COENOSES 9(2):75-79, 1994

A CONTINGENCY TABLE MODEL OF VEGETATION: THE MAPPING PROBLEM

A. R. Palmer[1] & L. Orlóci[2]

1 Roodeplaat Grassland Institute, PO Box 101, Grahamstown, 6140 South Africa
2 Department of Plant Sciences, University of Western Ontario, London, Ontario, Canada, N6A 5B7

Keywords: Canonical analysis, Distribution, Elevation, Gradient, Karoo, Plants, Populations, Rainfall, Topo-moisture.

Abstract. The contingency table model expresses vegetation on a specific site in terms of deviations of performance or sampling unit occupancy from an expected state. We define this state as global randomness that would prevail had the dispersion of the plant population not been constrained by trended environmental variation. With expectation so defined, the contingency table model is completely described by output from an enriched canonical contingency table analysis (CCTA). The output includes canonical correlations, canonical scores, and canonical partitions of deviations from expectations. Considering that the row categories in the contingency table in this case are species populations and the column categories are environmentally defined, the population deviations profiles of the sites and their partitions will be suitable as signatures to describe quantitatively the state of vegetation in a particular site at the time of the survey. With concatenation of these states consistent and anticipatory of environmental variation can provide clues of the direction and extent of vegetation change. The analysis has been described in detail elsewhere.

We discuss in this paper the problem of mapping new sites into the contingency table model without distortions of the original configuration. We propose CONAMAP, an application program, to derive the mappings. CONAMAP takes results from CCTA and derives canonical scores and deviations profiles for observational vectors from new sites. The species complement and population performance measures must be identical in the new vectors and in the contingency table. To illustrate the procedure we use floristic data from six topo-moisture classes of the eastern Cape midlands.

Introduction

The structure and floristic composition of the contemporary natural vegetation of the semi-arid regions of southern Africa has been explained as a consequence of the pastoral practices since the mid-nineteenth century (Tidmarsh 1948, Acocks 1953, Roux and Vorster 1983, Roux and Theron 1987). This explanation is based largely upon descriptive accounts of the early travelers in the region, isolated floristic records and fixed point monochromatic photographs. In order for the explanation to be adequately tested, two important weaknesses in the sampling approaches have to be considered. The first is that the climatic conditions preceding the collecting of the early data must be comparable with the conditions preceding the most recent sampling event. This is an essential pre-requisite considering the experiences with large variation in seasonal and long-term rainfall in the region. A reflection of this is the coefficient of variation in annual rainfall, an index of uncertainty of rainfall, which can be as high as 0.68 (Prince Albert Road, mean annual rainfall = 107mm, number of recording years = 109). Similarly, the coefficient of variation in monthly rainfall is also high, usually exceeding 1.00 in the drier months. The adequacy of replication (sensu Hurlbert 1987) of samples and the regional representativeness of sampling sites is also a problem. It is safe to say that neither the standardization of climatic conditions or the adequacy of replication have been addressed by the protagonists of vegetation change to the satisfaction of a scientific approach.

Some authors (Cowling and Hoffman 1990, Palmer *et al.* 1990) suggest that the direction of change may in fact be towards a more grassy condition. But it seems unlikely that the issue will be satisfactorily resolved with the current, largely descriptive methodology. It is therefore our aim in this paper to provide a technique by which conditions can be overlaid in analytical mappings suitable to evaluate vegetation change in comparative terms. We are of course aware of the need for suitable data and suggests that historical data on vegetation structure and composition is not sufficient on its own to make statements about the status of the natural vegetation in southern Africa.

Historically, one of the major failings resides with data collection and another with the lack of a testable model of potential or "expected" vegetation. We accept the point that the model may be qualitative or quantitative, but believe that it needs to be formulated before any survey work is undertaken. The preparations require good environmental data which can be related to the contemporary vegetation on the sites in no ambiguous terms. GIS (Geographic Information System) technologies and resource models can lay foundations for vegetation/environment models and for predictions of resource compartments where defined vegetation types are to be expected (Palmer 1991, Palmer and Van Staden 1992). What we propose is linked with this in idea, but gives versatility in defining "expected" states and conformity with these in mapping terms.

In southern Africa, Acocks (1953) developed the first qualitative model, providing a value perspective on vegetation pattern. This map was produced from some 3000 samples, providing 70 veld types (roughly equivalent to the syntaxonomic rank of order). Sampling intensity in veld types varied (e.g., 9 samples in VT2 to 102 samples in VT23). Although the map is extremely useful for agricultural planning, the sample data are inadequate for confidently evaluating changes in the structure and distribution of the vegetation. With the effect of climatic and anthropogenic influence on vegetation receiving much attention (IGBP, 1991), a more objective perspective to vegetation description and distribution studies is needed.

Stuart-Hill and Hobson (1991) suggest that a multivariate approach to modelling vegetation condition at a site would provide predictions, and enable management to manipulate vegetation towards pre-determined goals. One inadequacy of many of the methods is that when they map any new sample into the existing (ordination) configuration they change the position of other samples as well (Stuart-Hill and Hobson 1991). Our mapping method retains the original framework and locates the new point in the original ordination diagram without any scrambling of the original configuration.

There have been numerous efforts at developing linear models of community response to environmental variables (see review in Orlóci 1978; 1991a). One of the underlying weaknesses is the inability to incorporate information in the form of control sites with known history and treatment. An unfortunate consequence has been that many research efforts have not been taken seriously (Hurlbert 1984). In the spirit of earlier conceptualizations, epitomized by Whittaker's work (1973), efforts have been made to undertake direct gradient analysis to examine vegetation trends when some of the environmental variation is removed by placing samples in topo-moisture classes (Allen et al. 1991). Interestingly, time and time again, these pointed up the problems of which non-linearity looms greatest and still not adequately resolved. Coping with non-linearity in natural systems and its implications in statistical work has been the subject of many publications (see Orlóci 1978, 1988, 1991a, 1993, Hill and Gauch 1980, Kenkel and Orlóci, 1987 and references therein). The final outcome of the ongoing debate is uncertain, but it is clear that some current approaches are not favourably received (Wartenberg *et al.* 1987, James and McCulloch 1990). The linearity assumption in canonical contingency table analysis (Orlóci 1991b) necessitates careful stratification prior to sampling. Stratification has been recommended by many ecologists (e.g., Southwood 1978) and has been employed in many studies, with emphasis placed on chance sampling after stratification.

Materials and methods

Stratification

This affects the samples and also entails the development of primary ordination model. Using total floristic data from Werger (1980), we stratified the 536 5m x 5m total floristic samples from the study on the basis of elevation, median annual rainfall and lithology into thirteen topo-moisture classes and two substrate classes. Eight replicates were selected at chance localities within each topomoisture class which also yielded sufficient samples from each substratum (Table 1).

CCTA

Following stratification a classification table of the frequency of each species in each class was constructed and an ordination was performed for each substratum in the six topo-moisture classes, using canonical contingency table analysis (CCTA) by program CONAPACK (Orlóci 1991b). The latter supplies canonical correlations and canonical scores (ordination coordinates), prepares partitions of deviations from random expectations (related to marginal totals), and draws deviation graphs for the column (topo-moisture class) or row (taxa) entities.

In CCTA contingency tables are explored using eigenanalysis of double adjusted data. We draw attention to similar approaches that are gaining popularity (e.g., Goodman 1986; Moussa and Ouda 1988; Moser 1989).

Mapping

The next step is to locate analytically the position of external sampling units in the CCTA ordination. To illustrate this, one of us (A.R.P.) extracted floristic data from past study (Palmer 1988), conducted 250 km to the south west of the study area of Werger (1980), and selected eight replicated samples in topomoisture class C (1200-1300m; 250-<300 mm) to test the fit of the sample in the Werger classes. For this, the sample was mapped using CONAMAP (Orlóci 1992) into the ordination model generated by CONAPACK.

In CONAMAP, the new vector (relevé of a new site) is mapped into each canonical (ordination) plane generated by CONAPACK. If the mappable object is the external relevé $\mathbf{f}_E$ presented as a t-valued row vector with a characteristic element f_{Ej} (performance of species population j in site E)

Table 1. Topo-moisture classes examined during the study.

Elevation (m)	Median annual rainfall (mm) <250	250-<300	300-<350	350-<400	400-<450
<1100	A				
110-0<1200	B				
1200-<1300		C	D	E	
1300-<1400					F

and total $f_{E.}$, the map of $\mathbf{f}_E$ is an m-valued vector of X canonical scores,

$X_E = (X_{E1} \ldots X_{Em})$

There are m+1 deviations profiles associated with X_E:

$$\Delta_E = \begin{bmatrix} \Delta_{E11} & \ldots & \Delta_{Et1} \\ . & \ldots & . \\ \Delta_{E1m} & \ldots & \Delta_{Etm} \end{bmatrix}$$

The column sums of this are symbolically given by

$\Delta_{.E1}, \ldots, \Delta_{.Et}$

The X scores are computed in CONAMAP based on the equation

$$X_{Ei} = \sum_{j=1}^{t} \frac{f_{Ej}\, Y_{ij}}{f_{E.}R_i}$$

The data required to perform this transformation are stored in files by the CONAPACK run. The external relevé file $\mathbf{f}_E$ has to be created. The input files needed by CONAMAP thus include:

a. m x t column canonical scores

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & \ldots & Y_{1t} \\ . & \ldots & . \\ Y_{m1} & \ldots & Y_{mt} \end{bmatrix}$$

(stored in file COLS of Conapack.) **Y** is based on the analysis of the original data **f** (from Werger in the example).

b. m Eigenvalues (file EIG), the square roots of which are the canonical correlations (R_1, ..., R_m) from the same analysis that produced **Y**.

c. t column totals $f_{.1}$, ..., $f_{.t}$ (file COLT) also from the same analysis that produced **Y**.

d. grand total ($f_{..}$ of Werger data in the example).

e. relevé $\mathbf{f}_E$ that describes the new site (from Palmer 1988 in the example), using the same species and the same measure of species quantity in the site as those in the original **f**.

Based on the X-scores, deviations profiles are computed in CONAMAP according to

$$\Delta_{Eij} = \frac{X_{Ei}\, Y_{ij}\, R_i\, f_{E.}\, f_{.j}}{f_{..}} \; ; \; i=1,..,m.$$

Similar procedure is followed when the external vector $\mathbf{f}_E$ is perceived as a column vector, except that the output is a set

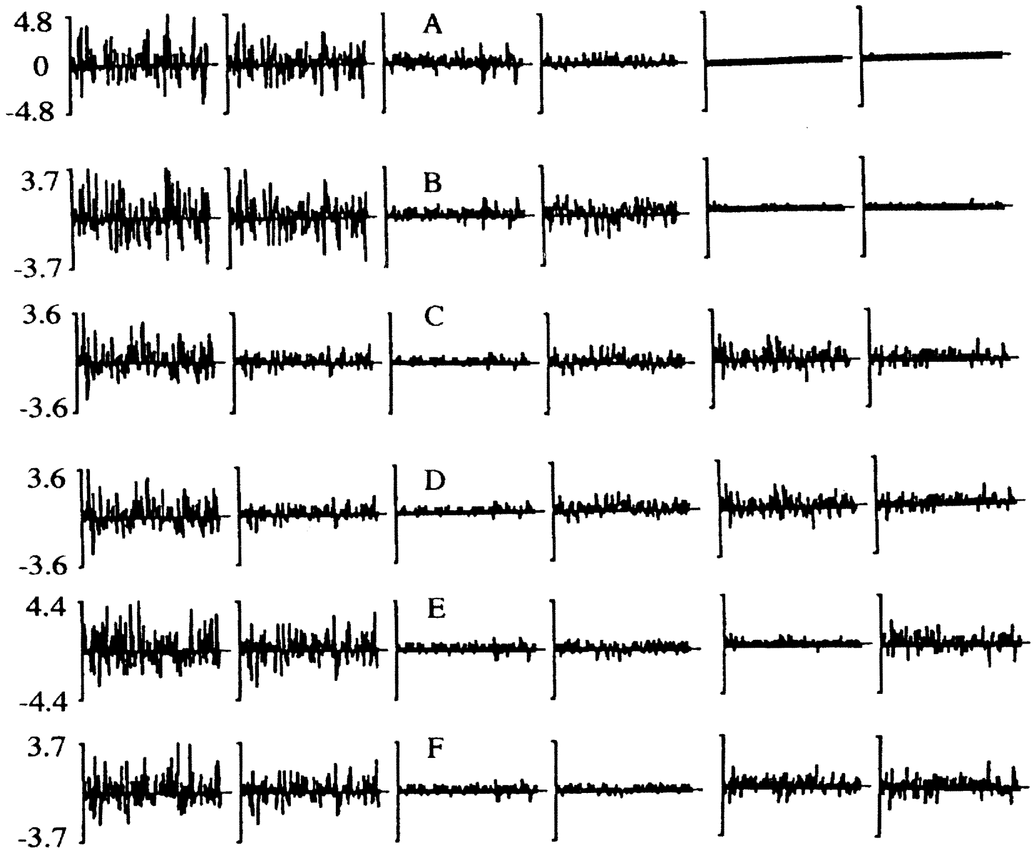

Figure 1. Deviations profiles showing 210 taxa (tick marks in horizontal axis) in Palmer's data (1988). In each topo-moisture class (A to F), the first diagram presents total deviations, with the second diagram giving the deviations in the first canonical plane, the third for the second canonical plane, and so on. The vertical axis measures the sense (positive up, negative down) and magnitude of deviations. The zero line (horizontal axis) represents the expected state (see explanations in the main text). Topo-moisture classes A to F are defined in Table 1.

of Y canonical scores

$\mathbf{Y}_E = (Y_{1E} \; ... \; Y_{mE})$,

deviations profiles

$$\Delta_E = \begin{bmatrix} \Delta_{11E} & \cdots & \Delta_{m1E} \\ . & \cdots & . \\ \Delta_{1qE} & \cdots & \Delta_{mqE} \end{bmatrix}$$

and rows sums

$$\Delta_E = \begin{bmatrix} \Delta_{1E} \\ . \\ \Delta_{qE} \end{bmatrix}$$

These $\mathbf{Y}_E$ and Δ_E are computed also in CONAMAP based on

$$Y_{iE} = \sum_{j=1}^{q} \frac{f_{jE} X_{ji}}{f_{.E} R_i}$$

and

$$\Delta_{Eij} = \frac{X_{ji} Y_{jE} R_i f_{j.} f_{.E}}{f_{..}} \; ; \; i=1,..,m.$$

which require:

a. $\mathbf{X}$ (CONAPACK run) and $\mathbf{Y}_E$ canonical scores (CONAMAP).

b. m canonical correlations (CONAPACK run).

c. the total of $\mathbf{f}_E$ (CONAMAP run) and q row totals $f_{1.}, ..., f_{q.}$ (CONAPACK run).

d. grand total ($f_{..}$) from CONAPACK run.

Step-by-step examples are given in Orlóci (1991b) and supplement (Orlóci 1992) available from L.O.

Results

CCTA supplied profiles of deviations (from expected) for each of 210 species in each topo-moisture class. The species set follows Werger (1980) and Palmer (1988) and are represented in the same order by tick marks in the diagrams given in Figure 1. The profiles are signatures 'as it were' and can be used in numerical comparisons. The ordination diagram (Figure 2) depicts the original data **f** (from Werger). Points A to F correspond to the model topo-moisture classes (Table 1). Point X is the mapping of an external relevé ($\mathbf{f}_E$ from Palmer) in the model. The mapping of this relevé is very close to topo-moisture class C, suggesting strong compositional affinity of the respective replicated samples in separate geographical regions from similar elevation, median annual rainfall and substratum.

Discussion

A refreshing perspective to vegetation modelling has been the use of expectations for populations based on community classification tables (Orlóci 1991b). The model in the present example come from published data (Werger 1980) and expectations are determined for a single external relevé from Palmer (1988) in communities along a topo-moisture gradient. Importantly the relevé is mapped into the model without modification of the model configuration. This is in contrast with other ordination procedures which necessitate reanalysis of the original data (Hill 1979, Ter Braak 1988).

A short-coming is that new taxa cannot readily be added after preparation of the model. If the new samples contain new taxa, the model must be constructed in a new CON-

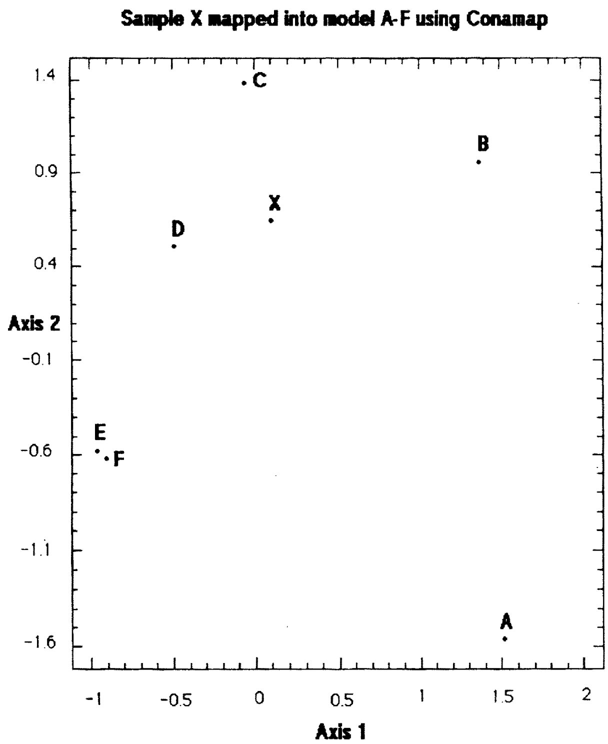

Figure 2. A plot of the first two ordination axes developed using CCTA (program CONAPACK), with a new relevé (X) mapped into the model. The new relevé incorporates frequencies from eight replicated quadrates in the same topo-moisture class as C, but from a study area 250 km to the southwest.

APACK run to incorporate the new taxa. This drawback notwithstanding, we feel that our mapping approach will provide vegetation scientists with a useful tool in comparative studies aimed to evaluate changes in vegetation composition.

Acknowledgments. The authors thank the Computing Centre for Water Research, Pietermaritzburg, for providing the median annual rainfall data; and the Department of Agricultural Development for supplying the Werger data in PHYTOTAB format. The research was funded by an overseas travel grant to ARP from the Roodeplaat Grassland Institute. An NSERC grant to L.O. supplied additional fund support.

References

Acocks, J.P.H. 1953. Veld types of South Africa. Mem. Bot. Surv. Sth. Afr. 28:1-192.

Allen, R.B., R.K. Peet and W.L. Baker. 1991. Gradient analysis of latitudinal variation in Southern Rocky Mountain forests. J. Biogeog. 18:123-139.

Goodman, L.A. 1986. Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. Int. Stat. Rev. 54: 243-270.

Hill, M.O. 1979. Decorana - A Fortran programme for detrended correspondence analysis and reciprocal averaging. Unpublished report, Ecology and Systematics, Cornell University, Ithaca.

Hoffman, M.T. and R.M. Cowling. 1990. Vegetation change in

the semi-arid, eastern Karoo over the last two hundred years: an expanding Karoo - fact or fiction? . S. Afr. J. Sci. 86: 286-294.

Hurlbert, S.H. 1984. Pseudo-replication and the design of ecological field experiments. Ecol. Monogr. 84:187-211.

James, F.C. and C.E. McCulloch. 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box ? Ann. Rev. Ecol. Syst. 21: 129-166.

Moser, E.B. 1989. Exploring contingency tables with correspondence analysis. Comp. App. Biosci. 5: 3-183.

Moussa, M.A.A. and B. A. Ouda. 1988. Correspondence analysis of contingency tables. Comp. Meth. Prog. biomed. 27: 2-11.

Orlóci, L. 1978. Multivariate analysis in vegetation research. W. Junk bv, The Hague.

Orlóci, L. 1988. Detecting vegetation pattern. ISI Atlas of Science 1: 173-177.

Orlóci, L. 1991a. Poorean approximation and Fisherian inference in bioenvironmetnal analysis. Research Trends, Advances in Ecology 1: 65-71.

Orlóci, L. 1991b. CONAPACK: Canonical analysis of classification tables. Ecological Computations Series. Vol. 4 SPB Academic Publishers, The Hague.

Orlóci, L. 1992. CONAMAP: Supplement to CONAPACK. Mimeographed.

Orlóci, L. and N.C. Kenkel. 1987. Data analysis in population and community ecology. Vol. 1. Static Systems. Pre-publication edition (Mimeographed).

Orlóci, L. 1993. The complexities and scenarios of ecosystem analysis. In: C.R. Rao (ed.), Multivariate Analysis: Future Directions, pp. 421-430. Elsevier, The Hague.

Palmer, A.R. 1988. Vegetation ecology of the Camdebo and Sneeuberg regions of the Karoo biome, South Africa. Ph.D. thesis. Rhodes University, Grahamstown.

Palmer, A.R., C. Hobson and M.T. Hoffman. 1990. Vegetation change in a semi-arid succulent dwarf shrubland in the eastern Cape, South Africa. S. Afr. J. Sci. 86:392-395.

Palmer, A.R. and J.M. van Staden. 1992. Predicting the distribution of plant communities using annual rainfall and elevation: an example from southern Africa. J. Veg. Sci. 3:

Southwood, T.R.E. 1978. Ecological Methods. Chapman and Hall, London.

Stuart-Hill, G.C. and F.O. Hobson. 1991. An alternative approach to veld condition assessment in the non-grassveld regions of South Africa. J. Grassl. Soc. Sth. Afr. 8: 179-185.

Ter Braak, C.J.F. 1988. Canoco - a FORTRAN program for canonical community ordination by (partial) (detrended) (canonical) correspondence analysis, principal component analysis and redundancy analysis (version 2.1). Technical report LWA-8802, Postbus 100, Wageningen.

Wartenberg, D., S. Ferson and F.J. Rohlf. 1987. Putting things in order - a critique of detrended correspondence-analysis. Am. Nat. 129: 434-448.

Werger, M.J.A. 1980. A phytosociological study of the Upper Orange River Valley. Mem. Bot. Surv. Sth. Afr. 46: 1-98.

Whittaker, R.H. 1973. Direct Gradient Analysis. In: R.H. Whittaker (ed.), Ordination and Classification of Communities, pp. 7-51. Junk bv, The Hague.

Manuscript received: April 1994