ORDINATION AS A TOOL FOR ANALYZING COMPLEX DATA SETS

Robert K. PEET*

Department of Botany, University of North Carolina, Chapel Hill, North Carolina 27514, USA

Keywords:
Forests, Gradient analysis, North Carolina, Ordination, Reciprocal averaging, Vegetation

Introduction

Ordination, or indirect gradient analysis, has been widely used in plant ecology as a tool for examining relationships between environment and vegetation. Since the classic work of Bray & Curtis (1957) which popularized such applications, the mathematical models underlying ordination methodology have become increasingly sophisticated. Principal Components Analysis provided increased conceptual rigor and objectivity. The realization that the assumptions of linearity implicit in PCA were not consistent with representation of nonlinear species responses along environmental gradients (Beals 1973, Jeglum, Wehrhahn & Swan 1971, Whittaker & Gauch 1973) led to introduction of a series of nonlinear methods including Reciprocal Averaging, RA, (Hill 1973, 1974), Gaussian Ordination, GO, (Gauch, Chase & Whittaker 1974, Ihm & Groenewoud 1975) and Nonmetric Multidimensional Scaling, NMDS, (Fasham 1977, Prentice 1977). Subsequent comparative studies (e.g., Fasham 1977, Gauch, Whittaker & Singer 1980) have suggested these nonlinear methods to consistently and accurately recover the structure of simulated data.

Simple, unidimensional coenoclines or compositional gradients appear to be readily recoverable using available nonlinear methods. Gauch et al. (1980) have also demonstrated reasonable success in recovery of second axes in simulated data, where the axes were roughly orthogonal and equal in length. What is rarely stated in such evaluations is that as the complexity of a data set increases, the efficiency of information recovery declines. This is in part a consequence of the underlying environmental factors not conforming to a consistent, geometric model. Multidimensional rectangular niche structures are theoretically attractive but rarely encountered. Rather, nonlinear interactions of factors are the rule; factors vary in importance with changes in other factors. The result is that beyond 2 or 3 dimensions, ordinations are rarely interpretable in terms of underlying environmental factors. In short, ordination can work well for identification of the most important and conspicuous environmental trends in a data set, but these would have been deduced by a competent field ecologist while collecting the data. Ordination has thus far proven much less effective at resolving environmental relations in complex data sets, precisely those data sets where much could be learned if effective ordination methods could be found.

As part of a study of forest succession on the North Carolina piedmont, I sampled, in collaboration with Dr. Norman Christensen (see Peet & Christensen 1979, 1980), 88 hardwood forest stands. One of our objectives was to characterize potential forest vegetation and to examine the environmental basis for its compositional variation. This characterization could then provide a framework for subsequent studies. Oosting's 1942 monograph had previously been the standard reference on the vegetation of this region, but it was based on only minimal sampling and recognized only three upland forest types: preclimax, climax and postclimax. The known edaphic and geologic complexity of the piedmont suggested to us that traditional gradient analysis methods wherein one or two master environmental complexes are recognized would be inappropriate. An ordination methodology was needed which would allow recognition of complex,

* The author gratefully acknowledges the continuing collaboration of Dr. Norman L. Christensen of Duke University. This research was supported by National Science Foundation grants DEB-7708743 and DEB-7804043 to R.K.P. and DEB-7707532 and DEB-7804041 to N.L.C.

multidimensional patterns which were not obvious at the time the data were collected.

The intent of the present contribution is to illustrate a possible solution to the difficulty of applying ordination methods to complex data sets. Although the approach was specifically developed for interpretation of North Carolina piedmont forests, it should be applicable to many complex data sets.

The basic methodology is one of progressive fragmentation, or removal of groups of stands. After initial ordination in 2 or 3 dimensions using a method such as RA or NMDS, the resulting pattern is examined for environmental correlations. Distinctive groups are removed (or a continuum is partitioned) and the remaining stands are reordinated, the impact of the initially observed factor having been greatly reduced. In effect, this resembles a subjective form of Hill et al.'s (1975) Character Species Analysis. The objective, however, is not simply to classify, but to understand the environmental relationships between groups of similar stands.

Methods

Samples were 0.1 ha quadrats, 20 × 50 m. Cover and frequency were recorded for all photosynthetic area below 1 m using a transect of 25 contiguous 0.5 × 2.0 m subquadrates located along the center of plot. All species occurring within the 0.1 ha were recorded and all trees (≧ 1.5 cm) were recorded by diameter at breast height. Importance values were calculated for understory (< 1 m) species as the average of relative cover and relative frequency. Species present in the 0.1 ha sample but absent from the subquadrates were assigned importance values of 0.01 (yielding a possible scale of 0.01 to 100). All 88 stands sampled were dominated by hardwoods, and had received little or no disturbance for over 50 years. Disturbance history was ascertained from records of the Duke Forest and Umstead State Park wherein all samples were collected. Five soil pits were dug in each plot and the upper 10 cm of the A_1 horizon were analyzed for Ca, Mg, K, PO_4, organic matter, pH, and available water capacity (see Peet & Christensen 1979 for further details).

RA was used for all ordinations. Species with less than four occurrences were consistently omitted from the data matrices. After removal of rare species the initial 88 stand matrix contained 202 species. The double relativization method of Bray & Curtis (1957) was consistently used. The tree data were not used in this analysis, the herbs (i.e., leafavea below 1 m) being considered more responsive to minor site differences and having fewer residual effects of past disturbances. Rather, the tree data were used in interpretation of the results by plotting dominant species on the ordinations.

Results

The first ordination revealed a pronounced gradient from alluvial bottomlands to dry, upland sites. The second axis was merely a polynomial version of the first, a result often encountered where the first axis is especially pronounced (Gauch, Whittaker & Wentworth 1976, Hill 1973). A plot of the first and third axes (Fig. 1) showed a distinct group of mesic, nutrient-rich coves. There is some suggestion of a third group characterized by the presence of *Quercus marilandica* and typically occurring on montmorillonitic clays, often derived from diabase dikes.

The distinctive alluvial (5) and cove forest (7) stands were removed from the data set leaving a heterogeneous set of mixed upland hardwoods. After removal of the rare species and relativization, these data were again ordinated (Fig. 2). Two distinct sets of stands emerged. *Quercus marilandica* stands over montmorillonitic clays dominate the lower right of the ordination while stands from north-facing bluffs with disjunct montane species such as *Kalmia latifolia*, *Rhododendron catawbiense* and *Galax aphylla* dominate the upper left. The remaining

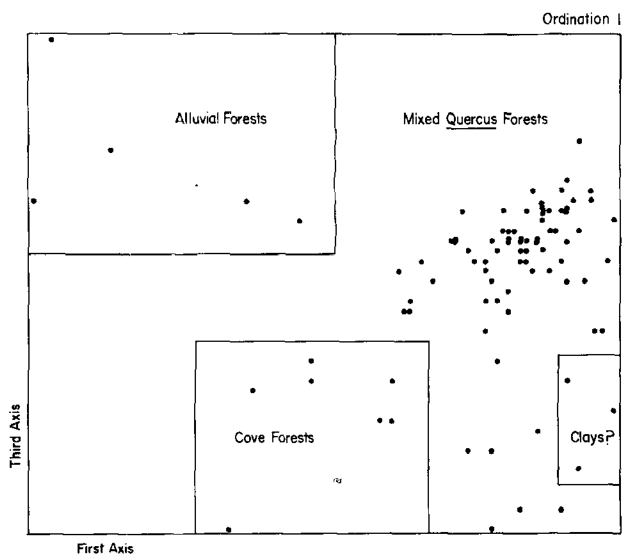

Fig. 1. First ordination of North Carolina piedmont hardwood stands showing segregation of alluvial and cove forest types.

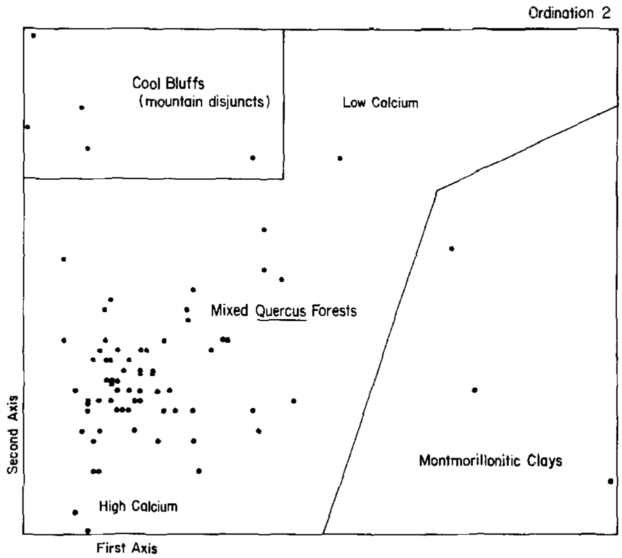

Fig. 2. Second ordination of North Carolina piedmont hardwood stands showing segregation of north-facing bluffs with montane elements, and of sites on montmorillonitic clays, usually with *Quercus marilandica* and *Q. stellata*.

stands again form a heterogeneous group of mixed hardwood stands, but with a suggestion of a gradient in soil pH and cations from base-rich soils in the lower left to more acidic, oligotrophic stands in the upper right.

A third ordination was performed after removal of the north-facing bluff (5) and *Quercus marilandica* (3) sites (not illustrated). The second axis of this ordination revealed a conspicuous group of stands dominated by *Quercus prinus* (3) characteristically occurring on isolated hills which rise up to 100 m above the surrounding landscape. These hills represent Jurassic intrusions of calcium poor volcanic rock. It is reassuring that the stands comprising the *Quercus prinus* group came from locations distant from each other, but similar in tree composition, topographic position and substrate.

The fourth ordination was performed after removal of the *Quercus prinus* stands (Fig. 3). Here pattern appears in what was previously a blurr of mixed hardwood stands. An apparent moisture gradient runs across the diagonal of the ordination from mesic stands dominated by *Fagus*, *Liriodendron* and *Acer rubrum* through mixed *Quercus* and *Carya* stands to xeric and ridge top stands dominated by *Quercus stellata*. Roughly orthogonal to this is a curving gradient in soil acidity and base content, the upper right portion of the gradient containing a group of stands dominated almost exclusively by *Quercus alba* and characterized by soils of extremely low pH ($\simeq$ 3.9) and exchangeable calcium content ($<$ 100 ppm).

A fifth ordination was performed after removal of the oligotrophic *Quercus alba* forests. As illustrated in Fig. 4, this ordination yielded two apparent complex-gradients. The first axis corresponded to a moisture gradient with

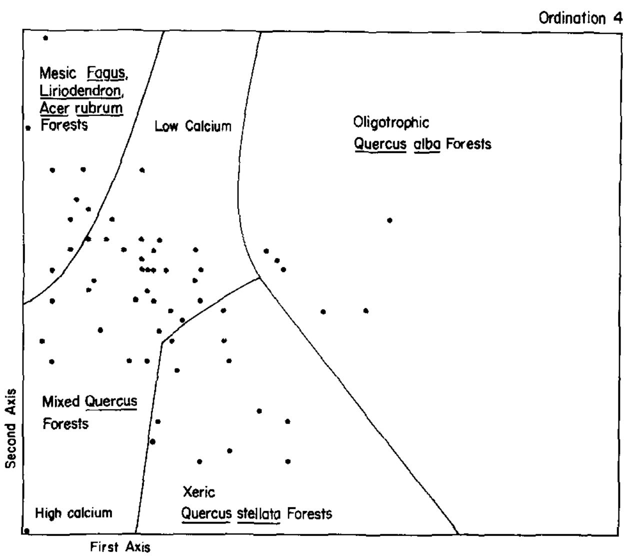

Fig. 3. Fourth ordination of North Carolina piedmont hardwood stands showing a soil pH and calcium gradient from lower left to upper right, and a moisture gradient from the upper left (mesic) to lower right (xeric).

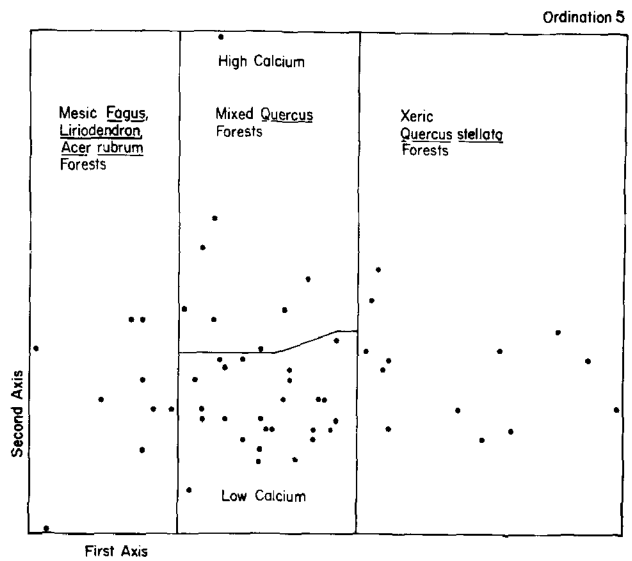

Fig. 4. Fifth ordination of North Carolina piedmont hardwood stands showing a moisture gradient from left (mesic) to right, and a soil calcium gradient from top to bottom.

mesic *Fagus*, *Liriodendron*, and *Acer* stands on the left and xeric *Quercus stellata* stands on the right. The second axis correlated strongly with soil calcium (r = 0.691), magnesium (r = 0.644) and pH (r = 0.515).

The mesic and xeric groups of stands were removed and the remaining set of 31 mixed *Quercus* and *Carya* stands was ordinated. In this sixth ordination (not illustrated) the first axis revealed the expected gradient in soil calcium content. The second axis revealed an axis in soil water storage capacity. This was paralleled by gradients in soil organic content and PO_4 suggesting low available water to correspond with low rates of decomposition.

Discussion and conclusions

The example presented illustrates the potential of a strategy of progressive data set fragmentation for determining environmental relations of vegetation when using ordination. While superficially the approach resembles classification, traditional classification methods do not allow ready examination of environmental relationships. If the stands were first classified into the final groups recognized and then ordinated, the environmental relationships would have again been obscured by the correlations, interactions and nonlinearities of the factors.

The approach suggested has an admitted and inescapable element of subjectivity (see Hill et al. 1975 for a more objective analogue), though careful use of correlation coefficients might reduce the severity of the limitation. It must be remembered that no classification scheme can be considered a priori correct; subjectivity will enter into all useful ecological classification schemes, Dale (1980) not withstanding (consistency can and should still be sought). Once the subjective nature of the classification process is recognized and accepted, it should be clear that progressive, subjective removal of stands should not greatly modify the perceived environmental relationships of the stands retained.

One of the major objectives of vegetation science is to understand the environmental patterns underlying vegetation composition. Ordination offers considerable potential as a tool for achieving this objective, if only the difficulties presented by complex data sets can be resolved. In this paper I suggest a simple but potentially useful strategy for coping with these difficulties.

Summary

Ordination has proven to be a useful tool for examining relationships between environment and vegetation in data sets with a simple underlying environmental structure. Complex data sets have proven much less tractable. A strategy is offered for dealing with complex data sets based on progressive removal of sets of stands along identified gradients, and subsequent reordination. This strategy is demonstrated using forests of the North Carolina piedmont.

References

Beals, E.W. 1973. Ordination: Mathematical elegance and ecological naivete. J. Ecol. 61: 23–25.

Bray, J.R. & J.T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27: 325–349.

Dale, M.B. 1980. Syntactic basis for classification. In: E. van der Maarel (ed.), Advances in vegetation science: Classification and Ordination. Vegetatio 42: 111–111.

Fasham, M.J.R. 1977. A comparison of nonmetric multidimensional scaling, principal components and reciprocal averaging for the ordination of simulated coenoclines, and coenoplanes. Ecology 58: 551–561.

Gauch, H.G., G.B. Chase & R.H. Whittaker. 1974. Ordination of vegetation samples by Gaussian species distributions. Ecology 55: 1382–1390.

Gauch, H.G., R.H. Whittaker & S.B. Singer. 1980. A comparative study of nonmetric ordinations. J. Ecol. (in press).

Hill, M.O. 1973. Reciprocal averaging: An eigenvector method of ordination. J. Royal Stat. Soc., Ser. C. 23: 340–354.

Hill, M.O., R.G.H. Bunce & M.W. Shaw. 1975. Indicator species analysis, a divisive polythetic method of classification, and its application to a survey of native pinewoods in Scotland. J. Ecol. 63: 597–614.

Ihm, P. & H. van Groenewoud. 1975. A multivariate ordering of vegetation data based on Gaussian-type gradient response curves. J. Ecol. 63: 767–777.

Jeglum, J.K., C.F. Wehrhahn & J.M.A. Swan. 1971. Comparisons of environmental ordinations with principal component ordinations for sets of data having different degrees of complexity. Can. J. For. Res. 1: 99–112.

Oosting, H.J. 1942. An ecological analysis of the plant communities of piedmont, North Carolina. Amer. Midl. Nat. 28: 1–126.

Peet, R.K. & N.L. Christensen. 1979. Hardwood forest vegetation of the North Carolina piedmont. Veröff. Geobot. Inst. ETH, Stiftung Rübel, 1979 (in press).

Peet, R.K. & N.L. Christensen. 1980. Succession: a population process. In E. van der Maarel (ed.), Advances in vegetation science: Succession. Vegetatio 43: ...–....

Prentice, I.C. 1977. Non-metric ordination methods in ecology. J. Ecol. 65: 85–94.

Whittaker, R.H. & H.G. Gauch. 1973. Evaluation of ordination techniques. In: R.H. Whittaker (ed.), Ordination and classification of plant communities, p. 287–321. Junk, The Hague.

Accepted 10 November 1979