



Promoting the Science of Ecology

---

An Examination of Some Ordination Problems By Use of Simulated Vegetational Data

Author(s): J. M. A. Swan

Reviewed work(s):

Source: *Ecology*, Vol. 51, No. 1 (Jan., 1970), pp. 89-102

Published by: [Ecological Society of America](#)

Stable URL: <http://www.jstor.org/stable/1933602>

Accessed: 06/04/2012 13:57

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Ecological Society of America* is collaborating with JSTOR to digitize, preserve and extend access to *Ecology*.

<http://www.jstor.org>

# AN EXAMINATION OF SOME ORDINATION PROBLEMS BY USE OF SIMULATED VEGETATIONAL DATA

J. M. A. SWAN

*Harvard University, Harvard Forest, Petersham, Massachusetts*

(Accepted for publication September 5, 1969)

*Abstract.* Hypothetical vegetation models were made to simulate numerical changes in species populations along a single environmental gradient. A single ordination procedure was evaluated by its ability to detect the ecological information in the hypothetical models. The procedure was reasonably successful when the data were drawn from a short length of the gradient but became progressively less so as longer lengths of the environmental gradient were included in the data. This parallels an increase in the number of stands from which each species is absent in the total data set. Zero values appear to mask ecological information, and an intuitive method of assigning "degree of absence" values to the data is described. After this adjustment, ordination patterns were easier to interpret because ecological information was concentrated in fewer axes.

## INTRODUCTION

Information on the behavior of plant-species populations comes from two main sources: observation of natural patterns and experiment. In a complex ecosystem where much experimentation is impractical, the onus for our plant ecological knowledge often rests with the analysis of vegetational patterns already visible in nature.

There is considerable evidence that a few environmental gradients (often composites of a number of variables) relate to quantitative changes in the population characteristics of many species (e.g. Bakuzis and Hansen 1960, Maycock and Curtis 1960, Loucks 1962, Waring and Major 1964). Thus, the relations among many species populations may be summarized in terms of a few underlying variables. However, the environment is often more difficult to sample than the vegetation and there may, in addition, be some uncertainty about what variables should be related to the population characteristics being studied. In such a case, the vegetation itself may be ordered according to compositional relatedness among samples (vegetational ordination of Austin 1968, indirect gradient analysis of Whittaker 1967). Other attributes of the ecosystem (the physical environment, succession, disturbance) can later be compared with this pattern to search for correlated features.

Principal components factor analysis (Dagnelie 1960, Orloci 1966) and analogous but simpler procedures, such as the Wisconsin methods (Bray and Curtis 1957, Beals 1960, Newsome and Dix 1968), are currently and frequently used techniques of vegetational ordination. However, since no two sets of ecological data are alike, it seems unlikely that vegetational ordination can be equally effective for all sets of data to which it is applied

(Williams and Dale 1962, Dix and Smeins 1967, McIntosh 1967, Whittaker 1967, Swan, Dix, and Wehrhahn 1969). In fact, Greig-Smith (1964) suggests that factor-analysis ordination may be most useful when applied to a narrow range of the vegetational and environmental spectrum available for study.

Attempts to evaluate this type of ordination for the display and interpretation of vegetational patterns have been largely based on analyses of natural data (e.g., Gittins 1965*a, b*, Gimingham, Pritchard, and Cormack 1966, Orloci 1966, Goff and Cottam 1967, Bannister 1968). However, relational patterns among many species in an area are obscured by chance circumstances, by local environmental effects, and by sampling errors. The value of ordination is hard to determine when the true character of the data is only partially known.

There is an alternative and hitherto little explored approach to evaluating ordination. A systematic appraisal of potential difficulties might be made from available studies. Species population characteristics thought to be critical to the success of ordination might then be simulated in vegetational models. When ordinated, the display of species interrelations might be compared with that from which it is derived to determine if the ordination patterns deviate significantly and in what way from the original patterns. As a step in this direction, Austin and Greig-Smith (1968) have recognized the importance of the character of the data matrix in phytosociological analysis. They have also observed differences in ordinations of artificial data previously transformed in various ways.

The specific aim of this paper is to use simulated data to indicate ways in which the interlocking trends of species behavior along gradients are de-

picted by vegetational ordination in metric space (term after Orloci 1968b).

#### THE PROBLEM

Some uncertainties about the function of ordination can be traced to the matrix construction methods that precede it. Based on available studies, an appraisal of some of these difficulties is possible.

For Wisconsin ordination (Bray and Curtis 1957, Beals 1960, Swan, Dix, and Wehrhahn 1969) the difference between the quantitative values for a species in two stands is a single test of compositional relationship between the stands. An ecological distance between these stands is calculated from the quantitative differences for all species. If these distances have ecological meaning, they should increase as the species composition and environment of two stands become increasingly different. Yet there are two characteristics of phytosociological data that may prevent this.

First, Lambert and Dale (1964) point out that quantitative data are truncated; numbers can be recorded to represent the degree of presence for a species, but there is no corresponding measure of the degree of absence for an unrecorded species. Hence, a species may contribute to the calculation of ecological distance in only a limited part of the data set since it is elsewhere represented by zeros between which no distance can be calculated. In fact, when  $2w/a + b$  similarities are calculated among samples distributed along a gradient, they drop to zero when the distance along the gradient has increased to the point where no two species are shared by the samples compared (Whittaker 1967). Zeros occur abundantly in boreal forest (Swan and Dix 1966) and prairie data (Dix and Smeins 1967), and all vegetational data are semi-quantitative in some degree. The effects of zero values on distance calculations may have been underestimated.

Second, species having their peak values at the ends of a sampled gradient are represented by "S-shaped" curves (values decrease in one direction only), while species with peaks progressively closer to the center of the gradient have increasingly evident and more complete "bell-shaped" curves (values decrease in both directions from the peak) until, about the center, the entire gradient distribution of a species may be encompassed by the measured gradient. For convenience, the terms S-shaped and bell-shaped will be used to categorize species-response curves throughout the paper. If stands have the same species values on opposite sides of a bell-shaped response curve, the stands are at different points in ecological space,

yet the species cannot contribute to a calculated ecological distance between the stands. In such a situation only S-shaped curves permit the calculation of interstand distances that increase as the population changes.

Now, as the range of gradient included in a study increases, each species is expected to occupy a progressively smaller part of the total gradient length. Thus, as samples are taken from a larger range of gradient, a larger number of species will probably have their modes between the extremes of the gradient as sampled. The proportion of species with S-shaped slopes decreases in the total set of species, while the proportion with bell-shaped curves increases (Whittaker 1967). Thus, for any data set, interstand distances are less likely to increase in proportion to population changes as the length of the sampled gradient increases, because this increase is accompanied by an increase in the number of bell-shaped curves. Since the bell-shaped curve has a sequel in  $n$ -dimensional space (the  $n$ -dimensional binomial solid, Whittaker 1967), the same problem is expected in the construction of multidimensional models.

It is reasonable to expect that stands known to occur in distinctly different habitats may not be separated by ecological distances commensurate with their environmental differences. In fact, for a set of grassland data from Nelson County, North Dakota, Dix and Smeins (1967) note that species were obviously disposed along a moisture gradient, yet calculated distances between the vegetations of stands did not increase in proportion to increasing distance along the gradient.

For factor-analysis ordination, coefficients of correlation are usually preferred (Whittaker 1967) to summarize population relations in a matrix. Here too there are difficulties. Data containing bell-shaped curves may have unfavorable implications for factor analysis because, as for the ordination of distances, the method is theoretically based on the assumption of linearity (Whittaker 1967). Also, species for which quantitative values occur in limited and different parts of the total data set, may appear to be uncorrelated in their behavior although they both respond to the same environmental gradient. Thus, in as much as they influence correlation and variance-covariance matrices, zero values and bell-shaped curves may prevent ordination by factor analysis from clearly illustrating population patterns.

Whittaker (1967) notes that the illustration of vegetational patterns by ordination may be regarded as a sufficient aim for the procedure. However, it seems that a major objective of ordination should be the illustration of population patterns in a form that can be related to vegetational processes

or other features of the ecosystem. To relate ordination patterns to changes in environmental variables, we must know how the environmental trends are to be recognized in the ordination diagram. It is tacitly acknowledged that spatial distances between stands should increase as their ecological relationships become more remote. Yet, species behavior patterns may obstruct such a relationship. Perhaps, in some cases, vegetational ordination may not provide a simplified model of species population patterns that will also have interpretive value.

To examine this problem I first tried to assess how ecological distances in an ordination reflected compositional differences in sets of artificial data. Then, because correspondence was frequently poor, ways were sought in which ordination patterns might be made to provide a clearer representation of vegetational interrelationships. The two analyses are treated separately since specific methods and results pertain to each.

#### ECOLOGICAL DISTANCE AS A MEASURE OF COMPOSITIONAL DIFFERENCE

##### *Methods*

*Vegetational models.*—Because of natural complexity, models constructed to represent natural data are likely to approximate reality most closely if they are multidimensional. However, unidimensional models include, in simplified form, many species population characteristics of multidimensional models. Thus, where overlapping population curves are arranged along a gradient, each species is located in a part only of the total gradient. This simulates the complex series of interlocking population patterns in multidimensional space. Zero values and bell-shaped curves along a single gradient both have a sequel in multidimensional patterns, and both are thought to cause problems in the construction of complex abstract models by ordination. Ordinations of unidimensional models may provide some insight into the limitations of ordination for more complex vegetations. Such models offer the simplest possible sets of test data and were therefore chosen for this study.

Bell-shaped curves and zero values are critical characteristics that may prevent distances between samples in an ordination from increasing in linear proportion to species population changes. If possible, the way these characteristics vary among different sets of natural data should be simulated.

Depending on the scope of a study, natural data vary from the almost entirely quantitative, where a limited range of vegetational composition and environmental variation is involved (considerable

overlap in population patterns among species), to a state in which the vegetation and environment are so disparate between samples that much information may be obtained from a purely qualitative analysis (little overlap in population patterns among species). All data to be ordinated fall somewhere within this spectrum. A broadening of the spectrum of vegetational composition associated with an increase in the sampled length of a gradient involves an increase in zero values and in the ratio of bell-shaped to S-shaped curves. Here, the aims of simulation can be narrowed to the construction of a series of single gradient models representing data that vary from the almost entirely quantitative to the highly qualitative with concomitant changes in the two critical characteristics. Next, the way in which these critical population characteristics are to be included in the models must be decided.

In natural data it is expected that species will vary in their ubiquity across a gradient as well as in the maximum quantitative value they achieve. Also the nature of the response curve—sometimes uniformly bell-shaped, sometimes U-shaped, often skewed, sometimes so poorly developed as to be unrecognizable—will vary among species within a single data set. The possible combinations of such features vary from one data set to the next. Since the study is a preliminary attempt to evaluate the effects of zero values and bell-shaped curves on ordination, it seemed inadvisable to introduce these sources of variability at this stage. Hence, I decided that for each model, all species should have identical bell-shaped response curves with the same maximum value and potential range across a gradient and that the peaks of these curves should be located at equidistant points along the gradient.

Zero values and bell-shaped curves both change in number simultaneously as the gradient length increases. To separate their effects for independent appraisal, the models are built in two stages. First, five models are made, each with the same number of species but differing from each other in the total range of gradient occupied by the species curves. In this way, five sets of data with different degrees of species overlap along the gradient (and consequently different numbers of zero values) can be derived and their effects on ordination examined. Second, variations in the ratio of bell-shaped to S-shaped curves can be added to these models so that the additional effect on ordination of changes in this ratio can be observed. The details of model construction are described with Fig. 1 to illustrate the process.

A series of S-shaped species population curves are drawn to extend different distances along a

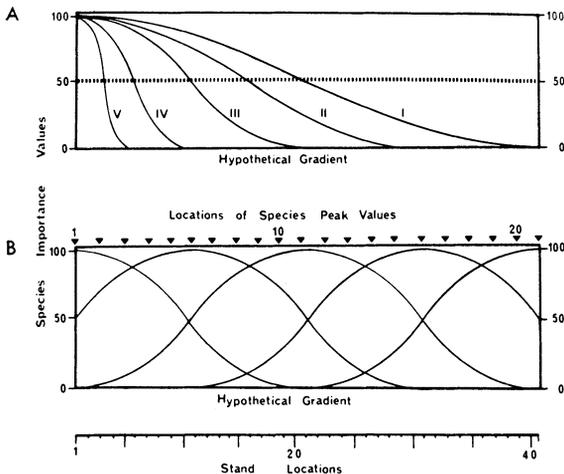


FIG. 1. Constructing models of plant species-habitat relations. The five S-shaped curves in Fig. 1A and their mirror images were each used to construct a model of the kind illustrated in Fig. 1B, derived from curve III. Here only five of the 21 species are shown in place. Peak locations for the other species curves are indicated by arrows above the diagram. All other models were constructed in the same manner.

gradient as in Fig. 1A. In this diagram the abscissa is the hypothetical gradient and the ordinate, a scale of importance values between zero and 100. Curve I extends the complete length of the gradient, while curves II, III, IV, and V extend  $\frac{3}{4}$ ,  $\frac{1}{2}$ ,  $\frac{1}{4}$ , and  $\frac{1}{8}$  of the gradient length, respectively. For each importance-value (species population) curve the upper and lower portions (above and below 50) have the same form, changing more rapidly about 50 and less so approaching the extremes of 100 and zero.

Second, five models (I to V) were constructed, each based on one only of the species curves, after the plan illustrated in Fig. 1B. This diagram illustrates the construction of the model based on curve III. All other models were developed in the same way. As before, the hypothetical gradient is established on the abscissa while the ordinate is a species importance scale between zero and 100. A decision had to be made about the number of species curves to be included. There were no a priori grounds for selecting a particular number of species. It seemed reasonable to choose a number large enough that it might fall within the range of the number of species commonly analyzed in natural data, yet small enough that it would not offer excessive computational problems in later analysis. Twenty-one equidistant points were arbitrarily chosen to represent the peaks of 21 species curves along each gradient. For all 21 species these points are marked by solid, inverted triangles along the top of the diagram in Fig. 1B. The S-shaped curve III in Fig. 1A and its mirror image

were combined to produce a symmetrical bell-shaped curve for each species. For simplicity, the completed curves for five only of the 21 species in model III are illustrated in Fig. 1B.

Next, I simulated the increase in bell-shaped relative to S-shaped curves that occurs as sampled data are drawn from a progressively wider range of the available spectrum. Of the available models, I and V represent the narrowest and broadest limits of the spectrum respectively. When species populations overlap widely (model I), a larger number of species with their modes outside a gradient segment can enter the segment as S-shaped curves than when there is little species overlap along the gradient (model V). If such species are added, the ratio of bell-shaped to S-curves will be lower for model I than for model V. This accords with Whittaker's statement (1967) that the proportion of species with S-shaped slopes decreases in the total set of species, while the proportion with bell-shaped curves increases, as samples are taken from a larger range of gradient.

For models I and V the ratio of bell-shaped to S-shaped curves was altered by extending the gradient beyond the measured portion; by drawing curves for species peaking along the added segment at the same intervals as those within the measured gradient; and by recording the portions of these curves that entered the measured gradient. For model I, 38 species slopes were added in this way (19 at each end of the gradient). For model V only four were needed to complete the sequence (two species slopes at each end of the gradient). The additional models are termed I' and V', respectively.

Samples taken from very narrow ranges of gradient may have a lower ratio of bell-shaped to S-shaped curves than represented in model I'. The present range has been chosen for this initial study because many natural data so far examined seem to include many species with modes within the sampled composition.

While the models do not represent the array of complexity found in some natural data, they do incorporate two key population characteristics thought to interfere with a clear display of relational patterns by ordination.

*Sampling and analytical procedures.*—In studies of natural vegetation, a sample or stand is an arbitrarily bounded area within which the vegetation and habitat are sufficiently uniform that the stand may be used as a vegetational unit to be compared with others (similarly bounded) in the search for relational patterns in populations. As many stands are sampled as the observer considers necessary to cover the variety in the compositional attributes

he is studying. This procedure has been employed to sample the hypothetical vegetation models.

Forty-one stands were used to sample each model. This number was arbitrarily judged to be large enough to permit an adequate sample of vegetational patterns yet small enough to avoid excessive computation in subsequent analysis. To sample each model, the stands were placed at equidistant intervals along each gradient (Fig. 1B), and the species values were read from the points on the appropriate curves directly above the stand locations. For each model, importance-value data for all species were summarized in a table. The overlap of species populations decreases in models I through V causing the number of species per stand to decrease correspondingly, from 21 to three. A small number of species per stand is not regarded as a model weakness, because ordinations have sometimes been applied to natural data with few species per stand.

The model data are to be used for two specific purposes: 1) to show how the calculation of a matrix of compositional relationships varies with differences in the critical characteristics of the models; and 2) to determine how this variation affects the representation of the vegetational models by abstract ordination models.

A multiplicity of vegetational ordination procedures are available. Bell-shaped curves and zero values are thought to adversely affect matrix-construction procedures for any ordination method. Ideas derived from the test of a single matrix-construction and ordination procedure on the models should suggest difficulties with and ways to evaluate the other procedures.

A relatively simple and mathematically precise method for the ordination of distances is available (Swan, Dix, and Wehrhahn 1969). It avoids the distant stand criterion and the oblique axes of the Wisconsin method. The euclidean distance index (Sokal and Sneath 1963, Orloci 1966) is appropriate for the matrix calculation preceding ordination by this technique. Both stand and species ordinations (the latter introduced by Gittins 1965*b*) can be accomplished by this combination of matrix construction and ordination. Stand ordination has been most widely used to date and was adopted here.

For each set of data a matrix of compositional relationships between stands was constructed using the euclidean index:

$$d_{ij} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ni} - X_{nj})^2}$$

where  $i$  and  $j$  are stands with species scores  $(X_{1i}, \dots, X_{ni})$ ,  $(X_{1j}, \dots, X_{nj})$  and  $d_{ij}$  is the calculated distance between stands  $i$  and  $j$ . Each

matrix was ordinated by the procedure of Swan, Dix, and Wehrhahn. All calculations were done on an IBM 3600 digital computer.

### Results

The results are presented in three parts. These treat the effects on distance calculation and ordination of (1) differently overlapping species populations along a gradient and (2) different ratios of bell-shaped to S-shaped curves along a gradient. The third part draws together the conclusions of these analyses.

*The effects of differently overlapping species populations.*—Ordination patterns can approximate the patterns in a model only if distances between stands calculated from vegetational characteristics match distances measured between the same stands on a gradient. In experimental terms, an increase in the number of intervals between stands along the gradient should be paralleled by an increase in the calculated distances between them. For each model there are 1,640 calculated matrix distances. However, the distances in each matrix diagonal are calculated between stands separated by the same interval, and each diagonal represents a different interval. The average diagonal values conveniently summarize the relations between all possible gradient intervals and corresponding calculated distances. For each matrix there are 40 such comparisons, the points of which describe a curve when plotted on a graph with increasing intervals along the abscissa and the corresponding calculated distances on the ordinate. In all cases the points described a curve so closely that, for simplicity, only the distance-interval curves from all models were summarized in Fig. 2.

The number of quantitative entries in the original data expressed as a percentage of the total number of entries gives, for each model, some idea of the range of gradient occupied by quantitative values for each species (degree of species overlap along a gradient). Through models I to V these values decrease as: 99.8, 91.6, 71.9, 40.8, and 20.6% of the number of entries in each data table.

If calculated and gradient distances exactly correspond, a linear plot should be obtained for each model. As expected, this is nowhere obtained. The curve for model I (99.8% quantitative values) most closely approximates a line, but gradually deviates from it as the greatest number of intervals between stands on the original gradient is approached. Curves for models II (91.6%) to V (20.6%) are progressively further from a straight line. Indeed, in models IV (40.8%) and V there is an abrupt change in the curve beyond which there are no further increases in calculated

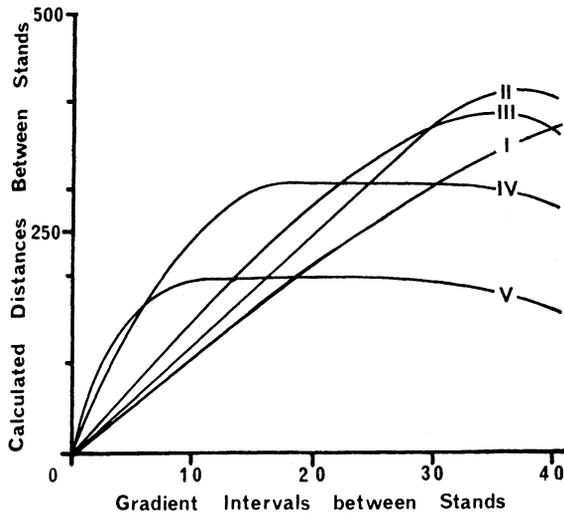


FIG. 2. The effects on distance calculations of differences in species behavior along hypothetical gradients. Distances were calculated from vegetational differences between stands (ordinate). Corresponding interstand distances were measured along hypothetical gradients (abscissa). Each curve is a comparison of calculated and measured distances for one hypothetical model of species responses along a gradient. Details in text.

distances for increasing gradient intervals. Finally, for models II to V, the calculated distances actually decrease as the maximum number of gradient intervals between stands on the gradient is approached. Clearly, a decrease in population overlap is accompanied by a progressively poorer representation of population changes through distance calculation.

Ordination arranges stands in space according to their calculated interstand distances, and the foregoing results indicate that these distances do not always increase in proportion to changes in species populations. This must surely affect the spatial disposition of stands derived by ordination. For example, a stand at an intermediate location on the gradient may be separated from stands at both ends of the gradient by ecological distances together greater than the calculated distance directly measured between the end stands. In this event a single axis cannot represent the distances between the three stands, but a triangle with the distances as sides might describe their relations. This is a two-dimensional pattern suggesting that the original gradient may be represented in more than one dimension by ordination and that a multidimensional ordination might be appropriate for the analysis. A three-dimensional ordination was applied to each data matrix.

Deviations from a linear stand order in the ordinations of models I to V can reasonably be attributed to the manner in which calculated distances reflect ecological differences. Yet, I wished

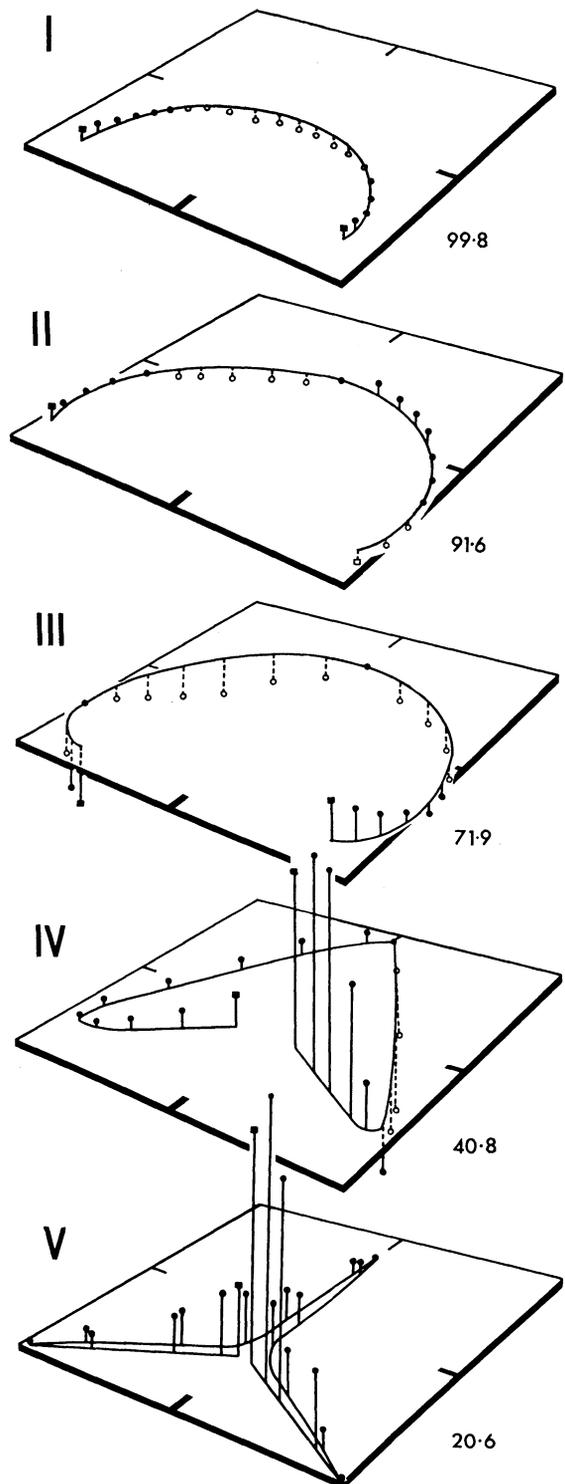


FIG. 3. The effects on vegetational ordination of differences in species behavior along hypothetical gradients. Each diagram is an ordination of one hypothetical model of species response to a gradient. Models I through V have progressively fewer species ranging widely across a gradient. To represent this change, the number of species values above zero in each data set are recorded as a percentage beside the appropriate diagram.

to ensure that such deviations could not be due to the ordination procedure itself. Forty-one stands (the number used in all models of the present study) were ordered at equidistant points on a line, and the exact distances between each and every stand were measured and recorded in a matrix. Upon ordination, a linear order of stands was obtained duplicating that of the original.

The ordinations of models I to V are illustrated in Fig. 3. For each ordination, stand distributions on the X, Y, and Z axes are summarized in a perspective diagram. The X and Y axes form a plane. Its front edge is the X axis while the Y axis is at the side. The Z axis elevates and depresses stands above and below the plane. To simplify the diagrams, alternate stands have been omitted (ordination patterns are unchanged by this omission) as well as axis scales (all axes are on the same relative scale for each ordination). Stands located above the plane are represented by solid circles and are linked to the plane by vertical lines, while those below the plane are illustrated as open circles and are linked to the plane by dashed lines. Terminal stands on the hypothetical gradients (stands 1 and 41) are marked with a square. For each ordination, a line is drawn to interconnect the locations of all stands on the plane by linking together those that are adjacent in the hypothetical model. In this way, the order of stands in the ordination diagrams can be visually compared with their linear order in the hypothetical models. For each model the percentage of quantitative values in the data set is recorded beside the appropriate ordination diagram.

For model I, stand order along the X axis is identical with that of the original gradient but it is distorted into two dimensions. In successive ordinations (I through V) the X-axis order is progressively less like that of the original models, and the distortion into two dimensions becomes more pronounced and ultimately extends significantly into three dimensions (models III, IV, and V). Hence, with decreasing overlap of species populations the ends of the gradient are drawn together by ordination and the whole gradient becomes progressively harder to detect.

This loss of visual value for ordination is accompanied by a progressively poorer representation, through models I to V, of the distance values in a matrix. A simple test of this is the sum of squares of stand positions on an axis expressed as a percentage of the total matrix sum of squares (Swan, Dix, and Wehrhahn 1969). These percentage extraction values are recorded for each ordination in Table 1.

The ordination technique is designed to extract axes in order of decreasing importance. A 100%

extraction value for the X axis would represent a completely effective ordination. In fact, the highest extraction value achieved for an X axis is 85% for the ordination of model I. The amount of matrix information extracted by the X axis decreases and the amount taken out by subsequent axes (Y and Z) increases from model I to V. In model V each of the three axes extract approximately the same amount of information from the matrix. This change is also paralleled by a decrease in the total percentage extraction for all three axes. In model V less than half of the matrix sum of squares is extracted by the ordination. A complete extraction of the matrix for this model would require many more axes than those presented here. As species populations overlap less and less on a gradient, there appears to be a loss or "drift" of information from the first to subsequent axes tending towards an even distribution of information among many axes. A low and even distribution of information among many axes might be expected where there is no gradient with which the performances of many species are correlated, a situation where ordination would be of little use. This seeming lack of species-gradient relationships is obtained from models each having a single linear gradient along which all species populations change in sequence.

TABLE 1. The percentage of matrix distances extracted by three-dimensional stand ordinations, for five hypothetical models of vegetational responses to an environmental gradient (The details of model construction and ordination are explained in the text.)

Model	% Quantitative values in data	X	Y	Z	Total
I	99.8	84.6	15.3	0.1	100
II	91.6	81.0	18.5	0.2	100
III	71.9	68.3	27.0	2.8	98
IV	40.8	32.7	22.2	16.1	71
V	20.6	13.2	13.1	10.9	37

*The effects of different ratios of bell-shaped to S-shaped curves.*—Models I' and V' differ from models I and V since they both have a lower bell-shaped/S-shaped curve ratio. To show the effect of the changed ratios on matrix construction, calculated distances between stands along the gradient are compared with the corresponding gradient intervals for all four models (Fig. 4). This figure was prepared in the same way as Fig. 2.

The curves for model I and I' differ little. The 38 added S-shaped species curves in model I' increase the short as well as the long distances. This may occur because the added curves for species peaking beyond the measured segment of the gradient can extend only part way across the mea-

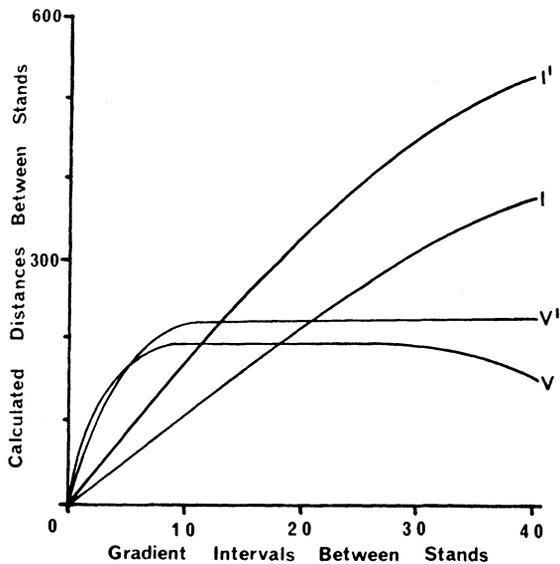


FIG. 4. The effects on distance calculations of differences in the ratio of bell-shaped to S-shaped species curves along hypothetical gradients. Distances were calculated from vegetational differences between stands (ordinate). Corresponding interstand distances were measured along hypothetical gradients (abscissa). Each curve in the diagram is a comparison of calculated and measured distances for one hypothetical model of species responses along a gradient.

sured segment. Zeros are introduced into the data between stands falling beyond the response curves of the added species, and these cannot contribute to distance calculation. Models V and V' have identical curves except for the drop at the end of the curve for V which is not observed for V'. Since the sole difference between these two models is the occurrence of four extra S-shaped species curves in model V' (two at each end of the gradient), the fall in the curve for model V is thought to result from the exclusion from the sampled gradient of species peaking beyond it that yet have a part of their response curves within the gradient. Similar falls in the curves for models II, III, and IV (Fig. 2) may perhaps be explained in these terms. For model IV the sequence of species curves within the gradient was extended beyond its confines, and S-shaped curves were added for species entering the gradient but with their modes beyond it. Eight species were added and the distance-interval curve for this model was compared with that for model IV and found to be identical except that it too remained horizontal at the extreme values—an analogous result to that derived from comparing V and V'. Thus, if the number of S-shaped species curves is less than might be expected from the degree of overlap of species along the gradient, distances calculated between stands separated by considerable distance

on the gradient of the original model may be less than those calculated between somewhat more adjacent stands.

Ordinations derived from models I' and V' were almost identical with those for the original models I and V and, for this reason, are not illustrated. Indeed this is to be expected since the distance matrices for I and I' and V and V' are much the same in character (Fig. 4). However, the ordinated stand positions for model I' were not quite as "bowed" as in model I, especially at ends of the gradient. This slight difference is reflected in the percentage extraction values which were 86.7, 12.5, and 0.5 for X, Y, and Z axes of model I' compared with 84.6, 15.3, and 0.4 for the X, Y, and Z axes of model I. A slight increase in the information extracted by the X axis of model I' is balanced by a corresponding decrease for the second axis. Models V' and V were practically identical in ordination patterns. Their percentage extraction values were 13.4, 13.0, 12.2 and 13.2, 13.1, 10.9, respectively.

After changes in the ratio of bell-shaped to S-shaped curves, distance-interval curves and ordinations are much the same as before. Changes in the overlap of species along gradients appear to be chiefly responsible for differences in ordination.

*Concluding statement.*—All these experiments have been conducted with the same basic number of species (21). Yet, the number of species in natural data varies with floral richness and the range of the compositional spectrum covered by a study. We may inquire if the ordination patterns described here can be expected where the number of species in the models is considerably different from that adopted. Expense and time preclude a detailed treatment of this question. To obtain some information about it, the same ordination procedure was applied to four linear models having the same species curves (Fig. 1A) as models I to IV but with five species only in each data set. For convenience, nine stands were used to sample each gradient in this preliminary trial. A model for species curve V could not be sampled with this number of stands and species because sufficient species had been removed from the gradient that the remaining curves did not overlap. Consequently some stands contained no species at all. The percentage of X-axis extractions for the new models I through IV were: 84.9, 76.9, 59.7, and 31.5 respectively, while the corresponding total (three axes) extractions for these models were: 100, 99, 97, and 80 respectively. These results are closely similar to those for corresponding models with 41 stands and 21 species. Decreasing the overlap of species populations along a gradient seems to have the same effect on the extraction of

matrix information by ordination, when there are few species in the data.

From these results the following conclusions can be made. As the data of a study are drawn from a progressively wider vegetational and environmental range, the illustration of species-gradient relations by ordination is obscured. This loss of efficacy is not due to a shift in the ratio of bell-shaped to S-shaped curves. In part, it may be due to a change in the range of overlap of species populations across the original gradient and associated changes in the zero-value complement of the data. The next section searches for ways to improve the representation of ecological relationships by calculated distances between stands.

MODIFYING ECOLOGICAL DISTANCES BY ADJUSTMENTS FOR SPECIES ABSENCE

Order in species population patterns might be regarded as a product of a feedback dialogue, continuous in time and space, between the genetics of the population and the nature of the external influences impinging on the species. This feedback relationship is presumably as active where a species is not quantitatively recorded as where it is. Thus, where a species does not occur, its absence may be a matter of chance only, the habitat being favorable for its occurrence. Alternatively, site conditions may be entirely unfavorable for the occurrence of the species. Since each unrecorded species is equally absent in the data, the zeros hide this part of the dialogue between species and habitat (Lambert and Dale 1964).

If appropriate adjustments can be made for the zeros in the data of models I through V, ordination patterns might correspond more closely with those of the models. In fact, the need for adjustment probably increases from models I through V since progressively more zeros are contained in the data through this sequence of models. The influence of zero values on ordination might be indicated by differences in ordination patterns derived from adjusted and unaltered data.

With these arguments as a conceptual framework, a simple method of adjusting data for species absence is described. The effect of the adjustments on distance calculation and ordination is then evaluated on models I to V.

Methods

A simple set of data helps to describe the method (Table 2). A sequence of five stands (1-5) samples a linear gradient along which five species (V to Z) peak at equidistant intervals from one end of the gradient to the other. Each species is quantitatively recorded (values between zero and 100) in some but not all of the five stands. The remain-

TABLE 2. Hypothetical numerical data for five species in five stands spaced at equidistant intervals along a hypothetical environmental gradient

Species	Stands				
	1	2	3	4	5
V.....	100	50	0	0	0
W.....	50	100	50	0	0
X.....	0	50	100	50	0
Y.....	0	0	50	100	50
Z.....	0	0	0	50	100

ing values are zeros. The zero-masked response of the species population might be empirically visualized on a quantitative "degree of absence" scale with the lowest degree of absence at zero and progressively greater degrees of absence as the site becomes less suitable. Because species V achieves its peak performance in stand 1 of the gradient and stands 2-5 are ordered at increasing distances along the gradient from this point, the degree of absence for species V in stand 3 is obviously lower than in stand 4 which in turn is lower than for stand 5. Similar observations can be made for the other species.

Now, species frequently found together are thought to share similar ecological requirements. Thus, the degree of absence for a species with a zero record in a stand might be assessed through its relations with others in the data set. If an absent species in a stand is usually found with the species complement of that stand in the total data set, then its degree of absence is low for the stand. On the other hand, if the absent species is rarely found with the recorded species complement, its degree of absence is high. To assess degrees of absence in this manner, Dice's (1945) reciprocal association indices are calculated between each and every species as:

$$\text{Association index } B/A = h/a$$

and the reciprocal:

$$\text{Association index } A/B = h/b$$

where *a* is the number of samples in which species *A* occurs, *b* is the number of samples in which species *B* occurs, and *h* is the number of samples in which both species *A* and *B* occur together. Both indices must be calculated since each species may serve as a reference in a number of stands (see below). The reference species is always the one for which a quantitative value is recorded in a stand, and the species being compared is always the one with a zero value.

By means of Dice's indices a sample calculation can be made from the data in Table 2. Species X has a zero value in stand 1 and species V and W represent the vegetational character of this stand.

We wish to find the association of species X with both V and W in all stands. According to Dice's index values, species X occurs in 50% of the stands containing species V and in 67% of the stands containing species W. The average Dice's index value of species X with V and W is 59% among all stands. This value can be used to represent the degree of absence for species X in stand 1. Theoretically, it falls on a scale ranging from 100, where the species in the comparison are always associated (no degree of absence), to zero (the highest degree of absence possible on this scale). The scale has a maximum possible value equivalent to the lowest possible quantitative value recorded in the data. Hence, 100 is added to the scores for species V and W in stand 1 so that they fall on a scale commencing at the uppermost limit of the derived scale. Now the average Dice's index value of species X with V and W is inserted in place of the zero for species X in this stand.

Likewise species Y has a zero in stand 1. It has an average association of 17% with species V and W, and this number is inserted in place of the zero for species Y in the stand. Species Z also has a zero in stand 1 but is not associated with either V or W in the data. Its degree of absence is clearly much higher than for species X or Y. This too can be represented numerically. Species Z is associated with X and Y which are now quantitatively represented in stand 1. Its average Dice's index value with these two species is 50%. This value replaces the zero for species Z and all other values in the stand are raised by 100 since the new rating falls on a scale with a maximum value equivalent to the lowest possible quantitative value of the previously derived scale. Likewise, the data in all other stands are completely quantified (Table 3).

The number of zero-to-100 scales added in this way may vary from stand to stand because all zeros may be replaced by the first comparison (e.g., species X and Y) or further comparisons may be required (e.g., species Z). To permit the direct calculation of distances from species values, the data may have to be adjusted so that the same number of zero-to-100 scales are added to all stand data.

After transformation, values for species V show a decreasing trend across the entire gradient from stand 1 to stand 5. This contrasts with the untransformed data (Table 2) in which species V has zeros in stands 3, 4, and 5, although the likelihood of finding it continues to decrease across these stands. After transformation, the four remaining species (W to Z) also have ordered trends across the entire gradient (Table 3).

TABLE 3. Absence-adjusted numerical data for five species in five stands spaced at equidistant intervals along a hypothetical gradient (details are in the text)

Species	Stands				
	1	2	3	4	5
V.....	300	250	133	111	50
W.....	250	300	250	133	117
X.....	159	250	300	250	159
Y.....	117	133	250	300	250
Z.....	50	111	133	250	300

Calculations were done on a CDC 3600 digital computer. Distance matrices were calculated for each set of transformed data (models I to V) and ordinated as before.

### Results

The adjustment for absence is expected to make calculated ecological distances more sensitive to changes in population composition. Thus, the effect of transformation on distance calculation is treated first, followed by an examination of the ordination patterns derived from transformed data.

*Distance calculations.*—If calculated distances increase in proportion to increasing distance along the gradient, a linear relationship should exist between the two. Distance-interval curves do not have this form for any model (Fig. 5). A progressive increase in calculated distances parallels an increasing number of gradient intervals for all models, although these distances gradually fall below a straight line as the number of intervals between stands increases. In Fig. 2 and 5 the distance-interval curves for model I are much the same for transformed and untransformed data, and the representation of gradient intervals by calculated distance steadily improves through models II to V after transformation.

Finally, distance-interval curves fall terminally in models II to V (Fig. 5). As for untransformed data, a brief experiment on models IV and V showed that this was a characteristic of the models used in the tests: a result of fewer species quantitatively recorded towards the ends of a gradient than towards its center.

*Ordination.*—The improved representation of gradient intervals by distances calculated from transformed data suggests that ordinations of transformed data might provide a clearer pictorial representation of species-gradient relations than direct ordination of the same data. To test this, the matrices of transformed data for all models (I to V) were ordinated and compared with corresponding ordinations of untransformed data. Ordinations for transformed data are illustrated in Fig. 6.

There is no detectable difference between the ordinations of complete and partial data for model I. Thus, where the data are almost completely quantitative (98%), the ordination pattern is un-

changed by transformation. For models II to V transformed data yield ordinations all of which are much the same—a semicircular pattern of stands

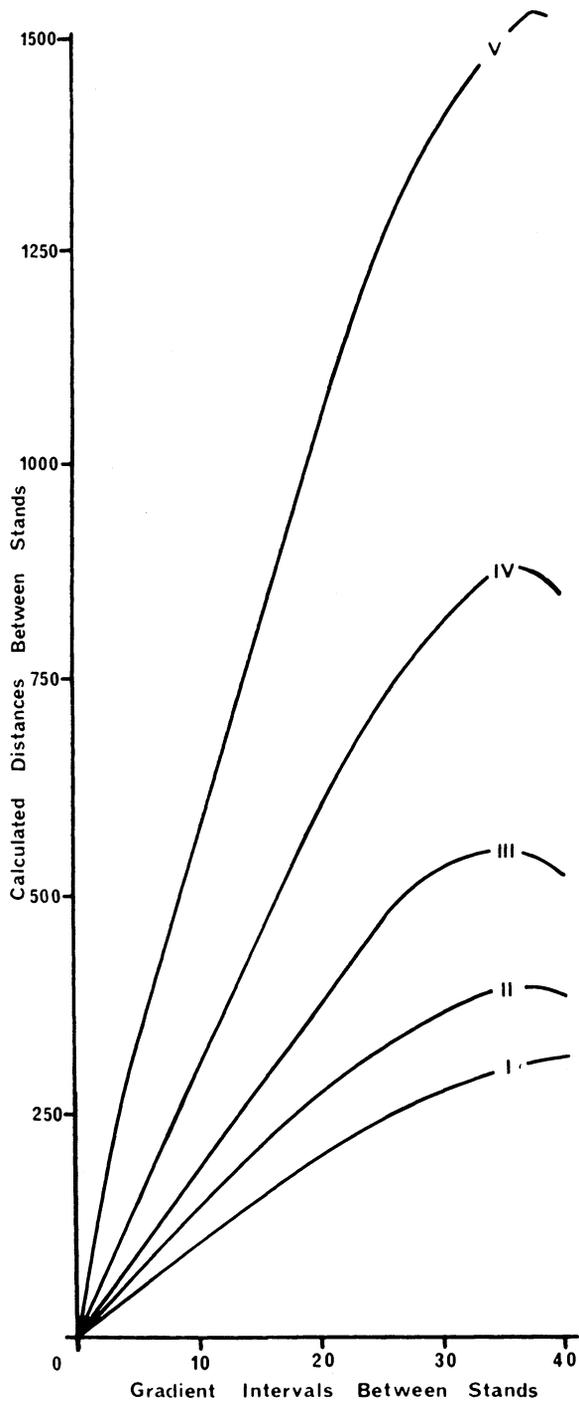


FIG. 5. The effects on distance calculations of adjustments for zero values in hypothetical data. Distances were calculated from vegetational differences between stands after adjustment for zeros (see text). Corresponding interstand distances were measured along the hypothetical gradient (abscissa).

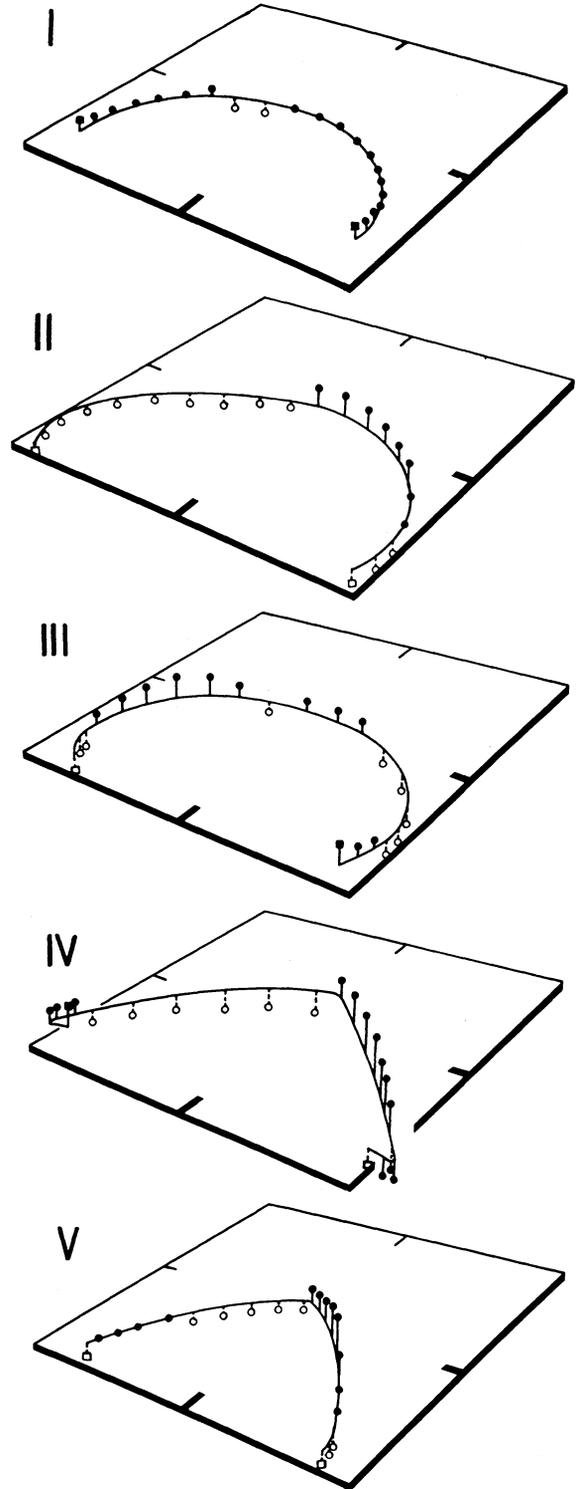


FIG. 6. The effects on ordination of adjustments for zeros in hypothetical data. Each diagram is an ordination of adjusted data for one hypothetical model of species responses along a gradient. Details in text.

with little stand separation in the third dimension. Thus, in models II to V, adjustments for species absence progressively improve gradient detection as the zero-value complement of the data increases. In all cases, the stand order of the original data is approximated by the X axis. Thus, a stand order that may be undetectable in untransformed data is retrieved for X-axis detection after transformation.

For untransformed data, the per cent extraction value for three axes decreases progressively from 100% for model I to 37% for model V. After transformation, the most information extracted (model I) is 100% as before but the least (model V) is 90%. Thus, through models I to V, the progressively improved representation of model patterns by ordinations of transformed data is paralleled by the fact that much information from the unidimensional model enters each matrix in a form now available for ordination in three or fewer dimensions. Before transformation, this information could only be ordinated in a progressively more complex multidimensional space as the data became increasingly qualitative.

All ordination patterns in Fig. 6 are similar. Distance-interval curves of transformed data for each model were also similar when expressed on the same relative scale. The representation of a linear gradient as a "bowed" pattern is probably due to the effect of bell-shaped curves on distance calculation, since zeros have been removed from the data.

Because of the similarities of distance-interval curves for all transformed data, a single curve can be drawn to illustrate the representation of compositional difference by ecological distance after transformation. The values on this curve each have a corresponding value on the straight line connecting the ends of the curve. By using these in place of the original values linear ordinations of models I and V (simulating the two extremes of variability included in the models) were obtained. However, this matter is not taken further here. The substitution creates specific ordination problems because it makes euclidean distances non-euclidean. In any case, distortion associated with bell-shaped curves might be reduced by the use of species in place of stand ordination, because distances are here calculated between species and not within binomially distributed populations.

## DISCUSSION

### *General comments*

A clear knowledge of the way ordination represents compositional relations as spatial patterns should aid an appraisal of the ecological meaning

of the patterns; this study is an initial assay of the problem.

Although much ecological information may be represented along a single gradient (Dix and Smeins 1967), natural data are often thought to require a multidimensional display. The implications of this study are probably limited by the simplicity of the models. Nevertheless, the function of ordination on natural data seems to be clarified by its function on simple sets of artificial data.

The tests support the predictions of Whittaker (1967) and Greig-Smith (1964) that the value of factor analysis is increasingly limited as the range of environments and communities represented in the sample set is increased.

For linear models that have a considerable overlap in species populations, the gradient is approximately represented along the ordination X axis. Here, bell-shaped curves are probably responsible for the representation of a linear stand order as a curvilinear pattern, but the distortion is not sufficient to obscure the gradient. Bell-shaped curves are represented by binomial solids in  $n$ -dimensional space. Hence, some distortion may be predicted in more complex models. In fact, curvilinear dispositions of samples have been found in natural data. Using phytosociological ordination, Beals and Cottam (1960) obtained a curvilinear relationship between forest plant-species distribution and moisture in the Apostle Islands of Wisconsin.

The failure of ordination to detect the single gradient in models that represent a wide range of composition (little overlap in species populations) was largely attributed to zeros in the data. Prairie and boreal forest stands often contain few species. As a consequence, there is a high percentage of zeros in these data. In such cases, it seems reasonable to predict that ordination is probably of limited value for natural data.

### *Specific comments on current uses of ordination*

The occurrence of stands falling close together on an axis, yet separated by considerable ecological distance, has sometimes been used to indicate the need for further axis extraction (Bray and Curtis 1957, Beals 1960, Buell et al. 1966). This useful idea should be regarded with caution since ecological distances may not increase commensurate with changes in composition. Hence, selected stands for further axes may not represent compositional extremes between which others can be usefully ordered.

In the present study a localization of species occurs in all ordination diagrams since similar stands are adjacent. However, stand order for each abstract (ordination) model is different be-

cause compositionally different stands are placed far apart in some models but close together in others. Thus, the localization of species patterns may not mean, as sometimes inferred, that species interrelationships are interpretably illustrated, although localized patterns are, of course, expected when this is so.

Recent studies have shown that essentially similar ordinations of natural data are obtained when different importance measures are used (Gimingham et al. 1966, Bannister 1966, Goff and Cottam 1967). However, a readily interpretable presentation of species relations cannot be implied from the consistency of the results alone because bell-shaped curves and zero values may distort all ordination patterns compared.

#### *Suggestions for future research*

Zero values seem to significantly influence ordination patterns. The effects of semi-quantitative data on other forms of matrix construction and ordination should receive attention since they bid fair to obscure the display of vegetational patterns. Theoretical population models seem to offer a useful preliminary method of evaluating ordination procedures.

Orloci (1968a) has proposed that a statistical decision might be used to determine whether quantitative or presence/absence data should be used in ordination. An alternative solution to the analysis of heterogeneous data is to classify and then ordinate within segments of the data (Greig-Smith 1964). In this report one further alternative is suggested. It may obviate the need for a decision between quantitative and qualitative data in some cases. All data are used after adjustments for species absence have been made. Based on the associations of species the transformation may retrieve, for ordination, information otherwise masked by the zeros in a set of data. Such a transformation has, in this study, concentrated ecological information into a few ordination axes, and hence simplified the interpretation of species patterns. Practical difficulties are expected in its application to natural data, and alternative methods of zero-value adjustment can probably be found. Nevertheless the subject warrants attention because most data are semi-quantitative in some degree.

#### ACKNOWLEDGMENTS

I gratefully acknowledge the help of Drs. R. L. Dix, J. K. Jeglum, and C. F. Wehrhahn for their constructive comments on the procedures and manuscript. I especially thank Dr. B. H. Walker for first drawing my attention to the zero-value problem. I also thank Mr. R. H. Gontter, Mr. R. A. Howard, and Dr. C. F. Wehrhahn for writing the requisite computer programs.

#### LITERATURE CITED

- Austin, M. P. 1968. An ordination study of a Chalk Grassland community. *J. Ecol.* 56: 739-757.
- Austin, M. P., and P. Greig-Smith. 1968. The application of quantitative methods to vegetation survey. II. Some methodological problems of data from rain forest. *J. Ecol.* 56: 827-844.
- Bakuzis, E. V., and H. L. Hansen. 1960. Use of ecographs in analysing species—environmental relationships in forest communities. *Minn. Forest Notes*, 91. 2 p.
- 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.
- Beals, E. W. 1960. Forest bird communities in the Apostle Islands of Wisconsin. *The Wilson Bull.* 72: 156-181.
- Beals, E. W., and G. Cottam. 1960. The forest vegetation of the Apostle Islands, Wisconsin. *Ecology* 41: 743-751.
- Bray, J. R., and J. T. Curtis. 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27: 325-349.
- Buell, M. F., A. N. Langford, D. W. Davidson, and L. F. Ohmann. 1966. The upland forest continuum in Northern New Jersey. *Ecology* 47: 416-432.
- Dagnelie, P. 1960. Contribution à l'étude des communautés végétales par l'analyse factorielle. *Bull. Serv. Carte Phytogéogr., Ser. B*, 5: 7-71, 93-195.
- Dice, L. R. 1945. Measures of the amount of ecologic association between species. *Ecology* 26: 297-302.
- Dix, R. L., and F. E. Smeins. 1967. The prairie, meadows, and marsh vegetation of Nelson County, North Dakota. *Can. J. Bot.* 45: 21-58.
- Gimingham, C. H., N. M. Pritchard, and R. M. Cormack. 1966. Interpretation of a vegetational mosaic on limestone in the island of Gotland. *J. Ecol.* 54: 481-502.
- Gittins, R. 1965a. Multivariate approaches to a limestone grassland community. I. A stand ordination. *J. Ecol.* 53: 385-401.
- . 1965b. Multivariate approaches to a limestone grassland community. II. A direct species ordination. *J. Ecol.* 53: 403-409.
- Goff, F. G., and G. Cottam. 1967. Gradient analysis: the use of species and synthetic indices. *Ecology* 48: 793-806.
- Greig-Smith, P. 1964. *Quantitative plant ecology*. 2nd ed. Butterworth and Co., Toronto. 256 p.
- Lambert, J. M., and M. B. Dale. 1964. The use of statistics in phytosociology, p. 55-99. *In* Advances in ecological research. Academic Press Inc., London.
- Loucks, O. L. 1962. Ordinating forest communities by means of environmental scalars and phytosociological indices. *Ecol. Monogr.* 32: 137-166.
- Maycock, P. F., and J. T. Curtis. 1960. The phytosociology of boreal conifer-hardwood forests of the Great Lakes Region. *Ecol. Monogr.* 30: 1-35.
- McIntosh, R. P. 1967. The continuum concept of vegetation. *Bot. Rev.* 33: 130-187.
- Newsome, R. D., and R. L. Dix. 1968. The forests of the Cypress Hills, Alberta and Saskatchewan, Canada. *Amer. Midland Natur.* 80: 118-185.
- Orloci, L. 1966. Geometric models in ecology. I. The theory and application of some ordination methods. *J. Ecol.* 54: 193-215.

- . 1968a. Information analysis in phytosociology: partition, classification and prediction. *J. Theoret. Biol.* 20: 271–284.
- . 1968b. Definitions of structure in multivariate phytosociological samples. *Vegetatio* 15: 281–291.
- Sokal, R. R., and P. H. Sneath. 1963. Principles of numerical taxonomy. W. H. Freeman and Co., San Francisco, Calif. 359 p.
- Swan, J. M. A., R. L. Dix, and C. F. Wehrhahn. 1969. An ordination technique based on the best possible stand-defined axes and its application to vegetational analysis. *Ecology* 50: 206–212.
- Swan, J. M. A., and R. L. Dix. 1966. The phytosociological structure of upland forest at Candle Lake, Saskatchewan. *J. Ecol.* 54: 13–40.
- Waring, R. H., and J. Major. 1964. Some vegetation of the California Coastal redwood region in relation to gradients of moisture, nutrients, light and temperature. *Ecol. Monogr.* 34: 167–215.
- Whittaker, R. H. 1967. Gradient analysis of vegetation. *Biol. Rev.* 42: 207–264.
- Williams, W. T., and M. B. Dale. 1962. Partition correlation matrices for heterogeneous quantitative data. *Nature* 196: 602.