Comparisons of Environmental Ordinations with Principal Component Vegetational Ordinations for Sets of Data Having Different Degrees of Complexity

J. K. Jeglum, C. F. Wehrhahn[1], and J. M. A. Swan[2]
Forest Research Laboratory, Canadian Forestry Service, Sault Ste. Marie, Ontario

Received February 2, 1971

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

Data from a survey of lowland, mainly peatland, vegetation were subjected to environmental ordination based on measurements of water level and water conductivity, and to vegetational ordination derived from principal component analysis (P.C.A.). Analyzed were the total set of the data ("all types"), half sets ("nonwoody" and "woody types") and quarter sets (stands of "marshes", "meadows", "shrub fens", and "other woody types"); the number of distinct physiognomic groups in a set of data, and presumably the amount of contained heterogeneity, decreased at each segmentation.

The effectiveness of the ordination models was tested by correlating measured distances in two-dimensional ordination models with $2W/(A + B)$ indices of vegetational similarity for randomly selected pairs of types or stands. As the physiognomic complexity decreased, the effectiveness of the P.C.A. vegetational ordination increased whereas that of the environmental ordination decreased. The environmental ordination seemed most appropriate to the data encompassing high complexity (total data set), while the P.C.A. vegetational ordination seemed most appropriate to data with low complexity (quarter sets of the data).

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

Les données obtenues lors d'une prospection d'un terrain bas, en grande partie tourbeux, ont été soumises à l'ordination. L'ordination des données de l'environnement portait sur les mesures du niveau phréatique et de la conductivité de l'eau, alors que celle des données de végétation était dérivée de l'analyse des composantes principales (ACP). Les analyses ont porté (1) sur l'ensemble des données, (2) sur les données réparties en deux groupes (types ligneux et non ligneux, (3) sur les données réparties en quatre groupes (stations de marais, de prairie, de fens arbustifs et d'autres types ligneux). Le nombre de groupes physionomiques distincts dans une série de données et, probablement, l'hétérogénéité intrinsèque, décroissait à chaque segmentation des données.

L'efficacité des modèles d'ordination a été testée en corrélant les distances mesurées dans des modèles d'ordination à deux dimensions avec les indices de similitude de végétation $2W/(A + B)$, pour des paires de types de végétation ou de stations tirées au hasard. A mesure que la complexité physionomique décroissait, l'efficacité de l'analyse des composantes principales augmentait pour l'ordination de la végétation mais décroissait pour l'ordination de l'environnement. L'ordination de l'environnement semble très appropriée les données comportant une grande complexité (ensemble des données), alors que l'ordination de végétation basée sur l'analyse de composantes principales convient mieux à des données de plus faible complexité (segmentation poussée).

Introduction

In recent years, principal component analysis (P.C.A.) has received increasing attention from ecologists as a technique for (1) reducing a mass of multivariate data to simpler patterns suitable for inspection and interpretation, and (2) formulating hypotheses for subsequent testing (Greig-Smith 1964). It has been suggested that vegetational ordinations, of which the currently most favored is derived from P.C.A., are not particularly useful when the data are too complex (*ibid.*), but few studies are available to assist the ecologist in judging whether or not a particular data set is "too complex" or "not too complex". It also has been suggested by Greig-Smith that an ordina-

---

[1] Institute of Animal Resource Ecology, University of British Columbia, Vancouver, British Columbia.

[2] Harvard Forest, Petersham, Massachusetts 01366, U.S.A.

tion based on vegetational characteristics (*cf.*, "indirect gradient analysis", Whittaker 1967) probably provides a better initial approach for interpreting vegetation–environment relationships in data from ecological survey than an ordination based on environmental features (*cf.*, "direct gradient analysis", *ibid.*), but that further comparisons of different ordinations for one set of data are needed before this is accepted as a guiding principle. This paper presents information that gives perspective to both of these tentative conclusions; environmental ordinations are compared with P.C.A. vegetational ordinations for a set of data, and then for progressively smaller segments of the same set of data to illustrate the effect on both ordination approaches of the amount of complexity contained in the data.

Methods

The Data

Available were data from an ecological survey by Jeglum (1968), consisting of 119 stands from the total range of lowland, mainly peatland, vegetation in a 30- × 60-km area around Candle lake, located in the southern boreal forest in Saskatchewan. Samples of density for trees (d.b.h. $\geqslant$ 10 cm), saplings (2.5 cm $\leqslant$ d.b.h. $\leqslant$ 9.9 cm) and seedlings (height $\geqslant$ 30 cm, d.b.h. $\leqslant$ 2.4 cm) were obtained with the linestrip method (Lindsey 1955); measurements of cover for shrubs, with the line intercept method; and frequency for herbs, in 20 to 30, 0.5- × 0.5-m quadrats. Individual species of moss, liverwort, and lichen were listed as present in a quadrat if the species covered 25% or more of the quadrat area, providing a measurement of dominance-frequency (*cf.* Loucks 1962). Unless otherwise noted, nomenclature of vascular species is according to Scoggan (1957) and nomenclature of bryophytes, according to Crum *et al.* (1965).

Environmental measurements for each stand included: the average depth to water level beneath the ground surface (taken during the summer drought of 1966 when no appreciable rains had fallen for at least 4 weeks); pH and conductivity from a water sample collected from surface water, small pools, or water filling a dug soil pit (taken at the same time as water level measurements); pH and von Post (1924) humification ratings for a peat sample from the upper 15 cm of the predominant peat horizon; depth of peat and, where the *Sphagnum* peat formed a continuous horizon, depth to the *Sphagnum*–fen peat junction; estimations of the Swedish drainage classes (*e.g.*, Sjörs 1948), in which limnogenous was assigned a value of 1, soligenous 2, topogenous 3, and ombrogenous 4; and frequencies of mud and sand bottoms among the stands placed in the dominance types.

The Original Analysis

The vegetational classification developed in the original analysis (Jeglum 1968) is presented in Table 1. The higher level units (A–I) are based on physiognomy, and hence are called "physiognomic groups". The lower level units (1–36), called "dominance types", are based on the leading dominant of the highest important stratum[3], except in those cases where it appeared that further subdivisions could be made on the basis of dominants or characteristic species of lower strata. In the latter cases, an oblique (/) separates the leading dominant of the uppermost stratum from the lower dominants or characteristic species.

The environmental ordination, presented in Fig. 1*a*, was constructed by plotting the vegetation units according to mean water levels and conductivities in a two-dimensional diagram. The choice of these single measurements was based on the assumption that they were good estimates of moisture and fertility regime, which from the literature and from our field experience appear to be the two most important factor-complexes in peatlands. The scale of conductivity was constructed to approximate that of a logarithmic scale as suggested by Whittaker and Fairbanks (1958). In the environmental ordination the axis for depth to water level was made slightly longer than that for conductivity because, in comparisons of vegetational similarities with $2W/(A + B)$ (see p. 102), the types at extreme ends of the moisture gradient had lower similarities than those at extreme ends of the nutrient gradient, although all values were very low. This suggests that greater vegetational variation occurs across the moisture regime than across the fertility regime. In subsequent environmental ordinations in this paper, the relative lengths of the water level to conductivity axes remain the same as in the original model in Fig. 1*a*.

To facilitate comparisons with the corresponding P.C.A. ordinations, we have indicated sequences I and II in the environmental ordination (Fig. 1*a*). These represent major zonal sequences observed in Candle lake wetlands (Jeglum 1968). Sequence I represents decreasing moistness in nutrient-rich wetlands; sequence II, decreasing moistness in nutrient-poor wetlands. These two sequences also can be interpreted as representing principal lines of vegetational variation.

Principal Component Analysis

P.C.A. is a mathematical technique for summarizing a set of related measurements as a set of derived variates, frequently fewer in number, which are definable as independent linear functions of the original measurements (Holland 1969). The derived variates are known as "principal components", and the first derived variate accounts for the largest possible proportion of the total variation, the second for the largest possible proportion of the remaining variance, and so on. The distribution of stands or species is plotted along the first few components to allow the

---

[3] The "highest important stratum" was judged subjectively as that stratum having a relatively continuous canopy and thereby giving the vegetation its principal visual impress.

TABLE 1. Vegetational categories

A. Marshes (26): Tall, perennial herbs occurring in zones or clumps adjacent to open water, with sedge-dominated zones landward, usually with standing water and small emergent, floating, and submergent species below the canopy.

- A-1. *Scolochloa festucacea*[a] (3)
- A-2. *Eleocharis palustris*[b] (4)
- A-3. Low sand beach (2)
- A-4. *Typha latifolia* (3)
- A-5. *Equisetum fluviatile* (1)
- A-6. *Sparganium eurycarpum* (3)
- A-7. *Scirpus acutus* (3)
- A-8. *Phragmites communis* (4)
- A-9. *Acorus calamus* (3)

B. Broad-leaved sedge meadow (19): Graminoid meadows dominated by sedges with leaves ⩾ 2 mm, less commonly by grasses with the broad-leaved sedges of secondary importance.

- B-10. *Carex rostrata* (4)
- B-11. *Calamagrostis inexpansa-neglecta* (1)
- B-12. *Carex aquatilis* (3)
- B-13. *Carex atherodes* (3)
- B-14. *Carex lacustris* (5)
- B-15. *Calamagrostis canadensis* (3)

C. Narrow-leaved sedge meadow (8): Sedge meadows dominated by sedges with leaves < 2 mm broad, the leading dominant in all stands being *Carex lasiocarpa*.

- C-16. *Carex lasiocarpa* (4)
- C-17. *Carex lasiocarpa* / small *Carex* (4). The small *Carex* include *C. chordorrhiza*, *C. limosa*, and *C. livida*.

D. Tall shrub fen (22): Shrubs with mature heights ≥ 135 cm (*cf.* LaRoi 1967), usually found as clumps of aggregated stems with spreading crowns, having > 25% shrub cover (*cf.* White 1965).

- D-18. *Salix maccalliana* (1)
- D-19. *Salix planifolia* (12)
- D-20. *Alnus rugosa* (4)
- D-21. *Salix bebbiana* (5)

E. Low shrub fen (12): Shrubs with mature heights < 135 cm (medium and low shrub categories of LaRoi 1967), having > 25% shrub cover.

- E-22. *Betula glandulifera*[c] (4)
- E-23. *Salix candida* (1)
- E-24. *Andromeda polifolia* (2)
- E-25. *Salix pedicellaris* (5)

F. Tamarack swamp (6): Stands dominated by tamarack (*Larix laricina*).

- F-26. *Larix laricina* / broad-leaved sedges (2)
- F-27. *Larix laricina* / narrow-leaved sedges (4)

G. Transitional muskeg (7): Stands usually have important levels for both fen and bog dominants, *i.e.*, one of the graminoid species characterizing the fen dominance types attains > 80% frequency, and *Sphagnum* spp. attains > 40% dominance-frequency.

- G-28. *Picea mariana-Larix laricina* / sedges / *Sphagnum* (3)
- G-29. *Salix planifolia* / broad-leaved sedges / *Sphagnum* (1)
- G-30. *Chamaedaphne calyculata* / narrow-leaved sedges / *Sphagnum* (3)

H. Treed muskeg (13): Stands with a tree stratum usually of *Picea mariana*, an ericaceous shrub (*Ledum* at Candle lake) dominating the shrub layer, and underlaid by a relatively continuous horizon of firmly consolidated *Sphagnum* peat, with *Sphagnum* spp. or various species of feather moss dominating the bottom layer (*cf.* Ritchie's (1956) definition for "muskeg").

- H-31. *Picea mariana* / *Ledum groenlandicum* / *Pleurozium schreberi* (7)
- H-32. *Picea mariana* / *Ledum groenlandicum* / *Sphagnum fuscum* (4).
- H-33. *Pinus banksiana* / *Ledum groenlandicum* / *Sphagnum fuscum* (2)

I. Moist forest (6): Forests united by their positioning on moist soils with shallow depths of peat accumulation (averaging 35 cm), usually adjacent to peatlands downslope.

- I-34. *Populus balsamifera* (3)
- I-35. *Picea glauca* (1)
- I-36. *Picea mariana* / *Pleurozium schreberi* (2)

NOTE: Parentheses ( ) enclose the number of stands in each category. An oblique (/) separates progressively lower strata where more than one stratum characterizes a type.

[a] Unless otherwise noted, nomenclature of vascular species is according to Scoggan (1957), that of moss species according to Crum *et al.* (1965).

[b] Includes *Eleocharis smallii* according to Gleason and Cronquist (1963).

[c] This is the lowland *Betula* species at Candle lake according to the range maps of Dugle (1966).

---

interpretation of relationships and formulation of hypotheses; one or more components are referred to as vegetational or species ordinations. It is emphasized that the principal components do not necessarily correspond to sensible biological concepts, and that they are used to generate hypotheses which must then be tested (*ibid.*; Pearce 1969).

P.C.A. as outlined by Thurstone (1947) and Anderson (1958) was employed with the R-type analysis and a variance–covariance matrix. The detailed calculation of P.C.A. will not be presented here because it has been adequately described elsewhere (*e.g.*, Pearce 1969). We employed the variance–covariance matrix on the assumption that P.C.A. would be sensitive to the scale of quantitative measurements used (*cf.* Seal 1964). Since the variances of individual species in the P.C.A. runs were all of similar magnitude, being around 2 to 5% of the total, the possibility is negligible that the component axes related to only a few species (*cf.* Austin 1968).

P.C.A. was applied to the total data set, which had already been subjected to the simple environmental ordination. The ordinations for the total data set were designated "all types". Then both P.C.A. and the simple environmental ordination were applied to approximate halves of the total data, designated "nonwoody types" and "woody types", to evaluate the effect of decreasing the number of physiognomic groups, and presumably the heterogeneity, contained in the data set. These type ordinations were based on means (based on the number of values summed) for the attribute measurements for stands within each dominance type. Finally, the two ordination approaches were applied to stands of "marshes" (A), "meadows" (B and C), "shrub fens" (D and E), and "other woody types" (F–I), each consisting of approximately one-quarter of the total data. These analyses were based on single species measurements from individual stands and hence are not strictly comparable to the "type" runs based on averages for stands within the dominance types. However, the complexity contained in the quarter-data segments, in terms of number of physiognomic groups (Table 1), was again less than that included in nonwoody types and woody types, which in turn was less than that contained in all types. Thus, we were able to apply the analyses of quarter sets of data to the question of influence of complexity of data on the suitability of the two ordination techniques.

The species used in the P.C.A. were the more quantitatively important. The all types P.C.A. was based on 42 species, including the following: trees—leading dominants; shrubs—leading dominants and species used to characterize a lower stratum of a type; herbs—leading dominants, species used to characterize a lower stratum, and those attaining 100% frequency in one stand; and mosses—species characterizing a bottom stratum of a type and those attaining 100% dominance-frequency in one stand. The woody types and nonwoody types P.C.A. were based on 56 and 48 species, respectively, including the following: trees—species exceeding 100 stems per acre in one of the 119 stands; shrubs—species exceeding 20% cover in one stand; herbs—species exceeding 80% frequency in one stand; and bryophytes and lichens—species exceeding 40% dominance-frequency in one stand. The P.C.A. was based on 26 species for marshes, 28 for meadows, 32 for shrub fens, and 35 for stands of other woody types. Criteria for choosing these species were the same as for the woody types and nonwoody types P.C.A.

If a value for a species exceeded 100, all values for that species were transformed to percentages of the maximum value, *e.g.*, for tree densities. No other transformations were made on the raw vegetational data. The raw data used here would not have changed significantly if each value had been expressed as a percentage of its maximum value (*cf.* Austin and Greig-Smith 1968).

In all P.C.A., five principal components were extracted. The program included subroutines that plotted stands or types in terms of their component values, and species in terms of their "adjusted" loadings, on various combinations of two and three axes. Component values for stands or types are the sum of the loadings for constituent species. "Adjusted" loadings are actually correlation coefficients between attributes and principal components, and are calculated from the expression $U\sqrt{\lambda/s_x^2}$, where U is the normalized loading for the species or habitat measurement on the axis, λ is the appropriate "latent root" and s_x^2 is the total variance in the species or habitat value. A "latent root" or "eigen-value" is an estimate of the variance accounted for by a principal component. The proportion of the total variation for species or habitat measurements that can be attributed to variation along a principal axis can be calculated by squaring the adjusted loadings, *i.e.*, $(U\sqrt{\lambda/s_x^2})^2$.

The ordinations for stands and types (*cf.* Greig-Smith *et al.* 1967; Orloci 1967) rather than for species (*cf.* Goodall 1954; Yarranton 1967*a*, *b*) were presented, since the former provide more information about vegetational variation, and type positionings could be compared with those of the environmental model. Only the first three axes of the P.C.A. were presented (Figs. 1*b*, 2, 3; Table 4), since they yielded the most useful comparisons with previous analyses (classification and environmental ordination). They are designated as axes 1, 2, and 3. In each ordination diagram (Figs. 1*b*, 2, 3) the first two axes are drawn with their lengths proportional to the amount of variation extracted by each. The distribution of types or stands along the third axis is represented by the degree of shading within the circles which represent them, varying from completely open at one end of the range of values along the axis to black at the other.

To explore more fully the possibilities offered by P.C.A. for detecting vegetation–environment relationships, habitat attributes—10, 9, and 9 in number, respectively—were included as dominance-type attributes for all types, nonwoody types, and woody types P.C.A. (Table 4). Mean values (based on the number of values summed) for environmental attributes were available for almost all types, but when missing, the type was assigned the mean value of the attribute for the stands in the physiognomic group to which it belonged. The range for each habitat attribute was then scaled from 0 to 100%. (Conductivity values, expressed in μmho's, were transformed to $\log_{10}$ values before being scaled from 0 to 100%.) So that habitat attributes would not contribute to the P.C.A. vegetational ordinations, each scaled value for habitat was divided by 1000, hence reducing to virtually nil the variances and covariances attained by them. It was still possible, however, to judge the relationships of habitat attributes to the vegetationally derived components, since adjusted loadings, *i.e.*, correlation coefficients, were still obtained for the habitat attributes on the various components.

Ordination Effectiveness

One means used to test ordination effectiveness was by comparing the percentages of variation extracted by the component axes of P.C.A. for the several runs. This could not be done, however, for the environmental ordinations, and we required an independent test for all ordinations to judge whether or not the types or stands which were vegetationally similar were positioned close in the ordination, and vice versa. To determine this, we measured actual distances on the diagrams between 75 randomly selected pairs of types (or stands) and correlated these with calculated $2W/(A + B)$ similarity indices (the coefficient of Czekanowski (Greig-Smith 1964)). (Because the pairs of variables being correlated may not be bivariate normal in distribution, Pitman's (1937) distribution-free test of the significance of r was used.)

Only the first two axes of the P.C.A. models were used for measuring distances because (1) we were limited to two axes in the environmental ordinations, and (2) the patterns portrayed in the first two dimensions should represent the most important variation and should portray this simply in order to be understood easily in visual terms. The main value of ordination, we feel, is that it provides a means of simplifying concomitant variation in vegetational and environmental attributes via orderings of the units (stands or types). From the viewpoint of conveying, simply, vegetation–environment relations, a single-axis ordering or a two-dimensional scatter diagram are easily comprehended. A third dimension of variation becomes more difficult to portray as well as to understand, and additional dimensions beyond this rarely offer unequivocal evidence for organism–habitat relationships.

The coefficient of Czekanowski was chosen as the independent measure of the vegetational similarities between pairs of types or stands because in heterogeneous systems, such as our total data represent, it is probably desirable to compare, as does this index, joint abundances (*cf.* Field 1969). The species utilized in calculating $2W/(A + B)$ indices were the same as those used in the corresponding P.C.A. ordinations. We realize that other indices could have been used for estimating stand or type similarity, *e.g.*, a Euclidean distance index (Orloci 1966, 1967). Although the Euclidean indices have properties that make them mathematically more suitable for dealing with geometrical constructions, biological data have several properties, *e.g.*, variable numbers of species and variable numbers of zeros in the data matrix, which vary between comparisons and contribute to an unpreciseness for all similarity indices. Furthermore, there is no parameter of vegetational similarities against which to judge the performance of the various similarity indices. In this light, and with the recognition that any coefficient has shortcomings and must be used only as an estimate of similarity, we decided that $2W/(A + B)$ was as satisfactory an estimate of vegetational similarity as any.

Results

Table 2 presents the percentages of total variance removed by the first three axes for the seven P.C.A. The least variation was removed for the analysis of all types, whereas the most was removed for the analyses of shrub fens and other woody types. The percentages suggest that P.C.A. extracts more variation as the complexity of the data decreases.

Total Data Set

Several kinds of evidence suggest that the arrangement of units in the environmental ordination of all types (Fig. 1*a*) is a good representation of vegetational affinities: (1) dominance types of the same physiognomic group (which are vegetationally similar) occur close together in the diagram; (2) there are bivariate normal patterns when values for species and structural attributes are plotted in the model (Jeglum 1968); and (3) there is a significant correlation at the 1% level between $2W/(A + B)$ similarity indices and measured distances in the diagram for random pairs of types (Table 3). Therefore, it is possible to utilize Fig. 1*a* as a reasonably accurate model with which to compare P.C.A. ordinations of the same data.

TABLE 2. Percentage of total variance removed by the first three axes of the seven P.C.A.

Total data set	
All types	45.3
Half sets	
Nonwoody types	47.3
Woody types	55.3
Quarter sets (stand basis)	
Marshes	51.6
Meadows	48.4
Shrub fens	63.7
Other woody types	57.4

Sequences I and II, which are straight lines in the environmental ordination (Fig. 1*a*), are curved in the P.C.A. vegetational ordination (Fig. 1*b*). Types of marshes (A) lying close to types of tall shrub fen (D) and moist forest (I) are an apparent anomaly and suggest the inappropriateness of P.C.A. for the range of variation included in all types. The correlations between measured distances on the models and similarity indices for pairs of types (Table 3) also indicate that the two-dimensional, environmentally based model (Fig. 1*a*) is a better vegetational ordination than the two-dimensional P.C.A. model (Fig. 1*b*). The environmental ordination removed twice as much variation as the P.C.A. ordination.

Although the P.C.A. ordination model presents a poor picture of relationships in two dimensions, the three-dimensional portrayal has, in fact, separated major clusterings of related types (Fig. 1*b*). The model can be interpreted as clusters of related types at the four apices of a solid geometrical triangle: (*i*) mainly types of marshes (A), (*ii*) types of groups C, E, F, and G which are drawn together by mutual possession of species char-

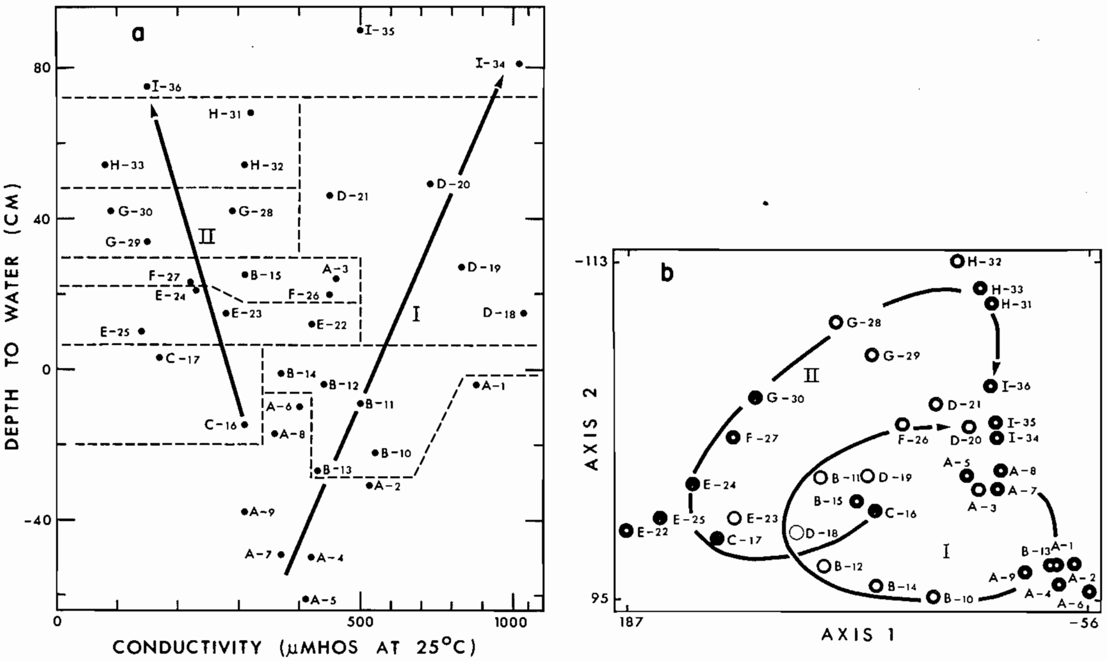

FIG. 1. Ordinations of all types. Sequence I in nutrient-rich wetlands: marshes (A), broad-leaved sedge meadow (B), tall shrub fen (D), *Populus balsamifera* (I-34). Sequence II in nutrient-poor wetlands: *Carex lasiocarpa* (C-16), *Carex lasiocarpa* / small *Carex* (C-17), low shrub fen (E), tamarack swamp (F), transitional muskeg (G), treed muskeg (H), *Picea mariana* / *Pleurozium schreberi* (I-36). Fig. 1*a*, environmental ordination utilizing gradients of water level and of water conductivity; Fig. 1*b*, P.C.A. vegetational ordination, based on 42 species.

TABLE 3. Comparisons of dominance type ordinations

	All types		Nonwoody types		Woody types	
	Environmental ordination	P.C.A. ordination (42 spp.)	Environmental ordination	P.C.A. ordination (48 spp.)	Environmental ordination	P.C.A. ordination (56 spp.)
r	−0.68**	−0.48**	−0.22	−0.36**	−0.66**	+0.60**
Percentage of total variance accounted for ($r^2 \times 100$)	45	23	5	13	44	36

NOTE: The measure of diagram distances was based on the first two axes of the P.C.A. ordinations; species used in calculating $2W/(A + B)$ indices were the same ones used for each P.C.A. run. Each correlation was based on 75 randomly selected pairs of types, with each pair used only once.

**Significant at the 1% level.

acteristic of narrow-leaved sedge meadow (C), (*iii*) types of group H, and (*iv*) types united by certain of the more nutrient-demanding species of broad-leaved sedge meadow (B) and tall shrub fen (D). Although interpretable, this disposition of types is not a simple or easily comprehended model for displaying relationships.

The correlations of habitat measurements on the various component axes (Table 4) also are interpretable, but not with ease. For example, the cluster of types of marshes (A) is located in Fig. 1*b* on the negative side of axis 1, the positive side of axis 2, and with values in the center of the range for axis 3. The high negative axis values for types of group A will yield negative correlations when correlated with habitat features which have their higher values associated with group A, positive correlations for habitat features which have their lower

TABLE 4. Proportion of total variance and adjusted loadings (correlation coefficients) for habitat conditions and selected species on the first three axes of all, nonwoody, and woody types P.C.A.

	All types			Nonwoody types			Woody types		
Ordination axis	1	2	3	1	2	3	1	2	3
Percentage of total variance accounted for	20.1	14.5	10.7	21.4	13.9	12.0	28.6	15.3	11.4
Loadings for habitat conditions:									
Water level	−0.14	0.71**[a]	−0.09	0.36	0.16	0.04	0.83**	−0.11	0.48*
Conductivity, water	−0.20	0.31	0.34*	0.58*	−0.45	−0.23	−0.08	−0.52*	0.00
pH, water	−0.38*	0.49**	0.26	0.42	−0.25	−0.12	−0.13	−0.63*	0.08
pH, moist peat	−0.15	0.56**	0.20	0.35	−0.45	−0.07	0.17	−0.67**	−0.12
Swedish drainage class	0.19	−0.56**	−0.35*	−0.53*	0.49*	−0.01	−0.16	0.70**	−0.08
Depth of peat	0.58**	−0.12	−0.25	−0.44	0.43	0.13	0.47*	0.28	−0.24
Depth of *Sphagnum*–fen peat junction	−0.09	−0.77**	0.02	A	A	A	−0.47*	0.68**	0.33
Humification of peat (von Post 1924)	−0.40*	0.67**	0.00	0.67**	−0.30	−0.18	0.23	−0.72**	−0.37
Mud bottom frequency	−0.56**	0.53**	−0.21	0.69**	0.36	0.11	−0.24	−0.26	−0.28
Sand bottom frequency	−0.23	0.11	0.04	0.03	−0.56*	0.22	A	A	A
Loadings for species:									
Trees									
Larix laricina[c]	0.13	−0.29	0.12	A	A	A	−0.03	−0.20	0.27
Picea mariana	−0.13	−0.33*	−0.08	A	A	A	−0.30	0.19	−0.12
Shrubs									
Salix pedicellaris	0.60**	0.03	−0.18	A	A	A	0.69**	0.05	−0.13
S. planifolia	0.22	−0.07	0.56**	A	A	A	0.09	−0.40	0.54*
Herbs									
Carex atherodes	−0.44**	0.36*	0.00	0.35	−0.01	−0.87**	−0.18	−0.44*	0.16
C. chordorrhiza	0.78**	0.06	−0.56**	−0.47*	0.76**	−0.12	0.82**	0.37	−0.30
C. lasiocarpa	0.79**	0.30	−0.42**	−0.71**	0.49*	0.15	0.92**	0.17	−0.25
Eleocharis palustris[d]	−0.53**	0.49	−0.13	0.78**	0.37	0.27	A	A	A
Scirpus acutus	−0.23	0.08	−0.03	0.08	−0.46	0.28	A	A	A
Mosses									
Drepanocladus aduncus	0.69**	0.35*	0.39*	−0.68**	0.08	0.10	0.63**	−0.46*	0.37
Pleurozium schreberi	−0.14	−0.45**	−0.07	A	A	A	−0.37	0.32	−0.05
Sphagnum fuscum	−0.05	−0.72**	−0.02	A	A	A	−0.33	0.73**	0.25

*Significant at the 5% level (d.f. = *n* (No. types) − 1).
**Significant at the 1% level (d.f. = *n* (No. types) − 1).
[a] Underlined habitat loadings are the highest positive and negative value for each axis.
[b] A, attributes not included in an ordination.
[c] Unless otherwise noted, nomenclature of vascular species is according to Scoggan (1957), that of moss species according to Crum *et al.* (1965).
[d] Includes *Eleocharis smallii* according to Gleason and Cronquist (1963).

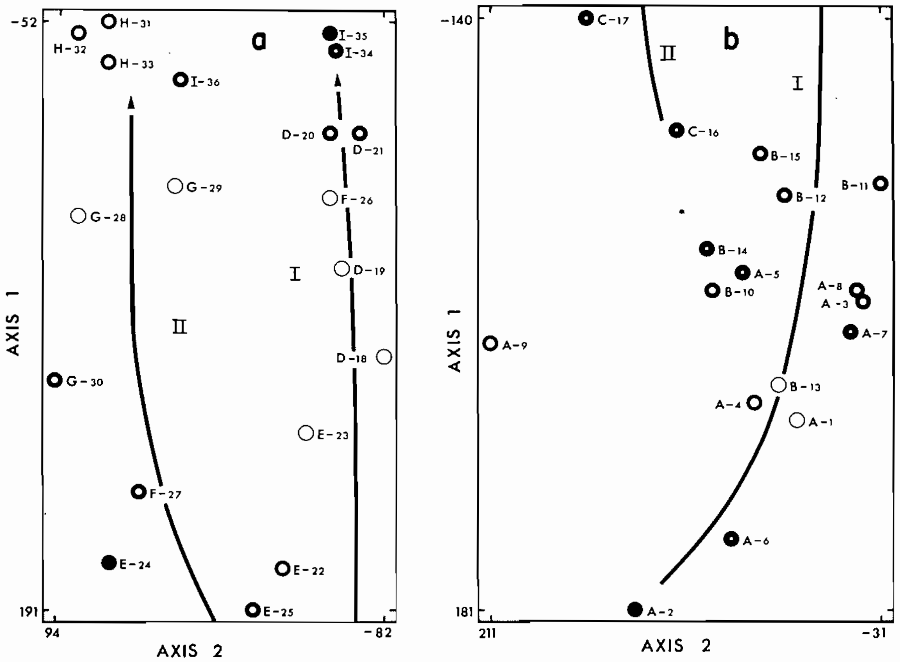

FIG. 2. P.C.A. vegetational ordinations. The figures are oriented to simulate the relative positionings of types in the environmental ordination (Fig. 1*a*) with woody types above nonwoody types. Fig. 2*a*, woody types, based on 56 species; Fig. 2*b*, nonwoody types, based on 48 species.

values associated with group A. By inspecting correlations of habitat measurements on axis 1 (Table 4), hence, it can be ascertained that types of group A are significantly correlated with high values for water pH, low peat depths, high peat humifications, and high mud bottom frequencies. Similar associations between habitat features associated with axis 2 can be extracted. Finally, since group A is located in a relatively central position on axis 3 it may be assumed that none of the highly correlated features of axis 3 are particularly characteristic of group A. Similar logic can be followed for each of the clusters of types, but the method of interpretation is rather laborious and quite difficult to comprehend for anyone unfamiliar with the data and P.C.A. We conclude that since vegetation–environment relationships are not simply portrayed nor easily interpretable, the data could more profitably have been segmented before analysis.

Half Sets of the Data

The P.C.A. ordinations for nonwoody and woody types in Figs. 2*a* and 2*b* are oriented so that comparisons of type locations can be made with the environmental ordination in Fig. 1*a*. The P.C.A. ordinations for half sets of data approximate the environmental model better than the P.C.A. of all types. There is still some slight curvature of the main sequences, but this does not seriously impede the ready interpretation of the models. Most of the same clusters of types that occur in the all types ordination (Fig. 1*b*) also occur in the P.C.A. for half sets of the data (Figs. 2*a*, 2*b*). However, new clusterings and better positionings do appear; for example, A-1, A-4, and B-13 occur together. This is significant because of the close association of *Scolochloa festucacea* and *Carex atherodes*, and the occurrence of *C. atherodes* in one stand of *Typha latifolia* (B-13). In another case, E-22, E-23, and F-26 are located close to tall shrub fen (D). This positioning reflects their common possession of nutrient-demanding dominants of broad-leaved sedge meadow (B). Comparison of Fig. 1*a* with Fig. 2 shows, therefore, that P.C.A. of segmented data is as useful as the environmental ordination of the full set of data.

The correlations between measured distances on the models and similarity indices for pairs of types (Table 3) also indicate that the P.C.A. models have increased efficiency relative to the environmental positionings of the same types. The P.C.A. for woody types has accounted for almost as much variance as the environmental ordination, whereas the P.C.A. for nonwoody types has extracted more than twice as much variance as the environmental model. Hence, with less heterogeneity the two-dimensional P.C.A. ordinations become more effective.

Adjusted loadings (correlation coefficients) for habitat attributes for the half sets of data (Table 4) yield interpretable vegetation–environment relationships. On the first three axes of the nonwoody types P.C.A., the highest loadings are for measures representing fertility and disturbance regimes, whereas depth to water level, the single measurement representing moisture regime, did not attain significance. The latter is somewhat surprising as moisture regime has usually been ascribed importance in studies of sedge- and reed-dominated vegetation. Hence, we would suggest further study to verify the hypothesis that moisture regime takes on less importance relative to fertility and disturbance regimes. For the woody types P.C.A., the habitat loadings suggest that moisture and fertility are the two most important regimes, and this concurs with our field judgements. Thus, habitat loadings confirm that the P.C.A. models for half sets of data are ecologically meaningful.

It appears that the P.C.A. for half sets of data, in contrast to the P.C.A. for the complete set of data, are more readily understood and interpreted, an important desideratum for effective ordination.

Quarter Sets of the Data

The four ordinations are compared in Fig. 3. Correlations of measured diagram distances with $2W/(A + B)$ indices of similarity (Table 5) show that P.C.A. for all four segments resulted in superior ordination models in comparison to the environmental models. This means that when the data are relatively homogeneous, a vegetational analysis is more effective for indicating vegetational relationships than is an environmental approach.

The variability accounted for by the P.C.A. models on the first two axes (Table 5) increases with decreasing wetness of the segment, being lowest in marshes (A) and highest in other woody types (F–I). This suggests that the most heterogeneity is found in emergent vegetation, and the least in the driest wetland types, since two principal components can remove progressively more of the total variation as the vegetation segment becomes drier.

The environmental models remove the least variation for the marshes (A), somewhat more for the shrub fens (D and E), and the highest variation for meadows (B and C) and other woody types (F–I) (Table 5). This shows increasing response of vegetational variation to water level and conductivity.

Discussion

It has been shown that P.C.A. presents distorted vegetational patterns when applied to a complex set of data which includes several physiognomic groups. Analysis of segments of half the data results in less distortion of the vegetational variation. For quarter segments of the data, distortion is minimal. Therefore, P.C.A. vegetational ordinations are increasingly useful for data with decreasing complexity. This agrees with current ideas that vegetational ordinations are most likely to be useful in data with low levels of heterogeneity (Greig-Smith 1964; Greig-Smith *et al.* 1967).

Conversely, the environmental ordination, based on depth to water level and conductivity, shows the reverse trend. The best vegetational positionings are obtained for the total data set; for half sets of the data, positionings are about as good as in the P.C.A. models; for the quarter-data segments, the environmental models are inferior to the vegetationally derived models. Of course, this conclusion may be influenced by shifting of importance of factors from the total data set to half sets to quarter sets; *e.g.*, the most important factors may be disturbance and substrate in marshes rather than water level and conductivity. This suggests that the quality of the environmental ordination approach will depend on the availability of a large number of environmental measures and the choice of the right ones for the model.

It must be acknowledged that in this series of comparisons the step between half segments and quarter segments of data entails a change from consideration of dominance types to consideration of stands. The major effect of this

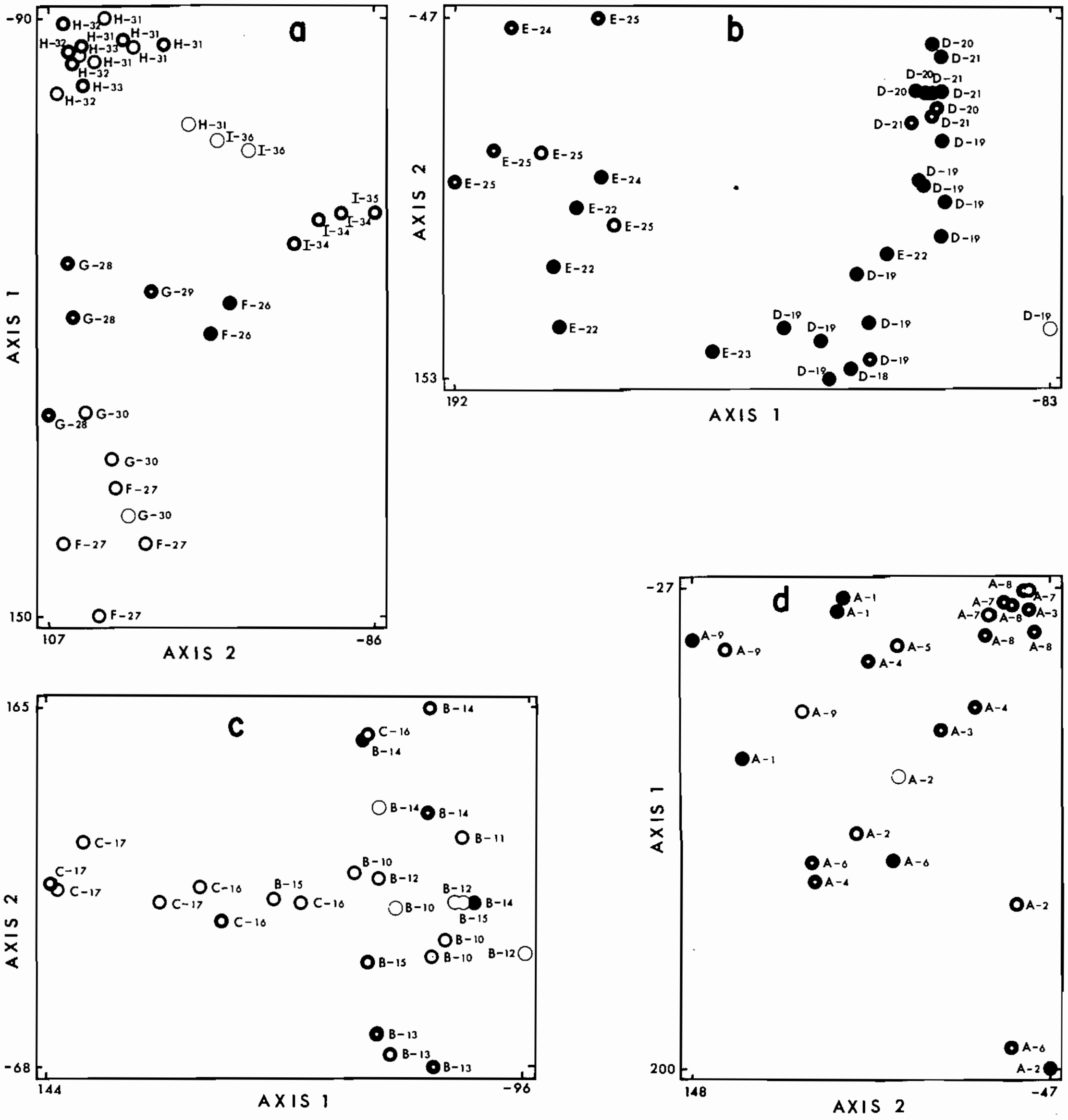

FIG. 3. P.C.A. ordinations for four sets of stands. The figures are oriented to simulate the relative positionings of types in the environmental ordination (see Fig. 1*a*). Fig. 3*a*, other woody types (F–I), based on 35 species; Fig. 3*b*, shrub fens (D and E), based on 32 species; Fig. 3*c*, meadows (B and C), based on 28 species; Fig. 3*d*, marshes (A), based on 26 species.

probably is on the environmental ordinations; the positionings of types according to *means* of depth to water level and conductivity probably yield much better positionings with respect to vegetational relationships than do the environmental positionings which use the more variable single measurements for depth to water level and conductivity. This does not alter the fact that P.C.A. has increased in effectiveness as the segments became less complex.

The step between half and quarter segments also entails a change in the numbers of species, there being around 50 for the type analyses and around 30 for the stand analyses. Further analyses of species numbers, not presented in the results, showed that a decrease in species from *ca.* 50 to 20 resulted in poorer ordina-

TABLE 5. Comparison of stand ordinations

	Marshes (A)		Meadows (B & C)		Shrub fens (D & E)		Other woody types (F–I)	
	Environmental ordination	P.C.A. vegetational ordination	Environmental ordination	P.C.A. vegetational ordination	Environmental ordination	P.C.A. vegetational ordination	Environmental ordination	P.C.A. vegetational ordination
r	−0.18	−0.32**	−0.47**	−0.63**	−0.33**	−0.74**	−0.45**	−0.81**
Percentage of total variance accounted for ($r^2 \times 100$)	3	10	22	40	11	55	20	66

NOTE: The measure of diagram distance was based on the first two axes of the P.C.A.; species used in calculating $2W/(A + B)$ indices were the same as those used for the P.C.A. runs, *i.e.*, 26, 28, 32, and 35. Each correlation was based on 75 randomly selected pairs of types, with each pair used only once.

*Significant at the 5% level.

**Significant at the 1% level.

tions for woody types and nonwoody types. Kershaw (1968) found that a reduction in species from 75 to 30 resulted in only slight differences in his plot ordinations whereas his species ordinations showed considerable improvements. Austin and Greig-Smith (1968) applied P.C.A. ordinations to a set of rain forest data using the 10, 15 . . . 50, 75, and 100 most abundant species. They found that less abundant species contributed little information, and that less than 25% (50 species) of the most abundant species of the flora could give a useful ordination. Further studies are required to understand better the effects of species numbers on P.C.A. ordination in sets of data with different levels of heterogeneity. Nonetheless, the present study indicates that P.C.A. increases in effectiveness as the segments become less complex physiognomically, in spite of the decrease in numbers of species for quarter sets of data.

In comparing the effectiveness of P.C.A. with environmental ordinations, we used only the first two axes of the P.C.A. models for measuring distances. Of course, P.C.A. continues to extract variation in subsequent components, thus offering the opportunity to interpret this further variation in terms of vegetation–environment relationships. Environmental models, on the other hand, are dependent on measured habitat features, and on the subjective, *a priori* choice of those measurements which are thought to be most strongly correlated with vegetational variation. In order to construct a third axis in the environmental ordinations, we must have another measurement which seems to be correlated with the variation not accounted for by the measurements used for the first two axes. P.C.A., therefore, offers greater objectivity in reducing variation in multivariate data; unfortunately, it has the disadvantage that curvilinear patterns of vegetational variation are produced for data which are too complex.

The lack of effectiveness of P.C.A. on data which are too complex seems due to the obtaining of curvilinear patterns of vegetational variation in the ordination models. Curvilinear patterns may be discerned in a number of ordination diagrams in the literature (*e.g.*, Loucks 1962; Gittins 1965; Ayyad and Dix 1964; Bannister 1966, 1968; Greig-Smith *et al.* 1967), although it is not readily verifiable whether or not a particular curved sequence of stands actually represents a linear vegetational gradient. However, when one can observe that a linear sequence of stands in an environmental ordering is portrayed as a curved sequence in a vegetational ordination, there is good reason to doubt the value of vegetational ordination in detecting vegetation–environment relationships. This seems to be demonstrated in the work by Loucks and by Ayyad and Dix, but recently more directly by Swan (1970).

Swan demonstrates with artificial sets of data that a sequence of stands along a single environmental gradient may appear as a curve in two-dimensional diagrams, oscillating in the third dimension. For the sets in which species had wide amplitudes and overlapped across the whole gradient, distortion in the two-dimensional vegetational ordinations was low. But in the sets in which all species had narrow amplitudes and correspondingly little overlap along the gradient, distortion in the ordination models was so pronounced that the original gradient could not have been detected without a thorough knowledge of the data. The reason for the increasing distortion was thought to be the influence of bell-shaped performance curves for species across the gradient and zero values in the raw data matrix (*cf.* Field 1969). (Although Swan uses an ordination technique other than P.C.A., Noy-Meir and Austin (1970) obtain essentially the same results on Swan's data with P.C.A.)

If both vegetational and environmental data are available, it seems clear that the choice of ordination, either vegetational or environmental, will be influenced greatly by the range of physiognomic complexity contained in the data. Hence, Whittaker (1956) found a direct gradient analysis very effective for portraying the great variation in vegetation and environment of the Great Smoky Mountains, whereas Bray (1961) concluded that for a relatively homogeneous set of data from transects radiating from open-grown trees in a Wisconsin prairie, vegetational ordinations were more informative than any single gradient ordering of stands. Similarly, Yarranton (1967*b*) found that a physiognomically uniform set of data for saxicolous bryophyte vegetation (species-contact data) was effectively analyzed by P.C.A., when treated either *in toto* or as two segments.

It would be useful for the ecologist to be able to decide *a priori* whether P.C.A. can be applied effectively to a given set of data. Available to assist the ecologist are empirical studies such as ours and the ones cited above with which the given set of data may be compared to judge the relative degree of physiognomic complexity included in the set. If the data in question are judged too complex, segmentation is advisable before analysis with P.C.A. It has been suggested by Swan (1970) that some inherent characteristics of complex ecological data, especially the zeros in the raw data matrix and the bell-shaped normal distribution patterns demonstrated in biological data, prevent the effective application of vegetational ordination. Swan tried to correct for one of these problems by assigning quantitative values to the zeros in the matrix. When zero-adjusted matrices of single-gradient theoretical models were ordinated, the resulting stand distributions, which had shown curvilinear patterns for ordinations of the unadjusted matrices, approximated the original linear gradient models. This approach of "completing" the quantitative data by assigning values to the zeros may prove one of the more profitable avenues of research into vegetational ordination technique.

In applying P.C.A. in this study, an extension of the technique was found extremely useful in interpreting vegetation–environment relationships. This was the calculation of adjusted loadings, actually correlation coefficients, for the attributes on the derived principal components. With these values, tables of significance were consulted to determine the degree of significance of the correlation. The value of inspecting correlations for numerous species on several components might be questioned. However, when habitat attributes are also correlated with the components, and certain components can be shown to be related to certain factor-complexes, then the correlations for species take on ecological meaning.

Habitat attributes were included in the type P.C.A., and made comparable to species attributes by scaling them from 0 to 100. Following this the habitat attributes could have been analyzed along with the species attributes without any further modification, in which case an ecosystemic ordination based on both kinds of attributes would have resulted. Although such an approach makes no initial assumptions, there are several important reasons for not including the habitat attributes in the ordination technique. First, the relative contribution of species versus habitat attributes to the analysis would be influenced by the relative numbers of each. Second, had the environmental attributes been included, a preponderance of measurements for one regime, *e.g.*, in our case measurements related to the fertility regime, would have yielded an ordination stressing that regime, whereas in fact another regime for which there might be few or no measurements, *e.g.*, in our case moisture regime, might be equally or more important than fertility in some of the segments which were considered. For these reasons we decided to allow the P.C.A. to be based only on vegetational variation, and then to find environmental correlates to the vegetationally derived components. To achieve this each scaled habitat attribute was divided by 1000, hence reducing to virtually nil their contribution to the variance–covariance matrix. It was still possible, however, to judge the relationships of habitat attributes to the vegetationally derived components, since adjusted loadings (correlation coefficients) were still obtained between the habitat attributes and the components.

The analysis of vegetational and environmental data by mathematical methods requires the skills of a mathematician and the ecologist's knowledge of the material. The unifying objective is the detection of ecological relationships. The results of this study suggest that too much physiognomic complexity in the data may prevent attaining this goal when using P.C.A. Further studies of the following nature are suggested: (1) the testing of P.C.A. on theoretical models of data with known characteristics and (2) the testing of P.C.A. on real data already analyzed in some detail. Heterogeneity, zero values in the raw data matrix, and numbers of species used in the study analysis stand out as key problem areas.

The data on which this work was based were collected by the senior author as part of a Ph.D. program in the Department of Plant Ecology, University of Saskatchewan, Saskatoon, Saskatchewan. The program was supported by grants from the National Research Council of Canada, the Saskatchewan Agricultural Research Foundation, and the Institute for Northern Studies, University of Saskatchewan. Sincere thanks are extended to R. L. Dix, Department of Botany and Plant Pathology, Colorado State University, Boulder, Colorado, and R. T. Coupland, Department of Plant Ecology, University of

Saskatchewan, for advice and support during the research program. Thanks also are extended to H. A. Crum, University of Michigan, Ann Arbor, for identifying bryophytes, and to G. W. Argus, Canadian Forestry Service, Ottawa, for identifying and verifying *Salix* specimens.

Anderson, T. W. 1958. An introduction of multivariate statistical analysis. John Wiley and Sons Inc., New York.

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

Austin, M. P., and Greig-Smith, P. 1968. The application of quantitative methods to vegetation survey. II. Some methodological problems of data from rain forest. J. Ecol. **56**, 827–844.

Ayyad, M. A. G., and Dix, R. L. 1964. An analysis of a vegetation–microenvironmental complex on prairie slopes in Saskatchewan. Ecol. Monogr. **34**, 421–442.

Bannister, P. 1966. The use of subjective estimates of cover-abundance as the basis for ordination. J. Ecol. **54**, 665–674.

——— 1968. An evaluation of some procedures used in simple ordinations. J. Ecol. **56**, 27–34.

Bray, J. R. 1961. A test for estimating the relative informativeness of vegetation gradients. J. Ecol. **49**, 631–642.

Crum, H., Steere, W. C., and Anderson, L. E. 1965. A list of the mosses of North America. Bryologist, **68**, 377–432.

Dugle, J. R. 1966. A taxonomic study of western Canadian species in the genus *Betula*. Can. J. Bot. **44**, 929–1007.

Field, J. G. 1969. The use of the information statistic in the numerical classification of heterogeneous systems. J. Ecol. **57**, 565–569.

Gittins, R. 1965. Multivariate approaches to a limestone grassland community. I. A stand ordination. J. Ecol. **53**, 385–401.

Gleason, H. A., and Cronquist, A. 1963. Manual of vascular plants of northeastern United States and adjacent Canada. D. van Nostrand Co., Princeton, N.J.

Goodall, D. 1954. Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. Aust. J. Bot. **2**, 304–324.

Greig-Smith, P. 1964. Quantitative plant ecology. 2nd Ed. Butterworth & Co. (Publishers) Ltd., London.

Greig-Smith, P., Austin, M. P., and Whitmore, T. C. 1967. The application of quantitative methods to vegetation survey. I. Association-analysis and principal component ordination of rain forest. J. Ecol. **55**, 483–503.

Holland, D. A. 1969. Component analysis: an aid to the interpretation of data. Exp. Agr. **5**, 151–164.

Jeglum, J. K. 1968. Lowland vegetation at Candle Lake, southern boreal forest, Saskatchewan. Ph.D. thesis. University of Saskatchewan, Saskatoon, Saskatchewan.

Kershaw, K. A. 1968. Classification and ordination of Nigerian savanna vegetation. J. Ecol. **56**, 467–482.

LaRoi, G. H. 1967. Ecological studies in the boreal spruce–fir forests of the North American taiga. I. Analysis of the vascular flora. Ecol. Monogr. **37**, 229–253.

Lindsay, A. A. 1955. Testing linestrip method against full tallies in diverse forest types. Ecology, **36**, 485–495.

Loucks, O. L. 1962. Ordinating forest communities by means of environmental scalars and phytosociological indices. Ecol. Monogr. **32**, 137–166.

Noy-Meir, I., and Austin, M. P. 1970. Principal component ordination and simulated vegetational data. Ecology, **51**, 551–552.

Orloci, L. 1966. Geometric models of ecology. I. The theory and application of some ordination models. J. Ecol. **54**, 193–215.

——— 1967. An agglomerative method for classification of plant communities. J. Ecol. **55**, 193–205.

Pearce, S. C. 1969. Multivariate techniques of use in biological research. Exp. Agr. **5**, 67–77.

Pitman, E. J. G. 1937. Significance tests which may be applied to samples from any population. II. The correlation coefficient test. Suppl. J. Roy. Statist. Soc. **4**, 225–232.

Post, L. von. 1924. Das genetlische System der organogen Bildungen Schwedens. Com. Int. Pedol. IV Comm. (Helsinki), **22**, 287–304.

Ritchie, J. C. 1956. The vegetation of northern Manitoba. I. Studies in the southern spruce forest zone. Can. J. Bot. **34**, 523–561.

Scoggan, H. J. 1957. Flora of Manitoba. Nat. Mus. Can. Bull. 140. pp. 1–619.

Seal, H. L. 1964. Multivariate statistical analysis for biologists. Methuen and Co., Ltd., London.

Sjörs, H. 1948. Mire vegetation in Bergslagen, Sweden (in Swedish, English summary). Acta Phytogeogr. Suec. **21**, 1–299 (plus appendices).

Swan, J. M. A. 1970. Examination of some ordination problems by use of simulated vegetational data. Ecology, **51**, 89–102.

Thurstone, L. L. 1947. Multiple factor analysis. University of Chicago Press, Chicago.

White, K. L. 1965. Shrub-carrs of southeastern Wisconsin. Ecology, **46**, 286–304.

Whittaker, R. H. 1956. Vegetation of the Great Smoky Mountains. Ecol. Monogr. **26**, 1–80.

——— 1967. Gradient analysis of vegetation. Biol. Rev. **42**, 207–264.

Whittaker, R. H., and Fairbanks, C. W. 1958. A study of plankton copepod communities in the Columbia Basin, southeastern Washington. Ecology, **39**, 46–65.

Yarranton, G. A. 1967*a*. Principal components analysis of data from saxicolous bryophyte vegetation at Steps Bridge, Devon. I. A quantitative assessment of variation in the vegetation. Can. J. Bot. **45**, 93–115.

——— 1967*b*. Principal components analysis of data from saxicolous bryophyte vegetation at Steps Bridge, Devon. II. An experiment of heterogeneity. Can. J. Bot. **45**, 229–247.