Vegetatio 75: 153–158 (1988)

Relating new information to a previous vegetation classification: a case of discriminant coordinates analysis

William B. Batista
Universidad de Buenos Aires, Facultad de Agronomia, Departmento de Ecologia. Av. San Martin 4453, 1417 Buenos Aires, Argentina

Accepted 8.4.1988

Keywords: Discriminant coordinate, Multiple discriminant analysis, Patagonia, Phytogeographic unit, Satellite data

Abstract

Discriminant coordinates analysis is an adequate technique for analyzing the linear relationships between a number of new variates (i.e. environmental or functional attributes) and a set of vegetational attributes already summarized in the form of a classification. It displays the principal differences among classes in relation to the new variates considered. The procedure and its rationale are equivalent to a special case of principal components analysis.

A case study on radiometer satellite data is presented. Two discriminant coordinates displayed the main differences in the seasonal dynamics of the NDVI (an index of standing green biomass) among broad phytogeographic units in the Patagonia region. The first coordinate can be interpreted as an index of height and convexity of the NDVI seasonal curve. It suggests that the principal difference among regions was the total seasonal growth. The second coordinate represents a contrast that discriminated between two already detected patterns of seasonal NDVI curve.

Abbreviations: DC – Discriminant Coordinate, NDVI – Normalized Difference Vegetation Index

Introduction: The problem

Current knowledge on the vegetation of a study area is usually based on a number of variates (i.e. species abundances) already analyzed and summarized. Summaries of vegetational survey data take the form or ordinations or classifications. A classification whose classes can be extended over large areas takes the form of a regionalization. New non-experimental studies may focus the attention on new variates (i.e. environmental or functional attributes) and on their relations with those previously studied.

The identification of interpretable relationships between species composition and additional (environmental) attributes was attempted by straight canonical correlation analysis (Austin 1968; Barkham & Norris 1970; Gauch & Wentworth 1976), where both vegetation and environment descriptions are treated as non-reduced sets of continuous variates. Later continuous approaches (Dargie 1984; Carleton 1984) replaced raw species data by sample scores over a few ordination axes in order to avoid difficulties derived from discreteness and non-linear relationships among those original variates and from a low samples: variates ratio.

The available data on potentially interesting new variates (i.e. climatic records, satellite scanning, soil surveys) often can only be matched with reduced vegetational data. In those cases a classification (regionalization) is frequently the available summary. A categorical correlation approach is then needed for the analysis of relationships. Besides, the above referred applications of canonical correlation made no formal use of previous knowledge for the analysis of the data set. Such a use would be required if results were expected to answer specific questions about the relationships between the new variates and already understood vegetation variation.

A categorical correlation approach for analyzing the linear relationships between m new variates and a set of vegetational attributes previously summarized in the form of a classification is provided by Discriminant Coordinates Analysis (Seber 1984). Within its framework the previous classification acts as a grouping criterion for m-variate observations. The data set is then regarded as a swarm of points in an m-dimensional space, each one labeled with the name of the class it belongs to. Unless the classes were derived from the same data set this technique provides answers to specific questions about features associated with already known vegetation differences.

A first question concerns whether the classes show differences for the new variates under study. If so, the leading differences among them would be those of interest. This approach stands for asking: 1°, How separate are the points belonging to the different classes in the m-dimensional space?, and 2°, Which are the main directions of their separation?. A third question would lead to inquire whether the classes occupy regions in the m-dimensional space at least roughly ordered along the principal axes of the swarm. If so, the differences among classes could be interpreted as the main source of variation for the newly studied attributes.

The Discriminant Coordinates, DC, (Gnanadesikan 1977; Seber 1984), or 'canonical' (*sensu* Jennrich 1977) functions used in Multiple Discriminant Analysis, answer the first two questions by displaying the principal differences among classes in relation to a given group of variates. They have already been used to analyze floristic differences among otherwise defined vegetation units (Norris & Barkham 1970; Grigal & Goldstein 1971; Goldstein & Grigal 1972; Matthews 1979; etc.), and to display vegetation-environment relationships (Mc Cune & Allen 1984; Tisdale & Bramble-Brodahl 1985; Gerdol *et al.* 1985). Their comparison with principal components or other suitable standard ordination axes should help to enlighten the third question.

Here we shall briefly discuss the rationale and solution of the problem of discriminant coordinates and present a case study on satellite measures of functional attributes of vegetation suitable to exemplify its use and interpretation.

The technique

The dimension reducing technique of DC, also called Multiple Discriminant Analysis (Pielou 1977; Lebart *et al.* 1984), includes three types of functions (Jennrich 1977; Seber 1984) used for classification problems. Another name, used in numerical taxonomy, is Canonical Variates (Blackith & Rayment 1971).

Given n m-variate observations $\mathbf{x}_{ij}$, each belonging to one of g groups, the aim of the discriminant coordinates technique is to replace them by lower (say q, where $q < m$ and $q \leq g$-1) dimensional vectors $\mathbf{z}_{ij}$ of independent coordinates z_{ijk} of the form

$$z_{ijk} = \mathbf{c}'_k \cdot \mathbf{x}_{ij} \quad , \quad k = 1, \ldots, q \tag{1}$$

in such a way that the separation among groups be maximum in the reduced space.

The rationale of the solution presented by Seber (1984) is to successively maximize the F statistics for testing the equality of means among groups of each z_{ijk}.

$$F_k = \text{constant} \cdot \frac{\mathbf{c}'_k \cdot \mathbf{B} \cdot \mathbf{c}_k}{\mathbf{c}'_k \cdot \mathbf{W} \cdot \mathbf{c}_k} \quad , k = 1, \ldots, q. \tag{2}$$

Where **W** and **B** are respectively the within-groups and the between-groups sums of squares and cross products matrices. The required maxima are attained when $\mathbf{c}_k$ are the first q eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$, being the corresponding eigenvalues proportional to the F_k statistic.

This procedure is equivalent to finding the principal components in the m-dimensional space with distance matrix $\mathbf{W}^{-1}$ for the g group means centered in the overall mean and loaded by the corresponding group sizes. This implicit choice of $\mathbf{W}^{-1}$ as the distance matrix for the observations in the variates space is only justified under a reasonable homogeneity of the group variances covariances matrices. When $g = 2$ the vector $\mathbf{c}$ gives the slope of the discriminant group classification function obtained by the likelihood ratio method.

The plots of the pairs $(z_{ijk}, z_{ijk'})$ permit visualizing the relative positions of the groups in the reduced space. Since the configuration obtained is deemed to be optimum in terms of discrimination among groups, wide overlaps are to be considered as a sign of no or small differences between the groups involved. The examination of the loadings for the original variates in each DC (elements of the vectors $\mathbf{c}_k$) gives the clue for the interpretation of the principal axes of among groups variation.

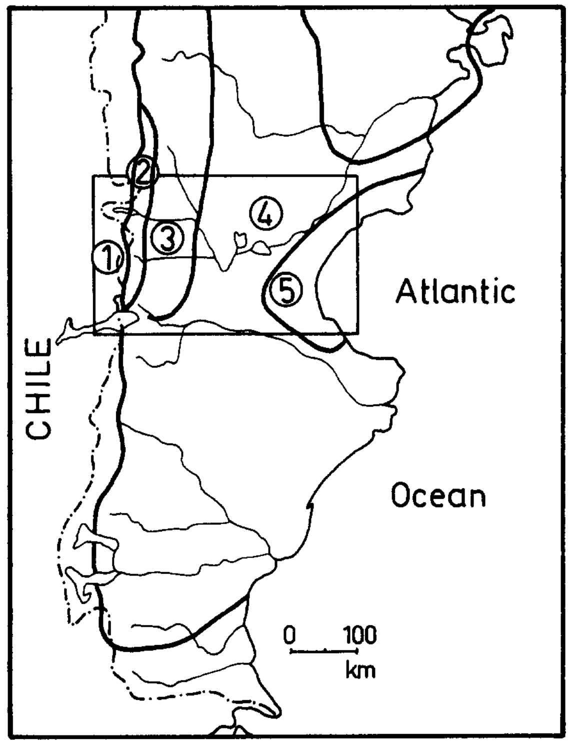

Fig. 1. Phytogeographic districts represented in the study area (rectangle). 1 Deciduous forest, 2 Subandean, 3 Occidental, 4 Central, 5 Gulf. Adapted from Soriano (1956).

Case study

Data set

Aguiar *et al.* (in press) reported the first analysis of a set of NOAA meteorological satellite data from the Patagonia region. Advanced very high resolution radiometer (AVHRR) measures were used to calculate an index of standing green biomass NDVI (Normalized Difference Vegetation Index, Curran 1981; Tucker *et al.* 1981; Tucker *et al.* 1985). The resulting input data matrix contains the NDVI values for 127 one sq. km sample points from a 230000 sq. km area on 6 dates spread over the growing season. The columns of this data matrix were regarded as 6-variate observations descriptive of the seasonal green biomass dynamics in each sample point. (For details on the sampling procedure, see Aguiar *et al.* in press).

Five phytogeographic (floristic) districts were represented in the study area (Fig. 1, Soriano 1956; Soriano 1983), from west to east: Subantarctic Forest, Subandean, Occidental, Central and San Jorge Gulf. Between the Subandean and the Occidental districts extends a conspicuous ecotone (León & Facelli 1981). Two units with different vegetation types can be distinguished within the San Jorge Gulf district: the *Colliguaya integerrima-Trevoa patagonica* scrub and the *Festuca pallescens-Poa ligularis* steppe (Soriano 1956; Bertiller *et al.* 1981). The predominantly wintertime rainfall of the Patagonia region is maximum in the Subantarctic Forest district, steeply decreases towards the Central district and has a slight increase again in the Gulf district.

The DC technique was applied to the data from the desert and semi-desert Subandean-Occidental ecotone, Occidental and Central districts, *Colliguaya integerrima* scrub and *Festuca pallescens* steppe. The phytogeographic units involved acted as grouping classes for the observations which were assigned to them according to the location of the corresponding sample points. After deleting the cases with missing values 84 cases remained in the data set. The results were expected to show if the phytogeographic units differ 'significantly' in their average seasonal NDVI dynamics, and eventually to display their leading differences.

Results

When the number of groups g is not greater than the dimension of the observations there can be at most g-1 DCs. The actual number is that of non zero eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$ which equals the dimension of the subspace spanned by the centered group means. The sum of the eigenvalues equals the total of squared $\mathbf{W}^{-1}$ distances between the group means and the overall mean each loaded by the corresponding group size, a homologue of the total between-groups variance for an oblique space. The eigenanalysis of $\mathbf{W}^{-1}\mathbf{B}$ for the present data set (Table 1) revealed the existence of a four-dimensional discriminant structure where the first two DCs accounted for 88.2% of that total variation. Consequently it was assumed that the between-class differences could be reasonably well approximated by the plane spanned by this two leading directions.

According to the loadings for the variates (eigenvectors **c**): the first DC can be described as a sum of the NDVI values in the four central dates, with special weight for the third and fifth, and slightly corrected by the first and last values (Table 1). Hence it can be roughly understood as an index of the height and convexity of the NDVI seasonal curve. The second DC can be regarded as a contrast between the sum of the second and the sixth NDVI values and the third one. Interpretation of such a contrast is less simple and leads to imagine various possible shapes of the NDVI seasonal curve. Let us consider just one case. Sample points which showed little change between the second and third dates and high NDVI at the end of the sampling period would get high values of this coordinate.

The plot of the first two DCs (Fig. 2) shows the Subandean-Occidental ecotone and the *Colliguaya integerrima* scrub well separated from the rest of the data. The Subandean-Occidental ecotone lay at the top of the first coordinate axis, this suggesting a higher average NDVI and/or a more convex seasonal curve. Average smoothed NDVI curves (Fig. 3)

Table 1. Results of the Discriminant coordinates analysis of the data set from Aguiar *et al.* (in press).

Discriminant coordinate	z_1	z_2	z_3	z_4	z_5	z_6
Eigenvalue	1.108	0.337	0.143	0.046	0.0	0.0
% total sq. distance explained	67.65	20.56	8.99	2.80	0.0	0.0
Eigenvector **c** (loadings)	−0.042	0.012				
	0.021	0.089				
	0.113	−0.120				
	0.035	0.014				
	0.104	0.029				
	−0.031	0.101				
Associated F statistic (4 & 79 d.f.)	21.88	6.66	2.88	0.91	0.0	0.0

adapted from Aguiar *et al.* (in press) confirm this description. The *Colliguaya integerrima* scrub showed the top scores for the second coordinate, except for one outlier. Its average smoothed curve was the only one that fitted the above posed description showing a small change between the second and third dates and high NDVI in the sixth (Fig. 3).

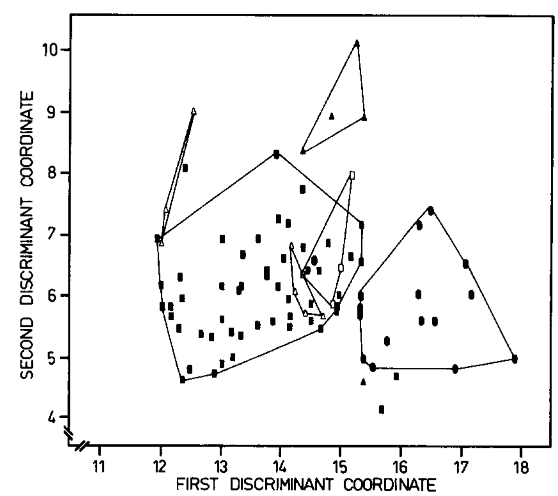

Fig. 2. Ordination plot of the sample points in the plane of the first two discriminant coordinates. ▮ Central district, ● Subandean-occidental ecotone, ▯ Occidental district, ▲ *Colliguaya integerrima* scrub △ *Festuca pallescens* steppe.

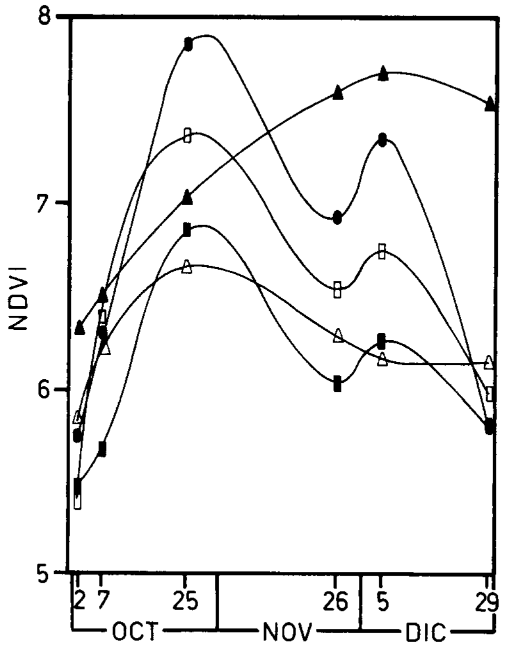

Fig. 3. Smoothed average NDVI change along the sampling period, October-December. ▮ Central district, ● Subandean-occidental ecotone. ▯ Occidental district, ▲ *Colliguaya integerrima* scrub. △ *Festuca pallescens* steppe. Adapted from Aguiar *et al.* (in press).

The *Festuca pallescens* steppe apparently shows two different NDVI curve models because it appeared split in two clusters in the plot (Fig. 2). According to this result the corresponding average smoothed curve (Fig. 3) would turn meaningless. The Central district shows lower scores on the first axis and similar scores on the second than the Subandean-Occidental ecotone (Fig. 2). Correspondingly, its average NDVI curve is lower but similarly shaped (Fig. 3). Two central district outliers with a high first DC were found to be near a dam. The Occidental district points lay in the plot merged with those of the Central district but clearly towards higher first DC scores (Fig. 2) while their average profile was intermediate between the Central and the Subandean-Occidental profiles (Fig. 3).

Discussion

The 'significant' differences among regions found in the sampling period average NDVI curve would be interpreted as an indication of differences in the seasonal standing green biomass dynamics (Curran 1981; Tucker *et al.* 1981; Tucker *et al.* 1985).

The first DC suggests that the principal difference among phytogeographic units is the amount of total seasonal growth. The special loadings for the third and the fifth dates values are related with the two peaks shown by the bulk of the data. Since the available field data on productivity in Patagonia concern only the Central (Bertiller 1984) and Occidental (Fernandez Alduncin 1986) districts, they can not be used to check this result although they do not disagree with it. The springtime occurrence of vegetation growth (Soriano *et al.* 1976; Soriano & Sala 1983) following the winter rainfall suggests that among regions differences in average total seasonal growth would be associated with vegetation adaptation to various starting levels of total water availability.

The second DC discriminated between two detected patterns of seasonal NDVI change (Aguiar *et al.* in press) suggesting that the difference between two shapes of seasonal green biomass curve is the second main variation among regions. The first pattern showed an explosive growth at the beginning of the sampling period, marked peaks and a rapid decay at the end. It corresponded to the Subandean-Occidental ecotone, and the Occidental and Central districts. The second pattern, fitted by the *Colliguaya integerrima* scrub, showed a slower but steadier increase of the NDVI with a delayed peak suggesting a longer growing season. It would indicate an effect of the Gulf district's marine climate over the vegetation.

Acknowledgements

The data set analyzed in this paper was obtained from M. R. Aguiar, J. M. Paruelo and R. A. Golluscio who kindly gave permission for its reanalysis as well as valuable suggestions for the interpretation of the results. I am grateful to Dra. G. Boente who introduced me to the algebra of the dimension reducing techniques, and to Dr. E. Ezcurra, Dr. W. B. Smith and an anonymous reviewer for useful comments on the manuscript.

References

Aguiar, M. R., Paruelo, J. M., Golluscio, R. A., León, R. J. C., Burkart, S. E. & Pujol, G. in press. The heterogeneity of the vegetation in arid and semiarid Patagonia: an analysis using AVHRR/NOAA satellite imagery. Annali Bot.

Austin, M. P. 1968. An ordination study of a chalk grassland community. J. Ecol. 56:739-757.

Barkham, J. P. & Norris, J. M. 1970. Multivariate procedures in an investigation of vegetation and soil relations of two beech woodlands, Cotswold Hills, England. Ecology 51:630-639.

Bertiller, M. B. 1984. Specific primary productivity in arid ecosystems: a case study in Patagonia, Argentina. Acta Oecol. Oecol. Gen. 5:365-381.

Bertiller, M. B., Beeskow, A. M. & Irrisarri, P. 1981. Caracteres fisonómicos de la vegetación del Chubut 1. Sierra de San Bernardo. Contribución N 40 CONICET Centro Nacional Patagónico.

Blackith, R. E. & Rayment, R. A. 1971. Multivariate morphometrics. Academic Press, London.

Carleton, T. J. 1984. Residual ordination analysis: a method for exploring vegetation-environment relationships. Ecology 65:469-477.

Curran, P. J. 1981. Multispectral remote sensing for estimating biomass and productivity. In: Smith M. (ed.), Plant and the day light spectrum. Academic Press, London.

Dargie, T. C. D. 1984. On the integrated interpretation of indirect site ordinations: a case study using semi-arid vegetation in south eastern Spain. Vegetatio 55:37-55.

Fernandez Alduncin R. 1986. Estimación de la productividad primaria neta area de pastos y arbustos en la estepa árida patagónica. Agr. Eng. Diss., Univ. de Buenos Aires.

Gerdol, R., Ferrari, C. & Piccoli, F. 1985. Correlation between soil and forest types: a study in multiple discriminant analysis. Vegetatio 60:49-56.

Gnanadesikan R. 1977. Methods for statistical data analysis of multivariate observations. Wiley, New York.

Gauch, H. G. & Wentworth, T. R. 1976. Canonical correlation as an ordination technique. Vegetatio 33:17-22.

Goldstein, R. A. & Grigal, D. F. 1972. Definition of vegetation structure by canonical analysis. J. Ecol. 60:277-284.

Grigal, D. F. & Goldstein, R. A. 1971. An integrated ordination classification analysis of intensively sampled oak-hickory forest. J. Ecol. 59:481-492.

Jennrich, R. I. 1977. Stepwise discriminant analysis. In: Enslein K., Ralston A. & Wilf H. S. (eds), Statistical methods for digital computers 3:76-95. Wiley, New York.

Lebart L., Morineau A. & Warwick K. M. 1984. Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices. Wiley, New York.

León, R. J. C. & Facelli, J. M. 1981. Descripción de una coenoclina en el SW del Chubut. Rev. Fac. Agron. 2:163-171.

Matthews, J. A. 1979. A study of the variability of some successional and climax plant assemblage-types using multiple discriminant analysis. J. Ecol. 67:255-271.

Mc Cune, B. & Allen T. F. H. 1984. Will similar forests develop in similar sites?. Can. J. Bot. 63:367-376.

Norris, J. M. & Barkham, J. P. 1970. A comparison of some cotswold beechwoods using multiple discriminant analysis. J. Ecol. 58:603-619.

Pielou E. C. 1977. Mathematical ecology. Wiley, New York.

Seber, G. A. F. 1984. Multivariate observations. Wiley, New York.

Soriano, A. 1956. Los distritos floristicos de la provincia Patagónica. Rev. Inv. Agr. X:323-347.

Soriano, A. 1983. Deserts and semideserts of Patagonia. In: West N. E. (ed.), Temperate deserts and semideserts, pp. 423-460. Elsevier, Amsterdam.

Soriano, A., Alippe, H. A., Sala O. E., Schlichter, T. M., Movia, C. P., León, R. J. C., Trabucco, R. & Deregibus V. A. 1976. Ecologia del pastizal de coirón amargo (*Stipa* spp.) del sudoeste de Chubut. Acad. Nac. Agron. Vet. 30(11):1-13.

Soriano A. & Sala O. E. 1983. Ecological strategies in a Patagonian arid steppe. Vegetatio 56:9-15.

Tisdale, E. W. & Bramble-Brodahl, M. 1983. Relationships of site characteristics to vegetation in canyon grasslands of west central Idaho and ajdacent areas. J. Range Man. 36:775-778.

Tucker, C. J., Holben, B. N., Elgin, J. H. & Mc Murtrey, J. 1981. Remote sensing of total dry matter accumulation in winter wheat. Remote Sens. Environ. 11:171.

Tucker, C. J., Townshend, J. R. G. & Goff, G. E. 1985. African land cover classification using satellite data. Science 277:369-375.