Private Bag X05, Lynn East, 0039

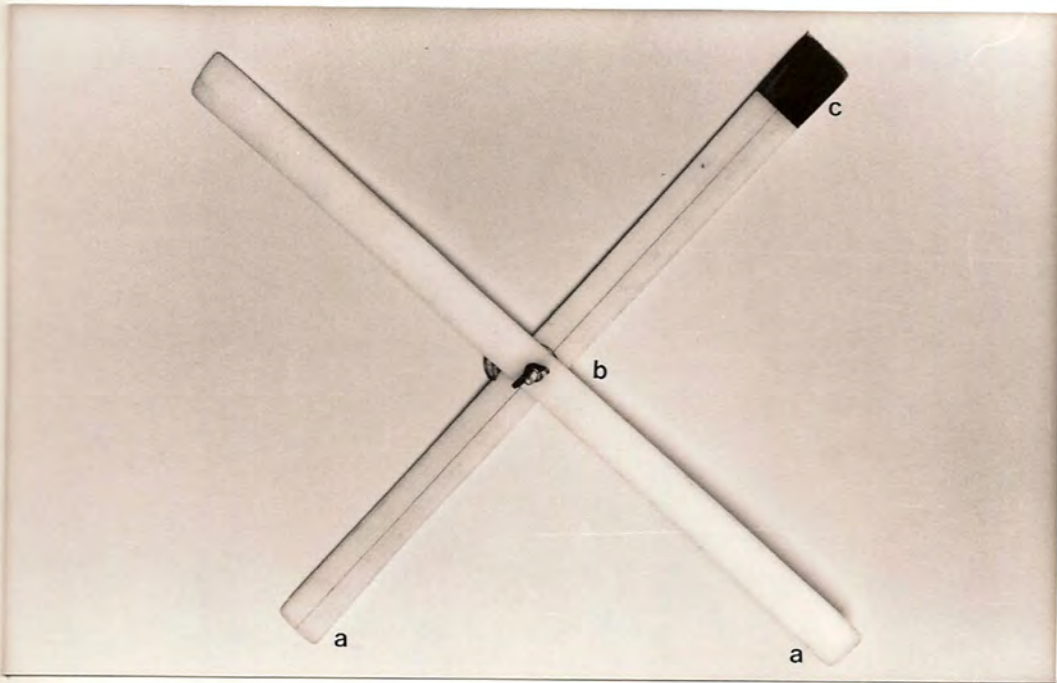


FIGURE 3.3. - A flexible plastic cross for selecting random pairs of plants in ground layer vegetation. a) intersecting arms, b) wing nut pivot - the first component of the plant pair is that which is nearest to the pivot, and c) marked - the second component of the plant pair is that which first intercepts the arm or continuation thereof.



FIGURE 3.4. - A tripod-mounted, manual spinner for selecting consecutive plants in pairs, at random, in vegetation above the ground layer. a) spinner, b) sight, and c) telescopic tripod.



TABLE 3.3. - A dual level system of land use categorization

<u>Land use:</u> (level 1)	<u>Primary</u> <u>symbol</u>	<u>Land use:</u> (level 2)	<u>Secondary</u> <u>symbol</u>
<i>Vegetation utilization</i>		<i>Production:</i>	
None observable	A	Cattle	A
Component mostly utilized:		Game	B
Grass	B	Goats	C
Herbs	C	Horses	D
Shrubs	D	Ostriches	E
Trees	E	Sheep	F
All	F	Other stock	G
<i>Vegetation replaced by:</i>		Bark	H
Pasture - dry land	I	Hay	I
Pasture - irrigated	J	Flowers	J
Crops - dry land	K	Fruits	K
Crops - irrigated	L	Leaves	L
Plantation	M	Root	M
Exotics (other)	N	Seeds	N
<i>Construction:</i>		Wood	O
Dam	Q	Cereals	P
Road - unsurfaced	R	Vegetables	Q
Road - surfaced	S	Alien	R
Footpath/track	T	<i>Waste:</i>	
Railway	U	Solid	W
Single storey	V	Liquid	X
Multi-storey	W	Gas	Y
Cleared	X		



according to Jansen (1982) with field verification of out-crops.

- ii. Altitude, being the altitude of the stand centre in metres, taken from the relevant 1:50 000 SA topo series map, to the nearest half-contour i.e. 10 m or 25 feet.
- iii. Geomorphology class, as visually categorized according to Table 3.2 .
- iv. Vegetation structure, being the total canopy cover for each of the height classes suggested by Edwards (1983) as determined by a cover meter (Westfall & Panagos 1984). Five random samples were taken in each height class. For the height classes up to one metre, a plastic cross (Figure 3.3) was arbitrarily thrown and for the height classes above one metre, a tripod-mounted spinner (Figure 3.4) was used to select random pairs.
- v. Land use, being a visual assessment according to the categories in Table 3.3 .
- vi. Insolation exposure, being 180 degrees minus the sum of the angles of the eastern and western horizons from the horizontal, at the stand centre and measured with an inclinometer. The directions of the eastern and western horizons, for the study area, were taken as due east and due west less the difference in degrees between the stand centre latitude and the Tropic of Capricorn. Maximum solar insolation is thus taken as being at the summer solstice.
- vii. Temperature of bare ground, ground layer vegetation and vegetation at chest height, being the mean of the measurements at five arbitrarily selected sites for each category around the stand centre. Temperatures were measured with an Instatherm



non-contact, infra-red thermometer. The temperature of a standard grey scale card was also measured at each stand for comparative purposes. These data are intended to assess the effect of vegetation cover on soil temperature.

viii. Grazing, as visually assessed according to the following four-class scale:

- 0, ground layer grazed short (less than 100 mm);
- 1, ground layer grazed short and patchy;
- 2, ground layer grazed evenly; and
- 3, no grazing evident.

ix. Browsing, as visually assessed according to a two-class scale:

- 0, browse line evident; and
- 1, browse line not evident.

x. Erosion, as visually assessed according to the following four-class scale:

- 0, donga erosion evident, by donga or gully formation;
- 1, sheet erosion evident, by general soil accumulation in a similar direction against plants;
- 2, donga and sheet erosion evident; and
- 3, no erosion evident.

xi. Fire, as visually assessed according to a two-class scale:

- 0, evidence of fire present; and
- 1, no evidence of fire present.

b) Sampling unit data

- i. Slope inclination, from the horizontal in degrees as determined with an inclinometer.
- ii. Aspect of inclination, in degrees as determined with a magne-



TABLE 3.4. - The "sausage" method of determining the clay percentage or texture of soil at field capacity (from F.S.S.A. 1974)

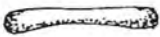
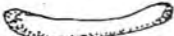



No sausage	Sand	<10% clay
	Loamy sand	10 - 15% clay
	Sandy loam	>15 - 20% clay
	Sandy clay loam	>20 - 35% clay
	Sandy clay	>35 - 55% clay
	Clay	>55% clay

TABLE 3.5. - The "finger test" method of determining the clay percentage or texture of soil at field capacity (from F.S.S.A. 1974)

Fingers stay clean	0 - 6% clay
Fingers slightly dirty	>6 - 12% clay
Fingers slightly sticky	>12 - 20% clay
Slight glossy smear and concretions	>20 - 35% clay
Highly glossy smear	>35% clay



- tic compass and allowing for magnetic deviation.
- iii. Percentage surface rock and outcrop cover, as a visual estimate.
 - iv. Soil surface compaction, being the mean of five measurements at the sampling unit centre and corners, using a pocket penetrometer, calibrated in kg.cm^{-2} .
 - v. Soil depth, as determined with a soil auger at the sampling unit centre, if the soil depth was greater than 300 mm or as the mean of five depth spike measurements at the sampling unit centre and corners, if the soil depth was less than 300 mm. These recordings are to the nearest 10 mm.
 - vi. Soil texture, being the mean of the class midpoints of the percentage clay in the A horizon, as estimated using both the "sausage method" (Table 3.4) and the "finger test" method (Table 3.5).
 - vii. Soil colour of the A and B horizons, determined at field capacity with a soil colour chart (Munsell Soil Color Charts 1954).
 - viii. Soil form, as determined by diagnostic horizon combinations according to MacVicar *et al.* (1977).
 - ix. Litter cover, as an estimated percentage of soil surface covered.
 - x. Litter depth, being the mean of five depths measured in mm to the nearest 10 mm at the sampling unit centre and the four corners.
 - xi. Relative biomass of the ground layer, being the mean of five measurements at the sampling unit centre and corners, using a disc pasture meter (Trollope & Potgieter 1983), calibrated



in cm. Conversion to $\text{kg}\cdot\text{ha}^{-2}$ requires correlations with clipped, dry material. However, relative comparisons can be made between stands directly, although less precise.

A program HABIMEAN written in BASIC for the Sharp PC 1500 converts the sampling unit habitat data to stand data. The conversion is generally based on mean values which are transformed to SI units, where applicable. In the case of soil form, the percentage occurrence of the soil forms in the stands is calculated. Mean aspect is calculated as the mean of the vectors of slope and aspect for each sampling unit. Mean soil colour is taken as the means of the Munsell values for brightness, saturation and hue, to the nearest Munsell class. This program does not form part of the PHYTOTAB-PC program because of the difficulty in standardizing environmental variables.

3.2.5.2 B) *Derived data*

Derived data are those habitat data for each stand which, in this case, were derived from topographic maps. The program SIDA is not included in the PHYTOTAB-PC package because current developments in extracting such information from geographic information systems will make this approach obsolete. The program description is, nevertheless, included here to show the type of information that can be derived from topographic maps. Some of this information is not usually recorded in the field because of difficulties in field determinations, such as stand position relative to watershed and drainage line.



The co-ordinates, in degrees, minutes and seconds generated with the program PHYTOLOC (section 3.1.6) for stand centre location were plotted on the relevant 1:50 000 SA topo series sheets. The circumference of each stand (5 mm radius at 1:50 000 scale) was circumscribed around the stand centres and the relevè number for each stand was noted on the sheets. The following was then determined from the topo sheets for input to the program SIDA written in BASIC for the Sharp PC 1500 computer:

a) Input

- i. Stand or relevè number.
- ii. Contour interval, in feet or metres, depending on the contour interval used on the relevant topo series sheet.
- iii. Sheet grid number, for reference purposes.
- iv. Latitude, in degrees, minutes and seconds of the stand centre.
- v. Altitude of the stand centre, in feet or metres (map dependent) to the nearest half-contour.
- vi. Aspect, in degrees from north of the stand inclination being the direction from the stand centre to the nearest point on the lowest contour, most adjacent to the stand circumference. This forms the aspect line.
- vii. Lowest contour altitude, in feet or metres of the lowest contour most adjacent to the stand circumference.
- viii. Lowest contour distance, in mm from the stand centre to the nearest point on the lowest contour, most adjacent to the stand circumference.
- ix. Highest contour altitude, in feet or metres of the highest contour most adjacent to the stand circumference.



- x. Highest contour distance, in mm from the stand centre to the nearest point on the highest contour, most adjacent to the stand circumference.
- xi. Watershed distance in mm, to the nearest half-contour from the stand centre along the highest contour distance line to the nearest crest or watershed. This is the watershed line. If the stand is situated on a crest then the watershed line is to the nearest drainage line approximately opposite the first drainage line (see xiii) so that the stand position can be related to two drainage lines.
- xii. Watershed altitude, in feet or metres, of the watershed or drainage line.
- xiii. Drainage distance, in mm, to the nearest half-contour, from the stand centre along the lowest contour distance line to the nearest drainage line. This is the drainage line. If the stand is situated in a depression then the drainage line is to the nearest watershed approximately opposite the first watershed line (see xi) so that the stand position can be related to two watersheds.
- xiv. Drainage altitude, in feet or metres of the drainage line or watershed.
- xv. Angle horizon east or west. The directions of the eastern and western horizons are due east and west plus the difference between the stand centre latitude and $23\frac{1}{2}^{\circ}$ south (as in section 3.2.5.2 A) (a) vi). The horizon angles are determined by a subroutine of the program SIDA when contour line altitudes and distances from the stand centre, in mm, along the directions of the eastern and



western horizons, are input. The subroutine displays extrapolated altitudes so that the user can determine the number of contour lines to be input. Contour lines lower than the stand are ignored.

- xvi. Cumulative distance, in mm, being the distance from drainage line to watershed or if watershed altitude is less than the stand centre then to the highest point along the watershed line.

b) Output

The output is based mainly on simple trigonometric functions of the input variables. The following are printed for each stand:

- i. stand or relevè number, as input;
- ii. grid number, as input;
- iii. altitude, in metres;
- iv. aspect, in degrees;
- v. mean soil depth, in mm, based on a linear regression of the combined slope and position between watershed and drainage line and measured soil depth;
- vi. mean annual rainfall, in mm, based on a linear regression of the combined altitude and latitude and the recorded rainfall for officially recognized rainfall stations in the study area where recordings generally exceeded 20 years;
- vii. insolation exposure, in degrees;
- viii. relative available moisture, in mm, as a function of mean annual rainfall, slopes, aspect, position in the landscape, altitude, latitude, insolation exposure and mean soil depth;

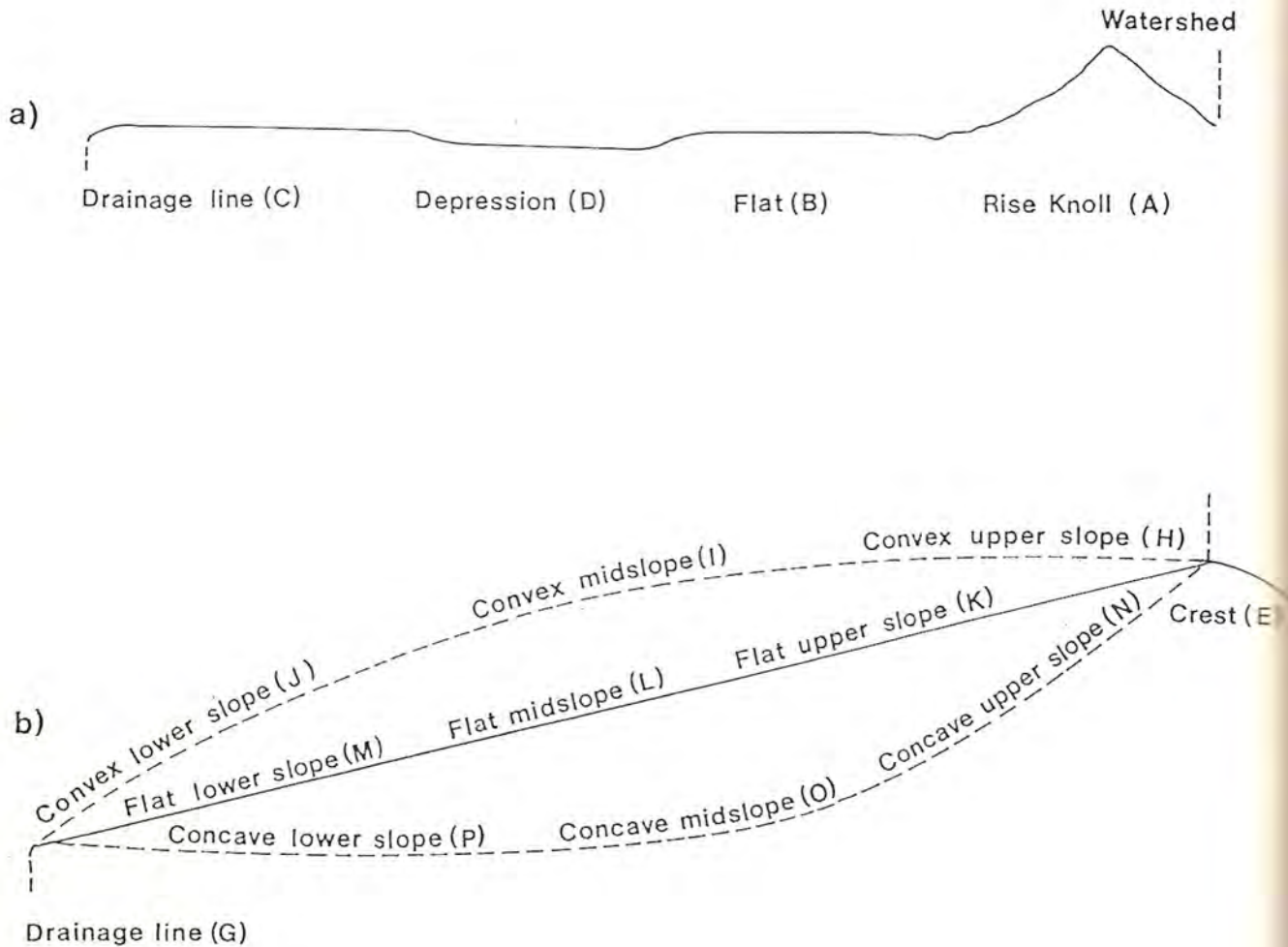


FIGURE 3.5. - Geomorphology classes determined with the SIDA program with class codes in brackets. a) Landscape slope $\leq 2^\circ$ b) Landscape slope $> 2^\circ$. Slope categorization is restricted if the slope is less than 12x stand radius (R) to: crest/koppie if slope $< 4R$; midslope if slope $< 8R$; lower slope and upper slope if slope $< 12R$.



- ix. topographic profile, from watershed to drainage line with the stand diameter proportionally indicated according to position;
- x. topographic profile slope (mean), in degrees;
- xi. stand slope, in degrees;
- xii. scale of topographic profile;
- xiii. geomorphology, in classes as listed in Figure 3.5; and,
- xiv. stand drainage class according to the topographic profile and stand slopes where:
 - A Outflow: Run-off > Run-in
 - B Throughflow: Run-off = Run-in
 - C No flow: Run-off = 0; Run-in = 0
 - D Inflow: Run-off < Run-in.

3.3 CLASSIFICATION

Classification in this sense refers to the ordering of species presence/absence data in a matrix form where columns represent relevès (sampled stands) and rows represent species. The matrix values at the intercepts of columns and rows indicate presence and quantity by a cover or cover-abundance symbol.

3.3.1 Background

Gauch (1982) defines classification as "grouping similar entities together in clusters". The New Collins Concise Dictionary of the English Language (1985) defines classification as "the placing of animals and plants in a series of increasingly specialized groups because of similarities in structure, origin, etc. that indicate



common relationships". The first definition seems more suited to a description of scatter diagrams, whereas, the second definition appears more appropriate to the needs for a vegetation classification, such as data reduction, identification and diagnosis of hierarchically grouped vegetation units and to show the relationships between and variation within, such units.

The procedure proposed by Mueller-Dombois & Ellenberg (1974) entails identification of "common species groups as a key to grouping relevès". The concatenation of relevès in which selected species occur, to achieve a relevè sequence, gives rise to the problem of diagnostic and non-diagnostic species recognition. In this sense, diagnostic refers to those species whose distributions are influenced by environmental gradients within the study area, at the scale concerned. They, therefore, have limited but definite distributions in the study area and can be used to delimit plant communities at any level in the hierarchy. Non-diagnostic species are not thus influenced and their limits within the study area are not clear. Clearly, reliable recognition of these two species groups can only be achieved after the classification process. If a set of species were chosen, for example, from the non-diagnostic species-group for initial relevè concatenation, a totally different classification could emerge, compared to that based on a set of diagnostic species. In practice, however, relevès are often grouped according to noticeable environmental gradients as a first step in sequencing relevès. Programs such as TWINSPAN (Hill 1979), PHYTO 20 (Westfall *et al.* 1982) and PHYTO 21 (Westfall & De Wet 1988) can also be used for initial classification but these programs do

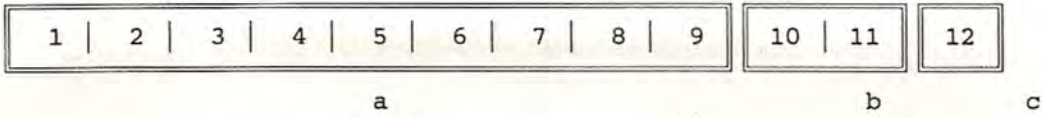


FIGURE 3.6. - Communities (numbered) sequenced according to the main environmental gradient responsible for their differentiation (a); subsidiary gradients (b and c) in which communities are sequenced similarly but inserted after the main environmental gradient.

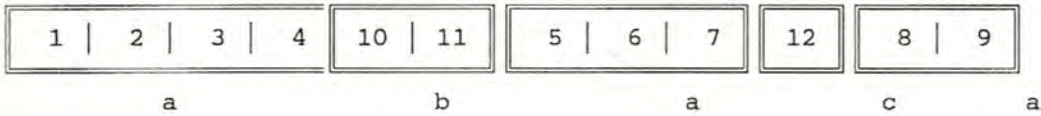


FIGURE 3.7. - Communities (numbered) sequenced according to the main environmental gradient responsible for their differentiation (a); subsidiary gradients (b and c) in which communities are sequenced similarly but inserted in the positions to which they correspond on the main environmental gradient.



not meet the needs for a final classification and further sequencing is generally required.

In a matrix classification the relevè number sequence is linear. This can pose problems where multiple environmental gradients are responsible for community differentiation. Should communities differentiated by subsidiary gradients be placed at the end of the communities differentiated by the main gradient or inserted within the communities of the main gradient at their respective positions on the main gradient? The last-mentioned possibility is preferred, as will be shown later. However, it is often likely that the two relevès adjacent to the point of insertion, of a community differentiated by a subsidiary gradient, are more similar to each other in terms of floristics than they are to the relevès to be inserted. These relationships are illustrated schematically in Figures 3.6 and 3.7.

In sequencing relevès according to similarity, an additional problem was found, namely, the selection of the first relevè with which the other relevès are to be compared. If all the relevès are tested in the first position and each relevè is compared with the first and subsequent relevès, then the number of tests is a factorial of the total number of relevès. The magnitude of this number is prohibitive even if mirror-image sequences are excluded. Furthermore, how can the "best" sequence be determined if only two relevès are compared at a time and not the entire sequence? What is required is a value relating to the effectiveness of a particular sequence of relevès.

The questions that arise from an objective approach to a Braun-Blanquet classification are:

- i. given that no two plant species have exactly the same distribution, can a particular classification be regarded as unique and, if not, what are the alternatives?
- ii. are the community relationships in the classified matrix data linear, as suggested by Poore (1956)?
- iii. what is the basis of the classification? Should species be sorted before relevès according to similarity in distribution as suggested by Werger (1974)?
- iv. is it possible in a data set with little or no observable discontinuities for two independent workers to achieve the same classification? It should be possible where discontinuities are obvious; and,
- v. can the classification process be automated thereby reducing decision-making in the classification process?

Subjective sampling followed by automated classification will not make the classification objective. For a classification to be objective and hence repeatable, the processes of stratification, sampling and classification should not have observer bias.

3.3.2 Methods applied

The approach adopted in this study, is in accordance with the needs for classification and has the following aims:

- i. to obtain a relevè sequence where (a) relevè-groups can be formed, based on floristic similarity; and (b) the relevè-



groups thus formed are also sequenced according to floristic similarity. This is not necessarily the same as simply sequencing the relevès according to similarity because of the possible insertions of subsidiary gradients;

- ii. to delimit relevè-groups; and,
- iii. to obtain a species sequence where the relevè-groups and their relationships are emphasized.

3.3.2.1 *Relevè sequencing*

The refinement of a preliminarily sequenced, relevè sequence often depends on an intuitive assessment and reduction of gaps in the matrix. The gaps are included blanks, between the first and last occurrence of a species in the matrix, caused by the absence of that species in the relevant relevès. The gaps before and after the first and last occurrence, respectively, of a species in the matrix are not taken into consideration, so that the sequence of relevès can affect the total. If all gaps were considered then the total would be the same irrespective of relevè sequence. The gaps can be caused by sampling omissions, outliers and, more often, the irregular spacing of many plant species. Visual sequencing of relevès according to gap minimization from the outset is, however, so difficult as to be virtually impossible.

Redundancy is the occurrence of species in two or more relevès to form pattern (Gauch 1982). Pattern can not be adequately quantified because value judgements have to be taken on pattern formation. Judgements can be based on a combination of number of



species and relevès and the constancy of species occurrence in the relevès forming the pattern. These judgements are subjective.

Gauch (1982) defines pattern as the co-ordinated occurrence and "noise", the unco-ordinated occurrence of species in the matrix. He regards "noise" and pattern as opposites, so that if "noise" is reduced in a matrix, then pattern is enhanced. Although Gauch (1982) does not regard "noise" as quantifiable, the effect of noise can be attributed to gaps in the matrix in terms of included blanks. The included blanks are species absences between the first and last occurrence of each species in the matrix, for a given relevè sequence. This can be quantified by addition.

For a given relevè sequence the included gaps for a single species are referred to as "separation units" and for all the species as "total separation units". The total separation units are inversely proportional to redundancy in the matrix and hence are related to classification adequacy. The optimum classification is that where the total separation units are a minimum. This presupposes that each relevè be tested in each possible position. The resulting possibilities are: $n!$, where n is the total number of relevès. A program was written to determine the time it would take to test these possibilities on a Burroughs B7900 mainframe computer, for a small matrix of 25 relevès x 33 species. The answer was in excess of 2 000 million, million years. Even excluding mirror-image sequences this is impracticable.

A heuristic approach, where not all, but only the best possibili-



ties are tested, was adopted. Clearly, a similarity index could also play a role in excluding many possible permutations. However, the initial relevè sequence could still affect the final sequence. It was found, for example, that the final sequence using TWINSPAN (Hill 1979) could change when the initial sequence was altered (see Chapter 4). A solution to this problem could be to standardize the initial sequence in as unique a manner as possible.

3.3.2.1 A) *Commonality sequence*

The sampling sequence, and hence relevè numbering, is often highly arbitrary. The aim of standardizing the relevè sequence is, therefore, to obtain a sequence in which the first relevè would be the same, irrespective of the sampling sequence. This implies that each relevè should have some sort of unique attribute independent of any relevè sequence. A method of achieving this aim, is to count the occurrences in the matrix, of all the species in a relevè. This is done for each relevè in the matrix. The relevès can then be sequenced, either ascending or descending, according to the totals for each relevè. The extremes of this sequence, have in tests, proved to be unique although some similarity can exist between some intermediate relevès. This sequence is called the commonality sequence because the one extreme of the sequence represents relevès which have the most generally occurring species in common and the other extreme the least generally occurring species in common.



TABLE 3.6.- Illustration of the possible permutations of species presence and absence (rows) in three relevès (columns). Presence is indicated by a "+" and absence by a "0".

		Columns		
		1	2	3
Rows	1	+	+	+
	2	+	+	0
	3	0	+	+
	4	+	0	+
	5	+	0	0
	6	0	+	0
	7	0	0	+
	8	0	0	0



3.3.2.1 B) *Similarity sequence*

A problem encountered with similarity co-efficients such as those of Jaccard, Sorensen, Ellenberg and Gleason (Mueller-Dombois & Ellenberg 1974) is that relevès are compared in pairs in isolation, with no cognizance taken of the matrix of which they form a part. For example two adjacent relevès in a matrix, representing two different communities could have an extremely low similarity co-efficient. This is because of the weight attached to negative association, generally found in similarity co-efficients.

If relevès are considered in three's (Table 3.6), then in terms of the central relevès position in the matrix, the following can be stated:

- i. positive associations exist in the case of the first three rows;
- ii. a possible positive association could exist if single absence resulted from sampling omission or irregular spacing, in the fourth row;
- iii. no positive association exists, but negative association can not be shown because of sampling omissions or irregular spacings in the fifth to seventh rows; and,
- iv. absence in all three columns could be considered a similarity but not a negative association, in the eighth row.

Therefore, if only the first two relevès are considered (Table 3.6) then positive associations exist for rows 1 and 2 (joint occurrences), possible positive associations exist for rows 3, 4, 5 and 6 (absences could be because of sampling omissions), and no



negative association exists in rows 7 and 8 (joint absence could indicate similarity).

Similarity weighting for a pair of relevès is, therefore, based on positive and possible positive associations with no negative values given, as negative associations cannot be shown for two relevès. The weightings given are as follows:

Positive association 2

Possible positive association 1.

The similarity co-efficient (C) for comparing two relevès in a matrix is, therefore:

$$C\% = \frac{1/2 S + J}{S + J} \times 100 \quad \text{where}$$

S = number of single occurrences in both relevès and

J = number of joint occurrences.

3.3.2.1 C) *Separation unit sequence*

After standardization of the sampling sequence of relevès and sequencing according to similarity, the total separation units for the matrix is calculated. The first relevè is then moved, successively, one position in the relevè sequence and the total separation units are calculated after each move. The relevè sequence with the lowest total separation units is retained and the process is repeated with the new relevè sequence. This results in $n^2 - n + 1$ calculations of total separation units, where n is the total number of relevès, after all the relevès have been moved in this manner.



The relevè sequence with the lowest total separation units is then reversed and the entire process is repeated successively until no decrease in total separation units is achieved. The reversal of the relevè sequence tends to average the position of a relevè where more than one position for that relevè exists.

The last retained sequence is then used to determine new similarity and separation unit sequences and these processes are iterated until the total separation units increase when the sequence with the lowest total separation units is reversed and the iteration then continues. When the lowest total separation units have been achieved in this manner, the iterations are terminated and the relevant relevè sequence is retained. This relevè sequence should then represent a matrix with a minimum of included blanks. Species occurring in all or a single relevè do not affect the position of the relevès in which they occur, in the relevè sequence.

3.3.2.2 *Relevè grouping*

After the relevès have been sequenced the next step in the classification process is to delimit relevè-groups, some of which could represent plant communities. If the optimum relevè sequence has been achieved then it could be expected that the relevès that best represent a relevè-group would be situated towards the middle of the relevant relevè-group, because of a sine curve effect of the increment and decrement of species. Furthermore, if the relevè-groups are differentiated by different species then the first and last occurrence of each species distribution in the matrix could be



significant, in terms of relevè-group delimitation. Thus a sine curve, or approximation thereof, could be expected from the values obtained by the difference between the number of species representing first and last occurrences in the matrix, for each relevè. Species occurring in all the relevès or in a single relevè are ignored.

The number of relevè-groups to be recognized depends on the degree of difference permitted between relevès. Clearly, this will relate to the scale of the study, and the number of relevè-groups should be closest to the original number of vegetation units stratified, for objectivity. Thus, scale, based on the stratification process, is taken into account in the classification process. However, the programs allow flexibility in the number of relevè-groups that can be recognized so that data can be classified, even if the stratification and sampling processes are not according to scale. Relevè-group delimiters are inserted into the relevè sequence automatically, at the appropriate positions.

3.3.2.3 *Species sequencing*

The final step in vegetation classification is to sequence species so that the relevè-groups and their relationships are emphasized. This is achieved by increasing pattern through species sequencing whereby species with similar distributions according to relevè-groups or combinations of relevè-groups are placed together in species-groups. A balance must also be achieved between outliers and redundancy or pattern. Various combinations of outliers to re-



dundancy were tested without definite ratios being evident. In terms of total separation units, however, single occurrences of species in the matrix are not significant as they cannot contribute to pattern. Similarly a single occurrence of a species in a relevè-group or combination of relevè-groups cannot be considered significant for the group or groups of relevès because no pattern is formed. Therefore, minimum species occurrence is taken as two, for a relevè-group or combination of relevè-groups.

Outliers are also limited to single occurrences in any relevè-group, provided that the total number of outliers does not exceed 38% of the total for that species in the matrix, because occurrences of more than one in a relevè-group could contribute to pattern. The 38% limit for outliers is based on tests and also corresponds with the Fibonacci fraction. The relationship between outliers and pattern can be considered to be based on minimum outliers. It must be emphasized that "noise" in terms of included blanks is present regardless of species sequence and that sequencing species does not affect relevè grouping or the relationships between relevè-groups. Furthermore, the maximum number of outliers that can be allowed in a given matrix is dependent on an appreciation of the pattern improvement that can be effected. This is obviously subjective, but does not affect the relevè grouping or the relationships between relevè-groups, as has been stated.

Species are also sequenced in descending order of occurrence in each species-group in the diagnostic species portion of the matrix. Species-group order is according to presence in:

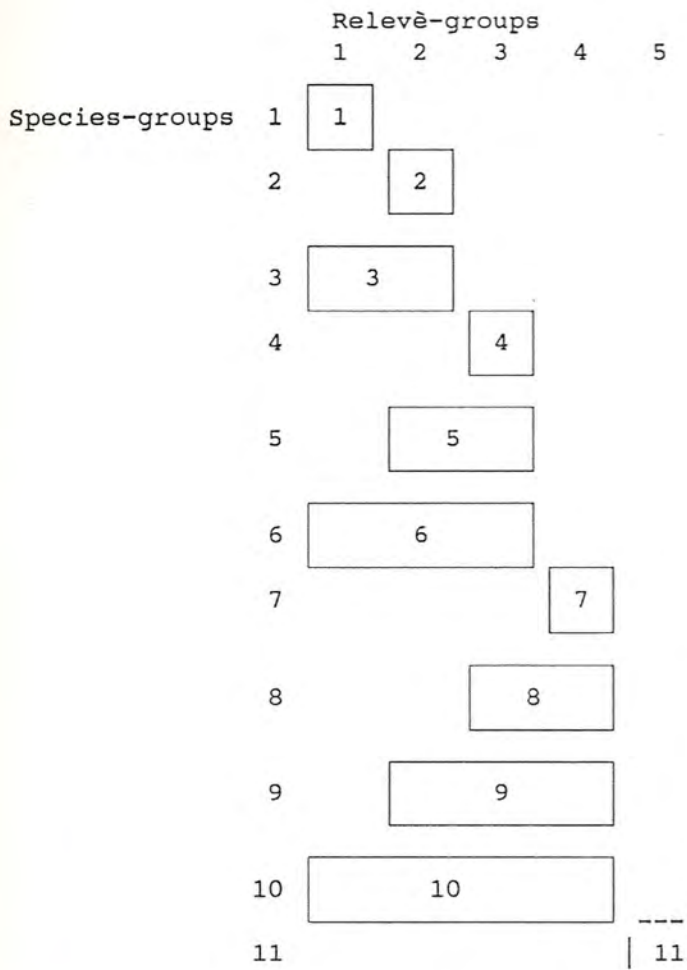


FIGURE 3.8. - Simplified schematic diagram showing the sequence of species-groups construction, according to the relevè group sequence.



- i. first relevè-group;
- ii. second relevè-group;
- iii. first and second relevè-groups;
- iv. third relevè-group;
- v. second and third relevè-groups; and,
- vi. first, second and third relevè-groups and so forth as illustrated in Figure 3.8.

Species-group delimiters are then inserted into the species sequence.

Species-groups with a single species can occur. The user should determine whether or not this is significant and either retain or combine such species-groups. The flexibility that is possible with species sequencing can be carried even further by combining species-groups, consisting of two or more relevè-groups, to correspond with the limits of environmental gradients (section 3.5.2.1 C). This simplifies the matrix, in terms of species-groups, without loss of information.

3.4 VERIFICATION

A classification should be tested to ensure its validity, because more than one solution to a classification could be possible. This testing procedure can be referred to as a verification process.

3.4.1 Background



Wenger (1974) stresses the need for confirmation of "coincidence between the pattern on the phytosociological table and specific habitat conditions". Apart from gaining insight into floristic and habitat relationships, the degree of "co-incidence" can also confirm a classification. However, if it is accepted that each species in the classified table has its own particular distribution according to factors which may limit it, then it is feasible that any combination of relevés could produce some or other correspondence with one or other environmental variable. Correspondence between community and habitat is, therefore, not necessarily verification of the validity of a classification, but can support a classification. The total separation units is a rating of the adequacy of a classification but the rating obtained is not necessarily the minimum total separation units because the relevés are not tested in every possible position. Other means should, therefore, be sought to jointly verify a classification.

3.4.2 Methods applied

The following sections describe some methods in which a classification can be verified.

3.4.2.1 *Classification efficiency*

A measure of the efficiency of a classification is the relationship between all the blanks (not just included blanks) in the matrix and the total separation units. This is a relative value, with which classifications can be compared, unlike the total separation



units, which is an absolute value. The classification efficiency (E) is calculated as follows:

$$E\% = 100 - \frac{(TSU \times 100)}{AG}$$

where TSU = total separation units and AG = all gaps in the matrix. A matrix with no noise and correctly classified would, therefore, have a classification efficiency of 100%. It appears that, in tests, adequately classified matrices have classification efficiency values of above 60%. Values markedly lower than this could indicate shortcomings in the classification process or data.

3.4.2.2 *Spatial relationships*

The next step suggested in verifying the adequacy of a classification is to determine the spatial relationships between relevè-groups. This can be done by using an overlay technique. In general relevè-groups should exhibit a relatively high degree of spatial integrity, except where the scale is such that mosaics can be formed by relevè-groups. A low degree of spatial integrity could indicate shortcomings in the classification process or data.

A measure of correspondence in the spatial relationships between the stratification process and the classification process can be indicative of classification adequacy. The degree of correspondence can be quantified by comparing the two number sequences; the one representing the relevès grouped according to the stratification; the other representing the relevès grouped according to the classi-



a)

Habitat variable ranges

1 Altitude (m ¹⁰)	40-45	35-42	25-42	31-47	55-68	53-70	52-65	60-83
2 Soil depth (cm)	10-14	11-13	15-20	16-18	10-13	12-16	17-18	17-19
3 Soil texture	Sand	Clay	Sand	Clay	Sand	Clay	Sand	Clay
4 Communities:	1	2	3	4	5	6	7	8

b)

Habitat variable ranges

2 Soil depth (cm)	10-14	15-20	11-13	16-18	10-13	17-18	12-16	17-19
1 Altitude (m ¹⁰)	40-45	25-42	35-42	31-47	55-68	52-65	53-70	60-83
3 Soil texture	Sand	Sand	Clay	Clay	Sand	Sand	Clay	Clay
4 Communities:	1	3	2	4	5	7	6	8

c)

Soil depth listed according to ascending range

10-13	10-14	11-13	12-16	15-20	16-18	17-18	17-19
-------	-------	-------	-------	-------	-------	-------	-------

FIGURE 3.9. - Hypothetical model illustrating the sequence of three environmental factors, namely, altitude, soil depth and soil texture to hierarchically differentiate eight communities (a). If soil depth is first in the sequence (b) then differentiation of only four groups of two communities each, occurs. This is because in (b) the ranges overlap so that no groups can be formed, as can be seen in (c), whereas, the division of soil depth by altitude in (a) creates non-intersecting sub-groups of soil depth.



ication. A program to determine correspondence between two sets of grouped number sequences is included in the PHYTOTAB-PC package. The degree to which the relevè-groups in both sequences intersect indicates the amount of correspondence. Furthermore, the stratification can form a basic hypothesis which is tested with the classification. A low degree of correspondence between the two processes can invalidate the hypothesis, whereas, the converse can also provide the basis for mapping.

3.4.2.3 *Floristic and habitat relationships*

Not only can an examination of floristic and habitat relationships aid in verifying the classification process but can also lead to an understanding of the causes of community differentiation. Obviously, the correspondence process is limited to those habitat factors which are available, such as those that have been recorded or derived. Factors not available, could cause problems in the correspondence process as the tested relationships between the habitat factors and floristics would then be incomplete.

Correspondence between floristics and habitat can be assessed with ordination techniques using programs such as DECORANA (Hill 1989b) and CANOCO (Ter Braak 1987). However, the former program requires overlaying selected environmental variables over the floristic scattergram to assess correspondence, and the latter program combines environment and floristics in a single classification process so that the classification is not based on floristics alone.

The following method is based on a simple hierarchical model which



tests direct correspondence between relevè-groups and the associated habitat factors available on a relevè basis. The first step is to construct a hypothetical model, as in Figure 3.9a. The hypothetical model is merely to determine the sequence of habitat factors, because the sequence, being hierarchical, is necessary in determining correspondence. This is illustrated in Figure 3.9b where fewer communities are differentiated because the sequence has changed. It is not possible because of permutation limitations to test all possible sequences programmatically, hence the necessity to construct the hypothetical model. Such a hypothetical model could also be of benefit, in terms of relevancy and economy, in determining which habitat factors to record in the field. The sequence of habitat factors should usually be from the general to the particular. For example, soil texture should not precede mean annual rainfall.

Starting from the first, most general, habitat factor, as determined with the model, the ranges of that habitat factor are determined for each relevè-group. Each unique set of ranges for one or more relevè-groups is allocated a different class symbol. If all the range sets have interceptions with other sets then a single class symbol is allocated and no correspondence for that habitat factor and the relevè-group is implied.

The next habitat factor is then tested in the same manner, but separately within each of the unique sets, determined with the previous habitat factor. The process is repeated for all the



habitat factors. Individual relevè-groups can be differentiated at various levels in the habitat factor hierarchy. If a poor differentiation of relevè-groups is achieved then the following conditions could apply:

- i. the classification is inadequate;
- ii. the habitat factor selection is inadequate;
- iii. the model is inadequate; or,
- iv. the limits for a habitat factor do not correspond precisely with the limits of a relevè-group. In other words, adjacent relevès, representing two communities have habitat factor values which do not correspond with the border between the two communities.

In the case of the last condition the process can be repeated using either nodal relevès or omitting relevès adjacent to relevè-group delimiters.

3.4.2.4 *Classification and field relationships*

The final step in the verification process is comparing the classified relevè-groups with the units to be mapped, on the ground. This step is not to confirm the occurrence, on the ground, of the plant species in the relevant relevè-groups, which should clearly be the case, but the following:

- i. to assess the degree to which the plant species of the relevant relevè-group are representative of the unit to be mapped;
- ii. to assess the variation represented by the relevès of the relevant relevè-group and the degree to which this corresponds



with the variation of the unit to be mapped, at the relevant scale;

- iii. to assess the value of plant species which can be used for plant community diagnosis as a means of identifying the plant community in the field. Here plant spacing, size and regularity of occurrence is important so that the community can be easily recognized;
- iv. to assess the reliability of the hypothesized floristic and habitat relationships;
- v. to assess the validity of the borders of the units to be mapped; and,
- vi. to assess which of the relevè-groups are mappable at the relevant scale.

If the classification has a classification efficiency of about 60% or more, does not invalidate the stratification, and corresponds well with the habitat data and the field assessment, then stratification, habitat correspondence and classification can all be deemed satisfactory. However, lack of correspondence in any of these processes does not necessarily invalidate the classification, but rather the process concerned could be questioned. The relevancy of the classification will ultimately depend on the uses derived therefrom.

3.5 DERIVATIVES

The term derivative is taken here, to mean information that can be derived directly (primary derivatives) or indirectly (secondary



derivatives) from the classification. Spatial relationships such as the locality of vegetation units or their co-occurrence with mappable factors such as geology or soils can be illustrated with maps, map overlays or integrated maps. Mapped vegetation units, as proposed in this study, are the product of the stratification process, which can be modified by the classification, rather than being derivatives of the classification.

3.5.1 Background

If a classification and a map illustrating spatial relationships of vegetation units are the end-products of a study then the value of such a study will be limited to those workers who are able to interpret and derive additional information from the classification. Those best able to interpret and derive information from a classification should be the compilers thereof. Information that can be derived from a classification will largely determine the uses for classifications and hence the need for such work. It is, therefore, of great importance to workers in this field to derive as much information as possible from classifications, so that the demand for such work can be increased.

3.5.2 Methods applied

The classified matrix can include a summary of coded variables for each habitat factor, above each relevè, to form a passive classification of habitat variables, (Deall & Westfall 1989) which can be visually correlated with the relevè-groups.



There is usually much redundancy in a classification and the information contained therein can be considerably reduced by constructing a synoptic matrix (Werger 1974) where relevè-groups are reduced to a single column according to constancy based on species presence. A five class scale is generally used (Mueller-Dombois & Ellenberg 1974). This tends to visually over-emphasize the importance of single occurrences which can be overcome by using:

- i. a "+" symbol for single occurrences; or,
- ii. the plant number scale described in section 3.2.3.

The classified habitat data and synoptic matrix, although dependent on the classification are not considered derivatives, in this study.

3.5.2.1 *Primary derivatives*

The primary derivatives that follow, can be derived from the synoptic matrix.

3.5.2.1 A) *Plant communities*

Braun-Blanquet (1928) defined a plant community as being characterized by "its own species". As more data were gathered, syntheses of matrices took place and community distinctions became less clear. Terms such as exclusive, selective and preferential character species as well as differential, territorial and regional species were introduced to clarify the situation (Werger 1974). Clearly, unless a plant community is restricted to a particular study area, its composition, diagnosis and description for



the study area, could differ considerably from that of the community of which it is a part. This relates to the proportion of the community variation present in the study area, which is usually not known where the community is not entirely included in a study area. The following is, therefore, suggested for individual study areas where no synthesis with other study areas is involved:

- i. a relevè-group represents a plant community only where the plant community is completely included in the study area concerned. A relevè-group cannot adequately represent a plant community that is only partially included in a study area, because the proportion of the variation represented by the relevè-group, is not known;
- ii. the species of the species-group that is restricted to the relevè-group representing the plant community, is the diagnosis of that plant community, in the study area, and are referred to as community diagnostic species;
- iii. formal ranking of plant communities should not be applied for individual study areas, but only with syntheses, because there is usually insufficient information to elucidate relationships;
- iv. informal ranking of plant communities within a study area is based on combined relevè-groups, where conditions i. and ii. apply for the combined group;
- v. relevè-groups without community diagnostic species can represent ecotones, extremes in variation for a particular community, or communities for which the particular scale or extent in the study area is inappropriate; or,

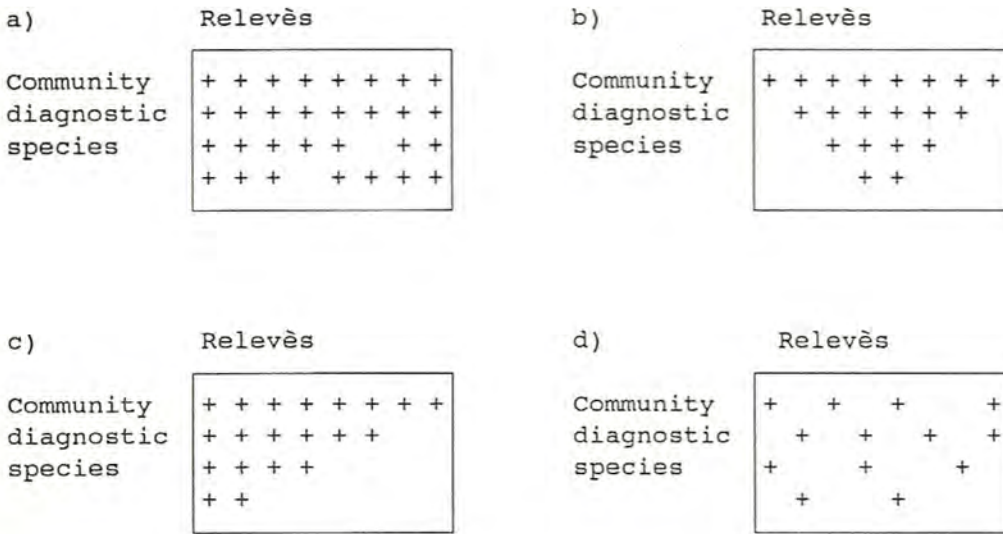


FIGURE 3.10. - Species presence in community diagnostic species-groups: a) few blanks indicating a well-defined community corresponding to an abrupt environmental change at the community limits; b) species presence approximates a normal curve, indicating a well-defined community but with a less abrupt change at the community limits than in (a); c) a variation of (b) where the relevès have been re-arranged so that the strongest community expression, in terms of species presence, is at the left of the group, decreasing towards the right. This pattern does not adequately show spatial relationships because the relevès on the extreme right can often be adjacent to different communities; and d) a poorly-defined community.



vi. plant community composition by all the species, both diagnostic and non-diagnostic, present in the relevè-group/s representing the plant community. This is, however, only a sample of the species present in the community.

This information as well as plant community names can be obtained directly from the synoptic matrix.

3.5.2.1. B) *Community definition*

Community diagnostic species pattern can be used to infer the degree to which the community is defined, if the relevè sequence is according to minimum total separation units. For example, community diagnostic species with no or few gaps, as illustrated in Figure 3.10a, should indicate a well-defined community which is differentiated by an abrupt environmental change at the community limits and can, therefore, be mapped with a high degree of precision.

A community where the community diagnostic species pattern approximates a normal curve (Figure 3.10b) should indicate a less abrupt environmental change at the community limits but the community is still well-defined. Mapping precision will be according to the width of the ecotone. Depending on sample skewness, the relevès best expressing community composition are those in the central part of the relevè-group. The relevès can be re-arranged within the relevè-group as illustrated in Figure 3.10c without affecting community relationships. However, the total separation units will then increase because the original relevè

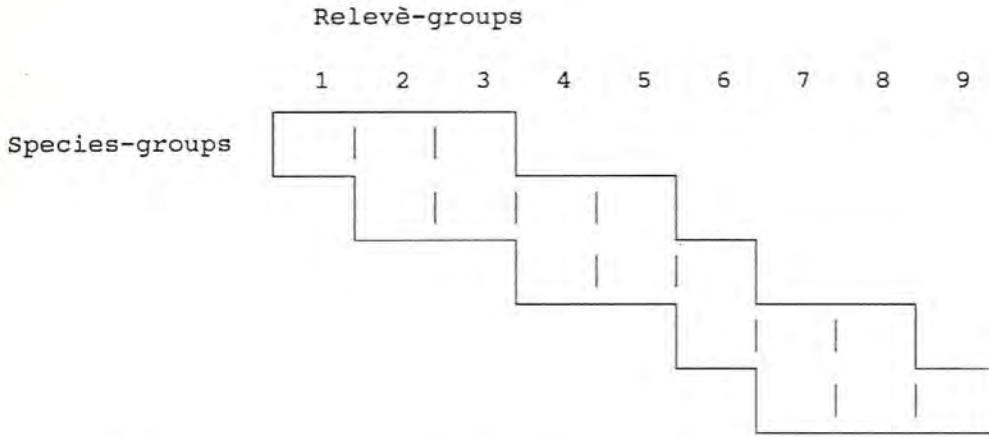


FIGURE 3.11. - Species-groups representing two or more communities, sequenced to correspond to an environmental gradient of which both the upper and lower limits for differentiated communities can be ascertained. This also applies to the horizontal mirror-image of the illustrated pattern.

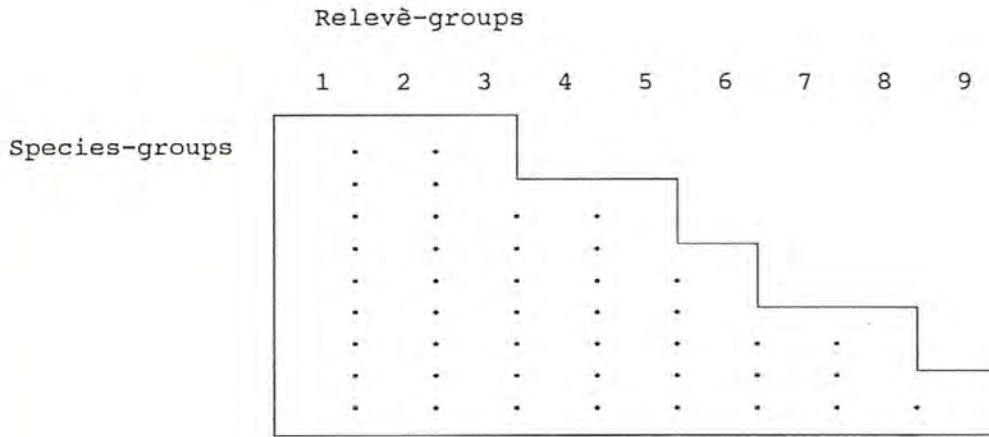


FIGURE 3.12. - Species-groups representing two or more communities, sequenced to correspond to an environmental gradient of which only the upper or lower limits for differentiated communities can be ascertained. This also applies to the horizontal mirror-image of the illustrated pattern.



sequence is based on minimum total separation units.

A community diagnostic species pattern with many blanks and irregular spacing of occurrences (Figure 3.10d) can be attributed to any of the following:

- i. a poorly defined community;
- ii. a gradual environmental change over the community limits;
- iii. heterogeneity caused by widespread disturbance over the community; or,
- iv. inadequate sample size.

Mapping precision of such a community would be low.

3.5.2.1 C) *Gradients*

The matrix can be further simplified by grouping overlapping species-groups, representing two or more relevè-groups, to correspond with environmental gradients. Two main patterns can emerge, namely, a) a gradient with both upper and lower environmental limits (Figure 3.11), or b) a gradient with either upper or lower environmental limits only (Figure 3.12). Both these types can be mirror-images of those illustrated in Figures 3.11 & 3.12. The limits should be evident after correlation of communities with habitat (section 3.4.2.3).

The simplification of the species-groups, in this way, will not affect the total separation units or the relationships between relevès or relevè-groups, so that information is not lost from the matrix but rather additional pattern becomes evident. The matrix

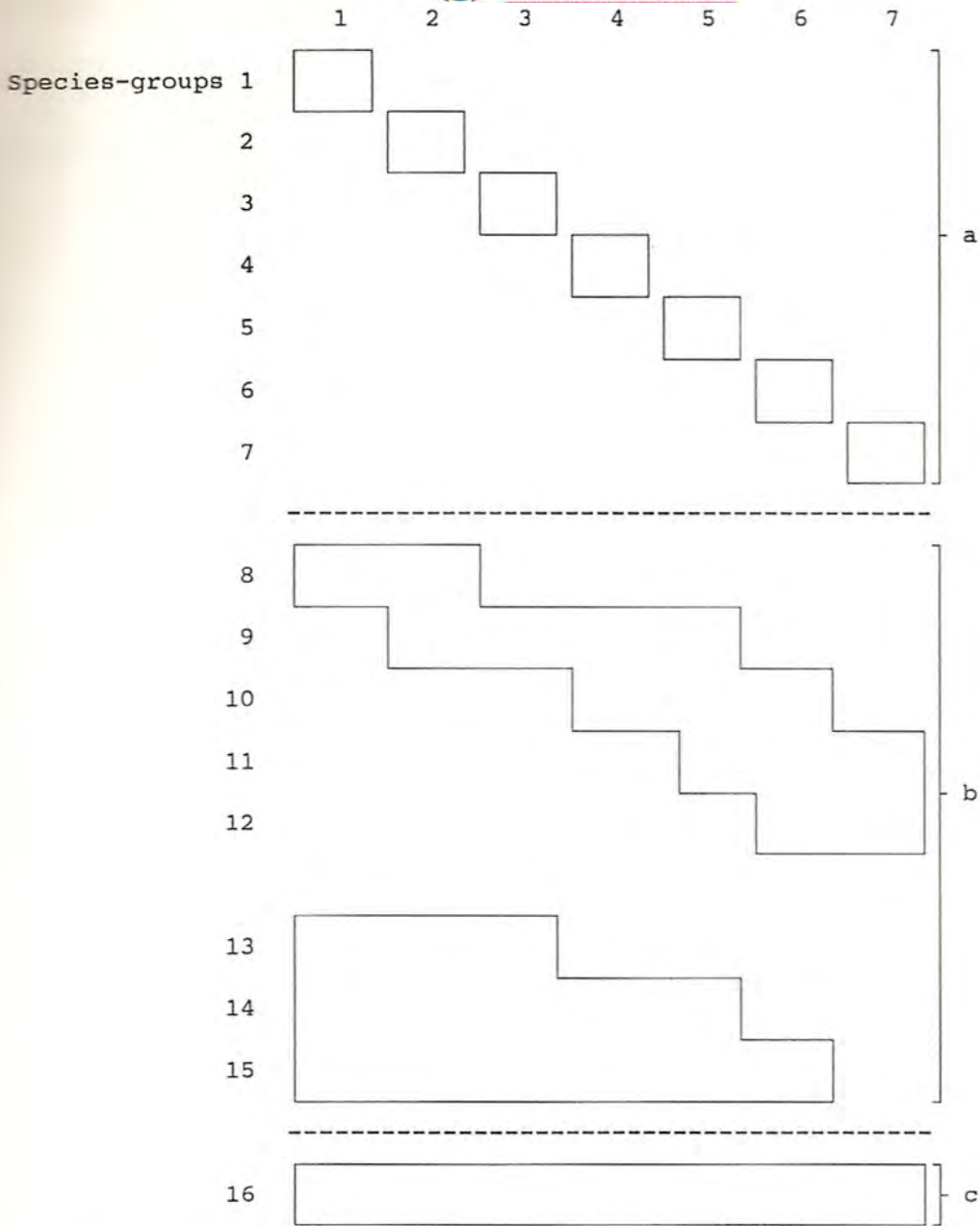


FIGURE 3.13. - Schematic illustration of the simplification of a phytosociological table into 3 sections: a) community diagnostic species-groups; b) species-groups corresponding to environmental gradients. Horizontal mirror-images of the illustrated pattern also apply; and c) non-diagnostic species.



will now consist of three parts, (Figure 3.13) namely, a) community diagnostic species-groups; b) species-groups indicating environmental gradients and c) non-diagnostic species.

3.5.2.2 *Secondary derivatives*

The secondary derivatives are those results computed from the recorded data but are dependent on the classification for grouping.

3.5.2.2 A) *Structure*

The general correspondence between the recorded growth forms and height (section 3.2.4), permit layer diagrams of vegetation structure, according to total recorded canopy cover for all species within each growth form class (after Ito 1979) for each community, to be made. This also allows differentiation of the ground layer into grasses and forbs which is not possible where structure is recorded as total cover per height class.

3.5.2.2 B) *Community composition analysis*

A plant species which utilizes a large resource-space per individual plant is likely to:

- i. have a correspondingly high crown cover;
- ii. have a high resource-space requirement, in the mature phase;
- and,
- iii. have a high cover in relation to its frequency.



The converse could also be valid, namely, that a plant species with a low cover in relation to its frequency of occurrence, requires a correspondingly small resource-space per individual plant in the mature phase. It is obvious that in comparisons of cover-to-frequency ratios, grasses for example, cannot be compared with trees as the disparity in resource-space requirements is too great in the mature phase. These statements are generalizations as individual plants can differ considerably in their reactions to external factors. However, the elucidation of intra-community dynamics in this study is dependent on species reactions, not that of individual plants.

Plant species within a defined plant community are, therefore, categorized according to the growth form classes defined in section 3.2.4. Cover-to-frequency ratios for each species, are then calculated and a linear regression for the cover and frequencies in each growth form class is determined. A linear relation between cover and frequency is assumed. The expected cover for the actual frequency of each species, according to the regressions is also calculated. The species are then arranged in decreasing order of the differences between actual and predicted cover.

Species outside the standard error of the mean, for the cover-to-frequency regressions, form two distinct groups, namely, those with a higher cover, and those with a lower cover than the standard error of the mean. The first group is referred to as strong competitor species because of their resource-space



requirements and those in the second group are referred to as weak competitor species. Generally most species fall within a range between these two groups forming a third group which can be referred to as the normal competition range. Thus, species within a community can be categorized into 15 different classes (i.e. three competitor groups in each of five growth form classes) which relate to the species present in the community and the effect of their differing resource-space requirements.

An extension of these programmatic calculations is the determination of cover proportions, of the growth form classes, within a community and the total cover within each growth form class.

3.5.2.2 C) *Stand phase analysis*

After the community composition analysis, it is then possible to compare the cover of any stand of vegetation within the community, at the scale at which the community was sampled, with that obtained for the community with the community composition analysis. For practical purposes the community frequency can be used and only cover differences taken into account. Thus the differences in cover between the stand and the community are emphasized.

Five distinct phases, in the comparison between stand and community, can be identified, namely:

Phase 0: all the species fall into the normal competition range.

i.e. there are no strong or weak competitor classes and,



therefore, no class is advantaged or disadvantaged;

Phase 1: weak competitors occur in the grass growth form class,
indicating possible selective utilization;

Phase 2: strong competitors occur in the grass growth form class,
indicating excessive grass utilization;

Phase 3: weak competitors occur in the woody growth form classes.

The class could be dependent on the biome or vegetation formation. In the case of savanna a progression through the classes from dwarf shrubs to trees is possible. Some of these species in the juvenile stage are potential strong competitors; and

Phase 4: Strong competitors occur in the woody growth form classes. The classes, as in Phase 3, could be biome or formation dependent. A tendency towards unidominance and/or moribundity could exist.

These main phases indicate the stage of vegetation development of the stand relative to the community of which the stand is a part. Dynamics is thus inferred, assuming progression through the stages. Reversal of the sequence is probably a rare occurrence in southern Africa.

3.5.2.2 D) *Community cover assessment*

The ideal vegetation composition for utilization is dependent on the type of utilization applied. For example, the vegetation composition ideally suited to cattle grazing is not necessarily the ideal for goats. It would appear far more efficient to adapt the type of animal utilization and degree of utilization to the



specific vegetation resource than to try to manage the resource to suit the form of utilization.

Vegetation condition can, therefore, be assessed from two perspectives, namely, a) suitability for a particular type of utilization and b) sustained production and resilience or ecological condition. In terms of the justification (section 1.3) ecological condition should primarily relate to the ability of the vegetation to protect the soil. This implies an adequate ground cover with increasing root penetration as slope increases, for soil binding. The following, accordingly, describes an elementary method for assessing tree cover adequacy.

Tree cover values, determined in the community composition analysis, and mean community slope are used as input. The relationship between tree cover and slope was taken after trial and error to be :

$$c = \tan s \times 100$$

where c = tree cover (%) and s = slope ($^{\circ}$).

This is merely a provisional indication of the minimum tree cover required for root soil binding which varies from 0% on level ground to 100% on a 45° slope. No assessment is made of the adequacy of ground cover.

3.6 PHYTOTAB-PC

PHYTOTAB-PC is a program package to facilitate the methods des-



cribed in this chapter and also facilitate handling of phytosociological data. The program package consists of 147 program and text files written in TURBO BASIC, a compiled version of BASIC, for an AT or 286 IBM-PC with 640 KB RAM, 20 MB hard disk and a 360 KB, 5¼" floppy disk drive, minimum configuration. The package is divided into two parts, both menu driven, namely,

- a) directory C:\PT for classification and processing, and
- b) sub-directory C:\PT\PD for a floristic data bank.

3.6.1 Classification and processing

The programs are executed from the C:\PT directory, or if ";C:\PT" is included in the path of the AUTOEXEC.BAT file, from any directory, by typing "PT". The main menu selection is then invoked. Menu items are executed by entering the symbol, which is bound by a single bracket, preceding the item. Application sequence is normally according to the numeric main menu items. Alphabetic main menu items are used for documentation and utility programs which can be executed when necessary. The functions and procedures for the program package are described, according to the menus, as follows:

D) Documentation

This details setup of directories, program transfer and program initiation as well as program usage by displaying or printing the on-line manual.

1) Matrix Data Input

This initiates keyboard input of the matrix in either PHYTOTAB format: relevè (species; cover-abundance) or TABIN format: species



relevè; cover-abundance). Provision is made for a new data set or continuation of input. Species are abbreviated to an eight digit or eight character code.

2) File Handling

This option can be used if PC-WRITE (Version 3.00) is resident, to view and edit sequential files, as well as to back up and restore files to diskette or, for file transfer.

3) Transform Matrix to QMAT

QMAT is the random access file used for further processing. This process must be completed before any further processing can be done. Outputs are: relevè sequence file, species sequence file, codename file for species and relevè computer code numbers, TABIN format file (if PHYTOTAB format is the input) and QMAT file. The last file is a random access file for speed of computation. Corrections should be done to either the PHYTOTAB or TABIN format files and transformed again. It should be noted that the relevè and species sequence files contain the computer allocated numbers for the species and relevès. Relevè numbers need not, therefore, be in sequence or range from 1 to n. However, relevè data should not be split. For example, relevè number 2 cannot occur in part before and after another relevè number.

4) Relevè and Species Sequencing

Eleven options are available in two submenus. On the first submenu, the first option is used for programmatically sequencing relevès. Matrix size limitation for relevè sequencing, using this option, is the product of species and relevès not greater than 186 000. This option can also utilize considerable processing time, in the order of hours to days, depending on matrix size,



noise (see section 3.3.1). and processor.

The second option, for sequencing species, is executed after the relevè sequence has been determined, either by the user or using the previous option. The species sequencing option is also used to group relevès where required (see sections 3.3.2 & 3.3.3). The third option sequences the species abbreviations alphabetically and the relevè numbers numerically and the codename file can be updated accordingly to facilitate abbreviation checking or to match full species names when downloading. The fourth option is for inserting or deleting numbers in a sequence. Separation characters for species and relevè-group delimiters can be inserted using this option or an editor can be used. Deleting a number decreases the value of subsequent numbers by one. It is important to remember that the numbers on the relevè file are computer codes which must be converted to the original numbers, processed and then converted back. The fifth option invokes the second submenu.

On the second submenu the first four options permit keyboard entry of a relevè or species sequence, corrections to an existing sequence (or an editor can be used), and generation of a number sequence from 1 to n; reversal of an existing sequence (for a mirror-image sequence), which is not affected by the presence of delimiters; and a random sequence from 1 to n, without replicates. The fifth and sixth options are sequence number to sequence code conversions and vice versa. These options must be used if any of the previous four options are executed. The final option is a permutation of relevès to obtain the optimum relevè sequence, exclud-



ing mirror images, based on minimum total separation units. This option should not be used for more than about 20 relevès, as it is extremely time-consuming.

5) Peripheral data input

This option invokes a submenu for input of species names, community names, table title, habitat data descriptions and habitat data classified codes. Other data which can be input are growth form codes for community composition analysis and check lists; cover codes and midpoint values used in the matrix and the actual values of environmental variables which can be either descriptive or numerical.

6) Print files

Options 1 to 7 are used to print files that have been created, for checking or reference purposes. Option 8 prepares an alphabetic sequence of species names with, optionally, growth forms, species frequency and separation units per species, in PC-WRITE format for annotating as a checklist. This option can also be used if an alphabetic species listing is required.

7) Print Working Table

This option permits printing of a matrix before habitat data and full species names have been added. The output includes total occurrences and separation units for each species; total species for each relevè; total separation units for all species; and percentage classification efficiency (see section 3.4.1).

8) Print Final Table

This option permits printing of a matrix with full species names and other peripheral data.



9) Process Data

This option invokes a submenu for the following:

1. Synoptic table production in which the matrix values can be according to a five class scale; a five class scale with single occurrences indicated with a "+" or the 33 class plant number scale (see section 3.5). All options available to the full matrix can be applied here.
2. Community composition analysis which includes community structure (see sections 3.5.2.1 & 3.5.2.2).
3. Stand phase analysis (see section 3.5.2.3).
4. Community cover assessment (see section 3.5.2.4)
5. Community and habitat correlation with relationships indicated by means of a dendrogram (see section 3.4.3).
6. Species cover relationships (see section 3.2.3.7).

U) Utility programs

The utility programs, being executable from the main menu, are referred to as internal utility programs. External utility programs do not form part of the PHYTOTAB-PC program package. The first three options with the invoked submenu are for correcting format errors that can occur after downloading matrix data or species names from a mainframe or data bank or to change the parameters for the species name file.

The fourth option is used to generate species sequences for plant identification keys (see section 3.2.2.8) and to split matrices for key or other purposes.

The fifth option converts the matrix data to the CONDENSED format



for the Cornell programs TWINSPAN (Hill 1979a) and DECORANA (Hill 1979b) or CANOCO (Ter Braak 1986, 1987). Should habitat data be required for the CANOCO programs these can be loaded, according to the necessary format, using an editor. Other format options available are those for Affinity Analysis (Scheiner & Istock 1987) and for the PHYTOTAB mainframe format (Westfall *et al.* 1982).

The sixth option is used for basic statistical applications on paired or unpaired variables. The applications include various means, range, standard deviation, variance as well as linear regressions. Optional output can be used as input for Harvard Graphics. Paired variables can also be sorted according to x or y values ascending.

The seventh option allows for fixed raster mapping according to grids, using the printer.

The eighth option is used to merge large sequential files, such as could be required for raster mapping.

The ninth option is used for inputting point co-ordinates for the ARC/INFO-PC geographic information system. This is required for mapping relevè localities.

The tenth option (C) converts cover codes to the plant number scale. This does not improve the precision of the original estimates.

The eleventh option (L) is used for generating and locating random



TABLE 3.7. - A model of the PHYTOTAB-PC program package, primary options with square brackets indicating keystroke sequences

Main output

Processes

A. Instructions

On-line manual

Documentation [D]

B. Classification

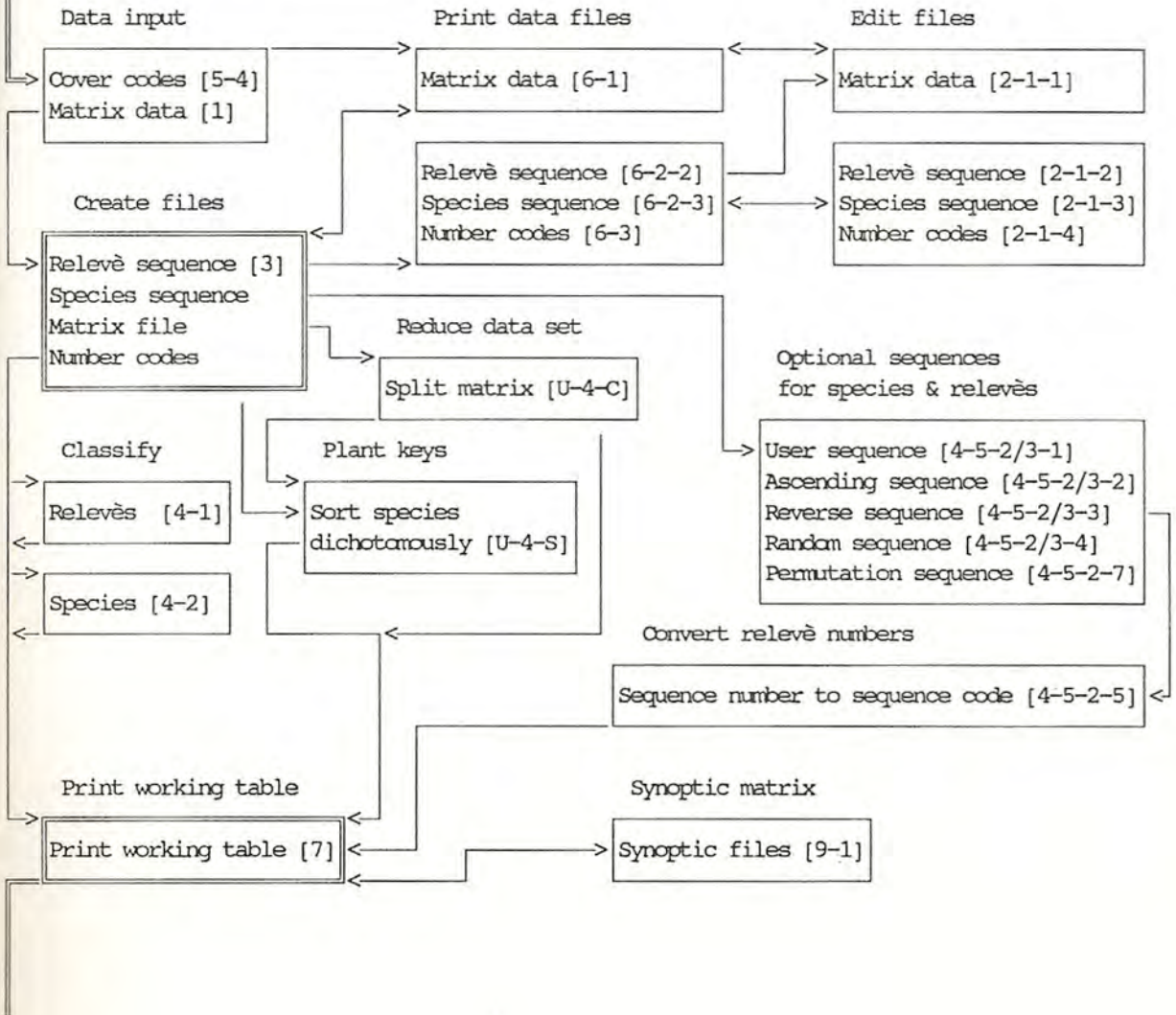




TABLE 3.7. (continued) - A model of the PHYTOTAB-PC program, package primary options with square brackets indicating keystroke sequences

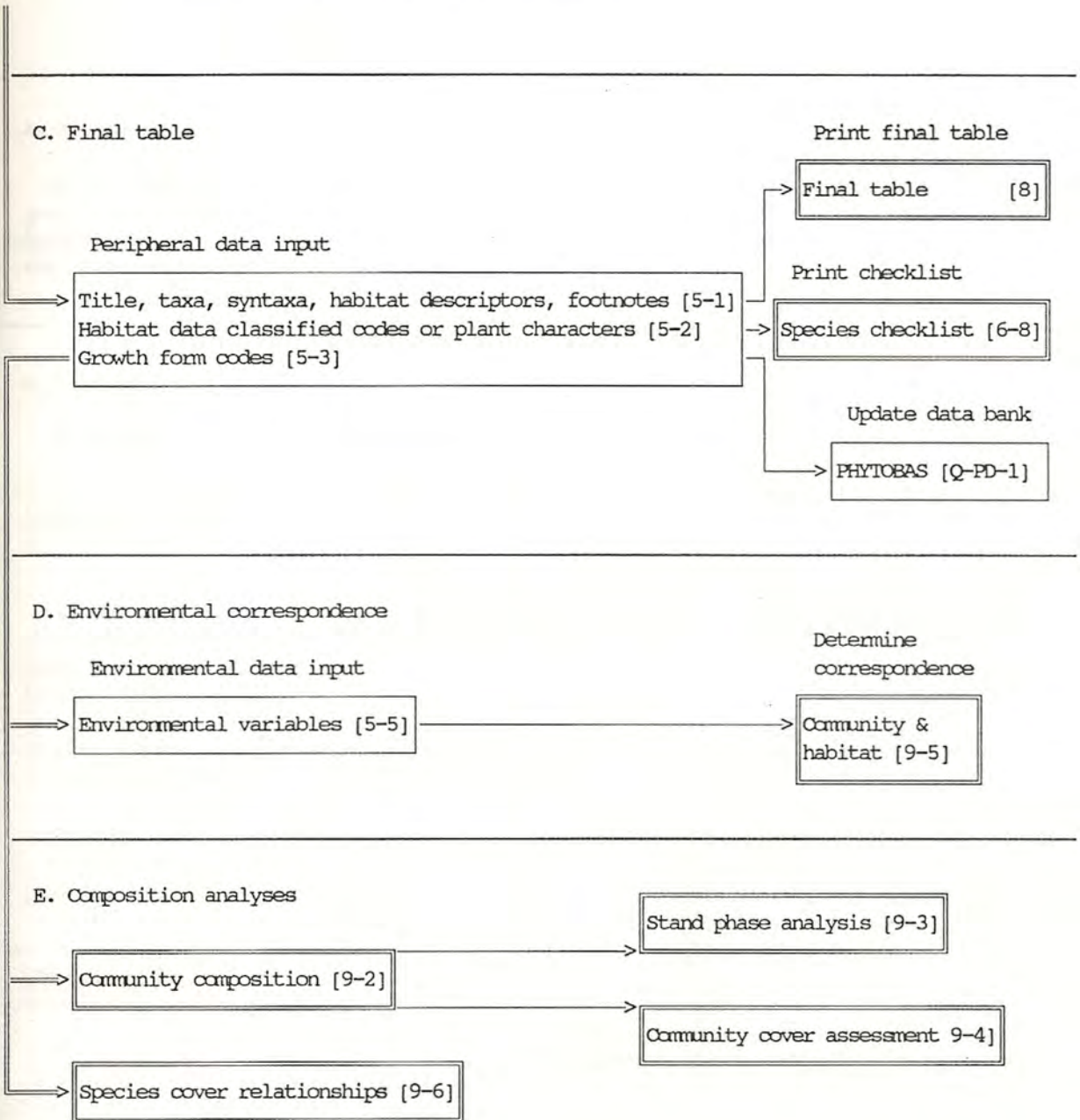




TABLE 3.7. (continued) - A model of the PHYTOTAB-PC program,
package primary options with square brackets
indicating keystroke sequences

F. Internal utility programs

Remove blanks from matrix file
(usually when transferring data) [U-1]

Convert PHYTOBAS SPPNAM to \PT\DESCRI
(for transfer from data bank) [U-2]

Redimension or remove names or blanks
from file \PT\DESCRI [U-3]

Convert to other formats (CANOCO,
AA, PHYTOTAB-mainframe) [U-5]

Statistics / Linear regressions [U-6]

Digital mapping [U-7]

Merge large sequential files [U-8]

Co-ordinate conversions or generate
random points [U-L]

Sample/stand dimensions [U-S]

Cover code conversions [U-C]

Number set comparisons [U-N]

G. External utility programs

3-D Ordination output -
Plotter is required [CAN3D]

Species comparisons with PRE data
set, search or add authors [SP]



points according to latitude, longitude co-ordinates and a 4 mm grid. Provision is also made for converting fractional degrees to or from decimal degrees and degrees, minutes and seconds.

The twelfth option (N) is for the comparison of two sets of grouped number sequences such as comparing two relevè sequences of the same data set.

The thirteenth option (S) is used for determining the following:

- i. stand dimensions from scale, stand area and stand radius;
- ii. sampling unit dimensions from species number and area; and,
- iii. species number estimation for a given area from species number per unit area.

Two function keys are operative for main menu items. The F1 key invokes the relevant section of the on-line manual and the F2 key invokes a problem solving guide.

All files generated with PT are saved to separate diskettes. This simplifies file maintenance with multiple users and if backup file copies are made both data integrity and security are ensured. File identification tags are printed automatically for each diskette. A flow chart of primary program options is shown in Table 3.7.

3.6.2 Data bank

This is a separate set of programs, on a subdirectory C:\PT\PD, functioning as a floristic data bank. The data bank programs are



executed from the C:\PT directory, or if ";C:\PT" is included in the path of the AUTOEXEC.BAT file, from any directory, by typing "PD". Four files are resident on a hard disk, namely, matrix data, species names, community names and an index. Data are loaded from diskettes generated with the classification programs. The main menu selections are the following:

1. Load data

This is from diskette in the PHYTOTAB-PC format, only after the final table has been printed and deemed satisfactory, if the programs are to be used for a data bank. The programs can also be used for analysing a single data set and splitting matrices.

2. List data

The following data can be listed: species names, both according to code or alphabetically; an index of data sets resident; relevè identifiers, being computer codes and original relevè numbers; species sequences being the order in which species should occur for a particular data set and community names, both numerically or alphabetically.

3. Search routines

Searches can be made starting with relevè identifiers, relevè numbers, data sets or multiple groups of relevè numbers. This permits subdivision and merging of data sets. The necessary files for classification are generated on diskette.

Searches can also be made starting from species codes, species names or truncated species names. The Boolean "and" logic applies when more than one name is entered for a search, i.e. all the species should be present in a relevè for the relevè to be select-



ed. Apart from creating diskettes for use with the classification programs, a list of species associated with selected species, on a relevè basis, can be obtained.

4. Correct random files

This applies to the species and community name files. The index and floristic data can be corrected with an editor.

5. Check species names

Species names on diskette, for table production, can be compared with those on the data base, for errors.

The data bank is limited by the capacity of the hard disk drive. Search and loading time is proportional to the amount of data present. For large data sets, an 80386 or 80486 processor is an obvious advantage.

3.6.3 External utility programs

Apart from an ASCII file editor such as PC-WRITE (version 3.00 or higher), the following programs can further facilitate vegetation analyses:

DECORANA: The CANOCO (Ter Braak 1986; 1987) version is preferred for ordination because of the options available and the inclusion of sample and species numbers in the scatter diagrams.

AA: Affinity analysis (Scheiner & Istock 1987) for analysing the variation of degrees of compositional relatedness among the communities of a data set.

CAN3D: This program is used for three-dimensional plotting of DECORANA output. A plotter is required.



SPECOM: (Westfall in prep.). This program is used for checking species names against the computerized National Herbarium synonym list and adding author names for checklists.

3.7 REFERENCES

- ACOCKS, J.P.H. 1975. Veld Types of South Africa. *Memoirs of the Botanical Survey of South Africa* 40: 1-128.
- BAKER, H.A. 1970. A key for the genus *Erica* using edge-punched cards. *Journal of South African Botany* 36: 151-156.
- BIANCHI, A.T.J. 1931. Een nieuwe determinatie-methode. *Tectona* 24: 884-893.
- BOLUS, H.M.L. 1958. Key to the genera of Mesembryanthemaceae. *Notes on Mesembryanthemum and allied Genera* 2: 392-403.
- BOUGHEY, A.S., BRIDGES, K.W. & IKEDA, A.G. 1968. An automated biological identification key. *Museum of Systematic Biology, University of California, Irvine, Research Series*, 2: i-xix, 1-35.
- BRAUN-BLANQUET, J. 1928. *Pflanzensoziologie*. 1 Aufl. Springer, Wien.
- COETZEE, B.J. 1975. A phytosociological classification of the Rustenburg Nature Reserve. *Bothalia* 11: 561-580.
- DEALL, G.B. & WESTFALL, R.H. 1989. Improving the resolution of floristic/habitat pattern correlations on phytosociological tables. *Bothalia* 19: 263-266.
- DUKE, J.A. 1969. On tropical tree seedlings. I. Seeds, seedling systems and systematics. *Annals of the Missouri Botanical Gardens* 56: 125-161.



- EDWARDS, D. 1983. A broad-scale structural classification of vegetation for practical purposes. *Bothalia* 14: 705-712.
- ELLIOTT, J.M. 1983. *Some methods for the statistical analysis of samples of benthic invertebrates*. Kendall, Wilson.
- F.S.S.A., 1974. Manual of soil analysis methods. *The Fertilizer Society of South Africa* 37: 1-65.
- GAUCH, H.G. 1982. *Multivariate analysis in community ecology*. Cambridge, New York.
- GIBBS RUSSELL, G.E., REID, C., VAN ROOY, J. & SMOOK, L. 1985. List of species of southern African plants. Edition 2. Recent literature and synonyms. Part 1. Cryptogams, Gymnosperms, Monocotyledons. *Memoirs of the Botanical Survey of South Africa* 51: 1-152.
- GIBBS RUSSELL, G.E., WELMAN, W.G., RETIEF, E., IMMELMAN, K.L., GERMISHUIZEN, G. PIENAAR, B.J., VAN WYK, M. & NICHOLAS, A. 1987. List of species of southern African plants. Edition 2. Recent literature and synonyms. Part 2. Dicotyledons. *Memoirs of the Botanical Survey of South Africa* 56: 1-270
- GLEN, H.F. 1974. *A taxonomic revision of the subtribe Gibbaeinae (Mesembryanthemaceae)*. M.Sc. thesis, University of Cape Town.
- GOODALL, D.W. 1968. Identification by computer. *BioScience* 18: 485-488.
- GYLLENBERG. H.G. 1965. A model for the computer identification of micro-organisms. *Journal of General Microbiology* 39: 401-405.
- HANSEN, B. & RAHN, K. 1969. Determination of Angiosperm families by means of a punched-card system. *Dansk botanisk Archiv* 26: 1-46 & 172 punched cards.



- HENKEL, J.S. 1934. *A field book of the woody plants of Natal and Zululand*. Robinson, Durban.
- HERRE, A.G.J. 1971. *Genera of Mesembryanthemaceae*. Tafelberg, Cape Town.
- HILL, M.O. 1979a. *TWINSpan - A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes*. Unpublished report. Cornell University, Ithaca, New York.
- HILL, M.O. 1979b. *DECORANA - A FORTRAN program for detrended correspondence analysis and reciprocal averaging*. Unpublished report. Cornell University, Ithaca, New York.
- ITO, K. 1979. A tentative study of stratification diagrams. In Miyawaki, A. & Okuda, S. (eds). *Vegetation und Landschaft Japans*. pp.357-368. The Yokohama Phytosociological Society, Yokohama.
- JANSEN, H. 1982. The geology of the Waterberg Basins in the Transvaal, Republic of South Africa. *Memoirs of the Geological Survey of South Africa* 71: 1-98.
- LAMARCK, J.B.A.P.M. de. 1778. *Flore Francaise* 3 vols. Imprimerie Royale, Paris.
- LAPAGE, S.P., BASCOMB, S. WILCOX, W.R. & CURTIS, M.A. 1970. Computer identification of bacteria. In A. Baillie & R.J. Gilbert (eds), *Automation, Mechanisation and Data Handling in Microbiology* 1-22. Academic Press, London.
- LAPAGE, S.P., BASCOMB, S. WILCOX, W.R. & CURTIS, M.A. 1973. Identification of bacteria by computer: general aspects and perspectives. *Journal of General Microbiology* 77: 273-290.
- LASSERRE, M., MALAN, O.G. & TURNER, B. 1983. *The application of*



principal component analysis to Landsat MSS data. Proceedings of seminar on Principal component analysis in the atmosphere and earth sciences, Pretoria 7-8 February 1983.

- LEENHOUTS, P.W. 1966. Keys in biology: a survey and a proposal of a new kind. *Proceedings of the Koninklijke Nederlandsche Akademie van Wetenschappen Amsterdam, Series C*. 69: 571-596.
- MACVICAR, C.N., DE VILLIERS, J.M., LOXTON, R.F., VERSTER, E., LAMBRECHTS, J.J.N., MERRYWEATHER, F.R., LE ROUX, J. VAN ROOYEN, T.H. & HARMSE, H.J.VON M. 1977. *Soil classification: A binomial system for South Africa*. Department of Agricultural Technical Services, Pretoria.
- MALAN, O.G. & LAMB, A.D. 1985. *Display of digital image data; quantitative and optimal*. Proceedings of the third South African symposium on Digital image processing, Durban 22-23 July 1985.
- MERXMUELLER, H. 1972. *Prodromus einer Flora von Suedwest Afrika*. Cramer, Lehre.
- METCALFE, C.R. & CHALK, L. 1979. *Anatomy of the Dicotyledons*. Vol. 1, 2nd edition. Oxford University Press, Oxford.
- MEYER, P.G. 1969. *Eintuehrung in die Pflanzenwelt Suedwestafrikas*. S.W.A. Wissenschaftliches Gesellschaft, Windhoek.
- MORSE, L.E. 1971. Specimen identification and key construction with time-sharing computers. *Taxon* 20: 269-282.
- MORSE, L.E. 1974. Computer programs for specimen identification, key construction and description printing using taxonomic data matrices. *Publications of the Museum, Michigan State University,, Biological Series* 5: 1-128.
- MORSE, L.E. 1975. Recent advances in the theory and practice of



- specimen identification. In R.J. Pankhurst (Ed.), *Biological identification with Computers*: 11-52. Academic Press, London.
- MUELLER-DOMBOIS, D. & ELLENBERG, H. 1974. *Aims and methods of vegetation ecology*. Wiley, New York.
- MUNSELL SOIL COLOR CHARTS, 1954. Munsell Color co., Baltimore.
- NEWELL, I.M. 1970. Construction and use of tabular keys. *Pacific Insects* 12: 25-37.
- OSBORNE, D.V. 1963. Some aspects of the theory of dichotomous keys. *New Phytologist* 62: 144-160.
- PANKHURST, R.J. 1971. Botanical keys generated by computer. *Watsonia* 8: 357-368.
- PANKHURST, R.J. 1978. *Biological identification*. Edward Arnold, London.
- PARTON, W.J. & INNIS, G.S. 1972. *Some graphs and their functional forms*. U.S. International Biological Program, Grassland Biome Project Technical Report 153: 1-41.
- RUTHERFORD, M.C. & WESTFALL, R.H. 1984. Sectors of the Transvaal Province of South Africa. *Bothalia* 15: 294-295.
- RUTHERFORD, M.C. & WESTFALL, R.H. 1986. Biomes of southern Africa - An objective categorization. *Memoirs of the Botanical Survey of South Africa* 54: 1-98.
- RYPKA, B.W., CLAPPER, W.E., BOWEN, I.G. & BABB, R. 1967. A model for the identification of bacteria. *Journal of General Microbiology* 46: 407-424.
- SCHEEPERS, J.C. (Unpubl.) *Geomorphology classes*. Grassland Research Centre, Pretoria.
- SCHEINER, S.M., & ISTOCK, C.A. 1987. Affinity analysis: methodologies and statistical inference. *Vegetatio* 72: 89-93.



- SNEATH, P.H.A. & SOKAL, R.R. 1973. *Numerical Taxonomy*. Freeman, San Francisco.
- TAINTON, N.M., BRANSBY, D.I. & BOOYSEN, P. de V. 1976. *Common Veld and Pasture Grasses of Natal*. Shuter & Shooter, Pietermaritzburg.
- TER BRAAK, C.J.F. 1986. Canonical correspondence analysis: A new Eigenvector technique for multivariate direct gradient analysis. *Ecology* 67: 1167-1179.
- TER BRAAK, C.J.F. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69: 69-77.
- TROLLOPE, W.S.W. & POTGIETER, A.L.F. 1983. *Characteristics of fire behaviour in the Kruger National Park*. Report. Department of Agronomy, University of Fort Hare, Alice.
- VOSS, L.G. 1952. The history of keys and phylogenetic trees in biology. *Journal of the Science Laboratory, Denison University* 43: 1-25.
- WEBER, W.A. & NELSON, P.P. 1972. *Random access key to the genera of Colorado mosses*. University of Colorado Museum, Boulder.
- WERGER, M.J.A. 1972. Species-area relationships and plot size: with some examples from South African vegetation. *Bothalia* 10: 583-594.
- WERGER, M.J.A. 1974. On concepts and techniques applied in the Zürich-Montpellier method of vegetation survey. *Bothalia* 11: 309-323.
- WESTFALL, R.H. 1981. *The plant ecology of the farm Groothoek, Thabazimbi District*. M.Sc. thesis, University of Pretoria.



- WESTFALL, R.H. (in prep.). SPECOM: A program for interrogating and comparing the PRECIS synonym list with species name data sets.
- WESTFALL, R.H., DEDNAM, G., VAN ROOYEN, N., & THERON, G.K. 1982. PHYTOTAB - A program package for Braun-Blanquet tables. *Vegetatio* 49: 35-37.
- WESTFALL, R.H. & DE WET, B.C. 1988. New programs for preliminary sequencing of relevès and species in phytosociological data sets. *Bothalia* 18: 122-123.
- WESTFALL, R.H. & MALAN, O.G. 1984. A comparison of vegetation units derived from vegetation-enhanced satellite imagery with the vegetation units derived from the floristic classification of the farm Groothoek, Thabazimbi District. Proceedings of the symposium on Pattern recognition in remote sensing and geophysics, Johannesburg, 2-3 July 1984.
- WESTFALL, R.H. & MALAN, O.G. 1986. A method for vegetation stratification using scale-related, vegetation-enhanced satellite imagery. *Bothalia* 16: 263-268.
- WESTFALL, R.H. & PANAGOS, M.D. 1984. A cover meter for canopy and basal cover estimations. *Bothalia* 15: 241-244.
- WESTFALL, R.H., VAN STADEN, J.M. & PANAGOS, M.D. 1987. Predictive species-area relations and determination of subsample size for vegetation sampling in the Transvaal Waterberg. *South African Journal of Botany* 53: 241-244.