GEOMETRIC MODELS IN ECOLOGY

II. AN EVALUATION OF SOME ORDINATION TECHNIQUES

BY M. P. AUSTIN AND L. ORLOCI*

Department of Botany, University College of North Wales, Bangor

CONTENTS

I. INTRODUCTION

A number of papers have recently been published on vegetation types as different as tropical forest in Brunei (Ashton 1964), tundra in Canada (Larsen 1965) and limestone grassland in Britain (Gittins 1965a, b, c), in which the ordination techniques of the Wisconsin school (Bray & Curtis 1957; Curtis 1959; Beals 1960; Maycock & Curtis 1960) have been used. These methods are reviewed by Greig-Smith (1964) and some comments on them can be found in the work of Dagnelie (1960), Dale (1964) and Lambert & Dale (1964).

The Wisconsin methods provide a simple spatial representation of stands of vegetation such that the major features of the stand relationships are apparent on inspection and their ecological interpretation correspondingly simplified. Their success in interpretation can be assessed from the published work where the phytosociological results are in accord with independent environmental data.

The purpose of this paper is an evaluation of the methods used by the Wisconsin workers, particularly the distance measure and axis construction technique introduced by Bray & Curtis (1957). This is achieved through theoretical considerations and by comparison with two alternative methods including (a) the use of geometric distance and perpendicular axes, and (b) principal components analysis of a particular similarity coefficient.

II. THE BRAY & CURTIS METHOD

A. *The coefficient* $C = 2w/(a+b)$

This coefficient is used by the Wisconsin workers as an expression of interstand similarity; w is the sum of the lesser values of the species scores in the two stands

---

* Now at: Department of Botany, University of Western Ontario, London, Ontario, Canada.

compared; *a* and *b* are the sum of the species scores in each stand. The complement $1-C$ or the corrected form $C_{\max}-C$ is assumed to be a Euclidean measure of distance (Bray & Curtis 1957) and used as such in axis construction.

In the work quoted above, and in others (e.g. Beals 1960), $C_{\max}$ is the average similarity value calculated from repeated sampling of the same stand. McIntosh & Hurley (1964), however, use as $C_{\max}$ the highest value actually found in the set of interstand comparisons. This definition of $C_{\max}$ is adopted in the present work. Maycock & Curtis (1960), and Ayyad & Dix (1964) apparently employ the similarity coefficient directly as a 'distance'.

Any measures to be used in ordinary geometric manipulations, as for instance the construction of ordinations in Euclidean space, must be either an angle or a distance. A definition of Euclidean distance (D_{jh}) between two stands (j, h) can be obtained, by applying the Pythagorean theorem to the species scores as rectangular co-ordinates in multidimensional space, as

$$D_{jh} = \sqrt{\left[\sum_{i=1}^{n} (X_{ij}-X_{ih})^2\right]}^{*}$$

which is explained in Orloci (1966)†.

If $C_{\max}-C$ is also a Euclidean measure of distance then it ought to bear an exact linear relationship to D. The ratio $D/(C_{\max}-C)$ should, therefore, remain constant regardless of the stands compared. This condition was tested in 528 paired measurements of D and $C_{\max}-C$ obtained from the comparison of thirty-three stands of dune slack vegetation (see Part I). The values of the ratio vary between 1·1 and 2·6, a range of more than 230%.

A direct comparison of the magnitude of $C_{\max}-C$ and D cannot be made unless they are transformed to a common scale. Fig. 1 shows the values plotted after transformation to zero mean and unit variance. Deviations from the 45° diagonal indicate the incidence and degree to which $C_{\max}-C$ fails to measure the actual distance (D) deviating from it in either a positive or negative direction.

These results disqualify $C_{\max}-C$ as a Euclidean distance. $C_{\max}-C$ violates the Euclidean definition of distance on at least three separate counts. (1) Negative distance may arise, e.g. when $C_{\max}$ [redacted] is determined by repeated sampling of a single stand. (2) Two stands may have zero distance although not identical; this occurs when $C_{jh} = C_{\max}$. (3) The sum of two sides of a triangle formed by three stands as points in space may be less than the third side, indicating that the sides are not straight lines.

These anomalies have been observed in several sets of data analysed in this department. They result from the coefficient alone and should not be confused with errors in the positioning of stands due to errors in the estimation of the species scores.

In order to study the effect of the 'distance' $C_{\max}-C$ on the arrangement of stands independently of other aspects of the ordination techniques, a comparison is made between $C_{\max}-C$ and D on the basis of stand ranks along three arbitrarily selected axes. Each axis is defined by two reference stands (R_1, R_2) as in Fig. 1 of Part I.

Stand positions along the axis (R_1, R_2) are calculated as

$$a_j = \frac{D_{1j}^2+D_{12}^2-D_{2j}^2}{2D_{12}}‡$$

---

* The use of Euclidean distance in taxonomy is described and discussed by Sokal & Sneath (1963).

† This contribution is subsequently cited as Part I.

‡ An expression also used by Beals (1960).

substituting the values of either $C_{max}-C$ or D in the equation. Spearman's rank correlation between stand positions obtained from the two different measures of interstand distance is shown in Table 1.

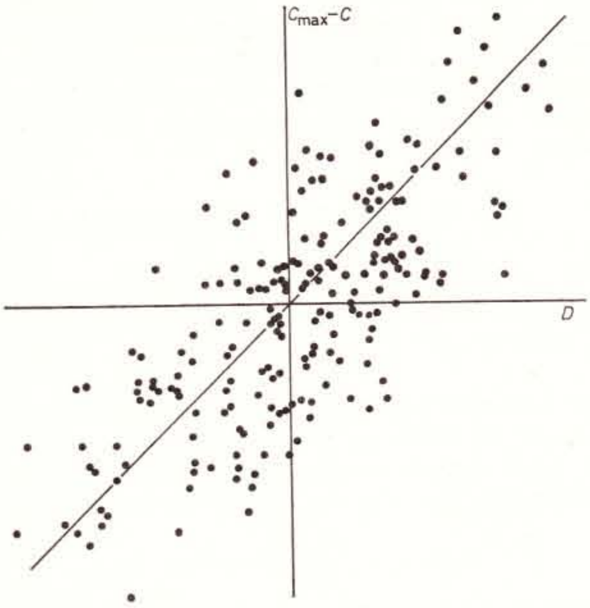

FIG. 1. Relationship between $C_{max}-C$ and D when expressed in standard deviation units. (Random sample from 528 observations.)

The coefficient indicates that no significant change in rank in these directions occurs regardless whether $C_{max}-C$ or D is used in defining stand positions, although the distances between stands do, of course, differ. The high rank correlation between positions on these arbitrarily selected axes should not be interpreted as an indication that potentially identical ordinations can be obtained from the two distance matrices. The coefficients may result in entirely different ordinations as a consequence of different end stands being chosen.

Table 1. *Rank correlations between* $C_{max}-C$ *and* D

Reference axes	Spearman's r
7, 30	0·977
9, 22	0·941
14, 18	0·974

B. *Axis construction*

If stands are represented as points in space whose co-ordinates are the individual species scores, then a maximum of $n-1$ axes may be required to account for all distances between n stands. Efficient simplification is achieved by extraction of a limited number of new axes positioned to account for a large portion of the interstand distances. When stands are plotted on these new axes, the resulting scatter diagrams represent a projection of the multidimensional point cluster into a space defined by the ordination axes. The distortion in the placement of the stands due to projection is minimal when projection is into a space where the first axis coincides with the direction of maximal variation between stands, and the subsequent axes are so positioned that each accounts for a maximal portion of the residual variation.

The technique used by the Wisconsin workers does not achieve this orientation of the axes. This is due to the choice of extreme stands as the reference points for the axes, which need not lie in the direction of maximal variation, or, more precisely, the position vectors do not coincide with this direction. In addition, their axes lie at oblique angles to each other and when plotted as perpendiculars another type of distortion in the arrangement of stands results.

III. ALTERNATIVE TECHNIQUES

A. *Perpendicular axes ordination*

An ordination technique based on Euclidean distance and defined perpendicular axes removes some of the deficiencies of the Bray & Curtis method. An outline of the technique is given in Part I.

This technique provides an orthogonal projection of the stands as points into a three-dimensional space (A, B, C) which the Wisconsin techniques fail to provide. However, the loss of efficiency which results from the selection of the extreme stands as the reference points of the ordination axes cannot be overcome by this simple procedure.

B. *Principal components analysis*

Application of more efficient ordination procedures requires the use of an appropriate similarity coefficient. A coefficient which leaves the original stand relationships unaltered and is suitable in Q-type analyses can be derived as

$$\sum_{i=1}^{n} (X_{ij} - \bar{X}_i)(X_{ih} - \bar{X}_i)$$

This is the weighted similarity coefficient (*WSC*). The derivation of and reasons for the use of this coefficient are discussed in Part I. The values X_i represent species scores and $\bar{X}_i$ is the average score of the ith species for all stands.

Principal component analysis of a matrix of the weighted similarity coefficients provides axes which lie in the direction of maximal vegetational variation among stands. The characteristic solutions (eigen or latent vectors, cf. Hohn 1964), after standardization in each set so that the sum of their squares is equal to the appropriate characteristic root (latent or eigen values), can be used as the stand co-ordinates on the principal axes.

IV. COMPARISON OF ORDINATIONS

A comparison of the three ordination techniques mentioned above was made, based on the original frequency scores of species in stands of dune slack vegetation on Newborough Warren (Part I). The necessity for standardization has been considered (cf. Greig-Smith 1964; Dale 1964), but no conclusive answer to this problem has yet been found. The original data have therefore been used to avoid the effect of incorporating modifications incidental to the technique of ordination.

Three two-dimensional ordinations of a single set of data were made using the techniques outlined previously. The Bray & Curtis ordination utilized $C_{\max} - C$ as a 'distance' measure and the co-ordinates were calculated from Beal's (1960) formula by means of a

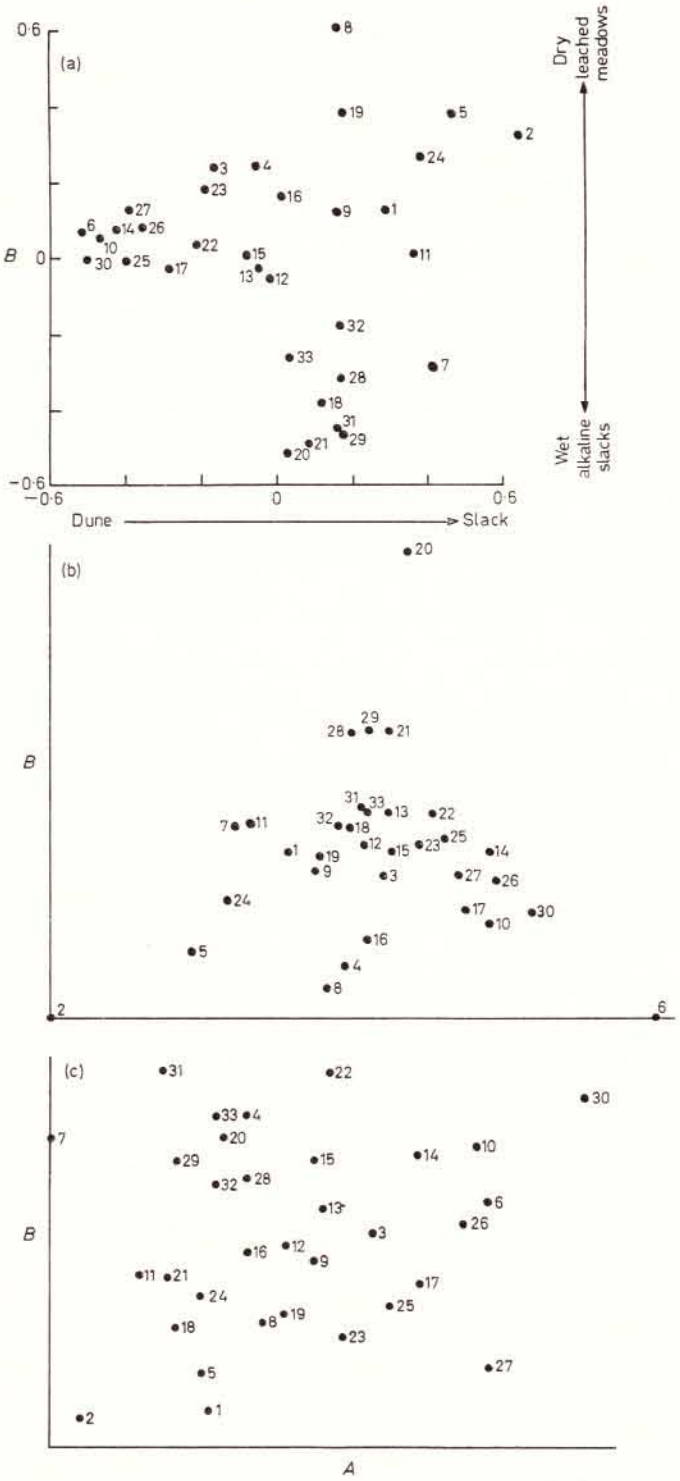

FIG. 2. Comparison of stand positions on axes *A* and *B*. (a) Principal component ordination, (b) perpendicular axes, and (c) Bray & Curtis.

specially written computer programme. The principal component analysis of the weighted similarity coefficient was carried out by a standard computer programme for determining the latent roots and vectors of a symmetric matrix, while the perpendicular axes ordination required a special programme. The scatter diagrams of these ordinations are shown in Fig. 2(a–c).

A direct comparison of these scatter diagrams is difficult as they represent planes of different orientation with respect to the multidimensional cluster of stands. However, the ordination distances between stands can be compared to the original distances defined by D. The coefficient of determination ($100r^2$) is used to measure the proportion of the original distances which is accounted for by the ordination distances (Table 2).

Although there is a high correlation between the original values of $C_{\max}-C$ and D ($r = 0{\cdot}684$), the values for the Bray & Curtis method, as indicated in Table 2, show that their ordination explains very little of the original interstand relationships. The perpendicular axes ordination accounts for 16% of the variation in the original distances, and the principal components 55%.

Table 2. *Comparison of ordinations*

	Method		
	Bray & Curtis	Perpendicular axes	Principal components
r	0·0278	0·4020	0·7431
r^2	0·0008	0·1616	0·5521
% of total variance accounted for	0·08	16·16	55·21

In the Bray & Curtis ordination (Fig. 2c) the stands appear more evenly dispersed than in the other two ordinations (Fig. 2a, b). There are also some obvious anomalies: the distances between certain pairs of stands (e.g. 2 and 30, 3 and 5, 1 and 30) are substantially greater in the scatter diagram (Fig. 2c) than originally calculated by $C_{\max}-C$. The floristic variation in the species poor stands (transitional dune types) is known to be much less than that between these community types and the more typical slacks. Examination of the Bray & Curtis scatter diagrams indicates the variation to be equal to that in the slacks. The distance, for instance, between stands 30 and 27 is almost equal to their distances from stand 28.

Some indication of the effectiveness of the Bray & Curtis technique is found in their paper (1957) in which they quote the correlation between the ordination distances and the original similarity coefficients (C) as $-0{\cdot}35$, corresponding to *c.* 12·3% of the total variance.

The perpendicular ordination is free from such anomalies but does not present as clear an interpretation of the gradients as does the principal components analysis. A further evaluation can be attempted in terms of the major ecological patterns and gradients which are revealed by the ordinations. The principal component analysis is interpreted as follows: the first axis indicates a vegetational gradient from the communities of the transitional and incipient dunes or dry slacks predominantly composed of *Ammophila arenaria* to the slack communities (Fig. 3a). The second axis distinguishes the communities of the wet alkaline slacks from the more stabilized, leached meadow types. The species plotted on the scatter diagrams indicate the positions and extent of these major types (Figs. 3–5). *Potentilla anserina* is characteristic of the meadow group, and *Carex serotina* of the wet alkaline slacks. These groups can be recognized in the other ordinations (Figs. 3b–5b, 3c–5c) but the separation of groups and clarity of the gradients is much less.

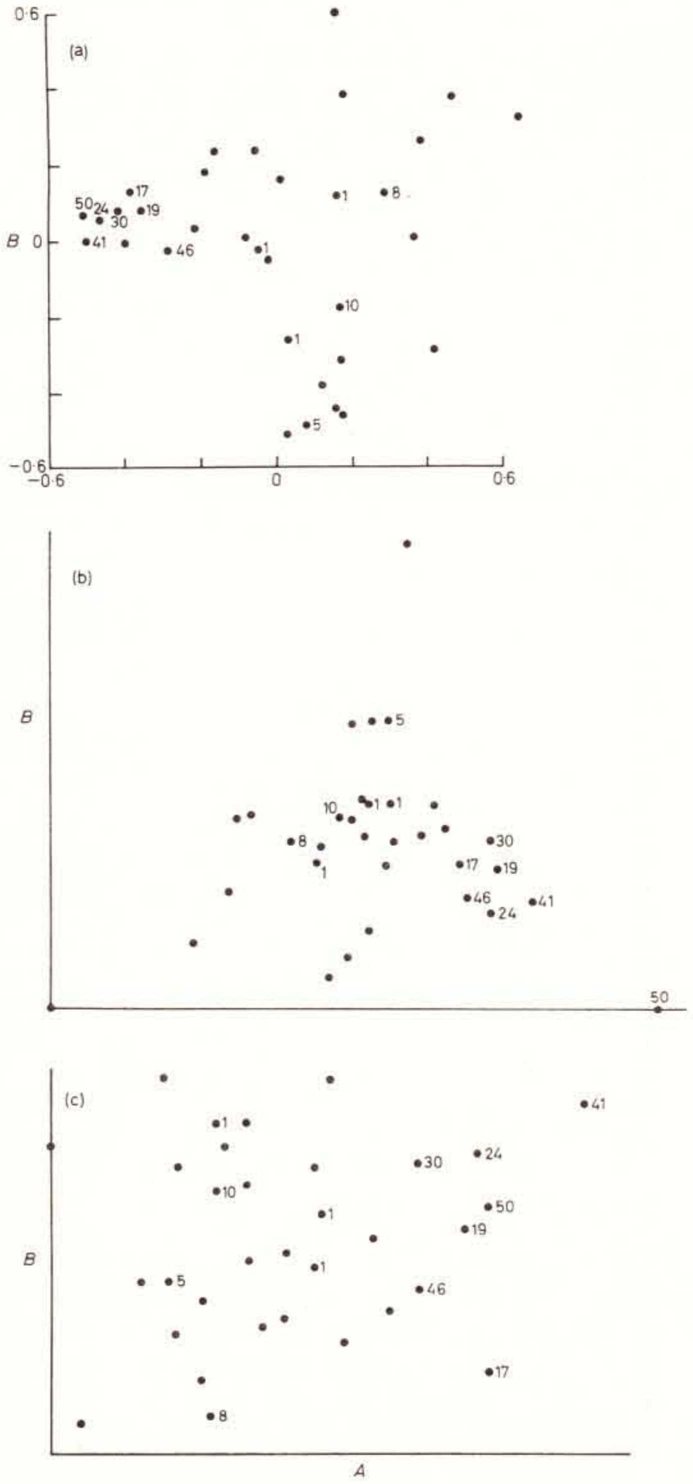

FIG. 3. Comparison of *Ammophila arenaria* distribution in the three ordinations (a)–(c) of Fig. 2. Figures are percentage frequencies of *A. arenaria*.

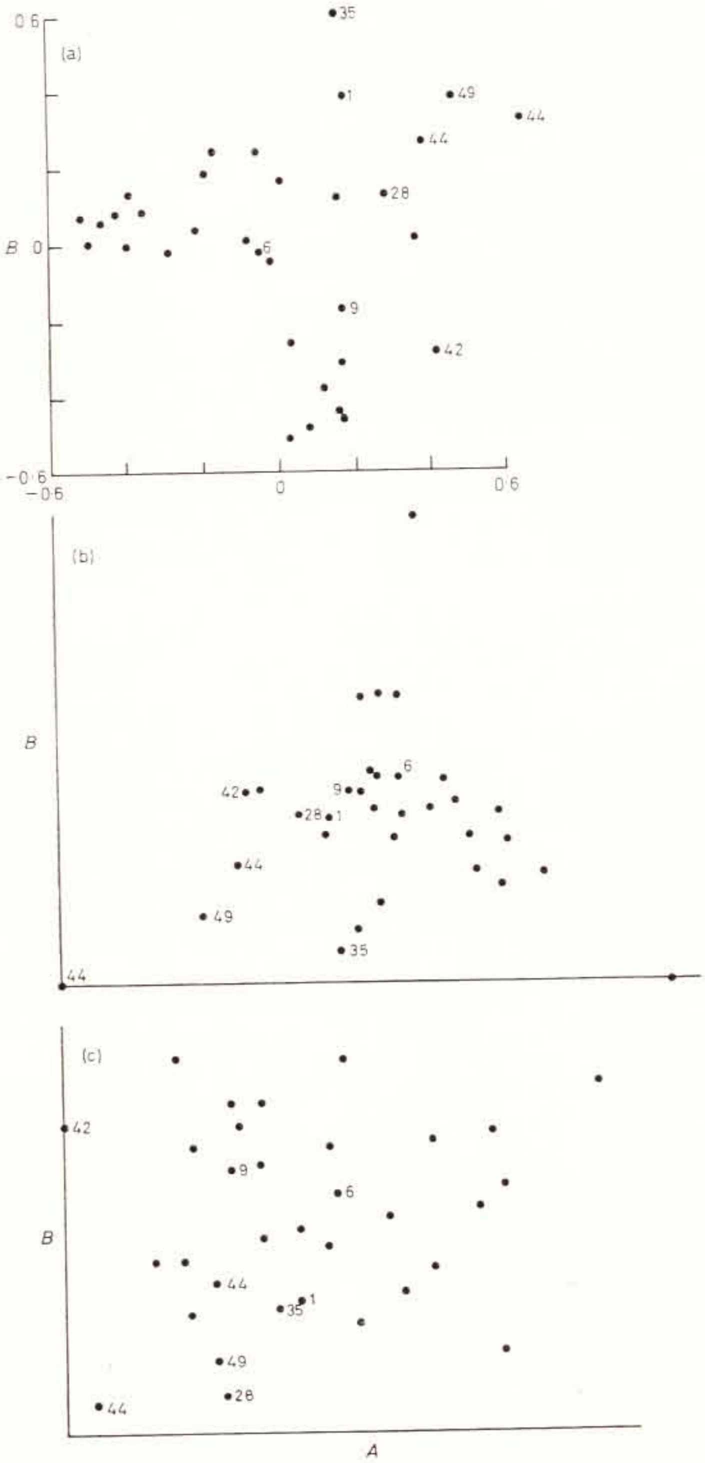

FIG. 4. Comparison of *Potentilla anserina* distribution in the three ordinations (a)–(c) of Fig. 2. Figures are percentage frequencies of *P. anserina*.

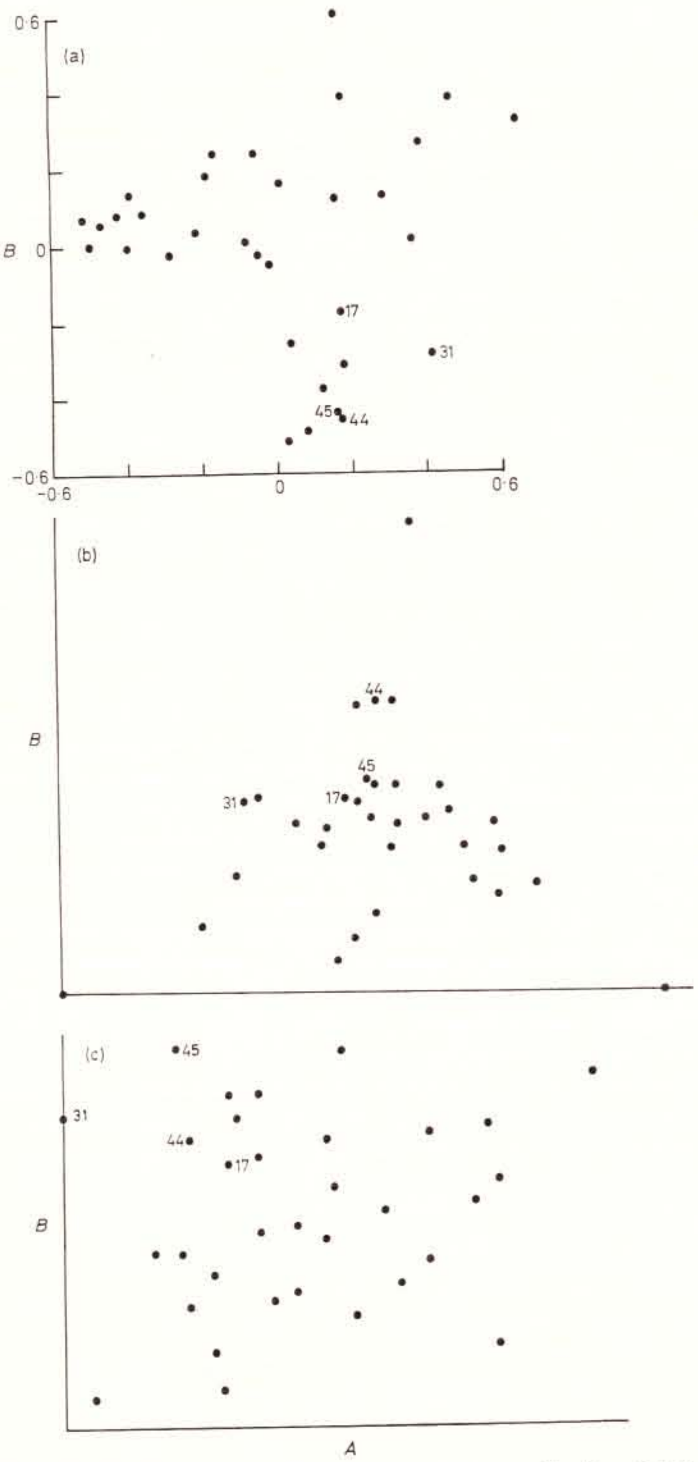

FIG. 5. Comparison of *Carex serotina* distribution in the three ordinations (a)–(c) of Fig. 2. Figures are percentage frequencies of *C. serotina*.

V. DISCUSSION

Bray & Curtis (1957) and Curtis (1959) mention some of the shortcomings of their methods but considered that sampling inadequacies did not make the use of more precise methods worthwhile. It should, however, be obvious that the measurement of interstand distance by an erroneous coefficient and a distorting method of axis construction will, if combined with sampling error, result in further loss of accuracy. The sampling error may well vary markedly from stand to stand but no satisfactory method is available for measuring the overall error in the spatial positioning of stands. Therefore, it is probably better to regard stands as fixed points in space and define their interrelationships by the most accurate measure available. The use of Euclidean distance or similarity coefficient in the manner outlined is to be recommended for this purpose. When a computer is available, these measures are less time-consuming to calculate than the Bray & Curtis measure of interstand similarity.

A comparison of the ordination results reveals a relatively clear separation of the major types by principal components analysis in contrast to the Bray & Curtis ordination, which promotes an appearance of a continuum. This exaggeration of continuity can be attributed to a methodological artifact but does not refute the vegetation continuum concept for which ordination results have been offered as evidence. Principal component analysis also shows continuity but of a more restricted nature.

When Fig. 2(a) is compared with Fig. 2(b) the effect of using the extreme stands as reference points for the ordination axes is apparent. None of the reference stand pairs define axes which lie exactly in the direction of maximal variation, and, as a consequence, the stands tend to cluster in one portion of the scatter diagram. An extreme case can be conceived in which none of the reference axes lies near the direction of maximal variation. The scatter diagram obtained in this case is of little ecological value as it obscures the major ecological gradients as well as the separation of clusters when they are present. Scatter diagrams of this unique case show an undifferentiated cluster of stands in the centre of the diagram with only reference stands at the extremes.

The success of simple ordinations depend on the spatial relationship of the extreme stands to the point cluster. A technique which does not give undue weight to the extreme stands is principal components analysis. The principal axes reflect the total variation in the sample and not that of the extreme stands.

Principal component analysis can provide as many axes as desired. These axes are extracted in a descending order of importance. Simple ordinations, however, are difficult to carry further than three axes and the sequence of extraction depends on the reference stands and not on their relative importance.

ACKNOWLEDGMENTS

We should like to thank Mr P. Greig-Smith for his supervision of this work, and Dr R. P. McIntosh for many helpful discussions, also members of the University College of North Wales Computing Laboratory and Mr R. Bunce, Mr B. Goldsmith and Mr J. Ogden who made their data available for comparison.

SUMMARY

1. The coefficient $C_{max} - C$ is not a Euclidean measure of distance. Its scale of measure varies relative to inter-stand distance, D, depending on the species composition of the

stands compared, and tends to overestimate D in species-poor communities and underestimate it in species-rich types. $C_{max}-C$ should not therefore be used as a substitute for D.

2. Oblique axes when plotted as perpendicular distort the relationships of the stands. This inadequacy is overcome by perpendicular axis construction as described by Orloci (1966).

3. The use of extreme stands as reference points for the axes results in a loss of ordination efficiency. This inadequacy is corrected in the principal component analysis of an appropriate similarity coefficient.

4. A comparison of the three ordinations suggests that the Bray & Curtis ordination exaggerates the appearance of a continuum.

5. The success of simple ordinations is dependent on the position of extreme stands relative to the point cluster. A better method is principal component analysis of an appropriate similarity coefficient.

REFERENCES

Ashton, P. S. (1964). Ecological studies in the mixed dipterocarp forests of Brunei State. *Oxf. For. Mem.* **25.**

Ayyad, M. A. G. & Dix, R. L. (1964). An analysis of a vegetation-microenvironmental complex on prairie slopes in Saskatechewan. *Ecol. Monogr.* **34,** 421–42.

Beals, E. (1960). Forest bird communities in the Apostle Islands of Wisconsin. *Wilson Bull.* **72,** 156–81.

Bray, J. R. & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27,** 325–49.

Curtis, J. T. (1959). *The Vegetation of Wisconsin: an Ordination of Plant Communities.* Madison, Wisconsin.

Dagnelie, P. (1960). Contribution á l'étude des communautés vegetales par l'analyse factorielle. *Bull. Serv. Carte. phytogeogr.* Sér. B. **5,** 7–71, and 93–195.

Dale, M. B. (1964). *The application of multivariate methods to heterogeneous data.* Ph.D. thesis, University of Southampton.

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

Gittins, R. T. (1965b). Multivariate approaches to a limestone grassland community. II. A direct species ordination. *J. Ecol.* **53,** 403–9.

Gittins, R. T. (1965c). Multivariate approaches to a limestone grassland community. III. A comparative study of ordination and association analysis. *J. Ecol.* **53,** 411–25.

Greig-Smith, P. (1964). *Quantitative Plant Ecology,* 2nd edn. London.

Hohn, F. E. (1964). *Elementary Matrix Algebra.* New York.

Lambert, J. M. & Dale, M. B. (1964). The use of statistics in phytosociology. *Adv. ecol. Res.* **2,** 59–99.

Larsen, J. A. (1965). The vegetation of the Ennadai Lake area N.W.T.: Studies in subarctic and arctic bioclimatology. *Ecol. Monogr.* **35,** 37–59.

McIntosh, R. P. & Hurley, R. T. (1964). The spruce-fir forests of the Catskill Mountains. *Ecology,* **45,** 314–26.

Maycock, P. F. & Curtis, J. T. (1960). The phytosociology of boreal conifer-hardwood forests of the Great Lakes region. *Ecol. Monogr.* **30,** 1–35.

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

Sokal, R. R. & Sneath, P. H. A. (1963). *Principles of Numerical Taxonomy.* London.

(*Received* 15 *June* 1965)